Research Article

Hyperspectral Band Selection Based on Adaptive Neighborhood Grouping and Local Structure Correlation

Xin Wang and Guoqiang Wang

College of Electronic Engineering, Heilongjiang University, Heilongjiang Harbin 150080, China

Correspondence should be addressed to Guoqiang Wang; wangguoqiang@hlju.edu.cn

Received 6 January 2021; Revised 7 April 2021; Accepted 12 April 2021; Published 22 April 2021

Academic Editor: Carlos Michel

Copyright © 2021 Xin Wang and Guoqiang Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Band selection is a direct and effective dimension reduction method and is one of the hotspots in hyperspectral remote sensing research. However, most of the methods ignore the orderliness and correlation of the selected bands and construct band subsets only according to the number of clustering centers desired by band sequencing. To address this issue, this article proposes a band selection method based on adaptive neighborhood grouping and local structure correlation (ANG-LSC). An adaptive subspace method is adopted to segment hyperspectral image cubes in space to avoid obtaining highly correlated subsets. Then, the product of local density and distance factor is utilized to sort each band and select the desired cluster center number. Finally, through the information entropy and correlation analysis of bands in different clusters, the most representative bands are selected from each cluster. Regarding evaluating the effectiveness of the proposed method, comparative experiments with the state-of-the-art methods are conducted on three public hyperspectral datasets. Experimental results demonstrate the superiority and robustness of ANG-LSC.

1. Introduction

Hyperspectral images (HSI), as a rich spectral information source, can accurately describe objects and are widely used in various fields such as marine exploration, military target detection, forestry, and hydrology [1–4]. However, the high-dimensionality of hyperspectral data and the scarcity of labeled samples make the classification of hyperspectral data a challenging subject. Redundant adjacent bands in a hyperspectral image cause storage and transmission challenges, increasing computational complexity, and commonly, can decrease the performance of the classifiers. Hence, it is still a challenging task to remove redundant and insignificant information within a reasonable time in these complex situations.

Feature extraction and feature selection (also known as “band selection”) are two of the most widely used dimensionality reduction strategies. Feature extraction mainly relies on finding an embedding and then projects the original high-dimensional data into a lower-dimensional feature space

[5]. However, through spatial transformation, the physical meaning of the original hyperspectral data is changed, and some key information is lost. The feature selection method [6] reduces the dimensionality of HSI by finding the most representative band to form a data subset, i.e., a group of most important bands is selected from all spectral bands to represent the entire spectrum. This way preserves the physical meaning of the original spectral data and facilitates the interpretation of the selected datasets. Therefore, the band selection method is very suitable for dimensionality reduction of hyperspectral data.

So far, a large number of band selection methods have been proposed. According to whether the training samples are utilized, these methods can be roughly grouped into supervised [7], unsupervised [8], and semisupervised [9]. Supervised and semisupervised methods rely heavily on supervised information to identify relevant bands. In general, unsupervised methods select representative bands by exploring the nature of samples without the prior information associated with labels. However, tagging samples is an expensive,

tedious, and time-consuming task. In contrast, unsupervised methods require no labeling. Therefore, unsupervised band selection becomes a highly popular research topic.

In the past few years, many unsupervised band selection methods have been proposed. They can be divided into four categories: ranking-based [10], clustering-based [11], searching-based [12], and sparsity-based [13]. Because the interaction between bands is fully considered in the clustering method, the band selection based on clustering can obtain accurate results, which attracts more and more attention. These methods can obtain satisfying results, but they always have two inherent shortcomings in the clustering process. On the one hand, most of the methods only consider the correlation between bands and ignore the subset information of the selected bands. On the other hand, according to the characteristics of the hyperspectral image cube, it is found that these bands are arranged in an orderly manner. In addition, for a certain band, within a certain range, the correlation with the adjacent band is stronger, and the correlation with the farther band is lower. Therefore, it can be concluded that discontinuous bands of different wavelength ranges cannot be grouped into clusters for band selection.

Motivated by the above descriptions, a band selection method based on adaptive neighborhood grouping and local structure correlation is proposed in this paper. Unlike the SNNC [14] and SNNCA [15] methods, we treat the bands of hyperspectral images as ordered. Moreover, without changing the original hyperspectral image data, the hyperspectral image cube is divided into several subcubes, and the relevant information bands in each group are obtained using the idea of sorting. The main contributions are as follows:

- (1) Based on the fact that adjacent bands have high redundancy, the ordered hyperspectral bands are partitioned into multiple subcubes by the clustering algorithm, which can effectively avoid obtaining a subset with high correlation
- (2) The product of local density and distance factor is used to sort and select the number of clustering centers required for each band to ensure the low redundancy of clustering centers. It can make better use of the local distribution feature and make the obtained subset have more recognition bands at the same time
- (3) Through the information entropy and correlation analysis of the bands in different clusters, the most representative bands are selected from each cluster. This method does not directly select the cluster center as the selected band but considers the combination of bands in different clusters to avoid falling into the local optimal solution

The rest of this article is organized as follows. In Section 2, the development of related technologies of the proposed method is briefly introduced. Section 3 introduces the proposed band selection method in detail, including the algorithm principles of adaptive subspace division and neighborhood grouping. Section 4 gives experimental results

and discussion. Conclusions and future goals will be given in Section 5.

2. Related Work

The clustering-based method groups the original bands and selects representative bands from each cluster to form the final subset of bands. The algorithm can minimize intraclass variance and maximize interclass variance simultaneously. Most clustering-based methods are derived from k-means, affinity propagation (AP), and graph clustering. A few typical methods are briefly introduced below.

The k-means algorithm is a widely used clustering technique. It is initialized with a randomly selected set of bands and iteratively optimizes the objective function until the optimal cluster centers are found. In [16], the authors proposed a new statistical method for band clustering and band selection based on k-means clustering. The algorithm is based on interquartile range, geometric mean, median absolute deviation, median, correlation coefficient, covariance, and mode. Hyperspectral data can be calculated by these approaches. The authors in [17] proposed a new spectral band selection method, i.e., representative band mining. In this method, disjoint information is adopted to measure the distance between two spectral bands. To tackle the inherent drawbacks of the clustering-based band selection method, Yuan et al. [18] proposed a new band selection framework for dual clustering, which included a dual clustering method with context information in the clustering process and considered the mutual influence of various bands. In [19], the authors proposed a fast clustering algorithm (FDPC) based on density peaks. According to the calculation of the local density of each point and the distance within the cluster, the product of the two factors is sorted in descending order, and the cluster center is identified as a point with an abnormally large value. The authors in [20] enhanced the FDPC algorithm to make it suitable for hyperspectral band selection. First, the score of each band is calculated by weighting the normalized local density and the intracluster distance of each point instead of considering them equally. Then, an exponential-based learning rule is employed to adjust the cutoff threshold for a different number of selected bands, which is fixed in the FDPC. Therefore, the proposed method is called Enhanced FDPC (E-FDPC). In [14], the authors proposed a hyperspectral optimal selection clustering method based on shared nearest neighbor (SNNC). They used the local density of each band to reflect the local distribution characteristics and used information entropy as a weighting factor to design a method to automatically select the optimal band subset. Inspired by the SNNC method, [21] proposed the SNNCA method, which considered the interaction of bands in different clusters, and obtained a set of bands with large amount of information and low redundancy.

Since the clustering of the k-means algorithm is sensitive to the initial conditions, therefore, the exemplar-based AP clustering algorithm was proposed to search for an appropriate set of exemplars as representative bands. [22] proposed a new method of semisupervised band selection based on affinity propagation (AP). This algorithm proposes a new

FIGURE 1: Schematic diagram of adaptive subspace principle partition.

normalized trivariable mutual information (NTMI) to measure band correlation for classification, taking into account not only band redundancy but also band synergy. The authors in [23] constructed the similarity matrix based on several distance measures proposed but did not consider the spatial structure information of the image. A structural method for evaluating image quality is proposed to establish the similarity between band images for band selection. [24] proposed a new affinity propagation (AP) technology based on distinguishing feature metrics. The spectral and spatial relationships between pixels are constructed by a new type of discrimination constraint. A discriminative feature measure (DFM) is proposed, and the discriminative constraint is modeled in terms of the best criterion for identifying the learning method of effective distance measurement, and then, the representative subset of bands is identified based on the AP clustering algorithm.

Graph clustering expresses band selection as a graph problem. The nodes in the graph represent HIS bands, and the edge connecting two nodes corresponds to the similarity between the two bands. A graph-based clustering method, by constructing an affinity matrix related to the similarity of the bands, the graphs are clustered into subgraphs to find representative bands. In [25], the authors proposed a novel method to address the problem of hyperspectral band selection. The principle is to create a band adjacency graph, where nodes represent bands, and edges represent similarity weights between bands. A series of random matrices are defined by the alternation of two operators on the affinity matrix to form different clusters of high correlation bands. In [21], the authors proposed a multigraph determinantal point process (MDPP) model to capture the complete structure between different bands. MDPP employs multiple graphs to capture the intrinsic relationship between hyperspectral bands and provides an effective search strategy to select the best band.

3. Proposed Method

In this section, we introduce the proposed method (ANG-LSC) in detail. Specifically, based on the fact that the hyper-

spectral bands are ordered, an adaptive subspace strategy is used to divide the hyperspectral image cube, and then, the bands with similar spectral characteristics are adaptively divided into a subcube, using local density, and the product of the distance factor sorts the bands and selects the required number of cluster centers. Finally, through information entropy and correlation analysis of the bands in different clusters, the most representative bands are selected from each cluster.

3.1. Adaptive Subspace Partition. There are two common problems in the selection of hyperspectral bands. First, if the band selection is directly performed on the entire hyperspectral image cube, it will consume a lot of time and computer resources. Second, the correlation between adjacent bands is higher than that of nonadjacent bands. Therefore, it is necessary to perform a simple segmentation of the hyperspectrum before selecting the band.

Suppose that the hyperspectral dataset is $X_i = [x_1, x_2, \dots, x_L] \in R^{N \times L}$, where N is the number of pixels and L is the total number of bands. $x_i = [x_{1i}, x_{2i}, \dots, x_{Ni}]$ is a vector composed of the i th band. In order to reduce the calculation time, we need to divide the hyperspectral image cube into equal-width subcubes, i.e., if the number of bands we need to select is K , then the number of bands of the subcube $P_i \in R^{N \times Z}$ is

$$Z = \frac{L}{K}. \quad (1)$$

Next, the Euclidean distance is used to construct a similarity matrix between the i th band and the j th band.

$$D(x_i, x_j) = \|x_i - x_j\|_2. \quad (2)$$

In conventional clustering algorithms, the final clustering results are obtained by distance between and within classes. Inspired by this, this idea is used to further subdivide the already segmented coarse subcube. Considering

FIGURE 2: Flow chart of the band selection algorithm.

Input: Hyperspectral image $X_i = [x_1, x_2, \dots, x_L] \in R^{N \times L}$, the number of bands to be selected K , the size of the Gaussian window r .

Output: The number of selected bands n .

- 1: The original hyperspectral image cube is segmented into subcube P_i by adaptive subspace partition method.
- 2: Calculate the information entropy H_i of each band.
- 3: The information entropy matrix E is convolved with the Gaussian kernel function, and the shape is the same.
- 4: According to Equation (9), the similarity matrix between the two bands is calculated.
- 5: Calculate local density p_i according to Euclidean distance and similarity matrix.
- 6: By calculating the minimum distance between each band and other high-density bands, the distance factor coefficient δ_i of each band is obtained.
- 7: Take the product of the three factors as the comprehensive weight w_i of each band. After sorting the weights in descending order, select the expected number of optimal bands as the clustering center to construct the desired band subset.
- 8: Calculate the local structural similarity index S of each band in the subcube.
- 9: A new weight W_i is defined to reevaluate the quality of each band. According to Equation (16), the band with the largest weight is selected from each cluster as the representative band.

ALGORITHM 1: ANG-LSC for band selection.

FIGURE 3: Classification results of different window sizes on three datasets (OA).

FIGURE 4: Classification results of different window sizes on three datasets (AA).

that two subcubes far apart have little correlation, only two adjacent subcubes (P_i and P_{i+1}) are considered. The mathematical expression is

$$\arg \max_t \frac{D_{\text{inter}}}{D_{\text{intra}}}, \quad (3)$$

where D_{inter} and D_{intra} are the interclass distance and the intraclass distance, respectively, and t is the dividing point. The maximum distance is selected as the interclass dis-

FIGURE 5: Classification results of different window sizes on three datasets (Kappa).

tance D_{inter} . The intraclass distance is composed of the sum of U_1 of P_i and U_2 of P_{i+1} , which can be expressed as

$$\begin{aligned} D_{\text{inter}} &= \max |D_{ij}|, \\ D_{\text{intra}} &= U_1 + U_2, \end{aligned} \quad (4)$$

where U_1 and U_2 are, respectively,

$$\begin{aligned} U_1 &= \frac{1}{t(t-1)} \sum_{i=1}^t \sum_{j=1}^t D(x_i, x_j), \\ U_2 &= \frac{1}{(2Z-t-1)(2Z-t-2)} \sum_{i=t+1}^{2Z} \sum_{j=t+1}^{2Z} D(x_i, x_j). \end{aligned} \quad (5)$$

A new partition point can be obtained through the above equation, and then, the initial point t is updated in the same way to obtain the final division point (as shown in Figure 1, the example divides the hyperspectral image cube (eight bands) into four subcubes. In the dotted line, the dark area means that only these adjacent bands are considered to update the current segmentation point).

3.2. Adjacency Clustering and Local Structural Correlation. After the hyperspectral bands are divided into subcubes, the subcubes are usually processed independently. When selecting the representative band of each cube, traditional methods usually select the most relevant or most informative band in each cube. However, this strategy makes the representative band may not be representative. To address this problem, an effective method, LSC, which simultaneously selects the information band and the most relevant band, is adopted.

The adaptive subcube partition method can reduce the correlation between the subcubes and effectively avoid the selection of redundant bands. Next, the subcube is selected

TABLE 1: Different number of bands to classify three datasets.

Dataset	Measuring	5	10	15	20	25	30	35	40	45	50
Indian Pines	OA	80.90	80.60	80.16	79.62	79.25	78.77	80.66	79.31	80.24	79.43
	AA	79.31	80.64	79.75	78.72	78.27	77.31	78.02	76.40	79.08	76.76
	Kappa	78.17	77.87	77.34	76.76	76.31	75.82	77.95	76.45	77.49	76.57
Pavia University	OA	87.35	90.52	91.34	92.04	92.31	93.27	93.44	93.37	93.47	93.48
	AA	82.91	87.80	88.69	89.72	90.03	91.26	91.18	91.26	91.28	91.41
	Kappa	82.91	87.32	88.45	89.39	89.76	91.04	91.26	91.18	91.32	91.32
Salinas	OA	87.20	88.40	90.49	92.57	92.59	92.66	93.04	93.15	93.23	93.34
	AA	95.96	96.15	96.10	96.36	96.32	96.32	96.54	96.65	96.67	96.74
	Kappa	91.29	91.52	91.61	91.78	91.73	91.81	92.23	92.36	92.45	92.58

FIGURE 6: Indian Pines dataset classification results and selected band information entropy distribution map: (a) ground truth images, (b) classification result map, and (c) information entropy distribution map of each band.

to select the most relevant wavelength with the largest amount of information.

First, information entropy is used to evaluate the amount of information in each band, i.e.,

$$H_i = - \sum_{z \in \Omega} p(z) \log p(z), \quad (6)$$

where Ω is a particular band of gray space and $p(z)$ denotes the probability of a certain gray level in the band. The information entropy H_i of each band can be calculated according to the gray histogram.

Next, a Gaussian filter is used to kernel transform the given input matrix E into E^* . Using the spatial information of the matrix, the transformed spectral value is calculated

FIGURE 7: Pavia University dataset classification results and selected band information entropy distribution map: (a) ground truth images, (b) classification result map, and (c) information entropy distribution map of each band.

through the matrix pixel value. The two-dimensional Gaussian distribution is

$$G(x, y) = \frac{1}{2\pi\sigma^2} \left[-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2} \right], \quad (7)$$

where σ is the standard deviation and (x_0, y_0) is the sample mean of the variables x and y . Two-dimensional Gaussian filtering is performed on the input matrix, and the Gaussian kernel width is set as R , then

$$E^*(x, y) = \frac{1}{\sum_{u=-r/2}^{r/2} \sum_{v=-r/2}^{r/2} G(u, v)}, \sum_{u=-r/2}^{r/2} \sum_{v=-r/2}^{r/2} E(x+u, y+v) G(u, v), \quad (8)$$

where $E(x+u, y+v)$ is the pixel value of the pixel in the neighborhood, $E^*(x, y)$ is the filtered result, and (u, v) is the position coordinate of the pixel in the core area. The Gaussian kernel function is a discrete approximation of con-

tinuous Gaussian. According to the characteristics of Gaussian distribution and data distribution, the width of the sliding window and the standard deviation σ are determined.

Then, calculate the similarity between the two bands, i.e.,

$$\text{SNN}(x_i, x_j) = |\text{KNN}(x_i) \cap \text{KNN}(x_j)|, \quad (9)$$

where $\text{SNN}(x_i)$ represents the number of elements shared by the x_i and x_j bands in the K -nearest neighbor sets. The local density ρ_i is calculated from the Euclidean distance and similarity matrix.

$$\rho_i = \sum_{j \in \text{KNN}(x_i)} \exp\left(-\frac{D(x_i, x_j)}{\text{SNN}(x_i, x_j) + 1}\right). \quad (10)$$

The distance factor σ_i of each band is obtained by calculating the minimum distance between each band and other

FIGURE 8: Salinas dataset classification results and selected band information entropy distribution map: (a) ground truth images, (b) classification result map, and (c) information entropy distribution map of each band.

high-density bands.

$$\delta_i = \begin{cases} \min_j (D(x_i, x_j)), & \rho_i < \rho_j, \\ \max_j (D(x_i, x_j)), & \rho_i = \max(\rho). \end{cases} \quad (11)$$

Then, take the product of the three factors as the comprehensive weight w_i of each band.

$$w_i = \rho_i \times \delta_i \times H_i. \quad (12)$$

After sorting the comprehensive weights of all bands in descending order, the expected number of optimal bands is selected as the clustering center to construct a subset of the desired bands.

Finally, the optimal band is selected by measuring the local structure similarity index and information entropy of image quality, where the local structure similarity index (SSIM, abbreviated as S below) is

$$S(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma, \quad (13)$$

where

$$\begin{aligned} l(x, y) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \\ c(x, y) &= \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \\ s(x, y) &= \frac{\sigma_{xy} + C_3}{\sigma_x + \sigma_y + C_3}, \end{aligned} \quad (14)$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y$, and σ_{xy} are the local means, standard deviations, and cross-covariance for images x, y . If $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$, then

$$S(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (15)$$

FIGURE 9: Classification results of the different number of bands in the Indian Pine dataset: (a) KNN and (b) SVM.

TABLE 2: OA and standard deviation of different methods on three datasets.

Dataset	Classifier (measure)	ONR	OCF	SOPSRL	ASPS_MN	MDSR	ANG_LSC
Indian Pine	KNN (OA)	70.46 ± 0.25	70.07 ± 1.03	68.23 ± 0.44	68.96 ± 0.11	56.59 ± 0.21	72.22 ± 0.55
	KNN (Kappa)	65.74 ± 0.34	64.62 ± 1.19	63.95 ± 0.57	64.34 ± 0.13	49.36 ± 0.33	68.09 ± 0.64
	SVM (OA)	79.56 ± 0.80	77.03 ± 0.54	74.59 ± 0.61	76.96 ± 0.29	67.59 ± 0.11	80.16 ± 0.11
	SVM (Kappa)	76.64 ± 0.88	73.08 ± 0.67	71.37 ± 0.70	75.13 ± 0.36	61.17 ± 0.09	77.34 ± 0.09
Pavia University	KNN (OA)	86.65 ± 0.16	85.58 ± 0.13	85.65 ± 0.26	85.66 ± 0.13	84.46 ± 0.26	86.98 ± 0.15
	KNN (Kappa)	83.04 ± 0.26	83.72 ± 0.40	82.87 ± 0.39	81.34 ± 0.18	78.49 ± 0.35	82.39 ± 0.22
	SVM (OA)	90.99 ± 0.08	91.71 ± 0.31	90.17 ± 0.04	88.56 ± 0.25	90.62 ± 0.18	91.34 ± 0.16
	SVM (Kappa)	88.39 ± 0.13	87.64 ± 0.46	88.31 ± 0.10	85.90 ± 0.36	87.32 ± 0.23	88.45 ± 0.21
Salinas	KNN (OA)	89.38 ± 0.12	89.24 ± 0.11	88.15 ± 0.04	89.27 ± 0.13	85.12 ± 0.16	87.87 ± 0.14
	KNN (Kappa)	87.48 ± 0.05	86.17 ± 0.32	85.03 ± 0.01	87.62 ± 0.14	83.46 ± 0.17	86.48 ± 0.16
	SVM (OA)	92.45 ± 0.08	92.06 ± 0.11	90.08 ± 0.01	91.88 ± 0.13	89.56 ± 0.06	92.49 ± 0.06
	SVM (Kappa)	90.45 ± 0.12	90.82 ± 0.14	89.21 ± 0.01	88.67 ± 0.15	86.58 ± 0.07	91.61 ± 0.07

In order to find a representative subset of bands with large information and low redundancy, a new weight W_i is defined to evaluate the quality of each band. The specific equation is

$$W_i = SUM(S) + H_i. \quad (16)$$

According to the above formula, the band with the largest weight is selected from each cluster as the representative band.

The algorithm flow chart of this paper is shown in Figure 2. For more details about the algorithm, the procedures are summarized in Algorithm 1.

4. Experiment

In this section, we will conduct a large number of experiments to evaluate the superiority of the proposed algorithm in hyperspectral image classification. First, we introduce three common hyperspectral image datasets. Then, the experimental environment, evaluation criteria, and comparison algorithm are described. Finally, we analyze the experimental performance of several methods in detail, and the

FIGURE 10: Continued.

FIGURE 10: Classification results of different algorithms on the Indian Pines dataset: (top) KNN and (bottom) SVM.

experimental results show the effectiveness and superiority of the proposed method.

4.1. Datasets

- (1) Indian Pines: this scene was gathered by the AVIRIS sensor over the Indian Pines test site in North-western Indiana and consists of 145×145 pixels and 224 spectral reflectance bands in the wavelength range $0.4\text{--}2.5\ \mu\text{m}$. Remove the severe water absorption and low signal-to-noise ratio of the bands (104~108, 150~163, and 220), and finally, the remaining 200 bands are used for the experiment
- (2) Pavia University: the scene of the University of Pavia was collected by the optical sensor of the reflection optical system imaging spectrometer (ROSIS-03), with a $0.43\text{--}0.86\ \mu\text{m}$ spectral range, 610×340 pixels, and 103 spectral bands. A total of 115 bands are included, 12 bands with severe water absorption are removed, and the remaining 103 bands are used
- (3) Salinas: Salinas is a famous subscene of Salinas image gathered by the 224-band AVIRIS sensor over Salinas Valley, CA, USA. The image pixel is 512×217 pixels, the spectral resolution is 10 nm, and the wavelength range is $0.4\text{--}2.5\ \mu\text{m}$. The water vapor absorption band

FIGURE 11: Classification results of the number of different bands in the Pavia University dataset: (a) KNN and (b) SVM.

and 20 bands with low signal-to-noise ratio are removed, i.e., 108-112, 154-167, and 224, the remaining 204 bands

4.2. Experimental Setup. To verify the feasibility and effectiveness of the algorithm in this paper, the following settings are required before the experiment.

4.2.1. Runtime Environment. The experimental environment is the tenth-generation Intel Core i7-10750H six-core processor, the main frequency is 2.60 Hz, the effective memory is 16 GB, and the development environment is in MATLAB R2016b.

4.2.2. Comparison Methods. In this paper, the proposed method (ANG-LSC) is compared with five state-of-the-art unsupervised band selection methods, ONR [26], OCF [27], SOPSRL [8], ASPS-MN [28], and MDSR [13].

4.2.3. Number of Selected Bands. Since the actual number of bands to be selected in the three public hyperspectral image datasets is unknown, experiments are carried out in the range of 5-50 bands (with an interval of 5) to explain the impact of the number of different bands on classification accuracy.

4.2.4. Classifier. In this experiment, KNN and SVM, two common classifiers, are used for classification. The parameter of the KNN classifier is set to 5. The SVM classifier employs the RBF kernel. Considering that these classifiers are supervised, we randomly select 10% of the samples from each category as the training set; the remaining 90% is the test set. To reduce the impact of randomly selecting 10% samples, the algorithm runs 10 times to obtain a mean result.

4.2.5. Accuracy Measures. This paper employs overall accuracy (OA), average overall accuracy (AA), and Kappa coefficient as the accuracy measure for hyperspectral image classification. The larger the OA, Kappa value, the better the image classification effect, and AA is a common indicator to measure the classification results of small categories.

4.3. Results. In this section, the effectiveness and superiority of the proposed algorithm are illustrated from three aspects, including Gaussian kernel size, number of selected bands, and classification performance to analyze the three datasets.

4.3.1. Gaussian Kernel Size. The size of the kernel is critical. If it is too large, the finer features of the image may be smoothed out. But kernel selection is too small to eliminate the noise. Therefore, in order to select the appropriate kernel size, the kernel was successively set to 3, 5, 7, 9, 11, 13, and 15 to conduct experiments on the three datasets. The experimental results of the three datasets are shown in Figures 3–5.

From Figures 3–5, we can draw this conclusion. As the window continues to increase, the changes of OA, AA, and Kappa of the Salinas dataset tend to be stable. However, it can be seen from the other two data that as the window continues to increase, the three indicators of OA, AA, and Kappa are constantly decreasing. In summary, the size of the kernel is chosen to be 3×3 in this experiment.

4.3.2. Number of Bands. In order to verify the band selection method proposed in this article, experiments are performed on three hyperspectral datasets in the range of 5-50 bands. Table 1 provides the classification results of the SVM classifier on the three datasets.

It can be seen from Table 1 that by setting the number of different bands, the values of OA, AA, and Kappa are gradually increasing. Although there is a temporary decline in

FIGURE 12: Continued.

FIGURE 12: Classification results of different algorithms on the Pavia University dataset: (top) KNN and (bottom) SVM.

some bands, the value of the decrease is almost negligible. Therefore, it can be considered that the evaluation indexes are still increasing with the increase of the band. Besides, the table shows that when the number of bands is 15-20, the best effect can be achieved for the three datasets. For the Indian Pines dataset, it is more appropriate to select a band of 15. For the Pavia University and Salinas datasets, when the number of bands is 20, OA has reached 92.04% and 92.57%, respectively. In addition, from the experiment of three hyperspectral image datasets, some key results can be summarized, i.e., the algorithm in this paper can realize the classification of different datasets through multiple classifiers and the information entropy distribution of each band,

and can obtain stable classification performance. The classification results are shown in Figures 6–8.

For the Indian Pines dataset, it can be seen from Figure 6(b) that the classification effect is not ideal, and there is a phenomenon of misclassification. Combined with the information entropy distribution diagram of the selected band, it can be seen that the number of bands in the interval 106-142 is too dense. The dense and large amount of information will easily affect the bands on both sides. Another reason is that for the Indian Pines dataset, there are too many features and features with small samples are not removed in the experiment. However, compared with previous methods for band selection, especially for low-dimensional data, the

FIGURE 13: Classification results of the different number of bands in Salinas dataset: (a) KNN and (b) SVM.

method presented in this paper has certain advantages in OA, AA, and Kappa. For the Pavia University and Salinas datasets, by comparing the truth map and the classification result map, it can be seen that the classification effect is very good, which exactly corresponds to the values in Table 1. It can also be seen from Figures 7 and 8 that there are few misclassifications, which basically meet the classification requirements.

4.3.3. Classification Performance. To prove the effectiveness of the proposed method, the method in this paper is compared with the four state-of-the-art algorithms. Two classifiers (KNN and SVM) are employed to analyze the hyperspectral image by using three accuracy evaluation criteria.

For the Indian Pines dataset, Figure 9 and Table 2 clearly show the fact that the method in this paper is better than other methods in the case of fewer bands, which exactly meets the purpose of dimensionality reduction. And in 5 bands, the OA of this method has reached 73.32%, which is 4.44% higher than the state-of-the-art ASPS-MN method. On the other hand, when the number of selected bands exceeds 25, although the OA of the ASPS-MN method in some bands will be higher than that of this method, it can be seen from Table 2 that the stability of this method on the SVM classifier is better than other ones. The corresponding standard deviations of OA and Kappa are ± 0.11 and ± 0.09 , respectively. In addition, the accuracy of this algorithm is better than other algorithms, and its advantages compared with MDSR are very obvious. In summary, the method in this paper has better stability and superiority, and the classification is really effective as shown in Figure 10.

For the data of Pavia University, Figure 11 and Table 2 show that the algorithm in this paper has good performance on both OA and Kappa indicators, reaching 91.34% and 88.45%, respectively (the classification result is shown in Figure 12). Compared with the ASPS-MN method, it is

improved by 2.78% and 2.55%, respectively. Compared with the other four methods, OA and Kappa can improve the maximum on the KNN classifier by 2.52% and 3.9%, respectively. When the number of selected bands is small, some algorithms show the instability of accuracy, especially MDSR and SOPSRL. All algorithms perform well when the number of selected bands exceeds 25. The classification advantage of this method in the Indian Pines dataset is not obvious, but it is well demonstrated in this dataset. Also, according to the SVM classification result curve, the advantages of this algorithm are more obvious than other algorithms. As for other algorithms, their accuracy increases steadily as the number of bands increases, and the ASPS-MN method can better illustrate this point.

For the Salinas dataset, Figure 13 and Table 1 reveal that the OA and Kappa of this method are significantly higher than other methods. It can be seen from Table 1 that the Kappa coefficient at this time can reach 92.58%, which shows that the method in this paper has higher classification accuracy and better effect for each ground feature. Furthermore, the following conclusion can be drawn from Figure 13, i.e., although the accuracy of the proposed method is lower than that of other algorithms in the initial stage, its superiority becomes very obvious when the number of bands exceeds 20. Particularly for the KNN classifier, the OA coefficient of this method is continuously rising and leading other algorithms. When the number of selected bands is small, SOPSRL and ASPS-MN fluctuate greatly. In addition, from the classification result map (Figure 14), it can be seen that there are fewer misclassifications in this method.

Through a large number of experiments in the three hyperspectral datasets, it can be seen that the performance of most comparison algorithms in the Indian Pines dataset is not very ideal. However, for the method proposed in this paper, it has a good performance in this dataset. In terms of

FIGURE 14: Continued.

FIGURE 14: Classification results of different algorithms on the Salinas dataset: (top) KNN and (bottom) SVM.

the classification performance of the other two datasets, the method proposed in this paper also has absolute advantages.

In order to further illustrate the advantages and disadvantages of each band selection algorithm, the band selection is carried out on three datasets (taking 15 bands as examples) and the classification accuracy of each ground object by different algorithms is given (as shown in Figures 15–17). The band selection results of different algorithms on three datasets are shown in Table 3.

The following conclusions can be drawn from Figures 15–17. For the Indian Pine dataset, due to the dataset itself, the classification accuracy of all current algorithms

is not high. Particularly for oats classification, the best performance is the ONR algorithm; other algorithms can not even make classification. Compared with other algorithms, it can be seen from the figure that the algorithm presented in this paper is superior in the classification of some ground objects (as shown in the yellow color in the figure). For the Pavia University dataset, the algorithm in this paper and other algorithms have a good classification performance, with 100% classification for shadows. But its classification effect is slightly better than that of other algorithms. At the same time, for the Salinas dataset, from the classification results, the algorithm in this paper does not show better classification

FIGURE 15: Schematic diagram of the classification effect of different algorithms on the Indian Pines dataset.

FIGURE 16: Schematic diagram of the classification effect of different algorithms on the Pavia University dataset.

performance than other algorithms, but the algorithm in this paper can have the same excellent classification performance as the latest algorithms. The classification performance of the

MDSR algorithm is far inferior to several other algorithms. The above experiment proves the effectiveness and superiority of the algorithm in this paper.

FIGURE 17: Schematic diagram of the classification effect of different algorithms on the Salinas dataset.

TABLE 3: The number of bands selected by different algorithms is on three datasets.

Dataset	Algorithm	Selected bands
Indian Pines	ONR	3, 10, 16, 44, 67, 72, 77, 99, 114, 131, 149, 162, 172, 182, 188
	OCF	15, 18, 22, 28, 33, 40, 47, 54, 62, 68, 72, 76, 81, 90, 100
	SOPSRL	14, 22, 27, 36, 38, 45, 54, 68, 82, 86, 87, 91, 92, 93, 98
	ASPS_MN	7, 27, 41, 52, 54, 78, 102, 110, 126, 143, 145, 165, 175, 179, 192
	MDSR	9, 28, 49, 69, 77, 92, 112, 116, 124, 140, 142, 150, 160, 164, 178
	ANG_LSC	9, 28, 42, 51, 54, 59, 82, 111, 119, 135, 150, 164, 169, 177, 185
Pavia University	ONR	15, 18, 22, 28, 33, 40, 47, 54, 62, 68, 72, 76, 81, 90, 100
	OCF	4, 10, 17, 22, 30, 36, 43, 49, 58, 63, 67, 78, 82, 91, 97
	SOPSRL	13, 21, 22, 23, 28, 30, 37, 38, 65, 67, 68, 78, 80, 83, 86
	ASPS_MN	9, 19, 20, 32, 34, 35, 51, 56, 60, 67, 73, 78, 88, 90, 100
	MDSR	7, 14, 15, 27, 30, 35, 41, 54, 65, 73, 80, 88, 91, 96, 97
	ANG_LSC	11, 19, 23, 33, 34, 40, 49, 58, 60, 65, 72, 83, 88, 90, 95
Salinas	ONR	6, 14, 19, 30, 44, 56, 70, 103, 113, 118, 133, 162, 173, 185, 194
	OCF	8, 15, 25, 34, 45, 58, 93, 120, 125, 135, 158, 167, 170, 180, 190
	SOPSRL	7, 8, 9, 10, 14, 15, 16, 17, 19, 33, 36, 40, 43, 90, 110
	ASPS_MN	12, 27, 28, 48, 79, 82, 106, 109, 133, 146, 148, 158, 176, 187, 203
	MDSR	2, 16, 54, 59, 74, 83, 95, 105, 117, 158, 162, 170, 179, 194, 204
	ANG_LSC	2, 27, 35, 54, 62, 81, 95, 112, 120, 141, 151, 163, 179, 190, 202

5. Conclusion

Recently, many clustering-based band selection methods have been proposed, but most of them only take into account redundancy between bands, neglecting the amount of information in the subset of selected bands. Furthermore, these algorithms never consider the hyperspectral bands as

ordered. We propose an unsupervised hyperspectral band selection method based on adaptive neighborhood grouping and local structure correlation. First, the method divides the hyperspectral cube into multiple subcubes through adaptive subspace partition and employs the product of the local density and the distance factor to sort the bands and select the desired number of cluster centers. Then, according to the

Euclidean distance matrix and the cluster center, all the bands are divided into several clusters. Finally, based on information entropy and correlation analysis, the most representative band is selected from each cluster. Different from other clustering methods ranking-based, this method employs the clustering method to divide the spatially ordered hyperspectral cube into multiple subcubes to generate an overall framework, which effectively avoids the selection of highly correlated subsets. Next, the LSC method is utilized to consider the combination of bands in different clusters to avoid falling into the local optimal solution. A large number of experimental results on three publicly available hyperspectral image datasets show that this method has better robustness and superiority than other methods.

The quality of the cluster centers has an important impact on the performance of the method. In future work, we will explore a better cluster center selection mechanism to improve the robustness and effectiveness of the selected band.

Data Availability

Data is available in http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (51607059) and Science and Technology Department, Heilongjiang Province (QC2017059).

References

- [1] C. M. Gevaert, J. Suomalainen, J. Tang, and L. Kooistra, "Generation of spectral-temporal response surfaces by combining multispectral satellite and hyperspectral UAV imagery for precision agriculture applications," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 3140–3146, 2015.
- [2] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: earth monitoring with statistical learning methods," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 45–54, 2014.
- [3] B. B. Damodaran and R. R. Nidamanuri, "Dynamic linear classifier system for hyperspectral image classification for land cover mapping," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2080–2093, 2014.
- [4] X. Zhang, Y. Sun, K. Shang, L. Zhang, and S. Wang, "Crop classification based on feature band set construction and object-oriented approach using hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4117–4128, 2016.
- [5] W. Sun, G. Yang, B. Du, L. Zhang, and L. Zhang, "A sparse and low-rank near-isometric linear embedding method for feature extraction in hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 4032–4046, 2017.
- [6] B. Barman and S. Patra, "Empirical study of neighbourhood rough sets based band selection techniques for classification of hyperspectral images," *IET Image Processing*, vol. 13, no. 8, pp. 1266–1279, 2019.
- [7] D. Saqui, J. H. Saito, D. C. de Lima et al., "NSGA2-based method for band selection for supervised segmentation in hyperspectral imaging," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 3580–3585, Bari, Italy, 2019.
- [8] X. Wei, W. Zhu, B. Liao, and L. Cai, "Scalable one-pass self-representation learning for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4360–4374, 2019.
- [9] X. Cao, C. Wei, Y. Ge, J. Feng, J. Zhao, and L. Jiao, "Semi-supervised hyperspectral band selection based on dynamic classifier selection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 4, pp. 1289–1298, 2019.
- [10] H. Li, C. Chang, L. Wang, and Y. Li, "Constrained multiple band selection for hyperspectral imagery," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 6149–6152, Beijing, 2016.
- [11] M. Zhang, J. Ma, and M. Gong, "Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 773–777, 2017.
- [12] L. C. B. dos Santos, S. J. F. Guimarães, and J. A. dos Santos, "Efficient unsupervised band selection through spectral rhythms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1016–1025, 2015.
- [13] F. Li, P. Zhang, and L. Huchuan, "Unsupervised band selection of hyperspectral images via multi-dictionary sparse representation," *IEEE Access*, vol. 6, pp. 71632–71643, 2018.
- [14] Q. Li, Q. Wang, and X. Li, "An efficient clustering method for hyperspectral optimal band selection via shared nearest neighbor," *Remote Sensing*, vol. 11, no. 3, p. 350, 2019.
- [15] R. Yang and J. Kan, "An unsupervised hyperspectral band selection method based on shared nearest neighbor and correlation analysis," *IEEE Access*, vol. 7, pp. 185532–185542, 2019.
- [16] M. Ahmad, I. U. Haq, Q. Mushtaq, and M. Sohaib, "A new statistical approach for band clustering and band selection using K-means clustering," *International Journal of Engineering and Technology*, vol. 3, no. 6, pp. 606–614, 2011.
- [17] R. Yang, L. Su, X. Zhao, H. Wan, and J. Sun, "Representative band selection for hyperspectral image classification," *Journal of Visual Communication and Image Representation*, vol. 48, pp. 396–403, 2017.
- [18] Y. Yuan, J. Lin, and Q. Wang, "Dual-clustering-based hyperspectral band selection by contextual analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1431–1445, 2016.
- [19] R. Mehmood, R. Bie, H. Dawood, and H. Ahmad, "Fuzzy clustering by fast search and find of density peaks," *International conference on identification, information, and knowledge in the internet of things (IIKI)*, 2015, pp. 258–261, Beijing, 2015.
- [20] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 88–102, 2016.

- [21] Y. Yuan, X. Zheng, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 51–64, 2017.
- [22] L. Jiao, J. Feng, F. Liu, T. Sun, and X. Zhang, "Semisupervised affinity propagation based on normalized trivariable mutual information for hyperspectral band selection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2760–2773, 2015.
- [23] S. Jia and L. Deng, "A CW-SSIM distance measure-based affinity propagation for hyperspectral band selection," in *2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS*, pp. 418–421, Melbourne, VIC, 2013.
- [24] C. Yang, Y. Tan, L. Bruzzone, L. Lu, and R. Guan, "Discriminative feature metric learning in the affinity propagation model for band selection in hyperspectral images," *Remote Sensing*, vol. 9, no. 8, p. 782, 2017.
- [25] S. Bhattacharya, S. Das, and A. Routray, "Graph manifold clustering based band selection for hyperspectral face recognition," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 1990–1994, Rome, 2018.
- [26] Q. Wang, F. Zhang, and X. Li, "Hyperspectral band selection via optimal neighborhood reconstruction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8465–8476, 2020.
- [27] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 10, pp. 5910–5922, 2018.
- [28] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 4940–4950, 2019.