A Novel Ranking-Based Clustering Approach for Hyperspectral Band Selection

Sen Jia, *Member, IEEE*, Guihua Tang, *Student Member, IEEE*, Jiasong Zhu, and Qingquan Li

***Abstract*—Through imaging the same spatial area by hyperspectral sensors at different spectral wavelengths simultaneously, the acquired hyperspectral imagery often contains hundreds of band images, which provide the possibility to accurately analyze and identify a ground object. However, due to the difficulty of obtaining sufficient labeled training samples in practice, the high number of spectral bands unavoidably leads to the problem of a "dimensionality disaster" (also called the Hughes phenomenon), and dimensionality reduction should be applied. Concerning band (or feature) selection, conventional methods choose the representative bands by ranking the bands with defined metrics (such as non-Gaussianity) or by formulating the band selection problem as a clustering procedure. Because of the different but complementary advantages of the two kinds of methods, it can be beneficial to use both methods together to accomplish the band selection task. Recently, a fast density-peak-based clustering (FDPC) algorithm has been proposed. Based on the computation of the local density and the intracluster distance of each point, the product of the two factors is sorted in decreasing order, and cluster centers are recognized as points with anomalously large values; hence, the FDPC algorithm can be considered a ranking-based clustering method. In this paper, the FDPC algorithm has been enhanced to make it suitable for hyperspectral band selection. First, the ranking score of each band is computed by weighting the normalized local density and the intracluster distance rather than equally taking them into account. Second, an exponential-based learning rule is employed to adjust the cutoff threshold for a different number of selected bands, where it is fixed in the FDPC. The proposed approach is thus named the enhanced FDPC (E-FDPC). Furthermore, an effective strategy, which is called the isolated-point-stopping criterion, is developed to automatically determine the appropriate number of bands to be selected. That is, the clustering process will be stopped by the emergence of an isolated point (the only point in one cluster). Experimental results on three real hyperspectral data demonstrate that the bands selected by our E-FDPC approach could achieve higher classification accuracy than the FDPC and other state-of-the-art band selection techniques, whereas the isolated-point-stopping criterion is a reasonable way to determine the preferable number of bands to be selected.**

***Index Terms*—Band selection, density-based clustering, hyperspectral imagery.**

I. Introduction

HYPERSPECTRAL imagery is produced by hyperspectral sensors (its wavelength range is typically 0.4–2.5 μm) at different spectral channels in a single collection. Clearly, the obtained hundreds of images can form a 3-D cube, which can be used for surface material classification and identification according to the continuous spectrum of each pixel [1]–[4]. However, compared with the large amount of spectral bands, it is difficult and laborious to obtain sufficient training samples in practice, which often leads to ill-conditioned problems, such as the Hughes phenomenon (i.e., for a limited number of training samples, the classification accuracy can be deteriorated in a higher dimensional feature space) [5]. Moreover, due to the high correlation among the neighboring bands, which not only increases the computational complexity of the classifiers but also may have a negative impact on the classification accuracy, dimensionality reduction (DR) should be applied as a preprocessing step to discard the redundant information [6]–[8].

Roughly speaking, DR methods can be divided into two main categories, i.e., feature extraction [9], [10] and band (or feature) selection [11], [12]. The former transforms high-dimensional data into a space of fewer dimensions through projections, such as projection pursuit [13], [14], principal component analysis (PCA) [15]–[17], and independent component analysis [18], whereas the latter identifies a subset of the original spectral bands that contains most of the characteristics [19]. Although feature extraction methods can generally achieve higher classification accuracy, the obtained features are not related to the original wavelengths. Alternatively, band selection methods can preserve the relevant original information of the spectral bands, which could significantly simplify the image acquisition process and save the data storage.

According to whether a training set is available or not, band selection algorithms can be further divided into supervised, unsupervised, and semisupervised methods. Specifically, supervised methods select a band subset based on the class separability measure of training samples and their class labels [20], [21], whereas the importance of a band obtained by unsupervised methods is evaluated by various statistical measures

Manuscript received October 28, 2014; revised February 2, 2015 and April 7, 2015; accepted June 20, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61271022, by the Guangdong Foundation of Outstanding Young Teachers in Higher Education Institutions under Grant Yq2013143, by the Shenzhen Scientific Research and Development Funding Program under Grant JCYJ20140418095735628 and Grant JCYJ20140828163633980, and by the Tencent "Rhinoceros Birds" Scientific Research Foundation for Young Teachers of Shenzhen University. *(Corresponding author: Jiasong Zhu.)*

S. Jia and G. Tang are with the Computer Vision Research Institute, College of Computer Science and also with the Shenzhen Key Laboratory of Spatial Information Smarting Sensing and Services, Shenzhen University, Shenzhen 518060, China (e-mail: senjia@szu.edu.cn; tangguihua@email.szu.edu.cn).

J. Zhu and Q. Li are with the Shenzhen Key Laboratory of Spatial Information Smarting Sensing and Services, Shenzhen University, Shenzhen 518060, China (e-mail: zhujiasong@gmail.com; liqq@szu.edu.cn).

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

Digital Object Identifier 10.1109/TGRS.2015.2450759

TABLE I
COMPARISON BETWEEN THE RANKING-BASED AND CLUSTERING-BASED UBS METHODS

	Ranking-based UBS	Clustering-based UBS
Stability of the selected band subset	High (better)	Low
Difficulty of adjusting the parameters	Low (better)	High
Time complexity	Low (better)	High
Correlation between selected bands	High	Low (better)
Classification accuracy	Low	High (better)

or clustering quality assessment [6], [22], which is unrelated to the labeled sample size. Due to the lack of labeled samples, semisupervised methods make use of both labeled samples and the large quantity of unlabeled samples to estimate band relevance [23], [24]. However, different training samples may exhibit different characteristics, leading to the instability of the bands chosen by the supervised or semisupervised methods, which is undesirable for hyperspectral data analysis. Therefore, unsupervised band selection (UBS) is preferable.

Similarly, based on the search strategy of representative bands, various existing UBS approaches can be further classified into two main parts, i.e., the ranking-based [19], [25], [26] and clustering-based methods [6], [27], [28]. As for the ranking-based UBS methods, it first quantifies the importance of each band according to a certain criterion, such as non-Gaussianity [19], [29], variance [25], mutual information [30], etc. Then, a given number of top-ranked bands in the sorted sequence are selected to form the subset. Clearly, the key point in the ranking-based UBS methods is how well the criterion characterizes the distinctiveness of spectral bands. Alternatively, the clustering-based UBS methods are performed in the similarity space defined on the bands. Through partitioning the bands into disjoint groups (clusters) such that bands in the same subset are more similar to each other than bands in different groups, the band close to the centroid of each cluster is picked out to form the chosen subset [31], [32].

In fact, the two kinds of methods have their own merits and shortcomings. Specifically, the ranking-based UBS methods only need to be executed once to obtain the selected band subset, which is stable and has a low computational cost. Nevertheless, the correlation between bands has not been considered during the sorting procedure, leading to the situation where the dependence among the chosen bands is still relatively high. On the contrary, the clustering-based UBS methods aim at finding the most representative bands from the whole set, which actually takes into account the interaction of the bands. Hence, the correlation in the chosen band subset is much less than that of the ranking-based methods, and the classification accuracy is also higher. However, due to the sensitivity of the clustering-based methods to the randomly chosen initial centroids, the selected band subset may be unstable. Meanwhile, in order to get a different number of clusters, the clustering-based UBS methods should be repeated multiple times, leading to the difficulty of adjusting the parameters and an increase in the computational load (see Table I, which summarizes the properties of both kinds of methods). It can be easily found that the two kinds of methods have complementary advantages; thus, both of them should be combined together to carry out the UBS task.

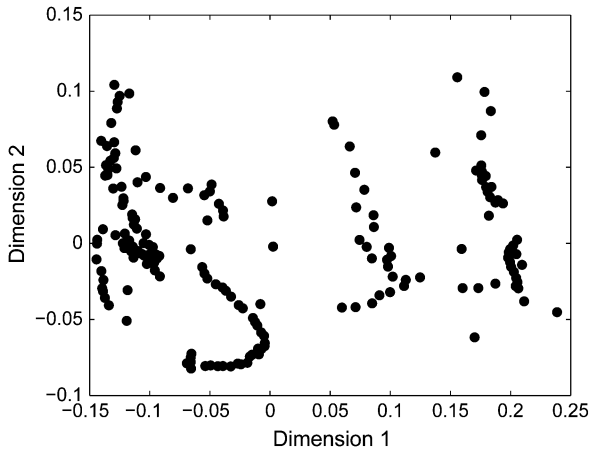

Fig. 1. Two-dimensional multidimensional scaling map based on the pairwise band similarity of the Indian Pines hyperspectral imagery.

Recently, a fast density-peak-based clustering (FDPC) algorithm has been proposed, which identifies cluster centers through investigating the local density and the intracluster distance of each point [33]. Based on a reasonable assumption that cluster centers are surrounded by neighbors with a lower local density and are at a relatively large distance from any points with a higher local density, a simple criterion has been adopted to find the independent density peaks. That is, each data point is ranked by the product of its local density and the distance from the points of higher density, and cluster centers are recognized as points with anomalously large ranking scores. Meanwhile, the FDPC approach only requires measuring the distance between all pairs of data points and does not need to parameterize a probability distribution [34] or a multidimensional density function [35], which is much easier to be computed. Obviously, due to the ranking behavior in the clustering process, the FDPC algorithm can be considered a ranking-based clustering method, which has shown particularly good results in several nonspherical clustering problems, such as face recognition and X-ray feature-based wheat analysis. To illustrate the distribution pattern of hyperspectral band images, the typical Indian Pines hyperspectral data are taken as an example. These data are obtained by the AVIRIS hyperspectral sensor [36], and 185 bands are retained from the original 224 bands after removing the water absorption and the noisy and zero channels (numbered 1–3, 103–112, 148–165, and 217–224) [37]. Detailed information about the data will be given later. Fig. 1 shows the 2-D multidimensional scaling map based on the pairwise distance measure (here, the Euclidean distance is adopted), which produces a low-dimensional configuration to visualize the pairwise similarities and reveals the multivariate structure of the band images [38]. As can be seen in Fig. 1, the distribution of the bands is clearly nonspherical (the bands actually live in a low-dimensional manifold [39]). It is also worth pointing out that, although the number of spectral bands is much smaller than that of spatial pixels, the FDPC is a good option for the hyperspectral UBS problem due to the high

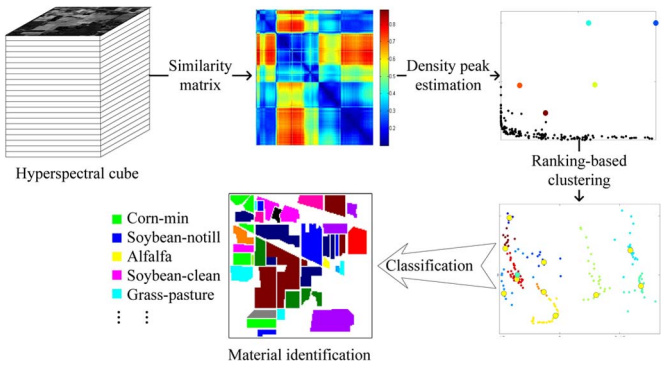

Fig. 2. System block diagram of the proposed ranking-based clustering E-FDPC algorithm for hyperspectral band selection.

structural similarity of hyperspectral imagery [40]. However, the main obstacle in using the FDPC is from two aspects. First, the FDPC is not quite appropriate for solving the problem that requires more cluster centers for subsequent analysis (such as the UBS), which is due to the two factors (the local density and the intracluster distance) not being scaled in a proper way. The other aspect is related to the cutoff threshold parameter that is unchanged with the increase in the cluster number, which eventually degrades the representativeness of the chosen cluster centers [26].

In this paper, through investigating the characteristics of the hyperspectral UBS problem, the FDPC method has been enhanced from the following two aspects. First, instead of taking the local density and the intracluster distance as equally important, the ranking score of each band is computed by weighting the two factors. Specifically, after the points with extremely large scores have been picked out, in order to choose the most informative and nonredundant bands from the rest, the intracluster distance should be more favorably considered than the local density, which could be done by normalizing the two factors and by increasing the weight of the intracluster distance. Second, in order to adjust the important cutoff threshold parameter properly, an exponential-based learning rule is introduced. The proposed approach is thus named the enhanced FDPC (E-FDPC). More importantly, through dividing the whole band set into two separated parts based on the clustering process of the E-FDPC, i.e., the representative and nonrepresentative regions, a simple strategy (called the isolated-point-stopping criterion) through examining the emergence of an isolated point (the only point in one cluster) has been developed to automatically determine the appropriate number of clusters to be selected. Fig. 2 illustrates the schematic of the proposed E-FDPC algorithm.

Our contribution is hence to extend a novel ranking-based clustering method (i.e., the FDPC) to perform the UBS for hyperspectral imagery. Different from the conventional UBS methods that only make use of band ranking or clustering alone, the proposed E-FDPC algorithm extracts the clusters by ranking a density-based measure, and the idea of ranking-based clustering is extremely attractive for the UBS problem due to the complementary nature of the two methodologies. From our knowledge, it is the first time that the two kinds of methods are combined together for the hyperspectral UBS. Moreover, after improving the FDPC method from the two aspects, the proposed E-FDPC algorithm can identify the most representative bands from the whole set. In particular, the chosen bands are actually the cluster centers in different scales for a different number of selected bands, showing the applicability of the proposed E-FDPC method for the UBS problem. As a result, the preferable number of clusters determined by the E-FDPC may be a reasonable indicator on the appropriate number of bands to be selected.

TABLE II
ABBREVIATIONS USED IN THIS PAPER

Abbreviation	Meaning
HSI	Hyperspectral imagery
UBS	Unsupervised band selection
SVM	Support vector machine
ID	Information divergence
MVPCA	Maximum-variance principal component analysis
AP	Affinity propagation
DBSCAN	Density based spatial clustering of applications with noise
FDPC	Fast density peak-based clustering
E-FDPC	Enhanced fast density peak-based clustering

Extensive experiments on three real hyperspectral data sets demonstrate that the proposed E-FDPC approach offers higher performance than several state-of-the-art methods, i.e., two ranking-based methods [information divergence (ID) and maximum-variance PCA (MVPCA)], two clustering-based methods [K-centers and affinity propagation (AP)], and two other density-based clustering methods (density-based spatial clustering of applications with noise (DBSCAN) [41], [42] and the FDPC), in terms of both the classification accuracy and the stability of the chosen band subset. Meanwhile, it has been also shown that the methodology of automatically determining the number of selected bands could achieve the balance between the classification accuracy and the data DR. Table II summarizes the abbreviations used throughout this paper.

The remainder of this paper is organized as follows. In Section II, we introduce some popular related band selection methods, including ranking-based and clustering-based methods, and Section III will present the recently proposed FDPC clustering algorithm and its enhanced version (E-FDPC), which could choose more reliable bands by using the strategy of weighting the ranking score for each band and shrinking the cutoff threshold gradually. Furthermore, the advantages of the proposed E-FDPC approach over the other methods are verified. Experimental results on real hyperspectral remote sensing data are showed in Section IV. Finally, conclusions are given in Section V.

II. RELATED WORKS

In order to facilitate the following description, some notations are introduced. Let $\mathbf{R} \in \mathbb{R}^{M\times N\times L}$ denote the hyperspectral image cube that contains two spatial dimensions M and N (pixels) and one spectral dimension L (wavelength). $\mathbf{R}_l \in \mathbb{R}^{M\times N}$ is the lth band image, and $\mathbf{R}_{mn} \in \mathbb{R}^{L}$ is the reflectance spectral signature at spatial location (m, n), which provides the ability to identify various materials in the ground scene.

Considering the hyperspectral UBS problem, most algorithms start with the construction of a similarity matrix $\mathbf{S} \in \mathbb{R}^{L\times L}$, in which entry $\mathbf{S}_{ij}$ indicates the similarity between the ith and jth bands. A common choice for the similarity computation is the Euclidean distance as follows:

$$\mathbf{S}_{ij} = \|\mathbf{R}_i - \mathbf{R}_j\|^2 = \sum_{m,n=1}^{M,N} (\mathbf{R}_{mni} - \mathbf{R}_{mnj})^2. \quad (1)$$

Based on similarity matrix $\mathbf{S}$, two main kinds of UBS methods, i.e., the ranking-based and clustering-based UBS methods, have been developed.

A. Ranking-Based UBS

The ID method [19] uses a divergence criterion [43] to assess the discriminative potential of each band. Since the class labels are not known in the UBS, the non-Gaussianity of a band image is usually used instead of its real discriminative ability of classification. The Kurtosis of the probability distribution [44] of a band image and the Kullback–Leibler divergence [43] to its associated Gaussian probability distribution are two popular measures of non-Gaussianity. After evaluating the difference between the probability distribution of each image and its associated Gaussian probability distribution, the whole band set can be sorted according to the resulting scores, and those with higher values are selected. However, the main disadvantage of the ID method is that the information concerning the spectral correlation is ignored, which causes the selected bands to still have a strong correlation, i.e., some bands can well represent the others, and some bands can be removed without a significant loss of information.

The MVPCA [25] is a joint band prioritization and band decorrelation approach. It ranks the bands by a criterion that comprises the importance of an individual band and its correlation with other bands. First, a loading factor matrix is constructed from the eigenvalues and eigenvectors obtained by the PCA transformation of image cube $\mathbf{R}$. After removing the bands whose discriminative capabilities are below a prescribed threshold, the remaining bands are decreasingly sorted according to their discriminative capability, which is evaluated by the variance-based band power ratio defined as the variance of this band divided by that of all bands. Second, all sorted bands are orderly decided to be kept or removed by a correlation criterion, which is measured by the Kullback–Leibler divergence. If the current visited band is highly correlated to any other band that has been decided to remain, it will be removed. Although redundant bands can be removed by applying band decorrelation, the band that contains the important information of some materials may be eliminated because of its low variance.

B. Clustering-Based UBS

On the other hand, the hyperspectral UBS problem can be conceptually formulated as a data clustering procedure, which partitions the data set into groups of similar objects (clusters) without any class label information [45], and has attracted increasing attention in the field of band selection. In clustering-based band selection approaches, each band is considered a data point, and the bands are separated into several clusters on the basis of similarity matrix $\mathbf{S}$. Through selecting a band in a cluster to represent all the bands in this cluster, the band decorrelation can be accomplished. If the similarity measure is the same (such as the Euclidean distance used here), the performance of the clustering-based methods is better than that of the ranking-based methods [6].

In the popular K-centers clustering technique, clusters are groups of data characterized by a small distance to the cluster center [46]. It begins with an initial set of randomly selected bands, and the objective function (typically the sum of the distance to a set of putative cluster centers) is iteratively optimized until the best cluster center candidates are found. Because the K-centers clustering technique is quite sensitive to the initial selection of bands, it usually should be replicated as many times as necessary with different initializations in an attempt to find a good solution. However, this only works well when the number of clusters is small, and chances are good that at least one random initialization is close to a good solution. To make matters worse, the whole clustering process should be repeated multiple times to obtain a different number of selected bands, thus having a higher computational cost than the ranking-based methods (which only need to be computed once). Moreover, the underlying concept of K-centers is to find the smallest radius such that the entire data set is contained within k (the given number of selected bands) balls of the radius, indicating that the method is not able to detect nonspherical clusters [47], which is just the case of the hyperspectral band distribution (see Fig. 1).

Recently, an exemplar-based clustering algorithm termed AP has been proposed, which initially considers all data points potential cluster centers (exemplars) and then exchanges messages between data points until a good set of exemplars and clusters emerges [32]. Clusters are formed by assigning each data point to its most similar exemplar. The AP approach has several advantages over the existing clustering-based methods, such as it is not sensitive to the initial selection of exemplars (because it considers all data points potential exemplars), the *a priori* information can be easily integrated into the AP (because the importance of each individual band can be tuned at first through its parameter of "preference" that indicates the possibility of a data point being exemplar), etc. [48]. The AP approach has been employed for hyperspectral band selection, and the generated bands have higher importance and less correlation/similarity than those obtained by several other band selection methods [22], [49]. However, due to the nature of the clustering-based methods, the band set chosen by the AP with a small number is not necessarily contained in the set with a large number, leading to the instability of the selected band set. In addition, although the magnitude of preference can be used to control the number of clusters, how to accurately and rapidly choose the parameter to attain a certain number of clusters is not a trivial task.

III. Ranking-Based Clustering Approach for Hyperspectral UBS

After introducing several popular ranking-based and clustering-based UBS methods in the previous section, a novel

ranking-based clustering algorithm, which is named the FDPC, is first presented in this section, and then, the E-FDPC approach is proposed. Finally, the main advantages of the proposed E-FDPC approach over the FDPC and other UBS methods are investigated.

A. FDPC Algorithm

By defining the cluster centers as the local maxima in the density of data points [50], the FDPC algorithm has presented a simple and effective criterion to rank the data points and automatically find the cluster centers. That is, for each band image i, $1 \leq i \leq L$, which can be considered a data point, its representativeness is decided by two factors, i.e., local density ρ_i and distance δ_i from points of higher density. Both of these quantities only depend on similarity matrix $\mathbf{S}$, which has been scaled down by the following transformation:

$$\mathbf{D}_{ij} = \frac{\sqrt{\mathbf{S}_{ij}}}{L}. \tag{2}$$

Then, the local density ρ_i of point i is defined as

$$\rho_i = \sum_j \chi(\mathbf{D}_{ij} - d_c) \tag{3}$$

where d_c is a cutoff distance used to keep a region for each data point, and $\chi(x)$ is a function to find which data point is its local neighbors, i.e.,

$$\chi(x) = \begin{cases} 1, & \text{if } x < 0 \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Thus, ρ_i is equal to the number of points that are closer than d_c to point i. Likewise, δ_i is measured by computing the minimum distance between point i and any other point with a higher density, i.e.,

$$\delta_i = \min_{j:\rho_j > \rho_i} (\mathbf{D}_{ij}). \tag{5}$$

Meanwhile, for the data point whose density is the highest among all the points, i.e., the global maximum in the density, δ_i should be much larger than the typical nearest neighbor distance, which is

$$\delta_i = \max_j(\mathbf{D}_{ij}), \quad \text{if } \rho_i = \max(\rho). \tag{6}$$

Thus, the cluster centers can be recognized as points with anomalously large values of δ, which is the core principle of the FDPC algorithm. To illustrate the idea of the method, Fig. 3(a) shows the plot of δ as a function of ρ for each band in the Indian Pines hyperspectral data, which is called a "decision graph." Clearly, the only points of high δ and relatively high ρ are the cluster centers (here, five representative points are picked out). After the cluster centers have been found (denoted by $\mathcal{C}$), each remaining point is assigned to the same cluster as its nearest neighbor of higher density, i.e.,

$$\mathcal{A}_j = \min_{i \in \mathcal{C}} \mathbf{D}_{ij}, \quad j \in \{1, \ldots, L\} - \mathcal{C}. \tag{7}$$

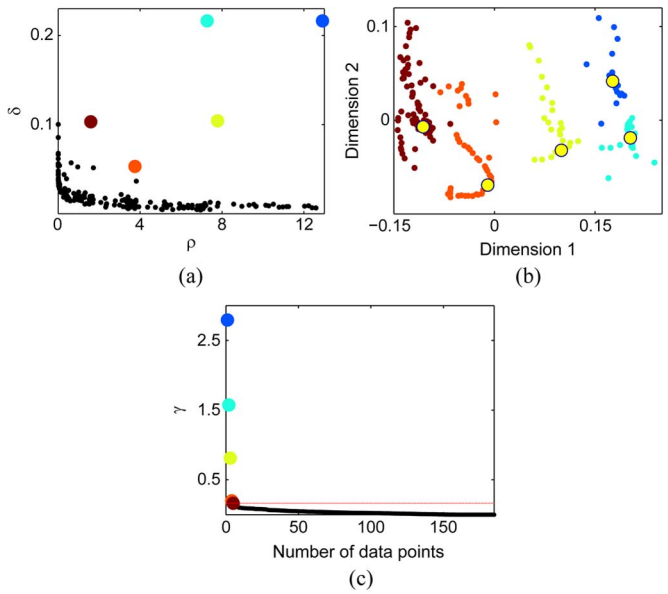

Fig. 3. Cluster analysis of the Indian Pines hyperspectral data. (a) Decision graph obtained by the FDPC, with the centers colored by cluster. (b) Band distribution in a 2-D space. The points are colored according to the cluster to which they are assigned, and the big yellow circles represent the cluster centers. (c) Value of $\gamma_i = \rho_i \delta_i$ in decreasing order.

Fig. 3(b) displays the band distribution in a 2-D space, in which points are colored according to the cluster to which they are assigned, and the big yellow circles represent the five chosen cluster centers. It can be easily observed from this figure that the cluster centers identified via the FDPC are able to reveal the nonspherical structure of this set. Furthermore, to make the clustering process more convenient, the two factors can be multiplied together to obtain a score for each point as follows:

$$\gamma_i = \rho_i \times \delta_i \tag{8}$$

and then, the importance of the data points can be determined by γ. Fig. 3(c) provides the plot of γ sorted in decreasing order, in which the cluster centers correspond to the large values. Therefore, the FDPC algorithm can be considered a ranking-based clustering method, which could make the chosen points both representative and stable. In addition, due to the simple calculation involved in identifying the cluster centers and the fact that the cluster assignment can be performed in a single step, the FDPC algorithm is very time efficient. Finally, it is worth pointing out that cutoff parameter d_c is often empirically chosen so that the average number of neighbors is around 1% to 2% of the total number of points in the data set.

B. E-FDPC Algorithm

As described earlier, the FDPC is a rapid and effective method to find the cluster centers according to local density ρ and intracluster distance δ. Inspired by the idea of using density-peak-based ranking scores to carry out data clustering, not only the stability of the chosen band subset can be guaranteed but also the difficulty of tuning parameters involved in the clustering process (such as AP) can be avoided; thus, the FDPC is a good candidate for hyperspectral band selection.

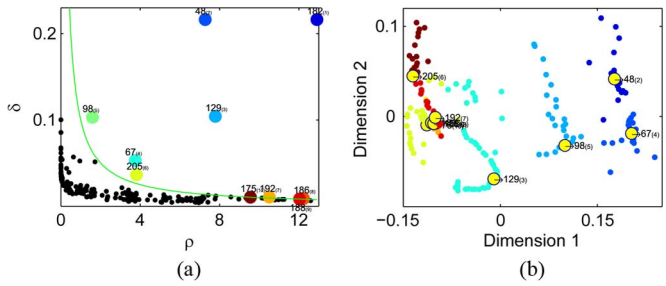

Fig. 4. Band selection analysis of the Indian Pines hyperspectral data by the FDPC. (a) Decision graph of the FDPC, in which ten bands are selected (the numbers denote the band channels, whereas those in the parentheses indicate the decreasing order of γ computed by (8). The green curve represents an inverse hyperbolic function in the form of $\rho\delta = \gamma_{10}$, where γ_{10} is the ranking score of the tenth selected band). (b) Clustering results by the FDPC with ten clusters (the big yellow circles represent the cluster centers).

However, it is infeasible to directly apply the FDPC method to cope with the UBS task of hyperspectral imagery due to the different goals of clustering and band selection. In particular, cluster analysis is aimed at grouping elements into categories on the basis of their similarity, whereas band selection aims to find a certain number of discriminative and nonredundant bands for the DR and subsequent analysis, such as pixel-oriented classification. In this respect, the appropriate number of selected bands is typically much larger than that of the clusters in the data. Concerning the FDPC clustering algorithm, when the number of selected bands k is small, the representative bands are obviously the cluster centers identified by the FDPC. As k increases, the FDPC may choose the bands with large ρ and relatively small δ, i.e., the points around the cluster centers, which are highly correlated to the bands already selected. To illustrate, Fig. 4 shows the band selection analysis of the Indian Pines hyperspectral data by the FDPC. It can be clearly seen in Fig. 4(a) that, after six bands (numbered 182, 48, 129, 67, 98, and 205, whose values of ρ and δ are extremely large) have been sequentially picked out (for cluster analysis, the FDPC method can be stopped), the rest of the chosen points (numbered 192, 186, 188, and 175) are strongly dominated by local density ρ and concentrated in the bottom right area (here, k is equal to 10), which are mostly around the first chosen band 182 [as depicted in Fig. 4(b)], leading to the situation where a high degree of redundancy exists in the chosen bands, and their representativeness can be considerably weakened. Therefore, the FDPC algorithm should be customized to be suitable for the UBS problem, and the proposed approach is named the E-FDPC, which can be explained in the following two aspects.

Weighted Ranking Score Scheme for Each Band: Although the idea of combining local density ρ and distance δ from points of higher density is far more reliable than using either of the two factors alone, how to effectively integrate the two factors for hyperspectral band selection is still an unsolved issue. The weight value is important to promote the performance of the algorithm, and the ranking score of each band should be computed through weighting the two factors. Specifically, after the points with anomalously large scores (considered the cluster centers) have been found by the FDPC, the following chosen bands should be those with large δ and relatively small ρ (corresponding to the lower left area in the decision graph), i.e., the points that are away from the chosen points and are the centers in smaller regions, which means that parameter δ should be given more weight than ρ. Based on this observation, instead of taking δ and ρ into account equally, the proposed E-FDPC algorithm increases the weight of δ through a normalization and square product process. Before detailing the improvement, it is worth pointing out that a Gaussian kernel function is adopted to estimate local density ρ_i for each band image to decrease the negative impact of the statistical errors caused by the small number of bands, i.e.,

$$\rho_i = \sum_{j=1, j\neq i}^{L} \exp\left(-\left(\frac{\mathbf{D}_{ij}}{d_c}\right)^2\right). \tag{9}$$

After δ has been computed by (5) and (6), it is normalized to the scale of [0, 1], which can be done by

$$\delta = (\delta - \delta_{\min})./(\delta_{\max} - \delta_{\min}) \tag{10}$$

where $./$ represents the elementwise division operator, and ρ is normalized in the same way. Because the range of ρ is much larger than that of δ [which can be seen in Fig. 4(a)], the weight of δ has been significantly increased after the normalization. In order to compensate ρ for the loss of weight during the normalization procedure, ranking score γ_i for any band i is finally obtained by

$$\gamma_i = \rho_i \times \delta_i^2. \tag{11}$$

Exponential-Based Heuristic Rule for Cutoff Threshold d_c: It can be easily found from (3) that the local density ρ_i of any data point is computed based on cutoff threshold d_c, which has a direct impact on the results of the cluster centers. In particular, if d_c is set too big, the overlapping neighborhood can even include the data from other clusters, and every point has a similar heavy density neighborhood; if d_c is set too small, every point has a similar sparse density neighborhood. In both cases, estimated density ρ has less significant discriminative power. As far as the band selection problem of hyperspectral imagery is considered, with the increase in the number of selected bands k, the connected region of each band should be correspondingly smaller, i.e., d_c should be gradually decreased. Therefore, an exponential-based heuristic rule is introduced to regularize the local region for each band based on k. Concretely, for each k, parameter d_c is updated by

$$d_c = d_{\text{ini}}/\exp(k/L) \tag{12}$$

where L is the spectral number of the whole data, and d_{ini} is the initial value of the cutoff threshold. According to the FDPC, here, d_{ini} is empirically defined as the value at the $2\% \times L \times (L-1)$th position in $\mathbf{D}_s$, where $\mathbf{D}_s$ is the sorted sequence of $\mathbf{D}$ from low to high (except the diagonal elements that are zero). In comparison, Fig. 5 shows the band selection analysis of the Indian Pines hyperspectral data by the E-FDPC. It can be observed in Fig. 5(a) that, after four bands (numbered 182, 48, 129, and 99) have been sequentially picked out by the E-FDPC, as expected, the rest of the selected

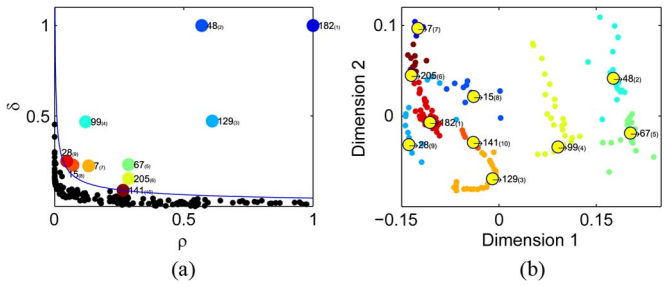

Fig. 5. Band selection analysis of the Indian Pines hyperspectral data by the E-FDPC. (a) Decision graph of the E-FDPC, in which ten bands are selected (the numbers denote the band channels, whereas those in the parentheses indicate the decreasing order of γ computed by the E-FDPC. The blue curve represents an inverse hyperbolic function in the form of $\rho\delta^2 = \gamma_{10}$, where γ_{10} is the ranking score of the tenth selected band). (b) Clustering results by the E-FDPC with ten clusters (the big yellow circles represent the cluster centers).

points (numbered 67, 205, 7, 15, 28, and 141) are concentrated in the lower left area. Correspondingly, Fig. 5(b) shows the clustering results by the E-FDPC with ten chosen bands. As can be seen, the points selected by the E-FDPC are scattered in a uniform manner and can be interpreted as the cluster centers in small regions, which is in line with the goal of the band selection for choosing independent and representative bands from the original hyperspectral data. Therefore, the proposed E-FDPC method can be more advantageous than the FDPC.

Finally, the pseudocode of the proposed E-FDPC band selection approach is outlined in Algorithm 1, in which the improvement of the E-FDPC over the FDPC is the rows with italic comments. It can be seen that the calculation time of the proposed E-FDPC approach can be divided into two parts. One part is related to the computation of the similarity matrix $\mathbf{S}$ that is used as the input of the subsequent clustering process, and the complexity is $O(2L^2MN)$. The other part is consumed for calculating the ranking scores of each band and the label assignment procedure, i.e., approximately $O(L^2)$ with a constant multiplier. Note that matrix $\mathbf{S}$ only needs to be computed once, and the E-FDPC clustering process is not affected by the number of pixels in the scene; thus, the approach presented here is computationally efficient, even for the hyperspectral imagery with a large size.

Algorithm 1 The E-FDPC Algorithm

INPUT: Scaled similarity matrix $\mathbf{D} \in \mathbb{R}^{L\times L}$, the number of selected bands k, and the initial value of d_c (denoted as d_{ini});
OUTPUT: The indexes of the selected bands $\mathcal{C} \in \mathbb{R}^k$ and the labels for each band $\mathcal{A} \in \mathbb{R}^L$.
BEGIN
$d_c = d_{\text{ini}}/\exp(k/L)$; *% the exponential-based heuristic rule*
for $i = 1$ *to* L **do**
  $\rho_i = \sum_{j=1,j\neq i}^{L} \exp(-(\mathbf{D}_{ij}/d_c)^2)$; *% a Gaussian kernel function is used to estimate* ρ
end for
$[\mathcal{V}, \mathcal{I}] = \text{sort}(\rho, \text{'descend'})$;
$\delta_{\mathcal{I}_1} = -1$;
for $i = 2$ *to* L **do**
  $\delta_{\mathcal{I}_i} = \max(\mathbf{D}_{ij})$;
  for $j = 1$ *to* $i-1$ **do**
    if $\mathbf{D}_{\mathcal{I}_i\mathcal{I}_j} < \delta_{\mathcal{I}_i}$ **then**
      $\delta_{\mathcal{I}_i} = \mathbf{D}_{\mathcal{I}_i\mathcal{I}_j}$;
    end if
  end for
end for
$\delta_{\mathcal{I}_1} = \max(\delta)$;
$\rho = (\rho - \rho_{\min})./(\rho_{\max} - \rho_{\min})$; *% normalize the two factors*
$\delta = (\delta - \delta_{\min})./(\delta_{\max} - \delta_{\min})$;
$\gamma = \rho. \times \delta^2$; *% a square weight measure is applied on each element of* δ
$[\mathcal{V}, \mathcal{I}] = \text{sort}(\gamma, \text{'descend'})$;
$\mathcal{C} = \mathcal{I}(1:k)$;
for $i = 1$ *to* L **do**
  $j = \mathcal{I}_i$;
  $\mathcal{A}_j = \min_{c\in\mathcal{C}} \mathbf{D}_{cj}$;
end for
END

C. Advantages of E-FDPC for Hyperspectral Band Selection

In this section, the discriminative ability, correlation, and stability of the selected bands have been examined to demonstrate the necessity and advantages of combining the ranking-based and clustering-based ideas together (i.e., the proposed E-FDPC approach) to achieve the hyperspectral UBS. Moreover, a simple and effective strategy to automatically determine an appropriate number of chosen bands, which is called the isolated-point-stopping criterion, has been developed.

First, to verify the discriminative power of the chosen bands, the pixel-oriented classification is considered. Similarly, the Indian Pines hyperspectral data set is used, which contains 10 366 labeled samples with 16 different classes (the detailed information of the data and the labeled samples is given in Section IV). To build the training set, a fixed number (here, the number is ten) of the labeled samples from each class are selected. In addition, the remaining samples are then used as the test set. With respect to the classification algorithms, the K-nearest neighborhood (KNN) is applied, and the parameter of the number of neighbors in the KNN is set to be three. Additionally, to assess the classification results, the overall accuracy (OA) and the kappa coefficient (κ) are used as the measures of accuracy. More precisely, the OA is defined as

$$\text{OA} = \frac{\text{number of correctly classified samples}}{\text{number of test samples}}. \tag{13}$$

κ is a statistical measure of the degree of agreement, which is computed by accounting for all the elements in the confusion matrix [51]. The greater the κ, the more accurate the results. Obviously, higher classification accuracy with less number of selected bands is preferable. To reduce the influence of random effects, the training and test data sets are randomly chosen

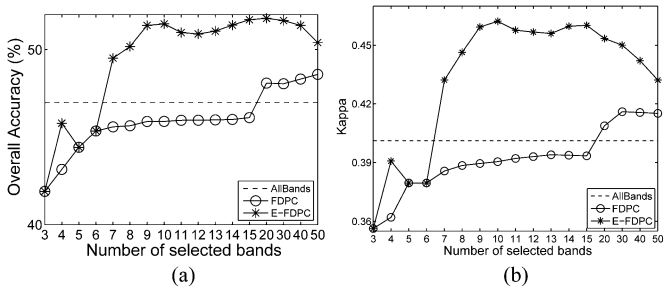

Fig. 6. Classification performance using the KNN versus the bands selected by the E-FDPC and the FDPC on the Indian Pines hyperspectral data. (a) OA. (b) Kappa.

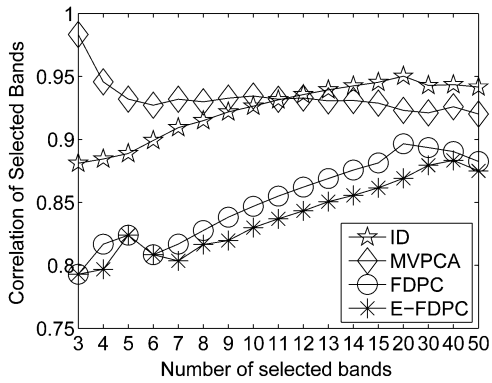

Fig. 7. Correlation measures between the selected bands that are produced by the ID, MVPCA, FDPC, and E-FDPC methods.

from the data ten times, and the average accuracy is computed. Fig. 6 displays the classification results using the KNN versus the bands selected by the E-FDPC and the FDPC (the results of other band selection methods are given in Section IV). To investigate the impact of a different number of selected bands on the classification accuracy, the number ranges from 3 to 50, as represented by the x-axis (here, the result with all bands is also given, with the legend "AllBands"). Evidently, the classification performance of the E-FDPC is more accurate than that of the FDPC when the number of selected bands is larger than six, validating the discriminative power of the bands chosen by the E-FDPC for classification.

Second, the correlation of the bands selected by the E-FDPC is studied. Just as the other ranking-based band selection methods, such as the ID and MVPCA methods, the correlation between the selected bands by the FDPC and the E-FDPC increases with the increase in the number of selected bands k. Fig. 7 gives the correlation between the selected bands that are produced by the different ranking-based band selection methods, which is computed by

$$\text{Cor} = 1 - \frac{2}{k(k-1)} \sum_{i,j=1,i<j}^{k} \mathbf{D}_{\mathcal{C}_i \mathcal{C}_j} \tag{14}$$

where $\mathcal{C}$ are the indexes of the selected bands. The smaller the value, the lower the correlation. Obviously, whatever k is taken, the bands selected by the ID and MVPCA methods always have the largest correlation, which is in accordance with

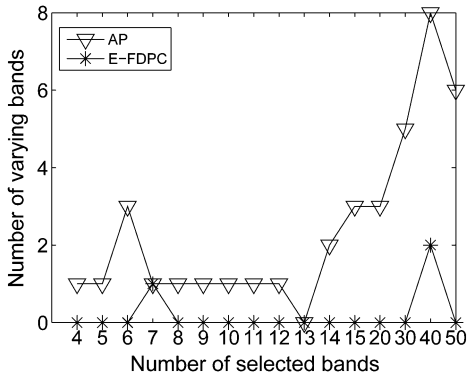

Fig. 8. Stability comparison between the AP and E-FDPC methods on the Indian Pines hyperspectral data.

the character that the spectral correlation (ID) or bands with a low variance (MVPCA) are ignored during the band selection procedure. Moreover, the correlation of the selected bands by the E-FDPC is lower than that by the FDPC, illustrating the representativeness of the bands chosen by the E-FDPC [note that the representativeness of the chosen bands by the E-FDPC can be also easily observed in Fig. 5(b)].

Third, the stability of the proposed E-FDPC method is investigated, which characterizes the varying degrees of the chosen bands and is important for the applicability of band selection methods. Clearly, the stability of the clustering-based methods is not as good as that of the ranking-based methods because the former should be applied multiple times to obtain a different number of selected bands, whereas the latter can be only executed once. As for the E-FDPC, due to the tuning of cutoff parameter d_c, which is related to the number of chosen bands k, the proposed E-FDPC method should be also carried out many times, and then, the stability of the chosen band subset can be degraded. For a given k, e.g., $k > 3$, the varying degree is measured by calculating the difference between the adjacent selected band subset, i.e.,

$$\nu = \text{length}\left(\text{setdiff}(\mathcal{C}_k, \mathcal{C}_{k-1})\right) - 1 \tag{15}$$

where $\mathcal{C}_k$ and $\mathcal{C}_{k-1}$ are the indexes of the selected bands containing k and $k-1$ elements, respectively, the setdiff operator returns the values in $\mathcal{C}_k$ that are not in $\mathcal{C}_{k-1}$, and the length operator returns the number of elements in the vector. Obviously, if $\mathcal{C}_{k-1}$ is totally contained in $\mathcal{C}_k$, then ν is equal to zero, and the stability of the chosen band subset can be guaranteed. The smaller the ν, the more stable the results. Fig. 8 shows the stability comparison between the AP and E-FDPC methods on the Indian Pines hyperspectral data. It can be observed in this figure that, as a typical clustering algorithm, the numbers of the varying bands obtained by the AP method are high. On the contrary, due to the ranking nature of the E-FDPC (inherited from the FDPC), the numbers of varying bands are equal to zeroes in most cases, verifying the strong stability of the proposed E-FDPC band selection approach.

Finally, a simple strategy, which is called the isolated-point-stopping criterion, is proposed to automatically determine a preferable number of chosen bands. In particular, as the number

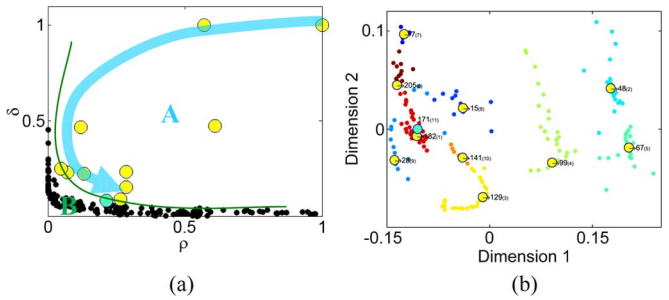

Fig. 9. Illustration of the isolated-point-stopping criterion for automatically determining the appropriate number of the selected bands by the E-FDPC approach on the Indian Pines hyperspectral data. (a) Sketch map of the clustering trends of the E-FDPC. After 10 points (the circles in yellow) have been picked out, the green point is the 11th chosen point that is an isolated point (the cluster only contains a single element). The blue curve with an arrow displays the clustering process of the E-FDPC for band selection. The green curve has divided the decision graph into two parts, i.e., the A region containing the cluster centers found by the E-FDPC and the B region containing the elements in the clusters. (b) Clustering results by the E-FDPC with 11 clusters (the big yellow and green circles represent the cluster centers). After picking out 10 points, the 11th point (band 171) is an isolated point that is the only element in one cluster.

of chosen bands k increases, instead of choosing the points with a large local density ρ, the proposed E-FDPC algorithm is prone to select those points with a large intracluster distance δ and a relative low ρ in the decision graph [as illustrated in Fig. 5(a)], which are actually the cluster centers in small regions [as shown in Fig. 5(b)]. To illustrate, Fig. 9 demonstrates the main idea of the automatic number determination of the selected bands on the Indian Pines data set. Specifically, according to the clustering process of the E-FDPC in Fig. 9(a) (the blue curve with an arrow), the decision graph can be divided into two parts by the green curve, i.e., the representative A region (containing the cluster centers) and the nonrepresentative B region (containing the elements in the clusters); thus, it can be a valid way to choose the bands in region A and eliminate those in region B for optimal band selection. Fortunately, the boundary between the two regions can be easily identified by *the emergence of an isolated point*, which is the only element in one cluster; thus, the E-FDPC algorithm should stop adding a new point to the selected band subset if an isolated point has been found. Fig. 9(b) shows the clustering results by the E-FDPC with 11 clusters. It can be easily observed that the 11th point (band 171) is an isolated point; hence, a proper number of selected bands for the Indian Pines data set can be 10. The rationality of the chosen number can be found in Fig. 6, in which the classification accuracy with 10 selected bands is better than that of the adjacent bands and is comparable with those with more than 15 selected bands. Therefore, the isolated-point-stopping criterion may be a reasonable indicator on the appropriate number of bands to be selected, which can achieve the balance between the classification accuracy and the DR. Fig. 10 and Algorithm 2 depict the block diagram and the procedure of automatically determining the preferable number of chosen bands by the E-FDPC, respectively. The following experimental results further demonstrate the effectiveness of the E-FDPC approach for hyperspectral band selection.

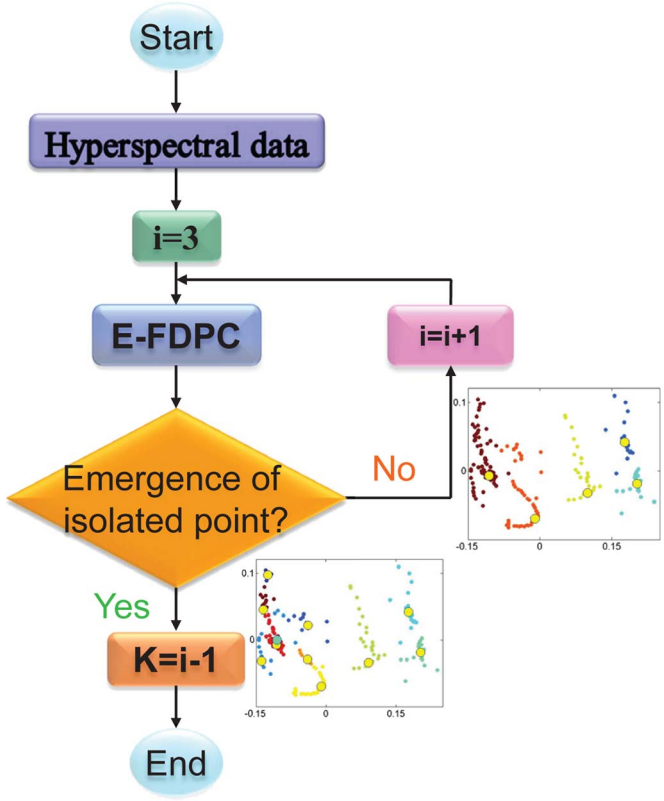

Fig. 10. Block diagram of automatically determining the preferable number of chosen bands by the E-FDPC algorithm.

Algorithm 2 The Procedure of Automatically Determining the Preferable Number of Chosen Bands by the E-FDPC algorithm

INPUT: Hyperspectral data cube $\mathbf{R} \in \mathbb{R}^{M\times N\times L}$;
OUTPUT: The optimal number of selected bands K.
BEGIN
Compute similarity matrix $\mathbf{S}$ by (1), and then, $\mathbf{S}$ is scaled down to $\mathbf{D}$ by (2);
Sort each entry in $\mathbf{D}$ from low to high (except the diagonal elements that are zero), which is denoted as $\mathbf{D}_s$, and then, the initial value of d_c is set as the value at the $2\% \times L \times (L-1)$th position in $\mathbf{D}_s$, which is denoted as d_{ini};
for $i = 3$ to L **do**
  $(\mathcal{C}, \mathcal{A}) = \text{E-FDPC}(\mathbf{D}, i, d_{\text{ini}})$;
  for $j = 1$ to length$(\mathcal{C})$ **do**
    $\Lambda(j) = \text{length}(\text{find}(\mathcal{A} == \mathcal{C}_j))$;
  end for
  if $\min(\Lambda) == 1$ **then**
    $K = i - 1$; % *eliminate the isolated point*
    break;
  end if
end for
END

IV. Experimental Results

Having presented our method in the previous section, we now turn our attention to demonstrating its utility for the

purpose of classification. Here, three real-world hyperspectral remote sensing data sets with different spatial resolutions have been employed. The experiments are conducted by comparing the E-FDPC algorithm with four popular band selection techniques, which include two ranking-based techniques (the ID and MVPCA methods) and two clustering-based techniques (the K-centers and AP methods). Meanwhile, the DBSCAN algorithm [41], which is a typical density-based clustering method and has shown to be effective in discovering clusters of arbitrary shape, and the FDPC are included in comparison. With respect to the classification algorithms, the KNN and a support vector machine (SVM) have demonstrated their status as the "state of the art" in the classification problem as applied to the data sets. In our experiments, the parameter of the number of neighbors in the KNN was set to be three, and the linear kernel and one-against-all scheme in the SVM was used for multiclass classification. In addition, the C parameter of the SVM is estimated by a tenfold cross validation.

The performance of the compared band selection methods is evaluated with different sample sizes, i.e., a fixed number of samples of the available labeled samples are selected from each class to form the training set. The remaining samples are then used as the test set for evaluation. Each experiment is repeated ten times with different training sets to reduce the influence of random effects, and the average results are reported. Likewise, the OA defined by (13) and the kappa coefficient (κ) measures are used to assess the classification results.

A. Indian Pines Data Set

The first real-world data set to be used is the commonly used Indian Pines data set acquired by the AVIRIS instrument over the agricultural area of Northwestern Indiana in 1992, which has a spatial dimension of 145 $\times$ 145 and 224 spectral bands. The spatial resolution of the data is 20 m per pixel. As aforementioned in Section I, after discarding 4 zero bands and the 35 lower SNR bands affected by atmospheric absorption, 185 channels are preserved. The data set contains 10 366 labeled pixels and 16 ground-truth classes, most of which are different types of crops (see Table III). Fig. 11 shows the ground-truth map containing 16 mutually exclusive land-cover classes.

Fig. 12 illustrates the classification results of the KNN and SVM classifiers using the OA and kappa measures, in which ten samples from each class are randomly picked out to form the training set. To investigate the impact of a different number of selected bands on the classification accuracy, the number ranges from 3 to 50, as represented by the x-axis. The results with all bands are also given for comparison, with the legend AllBands. In this figure, it can be seen that the results of the five clustering-based methods (the K-centers, AP, DBSCAN, FDPC, and E-FDPC methods) are better than those of the two ranking-based methods (the ID and MVPCA methods) in most cases, which are mainly due to the high correlation of the bands selected by the ranking-based methods. Compared with the five clustering-based methods, the E-FDPC delivers the most stable and accurate results, which are even higher than those of AllBands (when the number of chosen bands is larger than six for the KNN and eight for the SVM), proving the effectiveness of the proposed E-FDPC approach for hyperspectral band selection.

Based on the automatic number determination strategy of the selected bands by the E-FDPC given in the previous section, it has been shown that ten could be a proper number of selected bands for the Indian Pines data, and the chosen bands

TABLE III
LAND-COVER CLASSES WITH A NUMBER OF SAMPLES FOR THE INDIAN PINES DATA SET

Class	Land Cover Type	No. of Samples
$C1$	Alfalfa	54
$C2$	Buildings-Grass-Trees-Drives	380
$C3$	Corn	234
$C4$	Corn-min Till	834
$C5$	Corn-no Till	1434
$C6$	Grass-pasture-mowed	26
$C7$	Grass/Pasture	497
$C8$	Grass/Trees	747
$C9$	Hay-windrowed	489
$C10$	Oats	20
$C11$	Soybean-clean Till	614
$C12$	Soybean-min Till	2468
$C13$	Soybean-no Till	968
$C14$	Stone-Steel-Towers	95
$C15$	Wheat	212
$C16$	Woods	1294
	Total	10366

Fig. 11. Ground-truth map of the Indian Pines data set (16 land-cover classes).

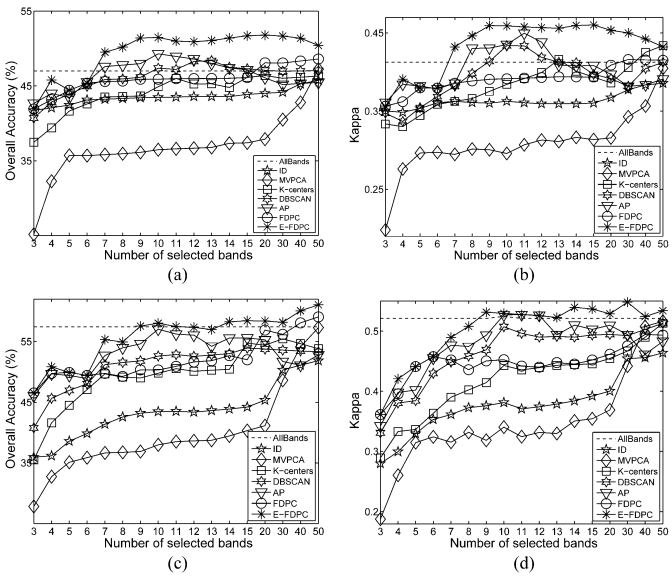

Fig. 12. Performance versus the selected bands of the different methods on the Indian Pines data set. (a) KNN using the OA. (b) KNN using kappa. (c) SVM using the OA. (d) SVM using kappa.

TABLE IV
COMPARISON OF THE SELECTED BANDS USING THE DIFFERENT BAND SELECTION METHODS

Band selection methods	Ten selected bands
ID	88/35/7/12/175/183/189/174/186/178
MVPCA	113/115/117/31/29/30/116/28/114/118
K-centers	5/10/60/85/92/94/99/115/145/191
DBSCAN	7/14/26/30/53/87/89/95/138/197
AP	15/26/48/67/95/118/129/141/182/209
FDPC	182/48/129/67/98/205/192/186/188/175
E-FDPC	182/48/129/99/67/205/7/15/28/141

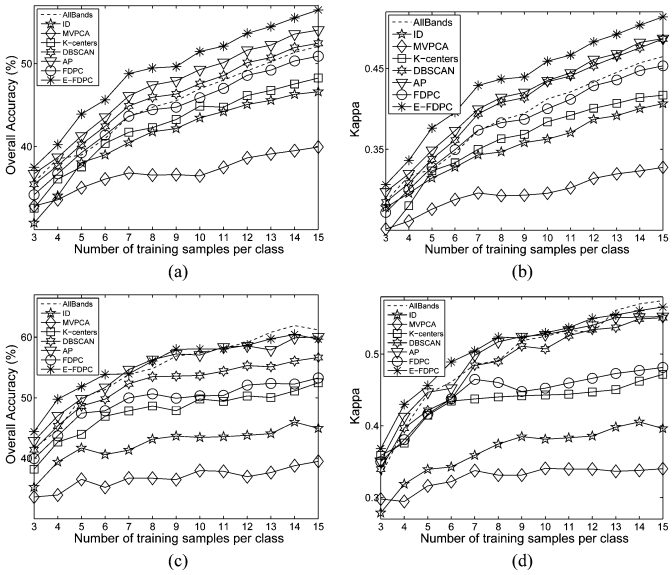

Fig. 13. Classification accuracy versus the different number of training samples per class on the Indian Pines data set. (a) KNN using the OA. (b) KNN using kappa. (c) SVM using the OA. (d) SVM using kappa.

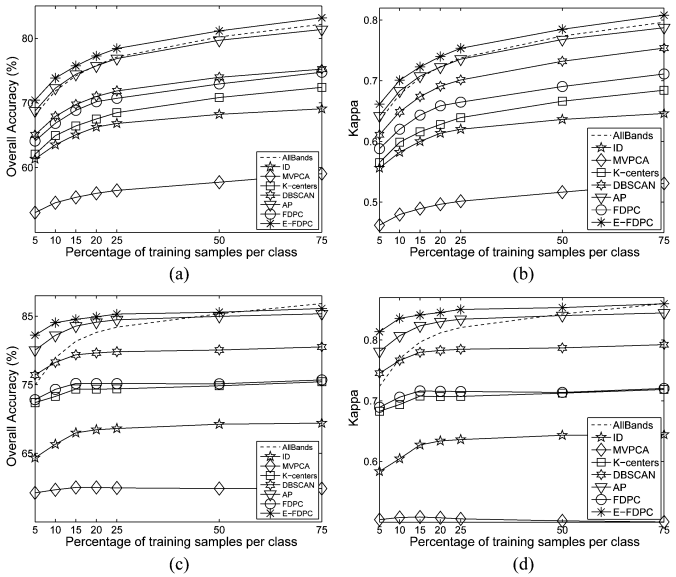

Fig. 14. Classification accuracy versus the different percentage of training samples per class on the Indian Pines data set. (a) KNN using the OA. (b) KNN using kappa. (c) SVM using the OA. (d) SVM using kappa.

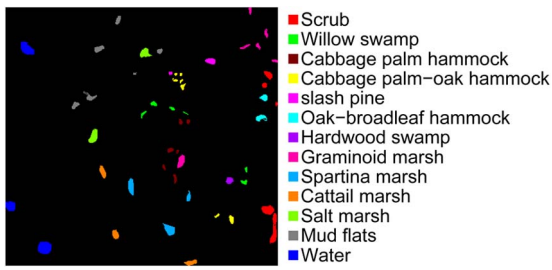

Fig. 15. Ground-truth map of the KSC data set (13 land-cover classes).

of the seven methods are listed in Table IV. Note that the bands selected by the ranking-based methods (the ID, MVPCA, FDPC, and E-FDPC methods) are presented in the order of importance, whereas those by the clustering-based methods (the K-centers, AP, and DBSCAN methods) are given in the sorted sequence. It can be noticed in Table IV that most of the bands selected by AP and E-FDPC methods are the same or close to each other. However, due to the instability of the chosen bands and the difficulty of tuning the preference parameter of the AP method, the proposed E-FDPC has a great advantage.

Finally, the classification accuracy level as a function of the number of training samples in each class for the interval $(3, \ldots, 15)$ is shown in Fig. 13. Meanwhile, the experiments of selecting the same percentage of samples from all labeled samples in each class as training samples for the interval $(5, \ldots, 75)$, i.e., each class has a different number of training samples, are conducted, as displayed in Fig. 14. The seven band selection methods with ten chosen bands and the results of all bands have been studied. As expected, the increase in the number of training samples has a substantial effect on the performance for all the eight methods. From these two figures, we can see that the performance of the E-FDPC is better than that of the alternatives in most cases, verifying the efficiency of the proposed approach. Due to the similar results obtained with the two kinds of selection strategies of training samples (i.e., a fixed number and a fixed percentage), the results with a different percentage of samples are not given in the following experiments.

B. KSC Data Set

The second data set that we used in our experiments was acquired by the AVIRIS sensor over the Kennedy Space Center (KSC), Merritt Island, FL, USA, on March 23, 1996 [52]. Fig. 15 depicts the ground-truth map of the land covers. In the original 224 bands, 48 bands are identified as water absorption and low-SNR bands (numbered 1–4, 102–116, 151–172, and 218–224), which are discarded, and only 176 bands remain. The spatial resolution of the data is 18 m per pixel. For classification purposes, 13 classes representing the various land-cover types that occur in this environment were defined for the site (see Table V). Classes 2 and 7 represent mixed classes, making the discrimination of the land cover more difficult.

Fig. 16 displays the classification accuracy of the KNN and that of the SVM using the selected bands obtained by the ID, MVPCA, K-centers, AP, DBSCAN, FDPC, and E-FDPC methods. Meanwhile, the result without a band selection procedure (i.e., AllBands) is also given for comparison.

TABLE V
LAND-COVER CLASSES WITH A NUMBER OF SAMPLES FOR THE KSC DATA SET

Class	Land Cover Type	No. of Samples
$C1$	Cabbage palm hammock	256
$C2$	Cabbage palm/oak hammock	252
$C3$	Cattail marsh	404
$C4$	Graminoid marsh	431
$C5$	Hardwood swamp	105
$C6$	Mud flats	503
$C7$	Oak/broadleaf hammock	229
$C8$	Salt marsh	419
$C9$	Scrub	761
$C10$	Slash pine	161
$C11$	Spartina marsh	520
$C12$	Water	927
$C13$	Willow swamp	243
	Total	5211

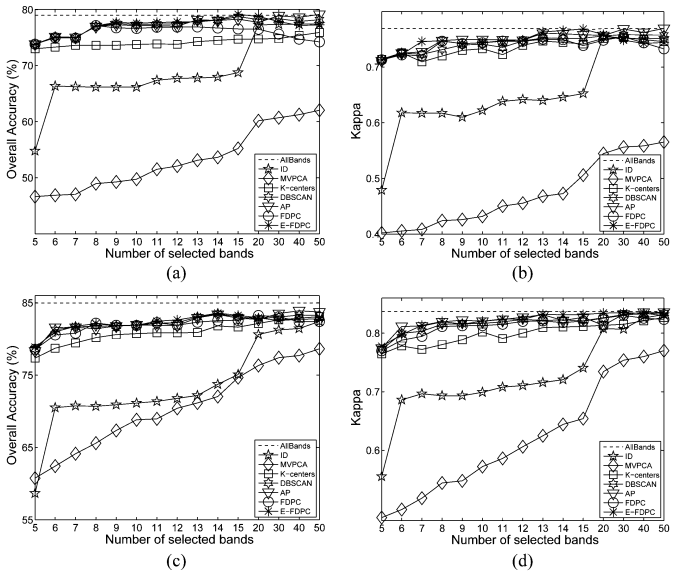

Fig. 16. Performance versus the selected bands of the different methods on the KSC data set. (a) KNN using the OA. (b) KNN using kappa. (c) SVM using the OA. (d) SVM using kappa.

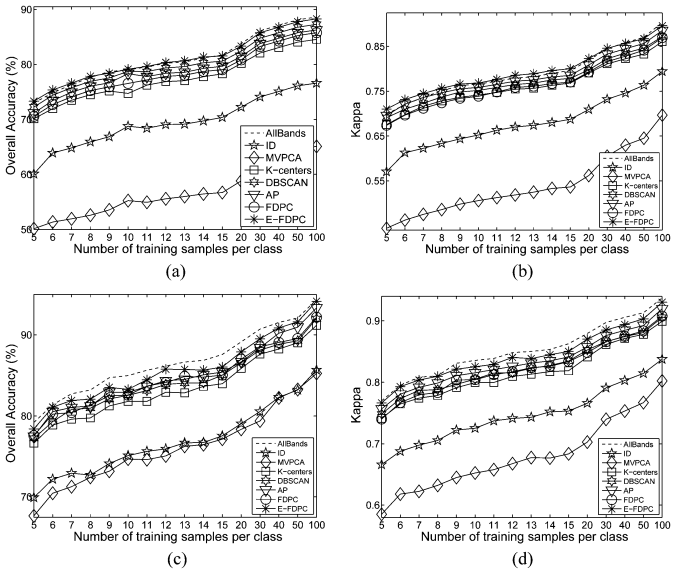

Fig. 17. Classification accuracy versus the different number of training samples per class on the KSC data set. (a) KNN using the OA. (b) KNN using kappa. (c) SVM using the OA. (d) SVM using kappa.

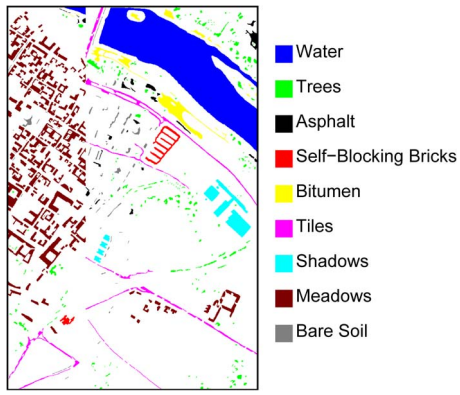

Fig. 18. Ground-truth map of the Pavia Center data set (nine land-cover classes).

As before, ten samples from each class are randomly chosen to form the training set. The classification accuracy values of the AP, DBSCAN, FDPC, and E-FDPC methods are close to each other, which are better than the other three methods, meaning that the E-FDPC is a valid option for band selection.

Same as the first experiment, the effect of varying the number of training samples on our method and the alternatives is conducted. According to the idea of automatically determining the optimal number of selected bands by the E-FDPC, the isolated band appears in the 16th point; thus, the preferable number of selected bands k should be 15. Fig. 17 displays the classification performance of the compared methods with 15 selected bands, in which the training samples per class range from 5 to 100. Likewise, the classification accuracy values improve as the number of training samples increases. In particular, both the ID and MVPCA methods yield the lowest results for both measures, showing the limits of the ranking-based methods for band selection. Meanwhile, the AP method is better than the K-centers method, which is due to the efficient message-passing procedure. Furthermore, our E-FDPC method consistently provides the best results with 15 selected bands, meaning that the ranking-based clustering approach is more effective for hyperspectral band selection.

C. Pavia Center Data Set

The third airborne data set was acquired by the Reflective Optics Spectrographic Imaging System 03 (ROSIS-03) sensor over the center of Pavia, Italy, with 115 spectral bands. The Pavia Center image was originally 1096 $\times$ 1096 pixels. A 381-pixelwide black stripe in the left part of the image was removed, resulting in an image with 1096 $\times$ 715 pixels. After removing the 13 noisy bands, the remaining 102 channels are processed. There are 148 152 labeled samples in total, and 9 classes of interest are considered. Fig. 18 shows the nine ground-truth classes of interest, whereas Table VI lists the land-cover classes with a number of samples.

TABLE VI
LAND-COVER CLASSES WITH A NUMBER OF SAMPLES FOR THE PAVIA CENTER DATA SET

Class	Land Cover Type	No. of Samples
$C1$	Water	65971
$C2$	Trees	7598
$C3$	Asphalt	3090
$C4$	Self-Blocking Bricks	2685
$C5$	Bitumen	6584
$C6$	Tiles	9248
$C7$	Shadows	7287
$C8$	Meadows	42826
$C9$	Bare Soil	2863
	Total	148152

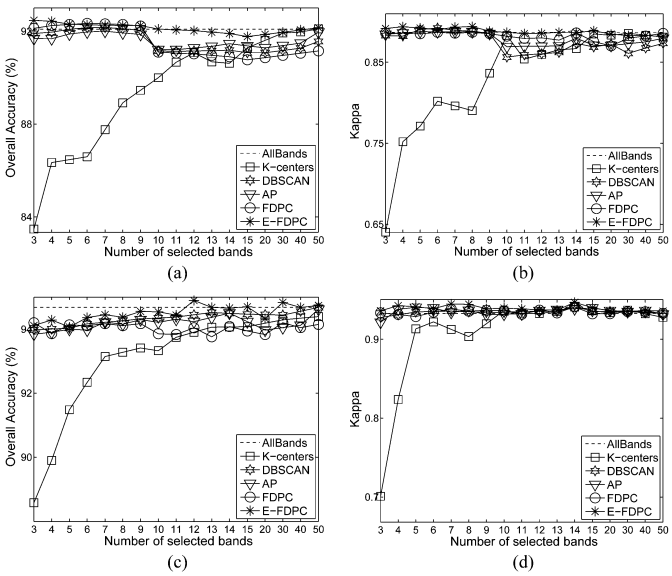

Fig. 19. Performance versus the selected bands of the different methods on the Pavia Center data set. (a) KNN using the OA. (b) KNN using kappa. (c) SVM using the OA. (d) SVM using kappa.

The classification results using the bands selected by the different methods, including the K-centers method, the AP method, the DBSCAN, the FDPC, and the proposed E-FDPC, are depicted in Fig. 19. The training sizes of all experiments are ten samples per class. It is worth noting that, because the accuracy values of the two ranking-based methods (the ID and MVPCA methods) are much lower than the others, they are not given in this figure. Apparently, the classification accuracy values are generally higher than those of the aforementioned two data sets, and the accuracy values are quite high when the number of selected bands is small. The reason may be due to the high spatial resolution of the data (1.3 m per pixel); thus, the spectral characteristic of each class is markedly different, and the pixel-oriented classification can be readily performed. Likewise, most of the accuracy values obtained by the E-FDPC are better than the others. Moreover, when the number of selected bands is 14 (which is the preferable number of selected bands), the classification performance of the compared methods is examined with a different number of training samples, as displayed in Fig. 20. Clearly, as the number of training samples increases, the accuracy values improve. Similarly, the proposed E-FDPC shows higher performance than the other band selection methods, and comparable with all bands, this justifies the efficiency of the proposed methodology.

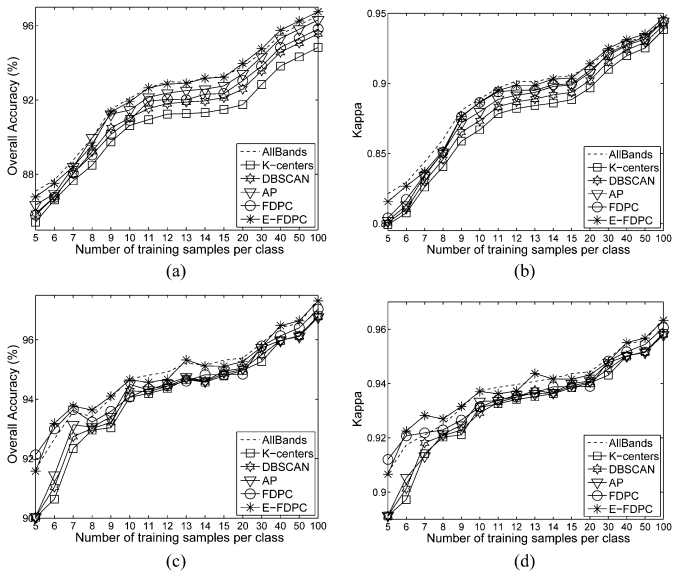

Fig. 20. Classification accuracy versus the different number of training samples per class on the Pavia Center data set. (a) KNN using the OA. (b) KNN using kappa. (c) SVM using the OA. (d) SVM using kappa.

From all the experiments on the three hyperspectral remote sensing data sets, some important results can be summarized. The spectral information of hyperspectral imagery contains large amounts of redundancy; thus, band selection is necessary and feasible. In general, the clustering-based methods usually achieve higher classification performance than the ranking-based methods, but the stability of the chosen band set and the relatively high complexity of the clustering-based methods are not as good as those of the ranking-based methods. Through extending the recently proposed FDPC clustering algorithm, the E-FDPC approach provides an artful band selection tool for combining the two kinds of methods. Compared with other methods, the E-FDPC always produces good and stable classification performance with any classifier and in any hyperspectral data set. As it can be seen in Table VII, with the same number of selected bands, the classification accuracy of the E-FDPC is higher than that of popular band selection methods such as the ID, MVPCA, K-centers, AP, and DBSCAN methods. Furthermore, through investigating the impact of a different number of selected bands on the classification accuracy of the E-FDPC (see Figs. 12, 16, and 19), the proper number of selected bands estimated by the E-FDPC (as shown in Table VII) can be a suitable candidate for hyperspectral band selection.

V. CONCLUSION

In this paper, we have presented a UBS approach for hyperspectral data. The method is based on the density of bands and the distance of each individual band to other bands. Different from conventional clustering-based methods (such as the K-centers and AP methods) and ranking-based methods (such as the ID and MVPCA methods), the proposed E-FDPC method is a ranking-based clustering band selection approach.

TABLE VII
ACCURACY AND STANDARD DEVIATION OF THE DIFFERENT METHODS (THE NUMBERS IN BOLD REPRESENT THE BEST CLASSIFICATION PERFORMANCE)

Data set (proper number)	Classifier (measure)	ID	MVPCA	K-centers	DBSCAN	AP	FDPC	E-FDPC
Indiana (10 bands)	KNN (OA)	43.46±0.22	36.47±0.23	44.89±0.56	47.37±0.21	49.27±0.23	45.90±0.15	**51.47±0.17**
	KNN (Kappa)	36.26±0.19	29.58±0.20	38.41±0.52	43.36±0.19	43.51±0.17	40.03±0.14	**45.89±0.15**
	SVM (OA)	43.44±0.27	37.96±0.28	49.78±0.33	53.65±0.32	57.30±0.45	50.07±0.12	**57.91±0.14**
	SVM (Kappa)	38.12±0.23	34.05±0.24	44.29±0.30	51.84±0.31	52.70±0.42	45.25±0.11	**52.95±0.13**
KSC (15 bands)	KNN (OA)	68.75±0.55	55.22±0.42	74.73±0.36	78.75±0.38	78.31±0.23	76.54±0.27	**79.15±0.24**
	KNN (Kappa)	65.25±0.53	50.60±0.37	74.16±0.32	75.19±0.35	75.85±0.21	73.38±0.25	**76.84±0.21**
	SVM (OA)	75.09±0.67	74.67±0.28	81.64±0.24	82.78±0.28	82.89±0.15	82.56±0.18	**83.27±0.20**
	SVM (Kappa)	72.51±0.64	65.24±0.23	79.96±0.22	80.51±0.24	81.55±0.14	80.41±0.17	**82.54±0.18**
Pavia Center (14 bands)	KNN (OA)	56.21±0.28	67.40±0.52	90.62±0.16	91.18±0.38	91.46±0.11	90.90±0.12	**91.90±0.08**
	KNN (Kappa)	54.28±0.21	65.39±0.48	86.69±0.13	87.33±0.35	87.98±0.09	88.64±0.10	**88.65±0.07**
	SVM (OA)	61.50±0.23	70.22±0.44	94.07±0.12	93.95±0.13	94.50±0.08	94.08±0.10	**94.66±0.06**
	SVM (Kappa)	47.69±0.21	55.39±0.38	93.13±0.10	92.92±0.11	93.34±0.05	93.14±0.08	**93.71±0.05**

The main idea of the E-FDPC is to pick out the bands that are actually the cluster centers in small regions, which has been achieved by introducing a weighted ranking score scheme for each band and an exponential-based heuristic rule for cutoff threshold d_c. Specifically, the reason that the proposed E-FDPC algorithm works well for hyperspectral band selection not only depends on the effectiveness of the original FDPC algorithm but also lies in the enhancement of the FDPC method for the hyperspectral UBS task. Furthermore, a simple and effective strategy to automatically determine the optimal number of selected bands by the E-FDPC has been presented.

We have illustrated the superiority of our E-FDPC approach on three real hyperspectral images and compared our method against a number of alternatives, i.e., the ID, MVPCA, K-centers, AP, DBSCAN, and FDPC methods. The experimental results consistently show that the E-FDPC exhibits better performance. We would like to emphasize that the E-FDPC method presented here is quite general in nature and can be readily applied to other features (such as wavelet and Gabor features), in which the ranking-based clustering combination is a valuable computational tool for the UBS of hyperspectral imagery.

ACKNOWLEDGMENT

The authors would like to thank Prof. M. Crawford for providing the AVIRIS Indiana Pines hyperspectral data, along with the training and test sets, and the anonymous reviewers for their comments and suggestions, which greatly helped us in improving the technical quality and presentation of our manuscript.

REFERENCES

[1] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–115, 2003.

[2] C.-I Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. New York, NY, USA: Plenum, 2003.

[3] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. S1, pp. 110–122, Sep. 2009.

[4] J. M. Bioucas-Dias *et al.*, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[5] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

[6] 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.

[7] P. Zhong, P. Zhang, and R. Wang, "Dynamic learning of SMLR for feature selection and classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sensing Lett.*, vol. 5, no. 2, pp. 280–284, Apr. 2008.

[8] J. Khodr and R. Younes, "Dimensionality reduction on hyperspectral images: A comparative review based on artificial datas," in *Proc. 4th CISP*, 2011, vol. 4, pp. 1875–1883.

[9] S. Kumar, J. Ghosh, and M. M. Crawford, "Best-bases feature extraction algorithms for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1368–1379, Jul. 2001.

[10] L. O. Jimenez-Rodrguez, E. Arzuaga-Cruz, and M. Velez-Reyes, "Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 469–483, Feb. 2007.

[11] 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.

[12] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.

[13] M. Jones and R. Sibson, "What is projection pursuit?" *J. R. Stat. Soc. Ser. A, Gener.*, vol. 150, no. 1, pp. 1–36, 1987.

[14] A. Ifarraguerri and C.-I Chang, "Unsupervised hyperspectral image analysis with projection pursuit," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 6, pp. 2529–2538, Nov. 2000.

[15] S. Kaewpijit, J. L. Moigne, and T. El-Ghazawi, "Hyperspectral imagery dimension reduction using principal component analysis on the HIVE," presented at the Science Data Processing Workshop, Greenbelt, MD, USA, 2002.

[16] A. Agarwal, T. El-Ghazawi, H. El-Askary, and J. Le-Moigne, "Efficient hierarchical-PCA dimension reduction for hyperspectral imagery," in *Proc. IEEE ISSPIT*, 2007, pp. 353–356.

[17] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 1, Mar. 2009, Art. ID. 783 194.

[18] 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.

[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] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, no. 5, pp. 1027–1061, May 2007.

[21] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

[22] 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.

[23] S. Chen and D. Zhang, "Semisupervised dimensionality reduction with pairwise constraints for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 369–373, Mar. 2011.
[24] X. Zhang, Y. He, N. Zhou, and Y. Zheng, "Semisupervised dimensionality reduction of hyperspectral images via local scaling cut criterion," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1547–1551, Nov. 2013.
[25] C.-I Chang, Q. Du, T. L. Sun, and M. L. G. 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.
[26] 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.
[27] Y.-Q. Zhao, L. Zhang, and S. G. Kong, "Band-subset-based clustering and fusion for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 747–756, Feb. 2011.
[28] 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.
[29] C.-I Chang, *Hyperspectral Data Exploitation: Theory and Applications*. Hoboken, NJ, USA: Wiley-Interscience, 2007.
[30] 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.
[31] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Hoboken, NJ, USA: Wiley-Interscience, 2012.
[32] J. F. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.
[33] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.
[34] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York, NY, USA: Wiley, 2008.
[35] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 1, pp. 32–40, Jan. 1975.
[36] AVIRIS NW Indiana's Indian Pines 1992 Data Set. [Online]. Available: https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html
[37] S. Bourennane, C. Fossati, and A. Cailly, "Improvement of classification for hyperspectral images based on tensor modeling," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 801–805, Oct. 2010.
[38] S. L. France and J. D. Carroll, "Two-way multidimensional scaling: A review," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 41, no. 5, pp. 644–661, Sep. 2011.
[39] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014.
[40] V. Talens, V. Laparra, J. Malo, and G. Camps-Valls, "Kernel structural similarity on hyperspectral images," in *Proc. IEEE IGARSS*, 2013, pp. 1214–1217.
[41] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2th ICKDDM*, 1996, pp. 226–231.
[42] K. Khan, S. Rehman, K. Aziz, S. Fong, and S. Sarasvady, "DBScan: Past, present and future," in *Proc. ICADIWT*, 2014, pp. 232–238.
[43] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.
[44] J.-F. Cardoso, "Dependence, correlation and Gaussianity in independent component analysis," *J. Mach. Learn. Res.*, vol. 4, pp. 1177–1203, 2003.
[45] A. K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Upper Saddle River, NJ, USA: Prentice-Hall, 1988.
[46] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ, USA: Wiley, 1990.
[47] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010.
[48] D. Dueck, "Affinity propagation: Clustering data by passing messages," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Toronto, Toronto, ON, Canada, 2009.
[49] S. Jia, Y. Qian, and Z. Ji, "Band selection for hyperspectral imagery using affinity propagation," in *Proc. DICTA*, 2008, pp. 137–141.
[50] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.
[51] J. A. Richards, *Remote Sensing Digital Image Analysis: An Introduction*. New York, NY, USA: Springer-Verlag, 2013.
[52] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

Sen Jia (M'13) received the B.E. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively.

He is currently an Associate Professor with the College of Computer Science and Software Engineering, Shenzhen University, Shenzen, China. His research interests include hyperspectral image processing, signal and image processing, pattern recognition, and machine learning.

Guihua Tang (S'15) received the B.E. degree in network engineering from Hunan University of Technology, Zhuzhou, China, in 2013. He is currently working toward the Master's degree in pattern recognition and intelligent system in the College of Computer Science and Software Engineering, Shenzhen University, Shenzen, China.

His research interests include hyperspectral image processing, pattern recognition, and machine learning.

Jiasong Zhu received the B.E. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1997; the M.E. degree from Wuhan University, Wuhan, China, in 2003; and the Ph.D. degree from The University of Hong Kong, Pokfulam, Hong Kong, in 2008.

He is currently an Associate Professor with the College of Civil Engineering, Shenzhen University, Shenzhen, China. His research interests include multisensor integration and data fusion, high-resolution image processing, and geographic information system applications in urban planning and transportation.

Qingquan Li received the M.S. degree in engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1988 and 1998, respectively.

From 1988 to 1996, he was an Assistant Professor with Wuhan University, where he served as an Associate Professor from 1996 to 1998 and has been a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing since 1998. He is currently the President of Shenzhen University, Shenzhen, China, where he is also the Director of the Shenzhen Key Laboratory of Spatial Information Smart Sensing and Services. He is an expert in modern traffic with the National 863 Plan. His research interests include photogrammetry, remote sensing, and intelligent transportation systems.

Dr. Li is an Editorial Board Member of the *Surveying and Mapping Journal* and the *Wuhan University Journal—Information Science Edition*.