

## Article

# A Spatial–Spectral Combination Method for Hyperspectral Band Selection

Xizhen Han <sup>1,2</sup>, Zhengang Jiang <sup>1,\*</sup>, Yuanyuan Liu <sup>3</sup>, Jian Zhao <sup>4</sup>, Qiang Sun <sup>3</sup> and Yingzhi Li <sup>4</sup>

<sup>1</sup> College of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130000, China; 2018200089@mails.cust.edu.cn

<sup>2</sup> Suzhou East Cloth Opto-Electronic Technology Co., Ltd., Zhangjiagang, Suzhou 215600, China

<sup>3</sup> Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130000, China; liuyuanyuan192@mails.ucas.ac.cn (Y.L.); sunq@ciomp.ac.cn (Q.S.)

<sup>4</sup> Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215000, China; zhaoj@sibet.ac.cn (J.Z.); liyz@sibet.ac.cn (Y.L.)

\* Correspondence: jiangzhengang@cust.edu.cn

**Abstract:** Hyperspectral images are characterized by hundreds of spectral bands and rich information. However, there exists a large amount of information redundancy among adjacent bands. In this study, a spatial–spectral combination method for hyperspectral band selection (SSCBS) is proposed to reduce information redundancy. First, the hyperspectral image is automatically divided into subspaces. Seven algorithms classified as four types are executed and compared. The means algorithm is the most suitable for subspace division of the input hyperspectral image, with the calculation being the fastest. Then, for each subspace, the spatial–spectral combination method is adopted to select the best band. The band with the maximum information and more prominent characteristics between the adjacent bands is selected. The parameters of Euclidean distance and spectral angle parameters are used to measure the intraclass correlation and interclass spectral specificity, respectively. Weight coefficient quantifying the intrinsic spatial–spectral relationship of pixels are constructed, and then the optimal bands are selected by a combination of the weight coefficients and the information entropy. Moreover, an automatic method is proposed in this paper to provide an appropriate number of band sets, which is out of consideration for existing research. The experimental results show, as compared to other competing methods, that the SSCBS approach has the highest classification accuracy on the three benchmark datasets and takes less computation time. These demonstrate that the proposed SSCBS achieves satisfactory performance against state-of-the-art algorithms.

**Keywords:** band selection; hyperspectral image; inter-spectral specificity; spatial–spectral combination; spatial correlation



**Citation:** Han, X.; Jiang, Z.; Liu, Y.; Zhao, J.; Sun, Q.; Li, Y. A Spatial–Spectral Combination Method for Hyperspectral Band Selection. *Remote Sens.* **2022**, *14*, 3217. <https://doi.org/10.3390/rs14133217>

Academic Editors: Saeid Homayouni, Edoardo Pasolli, Pedram Ghamisi, Amin Zehtabian, Ali Mousivand and Fardin Mirzapour

Received: 1 June 2022

Accepted: 1 July 2022

Published: 4 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Hyperspectral images are high-dimensional data containing rich spatial, spectral and radiation information, and have been widely used in fine agriculture, geological exploration, environmental monitoring, urban remote sensing, military reconnaissance and other fields [1]. These applications often require classifying each pixel in the scene. Because there are a huge number of features (or spectral bands) with only limited training samples available, HSI classification becomes a challenging task. Many spectral bands provide rich information for classifying various materials in the scene. However, with limited training samples, the performance of classifiers deteriorates as the dimensionality increases [2] (Hughes phenomenon [3]). High-dimensionality data processing also requires huge computational resources and storage capacity [4]. Meanwhile, the spectral bands are often correlated, and not all of them are useful for the specific classification task. Therefore, to achieve an excellent classification performance, a dimension reduction (DR) procedure is necessary.

Hyperspectral image dimensionality reduction methods are primarily divided into two categories: the feature extraction and the band selection [5]. Feature extraction transforms the feature attributes of all bands from high-dimensional space to low-dimensional space to generate a new feature space. In this way, the highest possible classification accuracy is retained. However, the original pixel values will be changed after the transformation, which changes the physical characteristics, and is not conducive to the automatic recognition and ground object inversion of hyperspectral remote sensing images [6,7]. Classic feature extraction methods include principal component analysis (PCA) [8], orthogonal subspace projection (OSP) [9], linear discriminant analysis (LDA) [10,11], locality preserving projection (LPP) [12], neighborhood preserving embedding (NPE) [13], local discriminant embedding (LDE) [14], non-parametric weighted feature extraction (NWFE) [15], local Fisher discriminant analysis (LFDA) [16] and marginal Fisher analysis [17]. Band selection involves directly selecting a representative band subset from all the bands of hyperspectral image data according to certain criteria or search strategy to express the features of the whole hyperspectral data to achieve dimension reduction [18–21]. Band selection methods can be categorized as supervised, semi-supervised and unsupervised methods. Supervised learning methods use labeled data to determine the importance of bands. For instance, Yang et al. [22] proposed a sequential forward selection (SFS) searching strategy for band selection. In [23], Archibald and Fann introduced an embedded feature selection algorithm that was tailored to operate with support vector machines (SVMs). Semi-supervised band selection methods combine the advantages of both supervised and unsupervised methods. Bai et al. [24] proposed a semi-supervised learning method that estimates the vectors of the mean values and covariance matrices for each class under the assumption of a Gaussian mixture model. Chen et al. [25] combined Fisher's criteria and a graph Laplacian to explore labeled and unlabeled samples simultaneously. Unsupervised methods do not need to learn a predictive model from training data. Classic unsupervised methods often use a priori knowledge of the scene to measure the statistical dependence between bands and evaluate the contribution of each band in the classification. Most methods are clustering-based and sorting-based. Clustering-based methods first separate all the bands into clusters firstly, and then select the most representative bands in each cluster to constitute the band subset. Sorting-based methods assign each band a rank value and simply select the top-rank bands with the desired number. Reza et al. [26] proposed a method for selecting distinct and informative bands in a prototype space, which was constructed by clustering raw image data. Bands are selected based on either their orthogonal distance to the diagonal of the prototype space or their angular distance related to the correlation between neighboring bands. Adaptive Band Selection algorithm (ABS) [27] obtains the ABS index by calculating the standard deviation of each band and the correlation coefficients of its adjacent bands. Then the ABS index is sorted from largest to smallest, and the bands that are larger than a certain threshold or the first  $m$  predetermined bands are selected. A. Rodriguez proposed a fast density-peak-based clustering (FDPC) algorithm. FDPC is a sorting-based clustering algorithm, which identifies cluster centers through investigating the local density and the intra-cluster distance of each band. Later, Sen Jia proposed an enhanced FDPC (E-FDPC) [28] that was more suitable for hyperspectral band selection. The fast neighborhood grouping method (FNGBS) [29] uses a coarse-fine strategy to segment a hyperspectral image cube in space, and then the best bands as a subset is obtained according to local density and information entropy. Both TRC\_OC\_FDPC and NC\_OC\_MVPCA algorithms [30] are implemented based on a clustering framework. The TRC\_OC\_FDPC algorithm uses TRC (top-rank cut) as the objective function and E-FDPC as the sorting method to realize the algorithm, and the NC\_OC\_MVPCA algorithm uses NC (normalized cut) as the objective function and MVPCA as the sorting method. TRC chooses the bands with the highest rank values in each cluster. NC is an effective graph-theoretic criterion. MVPCA evaluates the bands according to their variances.

Recently, spatial-spectral methods have played an important role in the dimensionality reduction. By fusing spatial and spectral information, the representation of the hyperspec-



tral image is improved and the classification performance is enhanced. Zhou et al. [31] proposed a spatial and spectral regularized local discriminant embedding (SSRLED) method. Huang et al. [32] proposed a spatial–spectral manifold reconstruction preserving the embedding (SSMRPE) method. Zhou et al. [33] proposed a spatial–spectral feature dimensionality reduction algorithm based on manifold learning. Zhao et al. [34] proposed a spectral–spatial feature-based classification (SSFC) framework. The aforementioned algorithms are the feature extraction methods. Feng et al. [35] defined discriminate spectral-spatial margins (DSSMs) to reveal the local information of hyperspectral pixels and explored the global structures of both labeled and unlabeled data via low-rank representation (LRR). Bai et al. [36] proposed a semi-supervised band selection method that allowed contributions from both labeled and unlabeled hyperspectral pixels, and then a linear regression model with group sparsity constraint was used for band selection.

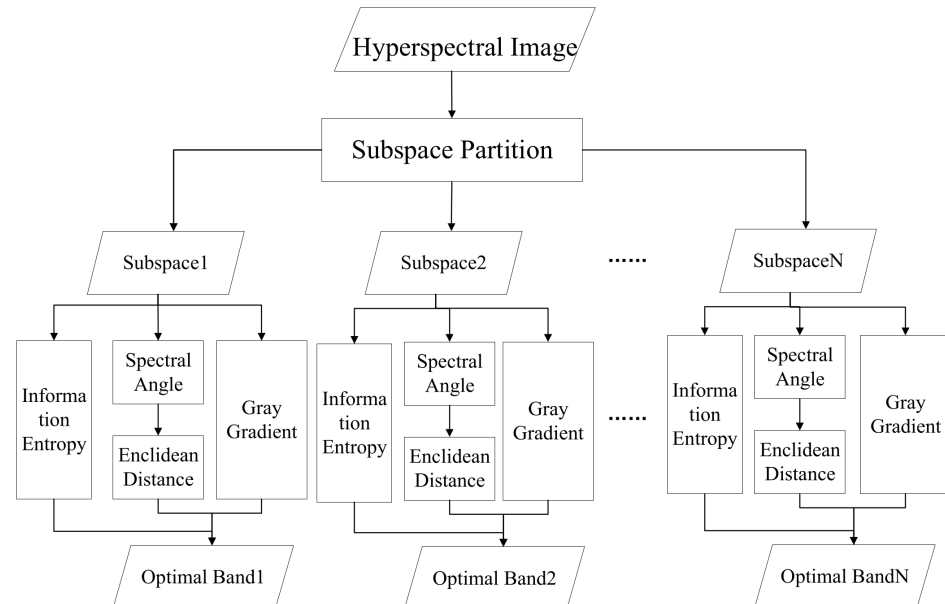
In recent years, many band selection methods have been proposed; most of them extract the band subset based on global information. Although the optimal solution can be found, the calculation is complex and usually trapped into local optimums. The main reason is that the solution space of the band selection problem is too large to attain the optimal solution in a limited time. In addition, some methods are very sensitive to hyperspectral bands containing noises, which makes the selected bands have a strong similarity. Usually, 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 [29,30]. To tackle these issues, the SSCBS method is proposed. It combines features based on spatial and spectral separability to provide more discernible characteristic spectral bands for subsequent image classification, and automatic detection and recognition of targets. The main contributions are as follows.

- (1) Subspace division is proposed to partition the hyperspectral image cube into multiple groups in space. The bands of high similarity in spectral dimension are assigned into one group. Seven algorithms are proposed. It is shown through a comprehensive comparison that the means algorithm is the most suitable, with the computation time being the shortest.
- (2) In each subspace, a spatial–spectral combination method is proposed. The band with maximum information and more prominent characteristics between different categories is selected. Compared with most existing methods about selecting representative band, this method can better exploit the spatial and spectral characteristics. The obtained band subset has more discriminative bands simultaneously.
- (3) An automatic method to determine the number of appropriate bands is proposed, which can better evaluate the information redundancy of band set. Experiments show that this algorithm can offer a promising estimation of band number for various data sets.
- (4) This method is an unsupervised band selection that does not require any label data.

## 2. Proposed Approach

The spectral resolution of hyperspectral images is extremely high, and the data have a strong correlation in the spectral domain. Thus, the correlation among adjacent bands is strong, and the bands have the characteristics of aggregation. To address the above-mentioned issues, the SSCBS method is proposed. SSCBS can be divided into two main components. In the first step, the hyperspectral image is automatically divided into subspaces. Automatic subspace division not only accelerates the selection of optimal bands, but also improves the rationality of band selection to avoid the selected bands too close. Then, for each subspace, a spatial–spectral combination method is proposed to select the best band. For each subspace, the spatial and spectral features can be integrated into a uniformed objective function by constructing weight coefficient. The Euclidean distance and spectral angle are used to quantify the internal spatial–spectral relationship of each point in the hyperspectral image. Combined with the information entropy algorithm, the

representative bands with a large amount of information, low spatial correlation and strong inter-class spectral specificity are extracted. Finally, the representative band of each subspace is combined to construct the optimized band set to achieve dimensionality reduction. An overall flowchart of the algorithm is presented in Figure 1.



**Figure 1.** Flowchart of algorithm.

### 2.1. Subspace Partition

The subspace partition divides the hyperspectral image dataset into multiple subspaces according to the correlation between different bands in the spectral dimension, so that the band correlation in each subspace is strong, and the correlation between different subspaces is weak. To select the best subspace partition method, this paper proposes several subspace partition algorithms, compares them, and finally chooses the best method.

#### 2.1.1. Means Method

The means method is the simplest division. We define  $X = \{x_1, x_2, \dots, x_L\} \in R^{N \times L}$  as a hyperspectral image, where  $N$  is the number of pixels in each band,  $L$  is the total number of bands, and  $x_i$  represents the data vector of the  $i$ th band. Suppose that  $M$  subspaces are divided, where  $M > 1$ , and the number of selected bands should also be set as  $M$ . All bands must be equally divided into  $M$  groups of  $\{X'_m\}_{m=1}^{m=M}$  with  $M + 1$  nodes, and the partition node  $G$  is defined as

$$G(g) = \begin{cases} 1, & g = 1 \\ \left\lfloor \frac{L}{M} \right\rfloor \times (g - 1), & 2 \leq g \leq M \\ L, & g = M + 1 \end{cases} \quad (1)$$

where,  $\lfloor \cdot \rfloor$  means casting the value to an integer. After partition node  $G$  is determined, the division of  $M$  group can be expressed as follows:

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

#### 2.1.2. Correlation-Based Method

A more precise spectral division is the characteristic of a hyperspectral image cube, which results in a strong correlation in the spectral domain. Spectral correlation means that images of adjacent bands of ground objects at the same position are similar in space. The main reason for the spectral correlation is that the light reflectivities of the same ground

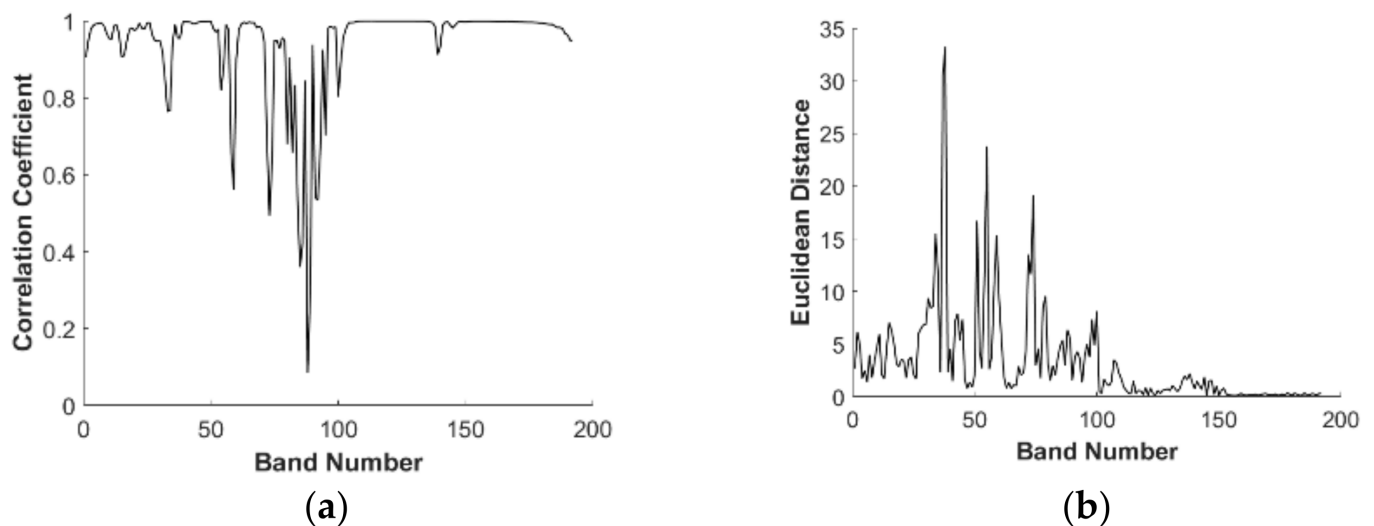
object in adjacent bands are very similar [37]. Such a correlation is often described by the correlation coefficient matrix and the Euclidean distance. The subspace division method based on the correlation coefficient is referred to as the correlation, and the subspace division method based on Euclidean distance is referred to as the Euclidean distance. The correlation coefficient  $r_{xy}$  between bands  $x$  and  $y$  is defined [33,38] as:

$$r_{xy} = \frac{Cov(x, y)}{\sqrt{D(x)}\sqrt{D(y)}} \quad (3)$$

where,  $Cov$  is the covariance and  $D$  is the variance. The Euclidean distance between bands  $x$  and  $y$  is defined as:

$$d_{xy} = ||x - y||^2 = \sum_{i=1}^N (x_i - y_i)^2 \quad (4)$$

where  $N$  denotes the total number of pixels in the image. Using the Indian Pines dataset as an example, a visualization of the calculated correlation coefficient and Euclidean distance is shown in Figure 2.



**Figure 2.** Visualization of the correlation between adjacent bands on the Indian Pines data set (a) Correlation coefficient (b) Euclidean distance.

The steps for automatic subspace partitioning based on correlation are as follows:

- (1) Convert a two-dimensional band image into one-dimensional band vector;
- (2) Calculate the correlation coefficient matrix  $R$  of adjacent bands, which is defined as  $R = [r_{1,2}, r_{2,3}, r_{3,4} \dots r_{i,i+1} \dots r_{L-1,L}]$ ; or calculate the Euclidean distance matrix  $D$  of adjacent bands which is defined as  $D = [d_{1,2}, d_{2,3}, d_{3,4} \dots d_{i,i+1} \dots d_{L-1,L}]$ .
- (3) The local minimum of correlation coefficient matrix  $R$  or the local maximum of the Euclidean distance matrix  $D$  is obtained by smoothing the correlation vectors. Suppose the local minimum or maximum is  $S$ . If  $S > M$ , take the previous  $M - 1$  values as nodes, and divide the hyperspectral data cube into  $M$  data subspaces. Otherwise, it can only be divided into  $S + 1$  data subspaces.

### 2.1.3. Coarse-Fine Partition Method

In hyperspectral images, the correlation of neighborhood bands is greater than that of non-neighborhood bands. Based on this idea, a coarse-fine strategy for subspace division is used. Compared with the correlation-based method [39,40], the influence of noise can be effectively suppressed by using the coarse division method. Only the correlation between the central band and neighboring bands needs to be calculated, which reduces the running time of the algorithm and fully mines the association information of adjacent bands.

The method in Sections 2.1.1 and 2.1.2 can be adopted for the coarse division. Because the correlation-based method only considers the correlation between adjacent bands without considering the correlation with the neighborhood, it is easily affected by noise bands and the division result is not desired. Therefore, fine division strategy is adopted to subdivide the bands further in order to obtain a more accurate subspace. For each subspace,  $x_{C_m}$  is defined as the center of the subspace. For coarse partitioning, the subspace center is defined as:

$$C_m = m \times \left\lfloor \frac{L}{M} \right\rfloor - \left\lfloor \frac{L}{2M} \right\rfloor, m = 1, 2, \dots, M \quad (5)$$

where  $m$  denotes the  $m$ th subspace. To accelerate the execution efficiency of the algorithm and fully mine the correlation between adjacent bands, for the center  $x_{C_m}$  of each subspace, we only consider its neighborhood band  $\{x_j\}_{j=a}^{j=b}$  and define the values of  $a$  and  $b$  using the following formula:

$$(a, b) = \begin{cases} [1, C_{m+1}), & m = 1 \\ (C_{m-1}, C_{m+1}), & 2 \leq m \leq M-1 \\ (C_{m-1}, L], & m = M \end{cases} \quad (6)$$

By updating the partition node  $G$  of the subspace, the bands with high correlation are divided into the same subspace by using the correlation among the bands in the neighborhood, so that the band redundancy of different subspaces is lower.

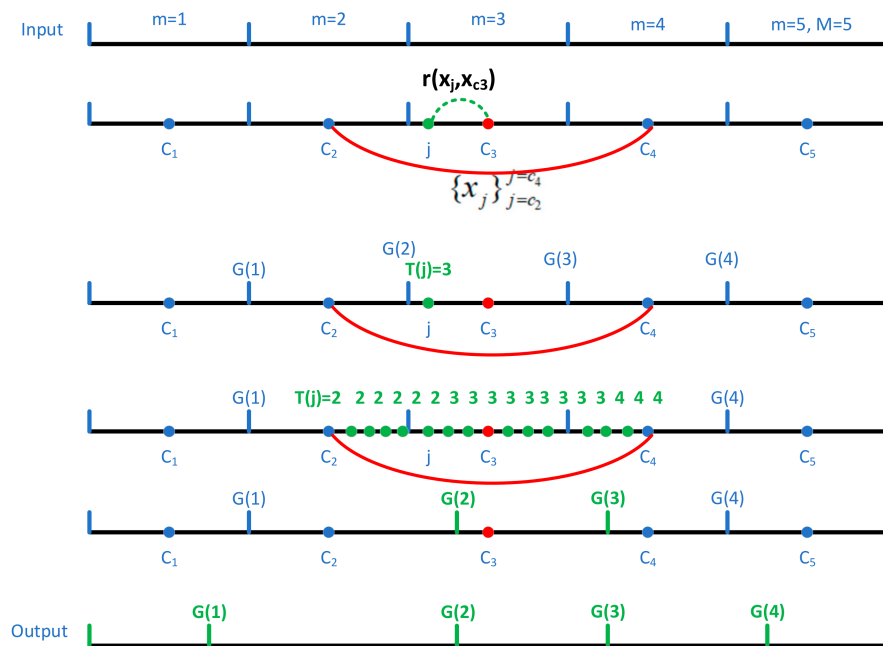
The fine partitioning strategy algorithm is described as follows:

**Input:** Coarse band grouping for hyperspectral data cube  $\{X'_m\}_{m=1}^{m=M}$ ;

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

- (1) Define  $R$  to record the correlation between the current band  $x_j$  and the central band  $x_{C_m}$ , and initialize  $R$  to zero. The subspace label is defined as  $T$  which is initialized to zero.
- (2) Calculate the correlation coefficient  $r(x_j, x_{C_m})$  between the current band  $x_j$  and the central band  $x_{C_m}$ . The correlation between the two bands can be regarded as the similarity between them. The greater the correlation, the higher the similarity.
- (3) If  $r(x_j, x_{C_m}) > R(j)$ , set  $R(j) = r(x_j, x_{C_m})$  and  $T(j) = m$ .
- (4) Traverse the neighborhood bands  $\{x_j\}_{j=a}^{j=b}$  of each central band  $x_{C_m}$ .
- (5) Remove the influence of the noise. The label value obtained by a neighborhood subspace may be  $[1, 1, 1, 1, 2, 2, 2]$ , and the intermediate label value of 2 may be influenced by noise. A singular value should be removed to avoid noise interference.
- (6) In subspace  $m$ , for the current band  $x_j$ , if  $r(x_j, x_{C_{m-1}}) > r(x_j, x_{C_m})$ ,  $T(j) = m - 1$ ; if  $r(x_j, x_{C_{m+1}}) > r(x_j, x_{C_m})$ ,  $T(j) = m + 1$ ; else  $T(j) = m$ ; for  $j \in (C_{m-1}, C_m)$ , if  $T(j - 1) = m$ , and  $T(j) = m + 1$ , the new band partition node of subspace  $m$  is  $G(m)$ ,  $G(m) = j$ .
- (7) Obtain the fine subspace of the hyperspectral data cube  $\{X_m\}_{m=1}^{m=M}$ .

Figure 3 illustrates an example of fine subspace partitioning of the hyperspectral data cube. To facilitate the expression of each algorithm in the experimental stage, in the coarse-fine partition method, the coarse partition using means is called as the means and correlation algorithm. The coarse partition using correlation is called as the correlation and correlation algorithm. The coarse partition using Euclidean is called as the Euclidean and correlation algorithm.



**Figure 3.** Example of hyperspectral image cube finely divided into 5 subspaces.

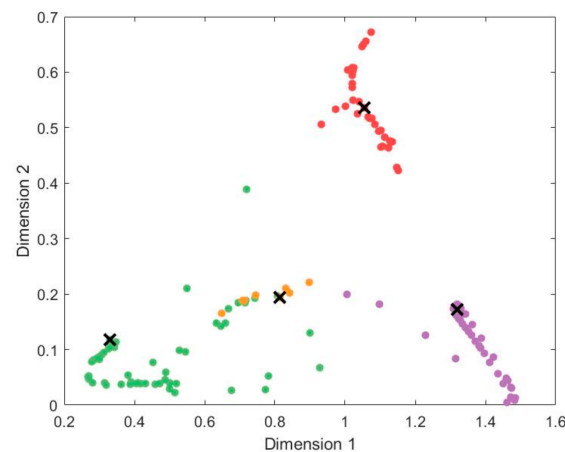
#### 2.1.4. K-Means Method

The K-means algorithm is a typical distance-based clustering algorithm, which uses distance as an evaluation index of similarity. In other words, the shorter the distance between two objects, the greater the similarity. The algorithm considers that clusters are composed of objects close to each other. Therefore, the ultimate goal is to obtain compact and independent clusters. The k-means algorithm adopted in this study is not exactly the same as traditional k-means algorithm. According to the characteristics of the hyperspectral image dataset, each band is regarded as a point, the similarity between points is evaluated by the Euclidean distance between bands, and a compact and independent cluster is sought as a subspace. The steps of subspace division based on k-means clustering are as follows:

- (1) From  $L$  bands, according to the principle of equal-spacing partition, each equal-spacing subset is taken as the initial cluster to find  $M$  cluster centers  $x_{C_m}$  ( $M = 1, 2 \dots M$ ) as the initial clustering center, and the cluster label  $T$  is defined.
- (2) For two adjacent cluster centers of  $x_{C_m}$  and  $x_{C_{m+1}}$ , and for each band of  $x_i$ , calculate the Euclidean distance  $d_{i,m}$  and  $d_{i,m+1}$  between  $x_{C_m}$ , and  $x_i$ , and between  $x_{C_{m+1}}$  and  $x_i$ , and take the label of the cluster center with a smaller Euclidean distance as the updated cluster label of this band.
- (3) Recalculate the center of each cluster  $x_{C_m}$ .
- (4) Repeat Steps (2) and (3) until all cluster centers do not change or the maximum number of iterations is reached.
- (5) The final  $M$  clusters are the desired  $M$  subspaces.
- (6) If the cluster label is not continuous (e.g., the cluster label is... 1,1,1,1,2,2,1,1,1,1,2,2,2,2,...), in other words, if there are noise bands, the discontinuous label must be updated to a label that is closer to the cluster. In the preceding example, 2,2 and 1,1,1,1 denote discontinuous labels. The distance between 2,2 and its cluster is 4, and the distance between 1,1,1,1 and its cluster is 2.

Figure 4 shows an example of dividing 4 subspaces by K-means method on the Indian Pines dataset. The different colors represent the different subspaces, and the black fork marks the clustering center.





**Figure 4.** Example of dividing 4 subspaces by K-means.

The comparison and analysis of the four subspace partition methods in this section are carried out in Section 3.3, based on classification performance in two dimension and computing time.

## 2.2. Spatial–Spectral Joint Information

After the subspace partition is completed, the method for finding the best band in each subspace is the focus of this study. Other band selection methods treat each image as a point and only consider the correlation between bands, but ignore the correlation between pixels in each band and the spectral specificity between the target and background. To make full use of the spatial and spectral information of the hyperspectral data cube, a spatial–spectral combination algorithm is proposed in this paper for band selection. By taking full consideration of the inner spatial correlation and spectral specificity, not only the dimension reduction is achieved, but also the accuracy of the image classification and the target detection probability can be improved. The main algorithms used in this study are as follows.

### 2.2.1. K-Nearest Neighbor of Pixel Points

The core idea of the k-nearest neighbor is the distance measurement, by which  $k$  points closest to the target point are obtained. According to the classification rules, the closer the pixel is to the target point, the higher the probability of the target being classified into the same category. The Euclidean distance is used for the distance measurement.

The choice of  $k$  is also important. If  $k$  is smaller, the approximate error will be smaller, but the random error will be larger. On the contrary, if  $k$  is larger, the approximation error will be larger, and the random error will be smaller.

### 2.2.2. Spectral Angle Mapping

The spectral angle mapping (SAM) algorithm was proposed by Kruse, et al., in 1993, which treated the spectrum of each pixel in the image as a high-dimensional vector. By calculating the angle between two vectors, the similarity between two spectra is measured. The smaller the angle, the more similar the two spectra and the greater the chance of belonging to the same object. Therefore, unknown data can be identified according to the size of the spectral angle. In the classification, the spectral angle forming by unknown data and known data is calculated, and the category of unknown data is classified into the category corresponding to the minimum spectral angle. The angle between the two vectors is calculated based on the cosine relation, as follows:

$$\theta(P, Q) = \cos^{-1}\left(\frac{P^T Q}{\|P\| \|Q\|}\right) = \cos^{-1}\left(\frac{P^T Q}{\sqrt{Q^T Q} \sqrt{P^T P}}\right) \quad (7)$$

where  $P$  is the spectral vector of the target point  $p$ , and  $Q$  is the spectral vector of its neighborhood point  $q$ . The spectral angle is represented by a small radian angle, which represents the majority of the spectral similarity between spectral curves and its variation range is  $[0, \frac{\pi}{2}]$ .

### 2.2.3. Information Entropy

According to Shannon's information theory, information entropy is used to measure the amount of uncertain information, which is a statistical form of a feature and reflects the average amount of information in an image. To ensure that the band selection contains a large amount of information, information entropy is considered in the algorithm. The information entropy is defined as follows:

$$H(x) = - \sum_{z \in \Omega} p(z) \log p(z) \quad (8)$$

where,  $x$  refers to the  $x$  band image,  $\Omega$  is the gray-level space of the band  $x$ , and  $p(z)$  is the probability of occurrence of each gray-level  $z$ .

### 2.2.4. Spatial-Spectral Combination Algorithm

In hyperspectral images, the closer the pixel is to the target point, the higher the probability of belonging to the same category. However, as hyperspectral images are greatly affected by the environment in the acquisition process, and there is spectral aliasing. These bring about problems of "homogeneous object with different spectrum" and "inhomogeneous objects with the same spectrum". Therefore, spatial information and spectral information of pixels should be considered in the classification of hyperspectral image. The Euclidean distance quantifies the spatial information between the target and surrounding pixels, while the spectral angle reflects the similarity between the target and surrounding pixels. In the  $k$ -nearest neighbor of the target point  $p$ , the closer the pixel is to the target point and the smaller the spectral angle forming with the target point, the higher the probability of the neighborhood point being the same as the target point, and the smaller the spatial-spectral weight coefficient. However, at the junction of the target and background, although the spatial distance is relatively short, the spectral angle is larger because of the different spectral characteristics of target and background, and the spatial and temporal spectral weight coefficients are larger. The hyperspectral data cube aims to obtain the bands with low spatial correlation, strong spectral specificity between classes and large image information as the best bands. To better reflect the joint function of spatial-spectral information, the following spatial-spectral combination function is proposed:

$$\Phi = \sum_{p=1}^N \sum_{q=1}^k \frac{\theta(P, Q)}{d_{pq}} (V_p - V_q)^2 \quad (9)$$

where,  $\theta(P, Q)$  is the spectral angle between the spectral vector  $P$  of the target point  $p$  and spectral vector  $Q$  of the neighboring point  $q$ . The more similar the spectra of the two points, the smaller the spectral angle.  $d_{pq}$  is the Euclidean distance between target point  $p$  and field point  $q$ , and the shorter the distance between two points, the smaller the Euclidean distance.  $\frac{\theta(P, Q)}{d_{pq}}$  represents the spatial-spectral weight coefficient between neighborhood point  $q$  and target point  $p$ .  $V_p$  and  $V_q$  represent the pixel value of point  $p$  and point  $q$ , respectively.  $k$  is the  $k$ -nearest neighbor of the target point, and  $N$  is the number of sample points of the single band image. In a certain spectral band, the greater the difference between the similar target and the surrounding background, that is, the more prominent the target, the larger the spatial-spectral combination function of the band.

The information entropy represents the amount of information in an image. The chosen band is expected to be not only a band with low spatial correlation and strong spectrum specificity, but also a band with a large amount of information. Therefore, the algorithm in

this study considers both the joint features of the spatial and spectral information. As the value of the spatial–spectral combination function is not in the same order of magnitude as the information entropy, it is normalized to  $[0, 1]$ .

$$\Phi = (\Phi - \Phi_{\min}) / (\Phi_{\max} - \Phi_{\min}) \quad (10)$$

$$H = (H - H_{\min}) / (H_{\max} - H_{\min}) \quad (11)$$

In each subspace, the band with a large amount of information and prominent spatial–spectral combination characteristics is found to be the optimal band, and the expression is as follows:

$$x'_m = \max_{x \in X'_m} \{\Phi \cdot H\}, m = 1, 2, \dots, M \quad (12)$$

where  $x$  is the image of a certain band,  $X'_m$  is the  $m$ th subspace,  $M$  is the number of subspaces, and  $x'_m$  is the optimal band image selected from the  $m$ th subspace.

### 2.2.5. Image Classification

In this study, the classical support vector machine (SVM) classifier [32] and KNN classifier are used to test the dimensionality reduction effect of the spatial–spectral combined band selection algorithm. SVM has many unique advantages in solving high-dimensional, nonlinear and small-sample pattern recognition problems, and can be successful in the classification of hyperspectral images. The KNN classifier is one of the simplest and most widely used non-parametric classifiers. When there is little or no prior knowledge of the data distribution, it is often used as the preferred classifier and is commonly used to evaluate the performance of band selection methods for classification accuracy [28,41].

### 2.2.6. Recommended Number of Bands

When selecting bands for hyperspectral images, it is usually unknown how many bands are appropriate to select [6,7,42]. To solve this problem, this paper proposes a method to recommend the number of bands based on information redundancy. For the hyperspectral image data, the information redundancy of the band is related to the correlation coefficient, and the greater the correlation coefficient of the band set, the higher the redundancy. The redundancy formula is defined as follows:

$$Rr = \frac{\sum_{i=1}^M \sum_{j=1}^M r_{i,j}}{\sum_{i=1}^L \sum_{j=1}^L r_{i,j}} \quad (13)$$

where  $M$  is the number of selected bands,  $L$  is the total number of bands, and  $r_{i,j}$  is the correlation coefficient between band  $i$  and band  $j$ . When the number of the selected bands is smaller and scattered, the correlation coefficient between the bands is smaller and the redundancy degree is smaller. When the number of the selected bands increases gradually, the band redundancy degree increases. The critical point, namely the number of recommended band sets, is determined by the maximum change of the slope of the redundancy curve. The discriminant conditions of the critical point are as follows:

$$S = \arg \max_{i \in (1, M)} \left( \frac{Rr^i - Rr^{i-1}}{Rr^{i+1} - Rr^i} \right) \quad (14)$$

## 3. Experiment

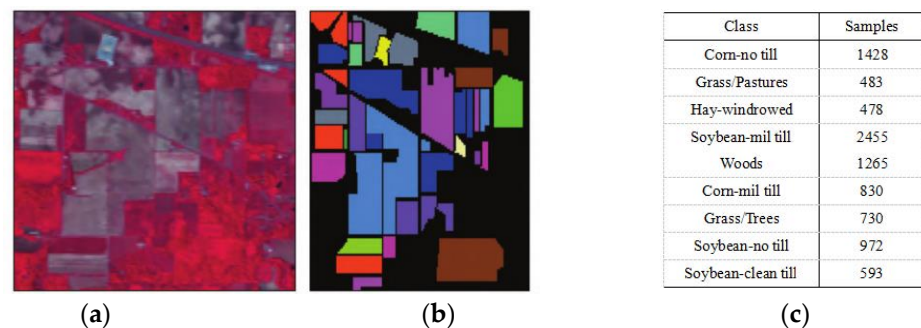
This section consists of three parts. The first part briefly introduces the basic features of the three classical hyperspectral image datasets. In the second section, the experimental setup, classifier setup and comparison methods are introduced. The third part presents a

comparative analysis of the four types of subspace partition methods. In the fourth part, three types of public datasets are used to verify the effectiveness of the proposed algorithm.

### 3.1. Data Sets

#### 3.1.1. Indian Pines

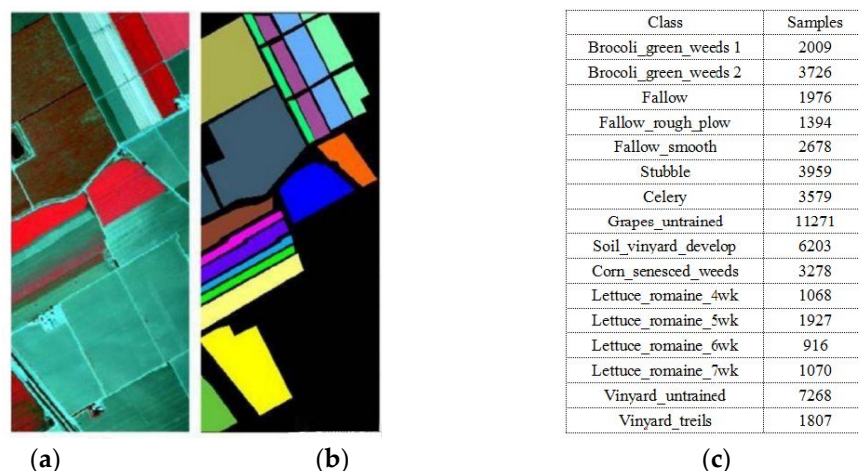
The Indian Pines hyperspectral image dataset was captured in 1992 with an airborne visible/infrared imaging spectrometer (AVIRIS) sensor in an agricultural region of north-western Indonesia ( $40^{\circ} 4' 3.4638''\text{N}$ ,  $116^{\circ} 17' 36.8982''\text{E}$ ). It contains 220 wavebands. After removing the water absorption and noisy bands, 200 bands are used in the experiments. The image size is  $145 \times 145$  pixels and there are 16 different types of ground objects. Due to the small number of seven kinds of samples, they are not representative. Figure 5a shows false color composite images of B50(R), B27(G), and B17(B) in the Indian Pines. Figure 5b shows the real ground object types in this data field. Only nine of these types are used in this experiment, and the number of sample sets for each type is shown in Figure 5c.



**Figure 5.** Indian Pines data set (a) false color image (b) corresponding ground object type (c) number of sample sets.

#### 3.1.2. Salinas

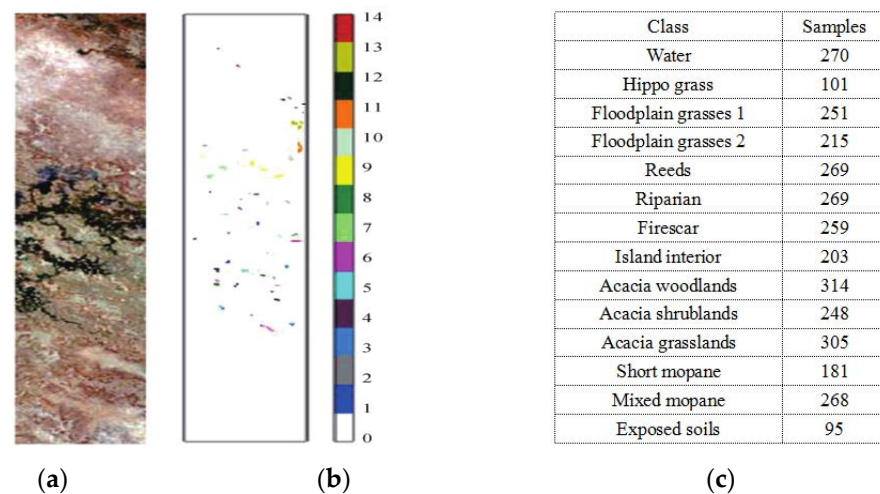
Similar to the Indian Pines dataset, the Salinas dataset was recorded by the AVIRIS sensor over the Salinas Valley in California ( $36^{\circ}42'48.3876''\text{N}$ ,  $121^{\circ}34'54.3972''\text{W}$ ), USA, with a spatial resolution of 3.7 m. There are 224 bands in this dataset, and 204 bands remain after removing bands 108–112, 154–167 and 224. The bands removed cannot be reflected by water. The image size of each band is  $512 \times 217$  pixels, with 16 types of ground features. Figure 6a is a pseudo-color image, Figure 6b is a real ground object type, and Figure 6c is the number of sample sets for each category.



**Figure 6.** Salinas data set (a) false color image (b) corresponding ground object type (c) number of sample sets.

### 3.1.3. Botswana

The Botswana dataset was obtained from NASA's EO-1 satellite in the Okavango Delta of Botswana ( $19^{\circ} 39' 3.6318''\text{S}$ ,  $22^{\circ} 54' 21.1674''\text{E}$ ) in the year from 2001 and 2004 with a spatial resolution of 30 m. After removing the water absorption and noise bands, there were 145 bands with an image size of  $1476 \times 256$  pixels and 14 categories for each band. Figure 7a is a pseudo-color image, Figure 7b is a real ground object type and Figure 7c is the number of sample sets for each category.



**Figure 7.** Botswana data set (a) false color images (b) corresponding ground object types (c) number of sample sets.

### 3.2. Experimental Setup

All the algorithms in this experiment run on a PC at the main frequency of 1.8 GHz CPU and 8 GB RAM, and are programmed with matlabR2015b (v8.6.0.267246). The k-nearest neighbor of SSCBS is set to  $3 \times 3$  neighborhood according to the empirical value, and the number of bands is set to be 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30. To reduce the processing time, the image is reduced to 0.1 times of the original when calculating the spatial-spectral combination function, which greatly improves the operation efficiency of the algorithm without destroying the spatial-spectral relationship between the target and the background in the image.

To verify the effectiveness of the band selection, SVM and KNN classifiers are used to classify three public hyperspectral datasets. In our experiment, the two classifiers have the same parameter settings on different datasets. The neighborhood parameter of the KNN classifier is set to be 5. SVM classifier uses an RBF kernel with penalty factors of C and gamma initialized to be  $1 \times 10^5$  and 0.5, respectively. Because both classifiers are supervised, 10% of all samples are randomly selected as the training set and the rest as the test set. Owing to the randomness of the training samples, the classification result is unstable. To reduce its influence, the final result is given by the average value of five calculations.

In this study, three classification accuracy measurements, OA (Overall Accuracy), AA (Average Accuracy), and Kappa coefficient, are adopted to evaluate the classification accuracy. To evaluate the effectiveness of the proposed method, the SSCBS is compared with five unsupervised band selection methods, including: ABS, E-FDPC, FNGBS, TRC\_OC\_FDPC and NC\_OC\_MVPCA. The evaluation is based on the three aspects of multiple dimensions of classification performance, optimal band set selection and calculation time.

### 3.3. Comparison of Different Subspace Partitioning Methods

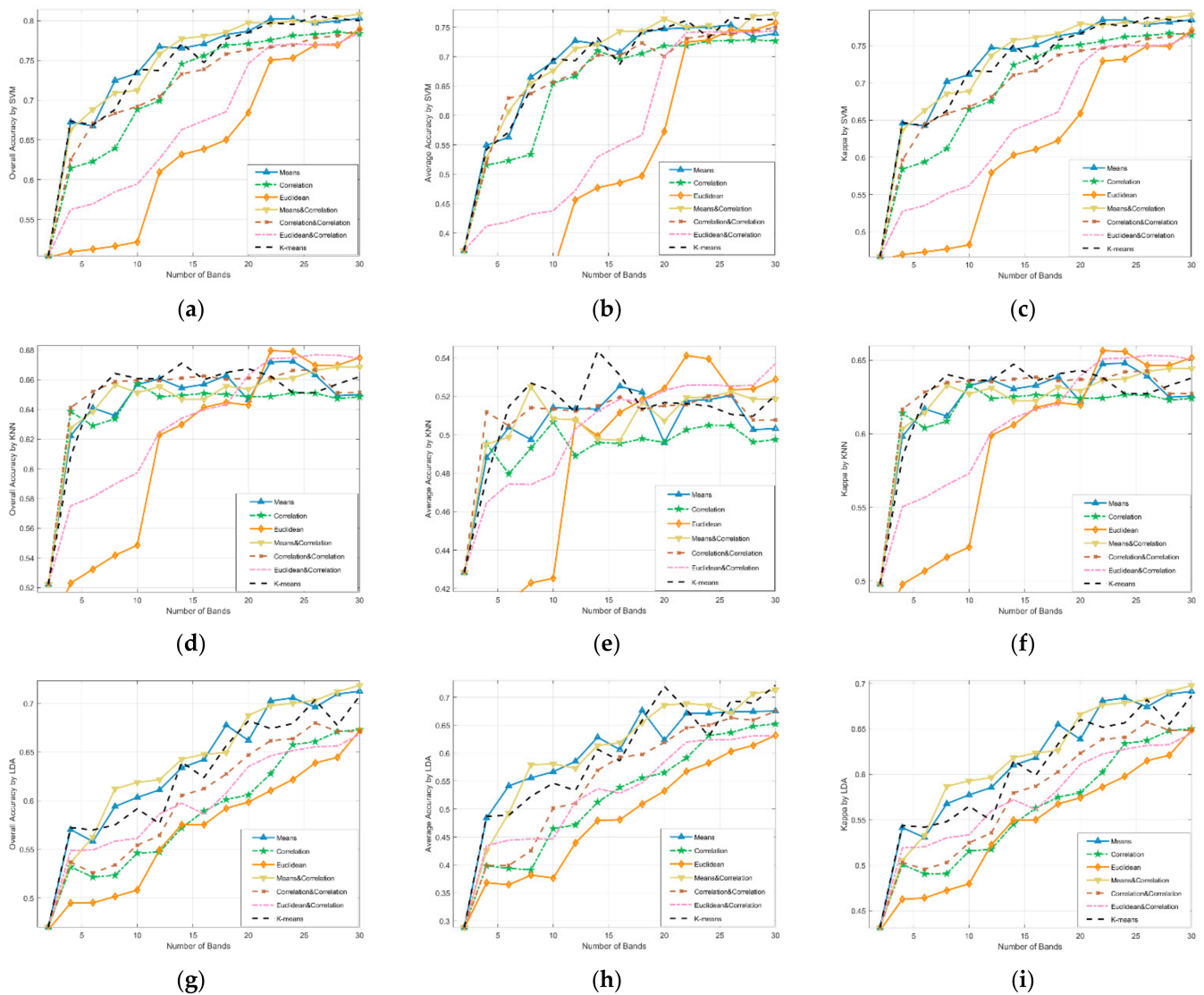
To compare the performance of the four types of subspace partitioning methods, means, correlation, Euclidean, means and correlation, correlation and correlation, Eu-



clidean and correlation and k-means algorithms are used to divide subspaces in the Indian Pines dataset. After division, the spatial–spectral combination algorithm described in Section 2.2.4 was used to select the optimal bands. Then, SVM, KNN and LDA classifiers are separately used for classification. The evaluation of classification performance is with OA, AA and Kappa coefficients, and the operating efficiency of the algorithm is compared with each other via computing time. The classification performances of the Euclidean, and Euclidean and correlation methods are shown in Figure 8a–i. It can be seen that the classification accuracy of Euclidean, Euclidean and correlation is significantly lower, regardless of whether the SVM, KNN or LDA classifier is adopted. The following is the correlation algorithm. However, means, means and correlation and k-means perform better. To further compare the performances of the three algorithms, the classification accuracy of each algorithm with the number of bands selected from 2 to 30 is calculated and the average of the acquired accuracy values is taken. The classification accuracy and calculation time after quantization are listed in Table 1. It can be seen that, when using the SVM classifier, the mean classification accuracy of the means algorithm is best. When using the KNN classifier, the mean classification accuracy of the k-means algorithm is best, and the means algorithm is second. From the perspective of the computing time, the means algorithm is significantly faster than the other algorithms. Considering the classification accuracy and computing time comprehensively, the means algorithm is selected as the best subspace partition method in this paper.

**Table 1.** Comparison of Different Subspace Partitioning methods on Indian Pines Data Set.

Classifier	Classification Accuracy	Means	Correlation	Euclidean	Means and Correlation	Correlation and Correlation	Euclidean and Correlation	K-Means
SVM	OA	<b>0.7447</b> ± 0.0134	0.7083 ± 0.0182	0.6371 ± 0.0203	0.7430 ± 0.0149	0.7182 ± 0.0097	0.6668 ± 0.0231	0.7398 ± 0.0189
	AA	0.6752 ± 0.0207	0.6333 ± 0.0156	0.5171 ± 0.0279	<b>0.6872</b> ± 0.0156	0.6675 ± 0.0221	0.5613 ± 0.0173	0.6747 ± 0.0129
	Kappa	<b>0.7230</b> ± 0.0245	0.6847 ± 0.0203	0.6078 ± 0.0168	0.7212 ± 0.0218	0.6950 ± 0.0147	0.6398 ± 0.0263	0.7177 ± 0.0136
	t(s)	<b>0.0001</b> ± 2E-5	0.0519 ± 0.0042	0.0296 ± 0.0022	9.7414 ± 0.068	9.9139 ± 0.093	9.9949 ± 0.075	9.8488 ± 0.084
KNN	OA	0.6506 ± 0.0276	0.6442 ± 0.0159	0.6089 ± 0.0145	0.6475 ± 0.0224	0.6497 ± 0.0178	0.6263 ± 0.0126	<b>0.6524</b> ± 0.0214
	AA	<b>0.5172</b> ± 0.0157	0.5069 ± 0.0166	0.4898 ± 0.0275	0.5152 ± 0.0243	0.5144 ± 0.0176	0.5110 ± 0.0189	0.5154 ± 0.0177
	Kappa	0.6271 ± 0.0173	0.6202 ± 0.0211	0.5852 ± 0.0156	0.6240 ± 0.0264	0.6262 ± 0.0166	0.6030 ± 0.0197	<b>0.6289</b> ± 0.0208
	t(s)	<b>0.0001</b> ± 3E-5	0.0536 ± 0.0025	0.0295 ± 0.0032	10.1750 ± 0.068	10.3461 ± 0.088	10.3837 ± 0.093	10.1713 ± 0.079
LDA	OA	0.6364 ± 0.0187	0.5829 ± 0.0245	0.5734 ± 0.0166	<b>0.6396</b> ± 0.0204	0.5989 ± 0.0198	0.6043 ± 0.0231	0.6307 ± 0.0187
	AA	<b>0.6134</b> ± 0.0239	0.5204 ± 0.0223	0.4889 ± 0.0169	0.6061 ± 0.0251	0.5468 ± 0.0297	0.5345 ± 0.0307	0.6040 ± 0.0163
	Kappa	0.6115 ± 0.0218	0.5546 ± 0.0266	0.5463 ± 0.0285	<b>0.6146</b> ± 0.0223	0.5710 ± 0.0153	0.5783 ± 0.0286	0.6055 ± 0.0266
	t(s)	<b>0.0001</b> ± 2E-5	0.0447 ± 0.0039	0.0267 ± 0.0080	7.6649 ± 0.089	7.8705 ± 0.069	7.8880 ± 0.082	7.6763 ± 0.057



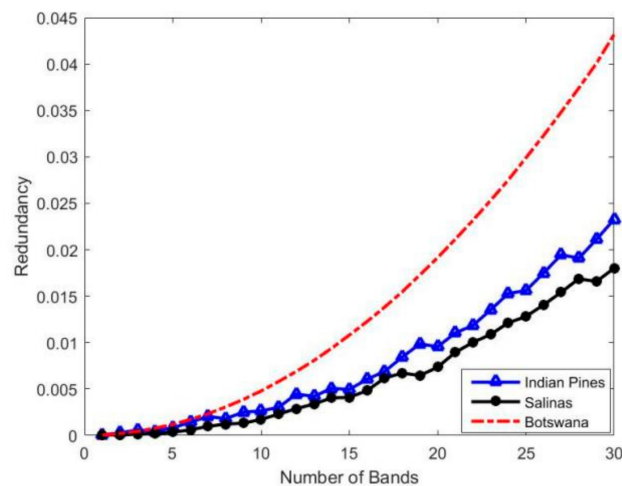
**Figure 8.** Classification accuracy of different subspace partitioning methods. (a–i) are OA, AA and Kappa Results by SVM, KNN and LDA classifiers.

### 3.4. Experimental Results and Analysis

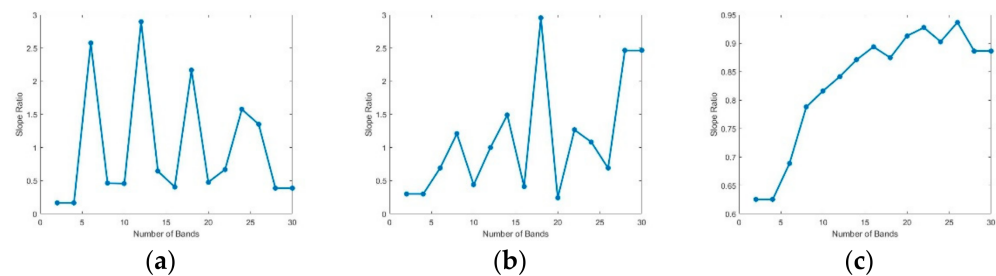
To investigate the performance of the proposed SSCBS algorithm at different latitudes, further analysis will be carried out in the aspects of the recommended bands, the classification performance and the computational time.

#### 3.4.1. Recommended Bands

According to the method based on information redundancy proposed in Section 2.2.6, the information redundancy curves are drawn for the Indian Pines data set, the Salinas data set and the Botswana data set separately, as shown in Figure 9. Through Equation (14), the change rate of the slope of information redundancy on each data set is obtained, as shown in Figure 10.



**Figure 9.** Information redundancy curves on different data sets.



**Figure 10.** Slope rate of information redundancy (a) Indian Pines data set (b) Salinas data set (c) Botswana data set.

The recommended number of band sets should be on the location of the local maximum points. As can be seen from Figure 10, the local maximum points are 6, 12, 18 and 24 on the Indian Pines data set. The local maximum points are 8, 14, 18, and 22 on the Salinas data set and 16, 22, and 26 on the Botswana data set. Combined with the classification performance of the proposed algorithm, the best number of band sets suitable for this paper is selected from the recommended number of band sets, and 12, 22 and 16 are selected for the Indian Pines, the Salinas and the Botswana data sets, respectively. Table 2 lists the band sets of different band selection algorithms in each data set, arranged in ascending order, and adjacent bands are displayed in bold. Hyperspectral images are characterized by strong correlation between adjacent bands and high redundancy. Generally speaking, the closer the adjacent bands are, the stronger the correlation and the worse the algorithm. As can be seen from Table 3, the bands selected by ABS, E-FDPC, FNGBS, NC\_OC\_MVPCA and TRC\_OC\_FDPC algorithms have adjacent bands, and the bands selected by TRC\_OC\_FDPC algorithm have an edge band 1, which are all unreasonable bands. However, the band subset selected by SSCBS in this paper is more dispersed, covering a larger spectral range, and has stable performance in the three data sets.

Table 2. Classification Results on Three Data Sets.

Dataset	M	Classifier	Method					
			E-FDPC	ABS	FNGBS	NC_OC_MVPCA	TRC_OC_FDPC	SSCBS
Indian Pines	6	SVM	0.4693 ± 0.0218	0.4392 ± 0.0196	0.5816 ± 0.0309	<b>0.6224</b> ± 0.0383	0.5598 ± 0.0276	0.5397 ± 0.0223
		KNN	0.475 ± 0.0133	0.4778 ± 0.0088	0.5417 ± 0.0332	0.5168 ± 0.0315	0.5098 ± 0.0235	<b>0.5568</b> ± 0.0187
		LDA	0.4515 ± 0.0143	0.3274 ± 0.0093	<b>0.5281</b> ± 0.0328	0.4707 ± 0.0421	0.4743 ± 0.0274	0.5276 ± 0.0258
	12	SVM	0.5147 ± 0.0147	0.5348 ± 0.0120	0.6895 ± 0.0356	0.692 ± 0.0302	0.6944 ± 0.0286	<b>0.7354</b> ± 0.0278
		KNN	0.481 ± 0.0279	0.47 ± 0.0056	0.5596 ± 0.0259	0.5677 ± 0.0409	0.5129 ± 0.0210	<b>0.5726</b> ± 0.0381
		LDA	0.5562 ± 0.0304	0.3938 ± 0.0027	<b>0.5984</b> ± 0.0251	0.5513 ± 0.0358	0.5674 ± 0.0246	0.5792 ± 0.0295
	18	SVM	0.536 ± 0.0215	0.6117 ± 0.0084	0.7587 ± 0.0195	0.7193 ± 0.0295	0.7454 ± 0.0247	<b>0.7663</b> ± 0.0216
		KNN	0.4872 ± 0.0294	0.4912 ± 0.0078	0.5705 ± 0.0155	0.5686 ± 0.0406	0.5338 ± 0.0224	<b>0.5829</b> ± 0.0224
		LDA	0.5532 ± 0.0217	0.444 ± 0.0108	0.6368 ± 0.0292	0.634 ± 0.0327	0.5935 ± 0.0268	<b>0.6553</b> ± 0.0218
	24	SVM	0.5722 ± 0.0210	0.6181 ± 0.0125	0.7546 ± 0.0357	0.7628 ± 0.0269	0.7566 ± 0.0394	<b>0.7836</b> ± 0.0169
		KNN	0.4899 ± 0.0341	0.4903 ± 0.0102	0.571 ± 0.0014	0.587 ± 0.0287	0.5439 ± 0.0281	<b>0.5745</b> ± 0.0052
		LDA	0.5558 ± 0.0157	0.439 ± 0.0047	0.6515 ± 0.0368	<b>0.7032</b> ± 0.0294	0.6338 ± 0.0357	0.6847 ± 0.0259
	30	SVM	0.6227 ± 0.0157	0.6586 ± 0.0102	0.7675 ± 0.027	0.7583 ± 0.0237	0.7598 ± 0.0248	<b>0.7728</b> ± 0.0213
		KNN	0.4963 ± 0.0425	0.5271 ± 0.0317	0.5645 ± 0.0073	0.5831 ± 0.0261	0.5727 ± 0.0413	<b>0.5957</b> ± 0.0187
		LDA	0.5662 ± 0.0158	0.4993 ± 0.0196	<b>0.6819</b> ± 0.0286	0.72 ± 0.0302	0.6528 ± 0.0267	0.6672 ± 0.0158
	6	SVM	0.9388 ± 0.0157	0.7484 ± 0.0014	0.9399 ± 0.0119	<b>0.9428</b> ± 0.0135	0.9321 ± 0.0237	0.94 ± 0.0153
		KNN	0.9237 ± 0.0132	0.7542 ± 0.0125	0.9257 ± 0.0147	<b>0.9267</b> ± 0.0204	0.9215 ± 0.0138	0.9218 ± 0.0149
		LDA	0.8415 ± 0.0179	0.6984 ± 0.0093	0.8684 ± 0.0148	0.8741 ± 0.0151	0.8437 ± 0.0197	<b>0.8863</b> ± 0.0104
	12	SVM	0.9501 ± 0.0178	0.8618 ± 0.0156	0.9527 ± 0.0147	0.9537 ± 0.0249	0.9529 ± 0.0193	<b>0.9545</b> ± 0.0172
		KNN	0.9328 ± 0.0138	0.8262 ± 0.0094	0.932 ± 0.0273	0.9308 ± 0.0192	0.9263 ± 0.0124	<b>0.9339</b> ± 0.0204
		LDA	0.9128 ± 0.0084	0.7474 ± 0.0047	0.905 ± 0.0138	<b>0.918</b> ± 0.0216	0.8973 ± 0.0273	0.9178 ± 0.0178
	18	SVM	0.958 ± 0.0197	0.8945 ± 0.0117	0.9538 ± 0.0180	<b>0.958</b> ± 0.0132	0.9572 ± 0.0208	0.9574 ± 0.0153
		KNN	0.9371 ± 0.0196	0.8644 ± 0.0148	0.9322 ± 0.0284	0.9327 ± 0.0235	0.929 ± 0.0148	<b>0.9349</b> ± 0.0150
		LDA	0.9236 ± 0.0146	0.7875 ± 0.0153	0.9198 ± 0.0186	0.9269 ± 0.0138	0.9178 ± 0.0196	<b>0.9292</b> ± 0.0153
	24	SVM	0.9591 ± 0.0205	0.9224 ± 0.0174	0.9569 ± 0.0208	<b>0.961</b> ± 0.0156	0.9603 ± 0.0195	0.9593 ± 0.0147
		KNN	0.9337 ± 0.0140	0.9027 ± 0.0196	0.9312 ± 0.0146	0.9348 ± 0.0207	0.9296 ± 0.0147	<b>0.9352</b> ± 0.0084
		LDA	0.9266 ± 0.0168	0.8409 ± 0.0147	0.9257 ± 0.0185	<b>0.932</b> ± 0.0296	0.9229 ± 0.0157	0.9306 ± 0.0161
	30	SVM	0.9603 ± 0.0159	0.939 ± 0.0102	0.9618 ± 0.0162	<b>0.963</b> ± 0.0147	0.9626 ± 0.0205	0.9616 ± 0.0159
		KNN	0.9351 ± 0.0149	0.9092 ± 0.0175	0.9342 ± 0.0256	0.9357 ± 0.0196	0.9316 ± 0.0135	<b>0.9377</b> ± 0.0115
		LDA	0.9301 ± 0.0154	0.8748 ± 0.0102	0.9298 ± 0.0157	<b>0.9346</b> ± 0.0179	0.9274 ± 0.0271	0.9328 ± 0.0143

Table 2. Cont.

Dataset	M	Classifier	Method					
			E-FDPC	ABS	FNGBS	NC_OC_MVPCA	TRC_OC_FDPC	SSCBS
Botswana	6	SVM	0.7413 ± 0.0234	0.6787 ± 0.0157	0.8439 ± 0.0275	0.8587 ± 0.0235	0.8379 ± 0.0186	<b>0.861</b> ± 0.0169
		KNN	0.7752 ± 0.0157	0.6319 ± 0.0177	0.8115 ± 0.0286	<b>0.8399</b> ± 0.0223	0.8223 ± 0.0288	0.8355 ± 0.0213
		LDA	0.818 ± 0.0286	0.7339 ± 0.0254	0.8821 ± 0.0271	0.8734 ± 0.0209	0.8642 ± 0.0374	<b>0.8763</b> ± 0.0196
	12	SVM	0.8671 ± 0.0237	0.7199 ± 0.0211	0.8732 ± 0.0297	<b>0.873</b> ± 0.0317	0.8711 ± 0.0211	0.861 ± 0.0169
		KNN	0.8485 ± 0.0260	0.6188 ± 0.0182	0.8335 ± 0.0299	<b>0.8497</b> ± 0.0301	0.8408 ± 0.0223	0.8329 ± 0.0297
		LDA	0.8894 ± 0.0280	0.7514 ± 0.0214	0.9002 ± 0.0166	<b>0.8951</b> ± 0.0183	0.8868 ± 0.0288	0.8876 ± 0.0209
	18	SVM	0.8714 ± 0.0281	0.8763 ± 0.0197	0.8712 ± 0.0212	0.8732 ± 0.0231	0.8797 ± 0.0375	<b>0.8777</b> ± 0.0280
		KNN	0.8548 ± 0.0297	0.8094 ± 0.0275	0.8365 ± 0.0269	0.8463 ± 0.0234	<b>0.8478</b> ± 0.0127	0.8464 ± 0.0251
		LDA	0.8938 ± 0.0286	0.8588 ± 0.0163	0.8964 ± 0.0218	0.8973 ± 0.0264	0.8922 ± 0.0188	<b>0.9001</b> ± 0.0157
	24	SVM	0.8764 ± 0.0175	0.8885 ± 0.0191	0.8798 ± 0.0144	<b>0.897</b> ± 0.0219	0.8937 ± 0.0157	0.8853 ± 0.0136
		KNN	0.861 ± 0.0188	0.8336 ± 0.0098	0.8441 ± 0.0127	<b>0.8596</b> ± 0.0274	0.8546 ± 0.0146	0.849 ± 0.0124
		LDA	0.8945 ± 0.0286	0.8624 ± 0.0189	0.8973 ± 0.0145	<b>0.9082</b> ± 0.0269	0.8922 ± 0.0128	0.9066 ± 0.0087
	30	SVM	0.8801 ± 0.0143	0.8824 ± 0.0186	0.8897 ± 0.0162	0.9025 ± 0.0121	0.8911 ± 0.0198	<b>0.9041</b> ± 0.0087
		KNN	0.8589 ± 0.0175	0.8241 ± 0.0188	0.8522 ± 0.0223	0.8607 ± 0.0129	0.8584 ± 0.0186	<b>0.861</b> ± 0.0213
		LDA	0.8945 ± 0.0166	0.8594 ± 0.0159	0.9022 ± 0.0035	0.9067 ± 0.0068	0.8906 ± 0.0082	<b>0.9068</b> ± 0.0034

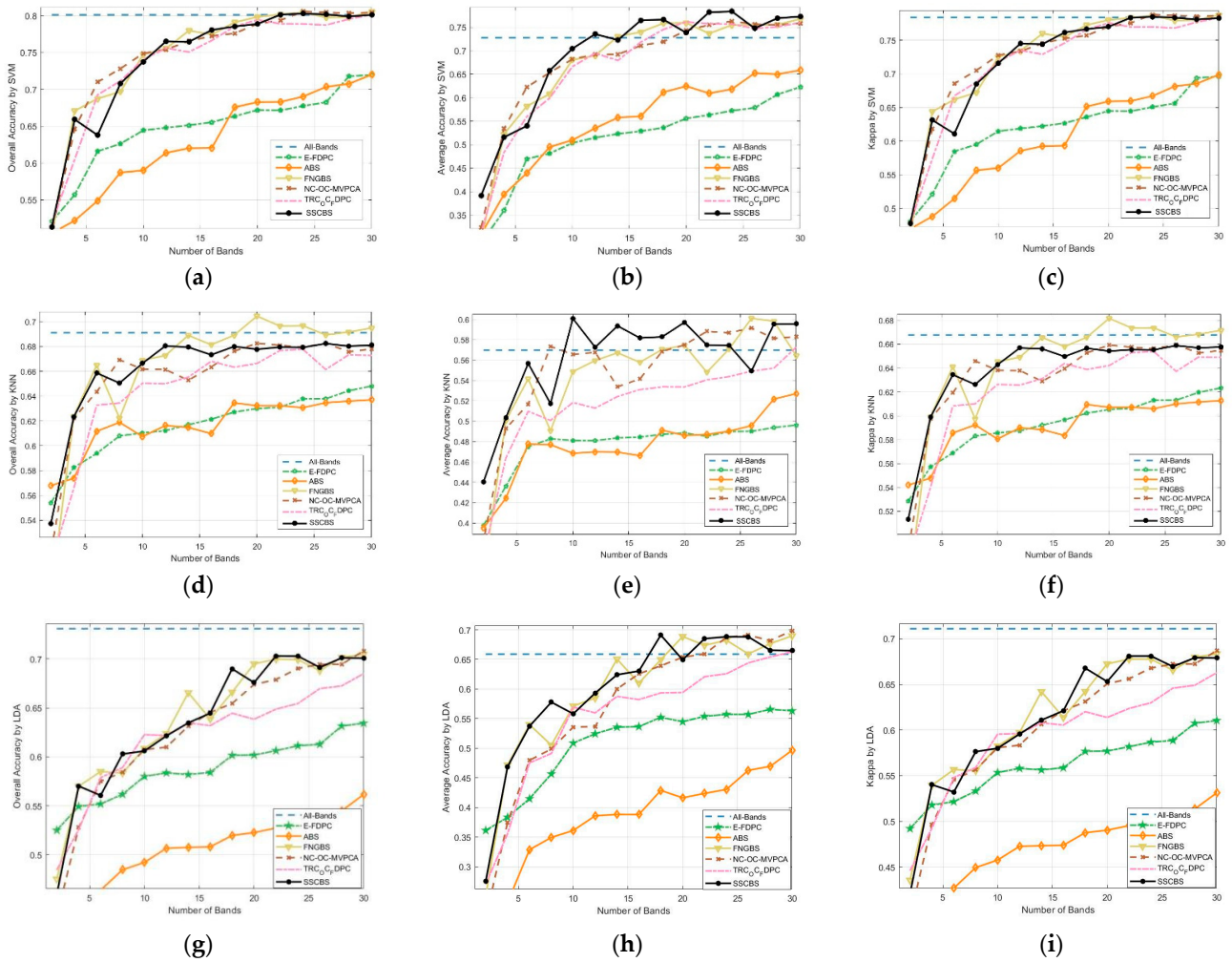
Table 3. Computational Time on Three Data Sets.

Data set	Classifier	E-FDPC	ABS	FNGBS	NC_OC_MVPCA	TRC_OC_FDPC	SSCBS
Indian Pines	SVM	0.0948	0.5413	0.4326	0.6399	0.7779	0.3918
	KNN	0.1083	0.5108	0.4285	0.6411	0.7841	0.4067
	LDA	0.0492	0.0371	0.2301	0.3749	0.5291	0.1897
Salinas	SVM	0.4819	2.8015	1.3425	1.8894	1.5748	1.2712
	KNN	0.6344	4.1058	1.7736	2.7571	1.8979	1.7373
	LDA	0.3145	0.2796	1.0651	1.1772	1.3049	0.8705
Botswana	SVM	1.1568	8.2198	4.8423	4.2913	2.638	5.256
	KNN	1.2409	5.6202	4.1123	4.088	2.5618	3.9562
	LDA	0.5855	0.4557	2.2882	1.8047	1.3911	2.1106

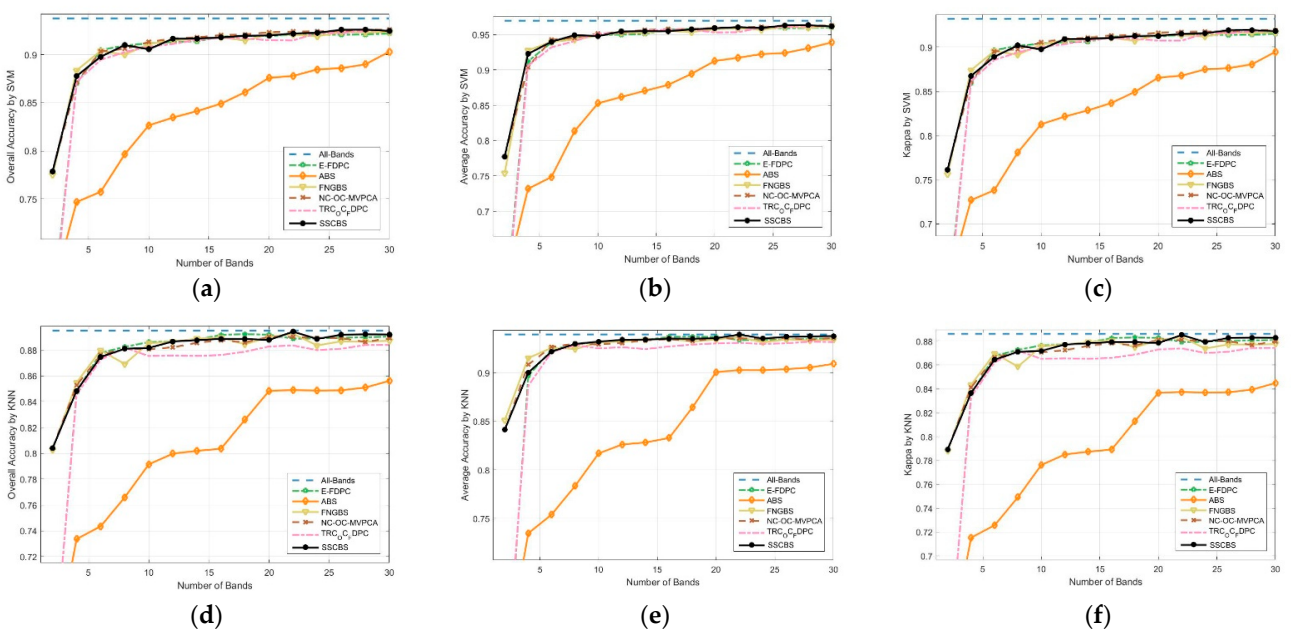
### 3.4.2. Classification Performance

To verify the effectiveness of the proposed algorithm, the SSCBS, SVM, KNN and LDA classifiers are used for classification on the three datasets: Indian Pines, Salinas, and Botswana. A comparison of the classification accuracies is shown in Figures 11–13.

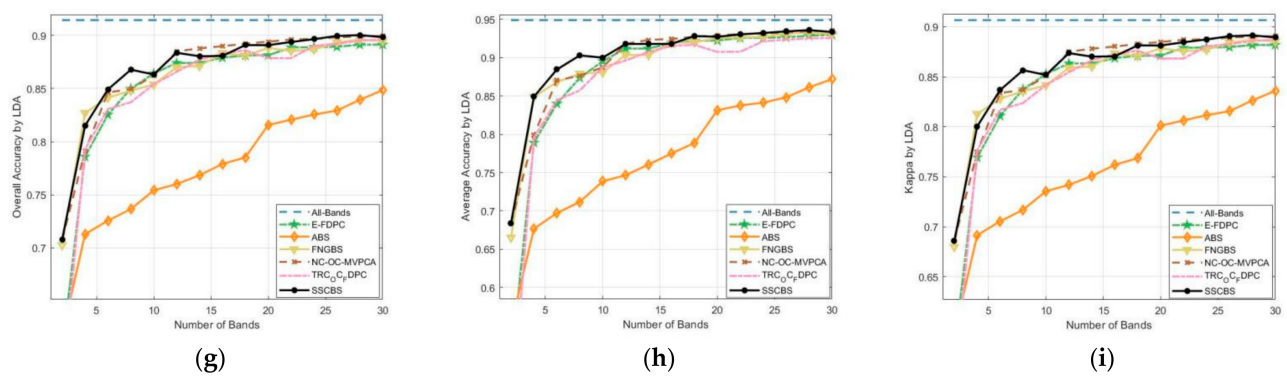




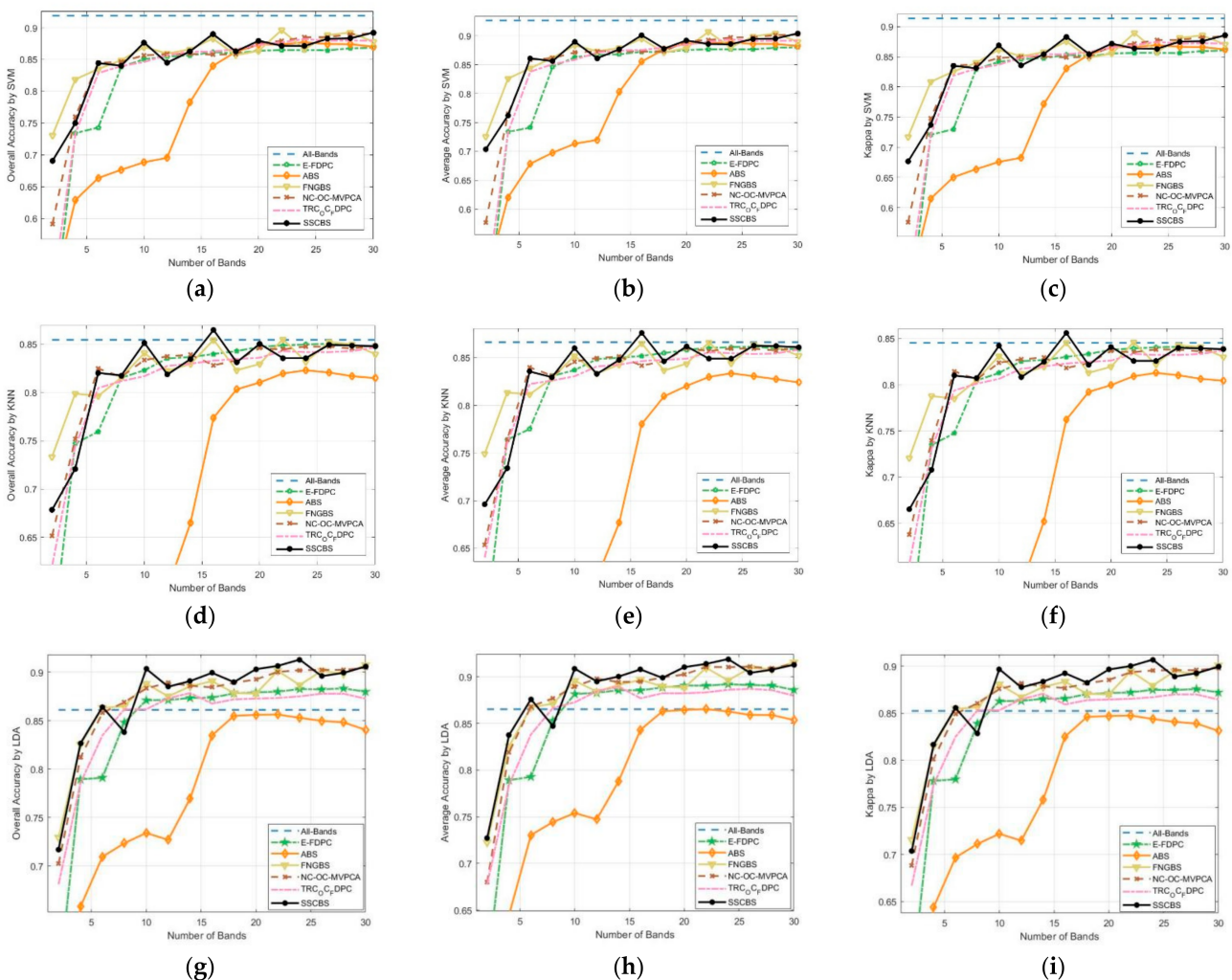
**Figure 11.** Classification results on Indian Pines data set. (a–i) are OA, AA and Kappa Results by SVM, KNN and LDA classifiers.



**Figure 12.** Cont.



**Figure 12.** Classification results on Salinas data set. (a–i) are OA, AA and Kappa Results by SVM, KNN and LDA classifiers.



**Figure 13.** Classification results on Botswana data set. (a–i) are OA, AA and Kappa Results by SVM, KNN and LDA classifiers.

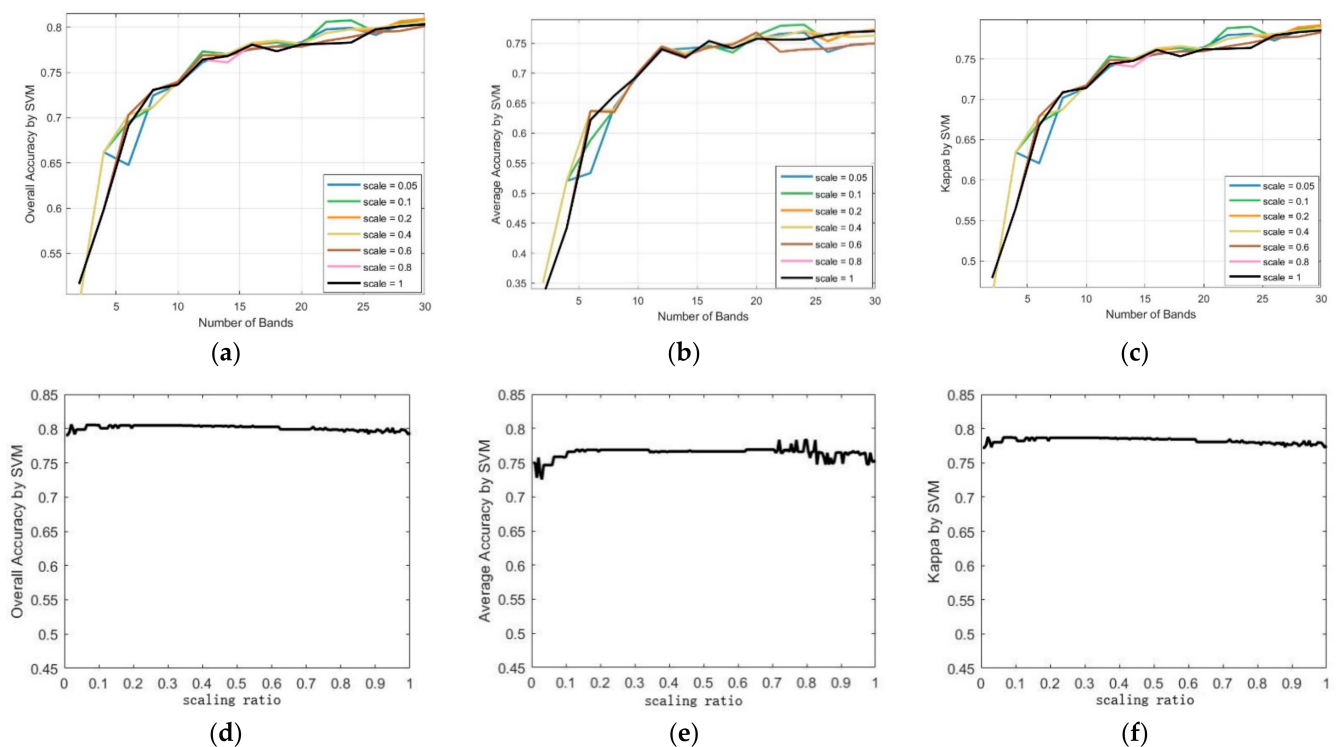
It can be seen from Figure 11a–i that the proposed SSCBS algorithm performs best on the Indian Pines dataset using the SVM classifier. Using the KNN classifier, SSCBS has the highest AA accuracy. FNGBS has the highest OA and Kappa coefficients, SSCBS is second, and ABS and E-FDPC have the worst performance. In the Salinas dataset, as shown in Figure 12a–i, except for the ABS algorithm, the performance of the other six algorithms

is good, and the SSCBS method tends to be best. As shown in Figure 13a–i, the SSCBS method performs best on the Botswana data set. Combining the three datasets, SSCBS is the best comprehensive performance. To quantify the advantages of the SSCBS algorithm in the classification accuracy of different datasets, the classification accuracy results of selecting the representative bands are listed in Table 2, and the maximum value of each row is represented in bold.

By comparison, we can notice that SSCBS maintained excellent classification performance on the different datasets. This shows that the algorithm in this paper fully considers the spatial correlation and spectral correlation of pixels, and the optimal bands extracted not only have low spatial correlation, but also have strong spectral specificity and obvious spectral characteristics between different classes, which is helpful to image classification.

### 3.4.3. Computational Time

To compare the operating efficiency of the six algorithms, the running times of the six algorithms on the three datasets are recorded. Thirty bands on the datasets are selected for the running time calculation. The results are listed in Table 3. It can be seen from the table that E-FDPC has the shortest computational time on the three datasets. Then on the Indian Pines and the Salinas datasets, generally SSCBS has the second short computational time. On the Botswana dataset, owing to the larger image resolution of this dataset, the calculation time of SSCBS algorithm is longer. This is because the SSCBS algorithm considers the spatial–spectral relationship between pixels, and the calculation time is related to the image resolution of the dataset. The higher the image resolution, the longer the calculation time of the algorithm. The operation effect of the SSCBS algorithm can be improved by setting a larger zoom ratio of the original image. Figure 14a–f shows a comparison of classification accuracy using the SVM classifier on the Indian Pines dataset with different scaling ratios in the proposed algorithm. In Figure 14d–f, the band is selected to be 30 and the scaling ratio range is from 0.005 to 1, with an interval of 0.005. As is shown in the figure, the results are stable regardless of the scaling ratio used.



**Figure 14.** Comparison of classification accuracy with different scaling ratios. (a–c) are OA, AA and Kappa Results by SVM. (d–f) are OA, AA and Kappa Results under different scaling ratios.

#### 4. Conclusions

For hyperspectral image data, the most valuable information is usually concentrated in limited spectral attributes. Selecting a band subset that is sensitive to different types of ground objects can classify images more effectively. The SSCBS method is proposed in order to extract representative bands with large information content, low spatial correlation and strong inter-class spectral distinctness. First, all the bands are divided into the required number of subsets through subspace division. In each subset, the spatial and spectral combination features are integrated into a uniformed objective function by constructing the weight coefficient. Then the band with the large and more discriminative information is selected. Third, in order to solve the problem of the appropriate number of bands to choose, an automatic method is proposed to determine the number of appropriate bands. Experimental results show that the optimal band subset extracted by SSCBS leads to excellent classification accuracy and less time consumption. If the image resolution is high, it can be solved by reducing the resolution of the original image. The experimental results show that reducing the image resolution does not affect the classification accuracy.

**Author Contributions:** Conceptualization, X.H. and Z.J.; methodology, X.H.; software, X.H.; validation, X.H., Z.J. and Y.L. (Yuanyuan Liu); formal analysis, Y.L. (Yuanyuan Liu); investigation, X.H. and J.Z.; resources, X.H. and Y.L. (Yingzhi Li); data curation, X.H.; writing—original draft preparation, X.H.; writing—review and editing, Z.J., J.Z. and Q.S.; visualization, X.H. and Y.L. (Yuanyuan Liu); supervision, J.Z. and Y.L. (Yingzhi Li); project administration, Q.S.; funding acquisition, Q.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China under Grant 61801455, and in part by the National Natural Science Foundation of China under Grant 62175233.

**Data Availability Statement:** Indian Pines dataset: [http://www.ehu.es/ccwintco/uploads/2/22/Indian\\_pines.mat](http://www.ehu.es/ccwintco/uploads/2/22/Indian_pines.mat) (accessed on 31 May 2022); Salinas dataset: <http://www.ehu.es/ccwintco/uploads/f/f1/Salinas.mat> (accessed on 31 May 2022); Botswana dataset: <http://www.ehu.es/ccwintco/uploads/7/72/Botswana.mat> (accessed on 31 May 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Lishuan, H. *Study of Dimensionality Reduction and Spatial-Spectral Method for Classification of Hyperspectral Remote Sensing Image*; China University of Geosciences: Wuhan, China, 2018.
2. Plaza, A.; Benediktsson, J.A.; Boardman, J.W.; Brazile, J.; Bruzzone, L.; Camps-Valls, G.; Chanussot, J.; Fauvel, M.; Gamba, P.; Gualtieri, A.; et al. Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* **2009**, *113*, S110–S122. [CrossRef]
3. Hughes, G.F. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *IT-14*, 55–63. [CrossRef]
4. Guo, B.; Gunn, S.R.; Damper, R.I.; Nelson, J.D.B. Band selection for hyperspectral image classification using mutual information. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 522–526. [CrossRef]
5. Sun, W.; Halevy, A.; Benedetto, J.J.; Czaja, W.; Li, W.; Liu, C.; Shi, B.; Wang, R. Nonlinear dimensionality reduction via the ENHLTSA method for hyperspectral image classification. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2014**, *7*, 375–388. [CrossRef]
6. Lu, T.; Li, S.; Fang, L.; Ma, Y.; Benediktsson, J.A. Spectral–spatial adaptive sparse representation for hyperspectral image denoising. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 373–385. [CrossRef]
7. Geng, X.; Sun, K.; Ji, L.; Zhao, Y. A fast volume-gradient-based band selection method for hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7111–7119. [CrossRef]
8. Luis, O.J.; David, A.L. Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 2653–2667.
9. Jia, X.; Richards, J.A. Segmented principal components transformation for efficient hyperspectral remote sensing image display and classification. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 538–542.
10. He, X.; Niyogi, P. Locality preserving projections. *Proc. Adv. Neural Inf. Process. Syst.* **2004**, *16*, 153–160.
11. He, X.; Cai, D.; Yan, S.; Zhang, H. Neighborhood preserving embedding. In Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, China, 17–21 October 2005; Volume 2, pp. 1208–1213.
12. Friedman, J. Regularized discriminant analysis. *J. Am. Stat. Assoc.* **1989**, *84*, 165–175. [CrossRef]



13. Bandos, T.V.; Bruzzone, L.; Camps-Valls, G. Classification of hyperspectral images with regularized linear discriminant analysis. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 862–873. [\[CrossRef\]](#)
14. Kuo, B.C.; Landgrebe, D.A. Nonparametric weighted feature extraction for classification. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1096–1105.
15. Sugiyama, M. Dimensionality reduction of multimodal labeled data by localFisher discriminant analysis. *J. Mach. Learn. Res.* **2007**, *8*, 1027–1061.
16. Chen, H.T.; Chang, H.W.; Liu, T.L. Local discriminant embedding and its variants. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; pp. 846–853.
17. Yan, S.; Xu, D.; Zhang, B.; Zhang, H.J.; Yang, Q.; Lin, S. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 40–51. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Serpico, S.B.; Bruzzone, L. A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 1360–1367. [\[CrossRef\]](#)
19. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182.
20. Li, Q.; Wang, Q.; Li, X. An efficient clustering method for hyper- spectral optimal band selection via shared nearest neighbor. *Remote Sens.* **2019**, *11*, 350. [\[CrossRef\]](#)
21. Wang, Q.; Li, Q.; Li, X. Hyperspectral band selection via adaptive subspace partition strategy. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2019**, *12*, 4940–4950. [\[CrossRef\]](#)
22. Yang, H.; Du, Q.; Su, H.; Sheng, Y. An efficient method for supervised hyperspectral band selection. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 138–142. [\[CrossRef\]](#)
23. Archibald, R.; Fann, G. Feature selection and classification of hyperspectral images with support vector machines. *IEEE Geosci. Remote Sens. Lett.* **2007**, *4*, 674–677. [\[CrossRef\]](#)
24. Bai, J.; Xiang, S.; Pan, C. Classification oriented semi-supervised band selection for hyperspectral images. In Proceedings of the 21st International Conference on Pattern Recognition, Tsukuba, Japan, 11–15 November 2012; pp. 1888–1891.
25. Chen, L.; Huang, R.; Huan, W. Graph-based semi-supervised weighted band selection for classification of hyperspectral data. In Proceedings of the 2010 International Conference on Audio, Language and Image Processing, Shanghai, China, 23–25 November 2010; pp. 1123–1126.
26. Reza, M.; Ghamary, M.; Mojaradi, B. Unsupervised feature selection using geometrical measures in prototype space for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3774–3787.
27. Chunhong, L.; Chunhui, Z.; Lingyan, Z. A new dimensionality reduction method for hyperspectral remote sensing image. *J. Image Graph.* **2005**, *10*, 218–223.
28. Jia, S.; Tang, G.; Zhu, J.; Li, Q. A Novel Ranking-Based Clustering Approach for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 88–102. [\[CrossRef\]](#)
29. Qi, W.; Qiang, L.; Xuelong, L. A Fast Neighborhood Grouping Method for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5028–5039.
30. Wang, Q.; Zhang, F.H.; Li, X.L. Optimal clustering framework for hyperspectral band selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5910–5922. [\[CrossRef\]](#)
31. Zhou, Y.; Peng, J.; Chen, C. Dimension Reduction Using Spatial and Spectral Regularized Local Discriminant Embedding for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1082–1095. [\[CrossRef\]](#)
32. Huang, H.; Shi, G.; He, H.; Duan, Y.; Luo, F. Dimensionality Reduction of Hyperspectral Imagery Based on Spatial-spectral Manifold Learning. *arXiv* **2018**, arXiv:1812.09530. [\[CrossRef\]](#)
33. Zhou, L.; Zhang, X. Discriminative spatial-spectral manifold embedding for hyperspectral image classification. *Remote Sens. Lett.* **2015**, *6*, 715–724. [\[CrossRef\]](#)
34. Zhao, W.; Du, S. Spectral–Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4544–4554. [\[CrossRef\]](#)
35. Feng, Z.; Yang, S.; Wang, S.; Jiao, L. Discriminative Spectral–Spatial Margin-Based Semi-supervised Dimensionality Reduction of Hyperspectral Data. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 224–228. [\[CrossRef\]](#)
36. Bai, X.; Guo, Z.; Wang, Y.; Zhang, Z.; Zhou, J. Semi-supervised Hyperspectral Band Selection Via Spectral–Spatial Hypergraph Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2774–2783. [\[CrossRef\]](#)
37. Liang, Z.; Ligu, W.; Danfeng, L. A subspace band selection method for hyperspectral imagery. *J. Remote Sens.* **2019**, *23*, 904–910.
38. Dehui, Z.; Bo, D.; Liangpei, Z. Band selection-based collaborative representation for anomaly detection in hyperspectral images. *J. Remote Sens.* **2020**, *24*, 427–438.
39. Yanlong, C.; Xiaolan, W.; En, L.; Mei-ping, S.; Hai-mo, B. Research and Application of Band Selection Method Based on CEM. *Spectrosc. Spectr. Anal.* **2020**, *40*, 3778–3783.
40. Fuquan, Z.; Huajun, W.; Liping, Y.; Changguo, L. Hyperspectral image lossless compression using adaptive bands selection and optimal prediction sequence. *Opt. Precis. Eng.* **2020**, *28*, 1609–1617.
41. Pal, M.; Foody, G.M. Feature selection for classification of hyperspectral data by SVM. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2297–2307. [\[CrossRef\]](#)
42. Kulesza, A. Determinantal point processes for machine learning. *Found. Trends Mach. Learn.* **2012**, *5*, 123–286. [\[CrossRef\]](#)