High-Resolution Image Classification Integrating Spectral-Spatial-Location Cues by Conditional Random Fields

Ji Zhao, *Student Member, IEEE*, Yanfei Zhong, *Senior Member, IEEE*, Hong Shu, and Liangpei Zhang, *Senior Member, IEEE*

***Abstract*— With the increase in the availability of high-resolution remote sensing imagery, classification is becoming an increasingly useful technique for providing a large area of detailed land-cover information by the use of these high-resolution images. High-resolution images have the characteristics of abundant geometric and detail information, which are beneficial to detailed classification. In order to make full use of these characteristics, a classification algorithm based on conditional random fields (CRFs) is presented in this paper. The proposed algorithm integrates spectral, spatial contextual, and spatial location cues by modeling the probabilistic potentials. The spectral cues modeled by the unary potentials can provide basic information for discriminating the various land-cover classes. The pairwise potentials consider the spatial contextual information by establishing the neighboring interactions between pixels to favor spatial smoothing. The spatial location cues are explicitly encoded in the higher order potentials. The higher order potentials consider the nonlocal range of the spatial location interactions between the target pixel and its nearest training samples. This can provide useful information for the classes that are easily confused with other land-cover types in the spectral appearance. The proposed algorithm integrates spectral, spatial contextual, and spatial location cues within a CRF framework to provide complementary information from varying perspectives, so that it can address the common problem of spectral variability in remote sensing images, which is directly reflected in the accuracy of each class and the average accuracy. The experimental results with three high-resolution images show the validity of the algorithm, compared with the other state-of-the-art classification algorithms.**

***Index Terms*— Conditional random fields, high resolution, image classification, remote sensing, spatial contextual information, spatial location.**

I. Introduction

CLASSIFICATION aims to assign a predefined semantic label to each pixel in an image. This procedure is also known as "semantic segmentation" in computer vision. Over the last few decades, classification using high-resolution remote sensing imagery has gained a lot of attention [1] as a result of the increasing availability of these remote sensing images. These high-resolution images can provide abundant available data for accurate earth observation from the different sensor platforms (such as IKONOS, QuickBird, and ROSIS). The land-cover labeling information of high-resolution remote sensing imagery obtained by classification plays an important role in various application domains, such as urban planning, precision agriculture, security applications, and damage assessment for environmental disasters.

The early classification methods always used a pixel-level processing strategy to assign each pixel a label. These methods, which are also called pixelwise classification methods, independently consider the spectral information of each pixel by the use of a statistical learning algorithm, such as support vector machine (SVM) [2], [3] or maximum likelihood classification (MLC) [4]. To obtain more representative features in the pixelwise classification, dimension reduction or manifold learning methods [5], [6] can be used as a preprocessing step. These pixelwise classification methods have been widely used in many applications, but they always result in a salt-and-pepper classification appearance in high-resolution image classification, since they do not consider the rich spatial contextual information of the images.

In order to improve the performance of classification, many researchers have made a great deal of effort to make full use of the spatial information, and there have been lots of studies of spectral-spatial classification [7], [8]. For example, probabilistic modeling based methods [9], [10] have been successfully used for spectral-spatial classification. For the probabilistic modeling based methods, the probability map is first obtained by pixelwise classification, and it represents the probability of each pixel belonging to each land-cover class. The probability map is then refined by the use of local filtering or an extended random walker method to consider the spatial information.

Another obvious way to consider the spatial information is an object-oriented classification method to enforce smoothness. The object-oriented classification [11] approaches combine the classification and segmentation algorithms to achieve the goal of spectral-spatial classification. The segmentation is first conducted to split an image into homogeneous objects, and then a majority voting strategy [12], [13] based on pixelwise classification or direct classification is used to obtain

Manuscript received November 28, 2015; revised March 29, 2016; accepted May 24, 2016. Date of publication June 7, 2016; date of current version July 1, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 4137134, in part by the Program for Changjiang Scholars and Innovative Research Team in University under Grant IRT1278, in part by the 863 High Technology Program of the People's Republic of China, under Grant 2013AA12A301, and in part by the Fundamental Research Funds for the Central Universities under Grant 2042016kf1035. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. David Clausi. *(Corresponding author: Yanfei Zhong.)*

The authors are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China, and also with the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: zhaoji2015@gmail.com; zhongyanfei@whu.edu.cn; shu_hong@whu.edu.cn; zlp62@whu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2016.2577886

the land-cover labels, taking objects as the basic unit. The segmentation is the key step in object-oriented classification, and can be achieved by many different methods, such as the mean shift segmentation approach (MSS) [14] and the fractal net evolution approach (FNEA) [15]. The object-oriented classification approaches intrinsically provide the spatial information by taking objects as the basic processing unit to alleviate the salt-and-pepper classification noise. However, the selection of the optimal segmentation scale is a challenging task in object-oriented classification, due to the scale diversity of the various land-cover types [16].

In recent years, random field methods have been widely used in remote sensing image classification, and are another useful classification approach that can consider the spatial interaction of pixels. As the most popular random field model, the Markov random fields (MRF) model was first introduced into image analysis in 1984 [17], and has been successfully used in many remote sensing image classification problems in recent years [1], [18], [19]. As an example, the multinomial logistic regression with active learning using a multilevel logistic spatial prior (MLRMLL) method [19] has been successfully used for remote sensing image classification. MLRMLL first learns the class posterior probability with a multinomial logistic regression model. The classification can then be obtained by optimizing the objective function of MRF with a multilevel logistic spatial prior.

MRF is a generative classification framework used to model the joint distribution, and it implies some assumptions about the image data for computational tractability. As the improved model of MRF, conditional random fields (CRF) directly models the posterior distribution, which is what we want to estimate in the classification task. The CRF model is thus more flexible and can incorporate the spatial contextual information in both labels and observed data. CRF was first introduced by Lafferty *et al.* [20] in 2001 for solving the labeling of 1D text sequences, and was then successfully applied in image processing by Kumar and Hebert [21], [22]. In the following years, CRF has been widely used for remote sensing image classification and processing [23]–[30]. For the image classification task, the potential functions are designed in the CRF model to consider the spectral and spatial contextual information. For example, the detail-preserving smoothing classification algorithm based on conditional random fields (DPSCRF) [28] was developed for high-resolution image classification. This approach uses an object-oriented strategy and the current labeling information in the CRF framework to consider the spatial information and preserve the detail information. For the support vector conditional random field classifier incorporating a Mahalanobis distance boundary constraint (SVRFMC) [26], SVM is used as the spectral term and a Mahalanobis distance boundary constraint is used to model the spatial term. These classification methods have been successfully used to overcome salt-and-pepper classification noise by considering the spatial information, and they have shown good classification performances. However, they pay less attention to the common problem of spectral variability in the remote sensing imagery, which may have an effect on the accuracy of some classes due to the confusion with other land-cover types.

In this paper, in order to alleviate the effect of the spectral variability, a conditional random field classification algorithm integrating spectral, spatial contextual, and spatial location cues (CRFSS) is presented for high-resolution images. In the CRFSS algorithm, the spectral, spatial contextual, and spatial location cues are respectively encoded by unary, pairwise, and higher-order potentials. Together, they can provide complementary information from varying perspectives, to improve the classification performance. The CRFSS algorithm is described in detail as follows.

1) The spatial location cues are directly modeled by higher-order potentials. The spectral variability can have a significant influence on the accuracy of certain classes in remote sensing image classification, which shows as the scenarios of "same material with different spectra" and "similar spectra from different materials" in remote sensing images. However, this phenomenon is not likely to happen in a local region, based on the spatial dependency principle. Therefore, spatial location cues in a local region are used to alleviate the effect of spectral variability. The higher-order potentials explicitly model the spatial location to consider the nonlocal range of interactions between the target pixel and its nearest training samples for all the classes, based on the spectral similarity of the same land-cover type in a local region. It should also be noted that the spatial location cues are different from the spatial information used in the spectral-spatial classification algorithms. The spatial location cues can make full use of the regularity of the image to model the nonlocal similarity of the land-cover type by the location information.
2) The spectral and spatial location cues can be fused in the CRFSS algorithm based on the class membership probabilities. The spectral cues are formulated in the unary potentials by the class membership probabilities. The spectral cues are direct and basic information for discriminating the various land-cover classes, due to the difference in their spectra. The spatial location cues encoded by the higher-order potentials can provide more useful information for the classes that are easily confused with other land-cover types, which can also be used to calculate the class membership probabilities from another point of view. The spectral and spatial location cues can provide complementary information about the land-cover type from the spectral and spatial location perspectives, which helps to alleviate the misclassification caused by the spectral similarity in remote sensing images. Therefore, they are integrated in the CRFSS classification algorithm by the potential functions based on the class membership probabilities.
3) The spatial contextual information is also used to overcome salt-and-pepper classification noise and obtain a satisfactory classification by the α-expansion inference algorithm. The CRFSS algorithm inherits the advantages of the CRF model to take the spatial contextual information into account, which is modeled by pairwise

potentials to consider the spatial interactions of neighboring pixels. With the definition of the potential functions in CRFSS to consider the spectral, spatial location, and spatial contextual information, inference by the graph-cut based α-expansion algorithm is performed to predict the optimal labeling by minimizing the energy function.

The efficiency of the proposed CRFSS method was confirmed with three high-resolution datasets: two multispectral images from different sensors (QuickBird and IKONOS) and a hyperspectral image (ROSIS). Compared to the other state-of-the-art classification methods, the experimental results confirm that the CRFSS algorithm has a competitive quantitative and qualitative performance.

The remainder of this paper is organized as follows. The CRF model used in image classification is first briefly introduced in Section II. The description of the CRFSS algorithm for high-resolution imagery is then presented in Section III. In Section IV, the experimental results and analysis are given. The sensitivity analysis is discussed in Section V. Finally, the conclusion is given in Section VI.

II. The CRF Model

To conveniently describe the classification problem, the notations and definitions used throughout the paper are first described. Consider an input high-resolution image $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$, where $\mathbf{y}_i$ is the spectral vector of pixel $i \in V = \{1, 2, \ldots, N\}$, and N is the total number of pixels in the image. A labeled land-cover map is represented as $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$, with x_i taking its value from the label set $\mathbf{L} = \{1, 2, \ldots, K\}$, where K denotes the number of classes.

The image classification problem can be considered as finding the image label set that maximizes the posterior probability $P(\mathbf{x}|\mathbf{y})$, which is what the discriminative approaches make an effort to model. The CRF model used in this paper is a discriminative classification approach, which directly models the posterior probability of the label $\mathbf{x}$ conditioned on the observed image data $\mathbf{y}$ [20], [22] as a Gibbs distribution with the following form:

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} \exp\{-\sum_{c \in C} \psi_c(\mathbf{x}_c, \mathbf{y})\} \quad (1)$$

where Z is the partition function, and C is the set of all the cliques, which represent the set of random variables with spatial connectivity. The function $\psi_c(\mathbf{x}_c, \mathbf{y})$ is the potential function, which is locally defined in the clique c to model the relationship of the random variables. The potential functions can be divided into unary, pairwise potentials, and even higher-order potentials, based on the various types of cliques. For the remote sensing image classification problem, the pairwise CRF model with unary and pairwise potentials is widely used [23], [24], [26]–[30], in which the size of the cliques is no larger than two. The pairwise CRF model has the ability to model the spatial interaction in both the labels and observed values, which is of importance in remote sensing image classification, as described in Section I.

However, the widely used pairwise CRF model has difficulty in capturing more wide-ranging spatial contextual information,

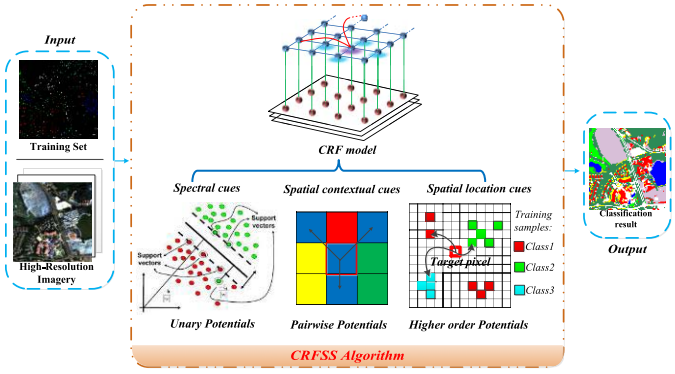


Fig. 1. Flowchart of the conditional random field classification algorithm integrating spectral, spatial contextual, and spatial location cues (CRFSS).

due to the limitation of the size of the cliques. In order to improve the performance of the model, the richer spatial interactions are beneficial for a more accurate model of the label problem, which can be considered in the CRF model by the higher-order potentials. The higher-order potentials are defined on the cliques with multiple random variables [31], so that they can have a higher modeling power. The CRF model with higher-order potentials is also called "higher-order CRF". In computer vision, there have been many successful applications of higher-order random fields in the last few years, such as image restoration [32], [33] and object segmentation [34]–[36]. However, the general higher-order potentials always increase the complexity of the model, and the corresponding inference algorithm also needs to be carefully designed to obtain the optimal result. This limits the wider use of higher-order random fields in remote sensing image classification. For remote sensing image classification, Zhong and Wang [25] used higher-order CRF to consider a smoothing prior in the higher-order cliques, based on the robust P^{N} model [36].

In this paper, the higher-order potentials are used to model the spatial location cues, based on the similarity of the same land-cover type in a local region. The spatial location information can help to alleviate the effect of spectral variability, which is a common problem affecting the accuracies of the classes in remote sensing image classification. Therefore, the higher-order CRF algorithm, integrating the spectral, spatial location, and spatial contextual information in high-resolution imagery classification, is proposed to improve the accuracy of each class and the average accuracy.

III. CRFSS for High-Resolution Imagery

In this paper, in order to consider the spatial contextual information and alleviate the effect of spectral variability, the CRFSS algorithm is proposed. As shown in Fig. 1, the spectral cues, the spatial contextual information, and the spatial location cues are integrated in the CRFSS algorithm by different potentials. Together, they can provide complementary information to distinguish the land-cover types from varying perspectives, and can help to alleviate the impact of spectral variability to improve the performance of remote sensing image classification.

As described in Section II, the CRF model is a discriminative classification framework that is used to directly model the posterior probability $P(\mathbf{x}|\mathbf{y})$, as shown in (1). The corresponding energy can be defined as:

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

The classification problem aims to find the label image $\mathbf{x}$ that maximizes the posterior probability $P(\mathbf{x}|\mathbf{y})$ (i.e., $\mathbf{x}_{MAP} = \text{argmax}_{\mathbf{x}} P(\mathbf{x}|\mathbf{y})$), based on the Bayesian maximum *a posteriori* (MAP) rule. Therefore, the MAP labeling $\mathbf{x}_{MAP}$ of the image classification problem using the CRF model is given by:

$$\mathbf{x}_{MAP} = \arg\max_{\mathbf{x}} P(\mathbf{x}|\mathbf{y}) = \arg\min_{\mathbf{x}} E(\mathbf{x}|\mathbf{y}) \tag{3}$$

The image classification problem using the CRF model is therefore equivalent to finding the minimization of the energy function $E(\mathbf{x}|\mathbf{y})$. In remote sensing image classification, the well-studied pairwise CRF can be formulated as a special case of (2) with the sum of the unary and pairwise potentials:

$$E(\mathbf{x}|\mathbf{y}) = \sum_{i \in V} \psi_i(x_i, \mathbf{y}) + \sum_{i \in V, \mathrm{j} \in \mathbf{N}_i} \psi_{ij}(x_i, x_j, \mathbf{y}) \tag{4}$$

where ψ_i is the unary potential term, and ψ_{ij} is the pairwise potential term, defined over the local neighborhood N_{i} of the site i. By convention, 8-neighborhood connectivity is used to encode the interactions of pairs of neighboring variables in the case of remote sensing image classification.

The pairwise CRF is a simplified model, only consisting of potentials of degree one and two. This can limit the expressive power of CRF in capturing the richer statistics of images. The problem of misclassification of different land-cover types due to spectral variability cannot be efficiently addressed. Therefore, the higher-order potentials, with the ability to encode higher-order spatial information, are employed in the CRF model to improve the classification performance. The higher-order CRF model can then be written as:

$$E(\mathbf{x}|\mathbf{y}) = \sum_{i \in V} \psi_i(x_i, \mathbf{y}) + \sum_{i \in V, \mathrm{j} \in \mathbf{N}_i} \psi_{ij}(x_i, x_j, \mathbf{y}) + \sum_{c \in C} \psi_c(x_c, \mathbf{y}) \tag{5}$$

where C represents the set of higher-order cliques, and ψ_c are the higher-order potentials defined over the cliques.

With the established CRF model (5) for image classification, a remaining problem is how to model the potential functions (i.e., the unary potentials, pairwise potentials, and higher-order potentials) and optimize the objective function to obtain the final labels by an inference algorithm.

A. Unary Potentials

The unary potentials ψ_i model the cost of a single pixel taking a particular land-cover label, and can be formulated by the class membership probabilities of the corresponding pixel. The class membership probability can be computed using a discriminative classifier based on the appearance features in the observed image, which gives a probability estimate of the label x_i taking a particular label. In our work, SVM is adopted to provide the discriminative information of the land-cover types based on the spectral cues, because of its excellent spectral classification performance, especially in the case of a small training set for remote sensing image classification. With this consideration, the used unary potentials can be defined as:

$$\psi_i(x_i, \mathbf{y}) = -\ln\left(P^{SVM}(x_i = l_k|\mathbf{y})\right) \tag{6}$$

where $P^{SVM}(x_i = l_k|\mathbf{y})$ denotes the class membership probabilities of pixel x_i taking label l_k, using SVM [37], [38] based on Platt's formulation. In the unary potentials, the spectral cues are directly used to discriminate the various land-cover types by the class membership probabilities.

B. Pairwise Potentials

The pairwise potentials try to take the prior knowledge of the spatial patterns of the land-cover types into account. The spatial prior knowledge is in accordance with human cognition, and is of importance in classification to help determine the type of certain pixels. The widely used spatial prior knowledge is the spatial smoothing prior in the classification task, based on the spatial dependence principle. The spatial smoothing prior expects that neighboring pixels will take the same label in the classification map, which can effectively alleviate the effect of the spectral noise. The pairwise potentials model this spatial smoothing prior to favor the neighborhood pixels in homogeneous image regions with the same land-cover class, and take the form of [39], [40]:

$$\psi_{ij}(x_i, x_j, \mathbf{y}) = \begin{cases} 0 & if\ x_i = x_j \\ g_{ij}(\mathbf{y}) & otherwise \end{cases} \tag{7}$$

$$g_{ij}(\mathbf{y}) = \lambda(1 + \theta \frac{\exp(-||\mathbf{y}_i - \mathbf{y}_j||^2/\pi)}{||i - j||^2}) \tag{8}$$

where i and j respectively represent the spatial location of neighboring pixels. λ and θ are free parameters to control the strength of the pairwise potentials, and parameter π is set to $(2\langle||\mathbf{y}_i - \mathbf{y}_j||^2\rangle)$, where $\langle.\rangle$ is the average over the image. $g_{ij}(\mathbf{y})$ represents a smoothing term related to the observed data $\mathbf{y}$, which is designed to measure the difference in appearance between the neighboring pixels. Based on the spatial dependence rule, the pairwise potentials model the spatial contextual information between each pixel and its corresponding neighborhood to encourage coherence in regions of similar appearance. Therefore, with the help of the spatial smoothing prior, the pairwise potentials encourage the neighboring pixels to take the same land-cover type and penalize the occurrence of pixels with different labels in the local neighborhood in the classification map.

C. Higher-Order Potentials

The higher-order CRF model extends the pairwise random fields with the incorporation of higher-order potentials to capture richer spatial interactions. In our work, the higher-order potentials directly model the spatial location cues. The different cues can provide useful complementary information about the land-cover types from varying perspectives. The spectral cues modeled by the unary potentials can help to

discriminate the various land-cover classes, due to the differences in their spectra. The spatial contextual information encoded by the pairwise potentials puts a great emphasis on the spatial relationship between neighboring pixels to favor spatial smoothness to help alleviate the salt-and-pepper classification noise. The spatial location in a local region can be used to help distinguish confused land-cover types to alleviate the effect of spectral variability. The spatial location cues are explicitly modeled by the higher-order potentials to consider the nonlocal range of spatial interactions, as follows:

$$\psi_c(x_c, \mathbf{y}) = (\|i-j\|^2/\delta_s + \left\|\mathbf{y}_i - \mathbf{y}_j\right\|^2/\delta_r + K_c) \tag{9}$$

$$K_c = \ln\left(\sum_{k=1}^{K} \exp(\frac{-\|i-j\|^2}{\delta_s}) * \exp(\frac{-\left\|\mathbf{y}_i - \mathbf{y}_j\right\|^2}{\delta_r})\right) \tag{10}$$

$$j = \arg\min_{j'}(||j'-i||^2), \quad j' \in S_k \tag{11}$$

$$k = \arg\min_{k'}(||\mu_{k'} - \mathbf{y}_i||^2), \quad k' \in M \tag{12}$$

$$(\mathbf{S}, \mu) = \arg\min_{\mathbf{S}', \mu'}(\sum_{t=1}^{M}\sum_{j \in S_t'} ||\mathbf{y}_j^l - \mu'_t||^2), \quad l = x_i \in L \tag{13}$$

where each target pixel i and the training samples constitute the higher-order cliques in C. $\mathbf{y}^l$ refers to the training samples with class label l. Formula (13) aims to cluster the training samples $\mathbf{y}^l$ into M pattern sets $\mathbf{S} = \{S_1, S_2, \ldots, S_{\mathrm{M}}\}$, and $\mu = \{\mu_1, \mu_2, \ldots, \mu_{\mathrm{M}}\}$ are the corresponding centers of each pattern. Considering the spectral variability in remote sensing images, the representative training samples of a single land-cover type may have different appearances in the spectral space. Thus, M pattern sets of each class can be obtained using the adaptive mean shift based clustering algorithm [41], which automatically determines the number of clusters. Each pattern of class is then considered to have similar spectra, which can be represented by the center of the pattern. For the computed target pixel i, the pattern that is similar to the target pixel for each class can then be found based on the distance of the spectral vector, as shown in (12). In the similar pattern of each class, (11) tries to search for a candidate pixel of each class by finding the spatially nearest pixel to the target pixel. The obtained candidate pixel of each class can then be used to encode the higher-order potentials by (9), based on its spatial location and spectral values, where i and j respectively represent the spatial locations. The pixel at spatial position i represents the target pixel, and j is the candidate pixel. The higher-order potentials are calculated by the use of the distance in the spatial location domain and in the spectral space. δ_s and δ_r are independent and are respectively set to the second smallest value of the spatial distances and the spectral distances between the target pixel and the corresponding candidate pixels for each class. K_{c} is a normalizing constant.

Although the general higher-order potentials face a severe challenge in facilitating an efficient inference, the specific higher-order potentials used in this paper allow for an efficient inference by the traditional optimization algorithm of pairwise CRF. Since the higher-order potentials encode the spatial interactions of each target pixel and the training samples, they can be rewritten by class membership probabilities and turned into unary potentials. The obtained candidate pixel of each class in the higher-order potentials can take the following form, to calculate the class membership probabilities of the target pixel:

$$\begin{aligned} P^{HP}(x_i = l_k) &= \exp(-\psi_c(x_c, \mathbf{y})) \\ &= \frac{1}{\exp(K_c)} \exp(\frac{-\|i-j\|^2}{\delta_s}) \\ &\quad * \exp(\frac{-\left\|\mathbf{y}_i - \mathbf{y}_j\right\|^2}{\delta_r}) \end{aligned} \tag{14}$$

The higher-order potentials can be readily changed to the class membership probabilities of the target pixel, which are calculated by the use of a Gaussian decreasing function to consider the distance in the spatial location domain and in the spectral space. In practice, the class membership probabilities obtained by the higher-order potentials are integrated in the unary potentials, to efficiently utilize the advantages of the spectral and spatial location cues in the classification. The enhanced unary potentials can then be written as:

$$\psi_i(x_i) = -\ln\left((1-\beta)P^{SVM}(x_i = l_k) + \beta P^{HP}(x_i = l_k)\right) \tag{15}$$

where β is a free parameter from 0 to 1, to make a balance between the spectral information and the spatial location cues emphasized by the higher-order potentials.

To alleviate the effect of spectral variability, the higher-order potentials explicitly model the spatial location to consider the nonlocal range of interactions between the target pixel and its nearest training samples, based on the spectral similarity of the same land-cover type in a local region. Compared to the spectral and spatial contextual information, the spatial location cues have a different focus and can provide complementary information about the land-cover type from the perspective of the nonlocal similarity of the land cover. Therefore, the higher-order potentials can help to distinguish the confused land-cover types through alleviating the spectral variability by encoding the spatial location cues.

D. The Inference of CRFSS

In the modeling of the potential functions in CRFSS, inference is used to optimize the objective function and obtain the final labels, which corresponds to finding the minimum value with respect to the energy function. The exact inference to obtain the global minimum value of the objective function is an NP-hard problem [42] for the multiclass image classification problem. Approximate inference algorithms such as iterated conditional modes, loopy belief propagation, and graph cuts can be used to obtain the classification labels. In this paper, the graph-cut based α-expansion inference algorithm [43] is used, which is an efficient approximate inference algorithm with a strong local minimum property that has performed well in various applications of computer vision [44].

The graph cuts inference method is efficient, and has been proven to be fast and to converge to a global energy minimum for the binary labeling problem. However, the high-resolution image classification problem we are concerned with in this paper always involves multiple class labels. The graph-cut based α-expansion algorithm [43], [45] was developed for energy minimization problems with multivalued variables. It designs a special local search strategy, which works by repeatedly computing the global minimum of a binary labeling problem via a graph-cut method in its inner loops. The binary labeling problem generated at every α-expansion step gives each pixel the following two choices: either keep its current label, or switch to a particular label $\alpha \in \mathbf{L} = \{1, 2, \ldots, K\}$. All the pixels of the image make this choice simultaneously, so there is an exponential number of possible changes with respect to any particular α, which ensures that the algorithm has a strong local minimum property. Therefore, the graph-cut based α-expansion algorithm can be considered to formulate the problem with multiple class labels to a sequence of optimization subproblems with binary labels, which can be easily optimized by the graph-cut method.

IV. Experiments and Analysis

A. Experimental Description

To evaluate the performance of the proposed CRFSS algorithm, experiments with three different high-resolution images were performed. The high-resolution images consisted of two multispectral mages (QuickBird and IKONOS) and a hyperspectral image (ROSIS). In the experiments, the CRFSS algorithm was compared with pixelwise classification algorithms, object-oriented approaches, probabilistic modeling based methods, and random field methods. As a widely used pixelwise classification algorithm, SVM implemented in LIBSVM [37] was used, which uses a radial basis function (RBF) as the kernel function. The object-oriented classification method performs a majority voting strategy for each segmentation region based on the same pixelwise SVM classification map. The segmentation was provided by mean shift segmentation (MSS) [14] and the multi-resolution segmentation algorithm in eCognition 8.0 (FNEA) [15], and the corresponding object-oriented methods are respectively denoted as MSS-OO and FNEA-OO. In the FNEA segmentation, the image layer weights were all set to 1, to ensure an equal level of importance for each band, and the shape and compactness parameters in the composition of the homogeneity criterion were set to the default (0.1 and 0.5). The probabilistic modeling based method used as a comparison algorithm was the method based on an edge-preserving filter (EPF) [9], which refines the probability images obtained by pixelwise SVM classification by the use of an edge-preserving filter. All the random field methods used in our comparison experiments were pairwise random field models that mainly pay attention to the spectral-spatial classification, without explicitly considering the spatial location cues. They included the multinomial logistic regression with active learning using a multilevel logistic spatial prior (MLRMLL) method [19], the detail-preserving smoothing classification algorithm based on conditional random fields (DPSCRF) [28], and a support vector conditional random field classifier incorporating a Mahalanobis distance boundary constraint (SVRFMC) [26].

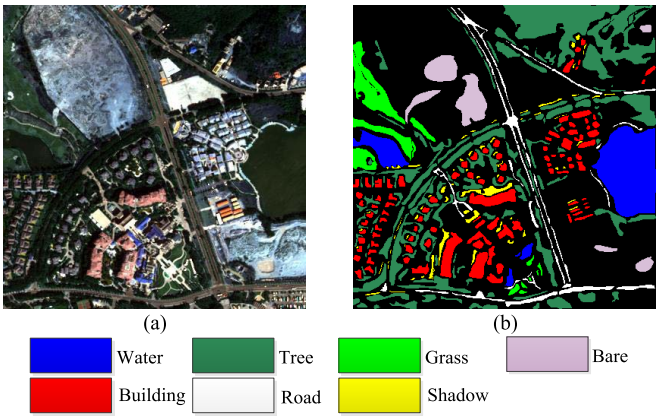


Fig. 2. Fancun QuickBird dataset. (a) RGB false-color image. (b) Ground-truth image.

TABLE I
Class Information of the Fancun QuickBird Image

Class name	Training samples	Test samples
Water	50	9303
Tree	50	25192
Grass	50	4415
Bare	50	4002
Building	50	8183
Road	50	3858
Shadow	50	1606

As in most studies, four kinds of accuracies were used to measure the quantitative performance: the accuracy of each class, the overall accuracy (OA—the percentage of correctly classified pixels), the average accuracy (AA—the average of correctly classified pixels for each class), and the Kappa coefficient (Kappa) [46]. Moreover, all the experiments were conducted using a personal computer of 3.1 GHz with 8 GB of RAM. In the next three sections, the experimental images and the corresponding classification results of the different algorithms with the best OA are presented.

B. Experiment 1: Fancun QuickBird Dataset

The first experiment involved a 2.4-m spatial resolution image from the Fancun area in Hainan province, China. This image was acquired in January 2010 by the QuickBird sensor. It contains 400×400 pixels and four spectral channels, whose appearance is shown in Fig. 2(a) by the use of a false-color image. This Fancun QuickBird dataset includes seven land-cover classes: water, tree, grass, bare, building, road, and shadow, and the distribution of the seven classes is given in Fig. 2(b). A total of 50 training samples for each class were randomly chosen from the reference ground-truth data, and the remaining samples were used to evaluate the accuracy. Table I gives the number of training and test samples for each class.

The classification results for the various classification algorithms (i.e., SVM, FNEA-OO, MSS-OO, EPF, MLRMLL, SVRFMC, DPSCRF, and CRFSS) with the Fancun QuickBird

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2577886, IEEE Transactions on Image Processing

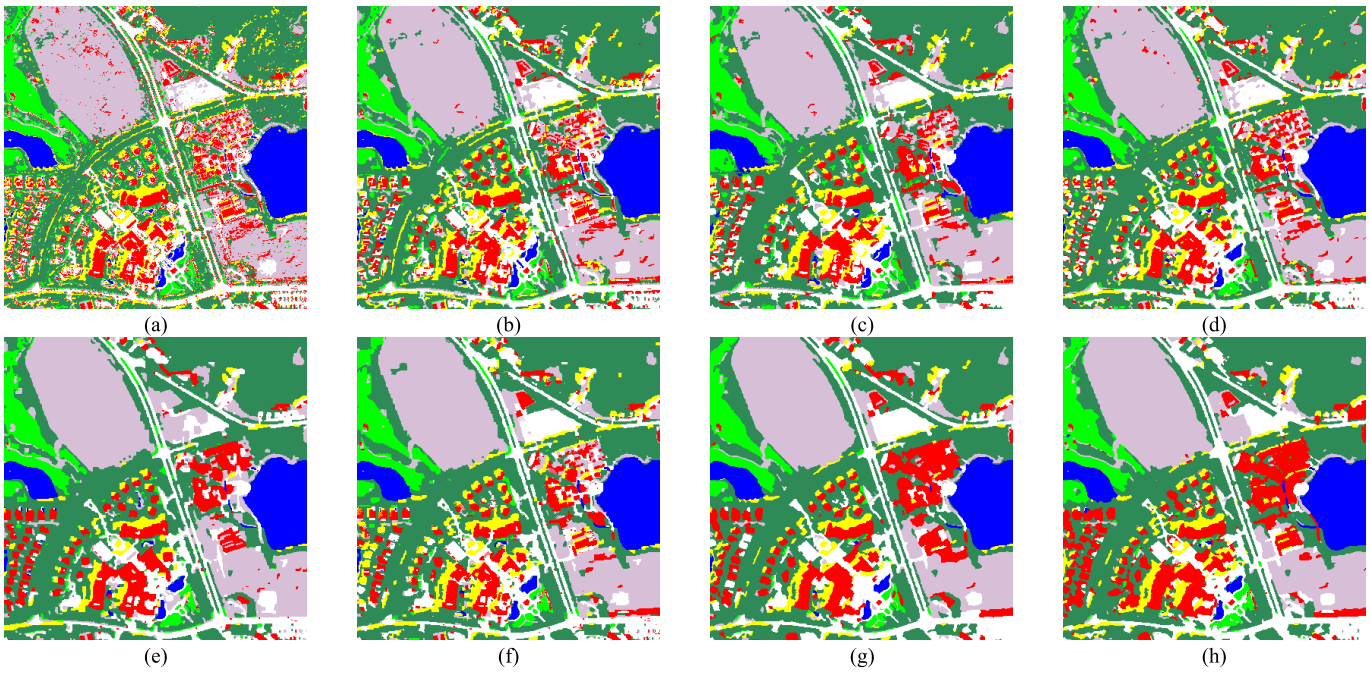


Fig. 3. The classification results for the Fancun QuickBird dataset. (a) SVM. (b) FNEA-OO. (c) MSS-OO. (d) EPF. (e) MLRMLL. (f) SVRFMC. (g) DPSCRF. (h) CRFSS.

TABLE II
CLASSIFICATION ACCURACIES FOR THE FANCUN QUICKBIRD DATASET

Methods	Accuracy (%)							OA (%)	AA (%)	Kappa
	Water	Tree	Grass	Bare	Building	Road	Shadow			
SVM	98.80	93.32	94.16	98.13	67.02	91.52	95.45	90.76	91.20	0.8768
MSS-OO	99.30	95.97	98.60	99.48	69.06	95.46	87.48	92.80	92.19	0.9034
FNEA-OO	98.86	95.85	93.73	99.33	71.04	91.96	97.14	92.61	92.56	0.9008
EPF	99.25	98.39	95.11	99.90	71.04	95.26	96.89	94.17	93.69	0.9213
MLRMLL	99.45	98.79	94.43	99.78	82.04	94.84	77.58	95.33	92.42	0.9368
SVRFMC	99.55	98.36	94.36	99.58	71.36	94.76	96.45	94.12	93.49	0.9207
DPSCRF	99.51	98.94	94.65	99.65	86.25	96.06	93.71	96.57	95.54	0.9535
CRFSS	99.72	98.43	99.30	99.63	91.26	95.02	94.77	97.42	96.88	0.9652

dataset are respectively shown in Fig. 3(a)–(h). As shown in Fig. 3(a), SVM exhibits a salt-and-pepper classification appearance due to the lack of consideration for the spatial contextual information. By considering the spatial information to alleviate the effect of noise, the spectral-spatial classification algorithms can deliver a smoother classification map. Although these spectral-spatial classification algorithms have a better visual performance than the pixelwise SVM method, their classification performances in keeping useful land-cover details are different, due to the effect of the spectral variability. The spectral variability always directly affects the accurate interpretation of the various land-cover classes. As shown in Fig. 2(a), the building land-cover type has several varying spectral appearances in the Fancun dataset, and some of them have a certain spectral similarity to the road class. Due to the effect of this spectral variability, it is a challenging task to accurately discriminate between these spectrally similar classes. For example, the road and building classes present in the Fancun QuickBird image are not easy to accurately classify with the other classification algorithms, which can result in a discontinuous shape or some classification noise. However, on the whole, the proposed CRFSS algorithm can achieve a complete shape feature with fewer meaningless regions and less salt-and-pepper classification noise. Therefore, the CRFSS algorithm has a competitive visual performance, and can both alleviate the salt-and-pepper classification noise and preserve useful boundary detail information.

To better evaluate the effectiveness of the classification methods, the quantitative results of the various algorithms (i.e., SVM, FNEA-OO, MSS-OO, EPF, MLRMLL, SVRFMC, DPSCRF, and CRFSS) are reported in Table II. As can be seen in Table II, the spectral-spatial classification algorithms show a great improvement of more than 2% over the pixelwise SVM classification in terms of OA and Kappa. This demonstrates that the spatial contextual information is very effective in the classification. Since the proposed algorithm also takes the spatial location cues into account, it leads to a better quantitative performance in the classification accuracies

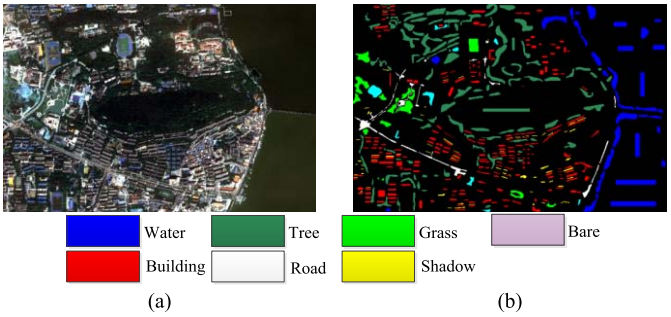


Fig. 4. Wuhan IKONOS dataset. (a) RGB false-color image. (b) Ground-truth image.

TABLE III
CLASS INFORMATION OF THE WUHAN DATASET

Class name	Training samples	Test samples
Building	50	6814
Grass	50	2788
Water	50	7720
Shadow	50	847
Bare soil	50	1080
Tree	50	13946
Road	50	1757

(OA, AA, and Kappa indexes). The spectral variability results in some classes being misclassified, which is directly reflected in the accuracy of each class and the AA. For the various land-cover types, CRFSS obtains a satisfactory accuracy for each class, which illustrates that the confused classes resulting from the spectral variability can also be correctly classified. In addition, the CRFSS algorithm takes only 6 s in this experiment, since the graph-cut based α-expansion inference algorithm is very efficient and is related to the size of the image and the number of classes. These experimental results demonstrate that the proposed CRFSS algorithm has a competitive classification performance for the Fancun QuickBird dataset.

C. Experiment 2: Wuhan IKONOS Dataset

A high-resolution image of a different sensor was used in the second experiment to assess the performance of the classification algorithms. The Wuhan IKONOS urban area image is of a high-resolution remote sensing image with a 4-m spatial resolution that was recorded by the IKONOS sensor over Wuhan in Hubei province, China. The image has a spatial dimension of 400 by 600 pixels and four spectral bands. An overview of this image is given in Fig. 4(a). As with the Fancun QuickBird dataset, this image also contains seven thematic classes, and their corresponding distribution is shown in Fig. 4(b). Again, 50 training samples for each class were randomly chosen from the reference ground-truth data to test the performance of the algorithm in the case of limited training samples. Table III reports the numbers of training and test samples in this image.

The classification results obtained by the various algorithms (SVM, FNEA-OO, MSS-OO, EPF, MLRMLL, SVRFMC, DPSCRF, and CRFSS) are shown in Fig. 5 for the Wuhan dataset, and the corresponding quantitative performances are reported in Table IV. As with the first experiment, the spectral-spatial classification algorithms can deliver smoother classification maps and an improvement in the classification accuracy, since they take the spatial contextual information into account. However, the spectral-spatial classification algorithms tend to greatly reduce the classification noise by the spatial smoothing, so that they may also eliminate some potentially useful details. In this experiment, the spectral difference between the road and building land-cover types is very small in this four-band high-resolution image, so that they are easily confused and are wrongly classified. As shown in Table IV, the accuracies of the road and building land-cover classes are limited for the spectral-spatial classification algorithms considering the spectral and local spatial interactions, including the pairwise CRF algorithms (MLRMLL, SVRFMC, and DPSCRF). However, the proposed CRFSS algorithm not only considers the spectral and spatial contextual information, but also fuses the spatial location cues, which can effectively alleviate the misclassification resulting from the spectral variability. Therefore, CRFSS shows an improvement for the road and building classes, both in the visual result and the quantitative metrics. In addition, the computation time of the CRFSS algorithm is only 8 s in this experiment. Overall, the proposed classification algorithm has a competitive performance in terms of both visualization and the quantitative metrics, especially in terms of AA.

D. Experiment 3: Pavia Center ROSIS Dataset

A high spatial-spectral resolution image was used in the third experiment. This is an urban image from the center of Pavia, Italy, acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) optical sensor. The image has a spatial resolution of 1.3 m and a size of 1096 by 715 pixels. The number of used spectral bands was 102. In this experiment, nine classes of interest were considered. The overview of the area is shown in Fig. 6(a) by the false-color image, and the corresponding reference ground truth is shown in Fig. 6(b). The numbers of randomly selected training and test samples for each class of interest are reported in Table V.

The classification maps obtained by the various algorithms (SVM, FNEA-OO, MSS-OO, EPF, MLRMLL, SVRFMC, DPSCRF, and CRFSS) in this experiment are respectively presented in Fig. 7, and the corresponding quantitative performances are reported in Table VI. As shown in Table VI, the accuracies of the classes for all the classification algorithms are higher than in the former experiments, since there are many spectral bands that lay the foundation to distinguish these land-cover types. In addition, considering the limitation of computer memory and the effectiveness of multispectral high-resolution images in the object-oriented classification, the seven bands are used by the averages of the contiguous spectral bands [47]. For the object-oriented classification methods, MSS-OO and FNEA-OO obtain a satisfactory visual result while keeping the obvious boundary information. For the

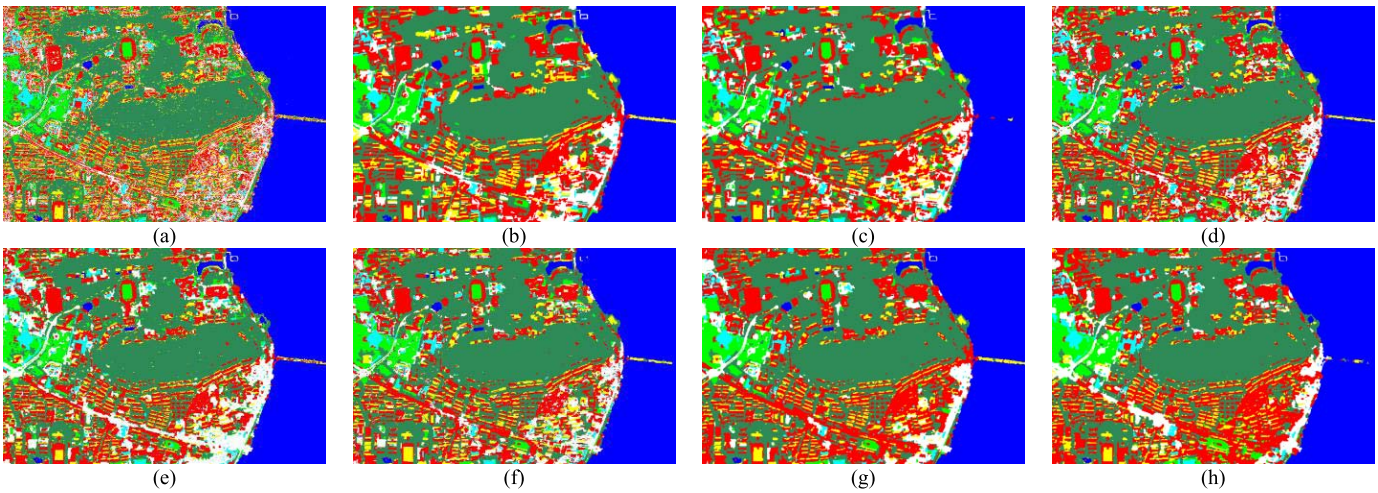


Fig. 5. The classification results for the Wuhan IKONOS dataset. (a) SVM. (b) FNEA-OO. (c) MSS-OO. (d) EPF. (e) MLRMLL. (f) SVRFMC. (g) DPSCRF. (h) CRFSS.

TABLE IV
CLASSIFICATION ACCURACIES FOR THE WUHAN IKONOS DATASET

Methods	Accuracy (%)							OA (%)	AA (%)	Kappa
	Building	Grass	Water	Shadow	Bare	Tree	Road			
SVM	64.19	92.65	97.72	94.33	88.98	95.22	70.06	88.04	86.16	0.8421
MSS-OO	71.10	94.87	98.85	91.62	96.85	98.59	74.62	91.56	89.50	0.8874
FNEA-OO	69.00	93.76	98.99	93.03	96.11	98.37	76.89	91.14	89.45	0.8819
EPF	72.79	95.44	98.64	98.11	96.20	98.82	70.06	91.89	90.01	0.8920
MLRMLL	79.45	93.79	99.97	95.87	93.98	97.73	81.84	93.39	91.80	0.9118
SVRFMC	69.03	95.37	99.00	98.35	96.30	99.04	69.72	91.31	89.54	0.8843
DPSCRF	85.98	95.91	99.00	97.17	89.35	99.64	64.77	94.41	90.26	0.9249
CRFSS	88.55	99.07	99.83	97.87	96.76	98.74	89.70	96.48	95.79	0.9530

TABLE V
CLASS INFORMATION OF THE PAVIA CENTER DATASET

Class name	Training samples	Test samples
Water	50	65921
Trees	50	7548
Meadows	50	3040
Bricks	50	2635
Bare soil	50	6534
Asphalt	50	9198
Bitumen	50	7237
Tiles	50	42776
Shadow	50	2813

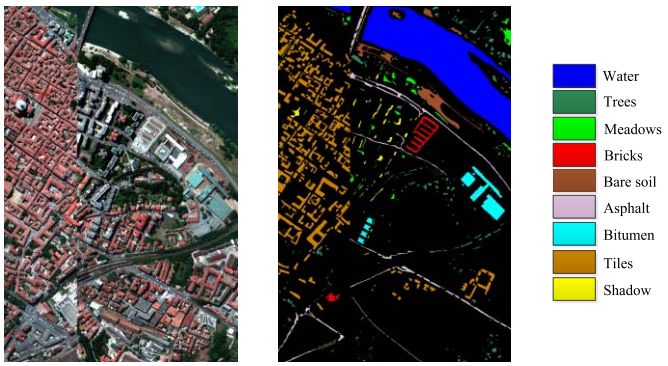


Fig. 6. Pavia Center ROSIS dataset. (a) RGB false-color image. (b) Ground-truth image

random field methods of MLRMLL, DPSCRF, and SVRFMC, the varying degrees of over-smooth visual performance can be clearly seen in the classification maps of Fig. 7(e)–(g). The CRFSS algorithm has a competitive visual performance, and it retains as much boundary shape as possible. In this experiment, all the classification algorithms obtain a high degree of accuracy, but the CRFSS algorithm shows an improvement in the quantitative metrics. For instance, the AA of CRFSS is equal to 99.52%, which demonstrates that all the land-cover types can be correctly classified in the CRFSS algorithm by integrating the spectral, spatial contextual, and spatial location information. Therefore, the CRFSS algorithm can effectively alleviate the effect of the spectral variability. For the computation time, the CRFSS algorithm is also acceptable at 33 s. Overall, the experimental results demonstrate that the CRFSS algorithm has a competitive performance for the Pavia Center ROSIS dataset.

V. SENSITIVITY ANALYSIS

In all the previous experiments, the CRFSS algorithm obtained a good classification performance. However, the three parameters of CRFSS play an important role in the classification result, and they are respectively denoted as β,

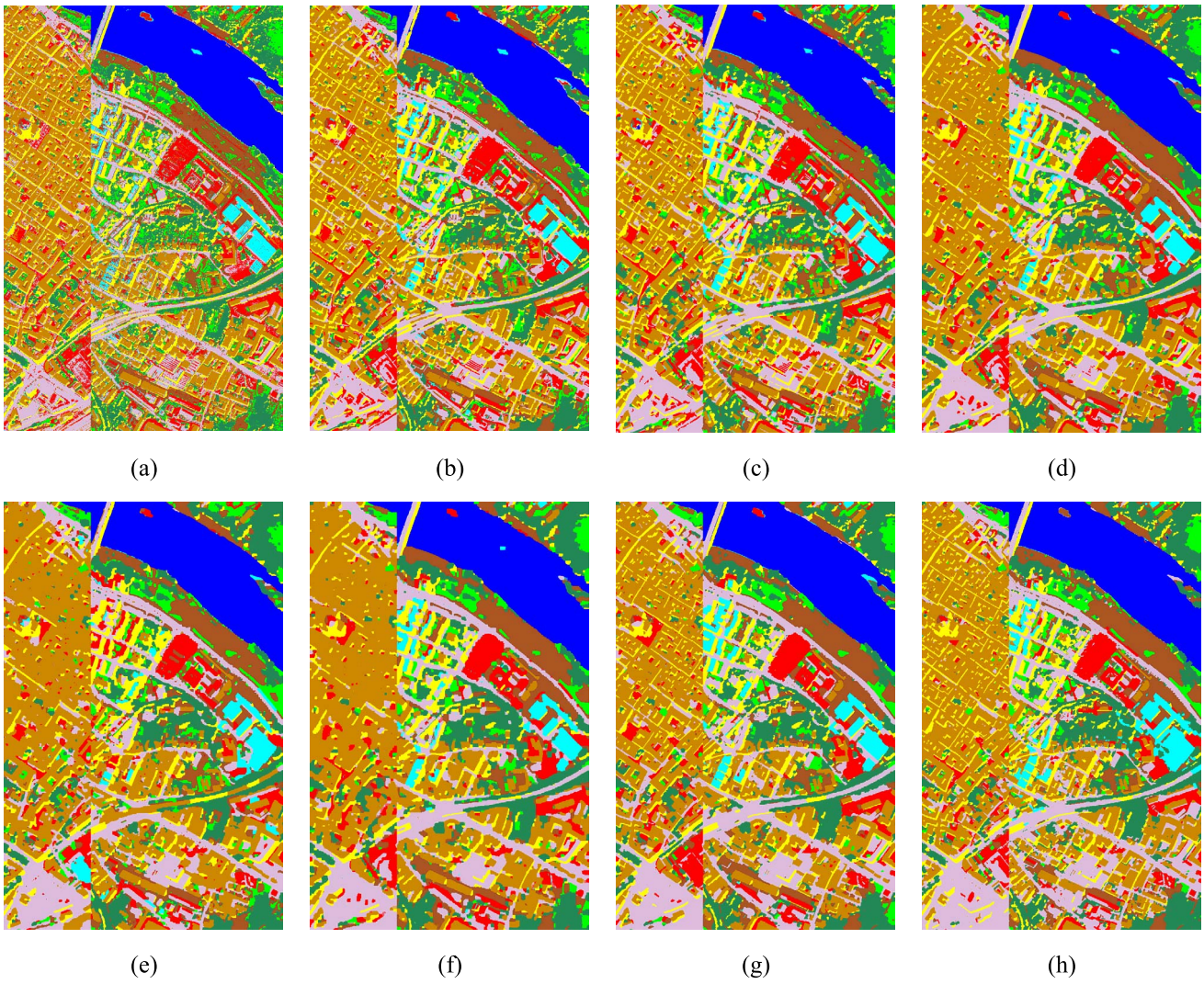


Fig. 7. The classification results for the Pavia Center ROSIS dataset. (a) SVM. (b) FNEA-OO. (c) MSS-OO. (d) EPF. (e) MLRMLL. (f) SVRFMC. (g) DPSCRF. (h) CRFSS.

TABLE VI
CLASSIFICATION ACCURACIES FOR THE PAVIA CENTER ROSIS DATASET

Methods	Accuracy (%)									OA (%)	AA (%)	Kappa
	Water	Trees	Meadows	Bricks	Bare soil	Asphalt	Bitumen	Tiles	Shadow			
SVM	99.61	91.56	96.55	94.08	86.65	98.02	87.81	98.63	100.00	97.51	94.77	0.9725
MSS-OO	99.68	94.29	98.75	99.73	91.49	98.78	96.71	99.54	100.00	98.79	97.66	0.9866
FNEA-OO	99.57	95.59	97.63	98.56	92.73	98.75	93.48	99.13	99.79	98.53	97.25	0.9838
EPF	99.73	95.76	97.27	100.00	95.15	99.79	95.32	99.93	100.00	99.13	98.11	0.9876
MLRMLL	100.00	96.78	94.64	99.92	96.30	99.57	89.82	98.71	97.87	98.62	97.07	0.9805
SVRFMC	99.80	95.79	96.12	100.00	94.84	99.73	95.16	99.94	98.15	99.08	97.73	0.9898
DPSCRF	100.00	96.85	94.08	100.00	96.22	100.00	96.59	99.96	99.89	99.37	98.18	0.9910
CRFSS	100.00	96.36	99.41	99.96	100.00	100.00	100.00	99.91	100.00	99.78	99.52	0.9975

λ, and θ. Parameter β mainly controls the effect of the spatial location cues provided by the higher-order potentials in the classification, while the λ and θ parameters control the spatial smoothing prior. In this section, a corresponding sensitivity analysis is given to test the effect of these parameters in the CRFSS algorithm. Therefore, additional experiments were conducted to analyze the effect of the parameters with the Fancun, Wuhan, and Pavia Center datasets.

A. Sensitivity Analysis for the λ and θ Parameters

In order to study the sensitivity of the λ and θ parameters for the CRFSS algorithm, parameter β was set to be a constant. In the additional experiments, it was set to 0.4, 0.3, and 0.3

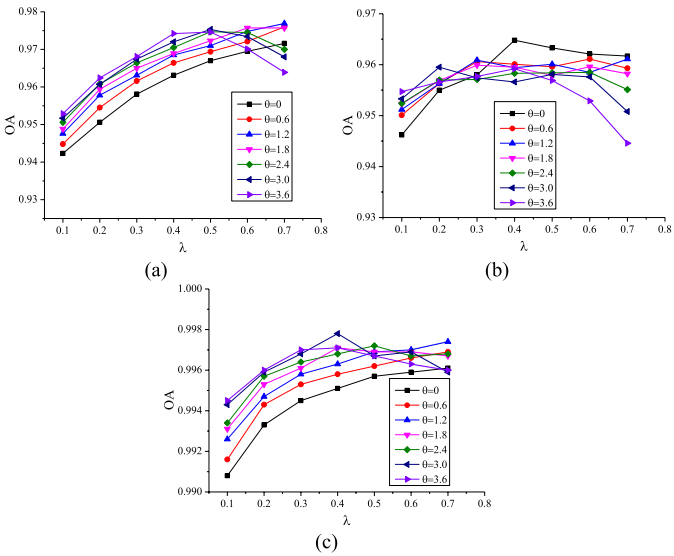


Fig. 8. Sensitivity analysis for the λ and θ parameters of CRFSS with the three datasets. (a) Fancun QuickBird image. (b) Wuhan IKONOS image. (c) Pavia Center ROSIS image.

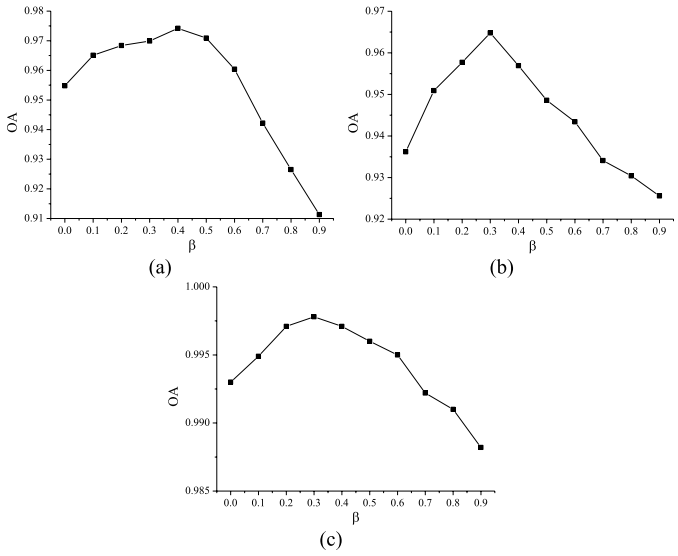


Fig. 9. Sensitivity analysis for parameter β of CRFSS with the three datasets. (a) Fancun QuickBird image. (b) Wuhan IKONOS image. (c) Pavia Center ROSIS image.

for the Fancun, Wuhan, and Pavia Center datasets, respectively. Parameter λ was varied from 0.1 to 0.7, with an interval of 0.1. Parameter θ was selected from 0 to 3.6, with an interval of 0.6. The sensitivity analysis for the λ and θ parameters of CRFSS is presented in Fig. 8.

As can be observed from Fig. 8, the λ and θ parameters have a great impact on the classification accuracies of CRFSS. The λ and θ parameters mainly control the effect of the pairwise potential term to consider the spatial contextual information. When keeping θ unchanged, the relative importance of the pairwise potential term is increased as parameter λ increases. Therefore, the classification accuracy first increases since the neighborhood spatial interactions can be properly utilized to alleviate the effect of noise. When parameter λ reaches a certain value, the accuracy no longer improves, and even shows a slight decreasing trend, since the large spatial smoothing effect can lead to varying degrees of over-smooth performance. For parameter θ, it has a certain fine-tuning function for the classification accuracy when keeping parameter λ constant, as presented in Fig. 8.

B. Sensitivity Analysis for Parameter β

The effect of parameter β for the CRFSS algorithm was examined with the Fancun, Wuhan, and Pavia Center datasets. In the additional experiments, the λ and θ parameters were set to be constant values. In particular, they were set to 0.4/3.6, 0.4/0, and 0.4/3.0 for the Fancun, Wuhan, and Pavia Center datasets, respectively. Parameter β was selected from 0 to 0.9, with an interval of 0.1. The trends of the classification accuracies (OA) with parameter β are illustrated in Fig. 9.

As shown in Fig. 9, parameter β has an influence on the classification accuracy of CRFSS. Parameter β mainly controls the importance of the spatial location emphasized by the higher-order potentials. The classification accuracy of CRFSS gradually increases in the beginning with the increase in parameter β, since the spatial location cues are properly considered. After parameter β reaches a certain value, the spatial location cues become dominant when compared with the spectral information in the unary potential. Therefore, the classification accuracy of the CRFSS algorithm tends to decrease with an increase in parameter β for all three image classifications, as shown in Fig. 9.

VI. Conclusion

In this paper, a conditional random field classification algorithm integrating spectral, spatial contextual, and spatial location cues is proposed for high-resolution images. These different cues each have a different focus and can provide useful and complementary information about the land-cover types from varying perspectives. Therefore, they are respectively formulated in unary, pairwise, and higher-order potentials within the unified probabilistic CRF framework, to alleviate the misclassification caused by spectral variability. The experiments with three real high-resolution remote sensing images confirm the competitive classification performance of the proposed algorithm in both the quantitative and qualitative evaluations. However, further improvements aimed at preserving more potentially useful details will be considered in the future. In addition, the use of spatial location cues in other remote sensing techniques such as semisupervised learning classification will also be considered.

Acknowledgment

The authors would like to thank the editor, associate editor and anonymous reviewers for their helpful comments and suggestions. They also would like to thank Beijing Panorama Space Technology Co., Ltd., for providing the free QuickBird image and Prof. P. Gamba, University of Pavia, Italy, for providing the ROSIS data.

REFERENCES

[1] G. Moser, S. B. Serpico, and J. A. Benediktsson, "Land-cover mapping by Markov modeling of spatial–contextual information in very-high-resolution remote sensing images," *Proc. IEEE*, vol. 101, no. 3, pp. 631–651, Mar. 2013.

[2] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[3] Y. Zhong, Q. Zhu, and L. Zhang, "Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6207–6222, Nov. 2015.

[4] J. D. Paola and R. A. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 4, pp. 981–996, Jul. 1995.

[5] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 260–274, Feb. 2009.

[6] T. Zhou and D. Tao, "Double shrinking sparse dimension reduction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 244–257, Jan. 2013.

[7] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[8] K. Schindler, "An overview and comparison of smooth labeling methods for land-cover classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4534–4545, Nov. 2012.

[9] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.

[10] X. Kang, S. Li, L. Fang, M. Li, and J. A. Benediktsson, "Extended random walker-based classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 144–153, Jan. 2015.

[11] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogram. Remote Sens.*, vol. 65, no. 1, pp. 2–16, Jan. 2010.

[12] Y. Zhong, B. Zhao, and L. Zhang, "Multiagent object-based classifier for high spatial resolution imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 841–857, Feb. 2014.

[13] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, 2010.

[14] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[15] M. Baatz and A. Schäpe, "Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation," in *Angewandte Geographische Informationsverarbeitung XII*, J. Strobl, T. Blaschke, and G. Griesebner, Eds. Heidelberg, Germany: Wichmann-Verlag, 2000, pp. 12–23.

[16] B. Johnson and Z. Xie, "Unsupervised image segmentation evaluation and refinement using a multi-scale approach," *ISPRS J. Photogram. Remote Sens.*, vol. 66, no. 4, pp. 473–483, Jul. 2011.

[17] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[18] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.

[19] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.

[20] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. Mach. Learn.*, 2001, pp. 282–289.

[21] S. Kumar and M. Hebert, "Discriminative random fields: A discriminative framework for contextual interaction in classification," in *Proc. 9th Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1150–1157.

[22] S. Kumar and M. Hebert, "Discriminative random fields," *Int. J. Comput. Vis.*, vol. 68, no. 2, pp. 179–201, 2006.

[23] G. Zhang and X. Jia, "Simplified conditional random fields with class boundary constraint for spectral-spatial based remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 856–860, Sep. 2012.

[24] P. Zhong and R. Wang, "Learning conditional random fields for classification of hyperspectral images," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1890–1907, Jul. 2010.

[25] P. Zhong and R. Wang, "Modeling and classifying hyperspectral imagery by CRFs with sparse higher order potentials," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 2, pp. 688–705, Feb. 2011.

[26] Y. Zhong, X. Lin, and L. Zhang, "A support vector conditional random fields classifier with a Mahalanobis distance boundary constraint for high spatial resolution remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 7, no. 4, pp. 1314–1330, Apr. 2014.

[27] Y. Zhong, J. Zhao, and L. Zhang, "A hybrid object-oriented conditional random field classification framework for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7023–7037, Nov. 2014.

[28] J. Zhao, Y. Zhong, and L. Zhang, "Detail-preserving smoothing classifier based on conditional random fields for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2440–2452, May 2015.

[29] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.

[30] J. Zhao, Y. Zhong, Y. Wu, L. Zhang, and H. Shu, "Sub-pixel mapping based on conditional random fields for hyperspectral remote sensing imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1049–1060, Sep. 2015.

[31] P. Kohli and C. Rother, "Higher-order models in computer vision," in *Image Processing and Analysis with Graphs*. Boca Raton, FL, USA: CRC Press, 2012, pp. 1–28.

[32] Y. Li and D. P. Huttenlocher, "Sparse long-range random field and its application to image denoising," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 344–357.

[33] S. Roth and M. J. Black, "Fields of experts," *Int. J. Comput. Vis.*, vol. 82, no. 2, pp. 205–229, Apr. 2009.

[34] P. Kohli, M. P. Kumar, and P. H. S. Torr, "P3 & beyond: Solving energies with higher order cliques," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[35] S. Vicente, V. Kolmogorov, and C. Rother, "Joint optimization of segmentation and appearance models," in *Proc. 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 755–762.

[36] P. Kohli and P. H. S. Torr, "Robust higher order potentials for enforcing label consistency," *Int. J. Comput. Vis.*, vol. 82, no. 3, pp. 302–324, May 2009.

[37] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, Apr. 2011, Art. no. 27.

[38] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *J. Mach. Learn. Res.*, vol. 5, pp. 975–1005, Dec. 2004.

[39] C. Rother, V. Kolmogorov, and A. Blake, "'GrabCut': Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, Aug. 2004.

[40] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "*TextonBoost*: Joint appearance, shape and context modeling for multi-class object recognition and segmentation," in *Proc. 9th Eur. Conf. Comput. Vis.*, 2006, pp. 1–15.

[41] B. Georgescu, I. Shimshoni, and P. Meer, "Mean shift based clustering in high dimensions: A texture classification example," in *Proc. 9th Int. Conf. Comput. Vis.*, Oct. 2003, pp. 456–463.

[42] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, 3rd ed. New York, NY, USA: Springer-Verlag, 2009.

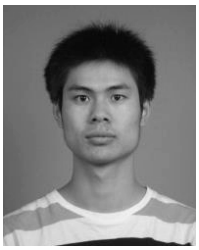
[43] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[44] R. Szeliski *et al.*, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008.

[45] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.

[46] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*, 4th ed. New York, NY, USA: Springer-Verlag, 2006.

[47] G. Licciardi *et al.*, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3857–3865, Nov. 2009.



Ji Zhao (S'14) received the B.S. degree in surveying from the Xi'an University of Science and Technology, Xi'an, China, in 2011. He is currently pursuing the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, China. His major research interests include high spatial resolution remote sensing image classification, scene analysis, and random field algorithms.

Yanfei Zhong (M'11–SM'15) received the B.S. degree in information engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, China, in 2002 and 2007, respectively. He has been with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University since 2007, where he is currently a Full Professor. His research interests include multi- and hyperspectral remote sensing data processing, high resolution image processing and scene analysis, and computational intelligence. He has authored over 100 research papers, including more than 50 peer-reviewed articles in international journals, such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART B, and the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, and the *Pattern Recognition*. He was a recipient of the National Excellent Doctoral Dissertation Award of China and New Century Excellent Talents in University of China. He was also a Referee of more than 30 international journals. He is serving as an Associate Editor of *International Journal of Remote Sensing*, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Hong Shu received the bachelor's from the Mathematics Department of Central China Normal University, the master's degree from the Computer Department, Wuhan University, China, and the Ph.D. degree from the Department of Photogrammetry and Remote Sensing, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He is currently a Professor of Geoinformatics with the State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research focuses on spatio-temporal geographical computation, mainly including spatio-temporal geostatistics, computational geometry and topology, spatio-temporal scale analysis, land surface data assimilation, and space-time common sense reasoning. He is a Principal Investigator of more than twenty research projects and an author of approximately 100 academic papers.

Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998. He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He is also a Chang-Jiang Scholar Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist with the China State Key Basic Research Project from 2011 to 2016, appointed by the Ministry of National Science and Technology of China to Lead the Remote Sensing Program in China. He has authored over 450 research papers and five books. He holds 15 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence. He is a fellow of the Institution of Engineering and Technology. He was a recipient of the 2010 Best Paper Boeing Award and the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing. He regularly serves as a Co-Chair of the series SPIE conferences on multispectral image processing and pattern recognition, conference on Asia remote sensing, and many other conferences. He edits several conference proceedings, issues, and geoinformatics symposiums. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geo-spatial Information Science*, and the *Journal of Remote Sensing*, and the Guest Editor of the *Journal of applied remote sensing and Journal of sensors*. He is currently serving as an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.