Hyperspectral Band Selection via Adaptive Subspace Partition Strategy

Qi Wang, *Senior Member, IEEE*, Qiang Li, and Xuelong Li, *Fellow, IEEE*

***Abstract*—Band selection is considered as a direct and effective method to reduce redundancy, which is to select some informative and distinctive bands from the original hyperspectral image cube. Recently, many clustering-based band selection methods have been proposed, but most of them only take into account redundancy between bands, neglecting the amount of information in the subset of selected bands. Furthermore, these algorithms never consider the hyperspectral bands as ordered. Based on these two facts, we propose a novel approach for hyperspectral band selection via an adaptive subspace partition strategy (ASPS). The main contributions are as follows: 1) the ASPS is adopted to partition the hyperspectral image cube into multiple subcubes by maximizing the ratio of interclass distance to intraclass distance; 2) unlike previous methods, we estimate the band noise and select the band containing minimum noise (high-quality band) in each subcube to represent the whole subcube; and 3) adaptive subspace partition is viewed as a general framework and thus forms the variant version. Experimental results on three public datasets show that the proposed method achieves satisfactory results in both accuracy and efficiency than some state-of-the-art algorithms.**

***Index Terms*—Adaptive subspace partition, band selection, hyperspectral image, noise estimation.**

I. Introduction

HYPERSPECTRAL images acquire information about objects of interest through large and narrow electromagnetic bands. Compared with an RGB image, these bands can provide more abundant spectral and image information, which can better describe the spectral characteristics of the object and improve the detection and recognition ability [1]. Therefore, it is widely used in various research fields, for instance, vegetation area estimation [2], environmental protection [3], and mineral exploration [4]. Although these bands present more information for relevant image processing, they also bring some obstacles, such as high redundant information and high computational complexity, which is not conducive to subsequent image analysis. Hence, it is extremely necessary to reduce lots of redundant information and time complexity.

Manuscript received May 23, 2019; revised July 29, 2019; accepted September 11, 2019. This work was supported by the National Natural Science Foundation of China under Grants U1864204 and 61773316 and in part by the Project of Special Zone for National Defense Science and Technology Innovation. *(Corresponding author: Xuelong Li.)*

The authors are with the School of Computer Science and the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: crabwq@gmail.com; liqmges@gmail.com; xuelong_li@nwpu.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2019.2941454

In general, there are two kinds of approaches, feature extraction [5], [6] and feature selection (band selection) [7]–[13], to reduce dimensionality for a hyperspectral image. These two types of methods select or extract data from all hyperspectral bands to represent the whole spectral cube, and the results are approximately equal to those of all bands. For the former, it maps high-dimensional spatial data into low-dimensional space based on certain criteria and extract a new feature subset to represent the original hyperspectral data. The typical methods include principal component analysis [14], linear discriminant analysis (LDA) [15], [16], maximum noise fraction [6], and independent component analysis [5]. However, through spatial transformation, the physical meaning of the original hyperspectral data is changed, and some key information is lost. For the latter, this type of methods selects a distinctive and representative subset from the original hyperspectral data without loss of physical meaning and information, which preserves the inherent properties of the hyperspectral data. In this article, we turn our attention to band selection.

Hyperspectral band selection is one of the key steps in the preprocessing of some algorithms, which has attracted many scholars' attention in recent years. Unlike the traditional feature selection methods, the principle of band selection is to make the band subset have low correlation and large amount of information, which can not only greatly reduce the data dimension of hyperspectral image cube, but also retain more complete and useful information. According to whether the sample is labeled or not, the existing band selection methods can be classified into two main parts: supervised [17]–[19] and unsupervised [20]–[24]. Supervised methods are to select the optimal bands by training and learning based on labeled samples. Although this type of methods can achieve better results, the training process is relatively complex, and labeled samples are often not available in practice. Unsupervised methods only need to select a subset from the hyperspectral bands through some evaluation criterion function without using labeled samples. The main criteria are as follows: information divergence [25], maximum ellipsoid volume [26], Euclidean distance [27], [28], etc. Overall, unsupervised methods are preferable.

Among these band selection methods, the clustering-based methods [8], [11], [24], [25], [29] appeal to our attention. These methods can obtain satisfying results, but they always have two inherent shortcomings in the clustering process. On one hand, most of them only take into account correlation between bands, neglecting the amount of information in the subset of selected bands [11], [24], [25]. Obviously, these do not satisfy

the principle of band selection. On the other hand, according to the characteristics of a hyperspectral image cube, we find that the bands are arranged in order. In addition, for a band, it has stronger correlation with the adjacent bands within a certain range and low correlation with farther bands. Therefore, it can be concluded that discontinuous bands with different ranges of wavelengths cannot be grouped into one cluster for band selection. For instance, in the Indian Pines dataset (see Section IV-A for a dataset detail), it totally has 200 bands. If we use a clustering algorithm to select a band subset, the band of index 1 and the band of index 200 cannot be classified into one class.

Motivated by the above descriptions, we believe that the hyperspectral image cube is divided into several subcubes, which can effectively avoid the above dilemmas. Moreover, the informative band should be selected as the representative band in each subcube. Based on these facts, in this article, we propose a novel approach for hyperspectral band selection via an adaptive subspace partition strategy (ASPS). The main contributions are as follows.

1) An ASPS is proposed to search unrelated hyperspectral subcubes. Based on the fact that adjacent bands have high redundancy, the ordered hyperspectral bands are partitioned into multiple subcubes by the clustering algorithm, which can effectively avoid obtaining a subset with high correlation.
2) An efficient band noise estimation method is used to select a band with minimum noise in each subcube as the representative band. The band image is divided into small blocks of the same size, and the noise level of the band image is estimated through the distribution of local variance to measure the contribution of the band, thereby realizing the selection of high-quality band images.
3) Through adaptive subspace partition, the obtained subcube can be regarded as a general framework, which means that the representative band in each subcube can be selected by other criteria. A variant method of selecting bands by using information entropy is presented in this article. Extensive comparative experiments verify that the variant method has outstanding performance and is robust enough across different datasets.

The remainder of this article is organized as follows. In Section II, several band selection methods are introduced, including clustering-based methods and ranking-based methods. Section III presents the detailed descriptions of the proposed method about the subspace partition strategy and band noise estimation. After that, the evaluation experiments, which are conducted on different public datasets, are described in Section IV. Finally, conclusions are given in Section V.

II. Related Work

The unsupervised band selection methods mainly include clustering-based and ranking-based methods. The methods based on clustering divide different properties into multiple groups so that there is low correlation between groups. The methods based on ranking use some criteria to assess the importance of each hyperspectral band and select top-rank bands as a subset. Next, the two kinds of typical methods will be described in detail.

A. Clustering-Based Methods

Clustering-based methods [25], [30] regard each band as a point and use Euclidean distance or other criteria to judge the similarity between any two bands. Then, some clustering algorithms are used to make the bands with high correlation into one class. Generally, the nearest band to the center is chosen as the representative band in each group. In the following, several typical methods will be briefly described.

In [25], the hierarchical clustering structure is adopted to make the ratio of interclass variance and intraclass maximum to cluster hyperspectral bands. Unlike other clustering methods, this method uses mutual information or Kullback–Leibler to measure the similarity among bands. Although this method considers both information and correlation, the results are not good, especially for noisy datasets. Wang *et al.* [30] propose an optimal clustering framework for hyperspectral band selection. This method is to partition hyperspectral bands into multiple groups by dynamic programming, and one band is selected in each group individually by applying the score that is computed by fast density-peak-based clustering or other approaches. Compared with most algorithms, this partition way requires appropriate computational time to perform dynamic programming when more bands are chosen. In addition, the recommended bands are calculated by the variance-based band-power ratio, but it is clearly not appropriate because the currently selected subset may not contain a subset of the previous bands.

B. Ranking-Based Methods

Ranking-based methods [31], [32] mainly have two steps. The first step is to quantify the importance of each band by function, such as information entropy [33]. The second step is to rank the quantized result of each band to obtain a subset according to the number of selected bands. The typical methods are information divergence [31], constrained band selection [31], and maximum-variance principal component analysis (MVPCA) [14]. Here, we will discuss three representative methods.

Information divergence in these methods is often used to evaluate image quality according to Gaussian probability distribution. It tries to select the most non-Gaussian bands as a subset, which can effectively reflect the amount of information contained in spectral bands. Constrained band selection sets the objective function to minimize the constrained correlation among bands. Specifically, the method changes a band to be target signature vector while considering other bands as unknown signature vectors. Then, the concept of constrained energy minimization is adopted to linearly constrain this band and minimize the correlation with other bands. MVPCA builds a covariance matrix in accordance with the hyperspectral data. The eigenvalues and eigenvectors are solved for the matrix and construct a loading factor matrix. Finally, the variances of all

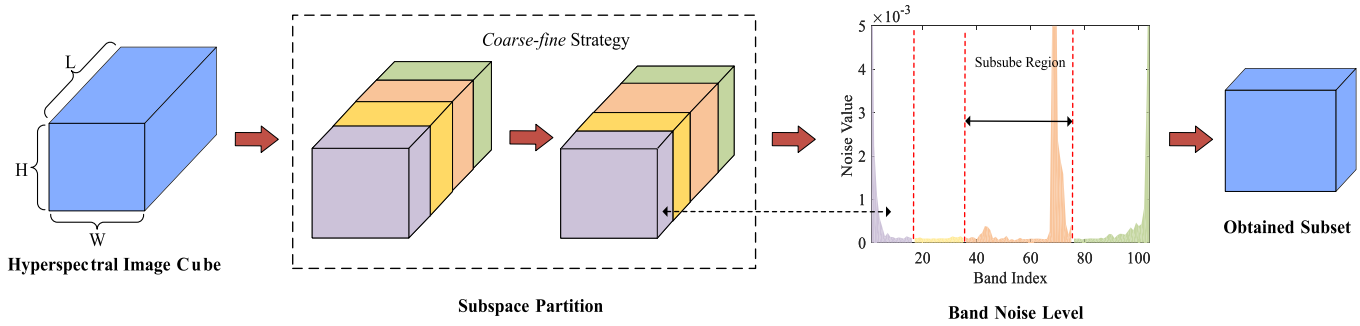

Fig. 1. Framework of ASPS_MN. Based on the fact that hyperspectral bands are ordered, the hyperspectral image cube in space is first partitioned to subcubes with the same size, and the clustering algorithm is adopted to accurately get subcubes with different sizes. Then, the band with minimum noise in subcube is selected as band subset.

bands are arranged in descending order. All of these methods have a common shortcoming that the selected bands may have highly correlated. Moreover, as for MVPCA, it has changed original features of the hyperspectral data.

As introduced in Section I, the selected bands should have lower correlation and more discriminative information. But each of the above methods only meets one of the principles for band selection, so we combine the advantages of the two kinds of methods to achieve band selection by dividing cube into multiple subcubes and selecting the representative bands in subcubes, It not only reduces the data dimension of hyperspectral image cube, but also retains more complete and useful information.

III. Proposed Method

In this section, we detail the proposed framework, namely ASPS_MN, whose flowchart is shown in Fig. 1. The basic idea is to treat hyperspectral bands as ordered and then adaptively divide the bands with similar spectral characteristics into one subcube. The band with minimum estimated noise is selected as the representative in each subcube. A detailed description of the two parts is given in the following sections.

A. Adaptive Subspace Partition Strategy

Considering that the correlation of adjacent bands is higher than that of nonadjacent bands, the ASPS is adopted to divide the hyperspectral image cube. As described in Section I, we can use the clustering algorithm to realize spatial partition. But if we directly employ the clustering algorithm to implement, it will result in greater time complexity. In order to partition the hyperspectral image cube faster, the proposed method uses the *coarse-fine* strategy, which mainly consists of the following two steps.

1) Coarse Subspace Partition: Let $X \in R^{W\times H\times L}$ denote the hyperspectral image cube, where L is the number of total spectral bands, and the width and height of each band are W and H, respectively. One of the purposes of band selection is to reduce computational time. Hence, in order to implement the clustering algorithm faster, the hyperspectral image cube is divided into finite subcubes by equal width according to the number of selected bands. Here, the number of bands in each subcube P_i is defined as

$$Z = \frac{L}{K} \tag{1}$$

where K is the number of selected bands. This way is an initial partition and can obtain subcubes $P_i \in R^{W\times H\times Z}$.

2) Fine Subspace Partition: To accurately represent spectral bands belonging to each subcube P_i, we propose a fine subspace partition method to acquire new subcubes. In detail, the matrix of each spatial band is stretched to a one-dimensional vector; thus, we get

$$X = [x_1, x_2, \ldots, x_L] \tag{2}$$

where $x_i \in R^{WH\times 1}$, and x_i is the stretched band vector of the ith band. Following the stretched band vector, a similarity matrix between the ith band and jth band is constructed by the Euclidean distance, which can be done by

$$D_{ij} = \sqrt{\sum_{n=1}^{WH} (x_{ni} - x_{nj})^2}. \tag{3}$$

In clustering algorithms, a common method is that the intra-class and interclass are applied to analyze problems. Specifically, the final clustering results are obtained by maximizing the ratio of interclass distance to intraclass distance. For the hyperspectral image cube that has been partitioned, we further utilize this idea to accurately partition this cube. Since the correlation between two subcubes far apart is small, we only consider the relationship between two adjacent subcubes (P_i and P_{i+1}). Therefore, a general form of the objective function is given as

$$\arg\max_{t} \frac{D_{\text{inter}}}{D_{\text{intra}}} \tag{4}$$

where D_{inter} and D_{intra} are interclass distance and intraclass distance, respectively, and t denotes the partition point. In the computation of interclass distance, we choose the maximum distance as a criterion to measure the relationship between the two classes. It is defined as

$$D_{\text{inter}} = \max |D_{ij}| \tag{5}$$

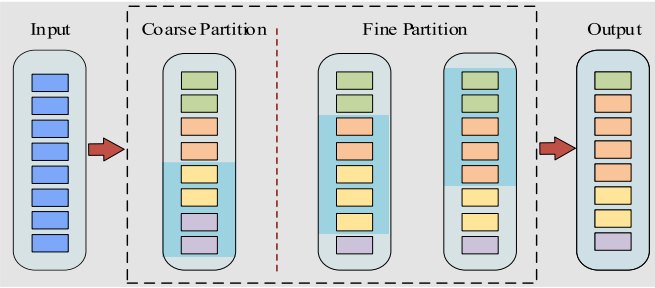

Fig. 2. Example to partition the hyperspectral image cube (eight bands) into four subcubes. Within the dashed line, the dark blue region indicates that only these adjacent bands are considered to update the current partitioned point.

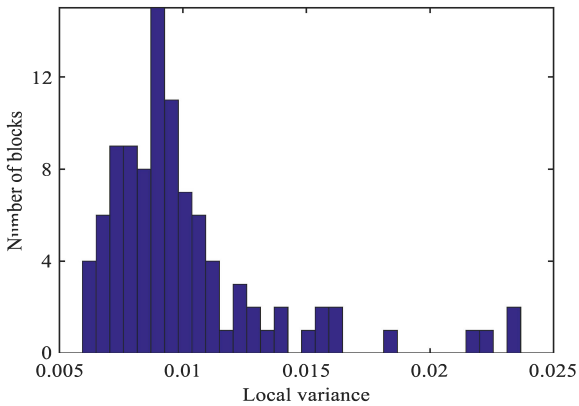

Fig. 3. Noise estimation for one band on Indian Pines dataset.

where $1 \leq i < j \leq 2Z$. For intraclass distance D_{intra}, it contains two parts: U_1 of the subcube P_i and U_2 of the subcube P_{i+1}, which can be denoted as the sum of intraclass distance of two subcubes, i.e.,

$$D_{\text{intra}} = U_1 + U_2. \tag{6}$$

Specifically, two parts are written as

$$U_1 = \frac{1}{t(t-1)} \sum_{i=1}^{t} \sum_{j=1}^{t} D_{ij} \tag{7}$$

$$U_2 = \frac{1}{(2Z-t-1)(2Z-t-2)} \sum_{i=t+1}^{2Z} \sum_{j=t+1}^{2Z} D_{ij}. \tag{8}$$

The first accurate partition point is obtained by the above equations instead of the original partition point. Accordingly, we use the previous partition point to update the initial point t in the same way to get the final segmentation point (see Fig. 2). This partition method can make the correlation low between obtained subcubes, which can effectively avoid the selection of redundant bands. It satisfies one of the principles of band selection.

B. Band Noise Estimation

Most algorithms select informative bands by information divergence, but when there are noises in the band image, reliable results cannot be obtained so as to affect the subsequent analysis. Inspired by this analysis, we can choose the band image with minimum noise in each subcube. The method for calculating noise is usually to decompose the image into two parts: clean image and noise image. However, this method takes a lot of time to conduct, which does not meet the purpose of low time complexity for band selection. Therefore, this article avoids direct image decomposition and uses local variance to estimate the noise level of each band image in accordance with Gao's method [34]. Then, the representative band with minimum noise in each subcube is chosen as the selected band.

1) Local Variance of the Hyperspectral Image: Following the approach of Coakley and Bretherton [35], each band image is first divided into small blocks of the size $B \times B$ pixels, such as $3 \times 3, \ldots,$ or 10×10 pixels. For the hyperspectral band image that cannot be completely partitioned, it is a common practice to remove some columns or rows of the band image until all the blocks of the same size are segmented, which does not influence the experimental results [36].

Since some blocks have similar spectral characteristics, in order to accelerate the processing speed of the algorithm, we randomly select M blocks from the original blocks. Although this treatment trades off speed for performance, overall, it has a little impact on the performance of our method. For a band, the local mean and local variance of each block are calculated according to

$$\text{LM} = \frac{1}{B^2} \sum_{i=1}^{B^2} S_i \tag{9}$$

$$\text{LV} = \frac{1}{(B^2-1)} \sum_{i=1}^{B^2} (S_i - \text{LM})^2 \tag{10}$$

where B^2 is the number of pixels, and S_i is the value of the ith pixel in the block.

2) Weight Computation: After calculating the variance, the corresponding values of M blocks are acquired for each band image. Next, the noise of band image is estimated from the variance values for the small blocks. Because the variance of some blocks is much greater than that of others (homogeneous blocks have small variance values, inhomogeneous blocks have large variance values), if these values are directly handled by calculating average value, this will result in a poor estimation. To deal with this trouble, the difference between the maximum and minimum values of M blocks for each band image is divided into k bins with equal width (see Fig. 3), i.e.,

$$k = (\text{maxV} - \text{minV})/\alpha \tag{11}$$

where maxV is the maximum variance, minV denotes the minimum variance, and α stands for the partition granularity ($\alpha = 3$ in our article). This approach is very much like acquiring mode. Then, M blocks are allotted into these bins in accordance with the values of local variance, and the number of blocks in each bin is counted. The bin with the largest number of blocks corresponds to the estimated noise of band image. All band images are treated in the same way, and the noise of each band

Algorithm 1: Framework of ASPS_MN.

Input: Hyperspectral image cube $X \in R^{W\times H\times L}$, the number of selected bands K, the size of block $B \times B$, the number of block M.
Output: Selected subset Y.
1: Coarsely divide hyperspectral image cube to obtain subcubes $P_i \in R^{W\times H\times Z}$.
2: Stretch each spatial band image into one-dimensional vector x_i.
3: Construct similarity matrix D_{ij} by the Euclidean distance.
4: Apply (4) to fine partition subspace.
5: Divide band image into the blocks of equal size with $B \times B$ pixels, and calculate local mean and local variance of each band image.
6: Estimate noise level N by the distribution of the minimum and maximum variance.
7: Select band with minimum noise in each subcube P_i to be subset Y.
8: **return** Y.

image is obtained, i.e.,

$$N = [N_1, N_2, \ldots, N_L]. \tag{12}$$

The band containing minimum noise is regarded as the most representative band in each subcube. Finally, a desired subset Y containing K bands is selected as the target bands.

For more details about the framework, the procedures are summarized in Algorithm 1.

C. General Framework

As shown in the above method, the band selection is performed in two parts: adaptive subspace partition and band noise estimation, which means that the representative band in each subcube can be selected in other ways to replace band noise estimation. From this point of view, the ASPS can be considered as a general framework for band selection. We can obtain the variant versions of ASPS_MN. For instance, the information entropy is one of the ways to measure the bands in subcube, where the name of the variant method is ASPS_IE. As for ASPS_IE, the information entropy is defined as follows:

$$H_i = -\sum_{z\in\Omega} p(z)\log p(z) \tag{13}$$

where Ω is the grayscale color space, and $p(z)$ represents the probability of the occurrence of event z appearing in the image, which can be obtained from the grayscale histogram. The information entropy is used to measure the average amount of information contained in an image. Generally speaking, the larger the information entropy, the richer the image information. Thus, based on this point, the band with the largest amount of information in subcube is to be the desired band.

D. Time Complexity Analysis

The implementation of the two versions of the proposed method mainly consists of two steps: adaptive subspace partition and representative band selection in each subcube. In the following, we analyze the time complexity of the proposed method for each step.

1) Subspace Partition: In this step, the subspace partition mainly consists of two parts: coarse subspace partition and fine subspace partition. Specifically, for the coarse subspace partition section, the hyperspectral image cube in space is divided into several subcubes $P_i \in R^{W\times H\times Z}$ by equal width, which takes almost no computing time. For fine subspace partition, a similarity matrix is first constructed by Euclidean distance. Equation (4) is then applied to perform fine subspace partition. The time complexity of a fine partition operation costs $O(L^2WH)$. For these two parts, the total time complexity of adaptive subspace partition takes $O(L^2WH)$.

2) Representative Band Selection: In order to obtain the desired band in each subcube, we propose two approaches. One is to select band with minimum noise. We conduct the mean and variance of M blocks picked from the original blocks. Additionally, the noise of the band image is estimated in accordance with the distribution of these variances. The time complexity of this part mainly depends on the size of block $B \times B$ and the number of selected blocks M. Therefore, $O(MB^2)$ time is needed for this part. The other is to select the band with the largest IE, whose time complexity is $O(LWH)$.

For ASPS_MN and ASPS_IE, their total time complexities are $O(L^2WH + MB^2)$ and $O(L^2WH + LWH)$, respectively. It can be seen that the time of both versions is acceptable for most the applications. Since $MB^2 < LWH$, it is apparent that ASPS_MN is the fast version that has much better computation efficiency than ASPS_IE.

IV. Experiment

In this section, we implement extensive experiments to assess the superiority of the proposed algorithm in the context of the hyperspectral image. First, three public hyperspectral image datasets are introduced. Then, we briefly explain the basic principle of the comparison methods. Next, the experimental setup is described from three aspects, including classification setting, number of selected bands, and accuracy measures. Finally, the experimental performance of several methods is conducted in detail to verify the effectiveness and advantages of ASPS_MN and ASPS_IE.

A. Datasets

1) Indian Pines Scene: The Indian Pines scene was collected by the AVIRIS sensor in 1992. AVIRIS captures 224 spectral bands with the size of 145× 145 pixels from 0.4 to 2.5 μm. The band consists of 16 classes of land cover objects of interest. Some bands (104–108, 150–163, and 220) produce large noises because of water absorption, so we discard these bands and finally get 200 bands in the experiment. The image and the corresponding groundtruth are shown in Fig. 4.

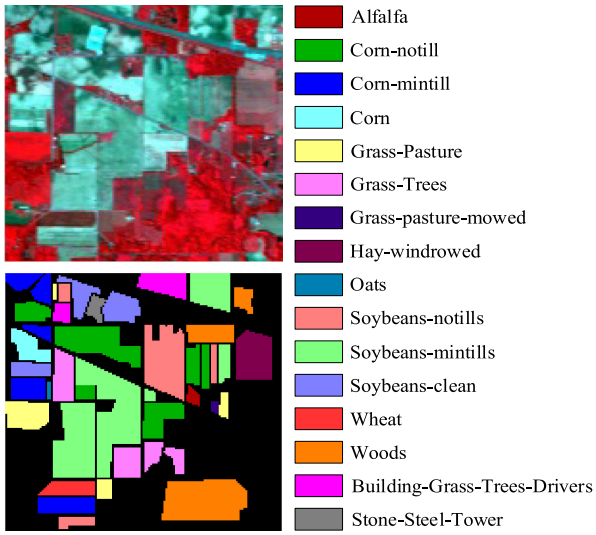

Fig. 4. Color image and groundtruth map on the Indian Pines dataset.

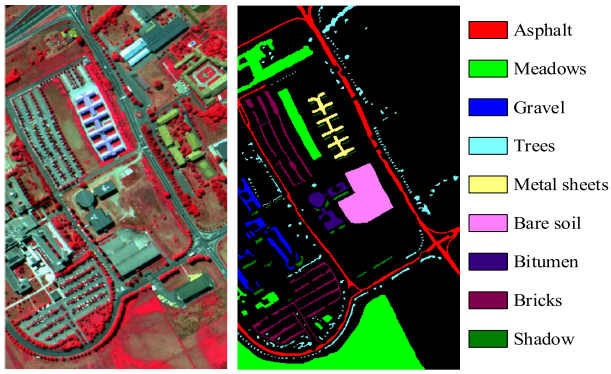

Fig. 5. Color image and groundtruth map on the Pavia University dataset.

2) Pavia University Scene: The Pavia University scene was acquired by the ROSIS sensor over Pavia in 2002. Similar to the Indian Pines dataset, the remaining number of spectral bands is 103 after removing some bands with lower SNR, and the size of each band is 610 × 340 pixels. The groundtruth of this dataset includes nine classes of interest, whose details are described in Fig. 5.

3) Salinas Scene: The Salinas scene was gathered by the AVIRIS sensor over Salinas Valley, CA, USA, in 1998. This dataset is characterized by a high spatial resolution (3.7-m pixels) and has 224 spectral bands with the size of 512 × 217 pixels. There are 16 classes of interest, which is introduced in Fig. 6.

B. Comparison Methods

To investigate the performance of the proposed method (ASPS_MN and ASPS_IE) in each dimension, five state-of-the-art algorithms that perform band selection using no labeled information are compared with the proposed method, where we briefly introduce these competitors in the following.

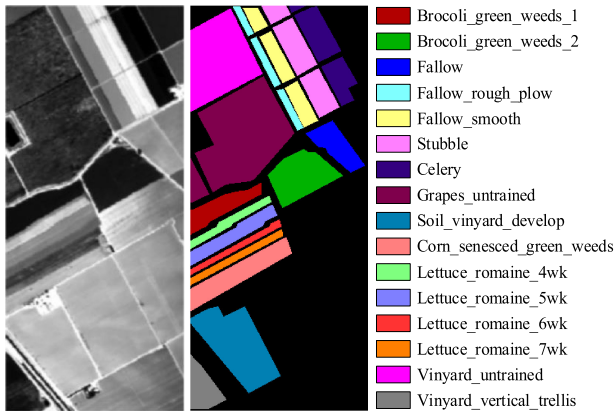

Fig. 6. Grayscale image and groundtruth map on the Salinas dataset.

1) TRC-OC-FDPC [30]: This method uses dynamic programming to divide hyperspectral image cube into multiple subcubes, and one band is selected in each class individually by applying the score that is computed by E-FDPC [11]. To simplify the name of the algorithm, TRC-OC-FDPC is written as TOF in our article.

2) MDSR [36]: Using a multidictionary learning mechanism, each band is expressed by the linear combination of other bands. The first several bands with large weight are chosen by reckoning sparse coefficient.

3) WaLuDi [25]: By constructing the Kullback–Leibler divergence matrix, hierarchical clustering is performed to acquire representative bands in accordance with intraclass and interclass variance.

4) RMBS [37]: A representation matrix is determined by the rank minimization constraint and is utilized to define the affinity matrix to pick out informative bands.

5) UBS [31]: According to the number of selected bands, the hyperspectral image cube is divided into multiple subcubes at equal width, and each segmentation point is viewed as the selected band.

C. Experimental Setup

1) Classification Setting: In the experiment, three typical classifiers, including k-nearest neighborhood (KNN), support vector machine (SVM), and LDA, are employed to assess the classification performance of several band selection algorithms from various dimensions. Next, we will briefly explain the characteristics and parameter setting of the three classifiers. KNN is the simplest classifier in machine learning, which determines the sample category according to the category of K similar training data. In the following experiments, the parameter K is set to 3 for all classification validation experiments about KNN. For the SVM classifier, it is a discriminant classifier defined by classification hyperplane. That is to say, given a set of labeled training samples, the algorithm will output an optimal hyperplane to classify the new samples (test samples). In our experiments, the SVM classifier is conducted with the RBF kernel, and the penalty C and gamma of the three datasets are

TABLE I
OA FOR ANALYZING THE IMPACT OF THE SIZE AND NUMBER OF BLOCKS ON THREE DATASETS

Data set	B×B	M	5	7	10	15	20	25	30	35	40	45	50	55	60
Indian Pines	3×3	10%	64.51	65.25	66.78	68.96	70.37	70.46	69.69	70.21	70.12	70.80	70.12	70.13	70.41
		20%	63.62	66.75	65.51	69.57	69.75	68.25	69.01	71.03	70.12	70.71	70.71	71.49	69.62
		50%	64.34	68.04	66.21	68.38	68.66	68.47	68.57	69.74	70.65	69.98	69.99	70.58	69.88
		100%	65.14	65.25	66.82	67.06	66.51	70.04	69.44	67.92	68.97	69.51	70.34	71.08	69.71
	10×10	10%	62.02	62.82	63.81	68.82	70.98	69.73	70.70	69.29	69.70	68.09	70.05	70.60	71.58
		20%	62.38	62.31	65.61	69.56	69.25	69.66	69.70	70.23	70.85	71.08	71.08	71.46	70.73
		50%	61.10	63.45	64.95	68.94	68.66	68.85	69.40	68.37	69.92	70.51	70.90	70.32	70.86
		100%	61.98	63.31	64.90	69.10	68.69	69.31	70.62	67.16	69.95	70.01	69.84	69.71	69.51
Pavia University	3×3	10%	83.19	84.56	85.95	85.66	85.70	87.01	87.42	86.35	86.74	86.82	87.03	86.79	86.63
		20%	83.64	83.55	85.36	86.55	85.34	86.55	85.97	86.96	86.62	86.90	87.69	86.69	86.88
		50%	82.53	84.47	86.01	85.71	85.07	86.57	86.43	86.98	86.86	87.09	86.89	86.62	86.77
		100%	82.87	82.77	85.31	85.73	85.30	86.33	86.63	86.13	86.77	86.41	87.21	87.12	86.32
	10×10	10%	82.65	83.60	85.66	86.83	85.54	86.03	87.35	86.42	86.92	87.66	86.96	87.13	86.83
		20%	84.30	85.13	84.33	85.53	87.38	86.42	86.55	86.46	86.49	86.49	87.61	86.62	86.95
		50%	82.14	82.58	86.32	87.73	85.60	85.85	86.65	86.21	86.86	87.09	87.75	86.64	86.52
		100%	84.03	83.26	87.07	87.06	86.59	87.01	87.16	86.99	87.27	86.91	87.50	86.99	86.86
Salinas	3×3	10%	84.66	86.39	88.40	89.27	89.48	89.55	89.78	89.40	89.33	89.91	90.01	89.96	89.45
		20%	84.54	88.79	88.38	89.15	89.09	89.38	89.89	89.92	89.54	89.93	89.84	89.78	89.74
		50%	84.43	86.64	87.91	89.25	89.17	89.41	89.34	89.64	89.35	89.49	89.70	89.82	89.60
		100%	84.95	85.75	88.05	89.63	89.34	89.06	89.61	89.69	89.36	89.91	89.67	90.04	89.74
	10×10	10%	83.94	86.18	88.35	89.72	89.37	89.15	89.51	89.45	89.34	89.40	89.77	89.28	89.77
		20%	84.66	85.92	87.88	89.43	89.20	89.44	89.80	89.92	89.28	89.88	89.67	89.69	89.43
		50%	83.80	87.51	88.16	88.92	89.43	89.11	89.14	89.54	89.73	89.88	89.92	89.71	89.49
		100%	84.36	86.64	87.71	89.52	89.23	89.13	89.77	89.47	89.51	89.41	89.63	89.84	89.71

initialized to 1×10^4 and 0.5, respectively. The principle of LDA is to project the labeled data (points) into the lower dimensional space through the projection method so that the projected points can be classified. Note that this classifier is parameter free.

Additionally, since we mention that these classifiers are supervised, 10% samples from each class based on selected bands are randomly chosen as the training set; the remaining 90% samples are used for test. Moreover, in order to reduce the influence of random selection of 10% samples, the algorithm runs ten times to obtain the average results. Figs. 4–6 display groundtruth maps containing each class for three datasets.

2) Number of Selected Bands: Because the desired number of bands that should be selected is unknown in practice for three public hyperspectral image datasets, we implement experiments in the range of 5–60 bands to explain the influence of different numbers of bands on classification accuracy.

3) Accuracy Measures: Three methods of accuracy measure are conducted to analyze the precision of the classified pixels. They are overall accuracy (OA), average overall accuracy (AOA), and kappa coefficient (Kappa).

D. Results

In this article, in order to investigate the effectiveness and advantages of the proposed method, we provide detailed analyses on three datasets from four aspects, including number and size of blocks, classification performance, and computational time.

1) Number and Size of Blocks Analysis: In the proposed band selection method, each band image needs to be divided into small blocks $B \times B$ with the same size during estimating the band noise. Moreover, in order to accelerate the execution of the algorithm, we randomly pick M blocks to perform our method. In this section, for illustrating the impact of the number and size of blocks on classification accuracy, we set the size of blocks to 3×3 and 10×10 pixels and choose 10%, 20%, 50%, and 100% blocks from all the blocks for the experiments. Table I provides the classification results for KNN classifiers on three datasets by selecting different numbers of selected bands (the range of 5–60). From this table, one can observe that the OA hardly changes through setting differently the number and size of blocks at every selected band, and only fluctuations occur while selecting a small number of bands. Overall, these two hyperparameters do not affect our algorithm very much. Therefore, from the experiments on three hyperspectral image datasets, some crucial results can be summarized. Whatever the size and number of the block are, the proposed method can obtain stable classification performance by several classifiers across different datasets on the whole. Consequently, during estimating the band noise, in order to facilitate the evaluation of the following comparative experiments, for Pavia University and Salinas datasets, we empirically set the size of block to be 10×10 pixels and randomly pick 10% blocks, and for the Indian Pines dataset, these parameters are set to 3×3 and 10%.

2) Classification Performance Comparison: To fully show convictive results of the proposed method compared with five state-of-the-art algorithms, three classifiers are employed to analyze the hyperspectral image by using three accuracy evaluation criteria. Moreover, all bands are also taken into account for performance comparison.

For the Indian Pines dataset, Table II reveals the AOA and Kappa by selecting different numbers of bands (the range of 5–60) for the classification results. To clearly show the

TABLE II
AOA AND STANDARD DEVIATION OF DIFFERENT METHODS ON THREE DATASETS

Data set	Classifier (Measure)	TOF	MDSR	WaLuDi	RMBS	UBS	ASPS_MN	ASPS_IE
Indian Pines	KNN(AOA)	**68.01±0.13**	56.00±1.03	66.85±0.10	67.04±0.15	67.10±0.18	**68.88±0.11**	67.22±0.10
	KNN(Kappa)	**65.70±0.14**	49.68±1.19	62.19±0.17	62.25±0.22	61.45±0.16	**64.34±0.13**	63.45±0.13
	SVM(AOA)	78.31±0.18	67.30±0.54	77.58±0.18	78.13±0.46	78.34±0.16	**78.87±0.29**	**78.34±0.14**
	SVM(Kappa)	75.00±0.18	61.72±0.67	74.24±0.20	74.07±0.56	75.00±0.19	**75.13±0.36**	**75.09±0.17**
	LDA(AOA)	**69.67±0.09**	59.58±0.39	69.02±0.08	69.43±0.24	69.93±0.10	**69.46±0.32**	68.66±0.21
	LDA(Kappa)	**66.55±0.09**	53.09±0.46	65.49±0.10	66.00±0.27	65.36±0.12	**66.93±0.37**	65.01±0.24
Pavia University	KNN(AOA)	84.58±0.04	84.36±0.29	84.97±0.06	83.99±0.35	85.08±0.05	**86.23±0.13**	**86.17±0.04**
	KNN(Kappa)	79.04±0.06	78.72±0.40	79.61±0.09	78.20±0.49	79.72±0.06	**81.34±0.18**	**81.26±0.07**
	SVM(AOA)	90.64±0.02	90.82±0.31	**91.36±0.07**	90.46±0.28	91.06±0.02	**91.44±0.25**	90.54±0.10
	SVM(Kappa)	87.39±0.03	87.64±0.46	**88.41±0.10**	87.09±0.42	87.97±0.03	**88.90±0.36**	87.21±0.15
	LDA(AOA)	80.74±0.08	80.17±0.28	80.94±0.05	81.04±0.22	**81.85±0.05**	**81.45±0.14**	80.97±0.08
	LDA(Kappa)	74.70±0.10	73.37±0.38	75.01±0.06	74.39±0.35	**75.60±0.06**	**75.55±0.19**	74.32±0.11
Salinas	KNN(AOA)	88.05±0.05	85.76±0.73	88.04±0.01	88.14±0.17	86.94±0.05	**88.27±0.13**	**88.16±0.03**
	KNN(Kappa)	86.48±0.05	83.17±0.82	85.03±0.01	84.53±0.19	85.38±0.06	**87.62±0.14**	**87.56±0.03**
	SVM(AOA)	90.92±0.03	89.76±0.72	91.03±0.01	91.08±0.28	90.25±0.03	**90.92±0.13**	**91.27±0.04**
	SVM(Kappa)	90.25±0.04	86.82±0.84	90.21±0.01	89.96±0.33	90.17±0.04	**90.67±0.15**	**91.11±0.05**
	LDA(AOA)	87.03±0.03	85.45±0.67	87.30±0.02	87.91±0.28	87.18±0.06	**87.45±0.05**	**87.67±0.01**
	LDA(Kappa)	85.94±0.03	83.78±0.77	86.08±0.02	85.86±0.33	86.02±0.07	**86.14±0.06**	**86.24±0.01**

The numbers in bold represent the two better classification performances.

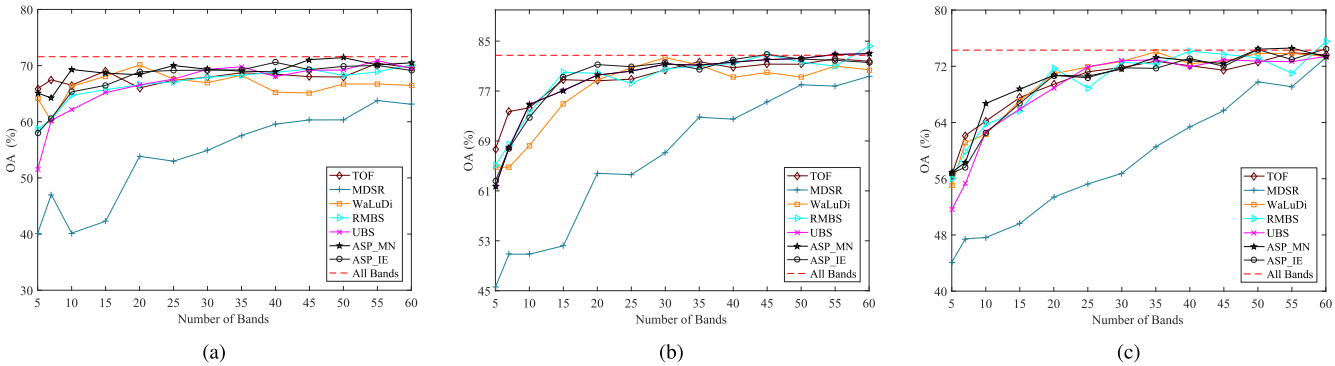

Fig. 7. OA for three classifiers by selecting different numbers of bands on the Indian Pines dataset. (a) OA by KNN. (b) OA by SVM. (c) OA by LDA.

influence of different numbers of selected bands, we provide OA results for every five bands (see Fig. 7). From this figure, it is explicitly observed that the proposed method can outperform the other methods at some selected bands, especially for ASPS_MN. More specifically, among all methods, the results selected by MDSR are significantly lower than other methods across different datasets. When the number of bands reaches 20, ASPS_MN can exceed stable performance. With respect to LDA and SVM classifiers, the difference in results is not obvious, and other competitors achieve satisfying results, except for MDSR. In addition, as can be seen from Table II, the superiority of ASPS_MN is more evident with AOA and Kappa, and ASPS_IE obtains comparable results, which is exactly consistent with the above analyses. To sum up, the proposed method is better than some compared methods in terms of three accuracy measures.

For the Pavia University dataset, it is clear from Fig. 8 and Table II that our method exhibits the better results in OA compared with most algorithms, and the classification accuracy of AOA and Kappa obtained by ASPS_MN and ASPS_IE also achieves relatively satisfying results. In Fig. 8(a), when the number of selected bands is small, the accuracy of some algorithms is unstable, particularly for RMBS and MDSR. But when the number of selected bands goes beyond 15, all competitors work very well. Meanwhile, our method still performs quite well, and its accuracy exceeds the accuracy of all bands at certain locations. When it comes to other classifiers, compared with the Indian Pines dataset, the advantage of the two versions of the proposed method is not greatly obvious. Overall, ASPS_MN also displays the best performance, but the results of ASPS_IE are not very ideal. As for other competitors, the OA grows steadily by increasing the number of selected bands.

For the Salinas dataset, the two versions of the proposed method also outperform the other methods in each dimension for three classifiers (see Fig. 9 and Table II). More specifically, in Fig. 9, our method produces higher OA compared with the other band selection methods for the same numbers of selected bands

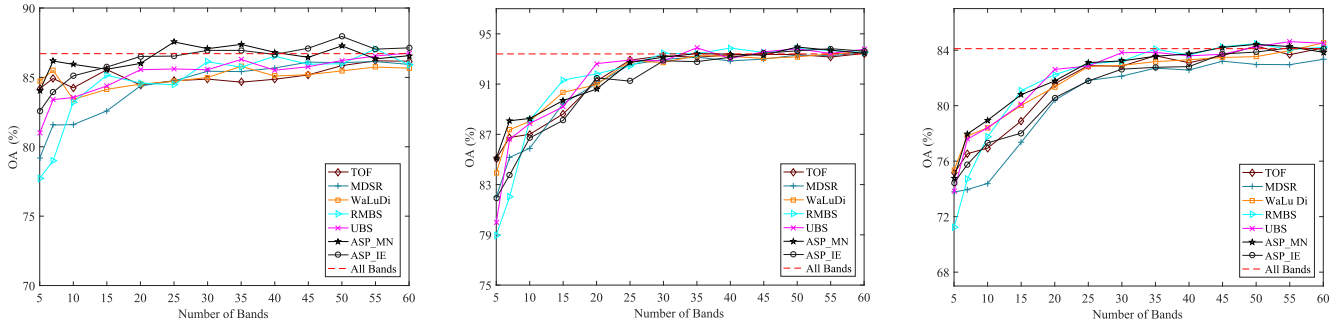

Fig. 8. OA for three classifiers by selecting different numbers of bands on the Pavia University dataset. (a) OA by KNN. (b) OA by SVM. (c) OA by LDA.

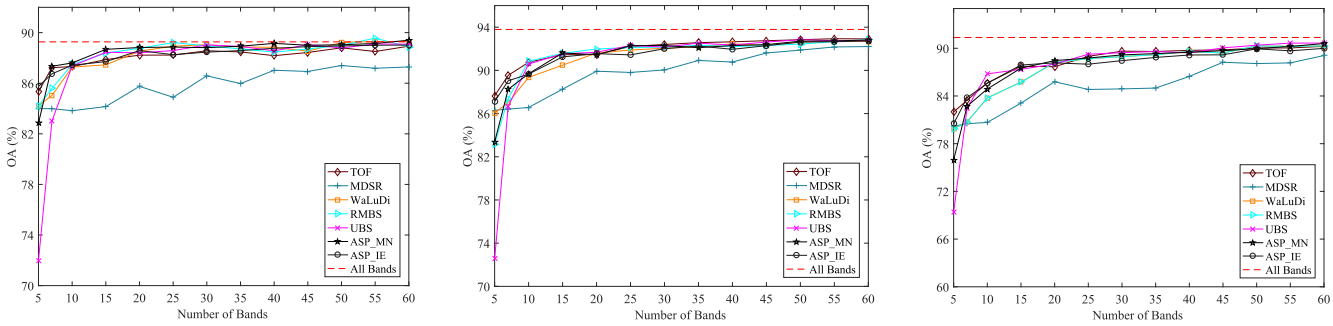

Fig. 9. OA for three classifiers by selecting different numbers of bands on the Salinas dataset. (a) OA by KNN. (b) OA by SVM. (c) OA by LDA.

TABLE III
PROCESSING TIME OF DIFFERENT BAND SELECTION METHODS TO SELECT PROPER NUMBER OF BANDS ON THREE DATASETS

Data set (Number of bands)	TOF	MDSR	WaLuDi	RMBS	UBS	ASPS_MN (10%)	ASPS_MN (100%)	ASPS_IE
Indian Pines (15 bands)	0.649s	0.205s	7.507s	43.618s	0.009s	0.915s	6.785s	0.409s
Pavia University (10 bands)	0.741s	0.208s	26.775s	200.396s	0.009s	0.895s	3.440s	1.321s
Salinas (15 bands)	1.356s	0.313s	40.357s	265.555s	0.003s	1.128s	5.884s	1.403s

when the KNN classifier is used. When the number of selected bands is small, the result curves of MDSR, UBS, and RMBS fluctuate greatly, especially for MDSR. While the SVM classifier is employed, the difference except MDSR is not so obvious. All the other algorithms have approximately the same results at every selected band. As for the LDA classifier, the results are similar to those obtained by the SVM classifier. However, when picking five bands, our method shows best classification performance. Furthermore, there is still a certain gap compared with the classification results of all bands at each location.

By conducting extensive experiments on three hyperspectral image datasets, most competitors exhibit bad accuracy on the Indian Pines dataset, but classify the pixels more precisely on other datasets. This is mainly due to the selection of noisy bands on the Indian Pines dataset, even if some bands are removed. As for our proposed method, it has absolute superiority on the Indian Pines dataset. It indicates that our method can deal with noisy dataset well. To sum up, the two versions of the proposed method overall obtain excellent and stable performance across different datasets, which verifies that our approach is robust enough for band selection.

3) Computational Time Comparison: In order to verify the effectiveness of the proposed method, the computational time of all competitors is conducted on three datasets. For fair comparison, all methods are performed in MATLAB 2016a with PC workstation (Intel Core i5-3470 CPU processor and 16-GB RAM). Table III shows the execution time by different band selection methods to select proper number of bands on different datasets. From the results, the processing time needed by WaLuDi and RMBS is far greater than that of other

algorithms, particularly for RMBS, which is caused by implementing Kullback–Leibler divergence and singular value decomposition, respectively. Additionally, compared with other methods, the computational cost of UBS is much less but has poor classification results. This is mainly because it is the simplest method of comparison; only selecting points with equal width are viewed as the selected bands. Although MDSR has a relatively faster processing time, it does not classify pixels very well for classification tasks. As for the version of the proposed method, namely, ASPS_IE, it is relatively time-consuming but more accurate than TOF, and thus, a little more time cost is acceptable. For some methods with higher accuracy, ASPS_MN not only can execute faster, but also achieve excellent classification performance than others. Moreover, we also show the influence of selecting 10% and 100% blocks on execution time. It can be seen that the number of blocks has a great impact on the algorithm execution time. To sum up, the proposed method has a great advantage in time efficiency while satisfying better performance in classification accuracy.

V. Conclusion

Many clustering algorithms only consider the redundancy between bands in the calculation. In addition, it is not found that these bands of the hyperspectral image cube are ordered. Therefore, in this article, we develop a novel approach via adaptive subspace partition to select some informative and distinctive bands from the original hyperspectral bands.

The ordered hyperspectral cube in space is partitioned into multiple subcubes by the clustering method to generate a general framework. It effectively avoids selecting a subset with high correlation. To obtain the informative bands, we estimate band noise in each subcube in terms of the local variance to select a band with minimum noise as a representative band. The above two contributions not only greatly reduce the data dimension of the hyperspectral image cube, but also retain more complete and useful information, which conforms to the principles of band selection. The last contribution is to produce the variant versions of the proposed method. This article only shows one variant method, ASPS_IE, to measure the band with the amount of information in subcube. Extensive experimental results on three public hyperspectral image datasets demonstrate that the two versions of the proposed method exhibit more robust and effective performance than other competitors.

In future work, we will extend the proposed method in two directions. One is to obtain the recommended number of selected bands. The second is to speed up the execution of the algorithm.

References

[1] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end two-dimensional CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019.

[2] D. Haboudane, J. R. Miller, E. Pattey, P. J. Zarco-Tejada, and I. B. Strachan, "Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture," *Remote Sens. Environ.*, vol. 90, no. 3, pp. 337–352, 2004.

[3] B. Gao *et al.*, "Additional sampling layout optimization method for environmental quality grade classifications of farmland soil," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5350–5358, Dec. 2017.

[4] M. H. Zadeh, M. H. Tangestani, F. V. Roldan, and I. Yusta, "Mineral exploration and alteration zone mapping using mixture tuned matched filtering approach on ASTER data at the central part of Dehaj-Sarduiyeh copper belt, SE Iran," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 284–289, Jan. 2014.

[5] J. Wang and C.-I Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.

[6] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.

[7] W. Zhang, X. Li, Y. Dou, and L. Zhao, "A geometry-based band selection approach for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4318–4333, Aug. 2018.

[8] Y. Yuan, J. Lin, and Q. Wang, "Dual-clustering-based hyperspectral band selection by contextual analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1431–1445, Mar. 2016.

[9] J. Feng, L. Jiao, T. Sun, H. Liu, and X. Zhang, "Multiple kernel learning based on discriminative kernel clustering for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6516–6530, Nov. 2016.

[10] R. Huang and M. He, "Band selection based on feature weighting for classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 156–159, Apr. 2005.

[11] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 88–102, Jan. 2016.

[12] H. Zhai, H. Zhang, and L. Zhang, "Laplacian-regularized low-rank subspace clustering for hyperspectral image band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1723–1740, Mar. 2019.

[13] P. Ramzi, F. Samadzadegan, and P. Reinartz, "Classification of hyperspectral data using an AdaBoostSVM technique applied on band clusters," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2066–2079, Jun. 2014.

[14] C.-I Chang, Q. Du, T.-L. Sun, and M. L. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.

[15] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[16] H. Zhai, H. Zhang, and L. Zhang, "Total variation regularized collaborative representation clustering with a locally adaptive dictionary for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 166–180, Jan. 2019.

[17] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 138–142, Jan. 2011.

[18] M. Riedmann and E. J. Milton, "Supervised band selection for optimal use of data from airborne hyperspectral sensors," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2003, pp. 1770–1772.

[19] F. Jie, L. Jiao, L. Fang, S. Tao, and X. Zhang, "Mutual-information-based semi-supervised hyperspectral band selection with high discrimination, high information, and low redundancy," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2956–2969, May 2015.

[20] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 911–923, Feb. 2019.

[21] Y. Yuan, X. Zheng, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 51–64, Jan. 2017.

[22] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.

[23] S. Jia, Z. Ji, Y. Qian, and L. Shen, "Unsupervised band selection for hyperspectral imagery classification without manual band removal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 531–543, Apr. 2012.

[24] Y. Qian, F. Yao, and S. Jia, "Band selection for hyperspectral imagery using affinity propagation," *IET Comput. Vis.*, vol. 3, no. 4, pp. 213–222, Dec. 2009.
[25] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.
[26] X. Geng, K. Sun, L. Ji, and Y. Zhao, "A fast volume-gradient-based band selection method for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7111–7119, Nov. 2014.
[27] N. Keshava, "Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1552–1565, Jul. 2004.
[28] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.
[29] Q. Li, Q. Wang, and X. Li, "An efficient clustering method for hyperspectral optimal band selection via shared nearest neighbor," *Remote Sens.*, vol. 11, no. 3, 2019, Art. no. 350.
[30] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.
[31] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.
[32] K. Sun, X. Geng, and L. Ji, "Exemplar component analysis: A fast band selection method for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 998–1002, May 2015.
[33] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 522–526, Oct. 2006.
[34] B. Gao, "An operational method for estimating signal to noise ratios from data acquired with imaging spectrometers," *Remote Sens. Environ.*, vol. 43, no. 1, pp. 23–33, 1993.
[35] J. A. Coakley and F. P. Bretherton, "Cloud cover from high-resolution scanner data: Detecting and allowing for partially filled fields of view," *J. Geophys. Res., Oceans*, vol. 87, no. C7, pp. 4917–4932, 1982.
[36] F. Li, P. Zhang, and H. Lu, "Unsupervised band selection of hyperspectral images via multi-dictionary sparse representation," *IEEE Access*, vol. 6, pp. 71632–71643, 2018.
[37] G. Zhu, Y. Huang, S. Li, J. Tang, and D. Liang, "Hyperspectral band selection via rank minimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2320–2324, Dec. 2017.

Qi Wang (M'15–SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively.

He is currently a Professor with the School of Computer Science and the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.

Qiang Li received the B.E. degree in measurement and control technology and instrument from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2015, and the M.S. degree in communication and transportation engineering from Chang'an University, Xi'an, in 2018. He is currently working toward the Ph.D. degree with the School of Computer Science and the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an.

His research interests include hyperspectral image processing and computer vision.

Xuelong Li (M'02–SM'07–F'12) is a Full Professor with the School of Computer Science and the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China.