

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Title: Superpixel-based Unsupervised Band Selection for Classification of Hyperspectral Images

This paper appears in: IEEE Transactions on Geoscience and Remote Sensing

Date of Publication: 2018

Author(s): Chen Yang, Lorenzo Bruzzone, Haishi Zhao Yulei Tan and Renchu Guan

Volume:

Page (s):

DOI:

# Superpixel-based Unsupervised Band Selection for Classification of Hyperspectral Images

Chen Yang, *Member, IEEE*, Lorenzo Bruzzone, *Fellow, IEEE*, Haishi Zhao  
Yulei Tan and Renchu Guan, *Member, IEEE*

**Abstract**—This paper presents an unsupervised approach to band selection in hyperspectral images that considers both spectral and spatial information in data dimensionality reduction. The approach exploits the concepts of superpixel and chunklets for identifying the spectral channels most suitable to be used in classification for discriminating land-cover classes. The segmented superpixels can be regarded as many small spectral homogeneous and spatial neighboring pixels chunklets. Based on the observation that the superpixels chunklets achieve high homogeneity and consistency within land-cover classes, a series of band criteria are identified by learning the optimal band transformation that results in low within class variability and high total variability. Then the learned band criteria, which are called band measures, are given in input to an efficient clustering algorithm, i.e., the affinity propagation, for selecting highly separable bands with low redundancy. The effectiveness of proposed approach was assessed on three hyperspectral datasets. The results point out the advantages of the proposed methods over five state-of-the-art unsupervised methods.

**Index Terms**—Affinity propagation, band criterion, band selection, superpixel, hyperspectral images.

## I. INTRODUCTION

HIGH dimensionality is the most significant characteristic of hyperspectral imagery (HSI). HSI can contain hundreds of narrow and continuous spectral channels with a very high spectral resolution spanning from visible to infrared region of the electromagnetic spectrum [1]. Therefore, they have very good potentialities for the identification of detailed land-cover classes. However, such adjacent spectral channels results in high redundancy and correlation among bands. Meanwhile, the

collection of a sufficient number of training samples is difficult compared with the large number of spectral bands and often leads to ill-conditioned problems (i.e., Hughes phenomenon, also known as curse of dimensionality) [2],[3]. This increases both the generalization error of automatic classifiers and the required computational complexity. Moreover, many hyperspectral systems acquire images that also have high spatial resolution, thus increasing the intraclass variability and decreasing interclass variability. This may lead to a reduction of statistical separation among different land-cover classes in the spectral domain [4]. The high intraclass variability makes the selection of representative training samples more difficult in many real applications.

To overcome the above issues, dimensionality reduction techniques are usually applied to HSI for removing redundancy, while retaining discriminant information in the feature space. Feature selection (FS) (also called band selection (BS) when only spectral bands are analyzed) techniques are commonly used for dimensionality reduction. Their goal is to identify a subset of the original spectral channels as small as possible, which allows optimization of the classification results [5],[6].

According to availability of prior information, BS techniques can be divided into three types: supervised, semi-supervised and unsupervised. Supervised methods [7]-[9] select a discriminative subset of bands by training models with a large number of representative labeled samples. However, land-cover class labels are difficult and expensive to obtain in many HSI applications. As a result, semi-supervised methods [10]-[12], which require a relatively small amounts of labeled data, have been recently applied to HSI analysis. Unsupervised methods [13]-[15] select features without any training sample or prior information. Despite their use is very attractive, it is challenging to define effective unsupervised BS methods. In this paper we focus on unsupervised band selection techniques.

Traditional unsupervised BS methods often select spectral bands in HSI according to a give criterion that ranks the spectral channels by considering either their amount of information or their degree of correlation (or both). Information divergence (ID) [16],[17] is a commonly used band prioritization criterion for measuring non-Gaussianity of bands [18],[19]. Pixels variability can be used for measuring the amount of information contained in spectral channels. Chang et al. [20] proposed a maximum-variance principal component analysis (MVPCA) criterion for band prioritization. In these methods, often the spectral correlation between bands is not considered, leading to the selection of bands containing similar information. In [20],

This work was supported in part by the National Natural Science Foundation of China under Grants NO. 61572228, by the National Key Basic Research Program of China under Grant No. 2015CB453000, by the Science-Technology Development Plan Project of Jilin Province of China under Grant No. 20160101247JC. (*Corresponding author: Renchu Guan and Chen Yang*)

C. Yang is with the College of Earth Sciences, Jilin University, Changchun, 130061, China and the Lab of Moon and Deepspace Exploration, National Astronomical Observatories, Chinese Academy of Sciences, Beijing, 100012, China (e-mail: yangc616@jlu.edu.cn).

L. Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38050 Trento, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it).

Y. L. Tan is with the College of Earth Sciences, Jilin University, Changchun, 130061, China.

H. S. Zhao and R. C. Guan are with the College of Computer Science and Technology, Jilin University, Changchun, 130012, China (e-mail: guanrenchu@jlu.edu.cn).

the divergence measure (DM) as a band decorrelation criterion was applied to remove redundant bands. Clustering-based BS methods focus on partitioning bands into similar groups based on their similarity or correlation, thus redundancy reduction can be obtained by selecting the centers of clusters as the representative bands. However, classical clustering methods, such as k-means, are sensitive to the initial conditions, i.e., the selected bands change depending on the particular initialization thus involving unstable results.

Recently, some efficient clustering algorithms were used for HSI band selection. The most representative algorithm is the affinity propagation (AP) [21]. AP has shown its advantages over traditional methods in HSI band selection [22]. Su *et al.* [23] proposed an adaptive AP (AAP) algorithm for selecting a fixed number of bands. For evaluating the image quality, a complex wavelet structural similarity (CW-SSIM) index has been developed as the similarity criterion of AP [24]. The above AP-based methods can select relatively low redundant and stable bands with little processing time. However, since they assume the same prior suitability for all the bands, they do not consider the discriminative capability of each band. Jiao, *et al.* [25] proposed a semi-supervised AP method with improved band preference based on Entropy and mutual information (MI) for selecting highly discriminative and low redundant bands subset. Moreover, a few density-based ranking-clustering band selection methods have been presented, such as enhanced fast density-peak-based clustering (E-FDPC) [26] and band selection using density based spatial clustering for applications with noise (DBSCAN) [27]. These methods select informative bands by computing ranking score of each band and removing redundancy in the clustering process. Despite their advantages, unsupervised BS methods usually focus on selecting diverse and representative bands for classification. The lack of labeled data does not make it possible to evaluate the discriminative capability of bands with respect to actual land-cover classes.

HSI represents real land surfaces that typically extend for few pixels in the image. Thus, nearby pixels in HSI have a high probability to belong to the same class, i.e., there is a high spatial correlation. The spatial information defined as spatial homogeneity has been exploited to improve the performance of BS. Zhu *et al.* [28] proposed a global unsupervised method based on structurally meaningful information for the measuring both band information and independence. In [29]-[31], hypergraph construction was introduced for semi-supervised band selection. Cao *et al.* [32] proposed a supervised band selection algorithm based on the local spatial information and a wrapper method. However, these related studies are mainly concentrated on the use of pixel-centered local or global fixed neighborhood systems, thus they cannot reflect complex spatial structure of HSI. Moreover, most of them relies on labeled data.

Adjustment learning (AL) [33] is a new learning paradigm introduced for image retrieval. In the AL scheme, data points can be identified as small group sets, i.e., "chunklets" with equivalence constraints that are known to originate from the same class (but the label is unknown). A non-iterative and efficient metric learning method, called relevant component analysis (RCA), based on the AL was proposed in [34],[35]. The goal of RCA is to find a transformation that amplifies relevant variability and suppresses irrelevant variability with

chunklets. The usefulness of RCA was demonstrated in metric learning [36] and representation learning [37]. In previous work, we introduced RCA into hyperspectral image BS and constructed a feature metric (FM) for assessing discrimination capability and spectral correlation associated with bands [38]. Chunklet groups spectrally close and spatially nearby pixels together with positive constraints. But the limitation of this method is that spectral and spatial information of HSI is not sufficiently utilized. Moreover, prior knowledge is required.

Superpixel [39] is identified as a coherent and local grouping of pixels in regions, which can be generated by over-segmenting of an image using a superpixel segmentation algorithm [40]. It can preserve most of the spatial structure information and align better with object edges than fixed image patches; moreover, it can reduce the complexity of subsequent image analysis tasks [41],[42]. Various superpixel segmentation algorithms have been proposed in the literatures [43]-[46]. These algorithms generate superpixels by varying parameters for capturing diverse visual contents of images. The size and shape of each superpixel can be adaptively adjusted according to local structures in most superpixel algorithms. Superpixels has been applied successfully to HSI classification [47]-[49], target [50], [51] and endmember detection [52] and image decomposition [53]. It has seldom used for dimensionality reduction in HSI [54],[55], especially for band selection.

In this paper, an unsupervised band selection approach is proposed that consider both spectral and spatial information by using superpixels for an effective HSI classification. Firstly, many small spectral homogeneous and spatial neighboring pixels chunklets, termed as superpixel chunklets (SC), are constructed by using a superpixel segmentation algorithm. Based on the observation that generated SC achieve high homogeneity and consistency within land-cover classes, two bands criteria (BCs) are defined by learning the optimal transformation through the RCA. Then the learned BC are modeled to band measures (BM) and given as input to the AP algorithm. Finally, AP offers an efficient search strategy for selecting highly-discriminative and low-redundant band subsets. Experimental results obtained on three HSI data sets show the effectiveness of proposed band selection approach.

The rest of this paper is organized as follows. The proposed superpixel-based BS approach is presented in Section II, where the learning BC with SC, the introduced BM and the selection process in AP are given. Section III reports the design of the experimental phase and experimental results of the proposed BS approach and five unsupervised methods; then it provides a discussion of performance comparison with different BCs. Section V draws the conclusion of this work.

## II. PROPOSED SUPERPIXEL-BASED BAND SELECTION

The proposed BS approach consists of three main parts: (i) a new spectral-spatial modeling of the information i.e., SC, is defined by a graph-based superpixels segmentation algorithm; (ii) two categories of BCs are defined by exploiting the optimal RCA with constructed SC; and (iii) a BM is developed based on the learned BC and given as input to AP for representative band subset selection. A detailed description of the three parts is given in the following subsections.

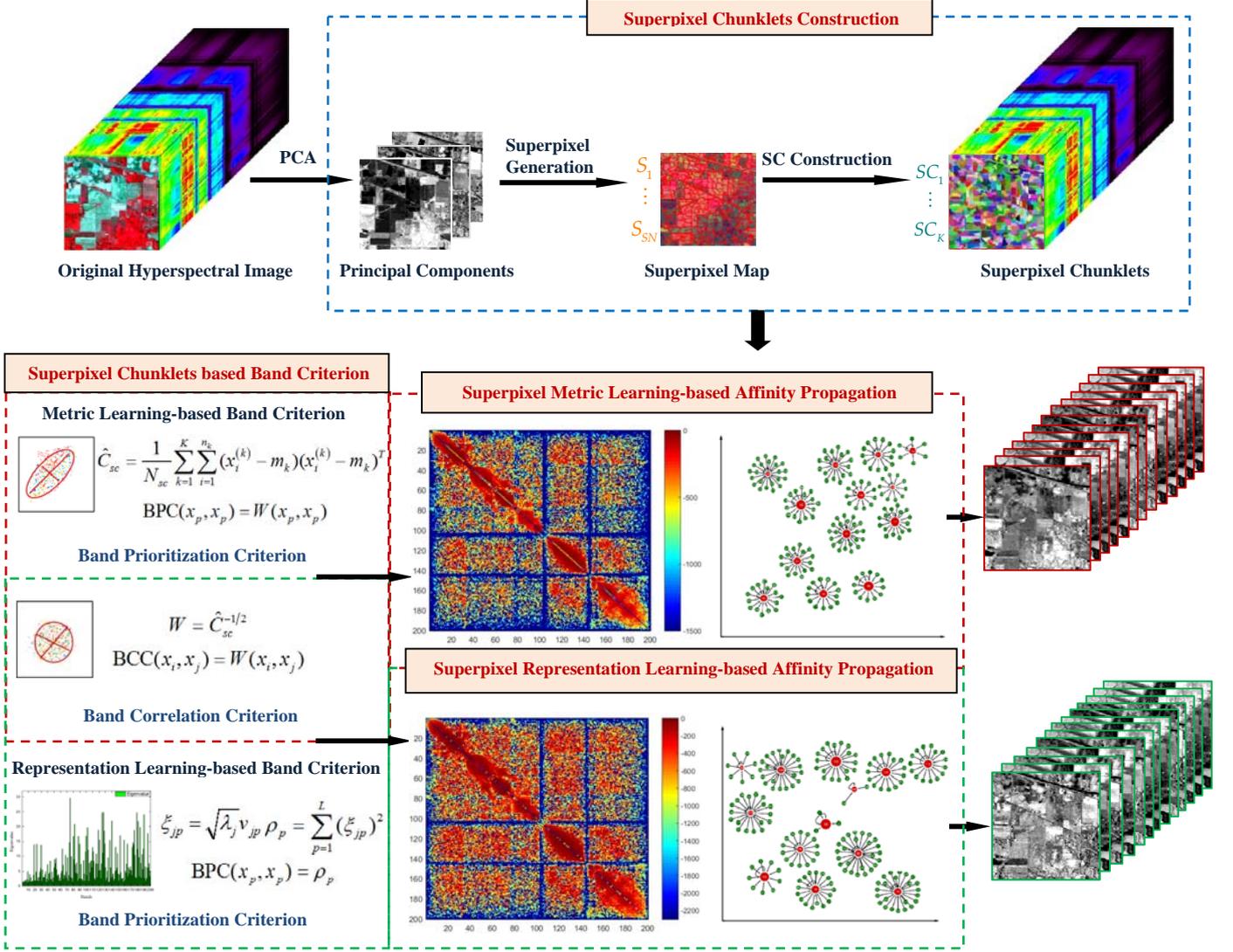


Fig. 1. Flowchart of the proposed superpixel-based band selection approach.

### A. Superpixel Chunklets

Traditional spatial information is extracted with fixed size or neighborhood windows. For HSI, a fixed window typically cannot meet the classification requirements due to the different sizes and complex shapes of ground objects throughout image. Therefore, in this paper, a new spectral-spatial structure, i.e., SC, is presented to improve the performance of band selection.

Superpixel algorithms can generally be categorized as graph-based and clustering-based on the basis of segmentation model adopted [41]. Two representative works of the two categories are entropy rate superpixels (ERS) [42] and simple linear iterative clustering (SLIC) [44], which are the most widely used superpixel segmentation algorithms in preprocessing or postprocessing of HSI. In ERS, the image is mapped to a graph in which each pixel is considered as a vertex and the pairwise similarities are defined as edge weights. The problem of superpixel segmentation is defined by a graph topology maximization with an objective function designed for getting compact and homogeneous superpixels with similar sizes. The SLIC can be considered as an adaptation of  $k$ -means for

superpixel generation. It can obtain regular superpixels which adhere well and efficiently to boundaries. However, the SLIC cannot capture global image properties and results in undersegmentation errors [44] as they only consider the similarity of pixels.

Three desired properties of SC are considered in BS methods: 1) Each SC should maintain homogeneity, i.e., overlap with only one ground object; 2) the shapes of SC should be relatively regular and compact; and 3) a number of SC as small as possible should be obtained satisfying the above desired properties. Therefore, ERS is adopted for SC construction in this paper. Let  $X = \{x_1, x_2, \dots, x_L\} \subset \mathbb{R}^{N \times L}$  be an HSI data set, where  $x_i = \{x_{i1}, x_{i2}, \dots, x_{iN}\}$ ,  $N$  is the number of pixel vectors present in the HSI,  $L$  is the number of spectral bands, and  $x_i$  represents the  $i$ th band spectral responses. Let  $SC = \{SC_1, SC_2, \dots, SC_K\}$  be the SC set,  $SC_k = \{x_1^{(k)}, x_2^{(k)}, \dots, x_{n_k}^{(k)}\}$  ( $k=1, 2, \dots, K$ ), where  $K$  is the number of SC,  $x_i^{(k)}$  is the  $i$ th pixel vector in the  $k$ th SC and  $n_k$  is the number of pixel vectors in the  $k$ th SC. The construction process of SC is described in Algorithm 1.

---

**Algorithm 1** Superpixel Chunklets Construction

---

**Input:**

Hyperspectral data set  $X = \{x_1, x_2, \dots, x_L\}$

**Output:**

Superpixel chunklets set  $SC_k = \{x_1^{(k)}, x_2^{(k)}, \dots, x_{n_k}^{(k)}\}$

**Procedure:**

1. *Segmentation base image generation.*
  - Principal component analysis (PCA) is applied to HSI to obtain the first three principal components;
  - Combination the first three principal components is used for generating the segmentation base image.
2. *Superpixel map and superpixels generation.*
  - Superpixel map generation: A graph is constructed on the segmentation base image. An objective function is defined consisting of the entropy rate of a random walk on the graph and a balancing term for obtaining compact, homogeneous, and balanced superpixels. By using an iterative greedy algorithm for optimizing the objective function, a subset of edges in the graph is selected resulting in a superpixel map  $S$  [42] in which the base image is oversegmented into different homogeneous regions.
  - Superpixels generation: The partition of map  $S$  refers to a division of all pixels into  $SN$  disjoint superpixels, i.e.,  $S = \{S_1, S_2, \dots, S_{SN}\}$ ,  $S_i \cap S_j = \emptyset$ ,  $i \neq j$ .
3. *Construction superpixel chunklets on HSI.*
  - The spatial location and the number of pixels  $n_s$  within each superpixel  $S_i$  can be obtained according to the superpixel map  $S$ ;
  - $SN$  non-overlapping pixel vectors can be extracted from the HSI. Then SC can be constructed by combining these extracted  $SN$  pixel vectors, i.e.,  $SC = S$ ,  $SC_k = S_k$ ,  $K = SN$ ,  $n_k = n_s$ .

---

**B. Superpixel Chunklets based Band Criterion**

The performance of BS methods critically depends on the function of the adopted band criterion. The pixels in selected bands should have high relevance within the same classes, i.e., the variability within same classes should be low. The selected bands should have high discrimination capability and low redundancy. In RCA, the feature space used for data representation can be projected and changed by a global linear transformation which assigns large weights to relevant dimensions and low weights to irrelevant dimensions. The goal of RCA is distance metric learning (DML) and dimensionality reduction when data are embedded in a high dimensional space. In this paper, two categories of BCs based on the solutions of RCA are considered. Each BC contains two criteria: a band prioritization criterion (BPC) and a band correlation criterion (BCC). The former models the discrimination score of a band by a quantitative measure; the latter analyzes spectral correlation between bands to select a low redundancy bands subset.

**1) Metric learning-based band criterion (ML-BC)**

The basic idea of RCA is to identify and down-scale global

unwanted variability within data for learning a distance metric, which can be regarded as learning a whitening transformation matrix to assign lower weights to directions having large variability. The directions of large variability are mainly due to within class changes and are “irrelevant” for the classification task. Accordingly, the first band criterion is derived from metric learning (ML). As mentioned before, also in [38] ML was used. However, the relevant dimensions in the previous work are estimated by getting within-chunklet covariance with positive constraints, in which only spectrally close and spatially nearby pixels are considered without the class concept. For HSI, pixels in each SC can be regarded as taken from a local spatial region having homogeneous spectral characteristics and corresponding to a given class. In order to model the variability in homogeneity of land covers and define metric learning-based bands criterion (ML-BC), the within-SC covariance matrix can be written with SC as follows:

$$\hat{C}_{sc} = \frac{1}{N_{sc}} \sum_{k=1}^K \sum_{i=1}^{n_k} (x_i^{(k)} - m_k)(x_i^{(k)} - m_k)^T \quad (1)$$

where  $m_k$  is mean vector of the  $k$ -th SC,  $N_{sc} = \sum_{k=1}^K |SC_k|$  and  $|\cdot|$  denotes the cardinality of a set.

The object of the whitening transformation is to estimate and minimize within-SC covariance of data for making the learned metric more effective for classification in new feature space. Therefore, the whitening transformation matrix associated with the within-SC covariance matrix can be computed as follows:

$$W^{ML} = \hat{C}_{sc}^{-1/2} \quad (2)$$

In terms of band selection, the selected bands should also have low variability in SC. Accordingly, the BPC of each band  $x_p$  based on metric learning (BPC<sup>ML</sup>) is defined as

$$\text{BPC}^{ML}(x_p, x_p) = W^{ML}(x_p, x_p) \quad (3)$$
$$p = 1, 2, \dots, L$$

The BCC between two different band  $x_i$  and  $x_j$  is given by:

$$\text{BCC}^{ML}(x_i, x_j) = W^{ML}(x_i, x_j) \quad (4)$$
$$i, j = 1, 2, \dots, L; i \neq j$$

**2) Representation learning-based band criterion (RL-BC)**

In the above metric learning process, the obtained whitening transformation essentially re-scales variability in all directions to equalize them. In this way, dimensions with small total variability may cause instability. In order to take this issue into account, the other BC needs to be designed for considering the within-SC variability and the total variability simultaneously. To achieve this goal, these relevant dimensions are estimated using SC, and emphasized with small within-SC covariance but large total-SC covariance. In this paper, the total-SC covariance matrix is defined as follows:

$$\hat{C}_{isc} = \frac{1}{N_{sc}} \sum_{i=1}^{N_{sc}} (x_i - m)(x_i - m)^T \quad (5)$$

where  $m$  is the mean vector of all the SC.

In this section, the solution of the RCA is derived from representation learning (RL) with Fisher theory (or criterion) [56], [57], in which an optimal transformation matrix is learned by simultaneously maximizing total covariance and minimizing

within-class covariance. Traditionally Fisher theory is computed from fully labeled training data, and falls within supervised learning. Therefore, we extend traditionally Fisher criterion to the unsupervised pattern with SC in the form of the same information theoretic criterion, and employ the following optimization function:

$$J(W^{\text{RL}}) = \arg \max_{W^{\text{RL}}} \frac{|(W^{\text{RL}})^{\text{T}} \hat{C}_{\text{isc}} W^{\text{RL}}|}{|(W^{\text{RL}})^{\text{T}} \hat{C}_{\text{sc}} W^{\text{RL}}|}. \quad (6)$$

in which  $W^{\text{RL}}$  is the optimal solution that can be obtained by eigenanalysis of  $\hat{C}_{\text{isc}} \cdot \hat{C}_{\text{sc}}^{-1}$ , and  $w_j, j=1,2,\dots,L$  is its  $j$ th column vector, i.e., a generalized eigenvector corresponding to the  $j$ th largest eigenvalue  $\lambda_j$  in the following generalized eigenvalue problem:

$$\hat{C}_{\text{isc}} w_j = \lambda_j \hat{C}_{\text{sc}} w_j, \quad (7)$$

in which  $\hat{C}_{\text{sc}}$  is nonsingular.

In [58] and [59], eigenanalysis has been suggested and used for band selection, but the relationship between eigenvalues and eigenvectors was not fully exploited. Tu *et al.* [60] defined the discriminant power of each band by maximizing the Fisher linear discriminant (FLD) function with eigenanalysis-based criteria, i.e., discrimination determined by the magnitude of nonzero eigenvalues and the corresponding eigenvectors (discriminant vectors). One of the best known BP criterion is to use band variances as a measure of priority of each band, which called PCA-based priority score [20], which is formulated as variance-based BP criterion [61].

In our proposed BP criterion, a loading-factors matrix is constructed with the eigenvalues and eigenvectors acquired by eigen (spectral) decomposition of  $W^{\text{RL}}$ . Then the loading factors are used to define the discrimination capability of each band. Let  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_L)$  be real and nonzero eigenvalues obtained by (8) or (9), and  $V_j = (v_{j1}, \dots, v_{jL})^{\text{T}}, j=1,2,\dots,L$ , their corresponding  $L$ -dimensional eigenvectors. Based on factor analysis, a loading factors matrix can be defined as

$$\xi_{jp} = \sqrt{\lambda_j} v_{jp} \quad (8)$$

$$j, p = 1, 2, \dots, L$$

For the  $p$ th band  $x_p$ , the band priority score  $\rho_p$  is defined as

$$\rho_p = \sum_{j=1}^L (\xi_{jp})^2 \quad (9)$$

$$j = 1, 2, \dots, L$$

and representation learning-based BPC ( $\text{BPC}^{\text{RL}}$ ) is formulated as

$$\text{BPC}^{\text{RL}}(x_p, x_p) = \rho_p \quad (10)$$

$$p = 1, 2, \dots, L$$

in which a larger value of  $\rho_p$  implies that the  $p$ th band has better discrimination capability.

Meanwhile, the BCC is also derived from the RL-based

optimization function. For getting the solution and solving the maximization of (6), let  $G$  be the solution matrix for the unconstrained problem that orthogonally diagonalizes both within-SC covariance and total-SC covariance, i.e.,  $G \hat{C}_{\text{sc}} G^{\text{T}} = \Lambda_1$ ,  $G \hat{C}_{\text{isc}} G^{\text{T}} = \Lambda_2$ , where  $\Lambda_1$  and  $\Lambda_2$  are diagonal matrices. In order to enforce the SC constraints, the matrix  $W^{\text{RL}} = \Lambda_1^{-1/2} G = \hat{C}_{\text{sc}}^{-1/2} = W^{\text{ML}}$  is defined as the optimal solution, where  $W^{\text{RL}}$  is the whitening transformation for learning a new data representation (feature space). Therefore, representation learning-based BCC ( $\text{BCC}^{\text{RL}}$ ) between different bands  $x_i$  and  $x_j$  is expressed as:

$$\text{BCC}^{\text{RL}}(x_i, x_j) = W^{\text{RL}}(x_i, x_j) = \text{BCC}^{\text{ML}}(x_i, x_j) \quad (11)$$

$$i, j = 1, 2, \dots, L; i \neq j$$

### C. Superpixel-based Band Selection Approach

AP [21] performs the clustering of data based on the measure of similarity. The process of clustering in AP consists in finding the optimal set of cluster centers, i.e., exemplars for which the sum of similarities of each point to its center is maximized. In conventional AP, a common choice for similarity is negative Euclidean distance (ED). The self-similarity represents the prior suitability of data points to be exemplars, and can be customized to a specific value for each data point or set to a global (shared) value. The AP does not require that the number of clusters is prespecified, as it is controlled by setting the self-similarity value. In this paper, the developed BPC, i.e.,  $\text{BPC}^{\text{ML}}$  and  $\text{BPC}^{\text{RL}}$ , and BCC, i.e.,  $\text{BCC}^{\text{ML}}$  and  $\text{BCC}^{\text{RL}}$ , are introduced into AP as BM for band selection. The BM can be formulated using (12):

$$\text{BM}(x_i, x_j) = \begin{cases} -SCFTS \cdot \frac{1}{\text{BPC}(x_i, x_j)} \cdot \frac{\text{Max}}{\text{Min}} & i = j \\ -\left| \frac{1}{\text{BCC}(x_i, x_j)} \right| & i \neq j \end{cases} \quad (12)$$

$$i = 1, 2, \dots, L; j = 1, 2, \dots, L$$

where *Max* and *Min* are the maximum and minimum values of  $\text{BPC}(x_i, x_j)$  ( $i=j$ ), i.e., the self-similarity of bands; *SCFTS* (called Feature Threshold Scalar with SC) is used to get expected spectral bands through setting of an appropriate value; The  $|\cdot|$  denotes the absolute value of the BCC. The BM is defined in a negative domain, i.e., a small value indicates a large similarity. By introducing the BPC and BCC into the BM construction simultaneously, high-discriminative and low-redundant bands subset can be selected and retained. Then the defined BM are used in the AP as input measures.

AP is derived from factor graph in which different types of messages need to be propagated. In AP, these messages can be reduced to two simple sets, i.e., responsibility  $r$  and availability  $a$ . The two kinds of messages consider different competitions that are iteratively updated to achieve the decision of exemplars. Let responsibility  $r(x_i, x_j)$  denote the degree of band  $x_j$  to serve as the exemplar for band  $x_i$  relative to other bands. Let availability  $a(x_i, x_j)$  indicate the suitability of choosing band  $x_j$  as exemplar of band  $x_i$ , taking into account the support from other

bands. The values of availability  $a(x_i, x_j)$  are initialized to zero, i.e.  $a(x_i, x_j)=0$ . The responsibility and availability between two bands  $x_i$  and  $x_j$  are updated by a max-product algorithm as follows:

$$a(x_i, x_j) = \begin{cases} \min \left\{ 0, r(x_j, x_j) + \sum_{p \neq j, i} \max \{ 0, r(x_p, x_j) \} \right\} & i \neq j \\ \sum_{p \neq j} \max \{ 0, r(x_p, x_j) \} & i = j \end{cases} \quad (13)$$

$$r(x_i, x_j) = BM(x_i, x_j) - \max_{p \neq j} \{ BM(x_i, x_p) + a(x_i, x_p) \} \quad (14)$$

where  $x_p$  is the  $p$ th band.

This search algorithm often leads to oscillations when computing responsibilities and availabilities with simple updating rules. In actual applications, damping is commonly used to avoid numerical oscillations. The responsibility and availability could be damped according to following equations:

$$\begin{aligned} \hat{R}^{t+1} &= \alpha \hat{R}^t + (1 - \alpha) \hat{R}^t \\ \hat{A}^{t+1} &= \alpha \hat{A}^t + (1 - \alpha) \hat{A}^t \end{aligned} \quad (15)$$

where  $\hat{R}$  and  $\hat{A}$  represent responsibility and availability vectors, respectively;  $\alpha$  is the factor of damping (which should satisfy  $0.5 \leq \alpha < 1$ ); and  $t$  is the number of iterations. Higher values of  $\alpha$  will lead to slower convergence.

For any band  $x_i$ , a larger sum of  $a(x_i, x_j)$  and  $r(x_i, x_j)$  means a greater possibility of band  $x_j$  to be the final cluster exemplar of band  $x_i$ . Band  $x_i$  determines its cluster exemplars according to the following equation:

$$\max_{x_j \in C} \{ a(x_i, x_j) + r(x_i, x_j) \} \quad (16)$$

where  $C = \{c_1, c_2, \dots, c_{nc}\}$  is the cluster exemplars set, and  $nc$  is the number of cluster centres.

Let us assume that  $Y = \{y_1, y_2, \dots, y_l\}$ ,  $y_l \in X$ , ( $l \ll L$ ) is a representative bands subset. The assignment of representative bands is done by the final cluster exemplars, i.e.,  $Y = C$ .

In the proposed BS approach, the ML-BC and RL-BC are used to determine the BM that is introduced into AP for choosing representative bands, and the corresponding two superpixel-based BS methods referred to as superpixel metric learning-based AP (SML-AP) and superpixel representation learning-based AP (SRL-AP). Fig. 1 shows the flowchart of the proposed superpixel-based BS approach. The procedure of the proposed BS scheme is illustrated in detail in Algorithm 2.

---

#### Algorithm 2 Superpixel-based Band Selection

---

##### Input:

Hyperspectral data set  $X = \{x_1, x_2, \dots, x_L\}$   
 Number of SC,  $K$   
 Feature Threshold Scalar with SC,  $SCFTS$

##### Output:

Representative band subset  $Y = \{y_1, y_2, \dots, y_l\}$  ( $l \ll L$ )

##### Procedure:

1. Extend ERS-based superpixel to HSI pixel vector, i.e., obtain the  $SC_k = \{x_1^{(k)}, x_2^{(k)}, \dots, x_{n_k}^{(k)}\}$  ( $k=1, 2, \dots, K$ ).
  2. Estimate the variability of SC, i.e., calculate within-SC covariance matrix  $\hat{C}_{sc}$  and total-SC covariance matrix
- 

---

$\hat{C}_{isc}$  using (1) and (5).

3. Calculate the whitening transformation matrix  $W$  with the constructed SC.

$$W = \hat{C}_{sc}^{-1/2}$$

4. Establish bands criteria, i.e. BPC and BCC:

- *BPC*:

SML-AP - Considering only the variability within the SC, the BPC is defined as:

$$BPC(x_p, x_p) = W(x_p, x_p)$$

SRL-AP - Considering the total variability of bands, BPC is calculated by integrating the eigenanalysis, i.e., diagonalize  $\hat{C}_{isc} \cdot \hat{C}_{sc}^{-1}$  to find  $L$  real and nonzero eigenvalues and their corresponding  $L$ -dimensional eigenvectors. Then band priority score  $\rho$  is computed by (8) and (9), and the BPC is defined as

$$BPC(x_p, x_p) = \rho_p$$

- *BCC*:

The BCC is given by (12)

$$BCC(x_i, x_j) = W(x_i, x_j)$$

5. Measure similarity BM between bands using (12).
  6. Update responsibility and availability according to (13), (14) and (15).
  7. Identify cluster exemplars  $C$  by the maximum value of the availabilities and responsibilities based on (16), and the number of cluster centre  $nc$  by *SCFTS*.
  8. Repeat steps 6-7 until decisions for cluster exemplars are unchanged, i.e., cluster boundaries are unchanged for some number of iterations. The final representative band subset  $Y = C$  and the number of bands  $l = nc$  can be used for HSI classification.
- 

### III. EXPERIMENTAL RESULTS AND DISCUSSIONS

#### A. Data Set Description

In order to assess the effectiveness of proposed BS method, three hyperspectral data sets are used in the experiments. In the following, the description of these data sets is given.

*Indian Pines 92AV3C* [62], consists in an image acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over northwest Indiana on 12 June 1992. It is extracted from the Indian Pines North-South image and has  $145 \times 145$  pixels. The original image contains 224 spectral channels with a wavelength range from 400–2500 nm with a spatial resolution of 20 m. After removing 4 zero and 20 water absorption bands (numbered 104–108, 150–163, and 220), it results in a total of 200 channels. 15 noisy bands (1–3, 103, 109–112, 148–149, 164–165, and 217–219) are often removed manually in advance from this data set [8]. In this paper, these noisy bands are not removed on purpose and act as a natural test for the two proposed BS methods. Fig. 2 shows a false color composition of the image and the map of the available reference labeled data for this data set. The data set includes 16 land-cover classes, which represent different crop types, vegetation and man-made structures with 10,366 ground labeled pixels.

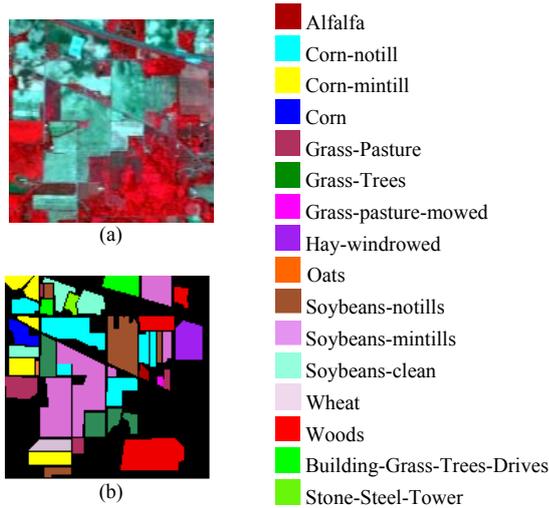


Fig. 2. (a) False color composite of image of bands 57, 27, 17; (b) available ground truth map (Indian Pines data set).

*University of Pavia*, was collected in 2003 by the Reflective Optics System Imaging Spectrometer (ROSIS-03) (spectral coverage ranging from 0.43 to 0.86  $\mu\text{m}$ ) surrounding the urban area of the University of Pavia, Italy. This data set has a size of  $610 \times 340$  pixels with 115 spectral bands and a very high spatial resolution channel (1.3 m per pixel). 12 spectral bands were removed from the data set due to noise, obtaining a total of 103 spectral channels to be used in the experiments. Nine land-cover classes are considered in the experiments: trees, asphalt, bitumen, gravel, metal sheets, shadow, bricks, meadows and bare soil. Fig. 3 shows the false color composite image and the available ground truth map of this data set.

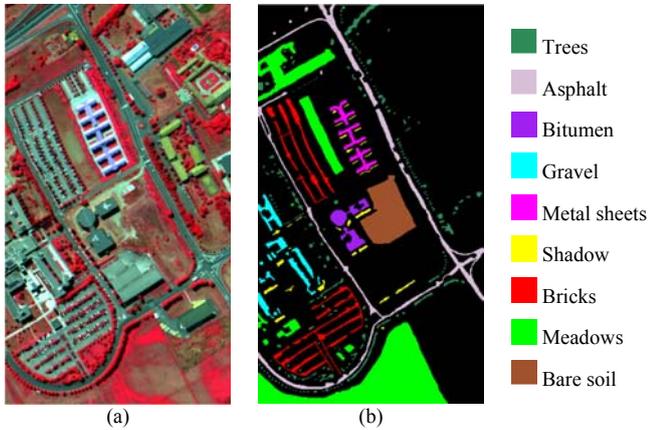


Fig. 3. (a) False color composite image of bands 102, 56, 31; (b) available ground truth map (University of Pavia data set).

*Pavia Center*, was recorded by the ROSIS-03 sensor over the centre of Pavia, Italy. The original image has 115 bands of size  $1096 \times 1096$  with a spatial resolution of 1.3 m per pixel and a spectral coverage ranging from 0.43 to 0.86  $\mu\text{m}$ . 13 most noisy channels were removed before experiments. Nine classes of interest are present in this image. The color composite of the Pavia centre image and the corresponding ground reference data are shown in Fig. 4.

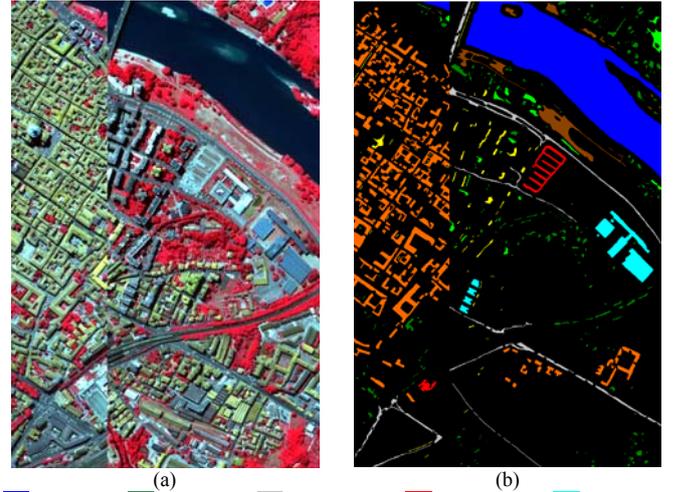


Fig. 4. (a) False color composite image of bands 102, 56, 31; (b) available ground truth map (Pavia centre data set).

## B. Design of Experiments

### 1) Compared algorithms

To assess the effectiveness of the proposed superpixel-based BS methods, i.e., SML-AP and SRL-AP, five unsupervised techniques, including three representative and two AP-based methods were compared in the experiments as described below.

- ID [16],[17]: All bands are measured by a band prioritization criterion, i.e., non-Gaussianity (NG) [18],[19].
- MVPCA [20]: Maximum variance criterion used to prioritize bands.
- E-FDPC [25]: The score of each band is computed by weighting the normalized local density and the intracluster distance obtained by ranking-based clustering method, i.e., a fast density-peak-based clustering (FDPC) algorithm.
- AP [21]: Negative Euclidean distance and median similarity were adopted as similarity measure between two different bands and self-similarity, i.e., preference, respectively.
- AAP [22]: AP with negative spectral angle mapper (SAM), and a fixed number of selected band obtained by an exemplar number determination procedure.

We also compared the results obtained with those achieved by using all of the original bands (Baseline).

### 2) Evaluation Indices

In the experiments, the following quantitative indices were considered in the evaluation of the proposed superpixel-based BS methods and reference methods.

- Accuracy measures:

To ensure an objective and quantitative evaluation process, two representative classifiers were implemented, i.e.,  $k$ -nearest neighborhood ( $k$ -NN) algorithm and support vector machine (SVM). The characteristics and parameter setting of the two classifiers are briefly described below.

$k$ -NN is one of the simplest non-parametric classifiers that does not depend on any data distribution assumption [63]. It commonly uses the Euclidean distance to measure the

similarity between a testing sample and the training data in the related neighborhood, and assigns the class by a majority voting scheme according to the most frequent class label in the  $k$ -nearest range.  $k$  is a positive integer that typically has a small value. The  $k$  value in our experiment was set to three.

*SVM* is a discriminative model used for classification and regression analysis. It is based on statistical learning theory developed by Vapnik [64] in which the classification model is defined by exploiting the concept of margin maximization. *SVM* do not require an estimation of the statistical distributions of classes to carry out the classification task. *SVM* appear to be especially advantageous in solving small training sample and nonlinear problems in high dimensional data, and has delivered state-of-the-art performance in HSI classification [65]. The *SVM* was implemented with radial basis function (RBF) kernel, and the parameters  $C$  and  $\gamma$  were optimized through tenfold cross validation.

For Indian Pines, University of Pavia and Pavia Center data sets 20%, 10% and 2% of available ground truth data were randomly selected as training set according to the characters of ground objects, and remaining samples used as test samples. For all data sets, five trials with different training samples were carried out to reduce the random effect of results. The overall accuracy (OA) and average overall accuracy (AOA) were used for analyzing the results. Moreover, also the standard deviation (SD) of overall accuracy over the five trials was also reported.

- Number of selected bands:

For the considered three HSI data sets, the desired number of bands to be selected is not known a priori. For demonstrating the effectiveness of proposed superpixel-based BS approach, for each data set we carried out trials with different numbers of selected bands (in the range between 5 and 60) for all the considered algorithms.

### C. Parameter Setting

In the proposed superpixel-based BS approach, three user defined parameters, i.e., convergence parameter  $\alpha$ , number of superpixel chunklets  $K$  and feature threshold scalar with SC  $SCFTS$  should be defined. The setting values and criteria are discussed below.

#### 1) Effect of convergence parameter $\alpha$

In AP-based BS methods, the parameter  $\alpha$  is introduced to avoid oscillations caused by "overshooting" the solution. The parameter  $\alpha$  should be at least 0.5 and smaller than 1. If the algorithm does not converge, the value can be increased, but the execution time increases as well. We suggest to fix parameter  $\alpha$  between 0.8 and 0.9. In our experiments, the  $\alpha$  values of AP, AAP and the proposed SML-AP and SRL-AP were set to 0.9. A larger value means a better convergence of AP.

#### 2) Effect of number of superpixel chunklets $K$

To obtain the desired properties of SC (especially the one for which every SC should overlap with only one ground object), the number of SC is selected based on ground objects distribution and structural texture complexness in the three considered HSI data sets. A larger number of SC results in SC containing few pixels, which causes that SC cannot provide

enough spatial information for the expected ground objects classification, whereas a smaller number of SC easily leads to some overlap with multi-ground objects (under segmentation errors). According to the characteristics of the Indian Pines, University of Pavia, Pavia Center data sets and the properties of SC, the ranges of number of SC used are {350, 500, 700}, {350, 600, 850}, {350, 600, 850} respectively.

Fig. 5 shows classification accuracy obtained by the proposed SML-AP and SRL-AP with different number of SC. From this figure, one can observe that the AOA values of the proposed BS methods have different behaviors compared to the Baseline versus the number of SC. These different trends are caused by the different image characteristics and sizes in the three considered data sets and are expected and common in super-pixel/segmentation based methods. The Indian Pines data set contains many large multiple homogeneous regions and few of small size. Accordingly, a large number of SC can provide higher classification accuracies in this data set. A good and stable accuracy was obtained when the number of SC was 500, whereas an unstable trend and drop in AOA were observed with a number of SC around 700. The University of Pavia data set presents complex structures and textures, thus a relative large number of SC should be used for maintaining homogeneity within the same ground objects. Results show that there is a strong fluctuation in AOA values, with more stable and accurate results around 600 SC. The Pavia Center data set contains many small size ground objects with wider coverage. Therefore, a larger value of number of SC is required to better reflect spatial information of ground objects. For this data set, the AOA provided by the proposed SML-AP and SRL-AP are relatively high when considering about 850 superpixel chunklets. On the basis of this analysis, the number of SC was set to 500 for the Indian Pines data set, to 600 for the University of Pavia data set and 850 for the Pavia Center data set in the following experiments.

#### 3) Effect of feature threshold scalar

For obtaining the desired number of bands, the  $SCFTS$  is introduced in the proposed BS methods. The value of  $SCFTS$  is monotonically related to the number of selected representative bands, i.e., relatively small values resulted in the selection of many bands, whereas high values led to a small number of bands in all sampling conditions of the three considered data sets (see Fig. 6). As done in other BS techniques, one can run SML-AP and SRL-AP several times with different  $SCFTS$  values searching for the desired number of bands.

### D. Results

In this section, consistent comparisons among the proposed SML-AP and SRL-AP and the other reference methods (i.e., ID, MVPCA, E-FDPC, AP, AAP, and Baseline) are presented. The AOA values obtained by using all the spectral channels and different bands subsets are compared in Fig. 7. From this figure, one can observe that in most cases the proposed SML-AP and SRL-AP resulted in higher AOA compared with the five considered BS methods for the same number of selected bands on the three data sets.

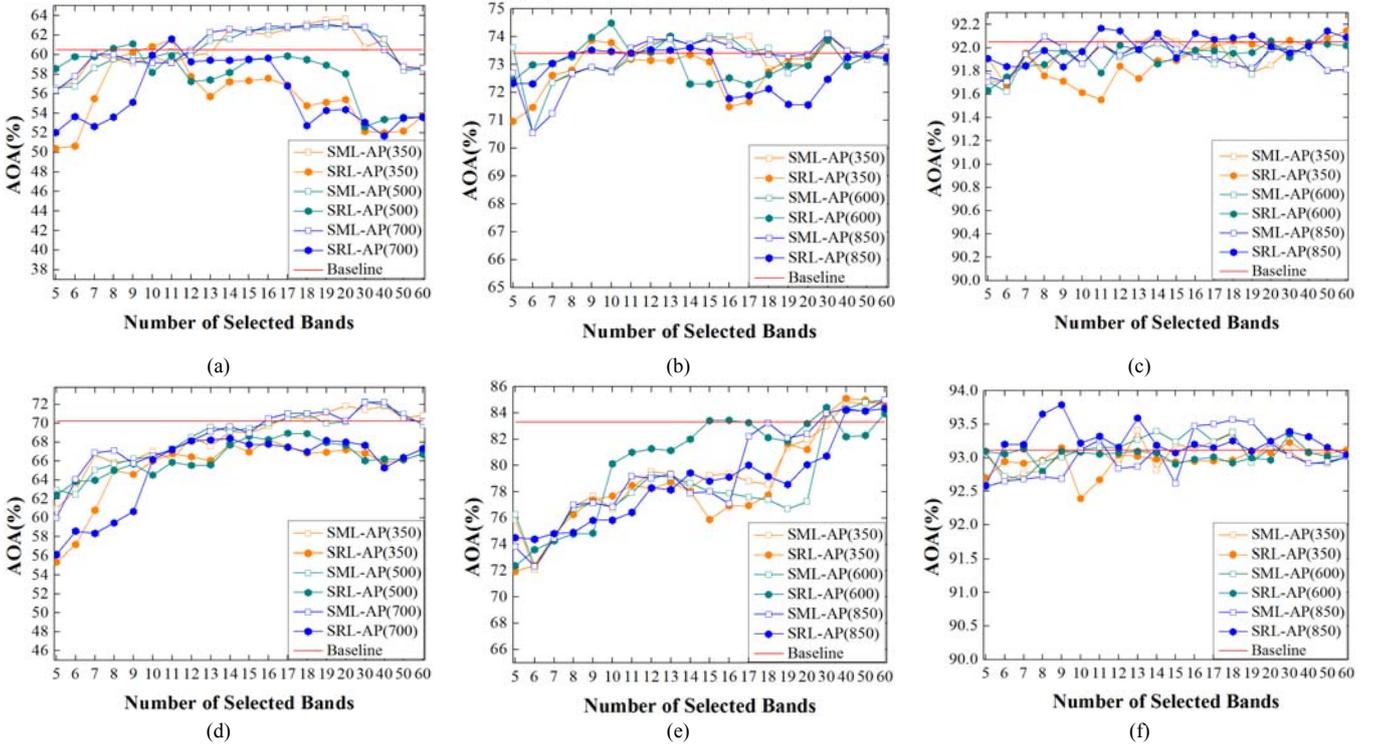


Fig. 5. AOA versus the number of selected bands obtained by proposed SML-AP and SRL-AP with different number of superpixel chunklets. The AOA results achieved by using all the spectral channels are also reported (Baseline). (a)  $k$ -NN on Indian Pines data set; (b)  $k$ -NN on University of Pavia data set; (c)  $k$ -NN on Pavia Center data set; (d) SVM on Indian Pines data set; (e) SVM on University of Pavia data set; (f) SVM on Pavia Center data set.

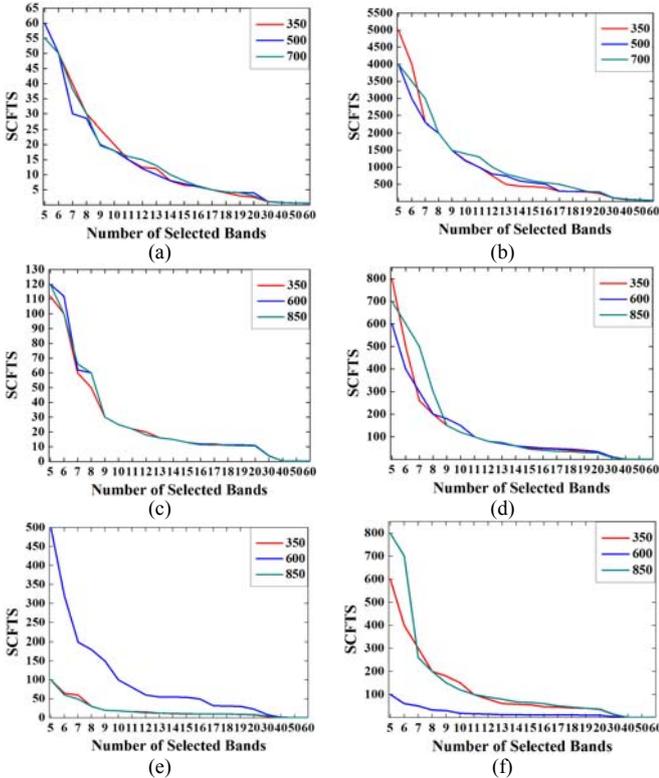


Fig. 6. Values of parameter  $SCFTS$  versus the number of selected bands. (a) SML-AP on Indian Pines data set; (b) SRL-AP on Indian Pines data set; (c) SML-AP on University of Pavia data set; (d) SRL-AP on University of Pavia data set; (e) SML-AP on Pavia Center data set; (f) SRL-AP on Pavia Center data set.

In Indian Pines data set experiments [Fig. 7(a), (d)], one can see that the proposed SML-AP and SRL-AP produced superior results as compared to other considered literature techniques. The accuracies obtained with SRL-AP were higher than those obtained with SML-AP when the number of selected bands was small. However, they did not achieve higher values when more bands were selected. In these conditions, the AOA obtained with SML-AP increased steadily by increasing the number of selected bands. In greater detail, SML-AP obtained an AOA of 61.42% for the  $k$ -NN and of 70.53% for the SVM when 13 and 17 bands were selected, respectively. These values are higher than those achieved by using all 200 bands (60.49% and 70.22%). However, SRL-AP delivers more stable performance, i.e., lower SD than SML-AP by increasing the number of selected bands [Fig. 8(a), (d)]. Although the E-FDPC and AAP achieved a few times higher classification accuracies than the proposed techniques, their performance is unstable. A decrease of classification accuracies in E-FDPC and AAP appear by increasing the number of selected bands. The MVPCA obtained lowest classification accuracy.

In University of Pavia data set experiments [Fig. 7(b), (e)], the proposed SML-AP and SRL-AP exhibited the highest accuracy and the lowest SD. The AOA obtained with SML-AP at first increased slowly by increasing the number of selected bands. Then it sharply growth when the number of selected bands was larger than 20. On the contrary, the SRL-AP obtained high accuracy with a small number of bands selected. One can see that the SRL-AP resulted in AOA values of 73.98% for the  $k$ -NN and 83.42% for the SVM by selecting 9 and 15 bands, respectively, which are higher than those

obtained by the baseline (73.41% and 83.32%). The E-FDPC and MVPCA obtained several times higher accuracy than the proposed methods, but with more selected bands and higher SD.

The AP and AAP provided similar AOA and growth trend by increasing the numbers of selected bands. The ID yielded the lowest AOA among the seven considered BS methods.

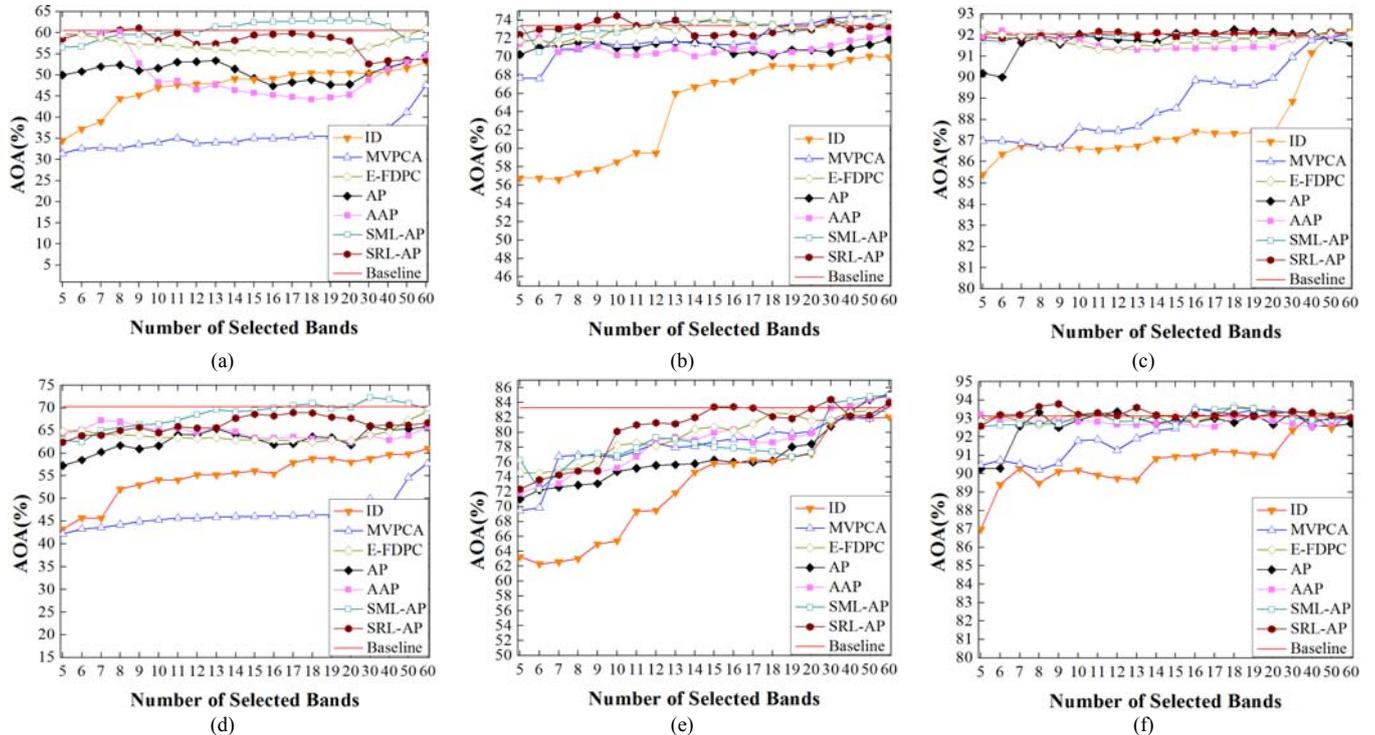


Fig. 7. AOA versus the number of selected bands obtained by the ID, MVPCA, E-FDPC, AP, AAP and the proposed SML-AP and SRL-AP methods. The AOA results achieved by using all the spectral channels are also reported (Baseline). (a)  $k$ -NN on Indian Pines data set; (b)  $k$ -NN on University of Pavia data set; (c)  $k$ -NN on Pavia Center data set; (d) SVM on Indian Pines data set; (e) SVM on University of Pavia data set; (f) SVM on Pavia Center data set.

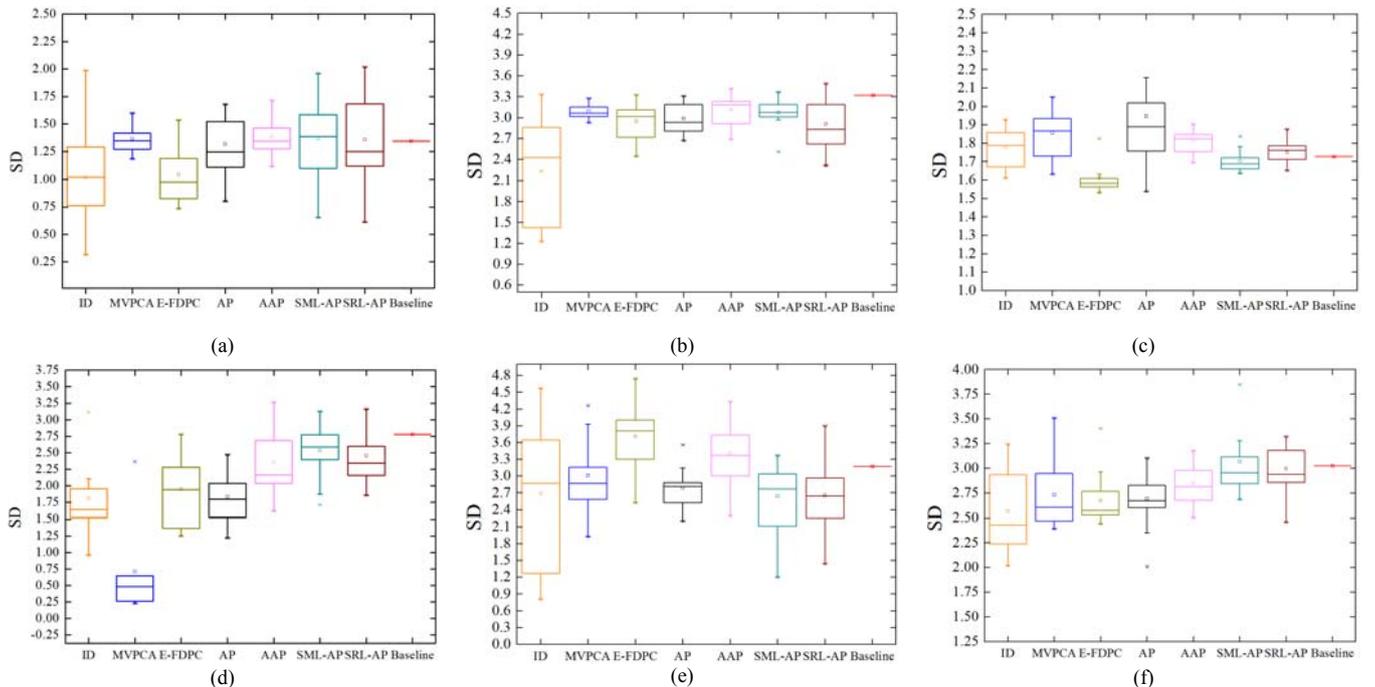


Fig. 8. SD of the ID, MVPCA, E-FDPC, AP, AAP and proposed SML-AP and SRL-AP methods with different selected bands. The SD results achieved by using all the spectral channels are also reported (Baseline). (a)  $k$ -NN on Indian Pines data set; (b)  $k$ -NN on University of Pavia data set; (c)  $k$ -NN on Pavia Center data set; (d) SVM on Indian Pines data set; (e) SVM on University of Pavia data set; (f) SVM on Pavia Center data set.

In the Pavia Center data set experiments [Fig. 7(c), (f)], the AOA provided by AP-based methods (i.e., the AP, AAP, SML-AP and SRL-AP methods) are close to each other and always better than the two ranking-based methods (i.e., the ID and MVPCA). Similarly to Indian Pines data set, the E-FDPC did not increase the accuracies when the number of selected bands increased. This is mainly due to the gain in the band selection process provided by an efficient message-passing scheme. In greater detail, the proposed SML-AP and SRL-AP can achieve more accurate results than those of Baseline with a small number of selected bands (i.e., 11 for the KNN and 9 for the SVM) and exhibited relatively stable performances [Fig. 8(c), (f)].

TABLE I  
SELECTED BANDS OBTAINED BY DIFFERENT BAND SELECTION METHODS ON INDIAN PINE, UNIVERSITY OF PAVIA AND PAVIA CENTER DATA SETS

Data sets	Methods	Selected bands
Indian Pines	ID	168, 169, 173, 174, 172, 178, 181, 117, 183, 118, 182, 184, 180
	MVPCA	214, 215, 208, 212, 206, 200, 216, 205, 195, 197, 194, <b>219</b> , 183
	E-FDPC	127, 49, 24, 88, 184, 68, 210, 209, 211, 208, 171, 213, 207
	AP	13, 33, 50, 57, 76, 79, 97, 102, 143, 193, 213, <b>217</b> , <b>219</b>
	AAP	1, 9, 17, 25, 48, 66, 88, 101, 104, 120, 129, 182, <b>219</b>
	SML-AP	64, 181, 13, 31, 145, 74, 53, 88, 117, 134, 122, 203, 168
	SRL-AP	80, 145, 100, 130, 75, 124, 19, 178, 53, 88, <b>165</b> , 198, 209
University of Pavia	ID	103, 95, 90, 92, 2, 91, 94, 88, 87
	MVPCA	74, 78, 75, 62, 97, 96, 83, 67, 102
	E-FDPC	55, 93, 18, 32, 82, 90, 84, 66, 81
	AP	13, 24, 39, 52, 58, 63, 75, 90, 93
	AAP	2, 16, 29, 40, 48, 59, 71, 76, 90
	SML-AP	6, 17, 27, 43, 49, 65, 71, 82, 96
	SRL-AP	19, 16, 28, 43, 45, 64, 76, 87, 93
Pavia Center	ID	91, 66, 92, 93, 94, 59, 101, 95, 85
	MVPCA	89, 95, 87, 70, 57, 94, 83, 90, 78
	E-FDPC	51, 92, 29, 62, 19, 90, 55, 82, 98
	AP	9, 12, 13, 15, 17, 30, 43, 58, 70
	AAP	2, 20, 31, 39, 49, 60, 71, 76, 91
	SML-AP	6, 17, 26, 36, 44, 65, 71, 86, 95
	SRL-AP	11, 46, 38, 14, 25, 58, 76, 87, 98

Different bands selected by proposed SML-AP and SRL-AP, and the compared BS methods for all data sets are reported in Table I. Note that the bands selected by the proposed SML-AP and SRL-AP, the ID, MVPCA and E-FDPC methods are presented in the order of prioritization, whereas those selected by the AP and AAP methods are given in the sorted sequence. From the comparison of numerical distributions, one can notice that the ID and MVPCA selected many adjacent bands on the three data sets as they only consider the bands prioritization, and the results obtained by  $k$ -NN and SVM are consistent with each other showing a similar trend. Some bands selected by the E-FDPC are still close to each other due to the fact that the intracluster distance is not adopted in bands ranking. The bands selected by AP and AAP are mostly the same or next to each other. In this condition, the proposed SML-AP and SRL-AP can select more dispersed band subsets, which cover large

intervals of the spectrum and show high discriminative capabilities on the three data sets. On the other hand, for the Indian Pines data set, the selected bands of a few BS methods (i.e., MVPCA, AP, AAP and SRL-AP) contain noisy bands (marked in bold). This points out that some noisy bands may contain useful spatial structure information and provide discriminative capabilities even if the low quality of spectral information may influence the classification performance.

### E. Discussions

From the above experimental results on the three considered HIS data sets, we can observe that superpixel-based methods (i.e., SML-AP and SRL-AP) are almost always superior to the other BS methods in terms of the AOA and SD. In this section we analyze the effectiveness and applicability of the SML-AP and SRL-AP with respect to the considered HSI information content and band criterion.

#### 1) Statistical Test

In order to further analyze the results, we focused on the relatively highest performance of the proposed SML-AP and SRL-AP methods obtained on the three data sets for comparison. We used McNemar's test [66] with variation to evaluate the classification results with 13 and 17, 9 and 15, 9 and 11 selected bands by using  $k$ -NN and SVM on Indian Pines, University of Pavia and Pavia Center data sets, respectively. McNemar's test is a statistical test based on standardized normal test statistic:

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (17)$$

where  $f_{ij}$  refers to the number of pixels correctly classified by learner  $i$ , but incorrectly by learner  $j$ . The difference between learners  $i$  and  $j$  is regarded as statistically significant at 5% level of significance if  $|Z| > 1.96$ . The results on five trials of the McNemar's test on the proposed SML-AP and SRL-AP and on other BS methods for the three data sets are shown in Table II. One can observe from the table that the proposed SML-AP and SRL-AP can always provide more accurate results ( $Z < -1.96$ ) than the spectral BS methods (i.e., the ID, MVPCA, E-FDPC, AP, AAP) and comparable with Baseline ( $|Z| < 1.96$ ) on the three considered data sets. Generally, spectral BS methods exploit pixelwise processing to chose informative and low-redundant bands. By contrast, with the introduced adaptive spatial information, the proposed SML-AP and SRL-AP methods will benefit from the separation of structured regions, and can improve classification performance effectively as also proven by the test statistic significance values. The statistical differences between SML-AP and SRL-AP are by 9.58 and 3.91, 6.21 and 30.28, 6.088 and 25.25 when using  $k$ -NN and SVM on Indian Pines, University of Pavia and Pavia Center data sets, respectively.

#### 2) Applicability

The proposed SML-AP and SRL-AP have been analyzed with two different types of bands criteria, i.e., ML-BC and RL-BC. This sub-section discusses the performances obtained with the two different criteria.

From Fig. 5, one can observe that SML-AP is not sensitive to the number of SC for the three considered data sets. This is

mainly due to the fact that the ML-BC only considers the within SC variability to learn a transformation matrix for establishing the band prioritization criterion. For SRL-AP, the number of SC has a certain impact on the band selection results for the RL-BC learning with a total SC variability via eigen (spectral) decomposition.

The proposed SML-AP and SRL-AP show different advantages on the three HIS data sets. For comparison purpose, OA versus the number of selected bands obtained by the proposed SML-AP and SRL-AP methods on five trials is given in Fig. 9. For Indian Pines data set [see Fig. 9(a), (d)], although SRL-AP provided few times the highest accuracy with  $k$ -NN, SML-AP generally achieved a good and stable performance by increasing the number of selected bands. This is motivated by the fact that the Indian Pines data set contains indistinguishable crops types, i.e., it has high intraclass variability which makes the discrimination of land cover more difficult. The SML-AP with ML-BC focus on finding spectral bands with low SC

variance, i.e., low intraclass variability, whereas the SRL-AP with RL-BC may emphasize this variability. Complex textures and high resolution contained in University of Pavia data set result in high intraclass variability and low interclass variability. In this condition, SRL-AP is more effective than the SML-AP [see Fig. 9(b), (e)]. For the Pavia Center data set [see Fig. 9(c), (f)], the classification accuracy provided by the SML-AP and SRL-AP increases quickly and is superior to the baseline when a small number of bands is selected. However, the accuracy obtained by the SML-AP decreases by increasing the number of selected bands showing relatively large fluctuations. On the contrary, SRL-AP remains relatively stable with small fluctuations by increasing the number of selected bands. This confirms that the RL-BC with total and within SC variance combination can balance the two kinds of variabilities and is more suitable than the ML-BC when facing high resolution HSI images.

TABLE II  
MCNEMAR'S TEST BETWEEN PROPOSED SML-AP AND SRL-AP, AND ID, MVPCA, E-FDPC, AP, AAP, BASELINE  
ON INDIAN PINE, UNIVERSITY OF PAVIA AND PAVIA CENTER DATA SETS WITH  $k$ -NN AND SVM

Data sets	Methods	Z (Trial1)		Z (Trial2)		Z (Trial3)		Z (Trial4)		Z (Trial5)		
		SML-AP	SRL-AP	SML-AP	SRL-AP	SML-AP	SRL-AP	SML-AP	SRL-AP	SML-AP	SRL-AP	
Indian Pines	ID	$k$ -NN	-27.52	-19.33	-22.81	-15.87	-20.57	-13.43	-21.01	-15.02	-25.85	-18.38
		SVM	-26.32	-20.80	-26.21	-19.86	-20.63	-20.59	-20.62	-20.00	-21.75	-18.63
	MVPCA	$k$ -NN	-44.14	-38.22	-41.62	-36.75	-40.12	-35.08	-39.18	-35.01	-43.21	-37.86
		SVM	-41.02	-37.83	-42.71	-37.55	-35.67	-35.85	-34.08	-33.82	-30.85	-28.50
	E-FDPC	$k$ -NN	-15.51	-3.81	-13.88	-3.92	-11.39	-1.11	-11.27	-2.82	-11.43	-2.27
		SVM	-21.21	-13.73	-20.88	-13.30	-16.42	-15.65	-13.46	-13.05	-14.73	-10.50
	AP	$k$ -NN	-17.04	-7.11	-20.08	-11.62	-14.10	-5.42	-15.56	-8.59	-15.97	-8.31
		SVM	-18.28	-13.27	-20.67	-13.40	-17.23	-17.23	-19.74	-20.26	-9.15	-5.50
	AAP	$k$ -NN	-27.28	-18.35	-26.44	-19.11	-24.00	-16.62	-24.22	-18.25	-24.86	-17.81
		SVM	-12.95	-8.28	-16.89	-9.75	-13.99	-15.06	-19.12	-20.03	-7.39	-3.80
	Baseline	$k$ -NN	-4.05	7.60	-1.76	8.13	-3.76	6.43	-2.24	5.67	-0.50	8.49
		SVM	-0.43	5.37	0.06	8.49	-3.32	-3.77	-1.06	-1.96	1.98	6.05
	SML-AP vs SRL-AP	$k$ -NN	11.54		9.88		9.90		7.80		8.80	
		SVM	5.60		8.30		0.20		0.80		4.65	
University of Pavia	ID	$k$ -NN	-42.39	-46.85	-66.31	-66.58	-55.29	-54.55	-51.55	-57.65	-57.26	-62.97
		SVM	-10.73	-50.02	-8.61	-29.90	-16.09	-32.41	2.46	-12.95	-8.07	-28.07
	MVPCA	$k$ -NN	-3.85	-10.59	-15.08	-16.31	-0.90	-0.12	-7.29	-16.74	-5.15	-13.99
		SVM	-0.12	-42.90	3.06	-20.91	-7.83	-24.99	5.02	-11.68	13.42	-7.86
	E-FDPC	$k$ -NN	-2.95	-10.64	-14.82	-16.13	-4.07	-3.40	-2.73	-13.38	-2.85	-12.40
		SVM	6.09	-38.61	10.79	-12.57	14.22	-2.89	18.78	1.34	12.41	-9.78
	AP	$k$ -NN	-5.79	-13.67	-10.15	-11.93	-5.06	-4.70	-6.82	-18.08	-8.87	-19.21
		SVM	-4.14	-47.48	-9.68	-33.00	-21.73	-36.73	-19.92	-36.01	2.64	-19.43
	AAP	$k$ -NN	-6.10	-13.45	-12.94	-14.26	-10.02	-9.30	-8.52	-19.01	-9.01	-18.29
		SVM	17.68	-35.98	10.27	-16.20	15.34	-1.59	2.60	-17.17	20.55	-1.46
	Baseline	$k$ -NN	0.90	-2.08	-1.26	-2.55	5.34	6.16	7.47	-3.95	1.72	-8.51
		SVM	38.29	-3.51	34.18	11.38	21.28	3.55	15.44	-5.31	44.41	-1.24
	SML-AP vs SRL-AP	$k$ -NN	-8.07		-1.36		0.90		-11.50		-10.41	
		SVM	-45.89		-24.41		-18.20		-19.69		-43.21	
Pavia Center	ID	$k$ -NN	-96.03	-96.44	-74.26	-77.74	-66.98	-68.20	-81.15	-83.82	-61.16	-65.46
		SVM	-46.33	-72.20	-100.75	-98.52	-53.61	-58.39	-39.51	-48.22	11.47	-53.06
	MVPCA	$k$ -NN	-96.02	-96.64	-65.42	-69.25	-64.95	-66.26	-63.59	-66.71	-55.75	-60.58
		SVM	-31.08	-60.77	-83.33	-80.58	-50.36	-55.02	-40.76	-50.75	38.18	-22.08
	E-FDPC	$k$ -NN	-30.06	-31.29	-9.93	-16.10	-12.40	-14.77	-23.50	-27.58	-11.60	-17.72
		SVM	5.29	-38.82	-32.75	-25.34	3.43	-4.83	4.28	-11.20	68.39	-0.49
	AP	$k$ -NN	-12.55	-13.74	6.51	-0.97	10.73	7.64	-13.58	-8.63	-8.34	-13.13
		SVM	-4.19	-43.24	-50.20	-46.67	-27.11	-32.89	-5.96	-17.48	59.86	-3.03
	AAP	$k$ -NN	-13.13	-14.56	-11.34	-17.08	-15.09	-16.71	-16.11	-18.89	-17.40	-22.97
		SVM	10.89	-34.89	-55.23	-51.63	2.64	-4.01	9.10	-7.86	65.73	6.62
	Baseline	$k$ -NN	-3.91	-6.37	-2.55	-12.05	0.13	-3.57	1.41	-4.00	11.18	4.13
		SVM	20.29	-25.02	-45.55	-41.32	2.34	-6.24	24.68	10.96	61.71	-24.84
	SML-AP vs SRL-AP	$k$ -NN	-2.72		-9.36		-4.14		-5.38		-8.84	
		SVM	-42.50		10.91		-8.78		-16.67		-69.22	

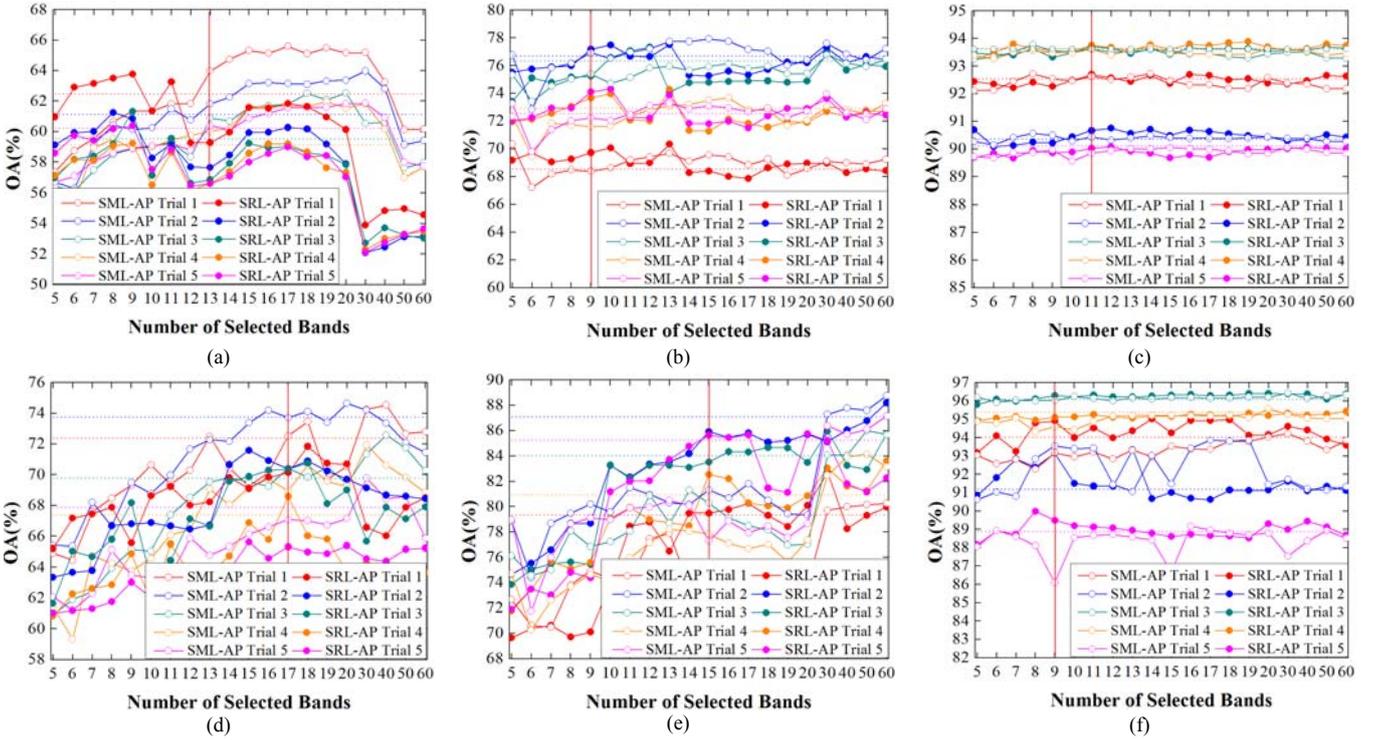


Fig. 9. OA versus the number of selected bands obtained by proposed SML-AP and SRL-AP methods on five trials. The OA results achieved by using all the spectral channels are also reported in dot line (a)  $k$ -NN on Indian Pines data set; (b)  $k$ -NN on University of Pavia data set; (c)  $k$ -NN on Pavia Center data set; (d) SVM on Indian Pines data set; (e) SVM on University of Pavia data set; (f) SVM on Pavia Center data set.

### 3) Execution Time

Fig. 10 shows the behavior of the execution time taken from the proposed superpixel-based BS approach, i.e., SML-AP and SRL-AP with two bands criteria, i.e., ML-BC and RL-BC, versus the number of selected bands on all the three considered data sets (software is implemented in Matlab with PC workstation (Intel(R) Core(TM) i7-3720QM CPU @ 2.60 GHz, 2.60 GHz with 16.0 GB of RAM)). The execution time of the proposed SML-AP and SRL-AP includes the computational cost of respective band criteria and band selection with AP. In this figure, navy blue and deep red denote the execution time of two bands criteria, i.e., ML-BC and RL-BC; while light blue and light red are the execution time of band selection in AP.

As expected, the computational times required by the ML-BC and RL-BC are larger than those required by band selection with AP and are proportional to the size of considered data set. The times taken by RL-BC were slightly smaller than those required by ML-BC. This is due to the computation complexity of the covariance matrix comparable in the two bands criteria. By analyzing the figure in greater detail, one can see that the execution time of band selection with AP is different on the three considered data sets. The execution time taken on Indian Pines data set is significant compared with those required by University of Pavia and Pavia Center data sets when increasing the number of selected bands. This can be explained by observing that the cost of discriminative bands subset selection is inversely proportional to the discriminability of land covers. Moreover, the execution time of the SML-AP showed a large variability on the three data sets, whereas the SRL-AP exhibits smooth performances.

## IV. CONCLUSION

In this paper, an unsupervised BS approach for HSI that exploits spectral and spatial information has been proposed. This approach is based on a new type of spectral-spatial information defined in terms of SC. By taking full advantage of the property of SC, metric learning-based and representation learning band criteria are built by learning the concept of the optimal RCA to measure bands prioritization and correlation. In metric learning-based scheme, the whitening transformation of RCA is used to estimate the within-SC covariance in HSI. Then, band prioritization and correlation is guided by the variability in each SC. In representation learning-based scheme, the within-SC variability and the total variability are exploited to determine the discrimination capability of each band. With the unsupervised extension of Fisher theory, the loading factors are constructed by eigen (spectral) decomposition of the optimal transformation of RCA for identifying the priority of each band and the correlation between any pair of bands. Two BCs (i.e., ML-BC and RL-BC) have been established and introduced into an AP-based search strategy, deriving two superpixel-based BS methods (i.e., SML-AP and SRL-AP).

To assess the effectiveness of the proposed superpixel-based BS approach, we have compared it with three representative and two AP-based methods by using three HSI data sets. The experimental results pointed out that the two proposed methods i.e., SML-AP and SRL-AP always provide higher accuracies and exhibit more stable performances compared to all of the considered BS methods. Furthermore, the accuracy results have illustrated the applicability of the two proposed BS methods on HSI data sets with different characteristics.

In future work, we plan to extend the proposed approach in order to take into account the possibility to perform domain adaptation when considering feature selection applied to different images acquired by the same sensor.

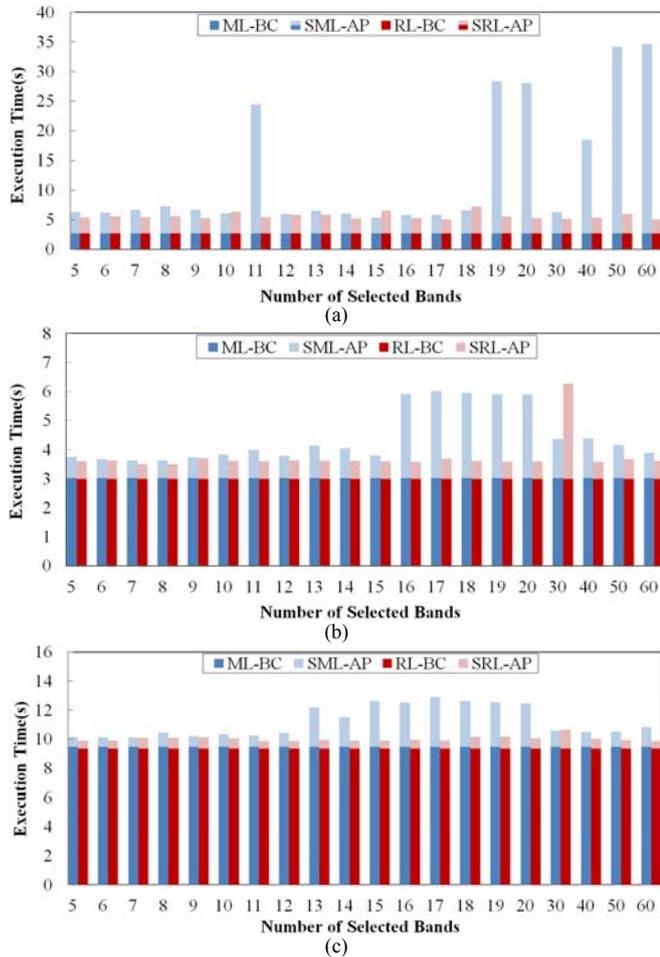


Fig. 10. Execution time (seconds) taken by the proposed superpixel-based BS approach with two bands criteria, i.e., ML-BC and RL-BC versus the number of selected bands. (a) SML-AP and SRL-AP on Indian Pines data set; (b) SML-AP and SRL-AP on University of Pavia data set; (c) SML-AP and SRL-AP on Pavia Center data set.

#### ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe from Purdue University for providing the free download of AVIRIS images, Prof. Paolo Gamba from the University of Pavia for providing the ROSIS images used in this paper.

#### REFERENCES

- [1] R. B. Smith, *Introduction to Hyperspectral Imaging*. Lincoln, NE: TNTmips; MicrolImages, Inc. 2006, pp. 1–24.
- [2] R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification (Second Edition)*, John Wiley & Sons Press, New York, UAS, 2001.
- [3] G. F. Hughes, “On the mean accuracy of statistical pattern recognizers,” *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1 pp. 55–63, Feb. 1968.
- [4] L. Bruzzone, and B. Demir, *A review of Modern Approaches to Classification of Remote Sensing Data*, In Land Use and Land Cover Mapping in Europe, ser. vol. 18, Jan. 2014, pp. 127–143.
- [5] B. S. Serpico and L. Bruzzone, “A new search algorithm for feature selection in hyperspectral remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1360–1367, Jul. 2001.

- [6] Y. Liu, H. Xie, L. Wang, and K. Tan, “Hyperspectral band selection based on a variable precision neighborhood rough set,” *Appl. Opt.* vol. 55, no. 3 pp. 462–472, Jan. 2016.
- [7] 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.
- [8] S. Patra, P. Modi and L. Bruzzone, “Hyperspectral band selection based on rough set,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5495–5503, May 2015.
- [9] X. Cao, T. Xiong and L. Jiao, “Supervised band selection using local spatial information for Hyperspectral image,” *IEEE Geosci. Remote Sens. Lett.* vol. 13, no. 3, pp. 239–233, Jan. 2016.
- [10] H. Su, H. Yang, Q. Du and Y. Sheng, “Semisupervised band clustering for dimensionality reduction of hyperspectral imagery,” *IEEE Geosci. Remote Sens. Lett.* vol. 8, no. 6, pp. 1135–1139, Jun. 2011.
- [11] J. Feng, L. Jiao, F. Liu, T. Sun, 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.
- [12] X. Bai, Z. Guo, Y. Wang, Z. Zhang and J. Zhou, “Semisupervised Hyperspectral band selection via spectral-spatial Hypergraph model,” *IEEE J. Sel. Top. App. Earth Obs. Remote Sens.* vol. 8, no. 6, pp. 2774–2783, Jul. 2015.
- [13] M. Gong, M. Zhang and Y. Yuan, “Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no.1, pp. 544–557, Aug. 2015.
- [14] X. Cao, B. Wu, D. Tao and L. Jiao, “Automatic band selection using spatial-structure information and classifier-based clustering,” *IEEE J. Sel. Top. App. Earth Obs. Remote Sens.* vol.9, no.9, pp. 4352–4360, Jan. 2016.
- [15] M. Zhang, J. Ma and M. Gong, “Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization,” *IEEE Geosci. Remote Sens. Lett.* vol. 14, no. 5, pp. 773–777, Mar. 2017.
- [16] T. M Cover and J. A. Thomas, *Elements of Information Theory (Second Edition)*. New York, USA: Wiley, 2006.
- [17] E. Arzuagacruz, L. O. Jimenezrodriguez and M. Velezreyes, “Unsupervised feature extraction and band subset selection techniques based on relative entropy criteria for hyperspectral data analysis,” in *Proc. Algorithms Technologies Multispectral Hyperspectral Ultraspectral Imagery*, 2003.
- [18] S. D. Stearns, B. E. Wilson, and J. R. Peterson, “Dimensionality reduction by optimal band selection for pixel classification of hyperspectral imagery,” *Applications of Digital Image Processing XVI*, vol. 2028, no. 330, pp. 118–127, Oct. 1993.
- [19] 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.
- [20] 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.
- [21] J.F. Frey and D. Dueck, “Clustering by passing messages between data points,” *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.
- [22] Y. Qian and F. Yao, “Band selection for hyperspectral imagery using affinity propagation,” *IET Comput. Vol.* 3, no. 4, pp. 213–222, Dec. 2009.
- [23] H. Su, Y. Sheng, P. Du, and K. Liu, “Adaptive affinity propagation with spectral angle mapper for semi-supervised hyperspectral band selection,” *Appl. Opt.*, vol. 51, no.14, pp. 2656–2663, May 2012.
- [24] S. Jia and L. Deng, “A CW-SSIM distance measure-based affinity propagation for hyperspectral band selection,” in *Proc. IEEE Int. Geosci. Remote Sens. Symposium, Melbourne, Australia*, 2013, pp. 418–421.
- [25] L. Jiao, J. Fang, F. Liu, S. Tao, and X. Zhang, “Semi-supervised affinity propagation based on normalized trivariable mutual information for hyperspectral band selection,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* Vol. 8, no. 6, pp. 2760–2773, Jun 2015.
- [26] 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. Jul. 2015.
- [27] A. Datta and S. Ghosh, “Combination of clustering and ranking techniques for unsupervised band selection of hyperspectral Images,” *IEEE J. Sel. Top. App. Earth Obs. Remote Sens.*, vol. 8, no. 6, pp. 2814–2823, May 2015.
- [28] G. Zhu, Y. Huang, J. Lei, Z. Bi and F. Xu, “Unsupervised hyperspectral band selection by dominant set extraction,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 227–239. Jul. 2015.

- [29] Z. Guo, X. Bai, Z. Zhang, and J. Zhou, "A hypergraph based semisupervised band selection method for hyperspectral image classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2013, pp. 3137–3141.
- [30] Z. Guo, H. Yang, X. Bai, Z. Zhang, and J. Zhou, "Semi-supervised hyperspectral band selection via sparse linear regression and hypergraph models," in *Proc. Geosci. Remote Sens. Symp. (IGARSS)*, 2013, pp. 1474–1477.
- [31] X. Bai, Z. Guo, Y. Wang, Z. Zhang and J. Zhou, "Semisupervised hyperspectral band selection via spectral-spatial hypergraph model," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 6, pp. 2774–2783, Jul. 2015.
- [32] X. Cao, T. Xiong and L. Jiao, "Supervised band selection using local spatial information for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.* vol. 13, no. 3, pp. 329–333, Jan. 2016.
- [33] N. Sental, T. Hertz, D. Weinshall and M. Pavel, "Adjustment learning and relevant component analysis," in *Proc. European Conf. Comput. vision*. 28–31 May 2002, pp. 776–790.
- [34] A. Bar-Hillel, T. Hertz, N. Sental and D. Weinshall, "Learning distance functions using equivalence relations," in *Proc. 20th Internat. Conf. Mach. Learn.* 2003, pp.11–18.
- [35] A. Bar-Hillel, T. Hertz, N. Sental, and D. Weinshall, "Learning a Mahalanobis metric from equivalence constraints," *Mach. Learn. Res.* vol. 6, no. 6, pp. 937–965, Dec. 2005
- [36] M. Sorci, G. Antonini, and J. Thiran, "Fisher's discriminant and relevant component analysis for static facial expression classification," in *Proc. European Signal Processing Conf.*, 2007, pp. 115–119.
- [37] C. Yang, L. Bruzzone, H. Zhao, Y. Liang, and R. Guan, "Decorrelation-separability-based affinity propagation for sem-isupervised clustering of hyperspectral Images," *IEEE J. Sel. Top. Appl. Earth Obs. Sens.* vol. 9, no. 2, pp. 568–582, Aug. 2016.
- [38] C. Yang, S. Liu, L. Bruzzone, R. Guan, and P. Du, "A feature-metricbased affinity propagation technique for feature selection in hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.* vol. 10, no. 5, pp. 1152–1156, Sep. 2013.
- [39] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. Computer Vision, 2003. Proceedings. Proceedings of the Ninth IEEE International*, 2003, pp. 10–17.
- [40] J. Shi, J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [41] P. Neubert, P. Protzel, "Superpixel benchmark and comparison," *Proc. Forum Bildverarb.*, Jan. 2012, pp. 1–12.
- [42] M. Wang, X. Liu, Y. Gao, X. Ma and N. Q. Soomro, "Superpixel segmentation a benchmark," *Signal Processing Image Communication*, vol. 56, Aug. 2017, pp. 28-39.
- [43] M.-Y. Liu, O. Tuzel, S. Ramalingam and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2097–2104.
- [44] M. V. D. Bergh, X. Boix, G. Roig, B. D. Capitani and L.V. Gool, "SEEDS: Superpixels extracted via energy-driven sampling," in *Proc. Computer Vision-ECCV 2012, Springer*, 2012, pp. 13–26.
- [45] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [46] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," in *Proc. Comput. Vision Pattern Recog. (CVPR)*, 2015, pp. 1356–1363.
- [47] L. Fang, S. Li, W. Duan, J. Ren and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Jul. 2015.
- [48] Li J, Zhang H and Zhang L. "Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 1–14, Apr. 2015.
- [49] J. Guo, X. Zhou, J. Li, A. Plaza and S. Prasad, "Superpixel-based active learning and online feature importance learning for hyperspectral image analysis," *IEEE J. Sel. Top. Appl. Earth Obs. Sens.*, vol. 10, no. 1, pp. 347–359, Oct. 2016.
- [50] A. Çalıřkan, E. Batı, A. Koza and A. A. Alatan, "Superpixel based hyperspectral target detection," in *Proc. Geoscience & Remote Sensing Symposium*, 2016, pp. 7010–7013.
- [51] Y. Liang, P. P. Markopoulos and E. S. Saber, "Subpixel target detection in hyperspectral images from superpixel background statistics," in *Proc. Geoscience & Remote Sensing Symposium*, 2016, pp. 7018–7021.
- [52] D. R. Thompson, L. Mandrake, M. S. Gilmore and R. Castano, "Superpixel endmember detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4023–4033, Sep. 2010.
- [53] X. Jin and Y. Gu, "Superpixel-based intrinsic image decomposition of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4285–4295, Apr. 2017.
- [54] X. Zhang, S. E. Chew, Z. Xu, and N. D. Cahill, "SLIC superpixels for efficient graph-based dimensionality reduction of hyperspectral imagery," in *Proc. SPIE*, 2015, pp. 947209.
- [55] J. Yi and M. Velezreyes, "Dimensionality reduction using superpixel segmentation for hyperspectral unmixing using the cNMF," *Proceedings of the Spie*, vol. 198, May 2017.
- [56] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*. Wiley-Interscience, John Wiley, 1992.
- [57] S. Mika, G. Ratsch, J. Weston, B. Scholkopf and K. Muller, "Fisher discriminant analysis with kernels," *IEEE Neural Networks for Signal Process.* vol.9, pp. 41–48, 1999.
- [58] P. W. Mausel, W. J. Kramber, and J. K. Lee, "Optimum band selection for supervised classification of multispectral data," *Photogramm. Eng. Remote Sens.* vol. 56, no. 1, pp. 55–60, Jan. 1990.
- [59] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis (Second Edition)*. New York: Springer-Verlag, 1993.
- [60] T. M. Tu, C.-H. Chen, J.-L. Wu and C.-I. Chang, "A fast two-stage classification method for high dimensional remote sensing data," *IEEE Trans. Geosci. Remote Sens.* vol. 36, no. 1, pp. 182–191, Jan. 1998.
- [61] C.-I. Chang, S. Wang, K.-H. Liu, M. Li, and C. Lin, "Progressive band dimensionality expansion and reduction via band prioritization for hyperspectral imagery," *IEEE J. Sel. Top. Appl. Earth Obs. Sens.* vol. 4, no. 3, pp. 591–614, Sep. 2010.
- [62] M. F. Baumgardner, L. L. Biehl and D. A. Landgrebe, *220 Band AVIRIS Hyperspectral image data set: June 12, 1992 Indian Pine Test Site 3*, Purdue University Research Repository: West Lafayette, Indiana, 2015.
- [63] J. Zou, W. Li, Q. Du, "Sparse representation-based nearest neighbor classifiers for hyperspectral imagery," *IEEE Remote Sens. Lett.*, vol. 12, no. 12, pp. 2418–2422, Dec. 2015.
- [64] C. Cortes and V. Vapnik, "Support-vector network," *Machine Learning* vol. 20, no. 3, pp. 273–297, sept. 1995.
- [65] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, 2004.
- [66] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, May 2004.



**Chen Yang** (M'11) was born in Jilin, China in 1981. She received the Ph.D. degree in digital geoscience from College of Earth Sciences, Jilin University, Changchun, China, in 2010. She was a visiting Ph.D. student at the Remote Sensing Laboratory, University of Trento, Italy in 2008.

She worked as a Post-Doctoral Researcher with the Remote Sensing Laboratory in the Department of Information Engineering and Computer Science, University of Trento, Italy, from 2011 to 2012. She is currently an Associate Professor in the College of

Earth Sciences, Jilin University, China. She has authored or co-authored over 30 scientific papers in refereed journals and proceedings, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the REMOTE SENSING and the IEEE Conference on Hyperspectral Image and Signal Processing, *etc.* Her research interests include remote sensing image processing and machine learning.



**Lorenzo Bruzzone** (S'95–M'98–SM'03–F'10) received the Laurea (M.S.) degree in electronic engineering (*summa cum laude*) and the Ph.D. degree in telecommunications from the University of Genoa, Italy, in 1993 and 1998, respectively. He is currently a Full Professor of telecommunications at the University of Trento, Italy, where he teaches remote sensing, radar, and digital communications.

Dr. Bruzzone is the founder and the director of the Remote Sensing Laboratory in the Department of Information Engineering and Computer Science, University of Trento. His current research interests

are in the areas of remote sensing, radar and SAR, signal processing, machine learning and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects. He is the Principal Investigator of many research projects. Among the others, he is the Principal Investigator of the *Radar for icy Moon exploration (RIME)* instrument in the framework of the *Jupiter ICy moons Explorer (JUICE)* mission of the European Space Agency. He is the author (or coauthor) of 218 scientific publications in referred international journals (157 in IEEE journals), more than 290 papers in conference proceedings, and 21 book chapters. He is editor/co-editor of 18 books/conference proceedings and 1 scientific book. His papers are highly cited, as proven from the total number of citations (more than 24500) and the value of the h-index (72) (source: Google Scholar). He was invited as keynote speaker in more than 30 international conferences and workshops. Since 2009 he is a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS). Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, July 1998. Since that he was recipient of many international and national honors and awards, including the recent *IEEE GRSS 2015 Outstanding Service Award* and the *2017 IEEE IGARSS Symposium Prize Paper Award*. Dr. Bruzzone was a Guest Co-Editor of many Special Issues of international journals. He is the co-founder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) series and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003 he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He has been the founder of the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE for which he has been Editor-in-Chief between 2013-2017. Currently he is an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He has been *Distinguished Speaker* of the IEEE Geoscience and Remote Sensing Society between 2012-2016.



**Haishi Zhao** received the M.S. degrees in digital geoscience from the College of Earth Sciences, Jilin University, Changchun, China, in 2017. He is currently pursuing the Ph.D. degree with the College of Computer Science and Technology, Jilin University, Changchun, China.

His research interests include machine learning and hyperspectral image processing.



**Yulei Tan** received the M.S. degrees in digital geoscience from the College of Earth Sciences, Jilin University, Changchun, China, in 2017, where she is currently pursuing the Ph.D. degree with Digital Geoscience Laboratory in the College of Earth Sciences, Jilin University.

Her research interests include hyperspectral image feature selection and classification.



**Renchu Guan** (M'11) received the Ph.D. degree in the College of Computer Science and Technology, Jilin University, Changchun, China, in 2010.

He was a visiting scholar in University of Trento, Italy from 2011 to 2012. He is currently an Associate Professor in the College of Computer Science and Technology, Jilin University, China. He has authored or co-authored over 40 scientific papers in refereed journals and proceedings. His research was featured in the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE, *etc.* His research interests include machine learning and knowledge engineering.