agriculture

MDPI

Article

Land Cover Classification from Hyperspectral Images via Weighted Spatial–Spectral Joint Kernel Collaborative Representation Classifier

Rongchao Yang [1,2], Qingbo Zhou [1,2,*], Beilei Fan [1,2], Yuting Wang [1,2] and Zhemin Li [3]

[1] Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[2] Key Laboratory of Agricultural Blockchain Application, Ministry of Agriculture and Rural Affairs, Beijing 100081, China
[3] Graduate School of Chinese Academy of Agricultural Sciences, Beijing 100081, China
* Correspondence: zhouqingbo@caas.cn

Abstract: The continuous changes in Land Use and Land Cover (LULC) produce a significant impact on environmental factors. Highly accurate monitoring and updating of land cover information is essential for environmental protection, sustainable development, and land resource planning and management. Recently, Collaborative Representation (CR)-based methods have been widely used in land cover classification from Hyperspectral Images (HSIs). However, most CR methods consider the spatial information of HSI by taking the average or weighted average of spatial neighboring pixels of each pixel to improve the land cover classification performance, but do not take the spatial structure information for pixels into account. To address this problem, a novel Weighted Spatial–Spectral Joint CR Classification (WSSJCRC) method is proposed in this paper. WSSJCRC only performs spatial filtering on HSI through a weighted spatial filtering operator to alleviate the spectral shift caused by adjacency effect, but also utilizes the labeled training pixels to simultaneously represent each test pixel and its spatial neighborhood pixels to consider the spatial structure information of each test pixel to assist the classification of the test pixel. On this basis, the kernel version of WSSJCRC (i.e., WSSJKCRC) is also proposed, which projects the hyperspectral data into the kernel-induced high-dimensional feature space to enhance the separability of nonlinear samples. The experimental results on three real hyperspectral scenes show that the proposed WSSJKCRC method achieves the best land cover classification performance among all the compared methods. Specifically, the Overall Accuracy (OA), Average Accuracy (AA), and Kappa statistic (Kappa) of WSSJKCRC reach 96.21%, 96.20%, and 0.9555 for the Indian Pines scene, 97.02%, 96.64%, and 0.9605 for the Pavia University scene, and 95.55%, 97.97%, and 0.9504 for the Salinas scene, respectively. Moreover, the proposed WSSJKCRC method obtains the promising accuracy with OA over 95% on the three hyperspectral scenes under the situation of small-scale labeled samples, thus effectively reducing the labeling cost for HSI.

Keywords: land cover classification; hyperspectral images; weighted spatial–spectral features; spatial structure information; kernel trick; collaborative representation

check for updates

Citation: Yang, R.; Zhou, Q.; Fan, B.; Wang, Y.; Li, Z. Land Cover Classification from Hyperspectral Images via Weighted Spatial–Spectral Joint Kernel Collaborative Representation Classifier. *Agriculture* **2023**, *13*, 304. https://doi.org/10.3390/agriculture13020304

Academic Editor: Ritaban Dutta

Received: 14 December 2022
Revised: 11 January 2023
Accepted: 25 January 2023
Published: 27 January 2023

Copyright: © 2023 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

With the rapid development of urbanization and the intervention of human activities, the forms of Land Use and Land Cover (LULC) are constantly changing [1]. The changes in LULC produce a significant impact on environmental factors such as climate, water balance, biodiversity, and terrestrial ecosystems [2]. Highly accurately monitoring and updating land cover information is not only crucial for environmental protection and scientific studies, but also plays an important role in land resource planning and management, landscape pattern analysis, sustainable development, and many others [3–6]. Remote Sensing (RS) technology, as one of the most important means for earth observation, has been

widely used in land cover classification and mapping through spaceborne or airborne visible, multispectral, hyperspectral, and other imaging sensors due to its promising accuracy and high efficiency [7–9]. By integrating image and spectral technology, hyperspectral RS technology provides abundant information for ground objects in both the spectral and spatial domains through hundreds of narrow and continuous spectral bands. Therefore, hyperspectral technology can effectively eliminate the phenomenon of "same objects with different spectra" and "foreign objects with same spectra" and improve the discrimination ability of ground objects [10–12]. As a consequence, hyperspectral RS technology has attracted great attention and interest from researchers.

In terms of land cover classification from Hyperspectral Images (HSIs), researchers mostly establish classification models based on statistical supervised machine learning algorithms, such as support vector machine [13,14], extreme learning machine [15,16], random forests [8,17], and so on. Such statistic-based classification algorithms not only need to assume that the sample data follow normal or multimodal distribution [18,19], but also need a large number of labeled samples for training to fit the models. However, it is time-consuming, laborious, and expensive to obtain the true label of samples (pixels) in the practical application of HSI [20–22]. The lack of labeled samples makes it difficult to meet the distribution assumption of data and affects the fitting performance of the statistic-based classification models [23]. Moreover, deep learning algorithms have been widely applied to HSI classification in recent years, such as convolutional neural networks [24,25]. Deep convolutional neural networks can automatically learn complex high-level semantic features in HSI and obtain strong classification performance. However, such algorithms contain a great quantity of weight parameters and require more complex training process. For the above reasons, the large-scale labeled samples are needed in the training process to ensure the fitting effect of the networks, which significantly increases the HSI labeling cost.

Recently, representation-based classification methods have attracted considerable attention due to not considering any prior distribution of data [26]. The essential idea of such classification methods is that each test sample is linearly represented and reconstructed by the available labeled samples, and then assigned to the class with the minimum reconstruction error [27,28]. Therefore, such methods can effectively avoid the complex training process and the influence of the number of labeled samples on the model fitting performance [23,29]. Sparse Representation Classifier (SRC), as one of the most representative representation-based classification methods, was initially developed for face recognition by employing $\updownarrow_0$-norm or $\updownarrow_1$-norm regularization to solve the representation coefficients [30]. Although SRC is superior to other traditional classification algorithms in classification performance, sparse norm constraint significantly increases the computational cost [19]. To solve this problem, Zhang et al. proposed a Collaborative Representation Classifier (CRC) for face recognition [31]. CRC can obtain a closed-form solution by adopting the $\updownarrow_2$-norm rather than sparse norm to solve the representation coefficients. As a consequence, CRC can achieve higher computational efficiency and better classification performance than SRC [31].

Given the superiority of the Collaborative Representation (CR) models, researchers have developed a series of improved CR-based methods and applied them to HSI classification. Li et al. introduced a distance-weighted Tikhonov regularization into two original CR models respectively, and proposed the Nearest Regularized Subspace (NRS) [32] and CR with Tikhonov regularization (CRT) [29] methods for HSI classification. The experimental results show that the Tikhonov regularization can effectively improve the classification performance of CR models. Therefore, many CR models with Tikhonov regularization have been proposed for HSI classification, such as Structure-aware CR with Tikhonov regularization (SaCRT) [33], Local Nearest Neighbor CR with Tikhonov regularization (LNNCRT) [23], Euclidean Distance-based Adaptive CR with Tikhonov regularization (EDACRT) [34], etc. However, different topographic structures and surface roughness make spectral curves of the same ground objects in the same hyperspectral scene generate amplitude shift [35], which causes the sample data to be presented in a nonlinear struc-

ture. The linear representation of the above-mentioned CR models is insufficient to represent the nonlinear structure of samples [26]. To solve this issue, researchers constructed nonlinear CR classification models based on kernel trick. Such methods can project the sample data into a nonlinear high-dimensional feature space, thereby enhancing the separability of samples in HSI, such as Kernel CRT (KCRT) [29], Discriminative Kernel CRT (DKCRT) [36], Diversity-driven multikernel CRC (DIV-KCRC) [37], and so on.

Furthermore, there is a high probability that each pixel and its spatial neighboring pixels belong to the same class in HSI. Due to this fact, Li et al. designed a spatial filter operator by averaging the neighboring pixels to mine the spatial–spectral features in HSI, and introduced this operator into the original NRS method to construct a joint within-class CR method for HSI classification [38]. The experimental results show that the spatial–spectral features mined by this spatial filter operator can effectively improve the HSI classification performance of the original NRS method. In view of this, the spatial filter operator has been successively introduced into many other CR models, such as KCRT with Composite Kernel (KCRT-CK) [29] and Joint DKCRT (JDKCRT) [36]. However, such methods also have certain drawbacks. On the one hand, there are usually different classes of pixels in the spatial neighborhood of each pixel in one hyperspectral scene. On the other hand, the hyperspectral sensor not only obtains the direct reflective power from each pixel, but also gathers the indirect diffuse reflective powers from its spatial neighboring pixels. Due to the above reasons, the spectral curve of each pixel produces spectral shift, which is called adjacency effect [39]. Therefore, the center pixels reconstructed by directly averaging the spatial neighboring pixels, i.e., spatial–spectral features, contain a lot of noise, thus affecting the classification performance of CR models. To address this problem, Yang et al. proposed a correlation coefficient-weighted spatial filtering operator to mine the spatial–spectral features, and integrated this operator into the KCRT and DKCRT methods to construct Weighted Spatial–Spectral KCRT (WSSKCRT) and Weighted Spatial–Spectral DKCRT (WSSDKCRT) methods [40]. The experimental results demonstrate that the proposed weighted spatial filtering operator can effectively alleviate the spectral shift caused by adjacency effect and significantly improve the classification performance of CR models for HSI.

From the above literatures, it can be seen that the existing CR models usually consider the spatial information of HSI by taking the average or weighted average of spatial neighboring pixels of each pixel. In essence, this is to only perform spatial smoothing filtering on all pixels of HSI without considering the spatial structure information for pixels. Yang et al. proposed a Joint CRC (JCRC) method, in which each test pixel and its spatial neighboring pixels can be simultaneously represented by the labeled training pixels to assist with classification of center test pixel in the spatial window [41]. This method takes advantage of the spatial structure information for pixels to improve the classification performance of CR model but does not carry out spatial filtering operation.

To fully mine the spatial–spectral features for HSI and further improve the land cover classification performance of CR model, a novel Weighted Spatial–Spectral Joint CRC (WSSJCRC) method is proposed in this paper. This method not only utilizes a weighted spatial filtering operator to perform spatial filtering on HSI to alleviate the spectral shift caused by adjacency effect, but also considers the spatial structure information of each test pixel. On this basis, a Weighted Spatial–Spectral Joint Kernel CRC (WSSJKCRC) method is also proposed for hyperspectral land cover classification, in which the kernel trick is introduced into the WSSJCRC method to enhance the separability of land cover objects.

2. Methodology

The proposed WSSJCRC and WSSJKCRC methods are inspired by the WSSKCRT [40] and JCRC [41] methods, so the principles of the WSSKCRT and JCRC methods are firstly introduced.

Suppose that $\mathbf{HSI} \in \mathbf{R}^{W \times H \times d}$ represents a three-dimensional hyperspectral scene with C land cover classes, where W is the width of image, H is the height of image, and d is the dimension of image (i.e., the number of hyperspectral bands). Then, the three-

dimensional hyperspectral scene **HSI** is transformed into a two-dimensional matrix, i.e., $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N] \in \mathbf{R}^{d \times N}$, where N represent the number of the labeled training samples. Additionally, $\mathbf{X}_l = [\boldsymbol{x}_{l,1}, \boldsymbol{x}_{l,2}, \cdots, \boldsymbol{x}_{l,N_l}] \in \mathbf{R}^{d \times N_l}$ represents the training set for the lth class, where $l \in \{1, 2, \cdots, C\}$ and N_l represents the number of the training samples for the lth class, i.e., $\sum_{l=1}^{C} N_l = N$. Consequently, $\mathbf{X}$ can also be expressed as $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_C]$. Moreover, the test sample is denoted as $\boldsymbol{y}$.

2.1. WSSKCRT

Firstly, a correlation coefficient-weighted spatial filtering operator is employed to perform spatial filtering on HSI to alleviate the spectral shift caused by adjacency effect. Suppose that the spatial neighboring pixel set for the pixel $\boldsymbol{x}$ under the window of $n \times n$ is $\boldsymbol{\Omega} = [\boldsymbol{x}_{0,0}, \boldsymbol{x}_{0,1}, \cdots, \boldsymbol{x}_{0,n\times n-1}]$, where $\boldsymbol{x}_{0,0}$ is the pixel $\boldsymbol{x}$ itself, called central pixel. The correlation coefficient between each pixel in $\boldsymbol{\Omega}$ and the center pixel is calculated, and the results can be expressed as $R = \{r_{0,0}, r_{0,1}, \cdots, r_{0,n\times n-1}\}$. The spatial neighboring pixel $\boldsymbol{x}_{0,i}$ (i = 0, 1, 2, . . . , $n \times n - 1$) with larger absolute value of correlation coefficient have higher probability of belonging to the same class as the central pixel $\boldsymbol{x}_{0,0}$. Therefore, the corresponding spatial neighboring pixel $\boldsymbol{x}_{0,i}$ should be provided larger weight. Moreover, the weights of spatial neighboring pixels can be expressed as

$$w_{0,i} = \frac{|r_{0.i}|}{\sum_{i=0}^{n\times n-1} |r_{0.i}|} \tag{1}$$

After obtaining the weights, the weighted average of spatial neighboring pixels can be calculated and considered as the reconstructed central pixel $\widetilde{\boldsymbol{x}}$, which is expressed as

$$\widetilde{\boldsymbol{x}} = \sum_{i=0}^{n\times n-1} w_{0,i} \boldsymbol{x}_{0,i} \tag{2}$$

By using Equations (1) and (2) to reconstruct all pixels for HSI, the training set can be expressed as $\widetilde{\mathbf{X}} = [\widetilde{\boldsymbol{x}}_1, \widetilde{\boldsymbol{x}}_2, \cdots, \widetilde{\boldsymbol{x}}_N] \in \mathbf{R}^{d \times N}$, the training set for the lth class can be expressed as $\widetilde{\mathbf{X}}_l = [\widetilde{\boldsymbol{x}}_{l,1}, \widetilde{\boldsymbol{x}}_{l,2}, \cdots, \widetilde{\boldsymbol{x}}_{l,N_l}] \in \mathbf{R}^{d \times N_l}$, and the test sample is denoted as $\widetilde{\boldsymbol{y}}$.

Then, a nonlinear mapping function Φ is utilized to map the reconstructed training samples $\widetilde{\mathbf{X}}$ and test sample $\widetilde{\boldsymbol{y}}$ to a high-dimensional feature space, i.e., $\Phi(\widetilde{\mathbf{X}}) = [\Phi(\widetilde{\boldsymbol{x}}_1), \Phi(\widetilde{\boldsymbol{x}}_2), \cdots, \Phi(\widetilde{\boldsymbol{x}}_N)] \in \mathbf{R}^{D \times N}$ and $\Phi(\widetilde{\boldsymbol{y}}) \in \mathbf{R}^{D \times 1}$, where $D >> d$ is the dimension of high-dimensional feature space. However, it is difficult to obtain the nonlinear mapping function Φ directly. When a kernel function satisfies Mercer's conditions, the inner product of any two samples mapped by nonlinear mapping function Φ can be represented by the kernel function. For WSSKCRT, the Gaussian radial basis function (RBF) is taken as the kernel function. Consequently, the mathematical relation between the nonlinear mapping function Φ and the kernel function can be formulated as

$$k(\widetilde{\boldsymbol{x}}_i, \widetilde{\boldsymbol{x}}_j) = \Phi(\widetilde{\boldsymbol{x}}_i)^T \Phi(\widetilde{\boldsymbol{x}}_j) = \exp(-\gamma \left\| \widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{x}}_j \right\|_2^2) \tag{3}$$

where γ $(\gamma > 0)$ is a parameter controlling the width of RBF and it is set as the median value of $1/(\|\widetilde{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}\|_2^2)$ (i = 1, 2, . . . , N), in which $\hat{\boldsymbol{x}} = (\sum_{i=1}^{N} \widetilde{\boldsymbol{x}}_i)/N)$ is the mean value of all available reconstructed training samples.

In the kernel-induced high-dimensional feature space, each reconstructed test sample is represented by all reconstructed training samples, and the representation coefficient vector $\boldsymbol{\alpha}$ is solved by $\updownarrow_2$-norm regularization, i.e.,

$$\boldsymbol{\alpha} = \arg \min_{\boldsymbol{\alpha}^*} \left\| \Phi\left(\widetilde{\boldsymbol{y}}\right) - \Phi(\widetilde{\mathbf{X}})\boldsymbol{\alpha}^* \right\|_2^2 + \lambda \left\| \boldsymbol{\Gamma}_{\Phi(\widetilde{\boldsymbol{y}})} \boldsymbol{\alpha}^* \right\|_2^2 \tag{4}$$

where λ is a global regularization parameter for balancing the minimization between the residual part and the regularization term, and $\mathbf{\Gamma}_{\Phi(\widetilde{y})}$ is the Tikhonov regularization matrix with the form of

$$\mathbf{\Gamma}_{\Phi(\widetilde{y})} = \begin{bmatrix} \|\Phi(\widetilde{\boldsymbol{y}}) - \Phi(\widetilde{\boldsymbol{x}}_1)\|_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|\Phi(\widetilde{\boldsymbol{y}}) - \Phi(\widetilde{\boldsymbol{x}}_N)\|_2 \end{bmatrix} \tag{5}$$

where $\|\Phi(\widetilde{\boldsymbol{y}}) - \Phi(\widetilde{\boldsymbol{x}}_i)\|_2 = [k(\widetilde{\boldsymbol{y}}, \widetilde{\boldsymbol{y}}) + k(\widetilde{\boldsymbol{x}}_i, \widetilde{\boldsymbol{x}}_i) - 2k(\widetilde{\boldsymbol{y}}, \widetilde{\boldsymbol{x}}_i)]^{1/2}$, $i = 1, 2, \ldots, N$. After that, the representation coefficient vector $\boldsymbol{\alpha}$ for Equation (4) can be obtained in a closed-form solution as follows:

$$\boldsymbol{\alpha} = \left(\mathbf{K} + \lambda \mathbf{\Gamma}_{\Phi(\widetilde{y})}^T \mathbf{\Gamma}_{\Phi(\widetilde{y})}\right)^{-1} k(\widetilde{\mathbf{X}}, \widetilde{\boldsymbol{y}}) \tag{6}$$

where $\mathbf{K} = \Phi(\widetilde{\mathbf{X}})^T \Phi(\widetilde{\mathbf{X}}) \in \mathbf{R}^{N \times N}$ is the Gram matrix formed from $\mathbf{K}_{i,j} = k(\widetilde{\boldsymbol{x}}_i, \widetilde{\boldsymbol{x}}_j)$ ($i, j = 1, 2, \ldots, N$), and $k(\widetilde{\mathbf{X}}, \widetilde{\boldsymbol{y}}) = [k(\widetilde{\boldsymbol{x}}_1, \widetilde{\boldsymbol{y}}), k(\widetilde{\boldsymbol{x}}_2, \widetilde{\boldsymbol{y}}), \cdots, k(\widetilde{\boldsymbol{x}}_N, \widetilde{\boldsymbol{y}})]^T \in \mathbf{R}^{N \times 1}$. Finally, the mapped test sample $\Phi(\widetilde{\boldsymbol{y}})$ is reconstructed by the mapped class-specific training samples $\Phi(\widetilde{\mathbf{X}}_l)$ and the corresponding class-specific representation coefficient vector $\boldsymbol{\alpha}_l$, and then assigned to the class with the minimum reconstruction error, i.e.,

$$\begin{aligned} \text{class}(\widetilde{\boldsymbol{y}}) &= \arg \min_{l=1,\cdots,C} \left\| \Phi(\widetilde{\mathbf{X}}_l)\boldsymbol{\alpha}_l - \Phi(\widetilde{\boldsymbol{y}}) \right\|_2 \\ &= \arg \min_{l=1,\cdots,C} \sqrt{k(\widetilde{\boldsymbol{y}}, \widetilde{\boldsymbol{y}}) + \boldsymbol{\alpha}_l^T \mathbf{K}_l \boldsymbol{\alpha}_l - 2\boldsymbol{\alpha}_l^T k(\widetilde{\mathbf{X}}_l, \widetilde{\boldsymbol{y}})} \end{aligned} \tag{7}$$

where $\Phi(\widetilde{\mathbf{X}}_l) = [\Phi(\widetilde{x}_{l,1}), \Phi(\widetilde{x}_{l,2}), \cdots, \Phi(\widetilde{x}_{l,N_l})]$ represents kernel sub-dictionary for the lth class, $\mathbf{K}_l = \Phi(\widetilde{\mathbf{X}}_l)^T \Phi(\widetilde{\mathbf{X}}_l) \in \mathbf{R}^{N_l \times N_l}$ denotes the Gram matrix for the lth class, and $k(\widetilde{\mathbf{X}}_l, \widetilde{\boldsymbol{y}}) = [k(\widetilde{\boldsymbol{x}}_{l,1}, \widetilde{\boldsymbol{y}}), k(\widetilde{\boldsymbol{x}}_{l,2}, \widetilde{\boldsymbol{y}}), \cdots, k(\widetilde{\boldsymbol{x}}_{l,N_l}, \widetilde{\boldsymbol{y}})]^T \in \mathbf{R}^{N_l \times 1}$.

The flowchart for the WSSKCRT method is shown in Figure 1.

2.2. JCRC

For taking the spatial structure information for pixels into account, JCRC utilizes the labeled training samples $\mathbf{X}$ to simultaneously represent each test pixel (center pixel) and its spatial neighboring pixels to assist with classification of center test pixel. Moreover, the test sample $\boldsymbol{y}$ and its spatial neighboring pixels $\boldsymbol{y}_{0,i}$ ($i = 0, 1, 2, \ldots, n \times n - 1$) under the window of $n \times n$ can be expressed as

$$\mathbf{M} = \left[\boldsymbol{y}_{0,0}, \boldsymbol{y}_{0,1}, \boldsymbol{y}_{0,2}, \cdots, \boldsymbol{y}_{0,n\times n-1}\right] = \underbrace{\mathbf{X}[\boldsymbol{\alpha}_{0,0}, \boldsymbol{\alpha}_{0,1}, \boldsymbol{\alpha}_{0,2}, \cdots, \boldsymbol{\alpha}_{0,n\times n-1}]}_{\boldsymbol{\psi}} = \mathbf{X}\boldsymbol{\psi} \tag{8}$$

where $\boldsymbol{y}_{0,0}$ is the test sample $\boldsymbol{y}$ itself, and $\boldsymbol{\psi}$ denotes the coefficient vector matrix composed of all representation coefficient vectors corresponding to the pixels under the window of $n \times n$, which can be solved by the Frobenius norm, i.e.,

$$\boldsymbol{\psi} = \arg \min_{\boldsymbol{\psi}^*} \|\mathbf{M} - \mathbf{X}\boldsymbol{\psi}^*\|_F^2 + \lambda \|\boldsymbol{\psi}^*\|_F^2 \tag{9}$$

And the solution for $\boldsymbol{\psi}$ can be analytically derived as

$$\boldsymbol{\psi} = \left(\mathbf{X}^T\mathbf{X} + \lambda \mathbf{I}\right)^{-1} \mathbf{X}^T \mathbf{M} \tag{10}$$

where $\mathbf{I}$ represents the identity matrix. Similarly, the test sample $\boldsymbol{y}$ and its spatial neighboring pixels are reconstructed by the class-specific training samples $\mathbf{X}_l$ and the corresponding

class-specific coefficient vector matrix $\boldsymbol{\psi}_l$, and the test sample $\boldsymbol{y}$ is assigned to the class with the minimum reconstruction error, i.e.,

$$\text{class}(\boldsymbol{y}) = \arg \min_{l=1,\cdots,C} \|\mathbf{M} - \mathbf{X}_l \boldsymbol{\psi}_l\|_F \tag{11}$$

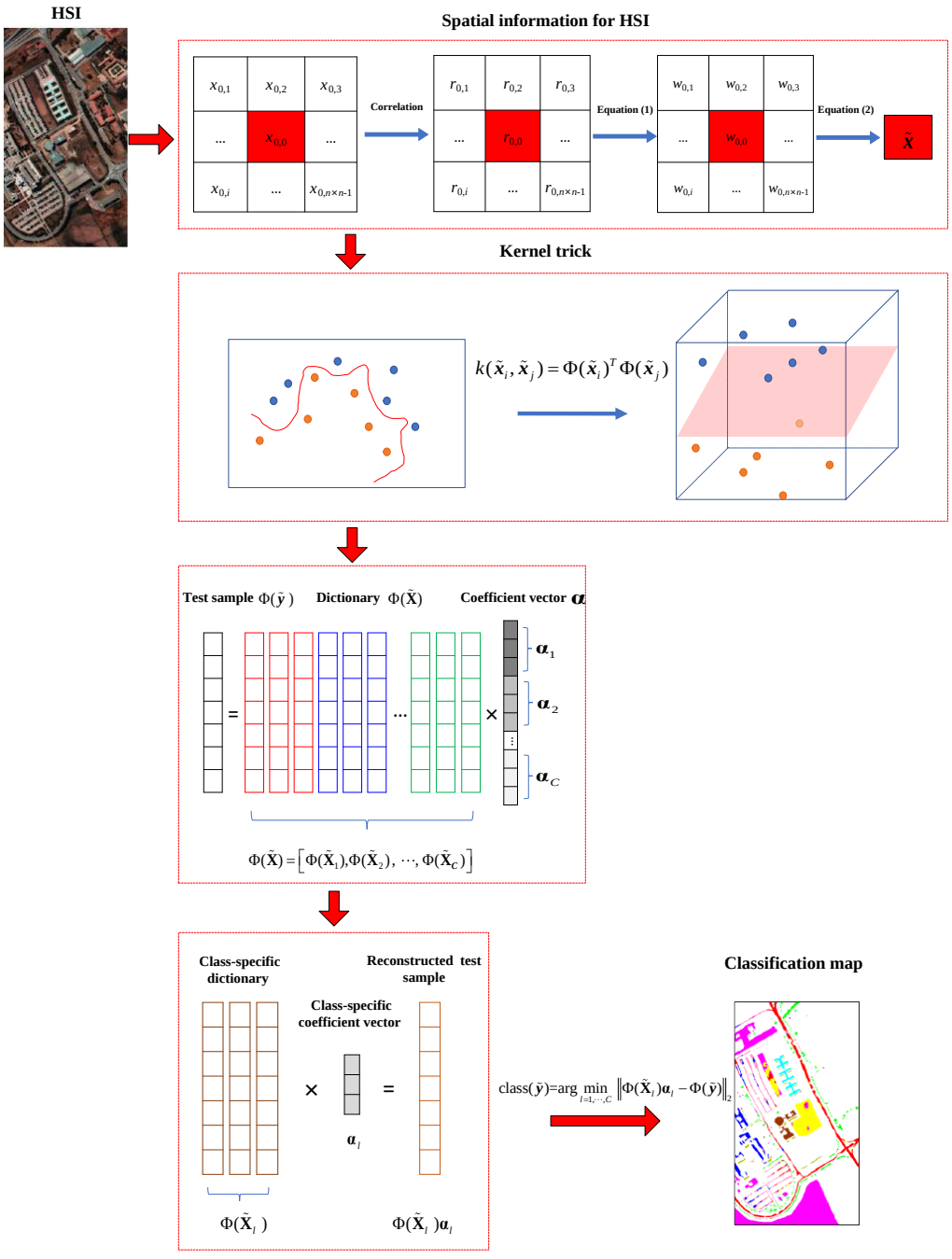


Figure 1. The flowchart for the WSSKCRT method.

The flowchart for the JCRC method is shown in Figure 2.

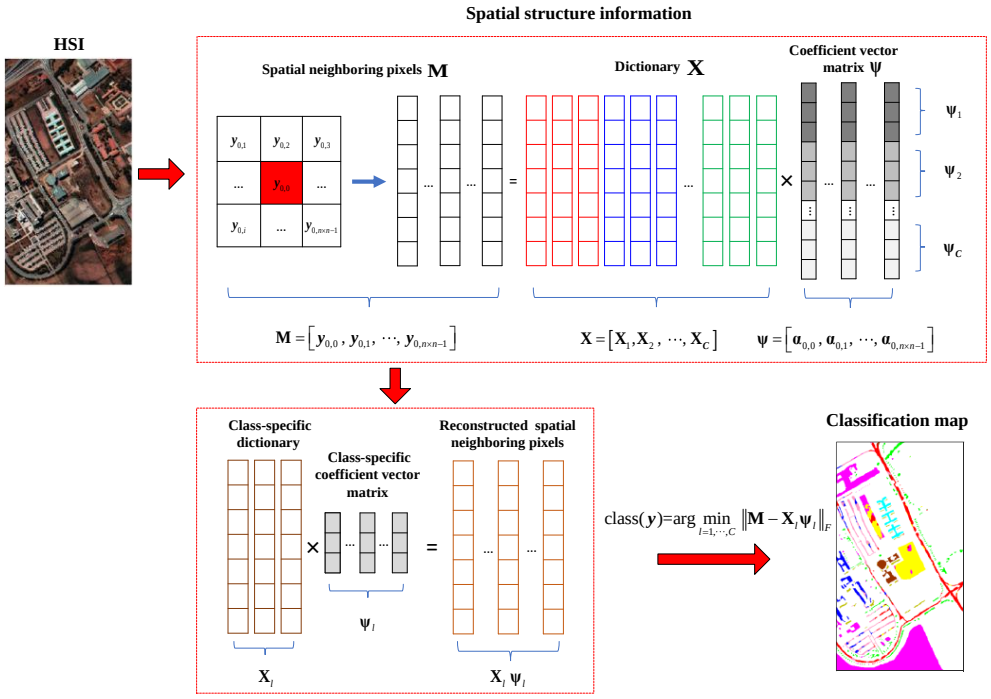


Figure 2. The flowchart for the JCRC method.

2.3. Proposed WSSJCRC

To fully mine the spatial–spectral features for HSI to assist with land cover classification, the proposed WSSJCRC method makes use of the superiorities of both WSSKCRT and JCRC methods. Firstly, as the same with WSSKCRT, the correlation coefficient-weighted spatial filtering operator constructed by Equation (1) and Equation (2) is utilized to perform spatial filtering on all pixels of HSI to alleviate the spectral shift caused by adjacency effect, so the training set $\mathbf{X}$ and the test sample $\boldsymbol{y}$ can be expressed as $\widetilde{\mathbf{X}} = [\widetilde{\boldsymbol{x}}_1, \widetilde{\boldsymbol{x}}_2, \cdots, \widetilde{\boldsymbol{x}}_N] \in \mathbf{R}^{d\times N}$ and $\widetilde{\boldsymbol{y}} \in \mathbf{R}^{d\times 1}$, respectively.

Then, through the same method as JCRC, to take the spatial structure information for pixels into account the training set $\widetilde{\mathbf{X}}$ is used to simultaneously represent the test sample $\widetilde{\boldsymbol{y}}$ and its corresponding spatial neighboring pixels under the window of $n \times n$, which can be written as

$$\widetilde{\mathbf{M}} = \left[\widetilde{\boldsymbol{y}}_{0,0}\,,\ \widetilde{\boldsymbol{y}}_{0,\,1},\ \widetilde{\boldsymbol{y}}_{0,2},\ \cdots,\ \widetilde{\boldsymbol{y}}_{0,n\times n-1}\right] = \underbrace{\widetilde{\mathbf{X}}\left[\boldsymbol{\alpha}_{0,0}\,,\ \boldsymbol{\alpha}_{0,1},\ \boldsymbol{\alpha}_{0,2},\ \cdots,\ \boldsymbol{\alpha}_{0,n\times n-1}\right]}_{\boldsymbol{\psi}} = \widetilde{\mathbf{X}}\boldsymbol{\psi} \quad (12)$$

where $\widetilde{\boldsymbol{y}}_{0,0}$ represents the test sample $\widetilde{\boldsymbol{y}}$ itself. Moreover, the coefficient vector matrix $\boldsymbol{\psi}$ is solved by the Frobenius norm, i.e.,

$$\boldsymbol{\psi} = \arg\ \min_{\boldsymbol{\psi}^*} \left\|\widetilde{\mathbf{M}} - \widetilde{\mathbf{X}}\boldsymbol{\psi}^*\right\|_F^2 + \lambda\|\boldsymbol{\psi}^*\|_F^2 \quad (13)$$

The solution for $\boldsymbol{\psi}$ can be calculated with a closed-form solution as follows:

$$\boldsymbol{\psi} = \left(\widetilde{\mathbf{X}}^T\widetilde{\mathbf{X}} + \lambda\mathbf{I}\right)^{-1}\widetilde{\mathbf{X}}^T\widetilde{\mathbf{M}} \quad (14)$$

Finally, the test sample $\widetilde{\boldsymbol{y}}$ is assigned to the class with the minimum reconstruction error, i.e.,

$$\mathrm{class}(\widetilde{\boldsymbol{y}}) = \arg\ \min_{l=1,\cdots,C} \left\|\widetilde{\mathbf{M}} - \widetilde{\mathbf{X}}_l\boldsymbol{\psi}_l\right\|_F \quad (15)$$

where $\widetilde{\mathbf{X}}_l = [\widetilde{\boldsymbol{x}}_{l,1}, \widetilde{\boldsymbol{x}}_{l,2}, \cdots, \widetilde{\boldsymbol{x}}_{l,N_l}] \in \mathbf{R}^{d\times N_l}$ represents the training set for the lth class after spatial filtering, and $\boldsymbol{\psi}_l$ denotes the corresponding coefficient vector matrix for the lth class.

The flowchart for the proposed WSSJCRC method is shown in Figure 3.

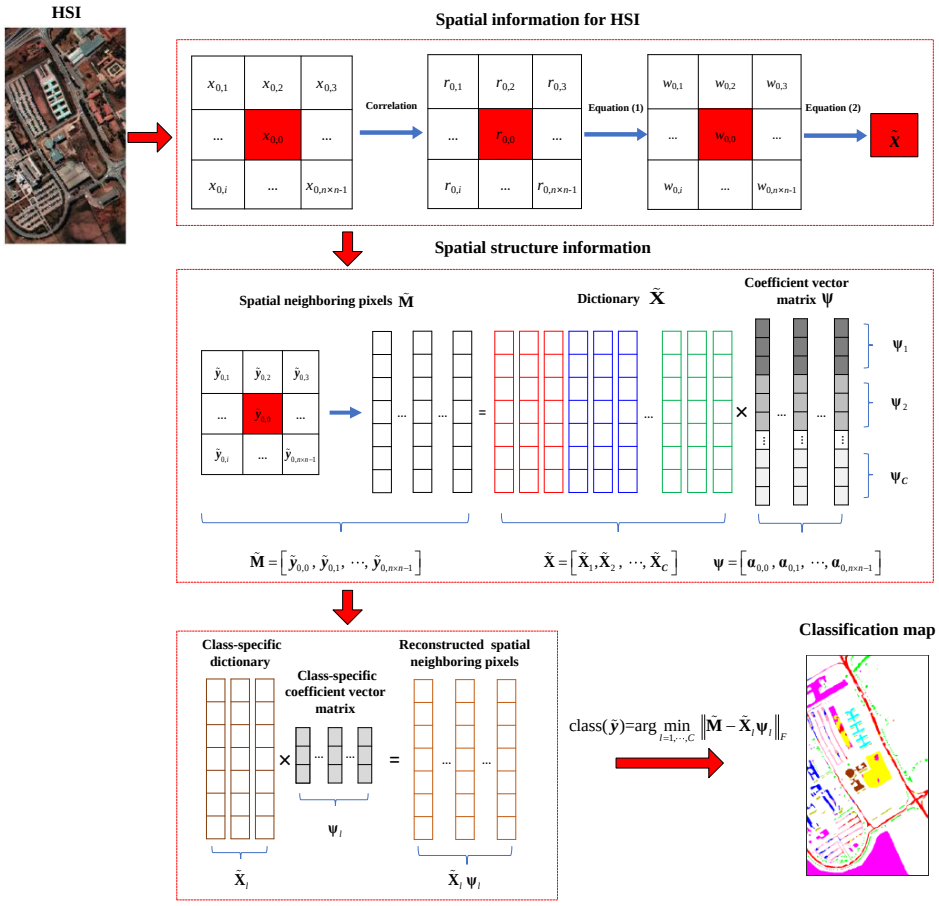


Figure 3. The flowchart for the proposed WSSJCRC method.

2.4. Proposed WSSJKCRC

As mentioned in the Introduction, the hyperspectral data usually present in a nonlinear structure due to different topographic structures and surface roughness [35]. To enhance the separability of nonlinear hyperspectral data, the kernel version of WSSJCRC (i.e., WSSJKCRC) is proposed in this paper to project the hyperspectral data into the kernel-induced high-dimensional feature space, in which the coefficient vector matrix $\boldsymbol{\psi}$ can be solved by the optimization problem as follows:

$$\boldsymbol{\psi} = \arg\min_{\boldsymbol{\psi}^*} \left\| \Phi(\tilde{\mathbf{M}}) - \Phi(\tilde{\mathbf{X}})\boldsymbol{\psi}^* \right\|_F^2 + \lambda \|\boldsymbol{\psi}^*\|_F^2 \tag{16}$$

And the closed-form solution for $\boldsymbol{\psi}$ can be denoted as

$$\boldsymbol{\psi} = (\mathbf{K}_{(\tilde{\mathbf{X}})} + \lambda \mathbf{I})^{-1} \mathbf{K}_{(\tilde{\mathbf{X}},\tilde{\mathbf{M}})} \tag{17}$$

where $\mathbf{K}_{(\tilde{\mathbf{X}})} = \Phi(\tilde{\mathbf{X}})^T \Phi(\tilde{\mathbf{X}}) \in \mathbf{R}^{N\times N}$ is the Gram matrix formed from $\mathbf{K}_{(\tilde{\mathbf{X}})i,j} = k(\tilde{x}_i, \tilde{x}_j)$ (i, j = 1, 2, . . . , N), and $\mathbf{K}_{(\tilde{\mathbf{X}},\tilde{\mathbf{M}})} = \Phi(\tilde{\mathbf{X}})^T \Phi(\tilde{\mathbf{M}}) \in \mathbf{R}^{N\times N'}$ represents the matrix composed of $\mathbf{K}_{(\tilde{\mathbf{X}},\tilde{\mathbf{M}})i,j} = k(\tilde{\boldsymbol{x}}_i, \tilde{\boldsymbol{y}}_{0,j})$, in which i = 1, 2, . . . , N, j = 0, 1, 2, . . . , $N' - 1$, and N' represents the number of spatial neighboring pixels under the window of $n \times n$, i.e., $N' = n \times n$. Finally, the mapped spatial neighboring pixels $\Phi(\tilde{\mathbf{M}})$ for the test sample $\tilde{\boldsymbol{y}}$ are reconstructed by the mapped class-specific training samples $\Phi(\tilde{\mathbf{X}}_l)$ and the corresponding class-specific

coefficient vector matrix $\boldsymbol{\psi}_l$, and then the test sample $\widetilde{y}$ is assigned to the class with the minimum reconstruction error, i.e.,

$$\begin{aligned} \text{class}(\widetilde{y}) &= \arg \min_{l=1,\cdots,C} \left\| \Phi(\widetilde{\mathbf{X}}_l)\boldsymbol{\psi}_l - \Phi(\widetilde{\mathbf{M}}) \right\|_F \\ &= \arg \min_{l=1,\cdots,C} \sqrt{tr\left(\left(\Phi(\widetilde{\mathbf{M}}) - \Phi(\widetilde{\mathbf{X}}_l)\boldsymbol{\psi}_l \right)^T \left(\Phi(\widetilde{\mathbf{M}}) - \Phi(\widetilde{\mathbf{X}}_l)\boldsymbol{\psi}_l \right) \right)} \\ &= \arg \min_{l=1,\cdots,C} \sqrt{tr\left(\mathbf{K}_{(\widetilde{\mathbf{M}})} + \boldsymbol{\psi}_l^T \mathbf{K}_{(\widetilde{\mathbf{X}}_l)} \boldsymbol{\psi}_l - 2\boldsymbol{\psi}_l^T \mathbf{K}_{(\widetilde{\mathbf{X}}_l, \widetilde{\mathbf{M}})} \right)} \end{aligned} \tag{18}$$

where the symbol *tr* means to find the trace of square matrix, i.e., the sum of all the main diagonal elements for a square matrix, and $\mathbf{K}_{(\widetilde{\mathbf{X}}_l)} = \Phi(\widetilde{\mathbf{X}}_l)^T \Phi(\widetilde{\mathbf{X}}_l) \in \mathbf{R}^{N_l \times N_l}$ denotes the Gram matrix for the *l*th class, $\mathbf{K}_{(\widetilde{\mathbf{M}})} = \Phi(\widetilde{\mathbf{M}})^T \Phi(\widetilde{\mathbf{M}}) \in \mathbf{R}^{N' \times N'}$ and $\mathbf{K}_{(\widetilde{\mathbf{X}}_l, \widetilde{\mathbf{M}})} = \Phi(\widetilde{\mathbf{X}}_l)^T \Phi(\widetilde{\mathbf{M}}) \in \mathbf{R}^{N_l \times N'}$.

The flowchart for the proposed WSSJKCRC method is shown in Figure 4.

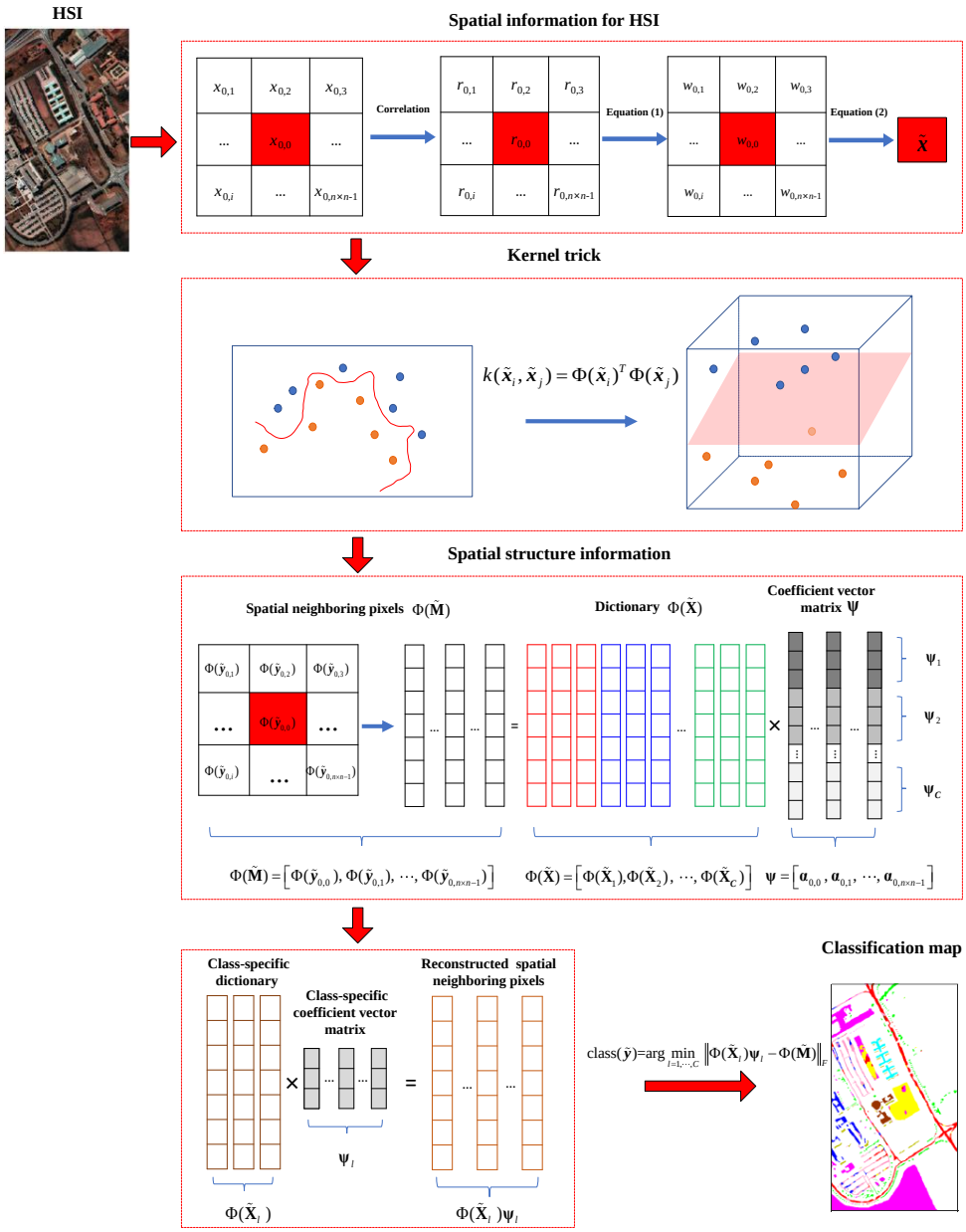


Figure 4. The flowchart for the proposed WSSJKCRC method.

3. Experimental Results and Analysis

3.1. Experimental Data

In this paper, three real hyperspectral scenes are employed to evaluate the performance of the proposed methods.

The first land cover hyperspectral scene is Indian Pines. This hyperspectral scene was gathered by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines test site in north-western Indiana. The scene image contains 145 × 145 pixels under a spatial resolution of 20 m, and provides 220 spectral bands in the range of 0.4–2.5 µm. After removing the bands covering the region of water absorption, 200 spectral bands are used for classification. The original ground truth map consists of 16 land cover classes. After eliminating some classes with very few samples, nine land cover classes are selected for the experiment in this paper. The false-color image and the ground truth map are shown in Figure 5a,b, respectively.

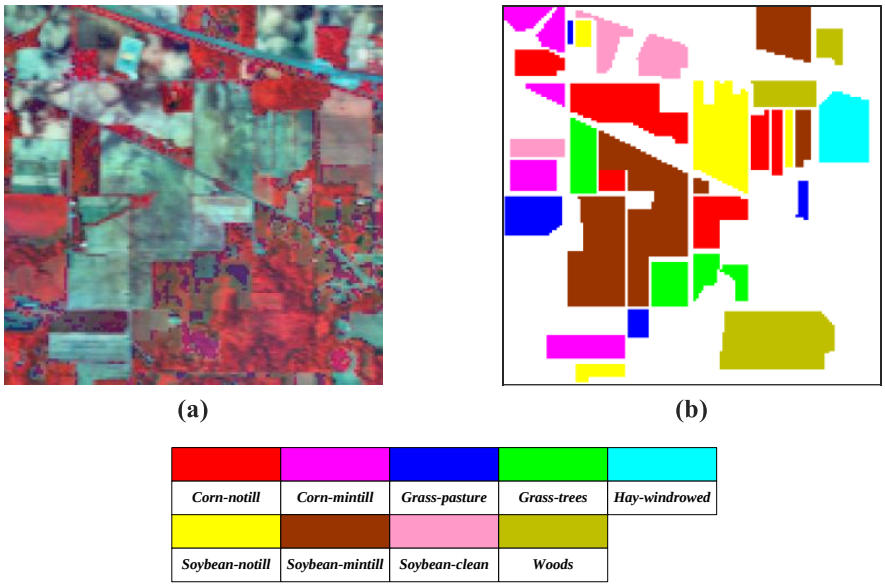


Figure 5. The Indian Pines scene: (**a**) false-color image and (**b**) ground truth map.

The second land cover hyperspectral scene is Pavia University. This scene was collected by Reflective Optics Spectrographic Imaging System (ROSIS) sensor in the University of Pavia, Pavia, Italy. Figure 6a shows the false-color image for the scene, which contains 610 × 340 pixels with a high spatial resolution of 1.3 m and consists of 103 spectral bands in the range of 0.43–0.86 µm. Moreover, there are nine land cover classes in the ground truth map shown in Figure 6b.

The last land cover hyperspectral scene is Salinas. This land cover scene was acquired by the AVIRIS sensor over Salinas Valley, California. The scene image contains 512 × 217 pixels with a spatial resolution of 3.7 m and provides 204 spectral bands after discarding the 20 water absorption bands. Its false-color image is shown in Figure 7a, and its ground truth map is shown in Figure 7b composed of 16 land cover classes.

In the ground truth maps as shown in Figures 5b, 6b and 7b, different colors represent different land cover types, in which the white areas represent unlabeled areas. It should be noted that it is time-consuming and expensive to label the pixels of HSI. Therefore, only major land cover types in the images are labeled, while non-major land types in the white areas are not labeled. The white areas are regarded as the background areas that are not considered as the classification targets. It should be emphasized that the white areas in Figure 5b contain not only unlabeled land types, but also some land types with very few samples. Moreover, each pixel in HSI is regarded as a sample.

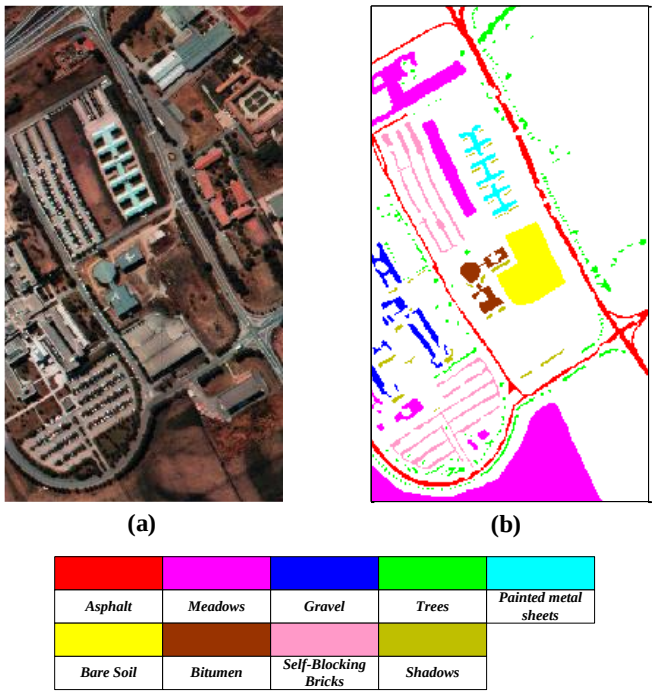


Figure 6. The Pavia University scene: (**a**) false-color image and (**b**) ground truth map.

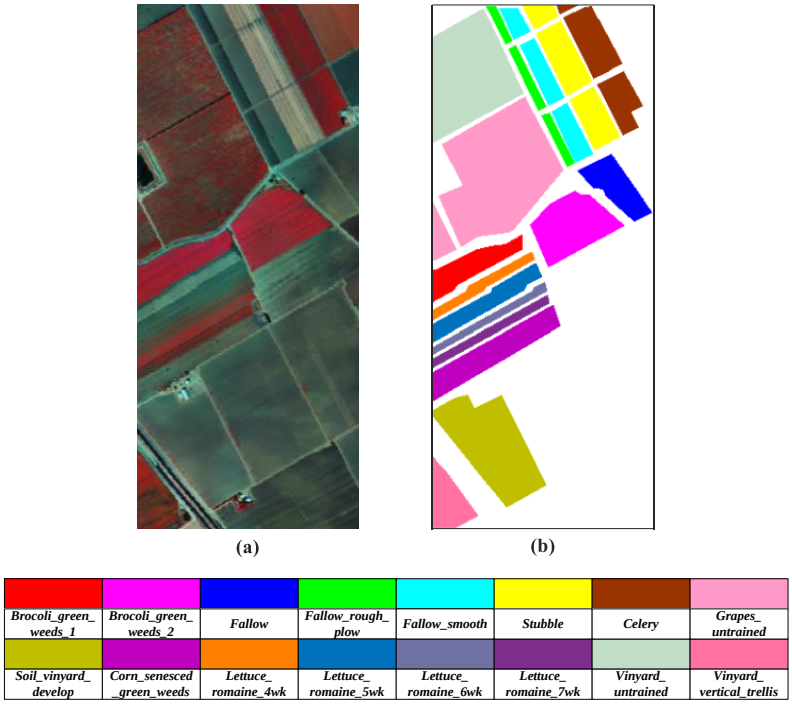


Figure 7. The Salinas scene: (**a**) false-color image and (**b**) ground truth map.

Given the difficulty of labeling samples in HSI, this paper attempts to use small-scale labeled training samples to classify large-scale unlabeled test samples. For this reason, 5% of the labeled samples for each class are randomly selected as training samples in the Indian Pines scene, 60 labeled samples in each class are randomly picked as training samples in the Pavia University scene, and 20 labeled samples in each class are randomly chosen as training samples in the Salinas scene. The remaining samples are used as test samples in the corresponding hyperspectral scene. Table 1 shows the division of samples for three hyperspectral scenes.

Table 1. The division of samples for the Indian Pines, Pavia University, and Salinas scenes.

Indian Pines				Pavia University				Salinas			
No.	**Class**	**Training Samples**	**Test Samples**	**No.**	**Class**	**Training Samples**	**Test Samples**	**No.**	**Class**	**Training Samples**	**Test Samples**
1	*Corn-notill*	72	1356	1	*Asphalt*	60	6571	1	*Brocoli_green_weeds_1*	20	1949
2	*Corn-mintill*	42	788	2	*Meadows*	60	18,589	2	*Brocoli_green_weeds_2*	20	3666
3	*Grass-pasture*	25	458	3	*Gravel*	60	2039	3	*Fallow*	20	1916
4	*Grass-trees*	37	693	4	*Trees*	60	3004	4	*Fallow_rough_plow*	20	1334
5	*Hay-windrowed*	24	454	5	*Painted metal sheets*	60	1285	5	*Fallow_smooth*	20	2618
6	*Soybean-notill*	49	923	6	*Bare Soil*	60	4969	6	*Stubble*	20	3899
7	*Soybean-mintill*	123	2332	7	*Bitumen*	60	1270	7	*Celery*	20	3519
8	*Soybean-clean*	30	563	8	*Self-Blocking Bricks*	60	3622	8	*Grapes_untrained*	20	11,211
9	*Woods*	64	1201	9	*Shadows*	60	887	9	*Soil_vinyard_develop*	20	6143
								10	*Corn_senesced_green_weeds*	20	3218
								11	*Lettuce_romaine_4wk*	20	1008
								12	*Lettuce_romaine_5wk*	20	1867
								13	*Lettuce_romaine_6wk*	20	856
								14	*Lettuce_romaine_7wk*	20	1010
								15	*Vinyard_untrained*	20	7208
								16	*Vinyard_vertical_trellis*	20	1747
	All class	466	8768		All class	540	42,236		All class	320	53,169

3.2. Data Preprocessing

In practical hyperspectral scenes, different topographic structures and surface roughness lead to the amplitude shift, that is the shape and trend of spectral curves of the same ground object is nearly invariant but the reflectance level (i.e., spectral amplitude) varies considerably. The amplitude shift may produce an impact on the performance of classification models [35]. For simplicity, two kinds of ground objects are selected separately from the above-mentioned three hyperspectral scenes to illustrate the amplitude shift phenomenon, as shown in Figure 8a–f, i.e., (a) *Hay-windrowed* and (b) *Woods* for Indian Pines, (c) *Painted metal sheets* and (d) *Self-Blocking Bricks* for Pavia University, and (e) *Fallow* and (f) *Stubble* for Salinas. To alleviate such spectral shift, the above-mentioned three hyperspectral scenes are pretreated by the Amplitude Normalization (AN) method proposed in the reference [35]. After pretreatment with AN method, the amplitude difference is significantly reduced, that is the spectral curves for the same ground object become compact and concentrated, as shown in Figure 8g–l. Therefore, the pretreated data are used for subsequent modeling analysis to improve the performance of classification models.

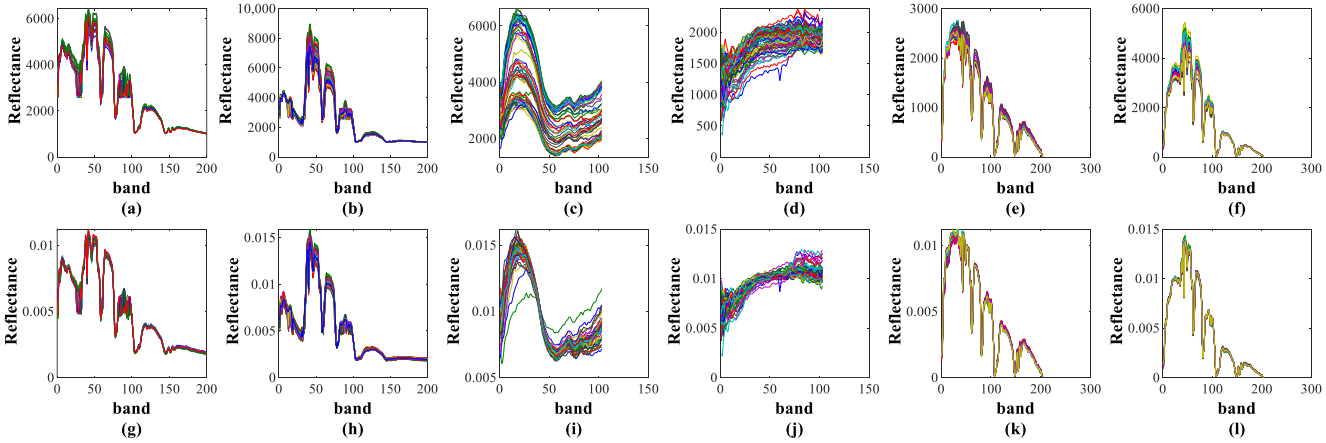


Figure 8. Original spectral curves: (**a**) *Hay-windrowed* and (**b**) *Woods* for Indian Pines, (**c**) *Painted metal sheets* and (**d**) *Self-Blocking Bricks* for Pavia University, and (**e**) *Fallow* and (**f**) *Stubble* for Salinas; and pretreated spectral curves with AN: (**g**) *Hay-windrowed* and (**h**) *Woods* for Indian Pines, (**i**) *Painted metal sheets* and (**j**) *Self-Blocking Bricks* for Pavia University, and (**k**) *Fallow* and (**l**) *Stubble* for Salinas.

3.3. Parameter Optimization

To fully verify the superiority of the proposed methods, the proposed WSSJCRC and WSSJKCRC methods are compared with the CRC, JCRC, CRC-M, JDKCRT, KCRT-CK, WSSDKCRT, and WSSKCRT methods on three real hyperspectral scenes for land cover classification performance. It should be noted that the essence of CRC-M is to introduce the directly averaging spatial filter operator into the CRC method, and other methods are all mentioned in the Introduction. To ensure the fairness of the experiment, the parameters of all methods are optimized, so that the classification performance of all methods can be compared under their respective optimal parameters.

3.3.1. Parameter Optimization of the Compared Methods

For the above-mentioned seven compared methods, namely CRC, JCRC, CRC-M, JDKCRT, KCRT-CK, WSSDKCRT, and WSSKCRT, there are four main parameters that affect the classification performance, i.e., global regularization parameter λ, positive regularization parameter β, spatial filtering window size W_f, and spatial structure window size W_s. According to the experience of relevant references in the Introduction, the optimization intervals of λ and β are set as $\{10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$, and the optimization intervals of W_f and W_s are set as $\{3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9, 11 \times 11, 13 \times 13, 15 \times 15, 17 \times 17, 19 \times 19, 21 \times 21\}$. It should be noted that only the optimization interval of λ for CRC-M on the Salinas scene is set as $\{10^{-15}, 10^{-14}, 10^{-13}, 10^{-12}, 10^{-11}, 10^{-10}, 10^{-9},$

10^{-8}, 10^{-7}, 10^{-6}}. The corresponding parameters of each method are optimized by using training samples and five-fold cross-validation strategy, and the Overall Accuracy (OA) is used to evaluate the classification performance of each method under each parameter. The optimal parameter settings for the compared methods on the three hyperspectral scenes are shown in Table 2, where NA means that the corresponding parameter is not applied.

Table 2. Optimal parameter settings for the compared methods on the Indian Pines, Pavia University, and Salinas scenes.

Method	Indian Pines				Pavia University				Salinas			
	λ	β	W_f	W_s	λ	β	W_f	W_s	λ	β	W_f	W_s
CRC	10^{-6}	NA	NA	NA	10^{-5}	NA	NA	NA	10^{-7}	NA	NA	NA
JCRC	10^{-7}	NA	NA	7×7	10^{-7}	NA	NA	5×5	10^{-8}	NA	NA	3×3
CRC-M	10^{-8}	NA	13×13	NA	10^{-8}	NA	13×13	NA	10^{-11}	NA	11×11	NA
JDKCRT	10^{-3}	10^{-2}	19×19	NA	10^{-3}	10^{-4}	5×5	NA	10^{-4}	10^{-4}	9×9	NA
KCRT-CK	10^{-5}	NA	15×15	NA	10^{-2}	NA	5×5	NA	10^{-3}	NA	9×9	NA
WSSDKCRT	10^{-4}	10^{-3}	19×19	NA	10^{-3}	10^{-4}	7×7	NA	10^{-3}	10^{-4}	5×5	NA
WSSKCRT	10^{-4}	NA	19×19	NA	10^{-2}	NA	9×9	NA	10^{-8}	NA	9×9	NA

3.3.2. Parameter Optimization of the Proposed Methods

The proposed WSSJCRC and WSSJKCRC methods simultaneously include three main parameters that affect the classification performance, i.e., global regularization parameter λ, spatial filtering window size W_f, and spatial structure window size W_s. The optimization strategy and the optimization intervals of parameters are the same as those of the above-mentioned seven methods. Figure 9 shows the optimization process of the proposed WSSJCRC and WSSJKCRC methods on the Indian Pines, Pavia University, and Salinas scenes. In the three-dimensional graph as shown in Figure 9, the surface of different colors is used to represent the corresponding global regularization parameter λ, and the asterisk (*) is utilized to represent the position of the optimal parameters. The optimal parameter settings for the proposed methods on the three hyperspectral scenes are shown in Table 3.

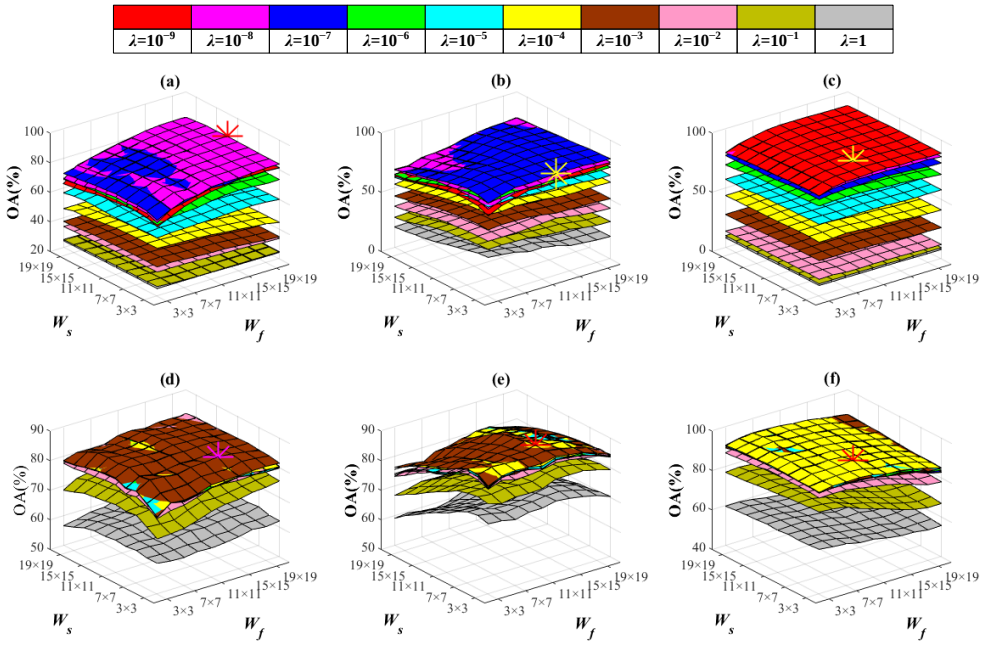


Figure 9. Classification performance under different parameters: (**a**) WSSJCRC and (**d**) WSSJKCRC on the Indian Pines scene, (**b**) WSSJCRC and (**e**) WSSJKCRC on the Pavia University scene, and (**c**) WSSJCRC and (**f**) WSSJKCRC on the Salinas scene.

Table 3. Optimal parameter settings for the proposed WSSJCRC and WSSJKCRC methods on the Indian Pines, Pavia University, and Salinas scenes.

Method	Indian Pines			Pavia University			Salinas		
	λ	W_f	W_s	λ	W_f	W_s	λ	W_f	W_s
WSSJCRC	10^{-8}	21×21	13×13	10^{-7}	13×13	3×3	10^{-9}	11×11	7×7
WSSJKCRC	10^{-3}	15×15	7×7	10^{-4}	13×13	7×7	10^{-4}	11×11	7×7

3.4. Classification Results and Discussion

3.4.1. Analysis of Classification Performance

After parameter optimization, the above nine CR methods perform land cover classification with their respective optimal parameters on the Indian Pines, Pavia University, and Salinas scenes. Moreover, four evaluation indicators are utilized to evaluate the performance of each method for land cover classification, i.e., individual class accuracy, Overall Accuracy (OA), Average Accuracy (AA), and Kappa statistic (Kappa). Moreover, each method is performed repeatedly ten times and takes the average value as the final result to avoid random error and any bias. During each running process, training samples and test samples are randomly selected according to the proportion and number of samples set in Table 1. The land cover classification results of different methods are shown in Table 4, Figures 10 and 11 for the Indian Pines scene, Table 5, Figures 12 and 13 for the Pavia University scene, and Table 6, Figures 14 and 15 for the Salinas scene, respectively. Among these figures, the curves of OA, AA, and Kappa values for each method on the three hyperspectral scenes are shown in Figures 10, 12 and 14. Moreover, the land cover classification maps generated by different methods on the three hyperspectral scenes are shown in Figures 11, 13 and 15, respectively.

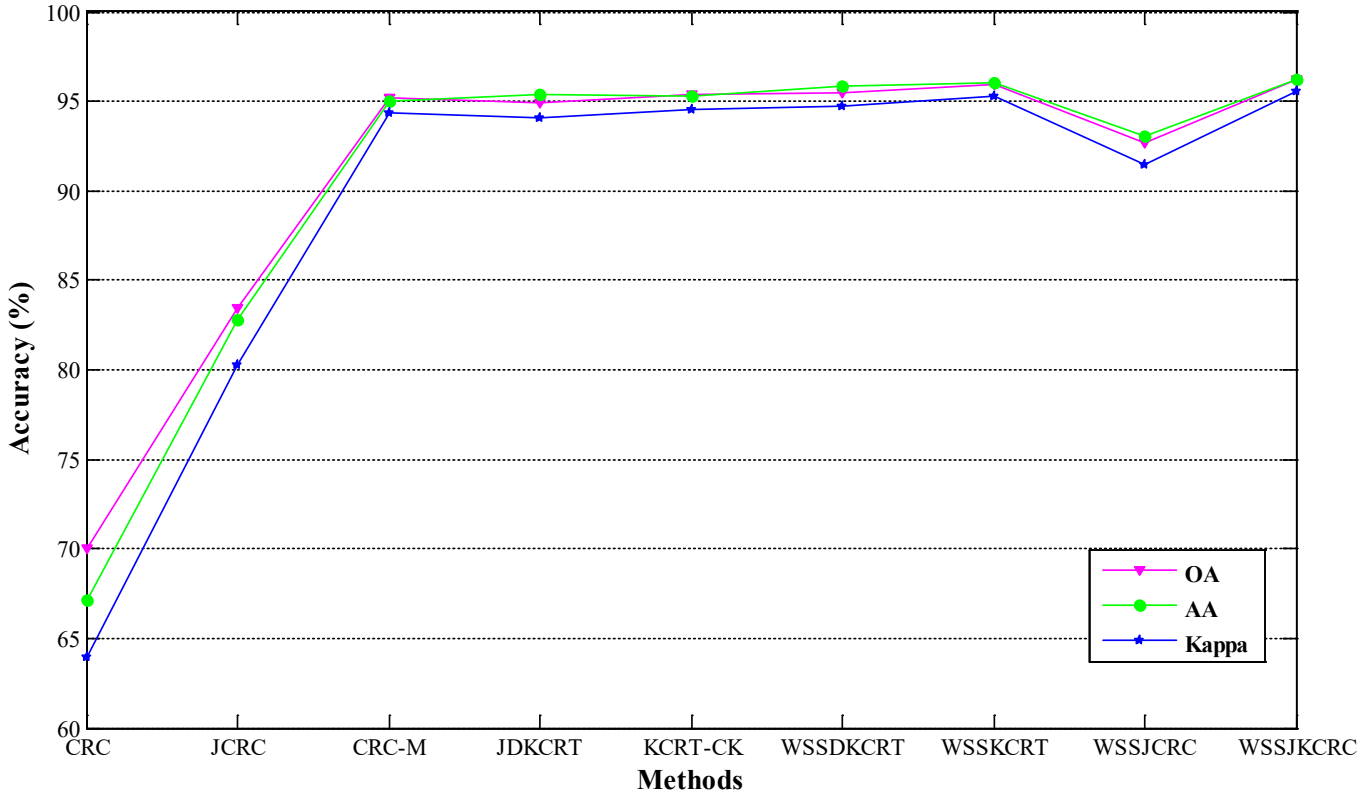


Figure 10. The OA, AA, and Kappa values for different methods on the Indian Pines scene.

Table 4. Land cover classification accuracy with STD for different methods on the Indian Pines scene.

Class	CRC	JCRC	CRC-M	JDKCRT	KCRT-CK	WSSDKCRT	WSSKCRT	WSSJCRC	WSSJKCRC
1	64.76 ± 4.26	83.33 ± 4.65	89.73 ± 2.24	89.59 ± 2.16	92.88 ± 1.41	91.57 ± 2.17	**93.25** ± 3.16	85.96 ± 3.90	92.43 ± 1.98
2	32.60 ± 6.43	60.36 ± 3.33	94.24 ± 2.59	94.71 ± 1.77	94.09 ± 2.98	94.49 ± 4.42	**95.66** ± 2.86	92.54 ± 2.63	94.51 ± 1.57
3	76.92 ± 5.27	88.47 ± 2.81	92.29 ± 4.66	95.87 ± 3.05	94.17 ± 4.50	96.83 ± 2.06	**97.88** ± 2.43	93.08 ± 5.74	95.44 ± 3.76
4	91.24 ± 2.36	98.56 ± 0.86	**98.82** ± 1.02	97.73 ± 0.83	97.59 ± 1.45	97.60 ± 1.69	96.22 ± 2.31	97.53 ± 1.12	98.25 ± 1.04
5	98.37 ± 1.03	**100.00** ± 0.00	99.91 ± 0.26	99.71 ± 0.36	98.96 ± 1.40	99.49 ± 0.69	99.01 ± 1.27	99.96 ± 0.13	99.98 ± 0.07
6	42.05 ± 5.23	62.95 ± 6.19	90.10 ± 2.61	92.09 ± 2.41	**93.43** ± 2.57	93.10 ± 2.54	93.38 ± 2.41	87.28 ± 2.26	92.29 ± 2.56
7	81.93 ± 1.82	86.42 ± 1.84	97.36 ± 0.76	95.19 ± 1.49	96.24 ± 1.35	95.81 ± 1.28	96.83 ± 1.63	93.66 ± 2.19	**97.94** ± 1.20
8	18.76 ± 3.63	65.56 ± 6.54	93.21 ± 3.62	94.33 ± 3.16	92.38 ± 4.05	94.03 ± 3.24	92.65 ± 6.42	88.56 ± 4.07	**95.74** ± 1.95
9	97.31 ± 0.67	**99.92** ± 0.07	99.74 ± 0.47	99.15 ± 0.76	98.23 ± 0.94	99.41 ± 0.42	99.13 ± 0.94	99.06 ± 0.76	99.20 ± 0.46
OA (%)	70.02 ± 0.66	83.41 ± 0.77	95.18 ± 0.58	94.91 ± 0.56	95.40 ± 0.52	95.51 ± 0.64	95.98 ± 0.92	92.71 ± 0.90	**96.21** ± 0.44
AA (%)	67.10 ± 0.80	82.84 ± 0.88	95.04 ± 0.83	95.38 ± 0.44	95.33 ± 0.67	95.82 ± 0.66	96.00 ± 1.14	93.07 ± 1.05	**96.20** ± 0.52
Kappa	0.6395 ± 0.0086	0.8031 ± 0.0092	0.9433 ± 0.0069	0.9404 ± 0.0065	0.9459 ± 0.0060	0.9474 ± 0.0074	0.9527 ± 0.0108	0.9145 ± 0.0106	**0.9555** ± 0.0052

Note: the best results are presented in bold font.

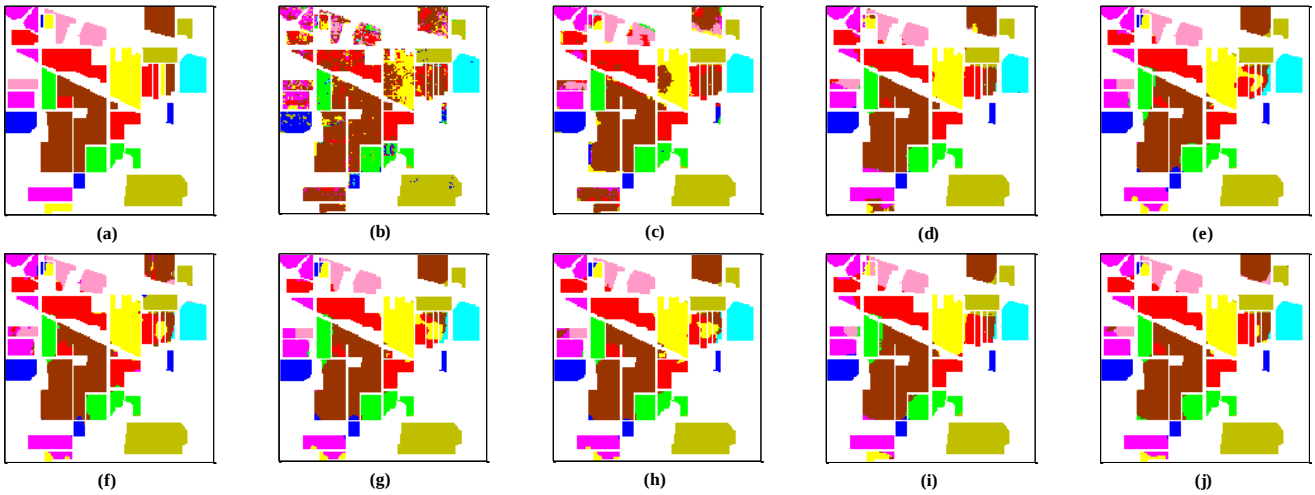


Figure 11. Land cover classification maps generated by different methods on the Indian Pines scene: (**a**) ground truth, (**b**) CRC, (**c**) JCRC, (**d**) CRC-M, (**e**) JDKCRT, (**f**) KCRT-CK, (**g**) WSSDKCRT, (**h**) WSSKCRT, (**i**) WSSJCRC, and (**j**) WSSJKCRC.

Table 5. Land cover classification accuracy with STD for different methods on the Pavia University scene.

Class	CRC	JCRC	CRC-M	JDKCRT	KCRT-CK	WSSDKCRT	WSSKCRT	WSSJCRC	WSSJKCRC
1	26.08 ± 6.00	41.70 ± 8.62	72.75 ± 4.04	92.15 ± 1.57	91.62 ± 2.39	93.68 ± 1.81	91.41 ± 2.63	72.42 ± 2.80	**94.03** ± 2.23
2	78.54 ± 3.75	82.10 ± 3.43	93.85 ± 3.21	95.76 ± 1.29	93.77 ± 2.84	97.34 ± 1.48	96.57 ± 1.44	91.42 ± 2.93	**98.20** ± 1.15
3	91.44 ± 1.39	95.52 ± 2.18	93.29 ± 2.94	95.27 ± 1.35	87.68 ± 1.55	**96.58** ± 1.29	90.30 ± 2.41	96.38 ± 1.22	95.23 ± 3.00
4	95.27 ± 1.74	96.69 ± 1.06	91.23 ± 1.58	96.77 ± 1.44	96.32 ± 1.23	96.82 ± 0.87	**97.15** ± 0.88	92.71 ± 1.27	95.32 ± 1.24
5	99.89 ± 0.07	**100.00** ± 0.00	99.84 ± 0.06	**100.00** ± 0.00	99.98 ± 0.03	99.77 ± 0.13	99.67 ± 0.07	**100.00** ± 0.00	99.98 ± 0.04
6	52.76 ± 6.60	69.48 ± 3.53	98.25 ± 1.52	94.90 ± 1.90	93.25 ± 1.34	96.88 ± 0.94	95.95 ± 1.38	98.73 ± 0.87	**99.29** ± 0.96
7	88.38 ± 3.22	98.17 ± 0.47	99.75 ± 0.26	98.77 ± 0.49	96.83 ± 1.22	99.15 ± 0.53	97.70 ± 0.77	99.96 ± 0.12	**99.98** ± 0.05
8	18.39 ± 3.95	22.73 ± 5.69	77.63 ± 3.06	67.10 ± 5.28	88.80 ± 2.21	71.48 ± 4.94	91.22 ± 1.56	73.96 ± 3.77	**94.50** ± 3.74
9	90.20 ± 2.07	98.56 ± 0.77	90.86 ± 1.66	**99.59** ± 0.22	**99.59** ± 0.16	98.55 ± 0.66	97.77 ± 0.67	93.45 ± 1.92	93.25 ± 1.57
OA (%)	65.19 ± 1.47	72.30 ± 1.43	89.78 ± 1.58	92.99 ± 0.60	93.24 ± 1.18	94.58 ± 0.56	95.13 ± 0.63	88.72 ± 1.53	**97.02** ± 0.77
AA (%)	71.22 ± 0.57	78.33 ± 0.94	90.83 ± 0.67	93.37 ± 0.52	94.20 ± 0.37	94.47 ± 0.37	95.30 ± 0.20	91.00 ± 0.78	**96.64** ± 0.70
Kappa	0.5547 ± 0.0148	0.6442 ± 0.0172	0.8656 ± 0.0199	0.9073 ± 0.0077	0.9109 ± 0.0150	0.9282 ± 0.0073	0.9355 ± 0.0082	0.8525 ± 0.0194	**0.9605** ± 0.0102

Note: the best results are presented in bold font.

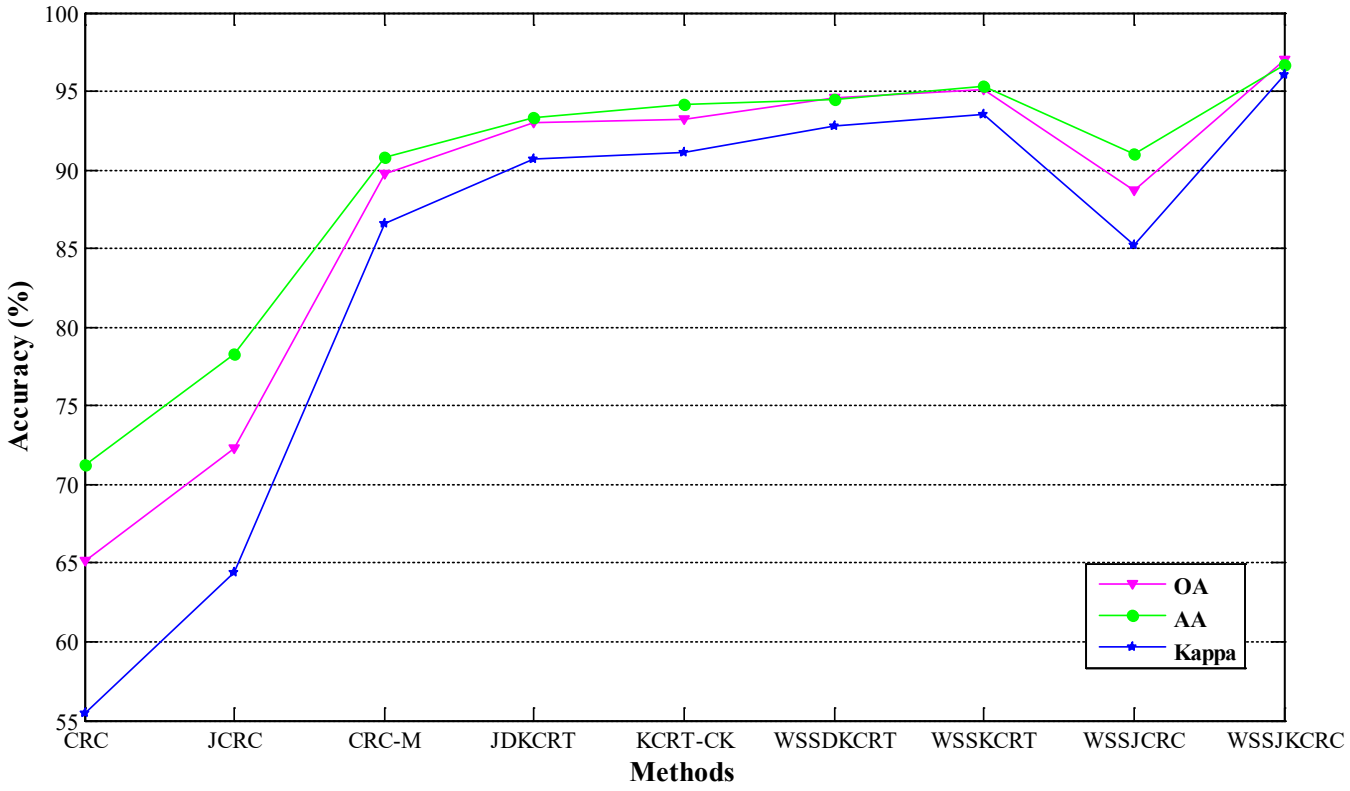


Figure 12. The OA, AA, and Kappa values for different methods on the Pavia University scene.

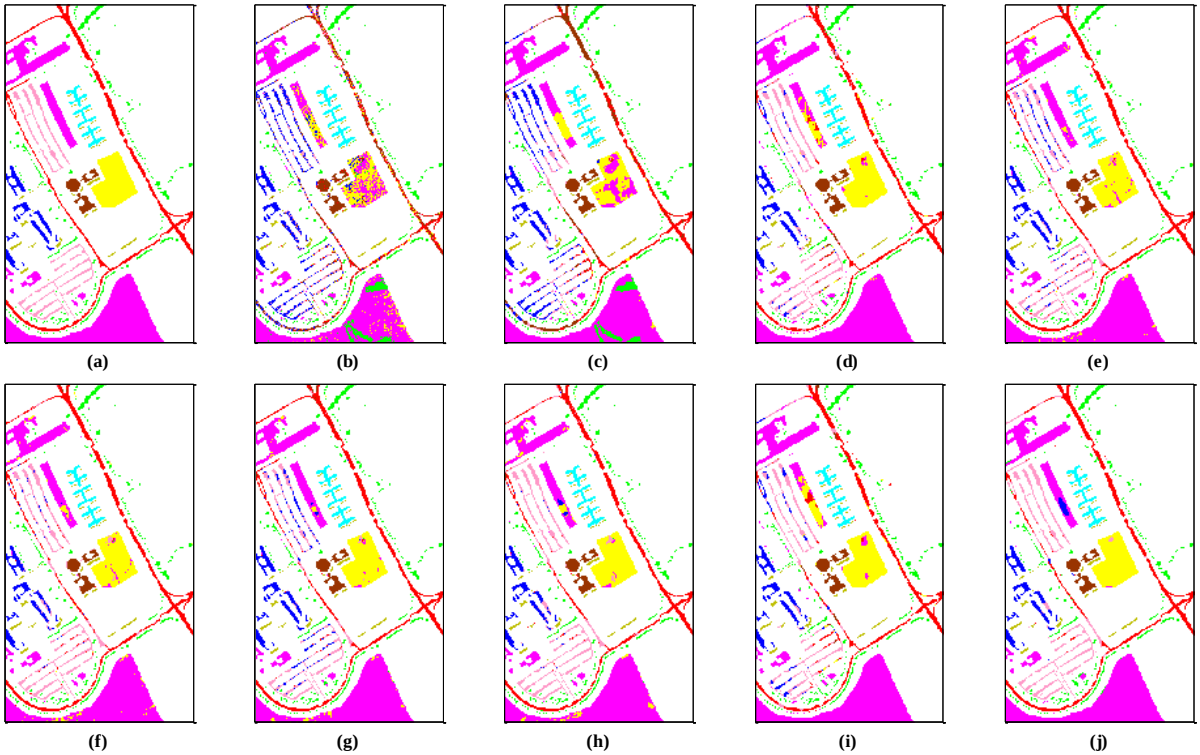


Figure 13. Land cover classification maps generated by different methods on the Pavia University scene: (**a**) ground truth, (**b**) CRC, (**c**) JCRC, (**d**) CRC-M, (**e**) JDKCRT, (**f**) KCRT-CK, (**g**) WSSDKCRT, (**h**) WSSKCRT, (**i**) WSSJCRC, and (**j**) WSSJKCRC.

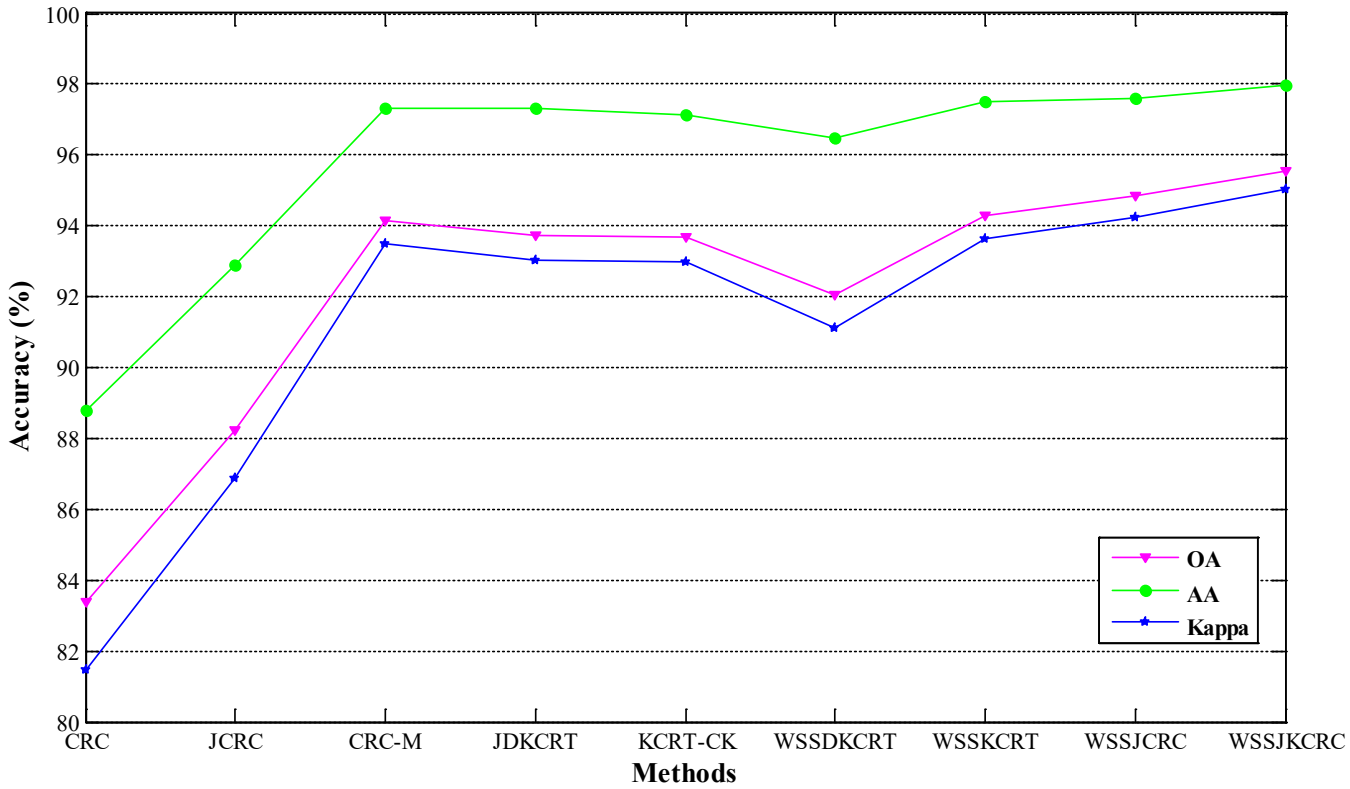


Figure 14. The OA, AA, and Kappa values for different methods on the Salinas scene.

Table 6. Land cover classification accuracy with STD for different methods on the Salinas scene.

Class	CRC	JCRC	CRC-M	JDKCRT	KCRT-CK	WSSDKCRT	WSSKCRT	WSSJCRC	WSSJKCRC
1	98.65 ± 0.43	99.17 ± 0.26	**100.00** ± 0.00	99.16 ± 0.96	98.79 ± 1.01	99.68 ± 0.39	98.76 ± 1.51	**100.00** ± 0.00	99.73 ± 0.44
2	99.20 ± 0.63	99.99 ± 0.01	99.99 ± 0.02	99.45 ± 0.61	99.25 ± 1.02	99.80 ± 0.19	99.88 ± 0.11	**100.00** ± 0.00	99.93 ± 0.07
3	92.62 ± 2.57	98.48 ± 1.59	99.87 ± 0.23	99.54 ± 0.36	99.92 ± 0.10	99.88 ± 0.11	99.78 ± 0.27	**100.00** ± 0.00	99.78 ± 0.16
4	96.26 ± 1.24	98.45 ± 0.91	99.03 ± 0.75	99.11 ± 0.67	99.03 ± 0.54	99.60 ± 0.32	99.00 ± 0.63	**99.74** ± 0.22	97.93 ± 0.39
5	96.53 ± 1.32	98.14 ± 0.79	95.91 ± 1.54	96.21 ± 2.18	96.48 ± 4.62	97.23 ± 1.88	97.46 ± 1.09	96.98 ± 1.55	**98.42** ± 1.00
6	99.91 ± 0.06	**99.99** ± 0.02	99.95 ± 0.09	99.69 ± 0.69	99.35 ± 1.40	99.26 ± 1.02	98.94 ± 1.60	99.86 ± 0.27	98.76 ± 2.80
7	99.82 ± 0.06	99.92 ± 0.04	99.99 ± 0.03	98.68 ± 0.91	98.10 ± 0.98	99.15 ± 0.45	99.24 ± 0.45	**100.00** ± 0.00	99.35 ± 0.44
8	65.41 ± 7.32	79.27 ± 6.64	82.62 ± 3.61	77.77 ± 8.09	80.38 ± 2.88	77.79 ± 4.44	81.82 ± 3.91	**87.24** ± 3.18	86.78 ± 3.51
9	99.71 ± 0.24	99.97 ± 0.10	99.98 ± 0.03	99.62 ± 0.38	99.45 ± 0.57	99.53 ± 0.51	99.81 ± 0.10	**99.99** ± 0.02	99.90 ± 0.19
10	86.79 ± 1.40	91.33 ± 1.18	95.93 ± 2.60	**97.96** ± 0.93	97.44 ± 1.47	95.99 ± 1.40	97.51 ± 1.67	96.14 ± 2.60	97.39 ± 1.54
11	95.46 ± 1.15	97.65 ± 1.24	99.76 ± 0.27	99.54 ± 0.94	**100.00** ± 0.00	99.20 ± 0.73	99.75 ± 0.30	99.66 ± 0.49	**100.00** ± 0.00
12	71.75 ± 5.26	89.04 ± 5.23	99.26 ± 0.43	99.87 ± 0.32	99.97 ± 0.04	**100.00** ± 0.00	99.77 ± 0.40	99.14 ± 1.10	99.96 ± 0.04
13	74.99 ± 4.38	83.45 ± 4.75	98.47 ± 1.02	**100.00** ± 0.00	99.72 ± 0.35	99.43 ± 0.83	99.92 ± 0.12	99.33 ± 0.57	99.82 ± 0.24
14	90.67 ± 1.92	95.89 ± 2.05	**99.39** ± 0.52	98.66 ± 1.04	98.82 ± 0.93	96.90 ± 3.07	99.16 ± 0.68	97.97 ± 2.05	99.19 ± 1.19
15	54.81 ± 5.17	57.01 ± 12.57	88.17 ± 5.19	**93.00** ± 3.22	89.84 ± 4.82	80.95 ± 5.27	90.36 ± 4.90	85.48 ± 3.35	91.44 ± 3.57
16	97.70 ± 0.60	98.44 ± 0.45	99.14 ± 0.76	99.11 ± 0.54	97.82 ± 2.43	99.14 ± 0.67	98.75 ± 0.98	**99.93** ± 0.08	99.07 ± 1.52
OA (%)	83.36 ± 1.24	88.23 ± 1.06	94.15 ± 0.71	93.72 ± 1.42	93.70 ± 0.65	92.04 ± 0.97	94.28 ± 0.89	94.85 ± 0.65	**95.55** ± 0.70
AA (%)	88.77 ± 0.60	92.89 ± 0.72	97.34 ± 0.27	97.34 ± 0.37	97.15 ± 0.48	96.47 ± 0.51	97.49 ± 0.38	97.59 ± 0.29	**97.97** ± 0.30
Kappa	0.8147 ± 0.0134	0.8686 ± 0.0120	0.9349 ± 0.0079	0.9303 ± 0.0157	0.9299 ± 0.0073	0.9114 ± 0.0108	0.9363 ± 0.0099	0.9426 ± 0.0073	**0.9504** ± 0.0078

Note: the best results are presented in bold font.

It should be noted that in the process of HSI classification and mapping, the row number and column number of each pixel in HSI are used as the position index of the corresponding pixel. The spectral information of the training pixels and test pixels, as well as the labeling information of the training pixels, are extracted through the corresponding position index. The land cover classification model can be established according to the training pixels and their corresponding class labels. Moreover, the established model is employed to classify the test pixels, and the obtained class labels are placed on the corresponding positions in HSI according to the position index of the test pixels to complete the land cover classification mapping.

It can be seen from Table 4 and Figure 10, Table 5 and Figure 12, and Table 6 and Figure 14 that the proposed WSSJKCRC method achieves the best land cover classification performance on the three hyperspectral scenes, in which OA, AA, and Kappa reach 96.21%, 96.20%, and 0.9555 for the Indian Pines scene, 97.02%, 96.64%, and 0.9605 for the Pavia University scene, 95.55%, 97.97%, and 0.9504 for the Salinas scene, respectively. Moreover, the classification maps generated by WSSJKCRC contain the least noise and the most homogeneous areas, as shown in Figures 11j, 13j and 15j. Among all the methods, the CRC method

produces the worst classification results on the three hyperspectral scenes, and there is a lot of noise in the classification maps produced by CRC shown in Figures 11b, 13b and 15b. The reason is that CRC only makes use of the spectral information of ground objects without considering any spatial information.

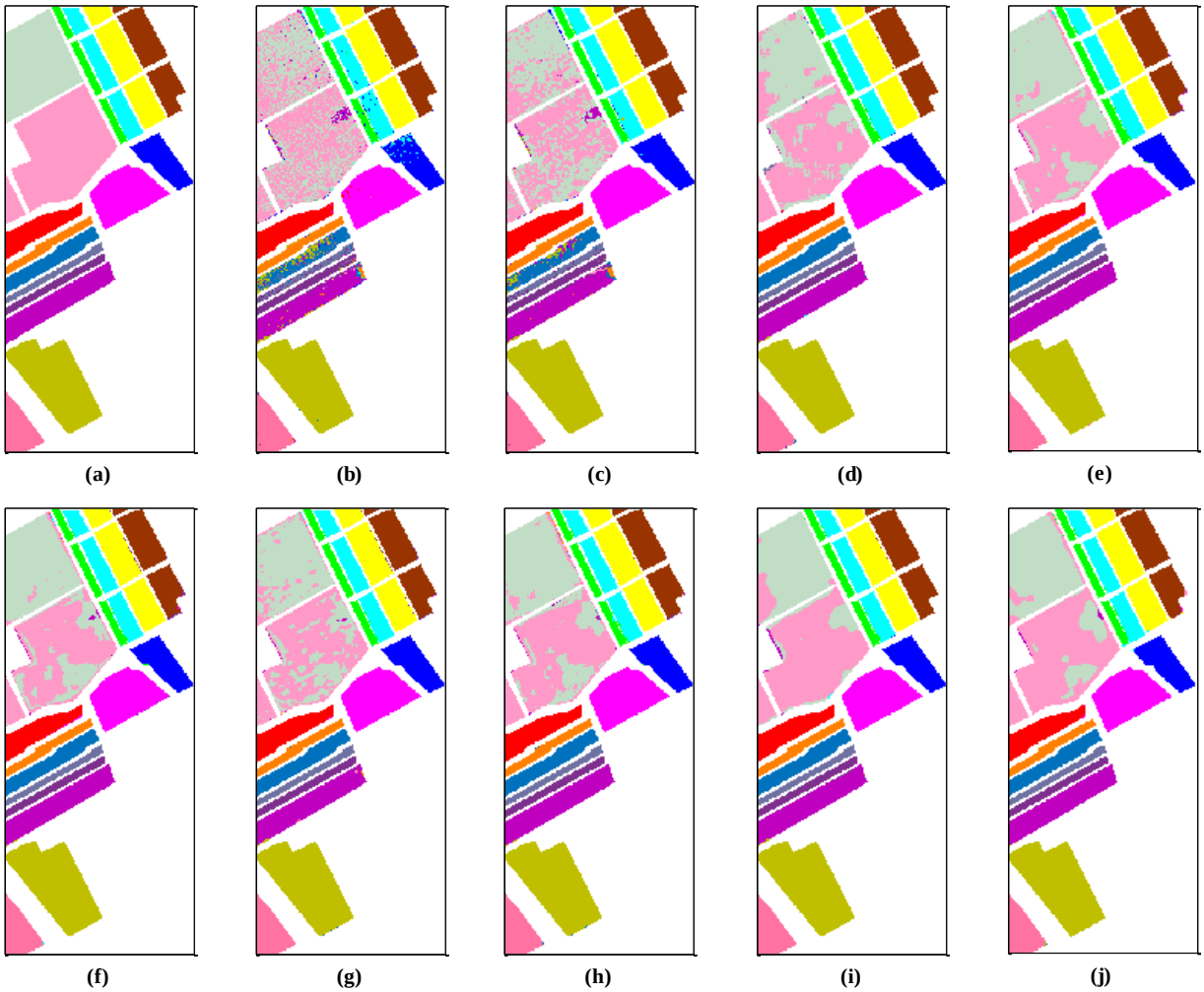


Figure 15. Land cover classification maps generated by different methods on the Salinas scene: (**a**) ground truth, (**b**) CRC, (**c**) JCRC, (**d**) CRC-M, (**e**) JDKCRT, (**f**) KCRT-CK, (**g**) WSSDKCRT, (**h**) WSSKCRT, (**i**) WSSJCRC, and (**j**) WSSJKCRC.

The CRC-M, JDKCRT, and KCRT-CK method perform spatial smoothing filtering on all pixels of HSI by directly averaging the spatial neighboring pixels, which reduces the noise in the original spectral curves of ground objects to a certain extent. Therefore, the classification performance of these three methods on the three hyperspectral scenes is significantly better than that of CRC, as shown in Table 4 and Figure 10, Table 5 and Figure 12, and Table 6 and Figure 14. However, these three methods do not consider the spectral shift caused by adjacency effect among the spatial neighboring pixels. As a result, the pixels in HSI contain a lot of noise. To alleviate such spectral shift, the WSSDKCRT and WSSKCRT methods utilize a weighted spatial filtering operator to perform spatial filtering on HSI. Moreover, the experimental results demonstrate that both WSSDKCRT and WSSKCRT achieve higher OA, AA, and Kappa than CRC-M, JDKCRT, and KCRT-CK on the Indian Pines and Pavia University scenes, as shown in Table 4 and Figure 10, and Table 5 and Figure 12. For the Salinas scene, WSSKCRT outperforms CRC-M, JDKCRT, and KCRT-CK, as shown in Table 6 and Figure 14. The CRC-M, JDKCRT, KCRT-CK, WSSDKCRT, and WSSKCRT methods consider the spatial information of HSI by taking the average or

weighted average of spatial neighboring pixels of each pixel, but do not take the spatial structure information for pixels into account.

The JCRC method simultaneously represents each test pixel and its spatial neighborhood pixels through the labeled training pixels to consider the spatial structure information of each test pixel to assist the classification of the test pixel. Consequently, JCRC achieves much higher OA, AA, and Kappa than CRC on the three hyperspectral scenes. As with CRC, the JCRC method does not carry out spatial filtering operation on HSI, which leads to a large amount of noise in the spectral curves of ground objects. The proposed WSSJCRC method not only considers the spatial structure information, but also employes the weighted spatial filtering operator to filter the pixels in HSI. It can be seen from the experimental results that WSSJCRC significantly outperforms CRC and JCRC for land cover classification on the three hyperspectral scenes. Especially on the Salinas scene, the land cover classification performance of WSSJCRC is only inferior to that of WSSJKCRC.

On the basis of the proposed WSSJCRC method, the proposed WSSJKCRC method utilizes the kernel trick to project the sample data into a high-dimensional feature space, thereby enhancing the separability for nonlinear samples. This is the reason that WSSJKCRC achieves the best land cover classification performance on the three hyperspectral scenes among all the compared methods.

Moreover, all methods use small-scale labeled training samples to classify large-scale unlabeled test samples in the experiment. Moreover, the OA of WSSJKCRC reaches more than 95% on all the three hyperspectral scenes, which indicates that WSSJKCRC can achieve the promising performance for land cover classification with small-scale labeled samples.

3.4.2. Stability Analysis for Classification Performance

In this section, the standard deviation (STD) of the results for ten times is utilized to evaluate the stability of classification performance for each method. The lower the STD values of OA, AA, and Kappa for a certain method are, the more stable the classification performance of the method is.

On the Indian Pines scene, the STD values of OA and Kappa for the proposed WSSJKCRC method are lowest among all the methods as shown in Table 4. The STD value of AA for WSSJKCRC is lower than that of other methods except for the JDKCRT method. Although the STD value of AA for WSSJKCRC is slightly higher than that of JDKCRT, the STD values of OA and Kappa for WSSJKCRC is lower than those of JDKCRT. On the Pavia University scene, it can be seen from Table 5 that the STD values of OA, AA, and Kappa for WSSJKCRC are lower than those of JCRC and WSSJCRC. Moreover, the STD values of OA and Kappa for WSSJKCRC are lower than those of CRC, CRC-M, and KCRT-CK. On the Salinas scene, the STD values of OA, AA, and Kappa for the proposed WSSJKCRC method are basically similar to those of the CRC-M, KCRT-CK, and WSSJCRC methods, but much lower than those of other methods as shown in Table 6.

Therefore, it can be seen from the above analysis that the stability of classification performance for the proposed WSSJKCRC method is satisfactory.

3.4.3. Significance Analysis for Classification Performance

It can be seen from Table 4 and Figure 10, Table 5 and Figure 12, and Table 6 and Figure 14 that the proposed WSSJKCRC method obtains the highest OA, AA, and Kappa values among all the compared methods on the Indian Pines, Pavia University, and Salinas scenes, which indicates that the proposed WSSJKCRC method has better classification performance than other methods. To further illustrate the significant difference for the classification performance of the proposed WSSJKCRC method and other methods, the *t*-test is utilized for significance analysis between the OA, AA, and Kappa obtained by WSSJKCRC for ten times and the OA, AA, and Kappa obtained by other methods for ten times at the significance level of 0.05. The *p*-values for OA, AA, and Kappa for different methods on the three hyperspectral scenes are shown in Table 7, in which *p*-value that is

less than 0.05 indicates that the classification performance for the proposed WSSJKCRC method is significantly different from that of the corresponding method.

Table 7. Significance analysis of classification performance for the proposed WSSJKCRC method and other methods.

Indicators	Datasets	CRC	JCRC	CRC-M	JDKCRT	KCRT-CK	WSSDKCRT	WSSKCRT	WSSJCRC
p-value for OA	Indian Pines	4.110×10^{-26}	1.082×10^{-19}	4.661×10^{-4}	3.155×10^{-5}	1.961×10^{-3}	1.423×10^{-2}	4.923×10^{-1}	4.526×10^{-9}
	Pavia University	7.786×10^{-22}	4.743×10^{-20}	3.202×10^{-10}	3.189×10^{-10}	2.293×10^{-7}	4.501×10^{-7}	2.141×10^{-5}	2.235×10^{-11}
	Salinas	1.214×10^{-15}	1.183×10^{-12}	5.536×10^{-4}	2.899×10^{-3}	1.747×10^{-5}	6.662×10^{-8}	3.532×10^{-3}	4.346×10^{-2}
p-value for AA	Indian Pines	1.860×10^{-25}	6.762×10^{-19}	2.346×10^{-3}	1.928×10^{-3}	6.447×10^{-3}	1.911×10^{-1}	6.467×10^{-1}	2.305×10^{-7}
	Pavia University	6.868×10^{-25}	2.903×10^{-20}	5.771×10^{-13}	1.280×10^{-9}	2.867×10^{-8}	1.531×10^{-7}	3.041×10^{-5}	3.740×10^{-12}
	Salinas	2.773×10^{-19}	1.637×10^{-13}	2.110×10^{-4}	8.763×10^{-4}	3.891×10^{-4}	5.278×10^{-7}	9.463×10^{-3}	1.542×10^{-2}
p-value for Kappa	Indian Pines	9.274×10^{-26}	1.168×10^{-19}	4.704×10^{-4}	3.291×10^{-5}	1.987×10^{-3}	1.534×10^{-2}	4.970×10^{-1}	4.409×10^{-9}
	Pavia University	4.011×10^{-23}	2.287×10^{-20}	1.875×10^{-10}	2.673×10^{-10}	1.705×10^{-7}	3.780×10^{-7}	1.969×10^{-5}	1.577×10^{-11}
	Salinas	8.783×10^{-16}	1.351×10^{-12}	5.487×10^{-4}	2.894×10^{-3}	1.839×10^{-5}	6.549×10^{-8}	3.596×10^{-3}	4.189×10^{-2}

As can be seen from Table 7, on the Indian Pines scene, the *p*-values for OA, AA, and Kappa between WSSJKCRC and WSSKCRT, and *p*-value for AA between WSSJKCRC and WSSDKCRT are higher than 0.05, but the *p*-values for OA, AA, and Kappa between WSSJKCRC and other methods are less than 0.05. On the Pavia University and Salinas scenes, the *p*-values for OA, AA, and Kappa between WSSJKCRC and other methods are all less than 0.05.

According to the classification results and the significance analysis of classification performance, the proposed WSSJKCRC method is not only better than other methods in classification performance, but also its classification performance is significantly different from other methods on the whole. Therefore, the proposed WSSJKCRC method is more advantageous than other comparison methods for HSI classification.

3.5. Comparison of Running Time

In this section, the running time for different methods is compared and analyzed on the Indian Pines, Pavia University, and Salinas scenes. All the methods are implemented using the MATLAB R2014a software on a computer with 2.90 GHz CPU and 32 GB RAM. To avoid random error and any bias, each method is performed repeatedly ten times and takes the average value of running time as the final result. The running time for these methods is shown in Table 8.

Table 8. Running time (seconds) of different methods for land cover classification.

Datasets	CRC	JCRC	CRC-M	JDKCRT	KCRT-CK	WSSDKCRT	WSSKCRT	WSSJCRC	WSSJKCRC
Indian Pines	0.47	10.75	13.49	24.33	15.36	27.34	27.90	4669.58	463.77
Pavia University	1.10	49.60	66.40	50.65	48.48	52.19	52.74	211.30	1094.92
Salinas	1.38	18.01	37.63	29.25	29.22	48.74	30.73	1136.04	1267.20

It can be seen from Table 8 that the proposed WSSJCRC and WSSJKCRC methods take more time than other methods on the three hyperspectral scenes. The main reason is that WSSJCRC and WSSJKCRC not only perform spatial filtering on HSI, but also consider the spatial structure information of each test pixel. Moreover, the larger the size of spatial filtering window W_f and spatial structure window W_s is, the more time WSSJCRC and WSSJKCRC take. However, the CRC-M, JDKCRT, KCRT-CK, WSSDKCRT, and WSSKCRT methods only perform spatial filtering on HSI, and JCRC only takes the spatial structure information of each test sample into account. Therefore, these methods take less time than the proposed WSSJCRC and WSSJKCRC methods. The CRC method only utilizes the spectral information of HSI for classification, but does not consider any spatial information, so its running time is the least.

4. Conclusions

To fully mine the spatial–spectral features for HSI, a novel WSSJCRC method is proposed for land cover classification in this paper, which not only utilizes a weighted spatial filtering operator to perform spatial filtering on HSI to alleviate the spectral shift caused by adjacency effect, but also considers the spatial structure information of each test pixel. On this basis, a WSSJKCRC method is also proposed, in which the kernel trick is introduced into the WSSJCRC method to enhance the separability of nonlinear samples. To verify the effectiveness of the proposed methods, the proposed WSSJCRC and WSSJKCRC methods are compared with seven traditional CR methods on three real hyperspectral scenes. According to the experimental results, it can be concluded as follows:

(1) The proposed WSSJCRC method significantly outperforms the traditional CRC and JCRC methods for land cover classification on the three hyperspectral scenes;
(2) The proposed WSSJKCRC method achieves the best land cover classification performance on the three hyperspectral scenes among all the compared methods. The OA, AA, and Kappa of WSSJKCRC reach 96.21%, 96.20%, and 0.9555 for the Indian Pines scene, 97.02%, 96.64%, and 0.9605 for the Pavia University scene, 95.55%, 97.97%, and 0.9504 for the Salinas scene, respectively;
(3) WSSJKCRC can achieve the promising performance for land cover classification with OA over 95% on the three hyperspectral scenes under the situation of small-scale labeled samples, thus effectively reducing the labeling cost for HSI.

The experimental results on the three hyperspectral scenes demonstrate that the proposed WSSJCRC and WSSJKCRC methods can effectively improve the classification performance of CR model by considering both spatial filtering and spatial structure information for HSI. The proposed WSSJKCRC method achieves the best land cover classification performance among all the compared methods by introducing the kernel technique, and obtains the promising accuracy with small-scale labeled samples. However, this framework increases the complexity of CR model and makes the proposed WSSJCRC and WSSJKCRC methods take more time than other compared methods. In the subsequent studies, based on this framework and according to the references [18,23,34], we will develop effective nearest neighbor cooperative representation mechanisms to optimize the dictionary of CR model and exclude the classes and samples unrelated to test samples from the dictionary. This can not only further improve the classification performance of CR model, but also improve its operation efficiency.

In this paper, only three small airborne hyperspectral scenes are used to verify the land cover classification performance of the proposed methods. In future research, the wider-scale hyperspectral scenes will be utilized to verify the generalization performance of the proposed methods for land cover classification. Moreover, the application of CR models in land cover classification based on spaceborne HSI will be carried out, such as GF-5 and Hyperion. Furthermore, all bands in HSI are employed to establish the land cover classification models in this paper. However, hundreds of continuous bands in HSI are highly correlated, which may lead to the "Hughes phenomenon", thus affecting the land cover classification performance of CR models. In the follow-up research, we will

construct dimensionality reduction algorithms for hyperspectral data to select the important bands with low redundancy and a large amount of information. Moreover, land cover classification models will be established based on CR methods and these important bands to further improve the operation efficiency and classification performance of the models.

Author Contributions: Data curation, R.Y.; Methodology, R.Y.; Supervision, Q.Z. and Z.L.; Validation, B.F. and Y.W.; Writing—original draft, R.Y.; Writing—review and editing, R.Y. and B.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Central Public-interest Scientific Institution Basal Research Fund (No. CAAS-ASTIP-2016-AII and No. JBYW-AII-2022-02) and the Project funded by China Postdoctoral Science Foundation (No. 2022M713415).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The three real hyperspectral scenes used in this paper can be obtained from http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes. And we obtained these datasets on 1 July 2021.

Acknowledgments: We acknowledge the Indian Pines, Pavia University, and Salinas hyperspectral scenes for the research of land cover classification.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Xie, S.; Liu, L.Y.; Zhang, X.; Yang, J.N. Mapping the annual dynamics of land cover in Beijing from 2001 to 2020 using Landsat dense time series stack. *ISPRS J. Photogramm. Remote Sens.* **2022**, *185*, 201–218. [CrossRef]
2. Phan, T.N.; Kuch, V.; Lehnert, L.W. Land cover classification using google earth engine and random forest classifier-the role of image composition. *Remote Sens.* **2020**, *12*, 2411. [CrossRef]
3. Agapiou, A. Land cover mapping from colorized CORONA archived greyscale satellite data and feature extraction classification. *Land* **2021**, *10*, 771. [CrossRef]
4. Akar, O.; Gormus, E.T. Land use/land cover mapping from airborne hyperspectral images with machine learning algorithms and contextual information. *Geocarto Int.* **2022**, *37*, 3963–3990. [CrossRef]
5. Wasniewski, A.; Hoscilo, A.; Chmielewska, M. Can a hierarchical classification of sentinel-2 data improve land cover mapping? *Remote Sens.* **2022**, *14*, 989. [CrossRef]
6. Yu, Y.T.; Guan, H.Y.; Li, D.L.; Gu, T.N.; Wang, L.F.; Ma, L.F.; Li, J. A hybrid capsule network for land cover classification using multispectral LiDAR data. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1263–1267. [CrossRef]
7. Hong, D.F.; Hu, J.L.; Yao, J.; Chanussot, J.; Zhu, X.X. Multimodal remote sensing benchmark datasets for land cover classification with a shared and specific feature learning model. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 68–80. [CrossRef]
8. Zafari, A.; Zurita-Milla, R.; Izquierdo-Verdiguier, E. Evaluating the performance of a random forest kernel for land cover classification. *Remote Sens.* **2019**, *11*, 575. [CrossRef]
9. Li, X.Y.; Sun, C.; Meng, H.M.; Ma, X.; Huang, G.H.; Xu, X. A novel efficient method for land cover classification in fragmented agricultural landscapes using sentinel satellite imagery. *Remote Sens.* **2022**, *14*, 2045. [CrossRef]
10. Lu, T.; Li, S.T.; Fang, L.Y.; Jia, X.P.; Benediktsson, J.A. From subpixel to superpixel: A novel fusion framework for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4398–4411. [CrossRef]
11. Gao, Q.S.; Lim, S.; Jia, X.P. Improved joint sparse models for hyperspectral image classification based on a novel neighbour selection strategy. *Remote Sens.* **2018**, *10*, 905. [CrossRef]
12. Yu, H.Y.; Shang, X.D.; Song, M.P.; Hu, J.C.; Jiao, T.; Guo, Q.D.; Zhang, B. Union of class-dependent collaborative representation based on maximum margin projection for hyperspectral imagery classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 553–566. [CrossRef]
13. Bigdeli, B.; Samadzadegan, F.; Reinartz, P. A multiple SVM system for classification of hyperspectral remote sensing data. *J. Indian Soc. Remote Sens.* **2013**, *41*, 763–776. [CrossRef]
14. Waske, B.; van der Linden, S.; Benediktsson, J.A.; Rabe, A.; Hostert, P. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2880–2889. [CrossRef]
15. Lopez-Fandino, J.; Quesada-Barriuso, P.; Heras, D.B.; Arguello, F. Efficient ELM-based techniques for the classification of hyperspectral remote sensing images on commodity GPUs. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2015**, *8*, 2884–2893. [CrossRef]
16. Huang, F.; Lu, J.; Tao, J.; Li, L.; Tan, X.C.; Liu, P. Research on optimization methods of ELM classification algorithm for hyperspectral remote sensing images. *IEEE Access* **2019**, *7*, 108070–108089. [CrossRef]

17. Zhang, Y.Q.; Cao, G.; Li, X.S.; Wang, B.S. Cascaded random forest for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2018**, *11*, 1082–1094. [CrossRef]
18. Li, W.; Du, Q.; Zhang, F.; Hu, W. Collaborative-representation-based nearest neighbor classifier for hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 389–393. [CrossRef]
19. Su, H.J.; Zhao, B.; Du, Q.; Du, P.J. Kernel collaborative representation with local correlation features for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1230–1241. [CrossRef]
20. Ye, M.C.; Qian, Y.T.; Zhou, J.; Tang, Y.Y. Dictionary learning-based feature-level domain adaptation for cross-scene hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1544–1562. [CrossRef]
21. Chen, H.; Ye, M.C.; Lei, L.; Lu, H.J.; Qian, Y.T. Semisupervised dual-dictionary learning for heterogeneous transfer learning on cross-scene hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2020**, *13*, 3164–3178. [CrossRef]
22. Shen, J.Y.; Cao, X.B.; Li, Y.; Xu, D. Feature adaptation and augmentation for cross-scene hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 622–626. [CrossRef]
23. Yang, R.C.; Zhou, Q.B.; Fan, B.L.; Wang, Y.T. Land cover classification from hyperspectral images via local nearest neighbor collaborative representation with Tikhonov regularization. *Land* **2022**, *11*, 702. [CrossRef]
24. He, X.; Chen, Y.S. Transferring CNN ensemble for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 876–880. [CrossRef]
25. Sun, H.; Zheng, X.T.; Lu, X.Q. A supervised segmentation network for hyperspectral image classification. *IEEE Trans. Image Process.* **2021**, *30*, 2810–2825. [CrossRef] [PubMed]
26. Du, P.J.; Gan, L.; Xia, J.S.; Wang, D.M. Multikernel adaptive collaborative representation for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4664–4677. [CrossRef]
27. Li, W.; Du, Q.; Zhang, F.; Hu, W. Hyperspectral image classification by fusing collaborative and sparse representations. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2016**, *9*, 4178–4187. [CrossRef]
28. Chen, X.; Li, S.Y.; Peng, J.T. Hyperspectral imagery classification with multiple regularized collaborative representations. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1121–1125. [CrossRef]
29. Li, W.; Du, Q.; Xiong, M.M. Kernel collaborative representation with Tikhonov regularization for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 48–52.
30. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227. [CrossRef] [PubMed]
31. Zhang, L.; Yang, M.; Feng, X.C. Sparse representation or collaborative representation: Which helps face recognition? In Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 471–478.
32. Li, W.; Tramel, E.W.; Prasad, S.; Fowler, J.E. Nearest regularized subspace for hyperspectral classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 477–489. [CrossRef]
33. Li, W.; Zhang, Y.X.; Liu, N.; Du, Q.; Tao, R. Structure-aware collaborative representation for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7246–7261. [CrossRef]
34. Yang, R.C.; Kan, J.M. Euclidean distance-based adaptive collaborative representation with Tikhonov regularization for hyperspectral image classification. *Multimed. Tools Appl.* **2022**. [CrossRef]
35. Liu, H.; Li, W.; Xia, X.G.; Zhang, M.M.; Gao, C.Z.; Tao, R. Spectral shift mitigation for cross-scene hyperspectral imagery classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2021**, *14*, 6624–6638. [CrossRef]
36. Ma, Y.; Li, C.; Li, H.; Mei, X.G.; Ma, J.Y. Hyperspectral image classification with discriminative kernel collaborative representation and Tikhonov regularization. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 587–591. [CrossRef]
37. Su, H.J.; Hu, Y.Z.; Lu, H.L.; Sun, W.W.; Du, Q. Diversity-driven multikernel collaborative representation ensemble for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2022**, *15*, 2861–2876. [CrossRef]
38. Li, W.; Du, Q. Joint within-class collaborative representation for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2014**, *7*, 2200–2208. [CrossRef]
39. Shaw, G.A.; Burke, H.K. Spectral imaging for remote sensing. *Lincoln Lab. J.* **2003**, *14*, 3–28.
40. Yang, R.C.; Fan, B.L.; Wei, R.; Wang, Y.T.; Zhou, Q.B. Land cover classification from hyperspectral images via weighted spatial-spectral kernel collaborative representation with Tikhonov regularization. *Land* **2022**, *11*, 263. [CrossRef]
41. Yang, J.H.; Qian, J.X. Hyperspectral image classification via multiscale joint collaborative representation with locally adaptive dictionary. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 112–116. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.