A Fast Neighborhood Grouping Method for Hyperspectral Band Selection

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

***Abstract*—Hyperspectral images can provide dozens to hundreds of continuous spectral bands, so the richness of information has been greatly improved. However, these bands lead to increasing complexity of data processing, and the redundancy of adjacent bands is large. Recently, although many band selection methods have been proposed, this task is rarely handled through the context information of the whole spectral bands. Moreover, the scholars mainly focus on the different numbers of selected bands to explain the influence by accuracy measures, neglecting how many bands to choose is appropriate. To tackle these issues, we propose a fast neighborhood grouping method for hyperspectral band selection (FNGBS). The hyperspectral image cube in space is partitioned into several groups using *coarse-fine* strategy. By doing so, it effectively mines the context information in a large spectrum range. Compared with most algorithms, the proposed method can obtain the most relevant and informative bands simultaneously as subset in accordance with two factors, local density and information entropy. In addition, our method can also automatically determine the minimum number of recommended bands by determinantal point process. Extensive experimental results on benchmark datasets demonstrate the proposed FNGBS achieves satisfactory performance against state-of-the-art algorithms.**

***Index Terms*—Hyperspectral image, band selection, neighborhood grouping, context information, determinantal point process.**

I. Introduction

THE hyperspectral sensor captures the target region by dozens to hundreds of consecutive spectral bands to obtain hyperspectral image cube. It has the characteristics of high resolution, large number of bands and large amount of data. Compared with multispectral image, the hyperspectral image not only has a great improvement in information richness, but also provides the possibility of more reasonable and effective analysis [1]. Therefore, it is widely used in ground object detection and recognition [2], environmental monitoring [3], disaster prediction [4], etc. Nevertheless, these bands also pose some issues. For instance, a large number of hyperspectral data increase the complexity of data processing. In addition, since the range of each band of hyperspectral image is narrowed and the spectrum is the same kind of wave in a certain range, there is remarkable feature that high similarity exits among adjacent bands [5] (see Fig. 1). It appears dimension disaster phenomenon easily, affecting subsequent image processing tasks. The above dilemmas are usually solved by reducing the dimension of hyperspectral image data.

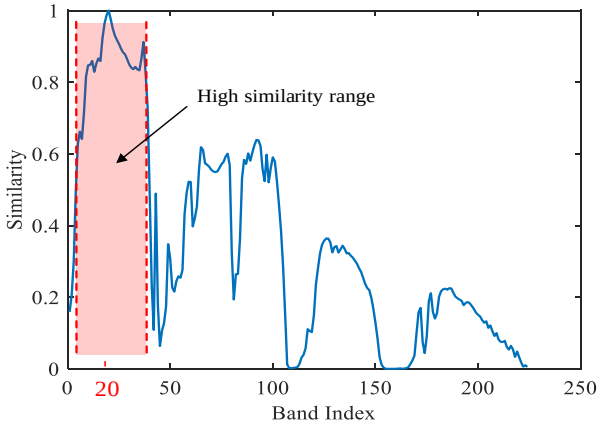


Fig. 1. The similarity for current band (band index 20) with other bands on Salinas dataset.

Band selection is regarded as an effective method for hyperspectral dimensionality reduction. Without changing the original hyperspectral data, it is to select a subset of bands containing most of the structure and information from dozens to hundreds of bands to remove some bands that are redundant or cause interference [6], [7]. When performing band selection, there are two principles to follow, that is, the selected bands have a large amount of information and the similarity among bands is small [8], [9]. By doing so, it can reduce the dimension of hyperspectral image and retain enough object information, which is more conducive to further image processing.

In recent years, many unsupervised band selection algorithms have been proposed. These methods are mainly divided into two categories, clustering-based [10]–[13] and ranking-based [14]–[17]. The clustering-based methods, e.g., fast density-peak-based clustering (E-FDPC) [10] and Ward's linkage strategy using divergence (WaLuDi) [11], treat each band as a point, and build a similarity matrix by Euclidean distance or other criteria. Using this matrix, some clustering methods [18] are then adopted to cluster these bands into groups so that the bands with high similarity are grouped into one group. The most relevant band with other bands in each cluster is finally selected as representative band [19], [20]. Unfortunately, this type of algorithms makes the similarity of obtaining subset small, but the selected band do not necessarily have the amount of information. Additionally, when perform-

This work was supported by the National Key R&D Program of China under Grant 2017YFB1002202, National Natural Science Foundation of China under Grant U1864204, 61773316, U1801262, and 61871470.

The authors are with the School of Computer Science and the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China (e-mail: crabwq@gmail.com, liqmges@gmail.com, xuelong_li@nwpu.edu.cn) (*Corresponding author: Xuelong Li.*)

ing clustering algorithm, the hyperspectral bands are usually considered to be unordered. It results in ignoring the context information of the whole spectral bands.

For ranking-based method, typical methods have maximum-variance principal component analysis (MVPCA) [15] and linearly constrained minimum variance based on constrained band selection (LCMV-CBS) [21]. This type of methods is based on a certain criterion to calculate a score for each hyperspectral band. The scores are then sorted by ascending or descending order, and the bands corresponding to the maximum or minimum score values are selected as the target subset in terms of the meaning of the score value [22]. Although these methods have achieved satisfactory results, they also have two shortcomings. First, these algorithms are very sensitive to hyperspectral bands containing noises, which makes that the selected band may have a strong similarity. Second, most algorithms change the original information in the hyperspectral image cube through spatial transformation, which results in the loss of some key information.

With respect to the above descriptions, it can be concluded that the hyperspectral bands should be viewed as ordered to deal with. Moreover, without changing the original hyperspectral image data, we can make use of the idea of ranking-based method to obtain the relevant and informative bands in each group. Importantly, when implementing band selection, the number of selected bands is unknown. In most cases, many scholars only pay attention to the different number of selected bands to explain the impact of precision measurement, ignoring how many bands to choose is appropriate [23], [24]. Therefore, this is one of the key issues that need to be solved urgently for band selection. To tackle these issues, we propose a fast neighborhood grouping method for hyperspectral band selection (FNGBS). The main contributions are as follows:

1) Neighborhood band grouping is proposed to partition the hyperspectral image cube into multiple groups in space. The bands of high similarity in certain spectrum range are assigned into one group via *coarse-fine* mechanism. It can fully mine the context information of the whole spectral bands so as to obtain the selected subset more dispersed.

2) The bands with the maximum product of local density and information entropy in groups are selected as band subset. Compared with most existing methods about selecting representative band, our method can better exploit local distribution characteristics and makes the obtained subset having more discriminative bands simultaneously.

3) An automatic strategy is designed to obtain the minimum number of recommended bands by determinantal point process. Through analyzing the redundancy among band subsets, we effectively reveal how many bands can be selected to get approximately the same results as all bands.

The remainder of this paper is organized as follows. Section II introduces two parts: ordered band partition and representative band selection. In Section III, the detailed descriptions about our proposed method are presented, including neighborhood bands grouping, maximum local density and informative band, and recommended number of bands. After that, extensive comparative experiments in Section IV are conducted on four public hyperspectral datasets. Finally, conclusions are given in Section V.

II. Related Work

As introduced in Section I, it makes low similarity in different groups by partitioning the hyperspectral image cube in space. Then the representative band is selected from each group by some evaluation criteria. Next, several representative methods are listed and discussed.

A. Ordered Bands Partition

The principle of uniform band selection [25] is based on the number of predetermined hyperspectral bands to evenly partition hyperspectral image cube. This method is often exploited to compare the performance of the proposed method. Due to the lack of theoretical foundation, the effect is not good. The hyperspectral data covers a wide range of spectral bands from visible light to near infrared and then to far infrared, some band selection methods [26] use this advantage to divide hyperspectral image cube. For instance, the cube is divided according to visible light (0.4-0.7 μm), near infrared (0.7-1.3 μm) and short infrared (1.3-2.5 μm). These methods have a relatively reasonable theoretical basis, but it does not take into account the impact of different scenes for data acquisition and the relationship of the change of object type and the structure of hyperspectral data.

Moreover, the researchers [23], [27] also propose automatic subspace partition based on neighboring transmissible correlation. Assuming that the local correlation of neighboring bands has approximate continuity and transmissibility, the correlation coefficients between any two bands are obtained. Then, the curve is generated by applying the correlation of adjacent bands. According to the curve, the local minimum values are attained to partition hyperspectral image cube into multiple groups. This treated way is extremely novel, but there is a threshold that needs to be set, which affects the outcome of subsequent processing. Furthermore, it only considers the relationship between two adjacent bands, while the influence of nearby bands is ignored. Based on this significant hyperspectral feature, Wang et al. [28] employ dynamic programming to segment hyperspectral image cube in space. Compared with above methods, it achieves promising performance.

B. Representative Band Selection

After dividing these bands, we need to select the most representative band in each group. For clustering-based methods, the band nearest to the cluster center is generally chosen as the selected band. This way can obtain the most relevant band, but it neglects the amount of information in the band, which not conforms to the principles of band selection. Unlike the above methods, some algorithms choose the informative band in each group by quantifying the importance of each band to obtain subset. In the following, several typical methods will be briefly described.

Information divergence [25] in these unsupervised methods is usually utilized to evaluate the recognition potential of each

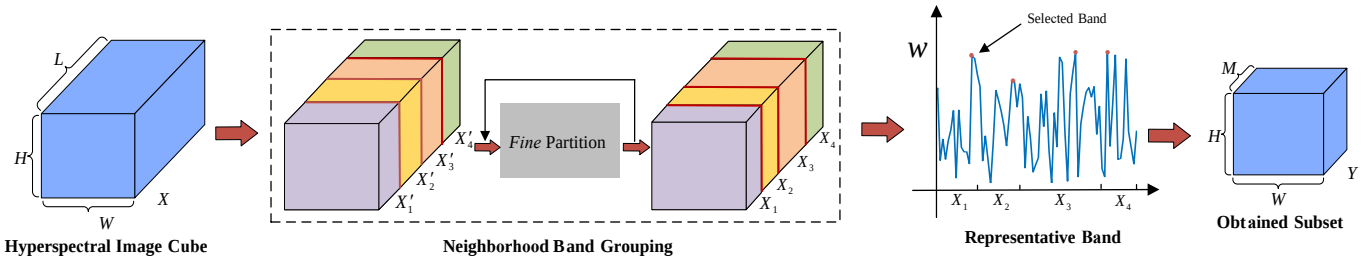


Fig. 2. Overall architecture of our proposed method.

band. It analyzes the difference of probability distribution in each band through non-Gaussian measurement (KL divergence [11] and kurtosis of probability distribution [29]), and acquires the band with greatest difference as the most informative band. Wang et al. [28] use three algorithms, MVPCA, E-FDPC, and information entropy [30], to evaluate the bands in each group. These methods either choose the most informative bands or cluster centers to be a subset, which makes it impossible for the selected subset to be discriminative and informative simultaneously.

III. Proposed Method

In this section, we detail the proposed method (FNGBS), whose flowchart is shown in Fig. 2. Specifically, based on the fact that there is large similarity between current band and adjacent bands, neighborhood band grouping is proposed to partition hyperspectral image cube into multiple groups. By calculating the two factors (local density and information entropy) of band in groups, we choose the band with maximum weight of the product of two factors in each group as band subset. In addition, an automatic strategy is designed to obtain the minimum number of recommended bands by determinantal point processes. Three parts are described in detail in the following sections.

A. Neighborhood Band Grouping

As we previously mentioned, the similarity between non-adjacent bands is usually lower than that between adjacent bands. Inspired by this phenomenon, we propose neighborhood band grouping to divide the hyperspectral image cube using *coarse-fine* strategy. Different from other algorithms, this method only considers the influence of adjacent bands, which can effectively reduce the execution time and fully mine the context information of spectral bands.

Let $X = \{x_1, x_2, ..., x_L\} \in R^{N \times L}$ be the hyperspectral image cube, where N represents the number of pixels in the band, L is the total number of bands, and x_i denotes the vector of the i-th band. Supposing that M bands are selected during band selection, the group number of bands also should be set as M. For the hyperspectral image cube X, they are first divided into M groups $\{X'_m\}_{m=1}^{m=M}$ in space as evenly as possible. Concretely, the partition point G is defined as

$$G(m) = \begin{cases} 1, & m = 1 \\ \dfrac{(L - \mathrm{mod}(L, M)) \times m}{M}, & 2 \le m \le M-1 \\ L, & m = M \end{cases}, \tag{1}$$

where $\mathrm{mod}(\cdot, \cdot)$ denotes mod operation. In terms of G, the $\{X'_m\}_{m=1}^{m=M}$ can be obtained by

$$X'_m = \{x_i\}_{i=G(m)}^{i=G(m+1)}, \quad m = 1, 2, ..., M-1. \tag{2}$$

The result of this partition is a coarse band grouping, which is similar to uniform band selection method.

Due to the lack of theoretical basis for the initial division and the results are not ideal, we adopt fine neighborhood grouping approach to repartition initial groups $\{X'_m\}_{m=1}^{m=M}$, thus obtaining more accurate grouping. The band x_{p_m} in the groups $\{X'_m\}_{m=1}^{m=M}$ is taken as the initial cluster center, which is defined as

$$p_m = \left\lfloor m \times \frac{L}{M} - \frac{L}{2M} \right\rfloor, \quad m = 1, 2, ..., M, \tag{3}$$

where m records that it currently belongs to the m-th group. To speed up the execution of the algorithm and fully dig the context information, for the current cluster center x_{p_m}, we only take into account these bands $\{x_j\}_{j=a}^{j=b}$. With respect to a, b, they can be obtained by

$$(a, b) = \begin{cases} [1, p_{m+1}), & m = 1 \\ (p_{m-1}, p_{m+1}), & 2 \le m \le M-1 \\ (p_{m-1}, L], & m = M \end{cases}. \tag{4}$$

Next, we will detail fine partition algorithm. For more description about this process, it is shown in **Algorithm 1**. Through multiple iterations, the number of bands in each group is no longer consistent. Finally, the bands of high similarity are assigned to one group, which makes the low redundancy between different groups. This way meets one of the principles of band selection. An example to fine partition the hyperspectral image cube is given in Fig. 3.

B. Maximum Local Density and Informative Band

After the hyperspectral bands have been separated into groups, the groups are often treated independently. When selecting the representative band in each group, the traditional methods usually select the most relevant or informative band in each group. However, this strategy makes representative band may not be discriminative. To address this problem, an

Algorithm 1 Process of fine partition algorithm

Input: Coarse band grouping for hyperspectral cube $\{X'_m\}_{m=1}^{m=M}$. Number of iterations $inter$.

Output: Fine band grouping for hyperspectral cube $\{X_m\}_{m=1}^{m=M}$.

1: Define R to record the similarity between the current band x_j and the cluster center x_{p_m}, and initialize the vectors R to infinity. Define group label T and initialize it to zero.

2: A similarity matrix D is constructed by Euclidean distance, and the distance between cluster center x_{p_m} and current band x_j is acquired.

3: If $D(x_{p_m}, x_j) < R(j)$, then set $R(j) = D(x_{p_m}, x_j)$ and $T(j) = m$.

4: After vectors R and T are updated by Step 3, the value in some groups is inconsistent due to the influence of factors, such as noise. For example, group label T is [1, 1, **2**, 1, 1, 2, 2, 2]. We process these outliers so that each group has the same value.

5: Start new iteration by returning to Step 3 until the number of iteration is reached or the group label T does not change.

6: According to group label T, we select the maximum value of the index in each group as the new partition point to update G.

7: Obtain fine division of hyperspectral image cube $X_m = \{x_i\}_{i=G(m)}^{i=G(m+1)}$.

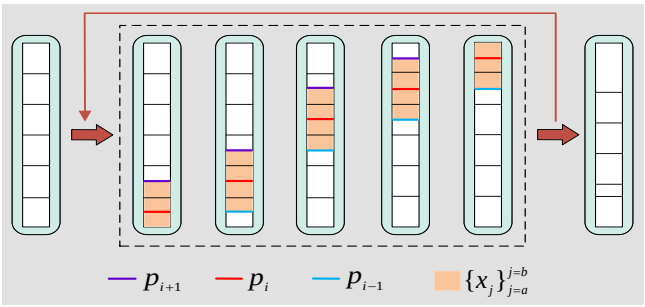


Fig. 3. An example to partition the hyperspectral image cube into six groups by multiple iterations.

effective method is proposed to select the informative and most relevant bands simultaneously. The basic idea is to rank the bands in each group in accordance with the product of local density and information entropy, and ensure that only one band with maximum weight is selected. Here, we give the details of the algorithm.

According to D, for a band x_u, the distances between it and other bands can be obtained and are arranged in ascending order. The k nearest neighbor set [31] of band x_u is defined as follows:

$$kNN(x_u) = \{x_v | D(x_u, x_v) \leq d_{uk}\}, \tag{5}$$

where d_{uk} is the k-th distance for band x_u. As can be seen from kNN, it effectively reveals the distribution between each band and other bands. In order to select representative band in each group, shared nearest neighbor [32], [33] is introduced

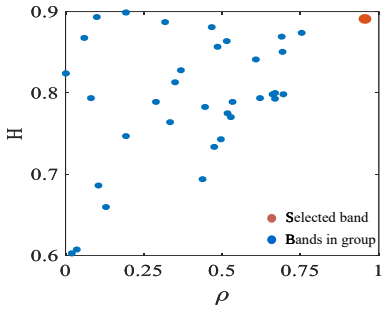


Fig. 4. Representative band by selecting the band corresponding to the maximum of two factors in each group.

to describe the number of shared neighbor elements between band x_u and band x_v, i.e.,

$$SNN(x_u, x_v) = |kNN(x_u) \cap kNN(x_v)|, \tag{6}$$

where $\cap$ denotes intersection for two bands. Shared nearest neighbor indicates the similarity between band x_u and k neighbors in local space, which can show the band distribution in space very well. In general, the larger this value is, the closer the distribution of band x_u and neighbor subset is.

Using the Eqs. (5) and (6), one of factors, local density, is expressed as the ratio of two variables. Since the number of bands is small in each group after partitioning the hyperspectral image cube, if the ratio of two variables is directly adopted to calculate the local density, this will cause statistical errors [10]. Accordingly, in this paper, Gaussian kernel function is used to accurately perform local density, i.e.,

$$\rho_u = \sum_{x_v \in kNN(x_u)} \exp\left(-\frac{D(x_u, x_v)}{SNN(x_u, x_v) + 1}\right). \tag{7}$$

Information entropy [34], [35] is a statistical form of feature, reflecting the average amount of information in the image. To make the selected bands having large amount of information, the other factor H_u in each group is measured by calculating information entropy, which is defined as

$$H_u = -\sum_{z \in \Omega} p(z) \log p(z), \tag{8}$$

where Ω denotes the grayscale space, and $p(z)$ is the probability that an event $z \in \Omega$ appears in the image. It is worth pointing out that the number of pixels N is usually extremely large, and if all pixels are utilized to calculate information entropy, the execution time of the algorithm will be significantly increased. Consequently, in order to avoid the case, we randomly pick Z pixels from N pixels ($Z \ll N$) to perform this operation. For the influence of the parameter Z, we will analyze it in detail in the following experiments.

Through the above analysis, it can be concluded that when both factors are the largest, the selected band will not only satisfy the large amount of information, but also be more representative (see Fig. 4). Note that the values of the two factors do not have the same order of magnitude, their data are normalized to the range of [0, 1] by

$$H = (H - H_{\min})./(H_{\max} - H_{\min}), \tag{9}$$

$$\rho = (\rho - \rho_{\min})./(\rho_{\max} - \rho_{\min}), \tag{10}$$

where $./$ denotes the element division operator, $H_{\min}$ and $\rho_{\min}$ represent the minimum of vector H and ρ, and $H_{\max}$ and $\rho_{\max}$ are the maximum of vector H and ρ. In addition, when the number of bands in group is less than the parameter k of nearest neighbor set, we only consider the amount of information of the band because the other factor ρ_u has little effect. Therefore, the weight w_u of each band in group can be denoted as

$$w_u = \begin{cases} H_u \times \rho_u, & count(m) > k \\ H_u, & otherwise \end{cases}, \tag{11}$$

where $count(m)$ is the total number of bands in group. According to Eq. (11), we pick the band with largest value w_u as the most relevant and informative band in each group and thus obtain the band subset Y containing M bands.

C. Recommended Number of Bands

In most cases, the scholars focus on different numbers of selected bands to explain the influence by accuracy measures, neglecting how many bands to choose is appropriate. Considering this problem, we propose an automatic strategy to obtain the minimum number of recommended bands by determinantal point process (DPP) [36].

DPP often estimates the redundancy between data, which is widely used in video summarization [37], [38]. To get the minimum number of recommended bands whose results approximately equal to the results of all the bands, DPP is introduced to measure the redundancy of selected bands. Actually, the aim of the minimum number of bands is to determine inflection point. This point is the recommended number of bands.

Suppose y is a subset of Y with M bands ($y \subseteq Y$, $y \neq \varnothing$), the discrete probability of a subset y can be defined as

$$P(y) = \frac{\det(B_y)}{\det(B+I)}, \tag{12}$$

where $det(\cdot)$ is the determinant operation, I is the identity matrix, $B = Y^T Y$, and B_y denotes the submatrix obtained by selecting the corresponding rows and columns from B in accordance with y. Considering that only two bands are selected, we have

$$P(y = \{x_u, x_v\}) \propto B_{uu}B_{vv} - B_{uv}^2. \tag{13}$$

If two bands are exactly the same, $P(y = \{x_u, x_v\}) = 0$. It indicates that the greater the similarity between bands is, the smaller the probability value is. In addition, in terms of above analysis. we can also acquire

$$\sum\nolimits_{y \subseteq Y} \det(B_y) = \det(B+I). \tag{14}$$

In our paper, compared with other subsets ($y \neq Y$), we assume that the redundancy of the subset ($y = Y$) contains M bands is minimum. Therefore, the above solution can be simplified to only calculate the redundancy of subset ($y = Y$), i.e.,

TABLE I
KEY INFORMATION ON FOUR PUBLIC DATASETS

Dataset	Bands	Size	Classes	Samples
Indian Pines	200	145 × 145	16	9323
Botswana	145	1476 × 256	14	3248
Pavia University	103	610 × 340	9	42776
Salinas	224	512 × 217	16	54129

$$P(y = Y) = \frac{\det(B_y)}{\det(B+I)}. \tag{15}$$

When the number of selected bands increases, overall, the discrete probability P will decrease in that many redundant bands are selected. Note that as the number of the selected bands reaches a certain value, adding more bands in subset will no longer significantly increase the redundancy. Hence, from this point, we determine the critical point r (recommended number) by the maximum change in the slope of the curve. Owing to setting an assumption, the curve of discrete probability may not be smooth. In our work, a simple exponential function is adopted to fit values of discrete probability to obtain a smooth curve before determining the required number. As a result, the discriminant condition of critical point is expressed as follows:

$$Q_M^{M-1} = P_{M-1} - P_M, \tag{16}$$

$$r = \arg\max \left(\frac{Q_M^{M-1}}{Q_{M+1}^M} \right), \tag{17}$$

where P_M is the discrete probability when selecting M bands. Of course, we also add a condition ($Q_{M+1}^M > 0$) to terminate this computation. According to the above equations, a suitable subset of the recommended bands is finally obtained from the original bands.

IV. EXPERIMENT

The experimental section mainly includes three parts. First, we briefly introduce four public hyperspectral datasets about basic characteristics. Then, the experimental setup is presented about classification setting, comparison methods, and number of selected bands. Finally, extensive experiments are employed to verify the effectiveness of our proposed method by different dimensions.

A. Datasets

1) Indian Pines: The Indian Pines dataset[1] was obtained by AVIRIS sensor in Indiana. This dataset contains 224 spectral bands and the size of each band is 145×145 pixels. Additionally, the ground truth consists of 16 classes (corn, grass, soybean, etc.). Since some bands are covered by the region of water absorption, we reduce these bands to finally get 200 bands, as shown in Table I.

[1]http://www.ehu.eus/ccwintco/uploads/2/22/Indian_pines.mat

TABLE II
CLASSIFICATION RESULTS BY SETTING DIFFERENT PARAMETERS ON INDIAN PINES DATASET USING SVM CLASSIFIER (%)

k	Z	5	7	10	12	15	20	25	30	35	40	45	50
2	1%	63.26	71.70	74.95	74.94	76.60	78.99	81.80	80.88	81.40	82.24	82.01	81.28
	10%	69.38	73.12	73.56	76.61	76.94	79.67	80.32	80.55	80.48	80.70	81.28	81.29
	50%	68.88	73.35	75.08	76.31	77.13	80.12	81.40	81.22	81.38	82.03	82.43	81.39
	100%	68.40	72.72	75.32	75.75	75.72	80.32	80.67	80.44	81.69	82.74	82.55	81.10
3	1%	64.40	72.59	77.00	75.36	76.75	80.91	80.76	80.16	81.52	81.03	82.09	81.89
	10%	64.75	71.22	74.78	76.77	76.34	79.89	81.66	80.67	81.54	81.88	82.70	82.39
	50%	63.27	71.20	74.70	75.21	76.49	79.38	79.82	79.48	81.68	82.37	81.96	81.36
	100%	64.39	72.39	75.91	76.34	75.75	79.74	80.16	80.95	81.49	80.24	82.93	81.01
4	1%	63.42	72.15	74.93	76.76	76.33	79.60	80.81	81.22	80.79	82.16	82.08	82.40
	10%	64.32	72.76	76.31	77.36	76.24	80.26	81.08	80.11	81.80	81.68	81.87	81.88
	50%	63.58	70.49	75.25	77.21	76.52	80.49	80.11	80.32	80.79	81.68	82.21	82.52
	100%	62.82	71.82	74.88	77.07	77.16	79.52	80.60	80.89	81.15	81.96	82.14	82.70

2) Botswana: The Botswana dataset[2] was acquired by NASA EO-1 satellite sensor over Okavango Delta, Botswana in May 31, 2001. This hyperspectral image cube at 30m pixel resolution has 242 bands covering the 400-2500 nm, and the size of each band is 1476 $\times$ 256 pixels. Moreover, this dataset contains 16 classes of land cover. Since water absorption and other factors lead to some bands with large noises, some bands are removed (10-55, 82-97, 102-119, 136-164, and 187-220) and we finally get 145 bands for experimental verification in our work. Table I displays some key information on Botswana dataset.

3) Pavia University: The Pavia University dataset[3] in 2002 was collected by ROSIS sensor during a flight campaign. The hyperspectral image cube has 115 bands with the size of 610 $\times$ 610 pixels. Unlike other datasets, this dataset discarded some samples without information before further analysis. Finally, the size of dataset becomes 610 $\times$ 340 pixels. Similar to Botswana dataset, we remove some bands with low signal to noise ratio and obtain 103 spectral bands with 9 classes (Table I).

4) Salinas: The Salinas dataset[4] was recorded by AVIRIS sensor over Salinas Valley, California. The original image cube has 224 spectral bands with the size of 512 $\times$ 217 pixels. The dataset contains 16 classes of land cover including bare soils, vegetables, etc. For this public hyperspectral dataset, we do not remove the bands of water absorption, whose details are described in Table I.

B. Experimental Setup

1) Classification Setting: In order to verify the quality of obtained bands, two classifiers, k-nearest neighborhood (KNN) [39] and support vector machine (SVM) [40], are adopted to classify the samples on four public hyperspectral datasets. In our experiments, two classifiers have the same parameter settings on different datasets. Specifically, the parameter of KNN classifier is set to be 5 for all classification validation experiments. With respect to SVM classifier, the RBF kernel is used, and the penalty C and gamma are initialized to 1×10^4 and 0.5, respectively.

Moreover, since the above two classifiers are supervised, 10% samples of selected bands are randomly picked as training set, and the rest of samples are employed to implement test task. Note that random selection of these samples will make the classification results unstable. In order to eliminate this effect, the final classification results are acquired by calculating average value after running 5 times (except for the section of parameter analysis). For all datasets, two methods of classification accuracy measure are conducted to evaluate the performance of algorithm, including overall accuracy (OA) and average overall accuracy (AOA).

2) Comparison Methods: To demonstrate the superiority of proposed method, four unsupervised band selection methods are compared with our method in each dimension. They are E-FDPC [10], WaLuDi [11], SNNC [8], and TOF [28]. Among these competitors, only SNNC has hyperparameter to set, and the rest are parameter-free. For the competitor SNNC, the hyperparameter is fixed at 5. Moreover, all bands in hyperspectral dataset are also considered for performance comparison.

3) Number of Selected Bands: When performing band selection, how many bands to choose in practice is unknown. Usually, different number of bands are set to analyze accuracy through classifiers. In our experiment, we set the range of [5, 50] to verify the impact of the different numbers of band.

C. Results

To investigate the performance of the proposed method in each dimension, in this paper, we will analyze it in detail from the following points, including the impact of parameter k and Z, recommended bands, classification performance, and computational time.

1) Hyperparameter Analysis: In order to reduce the processing time, Z pixels are randomly picked from N pixels when conducting the operation of information entropy. Moreover, as the number of bands in group is less than the parameter k of nearest neighbor set, we only consider the effect of the factor H. In this section, we set different parameters to study the impact on the performance of the

[2]http://www.ehu.eus/ccwintco/uploads/7/72/Botswana.mat
[3]http://www.ehu.eus/ccwintco/uploads/e/e3/Pavia.mat
[4]http://www.ehu.eus/ccwintco/uploads/f/f1/Salinas.mat

TABLE III
CLASSIFICATION RESULTS BY SETTING DIFFERENT PARAMETERS ON BOTSWANA DATASET USING SVM CLASSIFIER (%)

k	Z	5	7	10	12	15	20	25	30	35	40	45	50
2	1%	82.41	86.66	87.18	85.77	87.86	88.48	88.79	90.26	91.91	90.54	91.43	91.98
	10%	82.62	85.60	86.25	86.46	86.94	87.53	89.89	90.50	90.74	90.82	91.53	90.30
	50%	82.45	85.22	87.86	86.49	86.90	87.98	89.82	89.54	90.64	91.43	91.64	91.98
	100%	82.33	86.08	87.04	87.18	87.32	87.35	89.13	89.93	91.53	91.85	92.15	92.73
3	1%	80.08	83.48	84.20	86.90	87.49	87.45	88.17	89.44	90.81	90.26	90.71	90.13
	10%	80.29	85.46	85.43	85.98	85.77	88.65	89.41	90.16	91.53	90.81	90.78	91.12
	50%	81.45	83.75	85.40	87.08	87.56	86.84	88.79	89.78	90.89	92.08	91.02	92.53
	100%	79.91	84.23	84.81	86.36	87.25	87.52	89.37	90.20	91.43	89.85	90.61	91.22
4	1%	80.70	85.40	87.45	87.32	88.65	88.38	89.41	89.13	91.67	90.09	89.65	90.30
	10%	80.77	85.09	87.28	87.04	88.03	87.28	86.73	89.17	91.67	90.20	92.15	90.92
	50%	81.00	85.36	87.90	86.39	87.62	87.11	87.42	89.68	89.65	89.96	91.46	90.81
	100%	81.83	85.64	86.49	86.77	87.87	87.39	90.06	89.06	92.01	91.16	91.57	90.85

proposed method when different number of bands are selected (the range of [5, 50]). Specifically, as for parameter Z, 1%, 10%, 50%, and 100% of the pixels are picked from the original pixels. With respect to parameter k, it is set to 2, 3, and 4, respectively.

a) Indian Pines: The classification results about different parameter values are depicted in Table II. It can be seen that in this table these parameters greatly change the classification results, particularly as for k. Experiment reveals the parameters ($k = 2$ and $Z = 10\%$) can better superior performance. Therefore, in our work, we empirically set the parameters k and Z to 2 and 10% (maybe it is not optimal), without any further tuning parameters.

b) Botswana: Table III shows classification results by setting different parameters on Botswana dataset using SVM classifier. From this table, when the parameter k is fixed, the parameter Z has little influence on classification accuracy. On the contrary, it can be seen that setting different parameters k displays considerable change, especially when selecting a small number of bands. Overall, when the parameters k and Z are set to 2 and 1%, respectively, the classification performance is relatively good.

c) Pavia University: Likewise, we also set same parameters to analyze the impact on Pavia University dataset. Table IV provides the classification accuracy of selecting different bands. One can observe that no matter how many the parameter Z is set, there is little fluctuation in OA. However, the classification results have significant impact on the selection of 5 bands. Compared with Botswana dataset, the classification accuracy is not very different by setting several parameters k. In general, the OA is basically consistent during k equals 3 or 4. Thus, for Pavia University dataset, the parameters k and Z are determined to be 3 and 1% in our experiment, respectively.

d) Salinas: Similar to the OA results on Pavia University dataset, Table V reveals that the best performance can be achieved when $k = 3$ and $Z = 1\%$. Through the above analysis, it can be concluded that the values of two parameters have little influence on the whole, except for Botswana dataset.

2) Efficiency Analysis of Each Part: With respect to the proposed method, it mainly consists of two parts: neighborhood band grouping and maximum local density and informative band. To verify the efficiency of each part, we separate two parts and conduct the experiment separately. Specifically, without neighborhood band grouping, all the bands are viewed as one group. According to Eq. (11), the product of local density and information entropy for each band is calculated, and they are ranked in descending order. The bands with larger value are selected as a subset. This algorithm is named MLDIB for ease of description. After conducting neighborhood band grouping, one band is selected in each group. In order to effectively illustrate the effectiveness of ranking step, E-FDPC [10] is adopted instead of our ranking method. Similarly, the method is named NBG_E-FDPC. In addition, we combine the two factors of E-FDPC (local density ρ and distance D) with the two factors proposed in this paper (local density ρ and information entropy H). To distinguish the local density in the two methods, the local density in our paper is defined as $e\rho$. Finally, two competitors, NBG_ρ_H and NBG_$e\rho$_D, are added to the comparison algorithm.

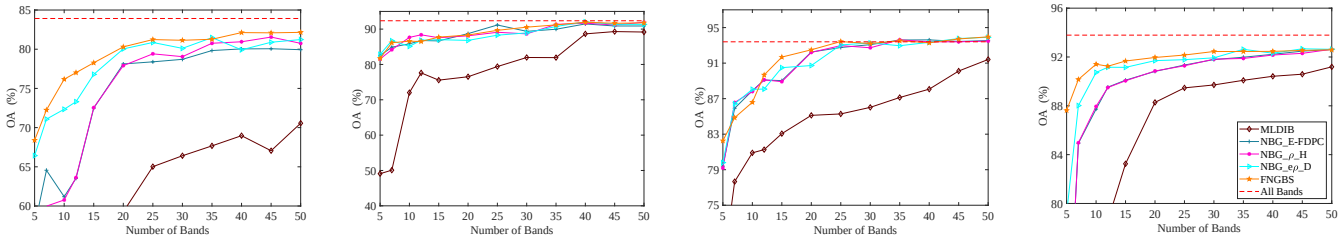
Fig. 5 shows the results for different parts by SVM classifier on four datasets. As can be seen from the figure, the performance of MLDIB is the worst and much lower than that of FNGBS. This also shows that neighborhood band grouping plays a key role in the algorithm, resulting in a significant improvement in performance. With respect to NBG_E-FDPC, the measure of ranking proposed in this study is superior to it. In our view, NBG_E-FDPC uses E-FDPC to ranking within each group. Although it can select representative band, it ignores maximum informative band, resulting in poor performance. As for the combinations about two factors, the results of NBG_$e\rho$_D is better than that of NBG_ρ_H and NBG_E-FDPC. It reveals the local density in our work can effectively explore the distribution between each band and other bands. When the combination of information entropy and local density proposed in [10], the superiority of its performance is not so obvious. Among these competitors, we can notice that our method still maintains excellent performance across different datasets. In particular, when a small number of bands are selected, the advantage of our algorithm is more clearer. To sum up, these two steps can significantly improve the performance of the algorithm.

TABLE IV
CLASSIFICATION RESULTS BY SETTING DIFFERENT PARAMETERS ON PAVIA UNIVERSITY DATASET USING SVM CLASSIFIER (%)

k	Z	5	7	10	12	15	20	25	30	35	40	45	50
2	1%	80.57	85.65	85.20	89.69	89.83	90.29	93.02	93.35	93.16	93.34	93.46	93.81
	10%	80.32	85.02	85.31	90.05	89.52	90.09	93.09	93.19	92.99	93.30	93.92	93.63
	50%	81.32	85.19	85.48	89.81	89.82	90.28	93.08	93.39	93.20	93.48	93.80	93.71
	100%	80.37	85.23	84.87	89.79	89.80	90.54	92.89	93.06	92.88	93.25	93.62	93.79
3	1%	82.39	84.09	86.63	89.54	91.71	92.46	93.22	93.21	93.17	93.63	93.52	93.65
	10%	82.22	84.06	86.31	89.07	91.61	92.31	93.14	93.03	93.43	93.51	93.82	93.89
	50%	82.12	84.09	86.31	89.33	92.04	92.19	93.06	93.12	93.32	93.68	93.87	93.58
	100%	82.08	84.52	86.82	89.24	91.84	92.45	92.91	93.30	93.48	93.55	93.69	93.89
4	1%	81.39	84.04	85.68	89.16	91.15	92.02	93.03	92.83	93.76	94.01	94.00	93.65
	10%	82.14	84.59	86.23	89.24	91.31	92.07	93.09	93.15	93.56	93.77	93.59	93.78
	50%	81.05	85.18	86.11	89.24	91.20	92.10	93.11	93.40	93.44	93.38	93.57	93.89
	100%	81.20	84.84	85.96	89.38	91.22	92.19	92.99	93.14	93.56	93.64	93.89	93.70

TABLE V
CLASSIFICATION RESULTS BY SETTING DIFFERENT PARAMETERS ON SALINAS DATASET USING SVM CLASSIFIER (%)

k	Z	5	7	10	12	15	20	25	30	35	40	45	50
2	1%	88.12	88.24	90.77	91.39	91.28	91.79	92.05	92.09	92.29	92.45	92.41	92.66
	10%	88.22	88.21	90.70	91.14	91.64	91.83	92.01	92.27	92.16	92.29	92.59	92.68
	50%	87.69	88.30	90.89	91.35	91.56	91.79	91.99	92.15	92.05	92.39	92.30	92.64
	100%	87.76	88.33	90.89	91.39	91.54	91.72	91.98	92.38	92.37	92.40	92.53	92.73
3	1%	87.91	89.25	91.01	90.93	91.71	91.85	91.81	92.07	92.24	92.45	92.73	92.70
	10%	87.68	89.08	91.17	90.90	91.86	91.82	91.87	92.27	92.55	92.43	92.68	92.59
	50%	87.92	89.21	90.88	91.21	91.77	91.81	91.98	92.10	92.54	92.58	92.67	92.51
	100%	87.86	89.31	90.75	91.20	91.53	91.63	92.10	92.08	92.71	92.64	92.60	92.63
4	1%	87.76	89.46	91.06	91.26	91.69	91.55	91.76	92.23	92.31	92.35	92.76	92.57
	10%	87.97	89.43	91.01	91.20	91.34	91.68	92.07	92.36	92.50	92.37	92.45	92.41
	50%	88.06	89.40	91.11	91.30	91.23	91.63	92.16	92.19	92.52	92.38	92.59	92.58
	100%	87.69	89.24	91.08	91.38	91.55	91.69	91.84	92.11	92.43	92.23	92.45	92.63



(a) (b) (c) (d)

Fig. 5. Classification results for different parts on four datasets. (a)-(d) Results by SVM classifier on Indian Pines, Botswana, Pavia University, and Salinas dataset, respectively.

3) Recommended Bands Comparison: Aiming at how many bands to choose in the process of band selection, the Eq. (17) is used as the discriminant condition to obtain the minimum number of recommended bands. Fig. 6 provides the number of recommended bands on four datasets, as shown by red dot. For four datasets, the number of recommended bands is 6, 8, 13, and 6, respectively. In this section, we will briefly analyze recommended bands.

Hyperspectral image cube has a remarkable feature that there is strong similarity among adjacent bands, which is usually adopted to subjectively evaluate the quality of selected bands. If many adjacent bands are acquired, it means that the algorithm is poor and the classification results are low. For four public datasets, the recommended bands by the proposed FNGBS are displayed in Table VI. Similarly, according to the number of bands recommended by the proposed method, we also show the bands of other competitors in Table VI when 6, 8, 13, and 6 bands are selected on four datasets. Moreover, the results of all competitors are ranked by ascending order. From the comparison of numerical distribution, some bands obtained by SNNC, E-FDPC, and WaLuDi are adjacent on Pavia University dataset. Unlike this dataset, all methods can avoid acquiring adjacent bands on Indian Pines and Botswana datasets. For Salinas dataset, most bands selected by each algorithm are the same or next to each other, which is consistent with the above results in Fig. 10. In this condition,

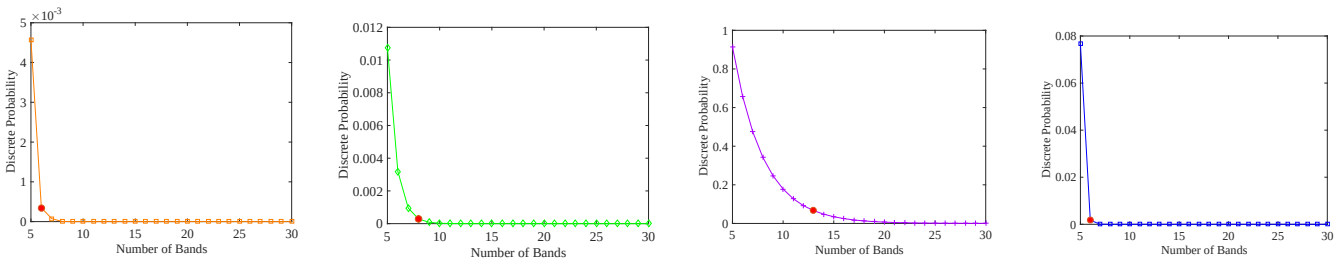


Fig. 6. Discrete probabilities of different band subsets. (a)-(d) Results on Indian Pines, Botswana, Pavia University, and Salinas dataset, respectively.

TABLE VI
SELECTED BANDS BY DIFFERENT BAND SELECTION METHODS ON FOUR DATASETS

Dataset	Method	Selected bands
Indian Pines (6 bands)	E-FDPC	28, 67, 83, 124, 167, 192
	WaLuDi	50, 67, 83, 128, 163, 189
	SNNC	28, 52, 67, 83, 124, 167
	TOF	16, 28, 50, 67, 90, 173
	FNGBS	28, 70, 92, 107, 129, 163
Botswana (8 bands)	E-FDPC	11, 20, 53, 65, 92, 106, 126, 138
	WaLuDi	2, 43, 50, 72, 92, 107, 116, 120
	SNNC	11, 19, 42, 53, 65, 71, 96, 129
	TOF	11, 20, 42, 53, 65, 92, 120, 128
	FNGBS	10, 20, 40, 52, 64, 97, 107, 136
Pavia University (13 bands)	E-FDPC	19, 33, 37, 42, 50, 52, 54, 56, 61, 82, 90, 92, 99
	WaLuDi	1, 9, 19, 23, 29, 33, 49, 61, 68, 72, 74, 78, 90
	SNNC	18, 28, 33, 48, 54, 62, 81, 84, 89, 92, 100, 102, 103
	TOF	4, 15, 19, 29, 33, 41, 48, 53, 61, 67, 79, 88, 99
	FNGBS	4, 12, 14, 29, 37, 40, 46, 53, 64, 70, 80, 87, 93
Salinas (6 bands)	E-FDPC	32, 55, 69, 88, 111, 185
	WaLuDi	68, 90, 136, 155, 194, 218
	SNNC	32, 53, 69, 90, 158, 184
	TOF	11, 32, 55, 68, 88, 109
	FNGBS	31, 55, 118, 126, 187, 212

FNGBS can select more dispersed subsets, which can not only cover a large spectrum range, but also has high recognition ability for objects on four datasets.

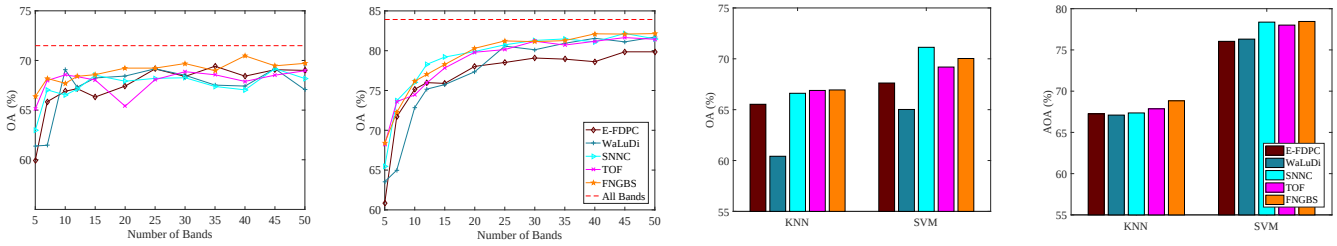
4) Classification Performance Comparison: In this section, we evaluate the proposed approach from three aspects by two classifiers, using four state-of-the-art methods. These includes 1) OA curves against different number of bands by two classifiers; 2) OA bars with respect to a certain number of bands; and 3) AOA bars about the range of 5 to 50 bands.

a) Indian Pines: The results about three aspects are demonstrated in Fig. 7. According to the curve, it can be concluded that our FNGBS can outperform better results in every aspect. Correctly, compare with the results collected by SVM classifier, all competitors except our method obtain unstable accuracy curves in Fig. 7(a), particularly as for TOF and WaLuDi. Moreover, the result (77%) for E-FDPC is significantly lower than other approaches after the number of bands exceeds 15 (see Fig. 7(b)). When 6 bands are selected, our method has comparable OA with SNNC using SVM classifier, but obviously superior to others. Through the above description, that is enough to illustrate our approach achieves superior performance over each dimension.

b) Botswana: Fig. 8 displays classification results by selecting different number of bands on Botswana dataset. Among all the competitors, our method achieves satisfactory results when using KNN classifier. Importantly, FNGBS can obtain stable results earlier than other competitors. With respect to SVM classifier, the classification curves demonstrate the classification accuracy increases with the increase of the number of selected bands. Overall, FNGBS also yields best performance than other methods. Unlike the results about KNN classifier, E-FDPC provides poor results (82%), which is about 5% lower than the classification accuracy of the other four algorithms. The remaining three methods (WaLuDi, SNNC, and TOF) can acquire better stable results on both classifiers. Fig. 8(c) reveals the classification accuracy when 8 bands are chosen. In Fig. 8(c), the proposed method displays a significant superiority.

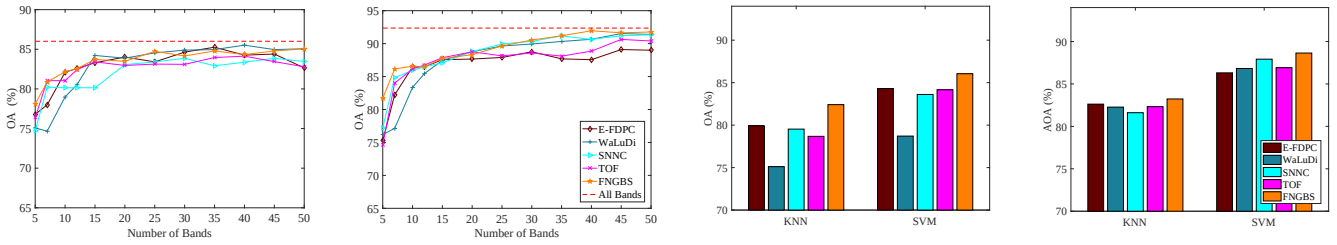
c) Pavia University: In this experiment, two classifiers are employed to classify these samples on Pavia University dataset, whose classification results are exhibited in Fig. 9. It can be observed that FNGBS has better performance when the number of selected bands surpasses 15. As for KNN classifier, the OA obtained by TOF and WaLuDi is about 3% lower than that of the other three algorithms on the whole. Besides, the results of SNNC, WaLuDi, and FNGBS even exceed the classification results of all bands. When it comes to SVM classifier, all competitors handle a small number of bands with lower results. Compared with the results by KNN classifier, WaLuDi and TOF display comparable performance to that of other algorithms. It demonstrates that the bands obtained by these two methods on this dataset are not very good. Fig. 9(c) presents selecting 13 bands acquire OA by five algorithms. Evidently, FNGBS attains approximately the same results as E-FDPC, but is more accurate than the other three approaches. In general, our method yields satisfactory classification results, which is comparable to the latest band selection methods.

d) Salinas: Fig. 10 shows the results using two classifiers on Salinas dataset. Different from the above two datasets, all band selection algorithms achieve approximately the same OA. However, our method can still better classify these samples,



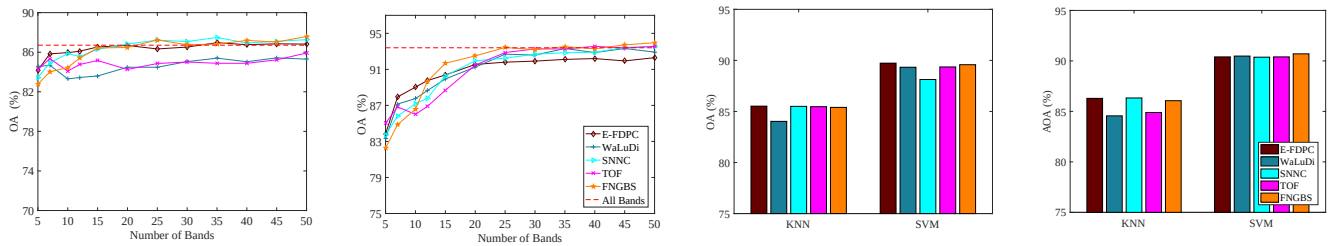
(a) (b) (c) (d)

Fig. 7. Classification results by selecting different number of bands on Indian Pines dataset. (a) and (b) Results by KNN and SVM classifiers. (c) Results about 6 bands. (d) Average classification results.



(a) (b) (c) (d)

Fig. 8. Classification results by selecting different number of bands on Botswana dataset. (a) and (b) Results by KNN and SVM classifiers. (c) Results about 8 bands. (d) Average classification results.



(a) (b) (c) (d)

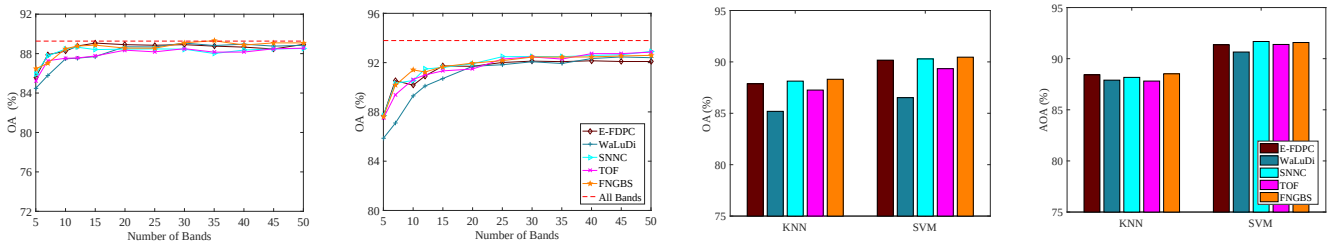
Fig. 9. Classification results by setting different number of bands on Pavia University dataset. (a) and (b) Results by KNN and SVM classifiers. (c) Results about 13 bands. (d) Average classification results.

but the difference is not very obvious. Fig 10(b) depicts that all methods are still very far from the accuracy of all bands. There are two main reasons for this issue. On one hand, this dataset has 224 bands, and we only show the results of some bands. On the other hand, for the four datasets, the parameters of SVM are set to the same. As one can see from the figure's results, these parameters are not appropriate for the dataset. In summary, we can draw some key conclusions that our method attains excellent and stable performance on different datasets, which demonstrates the proposed method is robust enough for band selection.

5) Computational Time Comparison: In this section, in order to evaluate the execution time of our proposed method, all competitors are conducted in MATLAB 2016a using PC workstation (Intel Core i5-3470 CPU processor and 16GB RAM). Table VII only presents processing time for different methods by selecting a certain number of bands on four datasets. From the results of this table, one can notice that WaLuDi yields very large execution time across different datasets, and the remaining methods take less time to process. Among these methods, only the running time of TOF will increase as the increase of the number of selected bands. In addition, two versions of the proposed method (1%, 100% pixels are picked from the original pixels) about processing time are displayed. Obviously, the time required to select part of the pixels of the band to perform the band selection is significantly small. Although our method is relatively time-consuming than E-FDPC, it can achieve promising results in classification performance and other aspects. To sum up, the proposed algorithm not only can execute faster, but also can yield better classification performance than others across four datasets.

V. Conclusion

Considering that the context information of the spectral information is rarely used for band selection, we propose a fast neighborhood grouping method for hyperspectral band selection (FNGBS), claiming the following contributions: 1) neighborhood band grouping is proposed to divide the hyperspectral image cube so that it can fully dig the context



(a) (b) (c) (d)

Fig. 10. Classification results by selecting different number of bands on Salinas dataset. (a) and (b) Results by KNN and SVM classifiers. (c) Results about 6 bands. (d) Average classification results.

TABLE VII
EXECUTION TIME OF DIFFERENT METHODS BY SELECTING A CERTAIN NUMBER OF BANDS ON FOUR DATASETS (S)

Dataset	E-FPDC	WaLuDi	SNNC	TOF	FNGBS (1%)	FNGBS (100%)
Indian Pines (6 bands)	0.121	7.430	0.4411	0.4165	0.2542	0.2995
Botswana (8 bands)	0.661	99.281	3.738	1.843	0.892	3.442
Pavia University (13 bands)	0.282	27.930	1.201	0.925	0.336	1.421
Salinas (6 bands)	0.381	40.382	1.561	1.276	0.465	1.464

information of spectral bands; 2) the representatives with maximum weight of the product of local density and information entropy are selected as band subset, which ensures band having informative and most relevant in each group; 3) determinantal point process is adopted to automatically determine how many bands are selected to achieve results similar to all bands. Extensive experiments on widely used benchmark datasets demonstrate that our FNGBS attains better accuracy against the state-of-the-art methods. In the future, we plan to study how to reconstruct these bands after partitioned groups.

REFERENCES

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

[2] X. Li, Z. Yuan, and Q. Wang, "Unsupervised deep noise modeling for hyperspectral image change detection," *Remote Sens.*, vol. 11, no. 3, 2019.

[3] R. J. Ellis and P. W. Scott, "Evaluation of hyperspectral remote sensing as a means of environmental monitoring in the st. austell china clay (kaolin) region, cornwall, uk," *Remote Sens of Environ.*, vol. 93, no. 1-2, pp. 118–130, 2004.

[4] Q. Tong, Y. Xue, and L. Zhang, "Progress in hyperspectral remote sensing science and technology in china over the past three decades," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 70–91, 2013.

[5] C. Yu, L.-C. Lee, C.I. Chang, B. Xue, M. Song, and J. Chen, "Band-specified virtual dimensionalit for band selection: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2822–2832, 2018.

[6] S. B. 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, 2001.

[7] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, no. 6, pp. 1157–1182, 2003.

[8] Q. Li, Q. Wang, and X. Li, "An efficient clustering method for hyperspectral optimal band selection via shared nearest neighbor," *Remote Sens.*, vol. 11, no. 3, pp. 350, 2019.

[9] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, 2019.

[10] 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, 2015.

[11] Adolfo A. MartÍnez-UsÓMartinez-Uso, F. Pla, j. m. Sotoca, P. José Martínez, and García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, 2007.

[12] Y. Yuan, J. Lin, and Q. Wang, "Dual clustering based hyperspectral band selection by contextual analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 1431–1445, 2016.

[13] J. C. Wu and G. C. Tsuei, "Unsupervised cluster-based band selection for hyperspectral image classification," *in Proc. Int. Conf. Adv. Comput. Sci. Electron. Inf.*, pp. 562–565, 2013.

[14] A. Datta, S. Ghosh, and A. Ghosh, "Combination of clustering and ranking techniques for unsupervised band selection of hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2814–2823, 2015.

[15] 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, 1999.

[16] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Geosci. Remote Sens.*, vol. 27, no. 6, pp. 1279–1289, 2017.

[17] 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, 2008.

[18] Q. Wang, Z. Qin, F. Nie, and X. Li, "Spectral embedded adaptive neighbors clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1265–1271, 2019.

[19] S. Jia, Y. Qian, and Z. Ji, "Band selection for hyperspectral imagery using affinity propagation," *IET Comput. Vis.*, vol. 3, no. 4, pp. 213–222, 2009.

[20] A. M. Uso, F. Pla, J. M. Sotoca, and P. Garcia-Sevilla, "Clustering-based multispectral band selection using mutual information," *in Proc. IEEE Int. Conf. Pattern Recog.*, vol. 2, pp. 760–763, 2006.

[21] C.-I Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, 2006.

[22] S. Kang, X. Geng, and L. Ji, "Exemplar component analysis: A fast band selection method for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 998–1002, 2015.

[23] T. Lu, S. Li, L. Fang, Y. Ma, and J. A. Benediktsson, "Spectral-spatial adaptive sparse representation for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 373–385, 2016.

[24] 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, 2014.

[25] C. I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, 2006.

[26] J. Yin, Y. Wang, and Z. Zhao, "Optimal band selection for hyperspectral image classification based on inter-class separability," in *Photonics and Optoelectronic (SOPO), 2010 Symposium on*, 2010, pp. 1–4.

[27] Y. Gu and Y. Zhang, "Unsupervised subspace linear spectral mixture analysis for hyperspectral images," in *Proceedings of the 2003 International Conference on Image Processing*, 2003, pp. 801–804.

[28] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, 2018.

[29] Cardoso and Fran Jean, "Dependence, correlation and gaussianity in independent component analysis," *Journal of Machine Learning Research*, vol. 4, pp. 1177–1203, 2003.

[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, 2006.

[31] L. Ma, M. M. Crawford, and J. Tian, "Local manifold learning-based$k$-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4099–4109, 2010.

[32] R. Liu, H. Wang, and X. Yu, "Shared-nearest-neighbor-based clustering by fast search and find of density peaks," *Inf. Sci.*, vol. 450, pp. 200–226, 2018.

[33] S. Singh and A. Awekar, "Incremental shared nearest neighbor density-based clustering," *22nd ACM International Conference on Information and Knowledge Management*, vol. 450, pp. 1533–1536, 2013.

[34] T. M. Cover and J. A. Thomas, "Elements of information theory," 1991.

[35] A. Webb, "Statistical pattern recognition, 2nd ed. hoboken, nj: Wiley," 2002.

[36] K. Alex, T. Ben, et al., "Determinantal point processes for machine learning," *Foundations and Trends® in Machine Learning*, vol. 5, no. 2–3, pp. 123–286, 2012.

[37] B. Gong, W.-L. Chao, Kristen. Grauman, and S. Fei, "Diverse sequential subset selection for supervised video summarization," *Advances in Neural Information Processing Systems*, pp. 2069–2077, 2014.

[38] K. Zhang, W.-L. Chao, F. Sha, and K. Grauman, "Video summarization with long short-term memory," *European conference on computer vision*, pp. 766–782, 2016.

[39] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, 1967.

[40] N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines and other kernel-based learning methods," *Cambridge, U.K.: Cambridge Univ. Press*, 2010.

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

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

Qiang Li received the B.E. degree in measurement & control technology and instrument from Xi'an University of Posts and Telecommunications, Xi'an, China, in 2015, and the M.S. degree in communication and transportation engineering from Chang'an University, Xi'an, China, in 2018.

He is currently pursuing the Ph.D. degree with the School of Computer Science and the Center for OPTical IMagery Analysis and Learning. His research interests include hyperspectral image processing and computer vision.

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