Article

Precise Crop Classification Using Spectral-Spatial-Location Fusion Based on Conditional Random Fields for UAV-Borne Hyperspectral Remote Sensing Imagery

Lifei Wei [1,2], Ming Yu [1,*], Yajing Liang [1], Ziran Yuan [1], Can Huang [1], Rong Li [1] and Yiwei Yu [1]

[1]   Faculty of Resources and Environmental Science, Hubei University, Wuhan 430062, China
[2]   Hubei Key Laboratory of Regional Development and Environmental Response, Hubei University, Wuhan 430062, China
*   Correspondence: 201711110811075@stu.hubu.edu.cn; Tel.: +86-173-173-99819

check for updates

Abstract: The precise classification of crop types is an important basis of agricultural monitoring and crop protection. With the rapid development of unmanned aerial vehicle (UAV) technology, UAV-borne hyperspectral remote sensing imagery with high spatial resolution has become the ideal data source for the precise classification of crops. For precise classification of crops with a wide variety of classes and varied spectra, the traditional spectral-based classification method has difficulty in mining large-scale spatial information and maintaining the detailed features of the classes. Therefore, a precise crop classification method using spectral-spatial-location fusion based on conditional random fields (SSLF-CRF) for UAV-borne hyperspectral remote sensing imagery is proposed in this paper. The proposed method integrates the spectral information, the spatial context, the spatial features, and the spatial location information in the conditional random field model by the probabilistic potentials, providing complementary information for the crop discrimination from different perspectives. The experimental results obtained with two UAV-borne high spatial resolution hyperspectral images confirm that the proposed method can solve the problems of large-scale spatial information modeling and spectral variability, improving the classification accuracy for each crop type. This method has important significance for the precise classification of crops in hyperspectral remote sensing imagery.

Keywords: hyperspectral remote sensing imagery; conditional random fields; spatial features; spatial location; precise crop classification; unmanned aerial vehicle

1. Introduction

China is one of the most populous countries in the world. With the rapid development of urbanization, crop protection in China is becoming more and more important [1,2]. Managing and monitoring agricultural areas and crop types in a timely manner is the basic requirement for ensuring food security [3]. The traditional methods of obtaining crop planting information include field surveys and statistical sampling. Although these methods are both accurate and objective, they have shortcomings in large-scale implementation and do not provide accurate spatial distribution information for crop areas [4]. With the development of remote sensing technology, the use of remote sensing imagery to classify crops is an effective way to monitor the spatial distribution of agriculture and obtain basic data for crop growth monitoring and yield forecasting [5,6].

Most Chinese agricultural systems are dominated by smallholder family farming, which provides 70–80% of Chinese food sources [7]. These smallholder family farms are characterized by a small

scale (less than 2 hectares, blurred boundaries), large distribution areas, mixed planting, etc. [8], meaning that traditional satellite-based multispectral remote sensing images and hyperspectral remote sensing images are unable to provide a sufficient data resolution for precision agriculture applications. However, the technology of unmanned aerial vehicles (UAVs) and light-weight sensors has provided an effective boost to the development of precision agriculture [9].

Compared with satellite or airborne data acquisitions, UAVs have a lower operating cost [10,11], are capable of more flexible operation [10,12], and allow more flexible flight route design [13]. UAVs can realize the rapid monitoring of crops and natural resources under custom temporal and spatial scales, and are now being widely used in the fields of precision agriculture, land-cover mapping [13], vegetation studies, environmental studies, and disaster monitoring [14]. Pádua et al. [15] used multi-temporal RGB images obtained by UAV to monitor grapevine growth; Zhou et al. [16] used two cameras in different spectral ranges installed on a UAV to obtain multi-temporal images to predict rice yield; and Marcial-Pablo et al. [17] used RGB and UAV-based multispectral images to estimate the vegetation fraction of corn. Hyperspectral sensors can be installed on UAVs to obtain hyperspectral remote sensing images with a high spatial resolution. Therefore, more and more spectral information and richer spatial information can be obtained in these images. At the same time, however, the simple use of spectral classification alone can no longer be applied to the classification of high spatial resolution hyperspectral data, and more spatial features need to be extracted for the precise classification of crops.

The random field model is one of the research hotspots of classification methods based on spatial context information, and it can construct scale-independent contextual information in a unified probability framework. The Markov random field (MRF) model is a typical random field model that can model the contextual information of the classification label, but it assumes that the observation information is conditionally independent, which limits the flexibility of spatial information utilization [18]. The conditional random field (CRF) model has flexibility of modeling, which overcomes the shortcomings of the MRF model and allows the posterior probability of given observation data to be directly modeled [19]. In the CRF model, the local potential function defines the relationship between random variables in the clique. Based on the different types of cliques, the potential function can be divided into unary potential, pairwise potential, and higher-order potential functions. The pairwise CRF model has been successfully applied to remote sensing processing and classification [20–25]. For example, the support vector conditional random field classifier incorporating a Mahalanobis distance boundary constraint (SVRFMC) can not only avoid the explicit modeling of observed data, but can also undertake appropriate smoothing with the consideration of contextual information by integrating SVM and a Mahalanobis distance boundary constraint [21]. The detail-preserving smoothing classifier based on conditional random fields (DPSCRF) considers the interaction of segmentation and classification in the CRF model, and adds spatial context information by segmentation [24]. For spatial–spectral fusion based on conditional random fields for the fine classification of crops in UAV-borne hyperspectral remote sensing imagery (SSF-CRF), the spatial and spectral features of the high spatial resolution hyperspectral data are fused by combining suitable potential functions in a pairwise conditional random field model to precisely classify crops [25]. However, these pairwise CRF models have difficulty in directly modeling large-scale spatial interaction information, resulting in the model often producing local over-smoothing and spectral variability in classification problems [23]. In the face of the problem of precise classification of crops with a wide variety and varied spectra, the traditional potential functions have difficulty in maintaining the details of the features in the classification, and will affect the accuracy of some classes due to the confusion with other classes.

Therefore, in this paper, in view of the problems of local over-smoothing and spectral variability faced by the CRF model in the precise classification of crops based on high-resolution hyperspectral remote sensing images, we propose a method for the precise classification of crops using spectral-spatial-location fusion based on conditional random fields (SSLF-CRF) for UAV-borne hyperspectral remote sensing imagery. The proposed method is further deepened on the basis of SSF-CRF, which constructs a unary potential function that combines spectral and spatial feature

information, a pairwise potential function of the spatial context information, and a higher-order potential function of the spatial location information. In the method of SSLF-CRF, the spectral information provides the most basic judgment for classification, based on the spectral diversity of different features; the spatial feature information offers more detailed information; the spatial context information uses the spatial relationship of neighboring pixels to spatially smooth and reduce noise; and the spatial location information provides complementary information for classification from the perspective of the non-local similarity of classes, which helps to distinguish similar land-cover types

2. Method

2.1. The SSLF-CRF Model

CRF is a probabilistic discriminant model. The maximum posterior probability $P(x|y)$ distribution is directly modeled as a Gibbs distribution [26,27], under the condition of given observation data, avoiding modeling the probability distribution of observed variables and having more flexible contextual information modeling capabilities, as shown in Equation (1):

$$P(x|y) = \frac{1}{Z(y)} \exp\left\{-\sum_{c \in C} \psi_c(x_c, y)\right\} \tag{1}$$

where the high spatial resolution hyperspectral remote sensing imagery is represented as $y = \{y_1, y_2, \cdots, y_N\}$; y_i is the spectral vector of the pixel at image position i, where $i \in V = \{1, 2, \cdots, N\}$ and N is the total number of pixels in the observed data; $x = \{x_1, x_2, \cdots, x_N\}$ is the class labels of the whole image; $x_i(i = 1, 2, \cdots N)$ comes from the label set $L = \{1, 2, \cdots, K\}$, where K is the number of classes; $Z(y) = \sum_x \exp\{-\sum_{c \in C} \psi_c(x_c, y)\}$ is the normalization function; and $\psi_c(x_c, y)$ is the potential function of clique c to model the relationship of the random variables, where C is the set of all the cliques, which represents a fully connected subgraph. Correspondingly, the Gibbs energy is as shown in Equation (2):

$$E(x|y) = -\log P(x|y) - \log Z = \sum_{c \in C} \psi_c(x_c, y) \tag{2}$$

Based on the CRF model, the goal of image classification is to find the largest label image x by the Bayesian maximum *a posteriori* (MAP) rule, so that the posterior probability $P(x|y)$ is the largest. Therefore, the MAP label X_{MAP} of the random field can be expressed as:

$$X_{MAP} = \underset{X}{\text{argmax}} P(x|y) = \underset{X}{\text{argmin}} E(x|y) \tag{3}$$

Therefore, in the classification problems of the CRF model, maximizing the posterior probability distribution $P(x|y)$ is equivalent to minimizing the energy function $E(x|y)$. In the proposed method, a higher-order CRF model, with the ability to encode higher-order spatial information, is used to improve the classification performance. The higher-order CRF model can be written as:

$$E(x|y) = \sum_{i \in V} \psi_i(x_i, y) + \sum_{i \in V, j \in N_i} \psi_{ij}(x_i, x_j, y) + \sum_{c \in C} \psi_c(x_c, y) \tag{4}$$

where $\psi_i(x_i, y)$ is the unary potential function; and $\psi_{ij}(x_i, x_j, y)$ is the pairwise potential function, which is defined over the local neighborhood *Ni* of the site of *i*. *C* represents the set of higher-order cliques, and $\psi_c(x_c, y)$ is the higher-order potential function defined over these cliques. In the proposed method, an eight-neighborhood system is used to encode the interactions of the neighboring variables. In view of the characteristics of high spatial resolution hyperspectral remote sensing imagery and the requirements for the precise classification of crops, the unary potential, the pairwise potential, and the higher-order potential are modeled on the basic framework shown in Equation (4).

2.1.1. Unary Potential

The unary potential function models the relationship between the label data and the observed data, and obtains the cost penalty of the specific labeled class by the class member probability of the corresponding pixel. Each pixel can be independently calculated using any discriminant classifier that gives the probability of the label, which is usually defined as:

$$\psi_i(x_i, y) = -\ln\left\{P\left[x_i = l_k \middle| f_i(y)\right]\right\} \tag{5}$$

where f is a feature mapping function that maps any arbitrary subset of a contiguous image cell to a feature vector, and $fi(y)$ represents the feature vector at position i. $P[xi = lk|fi(y)]$ is the probability of pixel i acquiring the label l_k that is based on the feature vector. Since the support vector machine (SVM) classifier performs well in remote sensing image classification with a small number of training samples [28,29], we use the SVM classifier, with a radial basis function as the kernel type, to provide the discriminant information of the classes based on the spatial-spectral feature vector. In the proposed method, the two parameters C and γ are set to the default values.

(1) Spectral Characteristics

Hyperspectral remote sensing imagery usually contains tens or even hundreds of spectral bands, so the spectral values of adjacent bands in hyperspectral images usually have a strong correlation [30], which results in data redundancy and can generate a "dimension disaster" problem during classification [31]. Minimum noise fraction (MNF) rotation is a commonly used method for extracting spectral features, and it is both simple and easy to implement. After MNF transformation, the components are arranged according to the signal-to-noise ratio, where the information is mainly concentrated in the first component. As the components increase, the image quality gradually decreases. Studies have shown that, compared with the original high-dimensional image data and the feature image obtained by principal component analysis (PCA) transformation, the low-dimensional feature image obtained by MNF transformation can extract the spectral information more effectively [32]. Therefore, we choose this method to extract the spectral information of the high spatial resolution hyperspectral imagery.

(2) Spatial Feature Characteristics

(A) Texture Features

Hyperspectral remote sensing images not only have continuous and abundant spectral information, but also rich texture information. Texture is the visual effect caused by spatial variation in tonal quantity over relatively small areas, which has been widely used in remote sensing image processing [33–36]. A number of studies have demonstrated the efficiency of texture for improving land-cover classification accuracy [37,38]. The gray-level co-occurrence matrix (GLCM) is a commonly used method for extracting texture information with a better discriminative ability [39,40]. The principle is to establish a GLCM between two pixels in a certain positional relationship in the image and to extract the corresponding feature quantity from this matrix for the texture analysis.

If we let $f(x, y)$ be a two-dimensional digital image of size $M \times N$, and the gray level is Ng, then the GLCM satisfying a certain spatial relationship is:

$$P(i, j) = \#\left\{(x1, y1), (x2, y2) \in M, N \middle| f(x1, y1) = i, f(x2, y2) = j\right\} \tag{6}$$

where $\#(x)$ is the number of elements in the set x, and P is the matrix of $Ng \times Ng$. If the distance between $(x1, y1)$ and $(x2, y2)$ is d and the angle is θ, then the GLCM $P(i, j, d, \theta)$ of various spacings and angles is:

$$P(i, j, d, \theta) = \#\left\{(x1, y1), (x2, y2) \in M, N \middle| f(x1, y1) = i, f(x2, y2) = j\right\} \tag{7}$$

In this paper, we use the six texture metrics of homogeneity, angular second moment, contrast, dissimilarity, mean, and entropy:

$$Hom = \sum_{i=0}^{L-1} \sum_{j=o}^{L-1} \frac{p(i,j)}{1 + (i-j)^2} \tag{8}$$

$$ASM = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p(i,j)^2 \tag{9}$$

$$Con = \sum_{n=0}^{L-1} n^2 \left\{ \sum_{\substack{i=0 \\ n=|i-j|}}^{L-1} \sum_{j=0}^{L-1} p(i,j) \right\} \tag{10}$$

$$Dis = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} |i-j| p(i,j) \tag{11}$$

$$Mean = \frac{1}{n \times n} \sum_{i} \sum_{j} f(i,j) \tag{12}$$

$$Ent = -\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p(i,j) \log_2 p(i,j) \tag{13}$$

(B)   Endmember Components

There are some mixed pixels in high spatial resolution hyperspectral images, which have a certain degree of influence on the classification. If a method of mixed pixel decomposition is used, the corresponding percentage of each class in the mixed pixel can be expressed, thereby obtaining an abundance image equal to the number of classes. The endmember is a physical quantity associated with a mixed pixel. It is the main parameter describing the linear mixed model, representing the characteristic feature with a relatively fixed spectrum. The endmember extraction can obtain more detailed information of the image. In the proposed method, the sequential maximum angle convex cone (SMACC) endmember model is used to extract the endmember spectra and the abundance image, to form the endmember component [41], which can be defined as:

$$H(c,i) = \sum_{k}^{N} R(c,k) A(k,j) \tag{14}$$

where H is the spectral endmember; c and i are the band index and the pixel index, respectively; k and j represent an index from 1 to the largest endmember; R is the matrix containing the endmember spectra; and A is the abundance matrix containing endmember j to endmember k in each pixel.

(C)   Morphological Features

Mathematical morphology is an effective image feature extraction tool that describes the local characteristics of images. The basic morphological operations are corrosion, expansion, and opening and closing operations, which act on the image through a series of shape regions called structuring elements (SEs). The morphological opening and closing reconstructions are another common kind of operator, which have a better shape preservation ability than the classical morphological filters. Since the shape of the SEs used in the filtering is adaptive, with respect to the structures present in the image itself, it nominally introduces no shape noise [42,43], as shown in Figure 1. In the proposed method, we extract the spatial information of the images based on "opening reconstruction followed

by closing reconstruction" (OFC), which can simultaneously smooth the bright and dark details of the structure while maintaining the overall feature stability and improving the consistency within the object area [44,45].
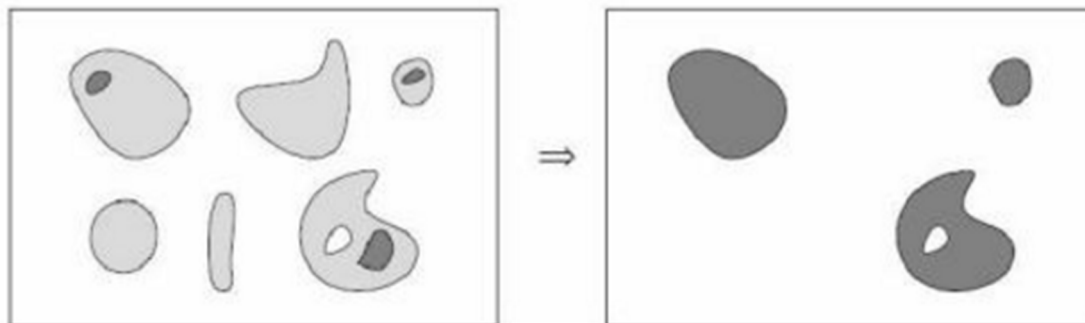

Figure 1. Morphological reconstruction.

The OFC operator is a hybrid operation of opening by reconstruction (OBR) and closing by reconstruction (CBR), which can be defined as:

$$OFC^{SE}(f) = \gamma_R^{SE}(\varphi_R^{SE}(f)) \tag{15}$$

where $\varphi_R^{SE}(f)$ indicates the closing reconstruction of image f, and $\gamma_R^{SE}(\varphi_R^{SE}(f))$ is the opening reconstruction of the closing reconstruction image.

2.1.2. Pairwise Potential

Based on the theory of spatial correlation, the pairwise potential function simulates the spatial context information between each pixel and its neighborhood by considering the spatial pattern prior knowledge of the land-cover classes, so that the spectral changes and noise can be mitigated. The pairwise potential function can model the spatial smoothing prior, and support adjacent pixels of the homogeneous region to obtain the same class label. In this paper, it is defined as shown in Equations (16) and (17) [46,47]:

$$\psi_{ij}(x_i, x_j, y) = \begin{cases} 0 & if\ x_i = x_j \\ g_{ij}(y) & otherwise \end{cases} \tag{16}$$

$$g_{ij} = \lambda \left(1 + \theta \frac{\exp(-\|y_i - y_j\|^2)/\pi}{\|i - j\|^2}\right) \tag{17}$$

where (i, j) is the spatial position of the adjacent pixel, and g_{ij} is the smoothing term associated with the data y, which is used to measure the difference between adjacent pixels. y_i and y_j are the spectral vectors representing the pixels of i and j. The parameters λ and θ are set to control the strength of the pairwise potential function, and π is twice the mean-square error between the spectral vectors of all the neighboring pixels in the image, i.e., $2\langle\|y_i - y_j\|^2\rangle$, where $\langle . \rangle$ represents the mean operation. The pairwise potential function can encourage adjacent pixels to have the same class in the classification result, thereby enhancing the spatial correlation.

2.1.3. Higher-Order Potential

Location information in local areas can help to distinguish confusing classes and mitigate the effects of spectral variability. The higher-order potential function in the CRF model can capture richer spatial interactions, it can explicitly model the spatial location information by considering the non-local spatial interaction between the target pixel and the nearest-neighbor training data, and it can provide

the complementary information of the feature class from the perspective of the non-local similarity of the class. It can be expressed as:

$$\psi_c(x_c, y) = \left(\|i - j\|^2 \right) / \delta_s + \|y_i - y_j\|^2 / \delta_r + K_c \tag{18}$$

$$K_c = \ln\left(\sum_{k=1}^{K} \exp\left(\frac{-\|i - j\|^2}{\delta_s} \right) * \exp\left(\frac{-\|y_i - y_j\|^2}{\delta_r} \right) \right) \tag{19}$$

$$j = \underset{j}{\operatorname{argmin}}(\|j - i\|^2), j \in S_k \tag{20}$$

$$k = \underset{k}{\operatorname{argmin}}\left(\|\mu_k - y_i\|^2 \right), k \in M \tag{21}$$

$$(S, \mu) = \underset{S, \mu'}{\operatorname{argmin}}\left(\sum_{l=1, j \in S_l}^{M} \|y_j^l - \mu_l'\|^2 \right), l = x_i \in L \tag{22}$$

The higher-order cliques in the higher-order potential function consist of the target pixel i and the training data. y^l is the training samples representing the class label l. The set $S = \{S_1, S_2, \cdots, S_M\}$ is a set of patterns formed by automatic clustering of training data y^l, and μ is the center corresponding to each pattern. In the proposed method, an adaptive mean shift clustering algorithm is used to obtain M patterns of the same feature class. Each pattern of the same feature has similar spectral characteristics and is represented by its corresponding mode center. For the target pixel i to be calculated, a pattern that is most similar to each feature class can be obtained based on the spectral vector distance. In the modeled higher-order potential function, the candidate pixels of each class are searched by finding the nearest-neighboring pixel to the target pixel space position. Equation (18) calculates the higher-order potential function based on the distance using the candidate pixels of each feature class in the spatial location and the spectral domain, where the pixel at position i represents the target pixel to be calculated, and the pixel at j is the candidate pixel. δ_s and δ_r are two independent parameters, which are respectively set as the second smallest value of the spatial distances and spectral distances between the target pixel and the corresponding candidate pixel for each class. K_c is a regularization constant.

Since the higher-order potential function can encode the spatial interaction between each target pixel and the training data, it can be transformed into a unary potential function for solving, and the higher-order potential function is re-described using the class member probability, as shown in Equation (23):

$$P^{HP}(x_i = l_k) = \exp(-\psi_c(x_c, y)) = \frac{1}{\exp(K_c)} \exp\left(\frac{-\|i - j\|^2}{\delta_s} \right) * \exp\left(\frac{-\|y_i - y_j\|^2}{\delta_r} \right) \tag{23}$$

After the higher-order potential function is transformed into the class member probability of the target pixel, the Gaussian decreasing function can be used to calculate the distance in the spatial location domain and in the spectral space. In this paper, the class membership probability obtained by the higher-order potential function is integrated with the unary potential function to effectively utilize the complementary advantages of the spectral and spatial position information, as shown in Equation (24):

$$\psi_i(x_i) = -\ln\left((1 - \beta)P(x_i = l_k) + \beta P^{HP}(x_i = l_k) \right) \tag{24}$$

where β is a free parameter from 0 to 1 for balancing the spectral information and the spatial location information.

2.2. Algorithm Flowchart

The flowchart of the algorithm for precise crop classification using spectral-spatial-location fusion based on conditional random fields (SSLF-CRF) for UAV-borne hyperspectral remote sensing imagery is shown in Figure 2.
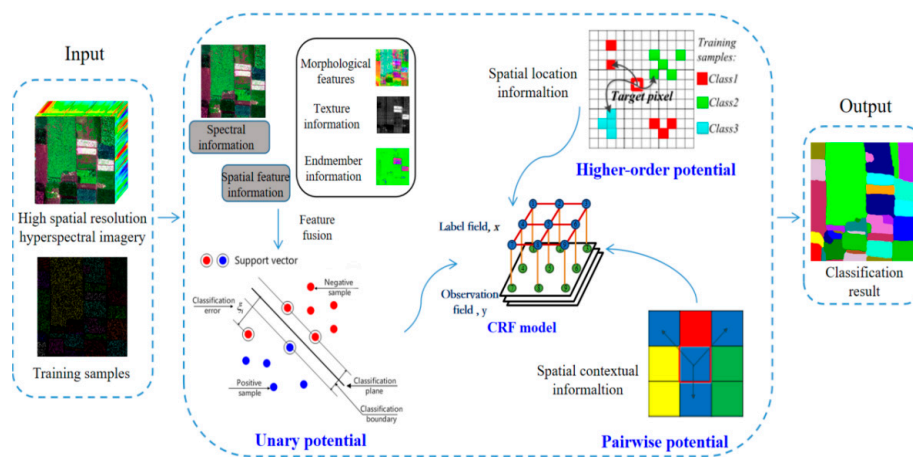

Figure 2. Flowchart of the spectral-spatial-location fusion based on conditional random fields (SSFL-CRF) method.

This method models the spectral information, the spatial feature information, the spatial context information, and the spatial location information through the unary potential, pairwise potential, and higher-order potential, respectively, which can integrate the complementary information from three different perspectives and make full use of the spatial information of the image, thereby improving the classification results. The SSLF-CRF method is described in detail as follows:

(1) According to the characteristics of high spatial resolution hyperspectral remote sensing data obtained by UAV, the original image is subjected to MNF rotation. The noise covariance matrix in the principal component is used to separate and readjust the noise in the data, to minimize the variance of the transformed noise data and bands that are not relevant.

(2) The unary potential function model the spectral and spatial feature information. The spectral information is the most basic information for judging the various land-cover classes, and the spatial feature information provides more details for the classification. Representative features are selected from the perspective of mathematical morphology, spatial texture, and mixed pixel decomposition, and then combined with the spectral information of each pixel to form a spectral-spatial fusion feature vector. The SVM classifier is used to model the relationship between the label and the spectral-spatial fusion feature vector, and the probability estimate of each pixel is calculated independently, based on the feature vector, according to the given label.

(3) The spatial smoothing term simulates the spatial context information for each pixel and its corresponding neighborhood through the label field and observation field, which is modeled by the pairwise potential function. According to spatial correlation theory, the spatial smoothing term encourages the neighboring pixels of the homogeneous regions to share the same class label, and punishes the class discontinuity in the local regions.

(4) Based on the characteristics of the similar spectral properties of the same feature class in a local region, the higher-order potential function is used to explicitly model the spatial location information by considering the training data of the target pixel closest to its spatial location, and it makes full use of the non-local similarity of the feature class to provide auxiliary information for the features that are difficult to distinguish. Differing from the spectral, spatial feature, and spatial contextual information, the spatial location information uses the regularity of the image to model the nonlocal similarity of the land-cover types through the location information.

3. Experimental Results

3.1. Study Areas

The two datasets used in the experiments cover the cities of Hanchuan (113°22′–113°57′E, 30°22′–30°51′N) and Honghu (113°07′–114°05′E, 29°39′–30°12′N) in Hubei Province, China (see Figures 3 and 4).

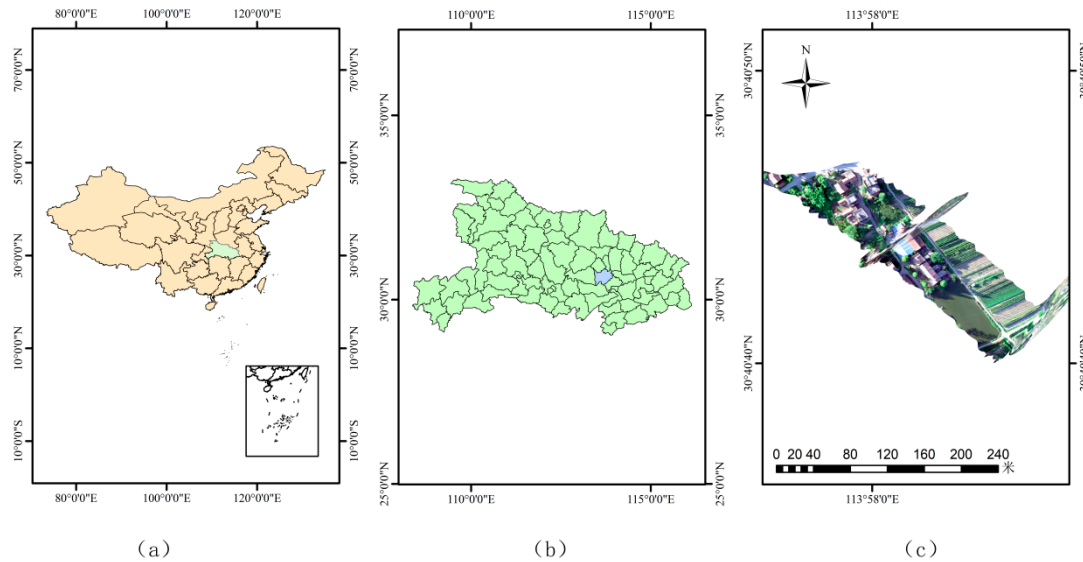(a)　　　　　　　　　　　　　　　　(b)　　　　　　　　　　　　　　　　(c)

Figure 3. (**a**) The location of Hubei Province in China. (**b**) Administrative area map of the city of Hanchuan in Hubei Province. (**c**) The study site.
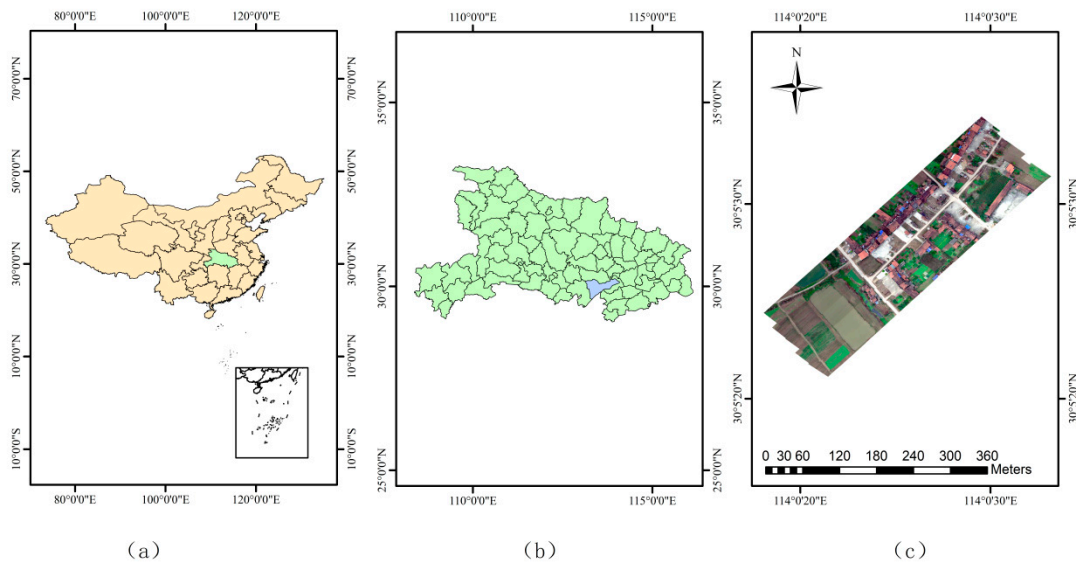

(a)　　　　　　　　　　　　　　　　(b)　　　　　　　　　　　　　　　　(c)

Figure 4. (**a**) The location of Hubei Province in China. (**b**) Administrative area map of the city of Honghu in Hubei Province. (**c**) The study site.

Located in the central part of Hubei Province, China, the city of Hanchuan is situated on the lower reaches of the Han River and in the middle of the Jianghan Plain, where the terrain is flat and low-lying. The climate of this city is subtropical humid monsoon. A wide variety of crops are grown in the area, including rice, wheat, cotton, and rapeseed.

The city of Honghu is located in the south-central part of Hubei Province, in the middle and lower reaches of the Yangtze River, and in the southeast of Jianghan Plain. The terrain is higher in the north

and south and is lower in the central part, forming a landform with a wide area and flat terrain. The climatic characteristics of the city of Honghu are similar to those of the city of Hanchuan, and both cities belong to the subtropical humid monsoon climate zone. The main crops of the Honghu area are cotton, rice, wheat, barley, broad beans, sorghum, and rapeseed.

3.2. Data Collection

The experimental data acquisition process consisted of two parts: (1) acquisition of aerial remote sensing images; and (2) acquisition of the ground truth. The drone model used for collecting the data was a DJI Matrice 600 Pro, and the hyperspectral imager was a Headwall Nano-Hyperspec hyperspectral imaging sensor manufactured by Headwall Photonics Inc. (Fitchburg, MA, USA) and provided by Xingbo Keyi Co., Ltd. (Guangzhou, China) ultra-micro airborne hyperspectral imaging spectrometer. This unit has a global positioning system/inertial measurement unit (GPS/IMU) navigation system and a complete data acquisition storage module. The integrated data acquisition system has Gig-E connection, which allows the data to be downloaded during flight. The synchronously acquired global positioning system/inertial navigation system (GPS/INS) data facilitate the subsequent geometric correction. Furthermore, the weight of the spectrometer is only 0.5 kg, which significantly reduces the burden on the UAV. The parameters of the Nano-Hyperspec imager are listed in Table 1.

Table 1. Nano-Hyperspec hyperspectral imaging sensor parameter information.

Class		Parameter		Class		Parameter	
Wavelength range		400–1000 nm		Field of view	33	22	16
Number of spectral channels		270		Storage		480 GB	
Lens focal length	8 mm	12 mm	17 mm	Camera type		COMS	
Operating temperature		0~50°C		Weight		<0.6 kg (no lens)	

3.3. Preprocessing of the UAV Images

Taking into account the hardware conditions and experimental requirements, the following pre-processing was performed on the high spatial resolution hyperspectral images acquired by the UAV: laboratory calibration, geometric correction, radiometric correction, and test sample production. In view of the low flight height of the drone, the complex atmospheric effects in flight could be ignored. The specific process of the preprocessing is described as follows:

(1) The laboratory calibration of the sensor was the first step, converting the output signal of each sensor unit into an accurate radiance value.

(2) The second step was geometric correction. The UAV integrates the sensors and the position and orientation system (POS) that combined with the differential GPS technology and Inertial Measurement Unit (IMU) technology, which can provide the position and attitude parameters of the sensor. For the geometric correction, the time standard of the POS data and hyperspectral image acquisition should be unified and corresponding. The correspondence between the image pixels and ground coordinates could thus be established by coordinate system transformation. Finally, the corrected pixels were resampled and the corrected image was reconstructed.

(3) The third step was radiometric correction. We used a calibration blanket with a third-order reflectivity with reflectances of 11%, 32% and 56%, respectively. Before the UAV took off, the calibration blanket was placed in a suitable position in the study area to ensure that the study area and the calibration blanket appeared in the same image at the same time. Three sets of ROIs were selected in the three reflectance regions of the calibration blanket by means of ENVI, and the average radiation value of each group of ROIs was calculated. Finally, the standard reflectance of the calibration blanket and the radiation values could be linearly regressed in the IDL, according to Equation (25).

$$reflectance = a \cdot DN(radiance) + b \tag{25}$$

1.　　The fourth step was test sample production. We obtained GPS linear feature information at where the data were acquired, and obtained ground marks with reference to the relative features synchronously, to generate the ground-truth data.

3.4. Experimental Description

In the experiments, the comparison algorithms were a pixel-based classification method, an object-oriented classification method, and random field based classification methods. The pixel-based classification method was the traditional SVM algorithm with a radial basis function as the kernel type. The object-oriented classification approach used a majority voting strategy for each segmentation region based on the same pixelwise SVM classification map. The segmentation was provided by mean shift segmentation (MS) [48]. The random field based methods were the Markovian support vector classifier (MSVC) [49], SVRFMC [21], DPSCRF [24], and SSF-CRF [25]. The MSVC algorithm integrates SVM into the MRF model, and obtains the final classification result through the iterated conditional modes (ICM) algorithm, using the Gaussian radial basis function and the Potts model as the kernel function and the local prior energy function, respectively. SVRFMC is a CRF classification algorithm based on Markov boundary constraints, where the spatial term is constrained by the Markov distance boundary, to maintain the spatial details of the classification results. DPSCRF considers the interaction of segmentation and classification in the CRF model, and adds large-scale spatial context information by segmentation. SSF-CRF fuses the spatial and spectral features of the high spatial resolution hyperspectral data by combining suitable potential functions in a pairwise conditional random field model.

In each experiment, three kinds of accuracies were used to assess the quantitative performance: the accuracy of each class, the overall accuracy (OA: the percentage of correctly classified pixels), and the Kappa coefficient (Kappa) [50]. For each algorithm, 1%, 3%, 5%, and 10% of the training samples were randomly selected to classify, and the remaining 99%, 97%, 95%, and 90% of the samples were used to verify the precision.

3.5. Experiment 1: Hanchuan Dataset

The dataset used in the first set of experiments was a UAV-borne high spatial resolution hyperspectral remote sensing image acquired over the city of Hanchuan, Hubei Province, China, in June 2016. The spatial resolution is 0.1 m. The image contains 303 × 600 pixels and 270 spectral channels, including nine land-cover classes: red roof, gray roof, tree, road, strawberry, pea, soy, shadow, and iron sheet. The true-color image is shown in Figure 5a and the corresponding ground-truth map is displayed in Figure 5b. Table 2 lists the number of training and test samples for each class at 1% training samples.

Table 2. The training and test sample information for the Hanchuan dataset.

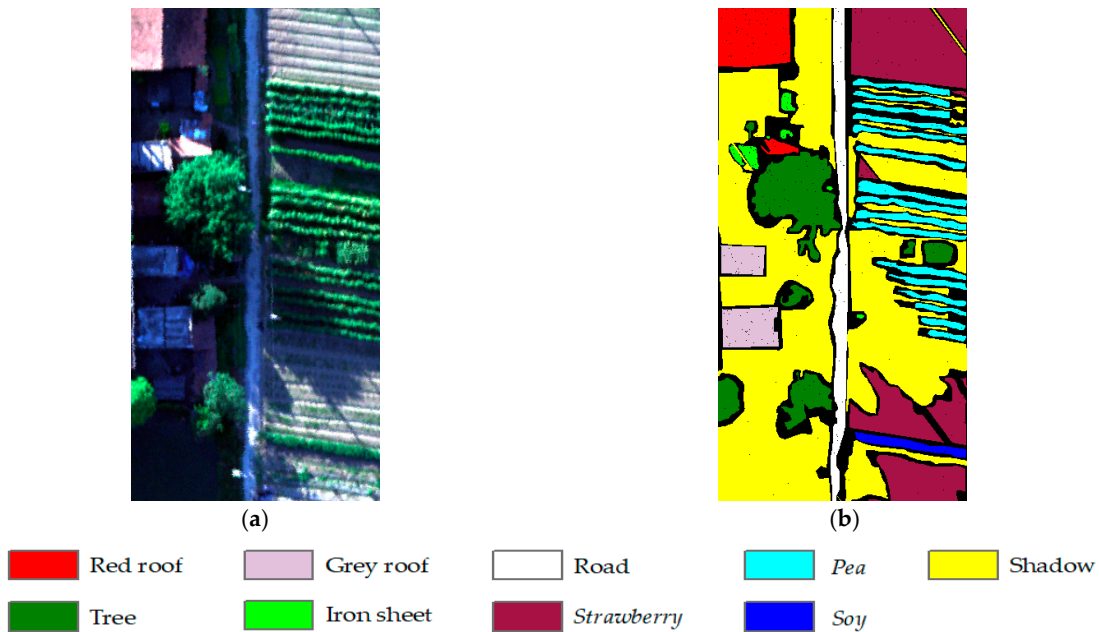| Class	Training Samples	Test Samples
Red roof	65	6525
Gray roof	48	4766
Tree	117	11664
Road	72	7138
Strawberry	226	22402
Pea	110	10937
Soy	13	1322
Shadow	763	75609
Iron sheet	10	1006

Figure 5. The Hanchuan dataset: (**a**) the true-color image; (**b**) the ground-truth map.

In the experiment with the Hanchuan dataset, the original image was reduced from 270 dimensions to 10 dimensions by MNF, according to the characteristics of the image. In order to introduce as little noise as possible, the most suitable spatial features were selected for the classification through a large number of experiments as follows: (1) the four endmembers of shadow, tree, strawberry, and red roof; (2) the four texture features of homogeneity, angular second moment, contrast, and mean were extracted with a window of 7×7; and (3) the morphological features were extracted with a disk-shaped SE with a size of 8. The parameters λ, θ, and β were set to 0.7, 2.4, and 0.4, to obtain the best results.

The experimental results corresponding to the different classification algorithms (SVM, MS, SVRFMC, DPSCRF, MSVC, SSF-CRF, and SSLF-CRF) are shown in Figure 6. Figure 6a is the classification result of SVM, which delivers a salt-and-pepper appearance because this algorithm does not consider the spatial information. Figure 6b–g are the results of the algorithms that consider the spatial context and spatial feature information, where the classification effects are smoother and the visual effects are better than those of the SVM algorithm. However, due to the influence of spectral variation, the different algorithms show different performances in terms of detail retention. For example, as shown in Figure 5a, the spectra of the iron sheet, gray roof, pea, tree, and soy classes have a certain similarity, so it is a challenging task to accurately distinguish these classes. In Figure 6b–e, the iron sheet is not easily identified, and for the most part is misclassified as gray roof and road, as shown by the red box in the figures. The features in the black box are pea, but this part was misclassified as tree or soy in most of the methods.

When compared with the result of SSF-CRF, the edge of this part in SSLF-CRF is the clearest. The features in the blue box are soy, but this part is mainly classified as tree by the methods of MS, SVRFMC, DPSCRF, and MSVC. In the proposed SSLF-CRF method, while alleviating the salt-and-pepper noise, the above misclassified classes can also be correctly identified. With the proposed method, most of the land-cover classes maintain a more complete boundary shape and useful feature details, with a better visual classification performance.
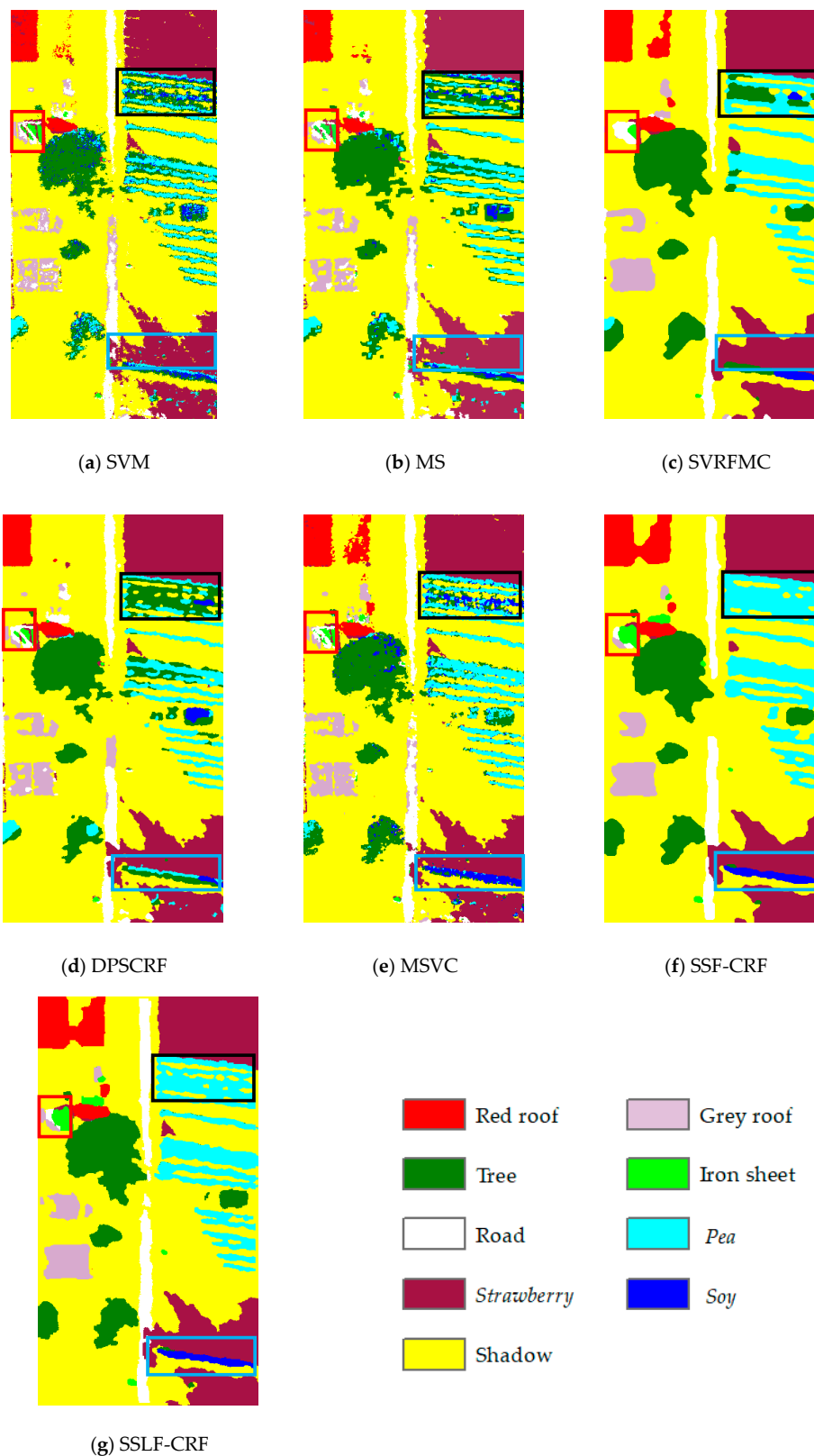
(**a**) SVM   (**b**) MS   (**c**) SVRFMC

(**d**) DPSCRF   (**e**) MSVC   (**f**) SSF-CRF

(**g**) SSLF-CRF

Figure 6. Hanchuan dataset classification results: (**a**) support vector machine (SVM); (**b**) mean shift segmentation (MS); (**c**) support vector conditional random field classifier with a Mahalanobis distance boundary constraint (SVRFMC); (**d**) detail-preserving smoothing classifier based on conditional random fields (DPSCRF); (**e**) Markovian support vector classifier (MSVC); (**f**) spatial-spectral fusion based on conditional random fields (SSF-CRF); (**g**) spatial-spectral-location fusion based on conditional random fields (SSLF-CRF).

The quantitative accuracy evaluation for the different algorithms (SVM, MS, SVRFMC, DPSCRF, MSVC, SSF-CRF, and SSLF-CRF) is provided in Table 3. As can be seen from Table 3, when the training samples are 1%, the OA and Kappa values of the algorithms merging spatial and spectral information are improved, compared to the pixel classification method of SVM. In particular, since the SSLF-CRF method combines spatial context, spatial features, spatial location information, and spectral information, the accuracy is improved by about 10%, which indicates the importance of spatial information for improving the classification accuracy. For the misclassification of classes due to spectral variability, the advantages of the SSLF-CRF method are also fully reflected in the accuracy evaluation. For example, the highest precisions for road and gray roof in the other algorithms are 76.42% and 76.88%, respectively, but with the SSLF-CRF method, the precision reaches 82.57% and 82.27%, representing increases of about 6%. That is to say, this algorithm can effectively distinguish the easily confused land-cover classes caused by spectral variation.

Table 3. The classification accuracies for the Hanchuan dataset.

Class	Accuracy (%)						
	SVM	MS	SVRFMC	DPSCRF	MSVC	SSF-CRF	SSLF-CRF
Red roof	49.72	48.89	64.75	49.96	67.43	82.16	84.08
Tree	67.30	73.95	92.47	80.38	84.33	96.12	97.12
Road	65.07	66.77	74.91	62.58	75.39	76.42	82.57
Strawberry	94.55	95.37	97.54	96.89	95.74	98.00	97.96
Pea	64.12	65.49	79.55	67.51	78.37	91.66	92.90
Soy	35.78	29.95	47.81	13.92	78.67	89.26	86.46
Shadow	97.19	97.41	98.84	97.53	97.83	98.07	98.04
Gray roof	53.90	53.67	74.21	64.06	72.05	76.88	82.27
Iron sheet	42.25	43.54	22.07	37.57	43.84	71.97	69.88
OA	85.51	86.41	91.98	87.40	90.91	94.60	95.29
Kappa	0.7757	0.7890	0.8760	0.8043	0.8607	0.9177	0.9286

3.6. Experiment 2: Honghu Dataset

The second group of experiments used a UAV-borne hyperspectral remote sensing image with a 0.4-m spatial resolution acquired in Honghu City, Hubei Province, China, in November 2017. This image has a spatial dimension of 400 × 400 pixels and 270 spectral bands. In these experiments, 18 classes of interest were considered: red roof, bare soil, *rape*, *cotton*, *Chinese cabbage*, *pakchoi*, *cabbage*, *tuber mustard*, *Brassica parachinensis*, *Brassica chinensis*, *small Brassica chinensis*, *Lactuca sativa*, *celtuce*, film-covered *lettuce*, *romaine lettuce*, *carrot*, *white radish*, and *spouting garlic*. The overview of the area is shown in Figure 7a by the true-color image, and the corresponding reference ground truth is shown in Figure 7b. The numbers of randomly selected training and test samples for each class are listed in Table 4.

In order to introduce as little noise as possible while improving the experimental efficiency, the original data were reduced from 270 dimensions to 10 dimensions by MNF. The most suitable spatial features were selected for the classification through a large number of experiments as follows: (1) the four endmembers of bare soil, rape, and film-covered lettuce; (2) the four texture features of homogeneity, dissimilarity, entropy, and mean were extracted with a window of 7×7; and (3) the morphological features were extracted with a disk-shaped SE with a size of 8. The parameters λ, θ, and β were set to 0.4, 3.6, and 0.3, to obtain the best results.

For the experiments with the Honghu dataset, the experimental results corresponding to the different classification algorithms (SVM, MS, SVRFMC, DPSCRF, MSVC, SSF-CRF and SSLF-CRF) are shown in Figure 8a–g, respectively. As with the Hanchuan dataset, the classification result of the SVM algorithm in Figure 8a contains a large amount of salt-and-pepper noise. After considering the spatial context information, the classification effects of the various algorithms are greatly improved. As shown in Figure 8b–g, the salt-and-pepper noise is significantly reduced, and the boundaries of the various types of features are relatively complete. However, several types of crops with similar spectra in the Honghu dataset, such as *romaine lettuce*/film-covered *lettuce* and *pakchoi/rape*, are apparently

misclassified due to the influence of spectral variation. As shown by the red box in Figure 8, the *romaine lettuce* is almost completely classified as film-covered *lettuce* in the classification results of the MS, SVRFMC, DPSCRF, and MSVC algorithms. The crop in the black box is *pakchoi*, but this is misclassified as *rape*, *small Brassica chinensis*, bare soil, and other features by the other algorithms. The SSLF-CRF algorithm considers the spatial features and the spatial location information, so for the above land-cover classes, not only are the types correctly identified, but the boundary shapes and details are also kept intact.
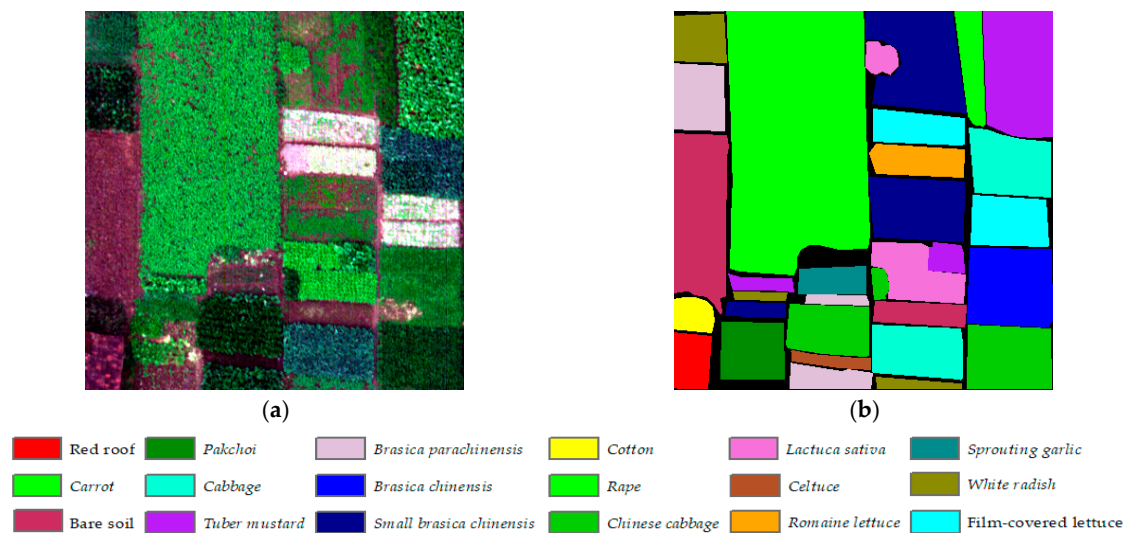
Figure 7. The Honghu dataset: (**a**) the true-color image; (**b**) the ground-truth map.

Table 4. The training and test sample information for the Honghu dataset.

Class	Training Samples	Test Samples	Class	Training Samples	Test Samples
Red roof	22	2182	*Brassica chinensis*	72	7181
Bare soil	118	11692	Small *Brassica chinensis*	159	15769
Cotton	14	1410	*Lactuca sativa*	52	5220
Rape	379	37541	Celtuce	10	993
Chinese cabbage	107	10688	Film-covered lettuce	72	7191
Pakchoi	40	4015	Romaine lettuce	30	2981
Cabbage	103	10204	Carrot	27	2766
Tuber mustard	114	11327	White radish	40	4042
Brassica parachinensis	63	6240	Spouting garlic	20	2046

The accuracy evaluation for the different algorithms (SVM, MS, SVRFMC, DPSCRF, MSVC, SSF-CRF, and SSLF-CRF) is provided in Table 5. From the quantitative accuracy analysis, it can be seen that when the training samples are 1%, compared with the SVM pixel classification algorithm, the methods combining spatial information show an improvement in both OA and Kappa. For example, the accuracy of the object-oriented classification method of MS is improved by about 8%, and the results of the random field based methods of SVRFMC, DPSCRF, MSVC, SSF-CRF, and SSLF-CRF are improved by about 15%, 5%, 5%, 21% and 22%, respectively. Due to the consideration of the spatial location in SSLF-CRF, a better quantitative performance is shown in the classification accuracy. However, for those species with spectral variability, which are difficult to distinguish and easy to misclassify, the poor classification effect is reflected in the accuracy of each class. From Table 5, we can see that SSLF-CRF obtains a satisfactory result for each class, which indicates that the method not only considers the spectral information, but also combines the spatial features and spatial location information, effectively reducing the misclassification of the various classes.
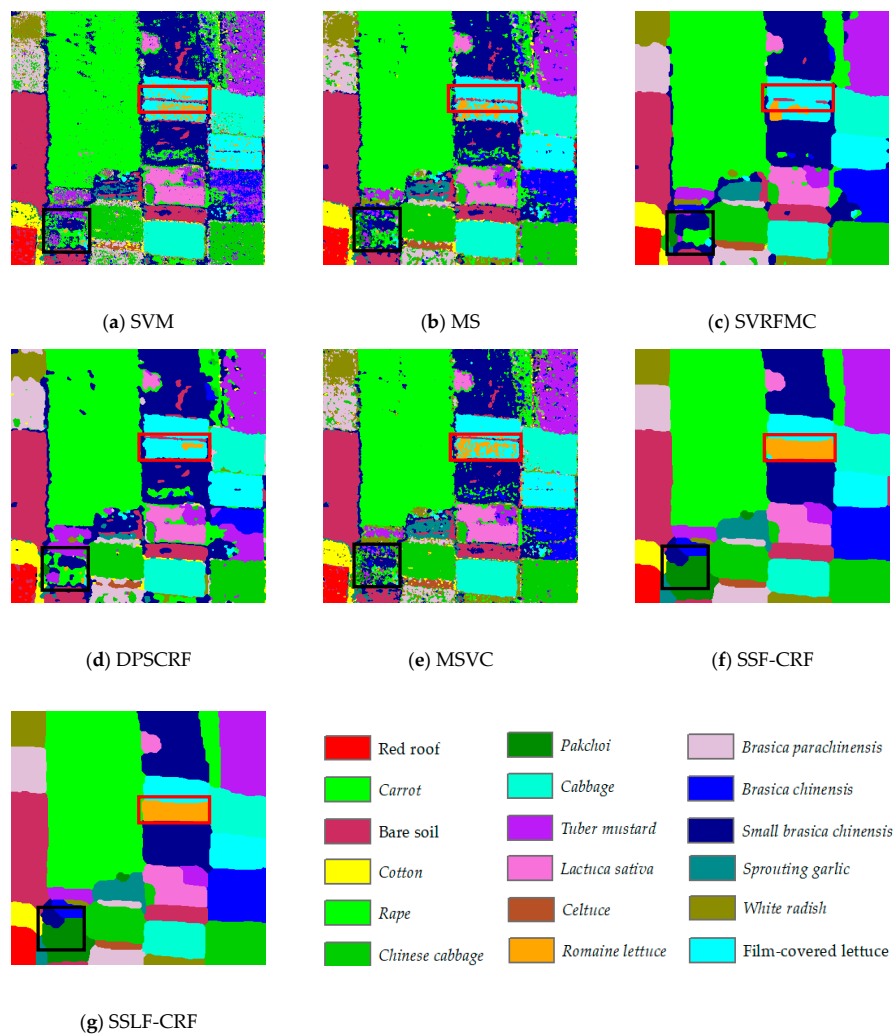
(**a**) SVM　　　　　　(**b**) MS　　　　　　(**c**) SVRFMC

(**d**) DPSCRF　　　　　(**e**) MSVC　　　　　(**f**) SSF-CRF

(**g**) SSLF-CRF

Red roof	Pakchoi	Brasica parachinensis
Carrot	Cabbage	Brasica chinensis
Bare soil	Tuber mustard	Small brasica chinensis
Cotton	Lactuca sativa	Sprouting garlic
Rape	Celtuce	White radish
Chinese cabbage	Romaine lettuce	Film-covered lettuce

Figure 8. Honghu dataset classification results: (**a**) SVM; (**b**) MS; (**c**) SVRFMC; (**d**) DPSCRF; (**e**) MSVC; (**f**) SSF-CRF; (**g**) SSLF-CRF.

Table 5. The classification accuracies for the Honghu dataset.

Class	Accuracy (%)						
	SVM	**MS**	**SVRFMC**	**DPSCRF**	**MSVC**	**SSF-CRF**	**SSLF-CRF**
Red roof	77.59	93.77	99.40	86.16	89.18	98.49	98.53
Bare soil	93.86	94.97	98.12	96.07	94.02	99.66	99.82
Cotton	83.55	95.89	98.58	97.09	91.77	99.01	99.08
Rape	96.19	98.90	99.80	98.11	98.73	99.91	99.91
Chinese cabbage	88.00	94.60	99.00	93.86	93.04	99.44	99.46
Pakchoi	1.79	14.92	13.87	3.76	10.76	87.50	87.90
Cabbage	94.13	97.28	99.30	97.29	96.32	99.57	99.64
Tuber mustard	63.15	77.96	90.17	80.52	70.80	98.54	98.75
Brassica parachinensis	62.36	72.72	93.69	83.51	67.32	97.63	97.63
Brassica chinensis	39.02	66.02	75.20	34.38	65.76	98.45	99.09
Small Brassica chinensis	77.68	82.67	92.68	84.31	83.46	94.98	95.03
Lactuca sativa	71.63	76.38	85.75	74.75	80.65	97.18	97.28
Celtuce	42.30	68.98	87.51	46.02	71.40	78.15	78.45
Film-covered lettuce	88.65	96.37	98.69	97.68	95.61	99.74	99.75
Romaine lettuce	31.23	36.30	27.31	8.45	43.17	95.64	96.04
Carrot	34.89	48.48	82.43	58.68	60.48	95.41	96.28
White radish	51.31	72.64	89.46	59.35	78.33	92.45	92.75
Sprouting garlic	39.20	61.29	82.94	21.80	71.16	97.21	97.17
OA	76.97	84.77	91.08	81.97	84.32	97.95	98.07
Kappa	0.7367	0.8262	0.8985	0.7936	0.8217	0.9768	0.9782

3.7. Sensitivity Analysis for the Training Sample Size

Training sample ratios of 1%, 3%, 5%, and 10% were used for each algorithm with each dataset to validate the SSLF-CRF method proposed in this paper. In all the experiments, we randomly selected 1%, 3%, 5%, and 10% training samples to validate each algorithm, and the remaining 99%, 97%, 95%, and 90% of the samples were used to evaluate the classification accuracy. The classification OAs for the different classification algorithms under different training sample sizes are shown in Figure 9.
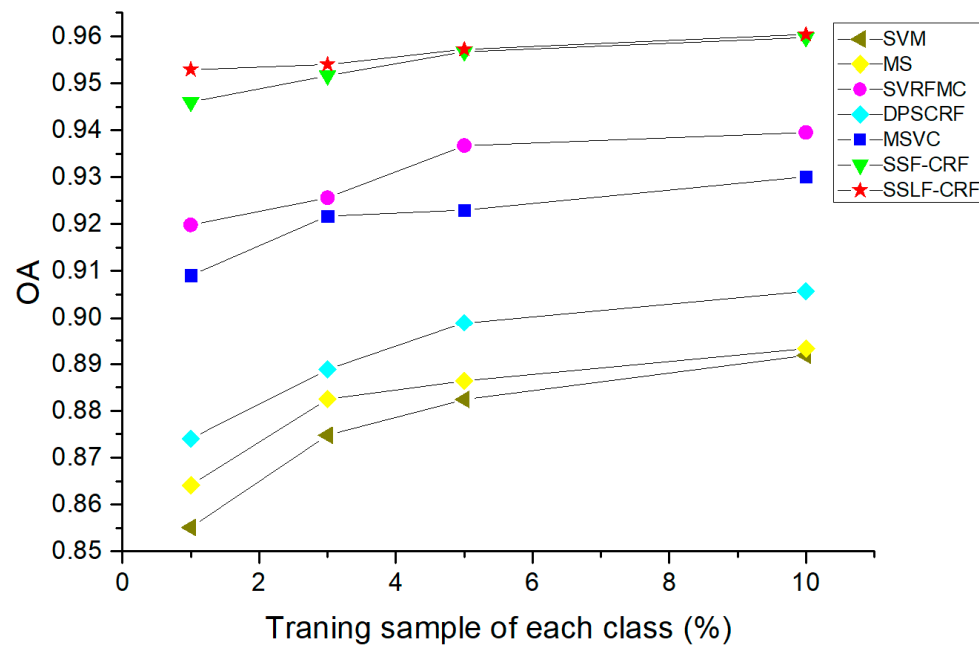
(**a**)
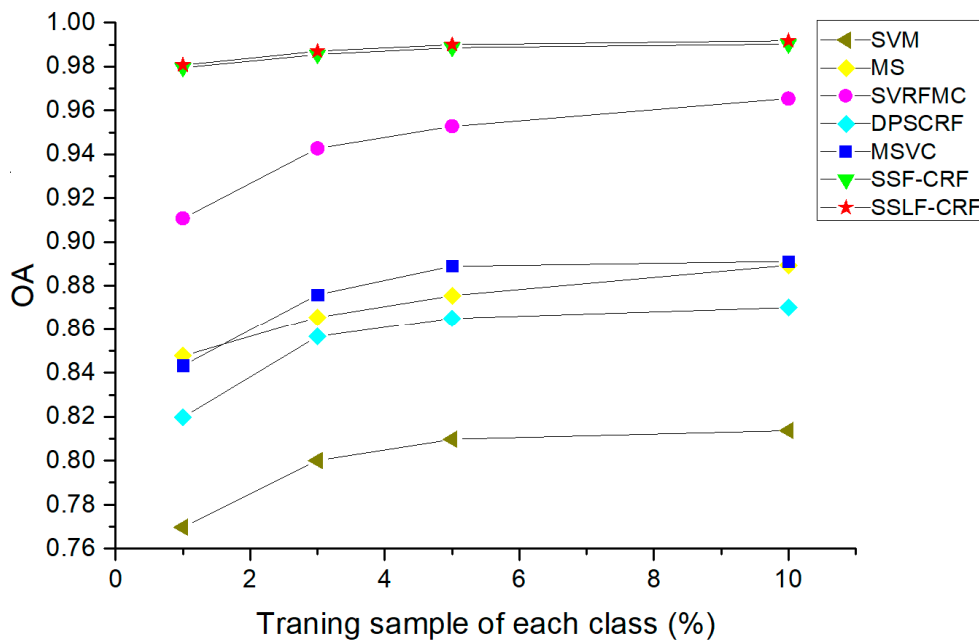
(**b**)

Figure 9. Sensitivity analysis for the training sample size: (**a**) Hanchuan dataset; (**b**) Honghu dataset.

We can see from Figure 9 that as the training samples increase, the accuracies of each algorithm increase significantly, for both datasets. However, given the different sizes of training samples, the performances of each algorithm in the two datasets are not exactly consistent, and are ranked as SSLF-CRF > SSF-CRF > SVRFMC > MSVC > DPSCRF > MS > SVM and SSLF-CRF > SSF-CRF > SVRFMC > MSVC > MS > DPSCRF > SVM, respectively. The object-oriented MS method performs better in the Honghu dataset than in the Hanchuan dataset.

4. Discussion

The important innovation of the proposed method is to integrate the spectral feature, spatial features, spatial context information and spatial location information of hyperspectral images by constructing higher-order potential function in the conditional random field model. The general crop classification methods are relatively simple, which may only use the above single or a few features, so that the image information cannot be fully utilized, and there is a certain room for improvement in classification accuracy. For UAV hyperspectral remote sensing image crop precise classification, spectral information provides the most basic information for their discrimination, based on the spectral differences of various land-cover classes; the spatial feature information provides more details for the classification from the perspective of mathematical morphology, texture and endmember components; the spatial context information takes into account the local spatial interaction of adjacent pixels, and provides spatial smoothing, which can effectively reduce the salt-and-pepper noise; the spatial location information fully exploits the spatial rule information of the image by considering non-local interactions of the same feature, providing auxiliary information for confusing classes. These information can offer complementary information for the discrimination of features from different angles, thus improving the classification effect to a certain extent and heightening the classification accuracy. Therefore, through constructing appropriate potential functions in the CRF model to fuse these features, this method can effectively improve the accuracy of crop precise classification of UAV hyperspectral remote sensing images, compared with the general classification methods.

In terms of sample selection, we randomly selected 1%, 3%, 5%, and 10% of training samples for classification. The remaining 99%, 97%, 95%, and 90% were used as test samples to verify classification accuracy. In the Hanchuan dataset, the accuracy of the proposed method increases with the training samples to a greater extent than with the Honghu dataset. This is because the Honghu dataset is more regular than the Hanchuan dataset. Therefore, even under 1% training samples, the proposed method can obtain a better classification accuracy, due to its strong ability to combine spatial information with spectral information.

In the UAV hyperspectral remote sensing image crop precise classification model of this paper, there are several aspects that can be optimized and further studied in the classification process. First of all, the extraction of spatial feature information is manually selected through a large number of experiments. This process can select the most suitable spatial features to participate in the classification, so that the classification effect is optimal, but it also takes a lot of time. So authors will perform automatic features extraction studies in the follow-up research to increase the efficiency of the method. Secondly, due to the limitation of the endurance of the drone, the experimental datasets obtained by us have small range, so this method is more advantageous when it is applied to small-and medium-scale crop classification applications. However, for common high resolution multi-spectral remote sensing images with rich spatial information, we can also combine them with hyperspectral remote sensing images to make them a data source with high spatial resolution and high spectral resolution. Therefore, it can be applied to the precise classification of crops in a wider range, and improve the applicability and application value of the research results.

5. Conclusions

With the rapid development of UAV technology, UAV-borne high spatial resolution hyperspectral data have become an ideal data source for the precise classification of crops. Based on the advantages

of the high spatial resolution and high spectral resolution of this type of data, we have proposed a spatial-spectral-location fusion higher-order CRF precise crop classification method for UAV-borne hyperspectral remote sensing imagery. The proposed method integrates spectral and spatial feature, spatial context, and spatial location information in the higher-order CRF model, providing complementary information for the land-cover type discrimination from different perspectives, and mining the large-scale spatial local similarity of crops. It not only resolves the limitations of the pairwise CRF model in large-scale spatial context information modeling, but also greatly alleviates the problems of local over-smoothing and spectral variability encountered by the CRF model in the precise classification of crops based on high-resolution hyperspectral remote sensing imagery. The experiments on UAV-borne high spatial resolution hyperspectral remote sensing images of Hanchuan and Honghu (China), confirmed that the SSLF-CRF method shows an excellent classification performance, in terms of both qualitative and quantitative evaluations.

References

1. Huang, J.; Ma, H.; Sedano, F.; Lewis, P.; Liang, S.; Wu, Q.; Su, W.; Zhang, X.; Zhu, D. Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST–PROSAIL model. *Eur. J. Agron.* **2019**, *102*, 1–13. [CrossRef]
2. Huang, J.; Sedano, F.; Huang, Y.; Ma, H.; Li, X.; Liang, S.; Tian, L.; Zhang, X.; Fan, J.; Wu, W. Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. *Agric. For. Meteorol.* **2016**, *216*, 188–202. [CrossRef]
3. Thenkabail, P.S. Global Croplands and their Importance for Water and Food Security in the Twenty-first Century: Towards an Ever Green Revolution that Combines a Second Green Revolution with a Blue Revolution. *Remote Sens.* **2010**, *2*, 2305–2312. [CrossRef]
4. Shen, K.; He, H.; Meng, H.; Guannan, S. Review on Spatial Sampling Survey in Crop Area Estimation. *Chin. J. Agric. Resour. Reg. Plan.* **2012**, *33*, 11–16.
5. Huang, J.; Tian, L.; Liang, S.; Ma, H.; Becker-Reshef, I.; Huang, Y.; Su, W.; Zhang, X.; Zhu, D.; Wu, W. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agric. For. Meteorol.* **2015**, *204*, 106–121. [CrossRef]
6. Huang, J.; Ma, H.; Su, W.; Zhang, X.; Huang, Y.; Fan, J.; Wu, W. Jointly assimilating MODIS LAI and ET products into the SWAP model for winter wheat yield estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4060–4071. [CrossRef]
7. Lowder, S.K.; Skoet, J.; Raney, T. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World Dev.* **2016**, *87*, 16–29. [CrossRef]
8. Du, Z.; Yang, J.; Ou, C.; Zhang, T. Smallholder Crop Area Mapped with a Semantic Segmentation Deep Learning Method. *Remote Sens.* **2019**, *11*, 888. [CrossRef]
9. Qi, J.; Sheng, H.; Yi, P.; Yan, G.; Ren, Z.; Xiang, W.; Yi, M.; Bo, D.; Jian, L. UAV-Based Biomass Estimation for Rice-Combining Spectral, TIN-Based Structural and Meteorological Features. *Remote Sens.* **2019**, *11*, 890.

10. Colomina, I.; Molina, P. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 79–97. [CrossRef]

11. Hugenholtz, C.H.; Moorman, B.J.; Riddell, K.; Whitehead, K. Small unmanned aircraft systems for remote sensing and Earth science research. *Eos Trans. Am. Geophys. Union* **2012**, *93*, 236. [CrossRef]

12. Zhong, Y.; Wang, X.; Xu, Y.; Wang, S.; Jia, T.; Hu, X.; Zhao, J.; Wei, L.; Zhang, L. Mini-UAV-Borne Hyperspectral Remote Sensing: From Observation and Processing to Applications. *IEEE Trans. Geosci. Remote Sens. Mag.* **2018**, *6*, 46–62. [CrossRef]

13. Pajares, G. Overview and Current Status of Remote Sensing Applications Based on Unmanned Aerial Vehicles (UAVs). *Photogramm. Eng. Remote Sens.* **2015**, *81*, 281–330. [CrossRef]

14. Shahbazi, M.; Théau, J.; Ménard, P. Recent applications of unmanned aerial imagery in natural resource management. *GISci. Remote Sens.* **2014**, *51*, 339–365. [CrossRef]

15. Pádua, L.; Marques, P.; Hruška, J.; Adão, T.; Peres, E.; Morais, R.; Sousa, J. Multi-Temporal Vineyard Monitoring through UAV-Based RGB Imagery. *Remote Sens.* **2018**, *10*, 1907. [CrossRef]

16. Zhou, X.; Zheng, H.; Xu, X.; He, J.; Ge, X.; Yao, X.; Cheng, T.; Zhu, Y.; Cao, W.; Tian, Y. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 246–255. [CrossRef]

17. Marcial-Pablo, M.d.J.; Gonzalez-Sanchez, A.; Jimenez-Jimenez, S.I.; Ontiveros-Capurata, R.E.; Ojeda-Bustamante, W. Estimation of vegetation fraction using RGB and multispectral images from UAV. *Int. J. Remote Sens.* **2019**, *40*, 420–438. [CrossRef]

18. Tappen, M.F.; Liu, C.; Adelson, E.H.; Freeman, W.T. Learning Gaussian condition random fields for low-level vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007.

19. Zhao, W.; Du, S.; Wang, Q.; Emery, W.J. Contextually guided very-high-resolution imagery classification with semantic segments. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 48–60. [CrossRef]

20. Bai, J.; Xiang, S.; Pan, C. A Graph-Based Classification Method for Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 803–817. [CrossRef]

21. Zhong, Y.; Lin, X.; Zhang, L. A support vector conditional random fields classifier with a Mahalanobis distance boundary constraint for high spatial resolution remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1314–1330. [CrossRef]

22. Zhong, Y.; Zhao, J.; Zhang, L. A Hybrid Object-Oriented Conditional Random Field Classification Framework for High Spatial Resolution Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7023–7037. [CrossRef]

23. Zhong, P.; Wang, R. Modeling and Classifying Hyperspectral Imagery by CRFs with Sparse Higher Order Potentials. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 688–705. [CrossRef]

24. Zhao, J.; Zhong, Y.; Zhang, L. Detail-preserving smoothing classifier based on conditional random fields for high spatial resolution remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2440–2452. [CrossRef]

25. Wei, L.; Yu, M.; Zhong, Y.; Zhao, J.; Liang, Y.; Hu, X. Spatial–spectral fusion based on conditional random fields for the fine classification of crops in UAV-borne hyperspectral remote sensing imagery. *Remote Sens.* **2019**, *11*, 780. [CrossRef]

26. Lafferty, J.; Mccallum, A.; Pereira, F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *Proc. ICML* **2001**, *3*, 282–289.

27. Kumar, S.; Hebert, M. Discriminative random fields. *Int. J. Comput. Vis.* **2006**, *68*, 179–201. [CrossRef]

28. Wu, T.; Lin, C.; Weng, R. Probability estimates for multi-class classification by pairwise coupling. *J. Mach. Learn. Res.* **2004**, *5*, 975–1005.

29. Chang, C.; Lin, C. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [CrossRef]

30. Li, W.; Feng, F.; Li, H.; Du, Q. Discriminant analysis-based dimension reduction for hyperspectral image classification: A survey of the most recent advances and an experimental comparison of different techniques. *IEEE Geosci. Remote Sens. Mag.* **2018**, *6*, 15–34. [CrossRef]

31. Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [CrossRef]

32. Simard, M.; Saatchi, S.; De Grandi, G. The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest. *IEEE Trans. Geosci. Remote Sens.* **2000**, *38*, 2310–2321. [CrossRef]

33. Anys, H.; He, D.C. Evaluation of textural and multipolarization radar features for crop classification. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 1170–1181. [CrossRef]

34. Laliberte, A.S.; Rango, A. Texture and scale in object-based analysis of subdecimeter resolution Unmanned Aerial Vehicle (UAV) imagery. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 761–770. [CrossRef]

35. Szantoi, Z.; Escobedo, F.; Abd-Elrahman, A.; Smith, S.; Pearlstine, L. Analyzing fine-scale wetland composition using high resolution imagery and texture features. *Int. J. Appl. Earth Obs.* **2013**, *23*, 204–212. [CrossRef]

36. Aguera, F.; Aguilar, F.J.; Aguilar, M.A. Using texture analysis to improve per-pixel classification of very high resolution images for mapping plastic greenhouses. *ISPRS J. Photogramm. Remote Sens.* **2008**, *63*, 635–646. [CrossRef]

37. Fu, Q.; Wu, B.; Wang, X.; Sun, Z. Building extraction and its height estimation over urban areas based on morphological building index. *Remote Sens. Technol. Appl.* **2015**, *30*, 148–154.

38. Zhang, L.; Huang, X. Object-oriented subspace analysis for airborne hyperspectral remote sensing imagery. *Neurocomputing* **2010**, *73*, 927–936. [CrossRef]

39. Maillard, P. Comparing texture analysis methods through classification. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 357–367. [CrossRef]

40. Beguet, B.; Chehata, N.; Boukir, S.; Guyon, D. Classification of forest structure using very high resolution Pleiades image texture. In Proceedings of the 2014 IEEE International Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; Volume 2014, pp. 2324–2327.

41. Gruninger, J.; Ratkowski, A.; Hoke, M. The sequential maximum angle convex cone (SMACC) endmember model. *Proc. SPIE* **2004**, *5425*, 1–14.

42. Pesaresi, M.; Benediktsson, J. A new approach for the Morphological Segmentation of high-resolution satellite imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 309–320. [CrossRef]

43. Benediktsson, J.; Pesaresi, M.; Arnason, K. Classification and feature extraction for remote sensing image from urban areas based on morphological transformations. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1940–1949. [CrossRef]

44. Yu, Q.; Gong, P.; Clinton, N.; Biging, G.; Kelly, M.; Schirokauer, D. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 799–811. [CrossRef]

45. Hu, R.; Huang, X.; Huang, Y. An enhanced morphological building index for building extraction from high-resolution images. *Acta Geod. Cartogr. Sin.* **2014**, *43*, 514–520.

46. Rother, C.; Kolmogorov, V.; Blake, A. 'GrabCut': Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **2004**, *23*, 309–314. [CrossRef]

47. Shotton, J.; Winn, J.; Rother, C.; Criminisi, A. Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; pp. 1–15.

48. Qiong, J.; Landgrebe, D. Adaptive Bayesian contextual classification based on Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2003**, *40*, 2454–2463.

49. Moser, G.; Serpico, S. Combining support vector machines and Markov random fields in an integrated framework for contextual image classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 2734–2752. [CrossRef]

50. Richards, J.; Jia, X. *Remote Sensing Digital Image Analysis: An Introduction*, 4th ed.; Springer: New York, NY, USA, 2006.