Detail-Preserving Smoothing Classifier Based on Conditional Random Fields for High Spatial Resolution Remote Sensing Imagery

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

***Abstract*—In the field of high spatial resolution (HSR) remote sensing imagery classification, object-oriented classification and conditional random field (CRF) approaches are widely used due to their ability to incorporate the spatial contextual information. However, the selection of the optimal segmentation scale in object-oriented classification is not an easy task, and some pairwise CRF models always show an oversmooth performance. In this paper, a detail-preserving smoothing classifier based on conditional random fields (DPSCRF) for HSR imagery is proposed to apply the object-oriented strategy in the CRF classification framework, thus integrating the merits of both approaches to consider the spatial contextual information and preserve the detail information in the classification. The DPSCRF model defines suitable potential functions based on the CRF model for HSR image classification, which comprise the spatial smoothing and local class label cost terms. Both terms favor spatial smoothing in a local neighborhood to consider the spatial information. In addition, the local class label cost also considers the different label information of neighboring pixels at each iterative step in the classification to preserve the detail information. In order to deal with the spectral variability of HSR imagery, a segmentation prior is used by the object-oriented processing strategy. This models the probability of each pixel based on the segmentation regions obtained by the connected-component labeling algorithm. The experimental results with three HSR images demonstrate that the proposed classification algorithm shows a competitive performance in both the quantitative and the qualitative evaluation when compared to the other state-of-the-art classification algorithms.**

***Index Terms*—Conditional random fields (CRFs), contextual information, detail-preserving smoothing, high spatial resolution (HSR), image classification, remote sensing.**

I. Introduction

THE classification of high spatial resolution (HSR) remote sensing imagery plays an important role in various applications such as damage assessment for environmental disasters, precision agriculture, security applications, and urban planning due to the rich geometric and detail information [1]. The traditional classification methods such as support vector machine (SVM) [2], [3], neural networks [4], and the maximum likelihood classifier (MLC) [5], which are also called pixelwise classification, use a pixel-level processing strategy to assign each pixel one thematic label independently. These approaches, which do not consider the correlations between neighboring pixels, have been found to be effective for medium or coarse resolutions in many applications. In addition, certain dimension reduction and manifold learning approaches [6]–[10] can also be used as a preprocessing step of the pixelwise classification to find the more representative features in the subspace. However, these methods have a certain limit in HSR image classification, due to the increase in the spatial resolution, since they take no account of the spatial correlation of the HSR imagery and thus result in a salt-and-pepper classification appearance.

In order to overcome this problem, the strong spatial correlations of HSR imagery can be used to deal with the spectral variability issue [1], [11], [12]. There are two main approaches used to take the spatial contextual information into account in classification: 1) object-oriented classification methods and 2) random field methods. The object-oriented classification methods [13] integrate the classification and segmentation algorithms. The first key step is to use a segmentation algorithm to split the image into relatively homogeneous regions named segmentation regions or objects, and then a majority voting strategy [14], [15] based on pixel-based classification or direct classification, using the extracted features in each segmentation region, is used to map the segmentation label map to the classification result. Many different approaches can be used to achieve image segmentation, such as the fractal net evolution approach (FNEA) [16], the mean shift segmentation approach [17], and watershed segmentation [18]. For the object-oriented classification algorithms, the basic analysis unit is a segmentation region, which intrinsically provides the spatial contextual information of a pixel to alleviate the effect of within-class spectral variability and noise [19]. Thus, the object-oriented classification methods have the ability to overcome salt-and-pepper classification noise. However, the segmentation scale, which directly affects the classification performance, is difficult to choose due to the lack of prior information and the scale diversity of the various land-cover types [20].

The random field method is another useful classification approach that can incorporate the spatial contextual information. As the most popular random field model, the Markov random field (MRF) model has been widely used in classification

Manuscript received February 3, 2014; revised May 15, 2014 and August 1, 2014; accepted September 11, 2014. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2011CB707105, by the National Natural Science Foundation of China under Grants 41371344 and 41171313, by the Foundation for the Author of National Excellent Doctoral Dissertation of P. R. China (FANEDD) under Grant 201052, by the Fundamental Research Funds for the Central Universities under Grant 2042014kf00231, and by the State Key Program of National Natural Science of China under Grant 41331175. *(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 (e-mail: zhongyanfei@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/TGRS.2014.2360100

problems in recent years [1], [21]–[23] after first being introduced into image processing in 1984 [24]. Recently, a novel classification framework based on MRF, called the Markovian support vector classifier (MSVC), has been proposed [25] and has performed well. MSVC integrates SVM and MRF in a unique formulation and uses iterated conditional modes (ICMs) to optimize the energy function for the spatial contextual classification.

MRF allows the user to incorporate the spatial information in the label image, but it is difficult to model the spatial interaction in the observed image data. As an improved model for MRF, conditional random fields (CRFs) have the ability to consider the spatial contextual information in both the labels and observed image data. CRF was first proposed to solve the segmentation and labeling of 1-D text sequences by Lafferty *et al.* [26] in 2001 and was then successfully applied in image analysis by Kumar and Hebert [27], [28]. In the following years, CRF has been widely applied in image segmentation [29], stereo vision [30], and activity analysis [31]. For the image classification task, the most commonly used CRF model is pairwise CRF, which uses the unary and pairwise potentials to incorporate the spatial interaction in the local neighborhoods (commonly using the 4- or 8-neighborhood of the pixel grid). The pairwise CRF model has been successfully used for remote sensing image classification [32]–[37], but some models may present an oversmooth performance when obtaining the best classification result [34], which can have a great impact on the HSR image classification due to the presence of small important structures in the HSR imagery.

In order to preserve the spatial details in the classification result, the high-order potentials [38], [39] use more complex statistics of the image and a larger neighborhood and have performed well in experiments. However, the wide use of these high-order potentials has been limited due to the difficulty of efficient inference and the complexity of the models. Another strategy to deal with the problem is to design suitable potential functions in the pairwise CRF model. As an example, a simplified CRF considering the boundary information constraint in the pairwise potential [35] was developed for hyperspectral image contextual classification. In addition, a support vector conditional random field classifier incorporating a Mahalanobis distance boundary constraint (SVRFMC) [36] has also been proposed for HSR image classification. Recently, a hybrid object-oriented CRF classification framework [37] has been developed to preserve the spatial details by fusing the different degrees of smooth classification results. These approaches have shown a good adaptability and performance, but they fail to sufficiently consider the different labeling information at each iterative step in the classification, limiting the performance in preserving the spatial details while considering the spatial smoothing.

In this paper, a detail-preserving smoothing classifier based on conditional random fields (DPSCRF) for HSR imagery is proposed to take the spatial contextual information into account in HSR imagery classification. In this aspect of the model, to consider the spatial contextual information and preserve the detail information, DPSCRF not only designs suitable potential functions to incorporate the SVM and label cost constraint but uses a segmentation prior to deal with the spatial heterogeneity issue by an object-oriented processing strategy. The effective α-expansion inference algorithm is then used to obtain the final classification map. The DPSCRF classification algorithm is described in the following.

1) *Potential functions with label costs*. The DPSCRF classification algorithm defines suitable potential functions (unary potential and pairwise potential) in the local neighborhood. The unary potential is computed independently for each pixel, using SVM to model the relationship between the observed image data and the label, since the discriminative SVM classifier performs very well in HSR image classification in the case of limited training samples. This approach also avoids explicit data distribution modeling. The pairwise potential is expressed as a linear combination of the spatial smoothing term and the local class label cost term to model the spatial interaction between each pixel and its corresponding local neighborhood. The spatial smoothing term and the local class label cost term both encourage the neighboring pixels to have the same label, based on the spectral and spatial correlation. Moreover, the local class label cost term also considers the different thematic labels of the neighboring pixels in the iterative process to preserve the detail information. With these potential functions, DPSCRF not only can consider the spatial contextual information but also can preserve an abundance of land-cover detail.
2) *Segmentation prior based on the object-oriented strategy*. In order to alleviate the effect of the within-class spectral variability and noise in the feature space, a segmentation prior is used. The segmentation prior copes with the spatial heterogeneity issue by an object-oriented processing strategy. The segmentation is first performed by the use of the connected-component labeling algorithm in a temporary classification result at each iterative step so that it avoids the selection of scale. The segmentation and classification is thereby integrated into the uniform framework. The segmentation prior models the probability of each pixel based on the segmentation information, which is obtained by using a majority voting strategy with the segmentation result to deal with the problems of noise and the spectral variability of the HSR imagery.
3) *Inference by the graph-cut-based α-expansion algorithm*. With the definition of the unary potential and pairwise potential in the DPSCRF classification algorithm, an inference algorithm is needed to perform the search for the optimal labeling of each pixel. However, the exact inference is an NP-hard problem [40] for the multiclass HSR imagery. In this paper, the graph-cut-based α-expansion inference algorithm [41] is used to obtain the final classification result. This is an efficient approximate inference algorithm that has performed very well in various applications of computer vision [42].

The experimental results of this study demonstrate the efficiency of the proposed DPSCRF classification algorithm with three HSR data sets, which consist of two multispectral HSR images from different sensors (QuickBird and IKONOS) and

a hyperspectral HSR image [Hyperspectral Digital Imagery Collection Experiment (HYDICE)]. Compared to the other state-of-the-art classification algorithms, the proposed algorithm shows a competitive performance in both visualization and quantitative evaluation.

The rest of this paper is organized as follows. The CRF model is briefly introduced in Section II. Section III presents the DPSCRF classification algorithm for HSR imagery in detail. In Section IV, the experimental results and analysis for the three HSR data sets are reported. The sensitivity analysis is discussed in Section V. Finally, the work is concluded in Section VI.

II. CRF Model

As described in Section I, the spatial contextual information of HSR remote sensing imagery is very important for the classification task. As a contextual classification model, the random field method has been successfully applied in remote sensing image classification [32]–[34]. As a widely used random field model, CRF models local neighborhood interactions between random variables in a unified probabilistic framework. CRF is a probabilistic discriminative framework, which directly models the posterior probability of the labels, given the observed image data [26], [28], as a Gibbs distribution with the following form:

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

where $\mathbf{y}$ represents the observed data from the input image and $\mathbf{x}$ is the corresponding class labels of the whole image. $Z(\mathbf{y}) = \sum_{\mathbf{x}} \exp\{-\sum_{c\in C} \psi_c(\mathbf{x}_c, \mathbf{y})\}$ is the partition function, and the function $\psi_c(\mathbf{x}_c, \mathbf{y})$ is the potential function, which locally models the spatial interactions of random variables, based on the neighborhood system and cliques in the image.

According to the types of cliques in the observed data and their corresponding labels, the potential functions can be divided into unary potentials, pairwise potentials, and even high-order potentials. For the classification problem, pairwise CRF is the widely used option, which includes both unary and pairwise potentials. In addition, CRF can also incorporate the high-order potentials, which can model more wide ranging contextual information defined by the high-order neighborhood system and the cliques. However, these general potentials make the inference difficult. Therefore, pairwise CRF with an 8-neighborhood is used in this paper.

In the CRF classification framework, not only is the contextual information in the label image considered, but the spatial interaction in the observed data can also be taken into account. The CRF model directly models the posterior distribution by relaxing the conditional dependences between observed variables so that it can model the contextual information in a flexible way. Although pairwise CRF has been widely used for remote sensing image classification, certain models may produce oversmooth classification results [34]. In order to preserve the spatial details, the high-order potentials, which allow us to model measures that involve more than two variables, are introduced into the pairwise CRF model to model the rich statistical information of the image for the hyperspectral image classification [38]. However, the efficient inference of these general high-order potentials is difficult, and they always increase the complexity of the model due to considering local or global additional information. Defining suitable potential functions in the pairwise CRF model can be another strategy to deal with the problem. Simplified CRF [35] uses the pairwise potential with a boundary information constraint for spectral–spatial-based remote sensing image classification. Moreover, an SVRFMC [36] has also recently been proposed to deal with the problem. This approach uses SVM as the spectral term and a Mahalanobis distance boundary constraint model as the spatial term in the HSR classification framework. However, these approaches fail to sufficiently consider the different labeling information at each iterative step in the classification, which has an effect on the performance in preserving the spatial details while considering the spatial smoothing.

III. DPSCRF for HSR Imagery

In this paper, in order to consider the spatial contextual information and preserve the spatial details in the CRF classification framework, a DPSCRF for HSR imagery is proposed. In the DPSCRF classification framework, not only are the appropriate potential functions defined to consider the different labeling information at each iterative step, but an object-oriented strategy is also applied, thus integrating the merits of both approaches to consider the spatial contextual information in the classification.

To clearly describe the proposed algorithm for HSR remote sensing classification, the notations and definitions are first described. Consider an observation field $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$, where $\mathbf{y}_i$ is the observed spectral vector of the input HSR image pixel $i \in V = \{1, 2, \ldots, N\}$ and N is the total number of pixels in the image. We also set a labeling field $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ with $\mathbf{x}_i (i = 1, 2, \ldots, N)$ in the domain of the label set $\mathbf{L} = \{1, 2, \ldots, K\}$, where K denotes the number of classes.

As a discriminative classification framework, CRF directly models the posterior distribution of labels $\mathbf{x}$, given the observations $\mathbf{y}$, as a Gibbs distribution in (1). The corresponding Gibbs energy is defined as

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

Based on the Bayesian maximum *a posteriori* (MAP) rule, the image classification corresponds to finding the label image $\mathbf{y}$ that maximizes the posterior probability $P(\mathbf{x}|\mathbf{y})$ (i.e., $\mathbf{x}_{MAP} = \arg\max_x P(\mathbf{x}|\mathbf{y})$). Therefore, the MAP labeling x_{MAP} of the random field is given by

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

Finding the maximization of the posterior probability $P(\mathbf{x}|\mathbf{y})$ is therefore equivalent to finding the minimization of the energy function $E(\mathbf{x}|\mathbf{y})$. Classification problems in remote sensing are typically formulated as pairwise CRF, which can be written as the sum of the unary and pairwise potentials

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

$\psi_i(x_i, \mathbf{y})$ is the unary potential term, and $\psi_{ij}(x_i, x_j, \mathbf{y})$ is the pairwise potential term, defined over the local neighborhood N_{i}

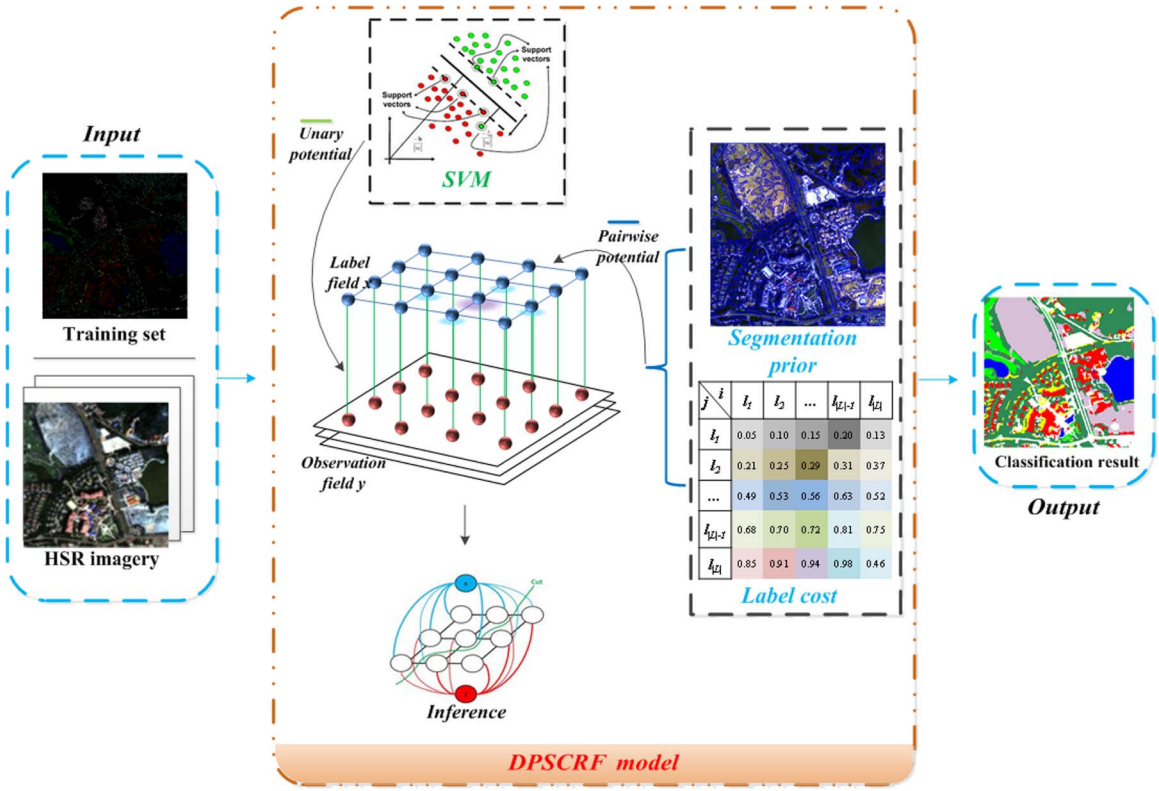

Fig. 1. Flowchart of the DPSCRF.

of the site i. The nonnegative constant λ is the tuning parameter for the pairwise potential term, making the tradeoff between the pairwise potential and the unary potential.

Based on the pairwise CRF model, the proposed DPSCRF algorithm for HSR image classification is described in Fig. 1. The DPSCRF method newly designs the unary potential term and pairwise potential term, which incorporate the SVM and label cost constraint for the HSR remote sensing image classification, to undertake appropriate smoothing while considering the spatial contextual information. In addition, in order to alleviate the effect of the within-class spectral variability and noise in the classification, a segmentation prior is applied by the object-oriented strategy to deal with this spatial heterogeneity issue. Finally, the optimal classification labels are obtained by the graph-cut-based α-expansion inference algorithm.

A. Unary Potential

The unary potential $\psi_i(x_i, \mathbf{y})$ models the relationship between the label and the observed image data, and it computes the cost of a single pixel taking a particular class label, based on the spectral feature vector. Therefore, it can be computed independently for each pixel, using a discriminative classifier that gives a probability estimate of the label x_i, given the feature vector. Typically, the unary potential $\psi_i(x_i, \mathbf{y})$ is defined as

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

where f is a feature mapping function that maps an arbitrary patch in an image to a feature vector and $f_i(\mathbf{y})$ represents the feature vector at site i, which denotes the spectral feature vector in this paper. $P(x_i = l_k | f_i(\mathbf{y}))$ is the probability of pixel x_i taking the label l_k, based on the feature vector, which is given by the pairwise coupling of the probability estimates from SVM [43], [44] in this paper. It should be noted that other classifiers that have a probabilistic output could also be used. However, the SVM classifier is used in this paper since it always shows a good performance in the case of a small training set for HSR remote sensing image classification.

The unary potential term independently estimates the classification labels based on the image features so that the minimization of this term's energy contribution by itself would be equivalent to a noncontextual Bayesian classification of the image. Therefore, the penalization of the unary potential is low in the case of the corresponding pixel being correctly classified with a high degree of confidence. This means that the corresponding land-cover class label of the pixel is expected to be kept in minimizing the energy when the confidence of the pixel taking the class label is high. In contrast, the penalization is high when the pixel is misclassified.

B. Pairwise Potential

The pairwise potential models the spatial contextual information between each pixel and its corresponding neighborhood by considering the labeling field and the observation field. Although the spectral values of the neighborhood pixels in homogeneous image regions may not seem to be the same, due to the effect of spectral variability and noise, they are always expected to share the same label due to the strong spatial correlation. The pairwise potential models this smoothness prior and takes the label constraint into account, which favors the neighborhood pixels in homogeneous image regions with the same

land-cover class and preserves the edges between two adjacent homogeneous regions. The form of the pairwise potential is

$$\psi_{ij}(x_i, x_j, \mathbf{y}) = \begin{cases} 0 & \text{if } x_i = x_j \\ g_{ij}(\mathbf{y}) + \theta^* \Theta_L(x_i, x_j|\mathbf{y}) & \text{otherwise} \end{cases} \quad (6)$$

where $g_{ij}(\mathbf{y})$ represents a smoothing term related to data $\mathbf{y}$ and $\Theta_L(x_i, x_j|\mathbf{y})$ is a local class label cost term with a size of $|\mathbf{L}| \times |\mathbf{L}|$, which represents the cost between the labels x_i and x_j within a neighborhood site. In addition, the parameter θ is the interaction coefficient that controls the degree of the label cost term in the pairwise potential.

In this paper, the function $g_{ij}(\mathbf{y})$ models the interaction between the neighboring pixels i and j, and it is designed to measure the difference in appearance between the neighboring pixels, as suggested in [45] and [46]

$$g_{ij}(\mathbf{y}) = dist(i,j)^{-1} \exp\left(-\beta \|y_i - y_j\|^2\right) \quad (7)$$

where the pair of (i, j) is the spatial location of neighboring pixels and the function $dist(i, j)$ is their Euclidean distance. y_i and y_j are the spectral vectors representing the appearance of the i and j pixels so that the strength of interaction within a neighborhood is related to the image data and encourages coherence in regions of similar appearance. Finally, the parameter β is set to the mean square difference between the spectral vectors of all the adjacent pixels in the image (i.e., $\beta = (2\langle \|y_i - y_j\|^2 \rangle)^{-1}$, where $\langle . \rangle$ is the average over the image).

The local class label cost term $\Theta_L(x_i, x_j|\mathbf{y})$ models the spatial relationship between different neighboring class labels x_i and x_j, based on the observed image data, and is defined as

$$\Theta_L(x_i, x_j|\mathbf{y}) = \frac{\min\left\{P\left(x_i|f_i(\mathbf{y})\right), P\left(x_j|f_j(\mathbf{y})\right)\right\}}{\max\left\{P\left(x_i|f_i(\mathbf{y})\right), P\left(x_j|f_j(\mathbf{y})\right)\right\}} \quad (8)$$

where the label probability $P(x_i|f_i(\mathbf{y}))$ based on the feature vector $f_i(\mathbf{y})$ is used in our work, which is also given by SVM, the same as the unary potential. This local class label cost term takes the current class label x_i of the neighboring pixels into account to measure the interaction between the labels at neighboring sites i and j, given the observed image $\mathbf{y}$. The term tends to change the label of the pixel, based on its neighborhood spatial label information, when there is a stronger overlap between the classes in the feature space, which affects the class discrimination. Thus, the local class label cost term, which is related to the current thematic labels, considers the spectral information through the form of an estimate of the probability distributions of the thematic class labels to undertake appropriate smoothing while considering the spatial contextual information. Note that an arbitrary matrix with a size of $|\mathbf{L}| \times |\mathbf{L}|$ can be used for determining $\Theta_L(x_i, x_j|\mathbf{y})$. Therefore, the label probability in this formula can also differ from the unary potential so that it can also be estimated by other probabilistic classifiers, which establishes the foundation for classification fusion. In addition, provided that all the label probabilities are set to be the same for all the labels and pixels, the label cost term will always be a constant. In this case, the local class label cost term is equivalent to the widely used Potts pairwise model.

The inclusion of the pairwise potential allows the spatial interaction of each pixel, expressed in terms of its neighborhood, to be considered in the classification. The pairwise potential is represented as a linear combination of the smoothing term and local class label cost term, and it models the spatial interaction between the class labels, as related to image features in the predefined neighborhood system. In the process of minimizing the energy function $E(\mathbf{x}|\mathbf{y})$, the pairwise potential encourages the neighboring pixels to have the same land-cover class in the output classification map, except for boundary regions between homogeneous image regions.

C. Segmentation Prior

The previous section gives a description of the pairwise potential, which benefits from the strong geometrical spatial information in the HSR image and is modeled by a linear combination of the smoothing term and the local class label cost term. The local class label cost term is expressed in the form of an estimate of the probability distributions of the thematic class labels. Thus, the pairwise potential depends on the accuracy of the probability estimation, to a certain degree. However, due to the spatial complexity and spectral variability of HSR imagery, it is difficult to arrive at an exact estimate, particularly in the case of limited samples, which may mean that the pairwise potential cannot effectively remove some of the salt-and-pepper classification noise. The object-oriented methods use the segmentation objects as the basic analysis units to alleviate the effect of the within-class spectral variability and noise in the classification. Therefore, in our work, the segmentation prior is used to deal with this spatial heterogeneity issue, which is modeled by the object-oriented processing strategy.

Similar to the flow of object-oriented processing, the segmentation is also the key step with the segmentation prior. The segmentation is produced by the connected-component labeling algorithm, which finds regions of connected pixels which have the same value in the classification result at each iterative step so that the segmentation avoids the selection of scale. The classical connected-component algorithm with an 8-neighborhood, using a union-find data structure, is used to assign labels to the segmentation objects [47], [48]. Then, for each segmentation region, the label of the region can be obtained by using the original SVM classification map to undertake majority voting. The segmentation prior is defined as

$$P(x_i = l_{seg}) = \max\left\{P(x_i = l_k)\right\}, k \in |L| \quad (9)$$

where l_{seg} represents the corresponding region label of the pixel.

The segmentation prior makes the probability estimate of the segmentation label to be the maximum of all the class labels for each pixel. This is similar to the object-oriented method, whose result can be obtained by the following processing, using the probabilistic map. The processing $P(x_i = l_{seg}) = \max\{P(x_i = l_k)\} + eps, k \in |L|$ (eps is a very small positive constant) is first applied for all the pixels, and then, the maximum probability rule is used to transform the probability map to the final classification result. Therefore, the segmentation prior uses a similar processing for the probability estimation to alleviate the effect of the within-class spectral variability and noise.

α-expansion algorithm

1 $x^p :=$ arbitrary labeling
2 Repeat
3 For each label $\alpha \in L = \{1, 2, ..., K\}$
4 $x^n := \arg\min_x E_\alpha(x^p)$
5 If $E(x^n) < E(x^p)$ then
6 $x^p := x^n$
7 End For
8 Until converged
9 Return x^p

Fig. 2. Graph-cut-based α-expansion algorithm.

D. Inference of DPSCRF

With the definition of the potential functions in DPSCRF, inference is performed to predict the optimal labeling, which corresponds to finding the minimum value with respect to the energy function. To find the optimal labeling, different approximate inference approaches have been proposed, such as ICM and graph cuts. However, ICM easily makes the solution stick at poor local minima, so it is extremely sensitive to the initial value. In this paper, the graph-cut-based α-expansion inference algorithm [41] is employed, which is a highly efficient inference algorithm that has gained a great deal of attention in various applications [42].

In the case of binary classification, graph cuts have been proven to be fast and to converge to a global energy minimum. However, HSR image classification is always a multiclass labeling problem. The graph-cut-based α-expansion algorithm [41], [49] designs a special local search algorithm for the energy minimization problem with multivalued variables to solve the problem of very small moves making the solution stick at poor local minima. The local search of the algorithm works by repeatedly computing the global minimum of a binary labeling problem via a graph-cut method in its inner loops. The α-expansion inference algorithm is described in Fig. 2.

Given a current label $x^p = \{x_i^p, i \in V\}$, the binary labeling problem comes from step 4 of the α-expansion algorithm for a particular label $\alpha \in \mathbf{L}$ in Fig. 2, which obtains the optimal solution $x^n = \{x_i^n, i \in V\}$ by optimizing the energy function $E_\alpha(x^p)$ by the use of a graph-cut method. This step gives each pixel the following two choices: either keep the current label or switch to a particular label $\alpha \in \mathbf{L} = \{1, 2, \ldots, K\}$. All the pixels make this choice simultaneously, so there are an exponential number of possible moves with respect to any particular α, which ensures that the algorithm has a strong local minimum property. Therefore, the α-expansion algorithm can be seen to reduce the problem with multivalued variables to a sequence of optimization subproblems with binary variables, which can be easily optimized by the graph-cut method. More details can be found in [41] and [49].

IV. Experiments and Analysis

A. Experimental Description

To evaluate the performance of the proposed DPSCRF classification framework, three HSR remote sensing data sets [14] are used in the experiments, with two multispectral HSR images (QuickBird and IKONOS) and a hyperspectral HSR image (HYDICE). Meanwhile, the comparison algorithms, including pixelwise classification algorithms, object-oriented approaches, and random field methods, are also performed with the same three data sets. The multiclass SVM approach implemented in LibSVM [43] is used to conduct the pixelwise classification experiment, which applies a radial basis function as the kernel function and a cross-validation approach to determine the optimal values of the parameters. The object-oriented classification method performs a majority voting strategy within the segments, using the same pixelwise multiclass SVM classification result. Mean shift segmentation (MSS) [17] and the multiresolution segmentation algorithm in eCognition 8.0 (FNEA) [16] are used to segment the image into relatively homogeneous segmentation regions, and they are respectively represented as MSS-OO and FNEA-OO. As for the FNEA segmentation method, the image layer weights of the segmentation are all set to be 1 because of the equal importance of each band, and the shape and compactness parameters in the composition of the homogeneity criterion are respectively set to 0.1 and 0.5. The random field methods in our comparison experiments include MSVC [25] and SVRFMC [36]. MSVC combines the SVM and Markov random field models in a general framework for spatial contextual classification to obtain the final labeling by the ICM algorithm. The MSVC framework respectively uses the Gaussian radial basis function and the Potts model as the kernel function and local prior energy function. SVRFMC considers the Mahalanobis distance boundary constraint in the spatial term to preserve the spatial details in the classification result.

In all the experiments, the quantitative performances are assessed by four kinds of accuracies, which are the accuracy of each class, the overall accuracy (OA is the percentage of correctly classified pixels), the average accuracy (AA is the average of the correctly classified pixels for each class), and the Kappa coefficient (Kappa) [50]. In addition, McNemar's test [51] is used to determine the statistical significance of the differences between the classification results obtained by the varying algorithms, using the same test sample set. Given two classifiers C_1 and C_2, the number of pixels misclassified by C_1 but not by C_2 is denoted as M_{12}, and M_{21} represents the number of cases misclassified by C_2 but not by C_1. If $M_{12} + M_{21} \geq 20$, the X^2 statistic can be considered as following a chi-squared distribution:

$$X^2 = \frac{(|M_{12} - M_{21}| - 1)^2}{M_{12} + M_{21}} \approx \chi_1^2. \tag{10}$$

This test can check whether the difference between varying classification results is meaningful. Given a significance level of 0.05, then, $\chi^2_{0.05,1} = 3.841459$. If X^2 is greater than $\chi^2_{0.05,1}$, the results of the two classifiers C_1 and C_2 are significantly different.

B. Experimental Data Sets

The first experiment is conducted using a 2.4-m spatial resolution HSR image, which was acquired in January 2010

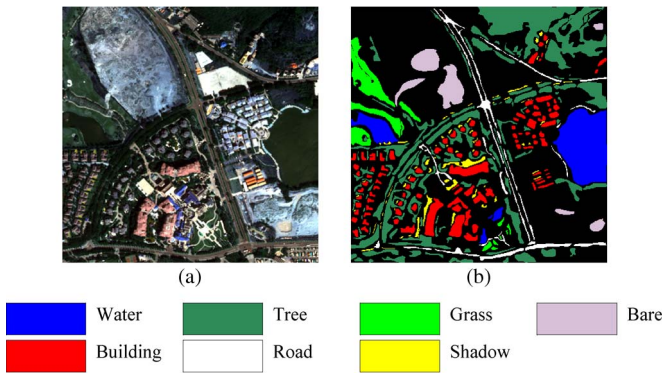

Fig. 3. Fancun QuickBird data set. (a) RGB false-color image (3, 2, 1). (b) Ground-truth image.

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

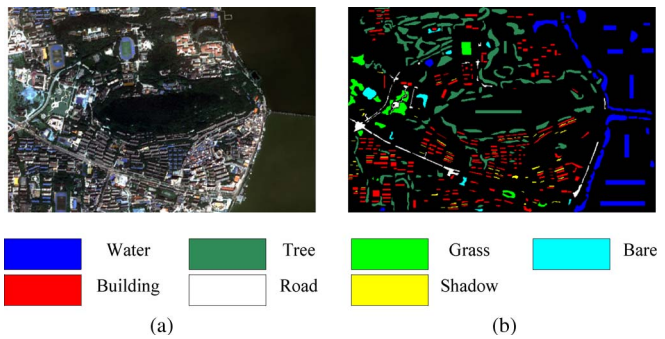

Fig. 4. Wuhan IKONOS data set. (a) RGB false-color image (3, 2, 1). (b) Ground-truth image.

by the QuickBird sensor from the Fancun area in Hainan province, China. This image contains 400 × 400 pixels and four multispectral channels. As an example, the false-color image by three bands is shown in Fig. 3(a). The Fancun QuickBird data set includes seven land-cover classes, i.e., water, tree, grass, bare, building, road, and shadow, and their corresponding thematic distribution is displayed in Fig. 3(b). The number of training and test samples for each class is given in Table I, and the training samples for each class are randomly chosen from the ground truth.

The second experiment uses the data of a different sensor to assess the performance of the proposed algorithm. The Wuhan image with a spatial resolution of 4 m is of an urban area that was recorded by the IKONOS sensor over Wuhan in Hubei province, China. The image has spatial dimensions of 400 × 600 pixels and blue, green, red, and near-infrared spectral channels. Fig. 4(a) and (b) respectively gives an overview of this data set by combining the (3, 2, 1) bands and the corresponding land-cover types. As with the Fancun data set, this image is also divided into seven thematic classes and uses the same number

TABLE II
CLASS INFORMATION OF THE WUHAN DATA SET

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

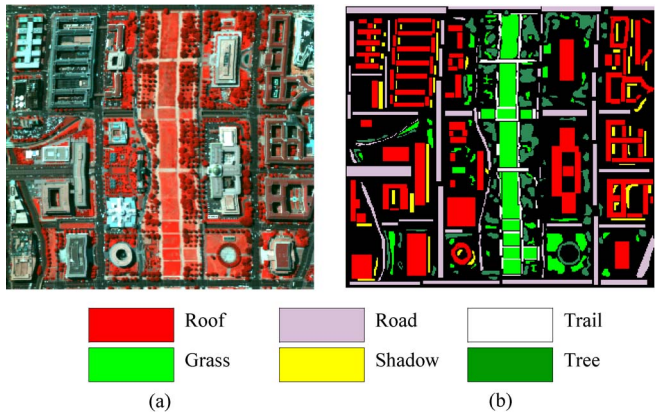

Fig. 5. Washington DC HYDICE data set. (a) RGB false-color image (60, 27, 17). (b) Ground truth.

TABLE III
CLASS INFORMATION OF THE WASHINGTON DC HYDICE IMAGE

Class name	Training samples	Test samples
Roof	50	12988
Road	50	7936
Trail	50	1303
Grass	50	6286
Shadow	50	1560
Tree	50	3870

of training samples for each class to test the performance of the algorithm in the case of limited training samples. Table II shows the list of the seven classes, including the training and test samples.

A high spatial–spectral resolution image is used in the third experiment, which is a subset of the Washington DC data set acquired by the HYDICE sensor. Since the roughly labeled ground truth of the original Washington DC data set is not sufficient to meet the requirement of evaluating the image details, the subimage with more detailed labels is used in this paper. The subimage contains 307 × 280 pixels and 191 bands. The false-color image shown in Fig. 5(a) represents the appearance of the area, and the corresponding ground truth is shown in Fig. 5(b). Table III gives the number of randomly selected training and test samples for each class of interest.

C. Experimental Setup

For the Fancun, Wuhan, and Washington DC data sets, the optimal parameters for all the classification approaches are set as follows to obtain the highest OA in the classification results. The parameters C/γ of SVM are respectively set to 256/0.5, 4096/0.5, and 16384/0.125. The spatial/spectral bandwidth parameters of the MSS algorithm are chosen as 15/20, 3/50, and

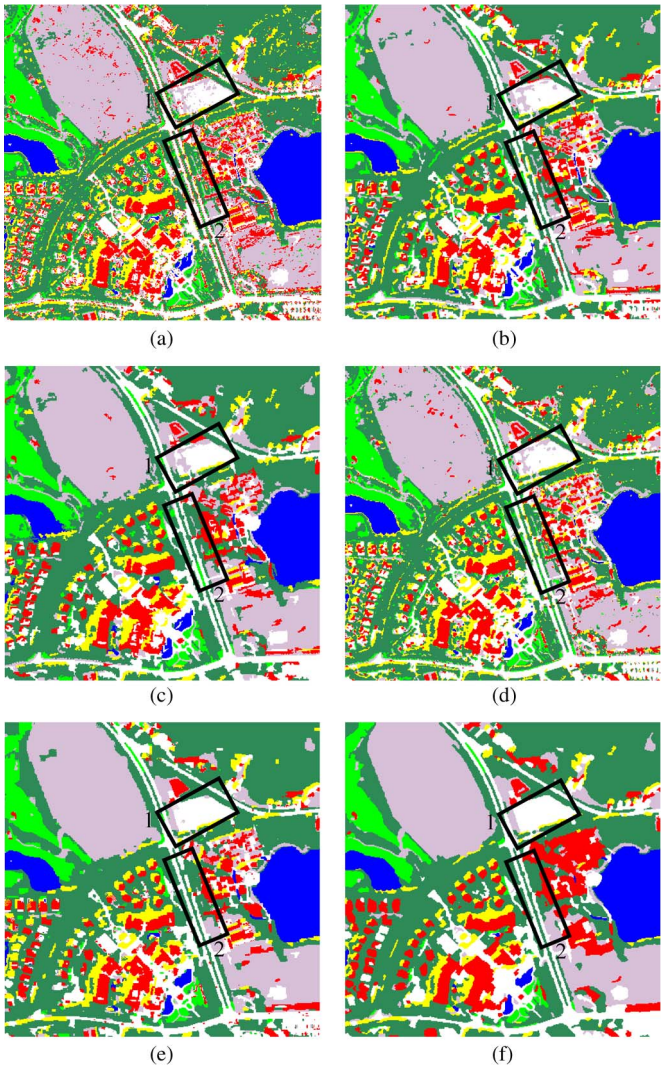

Fig. 6. Classification results for the Fancun QuickBird data set. (a) SVM. (b) FNEA-OO. (c) MSS-OO. (d) MSVC. (e) SVRFMC. (f) DPSCRF.

19/10, respectively. For the FNEA algorithm, the segmentation scale parameter is selected as 20, 40, and 80, respectively. For the SVRFMC algorithm, the fixed weight λ_1 for trading off the spatial term and spectral term is set as 0.2, 0.2, and 0.5, respectively. Meanwhile, the λ and θ parameters of DPSCRF are respectively selected as 0.7/3.4, 0.7/1, and 0.5/0.4.

D. Experimental Results and Analysis

For the Fancun, Wuhan, and Washington DC experiments, the classification maps obtained by the different classification algorithms (i.e., SVM, FNEA-OO, MSS-OO, MSVC, SVRFMC, and DPSCRF) are shown in Figs. 6–8, respectively. The corresponding quantitative results for the various classification methods are reported in Tables IV–VI.

As can be observed from the classification maps, the classification result of the SVM algorithm, without considering any neighborhood spatial contextual information, exhibits a lot of classification noise. By taking the neighborhood interactions into account, the object-oriented classification algorithms (FNEA-OO and MSS-OO) and the random field

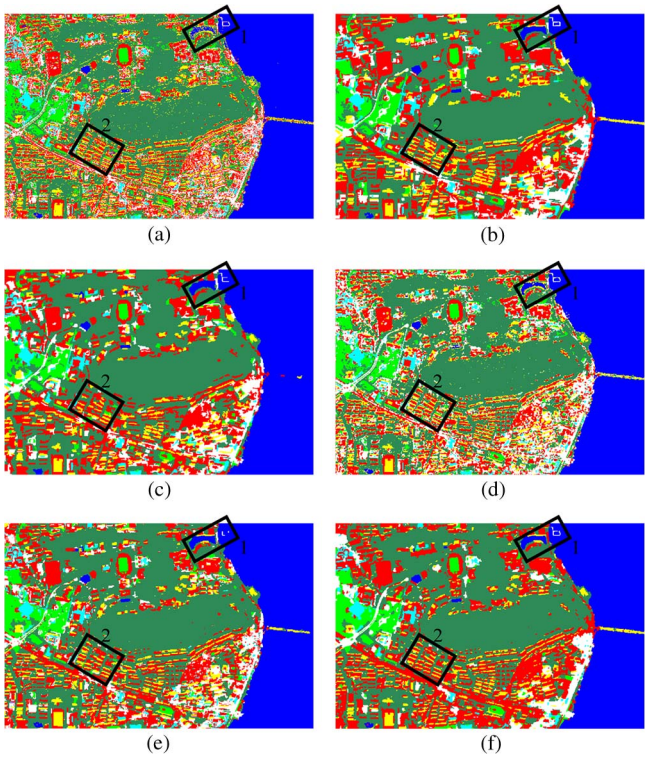

Fig. 7. Classification results for the Wuhan IKONOS data set. (a) SVM. (b) FNEA-OO. (c) MSS-OO. (d) MSVC. (e) SVRFMC. (f) DPSCRF.

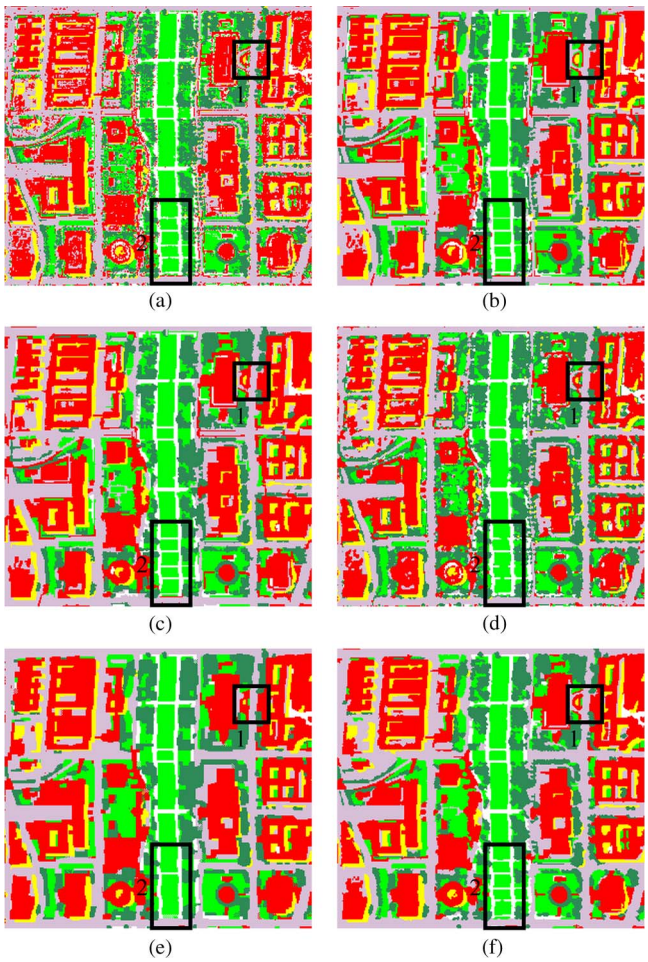

Fig. 8. Classification results for the Washington DC HYDICE data set. (a) SVM. (b) FNEA-OO. (c) MSS-OO. (d) MSVC. (e) SVRFMC. (f) DPSCRF.

TABLE IV
CLASSIFICATION ACCURACIES FOR THE FANCUN QUICKBIRD DATA SET

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	90.73	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
MSVC	98.86	96.34	94.65	99.25	67.43	94.01	96.95	92.51	92.50	0.8995
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

TABLE V
CLASSIFICATION ACCURACIES FOR THE WUHAN IKONOS DATA SET

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
MSVC	64.62	94.44	98.34	98.47	91.85	97.48	80.08	90.00	89.33	0.8676
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

TABLE VI
CLASSIFICATION ACCURACIES FOR THE WASHINGTON DC HYDICE DATA SET

Methods	Accuracy (%)						OA (%)	AA (%)	Kappa
	Roof	Road	Trail	Grass	Shadow	Tree			
SVM	91.55	92.49	97.62	95.53	96.86	91.29	92.95	94.22	0.9063
MSS-OO	95.18	93.83	98.93	96.96	98.46	94.01	95.35	96.22	0.9381
FNEA-OO	95.47	93.59	96.16	97.66	98.27	91.09	95.09	95.37	0.9345
MSVC	94.18	94.71	95.40	96.28	97.24	96.30	95.12	95.69	0.9351
SVRFMC	97.54	95.34	92.10	96.90	97.31	97.55	96.69	96.12	0.9557
DPSCRF	96.91	96.27	98.62	97.41	97.88	96.05	96.87	97.19	0.9582

methods (MSVC, SVRFMC, and DPSCRF) deliver a smoother classification result and exhibit a better visual performance. In Tables IV–VI, it can also be clearly seen that the algorithms considering the spatial interaction show an improvement of more than 2% over the pixelwise SVM classification, in terms of the OA and Kappa coefficient, which demonstrates the effectiveness of incorporating the spatial contextual information.

For the object-oriented classification approaches (FNEA-OO and MSS-OO), they always show a relatively good classification performance. However, the segmentation scale has a great impact on the classification performance, and it is difficult to choose because different size land-cover classes often have different optimal scales. Thus, some of the pixels for different land-cover types are wrongly classified in the MSS-OO and FNEA-OO methods, as shown in Figs. 6–8(b) and (c).

In Figs. 6(d)–8(d), the MSVC algorithm keeps the detail information as much as possible, but it also retains certain classification noise in the classification map. Therefore, as Tables IV–VI show, the improvement in the classification accuracy is very limited when compared to the SVM method.

SVRFMC is the latest CRF approach to be used for HSR image classification, and it shows an acceptable classification performance in all the experiments, as shown in Figs. 6(e)–8(e) and Tables IV–VI. The algorithm also aims at preserving the detail information so that the Mahalanobis distance boundary constraint is used in the pairwise potential. However, it fails to sufficiently consider the different labeling information at each iterative step in the classification, limiting the performance in preserving the spatial details. Therefore, an oversmooth classification performance can also be seen in the results, which is obvious in the Washington DC experiment, such as the road highlighted in black box 2 in Fig. 8(e).

Although all the object-oriented classification algorithms and random field methods can consider the spatial contextual information in the classification, the performances of the classification algorithms in trading off the smoothness and denoising are different. As shown in Figs. 6(f)–8(f), the DPSCRF method shows a competitive performance in alleviating the salt-and-pepper classification noise and keeping good boundary information, which not only results in a smooth appearance but also preserves useful information. As highlighted in black boxes 1 and 2, DPSCRF displays complete object visualization, while the classification results of the other classification approaches (FNEA-OO, MSS-OO, MSVC, and SVRFMC) contain some meaningless regions or salt-and-pepper classification noise in homogeneous areas. For the quantitative metrics (OA, AA, and Kappa) reported in Tables IV–VI, the DPSCRF classification algorithm again outperforms the other methods.

In addition, a pairwise comparison of the six classification approaches for the three different experiments is given in Tables VII–IX by the use of McNemar's test to evaluate the statistical significance, and the computation times of the varying algorithms are also presented in Table X by the use of a personal computer of 3.1 GHz with 8-GB RAM. It can be seen from Tables VII–IX that all McNemar's values between DPSCRF and the other methods are greater than the critical value of $\chi^2_{0.05,1}$ (3.841459), which means that the differences are significant. Moreover, it should also be noted that some of the McNemar's values for the compared methods are smaller than the critical value, which means that their differences

TABLE VII
MCNEMAR'S TEST VALUES FOR THE FANCUN QUICKBIRD DATA SET

Methods	SVM	MSS-OO	FNEA-OO	MSVC	SVRFMC	DPSCRF
SVM	NA	323.96	336.34	512.11	1031.40	2593.80
MSS-OO		NA	3.76	8.51	186.00	1310.80
FNEA-OO			NA	1.19	264.57	1554.00
MSVC				NA	329.54	1673.00
SVRFMC					NA	846.12
DPSCRF						NA

TABLE VIII
MCNEMAR'S TEST VALUES FOR THE WUHAN IKONOS DATA SET

Methods	SVM	MSS-OO	FNEA-OO	MSVC	SVRFMC	DPSCRF
SVM	NA	499.14	346.96	237.76	486.75	1438.30
MSS-OO		NA	11.28	135.20	5.07	478.93
FNEA-OO			NA	62.39	2.16	613.37
MSVC				NA	113.03	884.75
SVRFMC					NA	647.76
DPSCRF						NA

TABLE IX
MCNEMAR'S TEST VALUES FOR THE WASHINGTON DC HYDICE DATA SET

Methods	SVM	MSS-OO	FNEA-OO	MSVC	SVRFMC	DPSCRF
SVM	NA	321.80	272.06	382.05	830.87	1064.60
MSS-OO		NA	5.98	3.91	146.33	271.55
FNEA-OO			NA	0.05	224.14	333.21
MSVC				NA	250.13	332.65
SVRFMC					NA	5.63
DPSCRF						NA

TABLE X
COMPUTATION TIMES (IN SECONDS) FOR THE DIFFERENT METHODS

Methods	Fancun	Wuhan	Washington DC
SVM	16	19	30
MSS-OO	15	22	32
FNEA-OO	1	2	6
MSVC	4805	10177	2197
SVRFMC	1506	2520	1320
DPSCRF	9	12	15

are insignificant. For example, all McNemar's values between FNEA-OO and MSVC are smaller than the critical value in the three different data sets. For the computation times, the object-oriented and random field classification methods using the same initial classification do not contain the time cost of the initial SVM. It should be noted, however, that the MSVC method is a general classification framework endowed with efficient parameter optimization ability, and the time cost of MSVC is therefore computed with the parameter optimization. As shown in Table X, the computation times of FNEA-OO are the smallest, while the SVRFMC algorithm needs much more computation time since it needs dozens of iterations to converge to a stable energy value by loopy belief propagation. The computation times of the DPSCRF method are also acceptable for the three experimental data sets since the α-expansion inference algorithm is very efficient.

V. SENSITIVITY ANALYSIS

In the former section, the experimental results of the three data sets confirm that DPSCRF performs well. However, the λ and θ parameters of DPSCRF have an impact on the classification result, and they trade off the unary potential term and pairwise potential term to make the balance between the smoothing and the spectral information. In this section, a sensitivity analysis is provided to evaluate the performance of the DPSCRF algorithm. Finally, the effects of different training and test samples are also illustrated. Additional experiments are therefore performed to evaluate the effect of these parameters with the Fancun, Wuhan, and Washington DC data sets.

A. Sensitivity Analysis for the λ and θ Parameters

In order to study the effect of the λ and θ parameters of DPSCRF, parameter λ is varied from 0.1 to 1.9, with an interval of 0.2, and parameter θ is varied from 0.4 to 4.0, with an interval of 0.6, in the additional experiments for the Fancun, Wuhan, and Washington DC data sets. The relationship between the OA and the λ and θ parameters of DPSCRF is presented in Fig. 9.

As shown in Fig. 9, the classification accuracies of DPSCRF first increase and then slightly decrease with the increase in parameter λ. The λ and θ parameters of DPSCRF are both designed to control the relative importance of the unary potential term and the pairwise potential term. When parameter θ is kept unchanged, as parameter λ increases, the relative importance of the spatial term is increased so that the neighborhood spatial contextual information can be effectively utilized to alleviate the effect of the salt-and-pepper classification noise, which is the reason for the initial increase in the classification accuracy. However, when parameter λ reaches a certain value, the performance tends to slightly decrease. This is due to the inclusion of the large spatial smoothing effect, which leads to the varying degrees of oversmooth performance, relating to the size of parameter λ. For parameter θ, on the one hand, it has a similar function to parameter λ as one part of the spatial smoothing term; on the other hand, it also has a certain fine-tuning function for the classification accuracy when keeping parameter λ constant, as presented in Fig. 9. Therefore, DPSCRF is a relatively stable classification algorithm, whose classification performance changes regularly with the effect of its parameters.

B. Sensitivity Analysis for the Training Set Size

The effects of different training and test samples for all the aforementioned classification algorithms are examined with the Fancun, Wuhan, and Washington DC data sets. The training samples are randomly selected from the corresponding overall ground truth, and the numbers for each class are varied between 50, 100, 200, 300, 400, and 500 for the three data sets. The remaining samples are used as test samples to evaluate the classification accuracies. The classification accuracies (OA) for each classification method at each training number are illustrated in Fig. 10.

As can be observed from Fig. 10, the classification accuracies of all the classification approaches share a similar trend as the number of training samples increases. The classification performances of all the algorithms show a gradual increase with the increase in training number, except for certain training numbers over the three data sets. In addition, the classification approaches incorporating spatial contextual information (MSS-OO, FNEA-OO, MSVC, SVRFMC, and DPSCRF) always outperform SVM, in all cases, with different training samples. The object-oriented classification methods (i.e., MSS-OO and

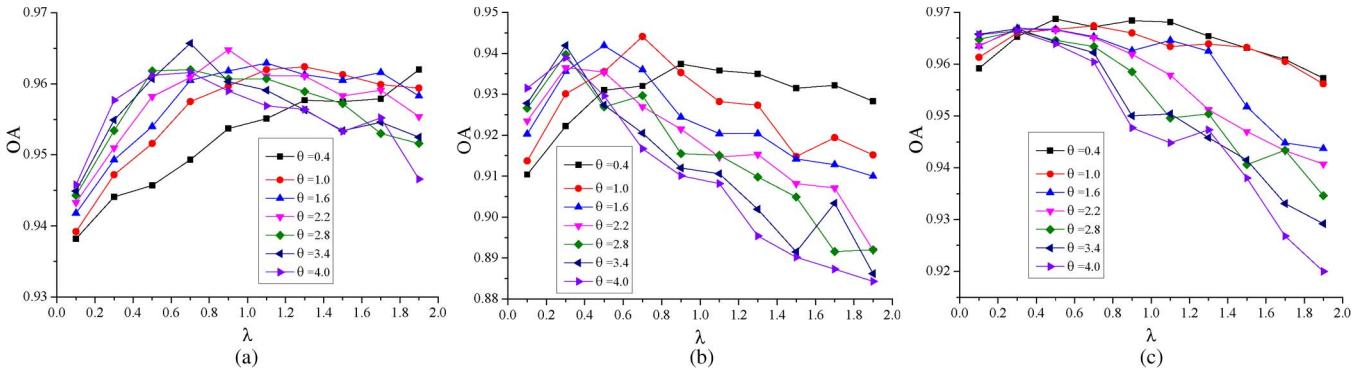

Fig. 9. Sensitivity analysis for the λ and θ parameters of DPSCRF with the three data sets. (a) Fancun QuickBird image. (b) Wuhan IKONOS image. (c) Washington DC HYDICE image.

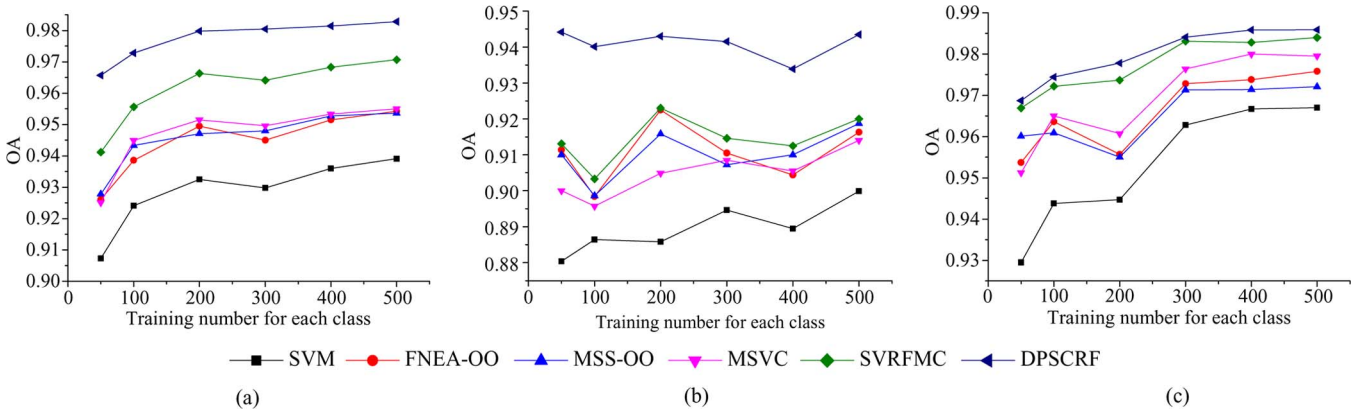

Fig. 10. Sensitivity analysis for the number of training samples, with the three data sets. (a) Fancun QuickBird image. (b) Wuhan IKONOS image. (c) Washington DC HYDICE image.

FNEA-OO) show a similar performance in all three image classifications. The MSVC and SVRFMC random field methods share similar trends in classification accuracy, and SVRFMC performs better than MSVC. MSVC has a similar ability to the object-oriented approaches for HSR image classification, in spite of being slightly better with the Fancun and Washington DC data sets and slightly poorer with the Wuhan data set. The DPSCRF algorithm has the best classification performance of all the approaches, in all conditions, with different training samples and all three data sets.

VI. Conclusion

In this paper, a DPSCRF for HSR imagery is proposed to consider the spatial contextual information and preserve the spatial details in classification. DPSCRF not only defines suitable unary and pairwise potentials, based on the pairwise CRF model, incorporating SVM and the label cost constraint to alleviate the oversmooth performance, but also uses a segmentation prior to deal with the spectral variability of HSR imagery. Probabilistic SVM is used in the unary potential to obtain an acceptable probability estimation with the limited training samples. The pairwise potential uses the spatial smoothing and local class label cost terms to favor spatial smoothing in the local neighborhood and to take the spatial contextual information into account. The local class label cost term also has the ability to alleviate an oversmooth classification result since it considers the different label information of the neighboring pixels at each iterative step in the classification. Moreover, based on an object-oriented processing strategy, a segmentation prior is used to adjust the probability of each pixel, based on the segmentation information, to cope with the spectral variability of HSR imagery. The segmentation is obtained by using the connected-component labeling algorithm in the temporary classification result at each iterative step so that it avoids the selection of scale. Three real data experiments using three types of HSR images from QuickBird, IKONOS, and HYDICE demonstrate that the DPSCRF classification algorithm has a competitive accuracy and visual performance, compared with the other state-of-the-art classification algorithms.

When using the local neighborhood in pairwise CRF, it is difficult to capture the rich global statistical information, which can result in an oversmooth performance. In our future work, high-order random fields will be studied to improve the ability of the model to capture these complex statistics.

Acknowledgment

The authors would like to thank the editor, associate editor, and anonymous reviewers for their helpful comments and suggestions. The authors would also like to thank Beijing Panorama Space Technology Company, Ltd., for providing the free QuickBird image and also the research group supervised by D. Landgrebe of Purdue University, West Lafayette, IN, USA, for providing the free downloads of the Hyperspectral Digital Imagery Collection Experiment image.

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] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2297–2307, May 2010.

[3] 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.

[4] F. Ratle, G. Camps-Valls, and J. Weston, "Semisupervised neural networks for efficient hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2271–2282, May 2010.

[5] 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.

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

[7] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.

[8] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General averaged divergence analysis," in *Proc. IEEE Int. Conf. Data Mining*, 2007, pp. 302–311.

[9] 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.

[10] W. F. Liu and D. C. Tao, "Multiview Hessian regularization for image annotation," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2676–2687, Jul. 2013.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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, Jul. 2010.

[16] 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.

[17] 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.

[18] D. Li, G. Zhang, Z. Wu, and L. Yi, "An edge embedded marker-based watershed algorithm for high spatial resolution remote sensing image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2781–2787, Oct. 2010.

[19] U. C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen, "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 3/4, pp. 239–258, Jan. 2004.

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

[21] A. H. Solberg, T. Taxt, and A. K. Jain, "A Markov random field model for classification of multisource satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 1, pp. 100–113, Jan. 1996.

[22] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.

[23] S. Y. Chen, H. Y. Tong, and C. Cattani, "Markov models for image labeling," *Math. Probl. Eng.*, vol. 2012, pp. 814 356-1–814 356-18, 2012.

[24] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI_6, no. 6, pp. 721–741, Nov. 1984.

[25] G. Moser and S. B. Serpico, "Combining support vector machines and Markov random fields in an integrated framework for contextual image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2734–2752, May 2013.

[26] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. Int. Conf. ML*, 2001, pp. 282–289.

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

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

[29] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation," presented at the Eur. Conf. Comput. Vis., 2006, pp. 1–15.

[30] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Minneapolis, MN, USA, 2007, pp. 1–8.

[31] C. Sminchisescu, A. Kanaujia, and D. Metaxas, "Conditional models for contextual human motion recognition," *Comput. Vis. Image Understand.*, vol. 104, no. 2/3, pp. 210–220, Nov. 2006.

[32] J. Bai, S. Xiang, and C. Pan, "A graph-based classification method for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 803–817, Feb. 2013.

[33] 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.

[34] 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.

[35] 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.

[36] 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 Observ. Remote Sens.*, vol. 7, no. 4, pp. 1314–1330, Apr. 2014.

[37] 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.

[38] 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.

[39] L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr, "Associative hierarchical CRFs for object class image segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 739–746.

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

[41] 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.

[42] 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.

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

[44] 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, Aug. 2004.

[45] Y. Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2001, pp. 105–112.

[46] 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.

[47] L. Shapiro and G. Stockman, *Computer Vision*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.

[48] K. Wu, E. Otoo, and K. Suzuki, "Optimizing two-pass connected-component labeling algorithms," *Pattern Anal. Appl.*, vol. 12, no. 2, pp. 117–135, Feb. 2009.

[49] 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.

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

[51] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, Jun. 1947.

Ji Zhao received the B.S. degree in surveying from the Xi'an University of Science and Technology, Xi'an, China, in 2011. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing in the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China.

His major research interests include high spatial resolution remote sensing image classification, scene analysis, and random field algorithms.

Yanfei Zhong (M'11) received the B.S. degree in information engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, 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 and is currently a Professor. His research interests include multi- and hyperspectral remote sensing data processing, high-resolution image processing and scene analysis, and computational intelligence. He has published more than 60 research papers, including more than 30 peer-reviewed articles in international journals such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART B, and *Pattern Recognition*.

Dr. Zhong was the recipient of the National Excellent Doctoral Dissertation Award of China (2009) and New Century Excellent Talents in University of China (2009). He was a Referee of more than 20 journals, including the IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and *Pattern Recognition*.

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 of 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 with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, as the Head of the Remote Sensing Division. He is also a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education, China. He is currently the Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He is an Executive Member (Board of Governor) of the China National Committee of International Geosphere-Biosphere Programme. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence, International Journal of Image and Graphics, International Journal of Digital Multimedia Broadcasting, Journal of Geo-spatial Information Science*, and *Journal of Remote Sensing*. He has more than 360 research papers and is the holder of five patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the Institution of Electrical Engineers, an Executive Member for the China Society of Image and Graphics, and others. He regularly serves as a Cochair of the series International Society for Optical Engineering (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 the Geoinformatics Symposiums. He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.