A Hybrid Object-Oriented Conditional Random Field Classification Framework for High Spatial Resolution Remote Sensing Imagery

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

***Abstract*—High spatial resolution (HSR) remote sensing imagery provides abundant geometric and detailed information, which is important for classification. In order to make full use of the spatial contextual information, object-oriented classification and pairwise conditional random fields (CRFs) are widely used. However, the segmentation scale choice is a challenging problem in object-oriented classification, and the classification result of pairwise CRF always has an oversmooth appearance. In this paper, a hybrid object-oriented CRF classification framework for HSR imagery, namely, CRF + OO, is proposed to address these problems by integrating object-oriented classification and CRF classification. In CRF + OO, a probabilistic pixel classification is first performed, and then, the classification results of two CRF models with different potential functions are used to obtain the segmentation map by a connected-component labeling algorithm. As a result, an object-level classification fusion scheme can be used, which integrates the object-oriented classifications using a majority voting strategy at the object level to obtain the final classification result. The experimental results using two multispectral HSR images (QuickBird and IKONOS) and a hyperspectral HSR image (HYDICE) demonstrate that the proposed classification framework has a competitive quantitative and qualitative performance for HSR image classification when compared with other state-of-the-art classification algorithms.**

***Index Terms*—Classification fusion, conditional random fields (CRFs), high spatial resolution (HSR), object-oriented classification, remote sensing.**

I. Introduction

THE availability of high spatial resolution (HSR) remote sensing imagery obtained from satellites (e.g., IKONOS, QuickBird, and WorldView-2) increases the possibility of accurate Earth observations [1]. Such HSR imagery provides valuable geometric and detailed information, which is important for applications such as damage assessment for environmental disasters, precision agriculture, security applications, and urban planning [2]. In these applications, mapping the predefined land cover type to the ground area denoted by image pixels (i.e., the classification task) is particularly relevant [3].

The traditional pixel-based classification methods, which are also called pixelwise classification, process each pixel independently without considering the correlations between neighboring pixels [4], [5]. These classification methods, e.g., support vector machine (SVM) [6], [7], neural networks [8], and the maximum likelihood classifier (MLC) [9], have been widely used for remote sensing imagery in many applications. One particular type of pixel-based classification, which is called probabilistic pixelwise classification (e.g., MLC), can produce the posterior probabilities of the samples belonging to the classes, which is particularly useful in practical recognition situations. In addition, certain dimension reduction and manifold learning [10]–[14] methods, as a preprocessing step of the pixel-based classification, can be first used to obtain the more representative features, particularly for medium or coarse spatial resolutions. However, with the gradual increase in the spatial resolution, the spectral variability within each land cover class increases, and the separability between classes decreases. As a result, these methods are less effective for HSR imagery, in which the salt-and-pepper appearance of the classification results in an increase in classification errors.

In order to overcome this problem, classification methods incorporating spatial information in the classification are necessary when considering the geometrical information in HSR imagery [15]. There are two main approaches to incorporating spatial contextual information: random fields and object-oriented classification methods. The object-oriented classification approaches [16] process each segmentation region as a whole rather than individual pixel, and the key step is to segment the image into relatively homogeneous regions called segmentation regions or objects [17]. Many different approaches, such as the fractal net evolution approach (FNEA) [18], the mean-shift segmentation (MSS) approach [19], and watershed segmentation [20], have been used to deal with this problem. A majority voting strategy after the pixel-based classification or the direct classification using region features is then applied to transform the segmentation map into the classification result [21]–[23]. Using the segmentation regions as the basic analysis units helps to overcome the salt-and-pepper appearance of the classification. The reason for this is that incorporating the spatial contextual information, such as size, shape, texture, and geometrical structure, helps to alleviate the within-class

Manuscript received August 21, 2013; revised November 20, 2013 and January 7, 2014; accepted February 3, 2014. Date of publication March 11, 2014; date of current version May 30, 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 Grant 41371344, and by the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant 201052.

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; zhaoji2015@gmail.com; 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/TGRS.2014.2306692

spectral variability [24]. However, the classification result is directly affected by the segmentation scale, which is important but challenging to choose [25]. For single-scale segmentation, the difficulty of choosing the scale is that there is no prior to determine what parameters will produce a good result, and varying land cover types often have different optimal scales due to the difference in size [17].

A further interesting approach is the random field method represented by Markov random fields (MRFs), which not only incorporates the spatial contextual information but avoids the scale choice problem for modeling the spatial dependencies of the pixels. MRF was first introduced into image processing in 1984 [26] and has been widely used in classification problems in recent years [3], [27]–[29]. In addition, a novel framework called Markovian support vector classifier (MSVC), integrating SVM and MRF models in a unique formulation for spatial contextual classification, has been recently proposed [30] and has performed well. However, MRF considers the spatial information only in the label image, not in the original observed image data. An improved model for MRF is the conditional random field (CRF) model, which directly models the class posterior probability, given observed data, rather than their joint distribution. This approach has the ability to incorporate the spatial information in both the labels and observed image data. CRF was first proposed by Lafferty *et al.* [31] for solving the labeling of 1-D text sequences and has been successfully applied in image segmentation [32], stereo vision [33], and activity analysis [34] after being first introduced into image analysis by Kumar and Hebert [35], [36]. The image classification problem is typically formulated as pairwise CRF, modeling the spatial dependencies in the local neighborhoods (where the neighborhood is typically defined as a 4- or 8-neighborhood in the pixel grid), which has been successfully applied to hyperspectral imagery [37]–[39]. However, the results of this model always have an oversmooth appearance [39], [40]. Although sometimes the oversmoothing phenomenon is not obvious in certain images, it is important to alleviate the effect for HSR imagery due to the presence of small important structures. In order to cope with this problem, high-order potentials [40], [41], modeling the more complex statistics of the image, have been used and have achieved good performances in experiments. However, it is difficult to achieve an efficient inference with these high-order potentials, and they always need to import local or global information, which increases the complexity of the model.

In this paper, a hybrid object-oriented CRF classification framework for HSR imagery (CRF + OO) is proposed to utilize both the pairwise CRF model and the object-oriented classification method, to incorporate the spatial contextual information for HSR imagery classification. The CRF + OO classification framework is described in the following.

1) A CRF classification framework containing two CRF models is designed. The two CRF models have the same pairwise potential to consider the neighborhood interactions but different unary potentials, which are called the log unary potential and the quasi-gamma unary potential, respectively; thus, the corresponding models are represented as CRF-LOG and CRF-QG, respectively. The unary potentials use probabilistic pixelwise classification to model the relationship between the observed data and the corresponding label. However, the quasi-gamma unary potential in CRF-QG is designed to give a larger weight to the spectral information, to impose a restriction on the less-possible pixel labels and to favor the most likely labels of the land cover types. Therefore, CRF-LOG and CRF-QG always have different degrees of smooth classification performance, and they provide complementary information about the land cover type since they also possess the same pairwise potential, which allows them to have the ability to incorporate spatial information.
2) Although the potentials have been modeled in the CRF classification framework, the inference searching for the best solution corresponding to the optimal pixel labeling is an NP-hard problem [42] for the multivalued variables in HSR imagery classification. To find the optimal labeling, different approaches, e.g., iterated conditional modes (ICMs), loopy belief propagation (LBP), and graph cuts, have been proposed. However, ICM easily gets stuck at poor local minima; thus, it is extremely sensitive to the initial estimate, and LBP is not guaranteed to converge. In this paper, the graph-cut-based α-expansion algorithm [43] is applied to the HSR imagery classification since it has a better performance (efficiency and accuracy) in computer vision [44].
3) To obtain a more robust and excellent classification result, a classification fusion scheme is proposed to make full use of the former classification information. Since the results of the two CRF models incorporate the spatial contextual information, the CRF-LOG and CRF-QG maps have the ability to obtain homogeneous regions. Therefore, they are selected to obtain the segmentation result by the connected-component labeling algorithm so that we can make a fusion of the classification information at the object level. In addition, the classification map can provide prior class information to address the scale choice problem because the classification result reflects the distribution information of the materials of interest, so that the scale of the segmentation meets the requirement of all the land cover classes. The final classification result is achieved by integrating the object-oriented classifications, using a majority voting strategy with a region size constraint at the object level.

The efficiency of the proposed CRF + OO classification framework is confirmed by performing experiments on three data sets, which consists of two multispectral HSR images (QuickBird and IKONOS) and a hyperspectral HSR image (HYDICE). Compared with other state-of-the-art algorithms, the proposed algorithm has a remarkable quantitative and qualitative performance.

The remainder of this paper is organized as follows. The general MRF and CRF models are briefly presented in Section II. In Section III, we give a detailed description of the CRF + OO classification framework for HSR imagery. Section IV provides the experimental results, and the sensitivity analysis is discussed in Section V. In the final section, the conclusion is given.

II. Markov and Conditional Random Field Models

In probabilistic schemes, undirected graphical models, also referred to as random fields, have been used to incorporate spatial contextual information in computer vision. As described in Section I, this random field method of modeling the neighborhood interactions is very important in HSR remote sensing image classification and can improve the performance of pixelwise classification. MRF is the most popular undirected graphical model, incorporating the local dependencies between random variables using probabilistic frameworks [27], [36]. In the MRF classification framework, the posterior probability of the labels, given the image data, is described in the following, which is based on Bayes' rule:

$$P(\mathbf{x}|\mathbf{y}) \propto P(\mathbf{x},\mathbf{y}) = P(\mathbf{x})P(\mathbf{y}|\mathbf{x}) \tag{1}$$

where $\mathbf{x}$ represents the corresponding labels of the whole image, and $\mathbf{y}$ are the observed data from an input image. The MRF framework models the joint probability of the observed data and the corresponding labels $P(\mathbf{x},\mathbf{y})$. In MRF, $P(\mathbf{x})$ is the prior distribution of the labels $\mathbf{x}$, which is formulated as a Gibbs distribution to consider the spatial information, i.e.,

$$P(\mathbf{x}) = \frac{1}{Z}\exp\left\{-\sum_{c\in C}\psi_c(\mathbf{x}_c)\right\} \tag{2}$$

where $Z = \sum_{\mathbf{x}}\exp\{-\sum_{c\in C}\psi_c(\mathbf{x}_c)\}$, which is usually named the partition function, is a normalization factor, and the term $\psi_c(\mathbf{x}_c)$, which is called the potential function, is locally defined in the clique c, which is a subset of variables $\mathbf{x}_c \subseteq \mathbf{x}$. C is the set of all the cliques. For computational tractability, the observed data are assumed to be conditionally independent, and the likelihood $P(\mathbf{y}|\mathbf{x})$ can have a factorized form, i.e., $P(\mathbf{y}|\mathbf{x}) = \prod P(y|x)$. However, it should be noted that the spatial contextual interaction modeled by the term $P(\mathbf{x})$ is restricted to the labeling field and does not depend on the observed data.

Compared with generative MRF expending efforts to model the joint distribution $P(\mathbf{x},\mathbf{y})$, CRF directly models the posterior probability of the labels, given the image data $P(\mathbf{x}|\mathbf{y})$ [31], [36] with the following form, which is that we want to estimate in the classification task:

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{Z(\mathbf{y})}\exp\left\{-\sum_{c\in C}\psi_c(\mathbf{x}_c,\mathbf{y})\right\} \tag{3}$$

where $Z(\mathbf{y}) = \sum_{\mathbf{x}}\exp\{-\sum_{c\in C}\psi_c(\mathbf{x}_c,\mathbf{y})\}$ is the partition function, and the term $\psi_c(\mathbf{x}_c,\mathbf{y})$ denotes the potential function. In the classification problem, the most common CRF model is known as pairwise CRF, which models the spatial dependencies of pairs of random variables in the local neighborhoods. In theory, the potential functions can include unary potentials, pairwise potentials, and even high-order potentials, based on the different types of cliques in the observed data and their corresponding labels. CRF can incorporate more wide-ranging contextual information, as defined by the high-order neighborhood system and the cliques for the high-order potentials. However, optimization methods for the general case are infeasible. Therefore, in this paper, we use pairwise CRF with an 8-neighborhood.

It can be seen from the formulations of MRF in (1) and (2) and CRF in (3) that both the MRF and CRF models have the ability to consider the spatial contextual information in the label fields, but the CRF model also has the ability to permit interactions in the observed data. The CRF model directly models the posterior distribution as a Gibbs field, which allows the model to incorporate the contextual information in a more flexible way. The classification problem is typically formulated as pairwise CRF, which has been successfully applied in various image analysis fields [32]–[34], [39], [40]. However, the results of this model will always have an oversmooth appearance, to some degree, in the classification application. Therefore, to utilize the pairwise CRF model and alleviate the oversmoothing, a hybrid object-oriented CRF classification framework (CRF + OO) is used for the spatial contextual classification of HSR imagery in this paper.

III. Hybrid Object-Oriented CRF Classification Framework

In this paper, in order to make full use of the spatial contextual information for the classification of HSR remote sensing imagery, a hybrid object-oriented CRF classification framework (CRF + OO) is proposed by integrating CRF and an object-oriented classification method, which both incorporate information on the spatial context of each pixel. The proposed CRF + OO for HSR imagery can be described in three main steps, as shown in Fig. 1.

In the probabilistic pixelwise classification step, for the input HSR imagery and the training samples, a probabilistic pixelwise classification is first performed to obtain the initial classification result and the corresponding probability map. In this paper, the probabilistic SVM classifier is applied for this purpose.

The CRF classification framework step, as the second step, is performed by combining the spectral and spatial information, the goal of which is to obtain the classification result of CRF by carrying out the graph-cut-based α-expansion algorithm. This step takes the probability map from the first step and the original HSR imagery as the input and consists of two CRF models, which have the same pairwise potential but different unary potentials, which are called the log unary potential and the quasi-gamma unary potential, respectively. The corresponding models are represented as CRF-LOG and CRF-QG, respectively, in the classification framework, and they export two corresponding classification maps. Since CRF-LOG and CRF-QG both have the ability to incorporate spatial information but possess different unary potentials, they always have different degrees of smooth classification performance and can provide complementary information about the land cover types.

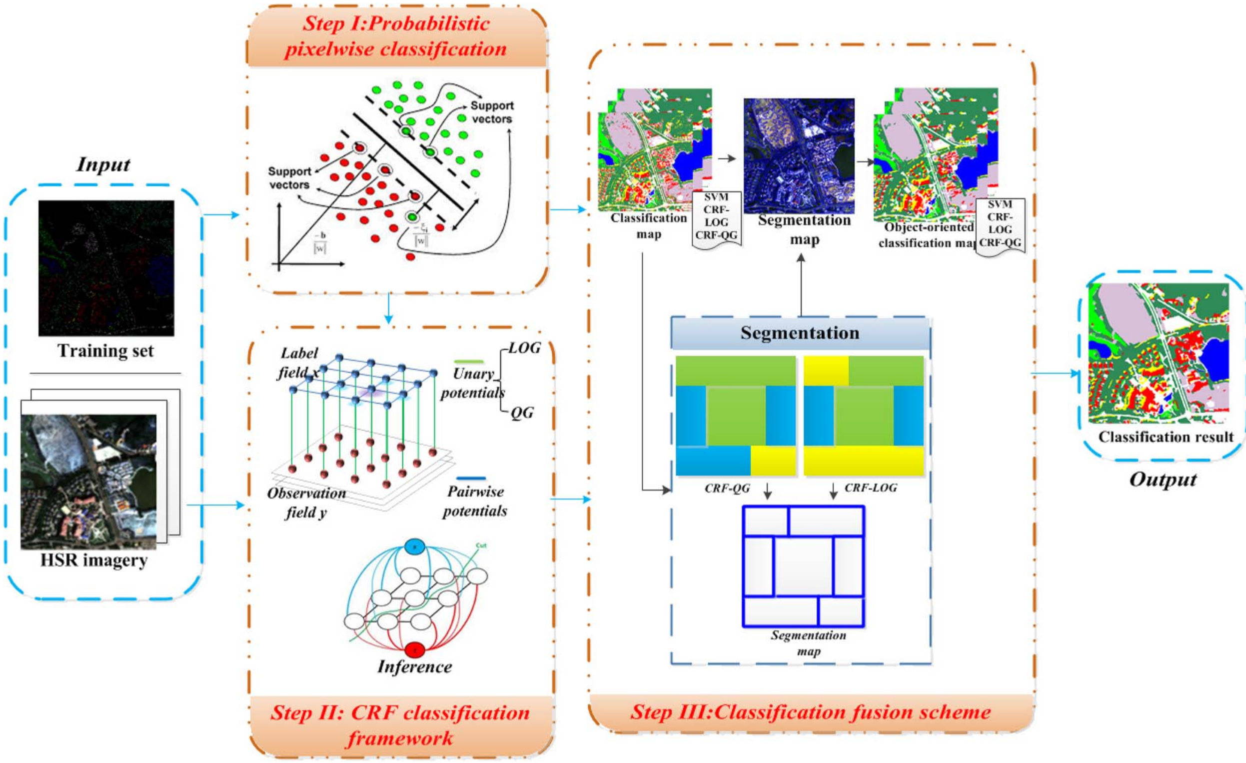


Fig. 1. Flowchart of the hybrid object-oriented CRF classification framework.

In the classification fusion step, the segmentation is first performed by the connected-component labeling algorithm, using the CRF-LOG and CRF-QG classification results obtained in the second step. Object-oriented classification maps can be then obtained by a majority voting strategy, using the segmentation map and the classification map. The final classification result is achieved by integrating the former object-oriented classifications at the object level.

A. Probabilistic Pixelwise Classification

The first step of the proposed framework consists of performing a probabilistic pixelwise classification of the HSR imagery. Various classifiers that have a probabilistic output can be used. However, the SVM classifier is suggested due to its high capacity and greater adaptability. In other words, whether in the case of high dimensions and a small training set or not, SVM always shows a good performance. More details about the SVM classifier can be found in [45]. In this step, the classification map and the corresponding probability map with the posterior probabilities that the samples belonging to the classes are expected to obtain. However, the output of general SVM, via the decision function, is only a class label without probability information. In order to obtain the probability estimates, Platt's formulation is used to map the output into the probability [45], [46]. Then, the remaining problem is how to expand the algorithm from binary classification to multiclass classification. The "one-against-one" approach is suggested in [45]. Assuming that k is the number of classes, we construct $k\ (k-1)/2$ classifiers so that each classifier trains data from two classes. Finally, in the classification, a majority voting strategy is used. More details can be found in [45].

B. CRF Classification Framework

Consider an ordered set of variables $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$, which denotes the set of spectral vectors of an HSR image, where N is the total number of pixels of the image, and $\mathbf{y}_i$ is the spectral vector of the image pixel $i \in V = \{1, 2, \ldots, N\}$ with length d (the number of spectral bands). The classification image is denoted by $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, where $\mathbf{x}_i$ in $\mathbf{x}$ takes the value of the label set $\mathbf{L} = \{1, 2, \ldots, K\}$. K is the number of classes.

Under Bayes' framework, the image classification can correspond to finding the maximum a posteriori (MAP) estimate of the label image, such that $\mathbf{x}_{\mathrm{MAP}} = \arg\max_{\mathbf{x}} P(\mathbf{x}|\mathbf{y})$. If $\mathbf{x}$ conditioned on $\mathbf{y}$ meets the Markov property, CRF directly models the posterior probability as a Gibbs distribution in (3). The corresponding Gibbs energy [40], [41] is defined as

$$E(\mathbf{x}) = -\log P(\mathbf{x}|\mathbf{y}) - \log Z(\mathbf{y}) = \sum_{c \in C} \psi_c(\mathbf{x}_c, \mathbf{y}). \quad (4)$$

Therefore, the MAP labeling x_{MAP} of the random field is given by

$$\mathbf{x}_{\mathrm{MAP}} = \arg\max_{\mathbf{x}} P(\mathbf{x}|\mathbf{y}) = \arg\min_{\mathbf{x}} E(\mathbf{x}). \quad (5)$$

Equation (5) shows that the maximization of the posterior probability $P(\mathbf{x}|\mathbf{y})$ is equivalent to the minimization of the energy function $E(\mathbf{x})$. A commonly used example of a CRF

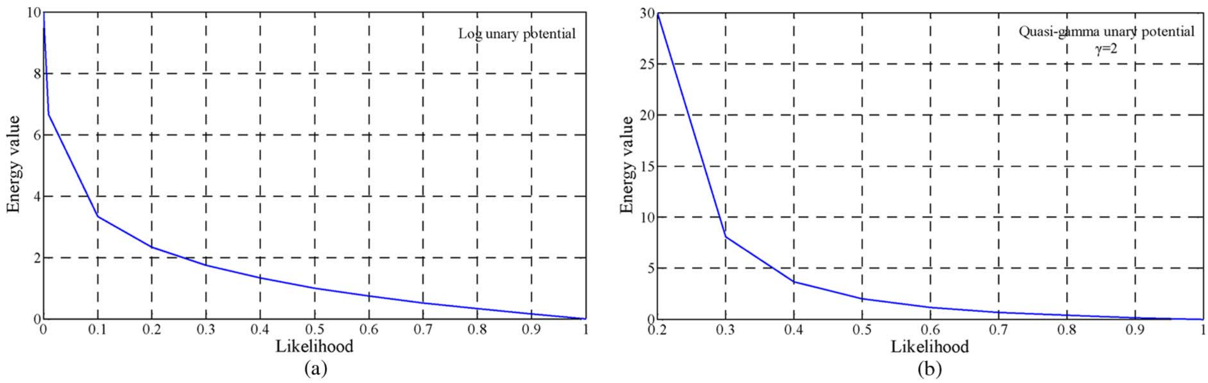


Fig. 2. Relationship between likelihood and energy value. (a) Log unary potential. (b) Quasi-gamma unary potential.

energy, which has been widely used for remote sensing image classification, can be written as the sum of the unary and pairwise potentials, i.e.,

$$E(\mathbf{x}) = \sum_{i \in V} \psi_i(x_i) + \lambda \sum_{i \in V, j \in N_i} \psi_{ij}(x_i, x_j) \tag{6}$$

where $\psi_i(x_i)$ is the unary potential term, and ψ_{ij} is the pairwise potential term, which is computed over the local neighborhood N_i of pixel i. The nonnegative constant λ trades off the strength of the pairwise potential against the unary potential.

After the establishment of the basic model (6), a remaining problem is how to formulate the two kinds of potentials (i.e., unary and pairwise potentials) and obtain the final labels by inference.

1) Unary Potentials: Unary potentials describe the cost of a single pixel taking a particular label, which depends on the local appearance features derived from the image. Typically, the unary potential $\psi_i(x_i)$ is defined as [3], [32]

$$\psi_i(x_i) = -\ln\left(P(x_i = l_k)\right) \tag{7}$$

where $P(x_i = l_k)$ is the probability of pixel x_i taking the label l_k, which can be given by various discriminative classifiers. In our work, these probability estimates are from "one-versus-one" SVM outputs, as described previously. In order to conveniently describe the details of the algorithm, we call the unary potential the "log unary potential" because of its log function form.

The unary potential's term is related to the pixelwise information, and the minimization of only this term's energy contribution by itself would be equivalent to a noncontextual Bayesian classification of the image. Therefore, it should be low when the corresponding pixel is correctly classified and high when it is misclassified. The log unary potential is formulated as the negative log likelihood of the pixel taking the related class labels. However, if a small object containing a few pixels with high likelihood is present in the image, these pixels tend to be assigned the class label of the surrounding objects because the log unary potential does not have enough ability to prevent the smoothing of the pairwise potentials. In order to suppress the oversmoothing of the classification result, the quasi-gamma unary potential is defined

$$\psi_i(x_i) = \gamma^{1/P(x_i = l_k)} - \gamma. \tag{8}$$

Fig. 2(a) and (b) shows the relationship between the likelihood and the energy value of the log and quasi-gamma unary potentials when $\gamma = 2$, respectively. As with the log unary potential, the quasi-gamma unary potential is also modeled as a decreasing function of likelihood. However, compared with the log unary potential, when the probability density estimate of the pixel taking the corresponding class label is high, the penalization of the quasi-gamma unary potential is still low. This means that the land cover class label of the pixel is maintained in minimizing the energy function when the confidence of the pixel taking the class label is high. In addition, the penalization is much greater when the likelihood is very low, and even when the likelihood is 0.1, this penalization can reach 1024, which implies that the land cover class label of the pixel is not expected to be kept when the confidence is less.

2) Pairwise Potentials: Pairwise potentials model a smoothness prior that encodes the fact that, in real images, neighboring pixels in homogeneous image regions usually take the same label. As is typical for pixel labeling with CRF, the contrast-sensitive smoothness prior $\psi_{ij}(x_i, x_j)$ takes the form of [32], [47], [48]

$$\psi_{ij}(x_i, x_j) = \begin{cases} 0, & \text{if } x_i = x_j \\ \frac{g(i,j)}{\|i-j\|^2}, & \text{otherwise} \end{cases} \tag{9}$$

where

$$g(i,j) = 1 + \theta_v \exp\left(-\theta_w \|x_i - x_j\|^2\right).$$

In this paper, the edge feature function $g(i,j)$ is designed to measure the difference in appearance between the neighboring pixels. The constant included in $g(i,j)$ is a bias term to be learned for removing small isolated regions; the pair of (i,j) represents the spatial location of neighboring pixels; parameter θ_v is a constant determining the degree of smoothness, and the quantity θ_w is the mean square difference between the spectral vectors over all the adjacent pixels in the image.

The pairwise potentials incorporate the spatial contextual information of each pixel, which is expressed in terms of its neighborhood, to be taken into account in the classification. When minimizing the energy function $E(\mathbf{x})$, the contrast-sensitive smoothness prior term penalizes the spatial inconsistencies among neighboring pixels with different class labels, while favoring, in the output classification map, the same land

α-EXPANSION algorithm
x^p := arbitrary labeling
Repeat
For each label $\alpha \in L = \{1,2,\ldots,K\}$
$x^n := \arg\min_x E_\alpha(x^p)$
If $E(x^n) < E(x^p)$ then
$x^p := x^n$
Until converged
Return x^p

Fig. 3. Graph-cut-based α-expansion algorithm.

cover class, except for boundary regions between homogeneous image regions. Furthermore, the penalization of different neighboring pixels with different class labels depends on the image data. According to the function $g(i, j)$, the penalization will be high when pixels i and j are similar, which means that the smoothness will be supported in minimizing the energy function. Furthermore, the penalization will be close to zero if they are very different, particularly in boundary regions, which means that their different land cover class labels will be kept in the process of optimization.

3) Inference by the Graph-Cut-Based α-Expansion Algorithm: Now that we have defined the class of the energy functions associated with each of the pixels' label assignments, we need to find a minimization method for the energy functions. However, global minimization of these energy functions is NP-hard [42] because searching for the best solution corresponding to the lowest cost pixel labeling (referred to as the inference problem) has an extremely large computational cost [41]. Fortunately, several researchers have proposed solutions to this inference problem, such as ICM, LBP, and graph cuts. Among these approaches, the graph cut methods, which formulate the energy minimization problem as a maximum flow problem over a suitable graph, have been the most popular [44].

Graph cuts have been proven to be fast and to converge to a global energy minimum in the case of binary classification. However, for HSR image classification, the label of an image is always a multivalued variable. The graph-cut-based α-expansion algorithm [43], which designs a special local search algorithm for the energy minimization, is more flexible. The local search of this algorithm works by repeatedly computing the global minimum of a binary labeling problem via a graph cut method in its inner loops. In this sense, the α-expansion algorithm reduces the problem with multivalued variables to a sequence of optimization subproblems with binary variables. The α-expansion algorithm is described in the following.

Given a current label $\mathbf{x}^p = \{x_i^p, i \in V\}$, to solve the problem of very small possible moves making the solution stick at a poor local minima, as with the ICM methods, the α-expansion 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; thus, 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. The α-expansion algorithm is summarized in the steps in Fig. 3.

For a particular label $\alpha \in \mathbf{L}$ in Fig. 3, an efficient way is needed to find the improved solution, which is denoted by $\mathbf{x}^n = \{x_i^n, i \in V\}$ with minimal $E_\alpha(\mathbf{x}^p)$, in one move of the expansion algorithm in line 4. Since the expansion moves are fundamentally binary, we can encode the moves of the expansion algorithm by the binary variables $t = \{t_i, i \in V\}$ as

$$x_i^n = \begin{cases} \alpha & \text{if } t_i = 0 \\ x_i^p & \text{if } t_i = 1. \end{cases} \tag{10}$$

Therefore, the original multilabel energy of (6) can be transformed to a binary energy E_α by (10), whose optimization process rapidly converges and results in a strong local minimum when E_α is a submodular function. It is generally known that a binary energy function can be efficiently minimized if it is submodular [41]. Therefore, which class of the multilabel energy functions can result in submodular E_α must be considered. Fortunately, Boykov *et al.* [43] explained that a sufficient condition for this submodularity is the metricity of the pairwise potentials. Pairwise potentials are called metric, if they satisfy

$$\begin{aligned} &\psi(l_a, l_b) = 0 \quad \Leftrightarrow \quad l_a = l_b \\ &\psi(l_a, l_b) = \psi(l_b, l_a) \geq 0 \\ &\psi(l_a, l_b) \leq \psi(l_a, l_c) + \psi(l_c, l_b) \end{aligned} \tag{11}$$

where $forall\ l_a, l_b, l_c \in L$.

It can be easily verified that the pairwise potential energy function (9) satisfies the submodularity condition (11) [32]. Therefore, the inference can be efficiently minimized using the graph-cut-based α-expansion algorithm.

C. Classification Fusion Scheme

The previous section gives a description of how to combine the spatial and the spectral information by modeling the unary potentials and the pairwise potentials of CRF. However, due to the land cover complexity of HSR imagery, many structures may be oversmoothed in the classification map when using CRF with log unary potential. In order to overcome this oversmoothing of the log unary potential, the quasi-gamma unary potential is proposed. The quasi-gamma unary potential gives the likelihood of a pixel taking a land class label a larger weight, which makes the quasi-gamma unary potential dependent on the accuracy of the probability estimation. However, due to the spatial complexity and the spectral variability of HSR imagery, it is difficult to arrive at an estimate, particularly in the case of limited samples. The oversmooth spatial classification map obtained by CRF with log unary potential can provide complementary information for the classification result by CRF with quasi-gamma unary potential. Therefore, a classification fusion scheme is put forward to take full advantage of both of these CRF results.

The classification fusion scheme results in a more robust and excellent classification result by making full use of the classification information obtained in the first two steps. Since the results of the two CRF models incorporate the pixels' spatial contextual information, the CRF-LOG and CRF-QG maps, which have a good ability to obtain homogeneous regions, have a better performance than the SVM map obtained in the

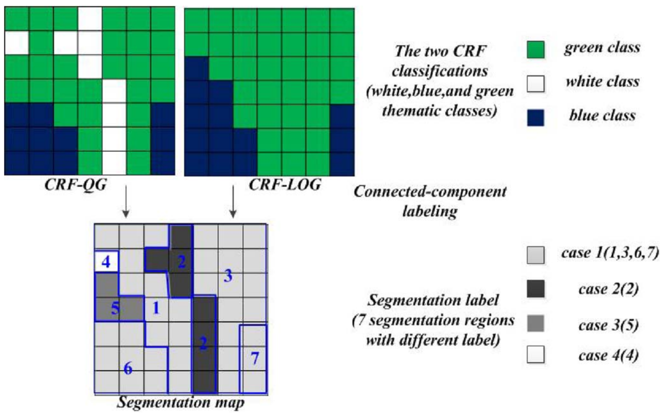


Fig. 4. Example of segmentation using the connected-component labeling algorithm.

first step. Therefore, the CRF-LOG and CRF-QG maps are selected to obtain the segmentation result. The segmentation is performed by the connected-component labeling algorithm, finding regions of connected pixels that have the same value in the CRF-LOG and CRF-QG classification results. The classical connected-component algorithm with an 8-neighborhood, using a union-find data structure, is used to assign labels to the objects [49], [50]. The scale of the segmentation meets the requirement of all the land cover classes in that a good classification result basically reflects the distribution of the materials of interest. As with the example shown in Fig. 4, which is the illustrative example of segmentation using the connected-component labeling algorithm in classification maps with various color thematic classes, the following can be found.

1) If the region is labeled to the same land cover type by CRF-QG and CRF-LOG, which means that its label is trusted, it will be labeled in the segmentation map, such as the 1, 3, 6, and 7 labels in Fig. 4, so that its label can remain in the next process.
2) If a small object with a relatively high possibility appears in the CRF-QG classification, it will be labeled in the same way as label 2 in Fig. 4 in the segmentation map, which means that it is not easily smoothed by the classification fusion scheme.
3) In the case of one region being smoothed to different class labels by CRF-QG and CRF-LOG, its label is untrusted. Therefore, this region is labeled in the same way as label 5 in Fig. 4 in the segmentation map, which allows it to be considered in the next step.
4) In the case of one region being heavily affected by noise and not being smoothed by CRF-QG or CRF-LOG, it will be assigned a label in the same way as label 4 in Fig. 4 in the segmentation map, so that it can be smoothed in the next process.

The segmentation map is achieved by the connected-component labeling algorithm. The object-oriented classification results, which are called the SVM-OO map, the CRF-LOG-OO map, and the CRF-QG-OO map, respectively, are then obtained by using the corresponding classification map to undertake majority voting. It is the segmentation that can allow the integration of the former classifications at the object level. In order to achieve the final classification result, the key problem is to design a classification fusion scheme that makes full use of all the object-oriented classification information, which is shown in Fig. 5. We analyze each segmentation region as follows.

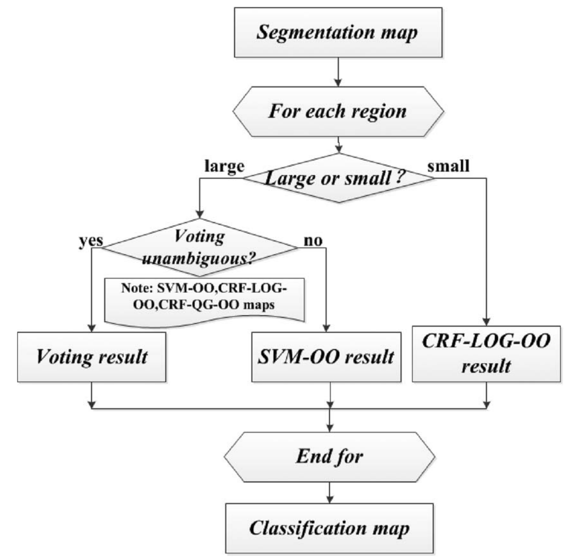


Fig. 5. Flowchart of the classification fusion scheme.

1) If a region is very small, which is controlled by a defined region size represented by symbol S, it is expected to be labeled by a more smooth result to alleviate the noise effects. This is because the small regions are always a result of the difference between the CRF-LOG and CRF-QG maps, due to the spectral variability and noise.
2) If a region is large enough, the label of the region is considered to be a voting of the SVM-OO, CRF-LOG-OO, and CRF-QG-OO maps. Finally, the region is designated to be in the class with the maximum number of votes. However, in the case of three classes having identical votes, it may be difficult to determine the label. Analyzing the source of the whole data, we can find that the CRF-LOG-OO and CRF-QG-OO maps both have the possibility of being smoothed by neighboring region pixels. Therefore, the SVM-OO result is used in this case.

IV. Experimental Results

Three HSR remote sensing image data sets [22], consisting of two multispectral HSR images (QuickBird and IKONOS) and a hyperspectral HSR image (HYDICE), are presented to test the performance of the proposed CRF + OO classification framework. The comparison experiments are conducted by pixelwise classification, object-oriented classification, and random field methods. The SVM classifier implemented in LibSVM [45] is used to make the pixelwise classification. The radial basis function kernel is selected in all the experiments on account of its excellent performance in HSR image classification, and the strategy of cross-validation is applied to determine the optimal parameters. As for the parameters C and γ of SVM, the range of C is set from 2^0 to 2^{10}, while the range of γ is set from

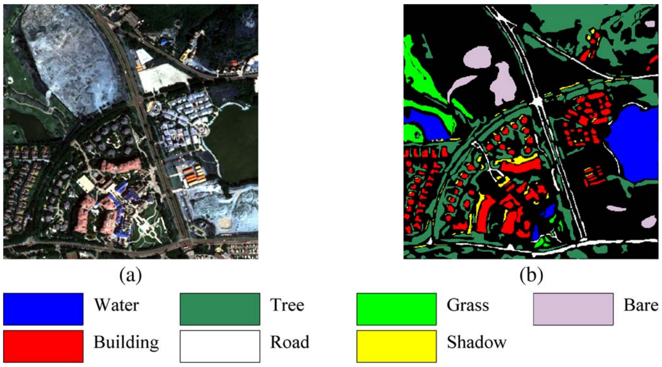


Fig. 6. Fancun QuickBird data set. (a) RGB false-color image (3, 2, 1). (b) Ground truth image.

2^{-10} to 2^{10}. The object-oriented classification methods used in our comparison experiments are methods based on MSS [19] and the multiresolution segmentation algorithm in eCognition 8.0 (FNEA) [18], using a majority voting strategy [21], [22] with the same pixelwise SVM classification result, and are denoted by MSS-OO and FNEA-OO, respectively. For the MSS algorithm, the spatial/spectral bandwidth parameters are chosen from the range [1, 19] and [10, 200], respectively. For the FNEA algorithm, the importance of each band is supposed to be equal in the experiments; thus, the image layer weights of the segmentation are all set to be 1; the shape and compactness parameters in the composition of the homogeneity criterion are set to 0.1 and 0.5, respectively. The remaining parameter, i.e., the scale parameter, has a major impact on the segmentation and is selected from the range [10, 100] by a grid search method. Moreover, the MSVC, as a random field method combining SVM and MRF in an integrated framework for contextual image classification, is also used. More details can be found in [30].

In our experiments, not only is the classification result of CRF + OO shown, but the intermediate segmentation result of the CRF + OO classification framework is also presented and is represented by CRF-SEG, to prove its good performance when compared with the traditional object-oriented classification methods. Moreover, the two CRF classification results (CRF-LOG and CRF-QG) in the CRF classification framework are also shown in the series of experiments.

To assess the experimental results, four kinds of accuracies are used, which are the accuracy of each class, the overall accuracy (OA), the average accuracy (AA), and the kappa coefficient (Kappa). OA is the fraction of correctly classified pixels, with regard to all the pixels of that ground truth class, and AA is the average of all the class accuracies. To allow a fair comparison, the classification results with the highest OA are selected for all the classification algorithms.

A. Experimental Data Sets

The first experiment image is from the Fancun area in Hainan Province, China, and was acquired in January 2010 by the QuickBird sensor. The image is of 400 × 400 pixels, with a spatial resolution of 2.4 m, and four multispectral channels. Fig. 6(a) gives an overview of this data set by combining the first, second, and third bands. The corresponding ground truth is shown in Fig. 6(b). Seven classes of interest are considered, as detailed in Table I, which also shows the number of the training and test samples for each class. The training samples are randomly chosen from the reference ground truth data.

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

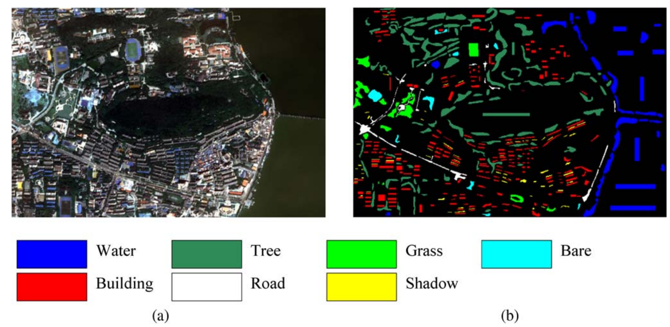


Fig. 7. Wuhan IKONOS data set. (a) RGB false-color image (3, 2, 1). (b) Ground truth image.

TABLE II
CLASS INFORMATION OF THE WUHAN DATA SET

Class name	Training samples	Test samples
Building	320	6544
Grass	142	2696
Water	155	7615
Shadow	90	807
Bare soil	90	1040
Tree	294	13702
Road	271	1536

Data from a different sensor are used in the second experiment to confirm the validity of the proposed algorithm. The image is of an urban area, with a spatial resolution of 4 m, and was acquired by the IKONOS satellite from Wuhan in Hubei Province, China. Fig. 7(a) and (b) presents an intuitive view of this image and the corresponding land cover types, respectively. The image size is 400 × 600 pixels, with blue, green, red, and near-infrared spectral channels. Like the Fancun data set, this image also contains seven thematic classes of interest. Table II shows the training and test samples and their labels.

The third experiment is performed using a subset of the Washington DC data set, which is a hyperspectral image acquired by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor. The image size is 307 × 280 pixels, with 191 bands. Fig. 8(a) shows its false-color appearance, and the corresponding land cover types of interest are shown in Fig. 8(b). Table III gives the number of training and test samples for each class of interest.

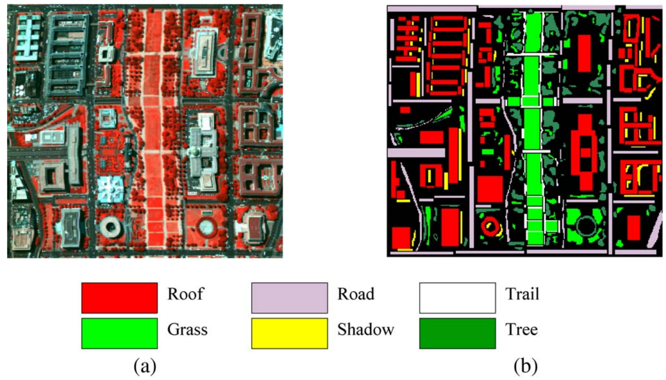


Fig. 8. Washington DC HYDICE data set. (a) RGB false-color image (60, 27, 17). (b) Ground truth image.

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

B. Experimental Setup

For the Fancun, Wuhan, and Washington DC data sets, the optimal parameters giving the highest accuracies for all the classification approaches are set as follows. The parameters C/γ of SVM are set to 256/0.5, 4096/0.5, and 16 384/0.125, respectively. For the MSS algorithm, the spatial/spectral bandwidth parameters are chosen as 15/20, 1/140, and 19/10, respectively. For the FNEA algorithm, the segmentation scale parameters are selected to be 20, 20, and 80, respectively. The λ and θ_v parameters of CRF-LOG are set to 1.2/0.2, 0.6/1.8, and 0.3/1.5, respectively; whereas they are selected as 190/2.1, 160/2.1, and 13/1.8 for CRF-QG. Finally, for CRF + OO, the size parameters represented by the symbol S are set to be 25, 5, and 20, respectively. In addition, it should be noted that the computation times of CRF + OO are only 10, 15, and 10 s, respectively, without considering the time cost of SVM, using a computer of 3.1 GHz with 8-GB random access memory, because the α-expansion inference algorithm based on graph cut is very fast.

C. Experimental Results and Analysis

For the Fancun, Wuhan, and Washington DC data sets, the classification results are shown in Figs. 9–11; the classification maps of the different algorithms (i.e., SVM, FNEA-OO, MSS-OO, CRF-LOG, CRF-QG, CRF-SEG, MSVC, and CRF + OO) are respectively presented in the subfigures of the corresponding classification results. From these classification maps, it can be seen that the SVM algorithm, which does not consider any neighborhood spatial contextual information, results in a mass of isolated salt-and-pepper classification noise, whereas the algorithms taking into account the neighborhood interactions (i.e., FNEA-OO, MSS-OO, CRF-LOG, CRF-QG, CRF-SEG, MSVC, and CRF + OO) exhibit much better visual classification results.

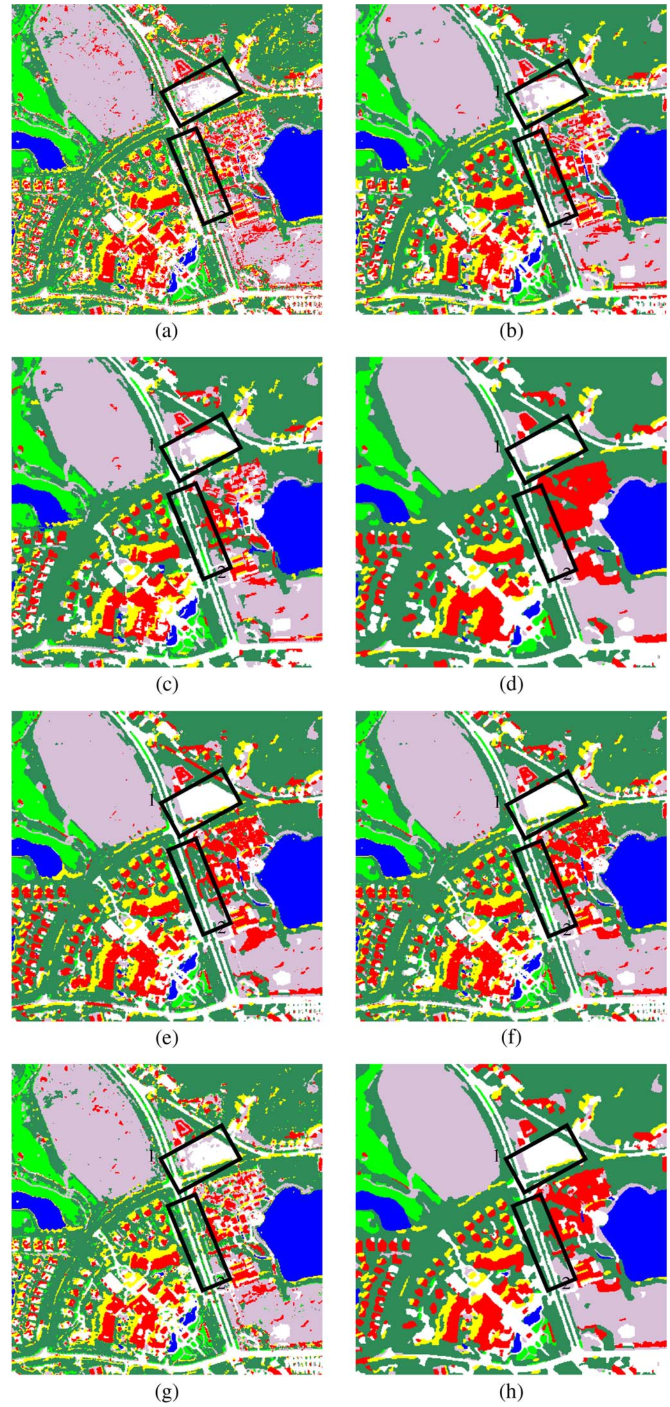


Fig. 9. Classification results for the Fancun QuickBird data set. (a) SVM. (b) FNEA-OO. (c) MSS-OO. (d) CRF-LOG. (e) CRF-QG. (f) CRF-SEG. (g) MSVC. (h) CRF + OO.

As for the CRF-LOG algorithm, it produces an oversmooth classification result, and it is the λ parameter that mainly controls the strength of the spatial contextual information. As shown in Figs. 9(d), 10(d), and 11(d), this oversmoothing phenomenon is obvious for the more complex man-made objects for the Fancun and Wuhan data sets. However, the oversmooth performance is less obvious in the result of the Washington

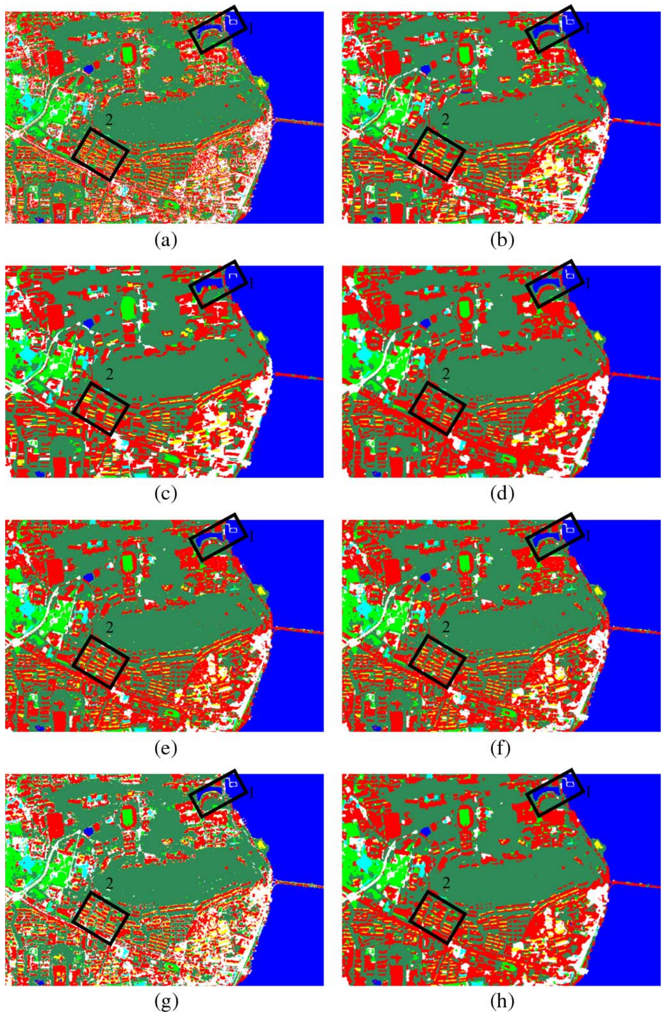


Fig. 10. Classification results for the Wuhan IKONOS data set. (a) SVM. (b) FNEA-OO. (c) MSS-OO. (d) CRF-LOG. (e) CRF-QG. (f) CRF-SEG. (g) MSVC. (h) CRF + OO.

DC experiment since there are many spectral bands that have a certain ability to distinguish the varying land cover types. This means that the optimal λ parameter, which mainly controls the strength of the spatial contextual information, is very small, i.e., the neighborhood spatial interactions are less.

For the CRF-QG algorithm, which is proposed to alleviate the oversmoothing, it is shown that the oversmoothing phenomenon is less serious in Figs. 9(e), 10(e), and 11(e). However, it was noted previously that CRF-QG puts more emphasis on the spectral information; thus, the classification depends on the pixelwise classification result to a great extent, which perhaps explains the misclassification and the undersmoothing of CRF-QG, as shown in black box 2 in Fig. 9(e).

CRF-SEG uses the complementary information of CRF-LOG and CRF-QG to obtain the segmentation maps presented in Figs. 9(f), 10(f), and 11(f), which lays a solid foundation for the information fusion on the object layer for the CRF + OO algorithm. The classification result of CRF + OO is presented in Figs. 9(h), 10(h), and 11(h), and it is shown to have a competitive performance, as highlighted in black boxes 1 and 2.

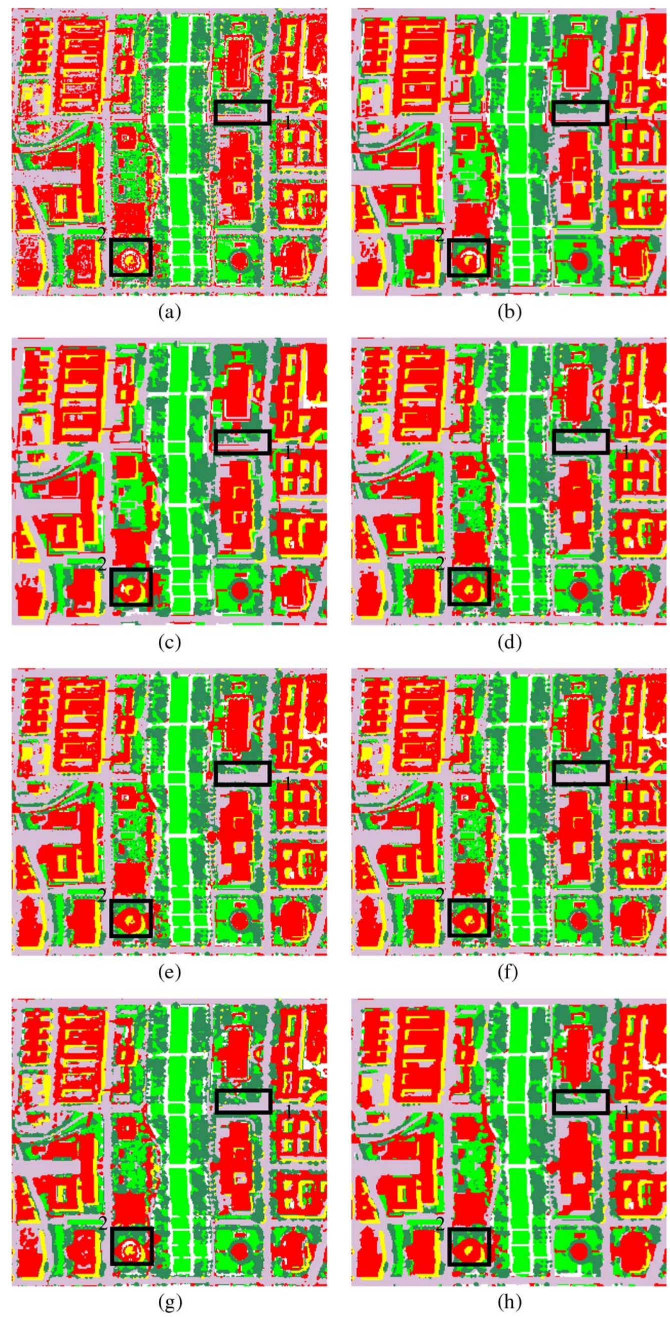


Fig. 11. Classification results for the Washington DC HYDICE data set. (a) SVM. (b) FNEA-OO. (c) MSS-OO. (d) CRF-LOG. (e) CRF-QG. (f) CRF-SEG. (g) MSVC. (h) CRF + OO.

The quantitative performances with the highest classification accuracies obtained by SVM, FNEA-OO, MSS-OO, CRF-LOG, CRF-QG, CRF-SEG, MSVC, and CRF + OO are reported in Tables IV–VI. From the tables, a similar conclusion can be reached, in that the algorithms taking spatial contextual information into account show a great improvement over the pixelwise SVM classification in classification accuracy. Moreover, the accuracy of CRF-SEG is higher than the two other object-oriented classification methods (i.e., FNEA-OO and MSS-OO), which confirms that the segmentation scale of CRF-SEG can be adaptively obtained for varying land cover types. For the Wuhan data set, the quantitative performance

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
CRF-LOG	99.28	99.23	94.63	99.73	84.03	96.22	88.36	96.20	94.50	0.9485
CRF-QG	98.53	97.80	96.38	99.50	86.96	86.81	96.89	95.58	94.70	0.9404
CRF-SEG	99.40	97.87	96.60	99.50	80.79	95.41	96.76	95.47	95.19	0.9389
MSVC	98.86	96.34	94.65	99.25	67.43	94.01	96.95	92.51	92.50	0.8995
CRF+OO	99.68	98.96	96.74	99.68	85.62	96.11	94.40	96.70	95.88	0.9554

TABLE V
CLASSIFICATION ACCURACIES FOR THE WUHAN IKONOS DATA SET

Methods	Accuracy (%)							OA (%)	AA (%)	Kappa
	Building	Grass	Water	Shadow	Bare	Tree	Road			
SVM	72.30	92.03	98.06	88.48	81.06	96.97	81.90	90.69	87.26	0.8756
MSS-OO	78.47	94.88	98.96	80.79	89.04	98.91	82.36	93.18	89.06	0.9081
FNEA-OO	77.74	94.51	98.35	89.59	86.92	98.47	81.71	92.80	89.61	0.9034
CRF-LOG	94.27	94.73	98.65	69.14	90.10	98.96	68.10	95.27	87.71	0.9359
CRF-QG	94.12	94.21	98.57	91.33	85.29	98.66	67.84	95.43	90.00	0.9382
CRF-SEG	94.25	94.58	98.62	89.22	86.15	98.79	67.64	95.52	89.89	0.9394
MSVC	75.47	93.66	98.46	92.69	85.10	98.09	87.57	92.46	90.15	0.8990
CRF+OO	94.67	94.99	98.62	87.98	86.83	99.04	67.51	95.72	89.94	0.9420

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
CRF-LOG	96.88	96.33	97.31	97.06	97.69	95.53	96.69	96.80	0.9558
CRF-QG	97.21	96.27	97.70	96.68	97.76	94.03	96.57	96.61	0.9543
CRF-SEG	97.21	96.42	97.70	96.98	97.69	94.06	96.66	96.68	0.9554
MSVC	94.18	94.71	95.40	96.28	97.24	96.30	95.12	95.69	0.9351
CRF+OO	97.88	97.49	94.63	97.04	97.88	96.56	97.36	96.91	0.9647

of CRF-QG reported in Table V shows a great improvement in AA, and the 20% accuracy improvement (from 69.14% to 91.33%) of shadow for CRF-QG compared with CRF-LOG shows that CRF-QG puts more emphasis on spectral information; thus, it has the ability to restrain some land cover types from smoothing. Finally, from the classification accuracy tables, it can be seen that CRF + OO obtains the highest accuracy.

V. SENSITIVITY ANALYSIS

In Section IV, the three sets of experimental results indicate that CRF + OO performs well. Furthermore, it has only one parameter that has an impact on the classification performance. Here, an additional analysis of the effect of the size parameter represented by S is given. This parameter is the fusion parameter reflecting the confidence for the classification of CRF-LOG. In addition, in the hybrid classification framework, the CRF parameters (CRF-LOG and CRF-QG) indirectly affect the classification performance of CRF + OO; thus, a corresponding sensitivity analysis is also given. Finally, a sensitivity analysis of the number of training samples is conducted to further investigate the robustness of CRF + OO. Additional experiments are also conducted to evaluate the effect of these parameters on the Fancun, Wuhan, and Washington DC data sets.

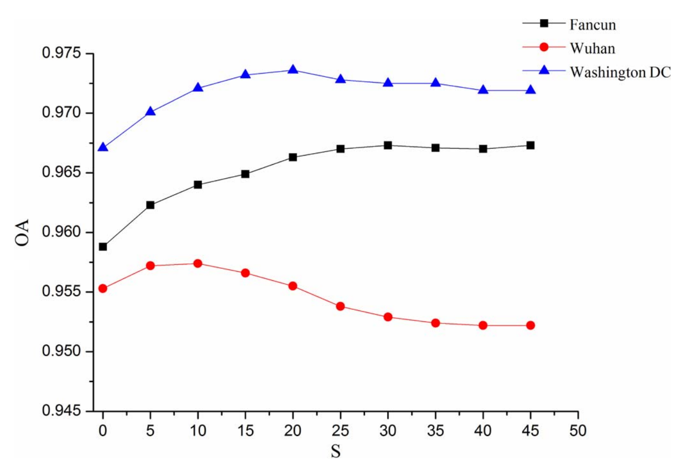


Fig. 12. Sensitivity analysis for the fusion parameter size, with the three data sets (Fancun QuickBird, Wuhan IKONOS, and Washington DC HYDICE images).

A. *Sensitivity Analysis for the Fusion Parameter S*

The parameter S is the fusion parameter size, which reflects the confidence for the classification of CRF-LOG. In order to analyze its sensitivity, another set of experiments is performed for the Fancun, Wuhan, and Washington DC data sets by varying the parameter S from 0 to 45, with an interval of 5. The result is reported in Fig. 12.

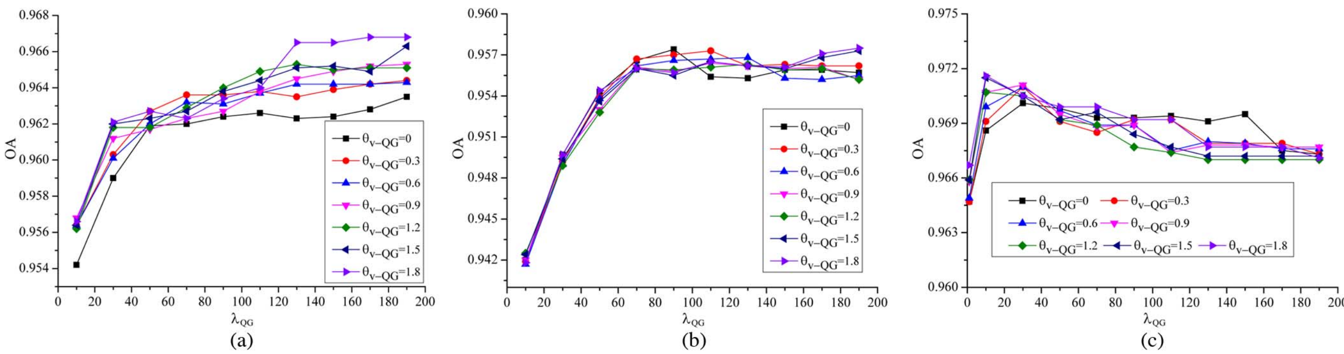


Fig. 13. Sensitivity analysis for the parameters of CRF-QG in the CRF + OO classification fusion framework, with the three data sets. (a) Fancun QuickBird image. (b) Wuhan IKONOS image. (c) Washington DC HYDICE image.

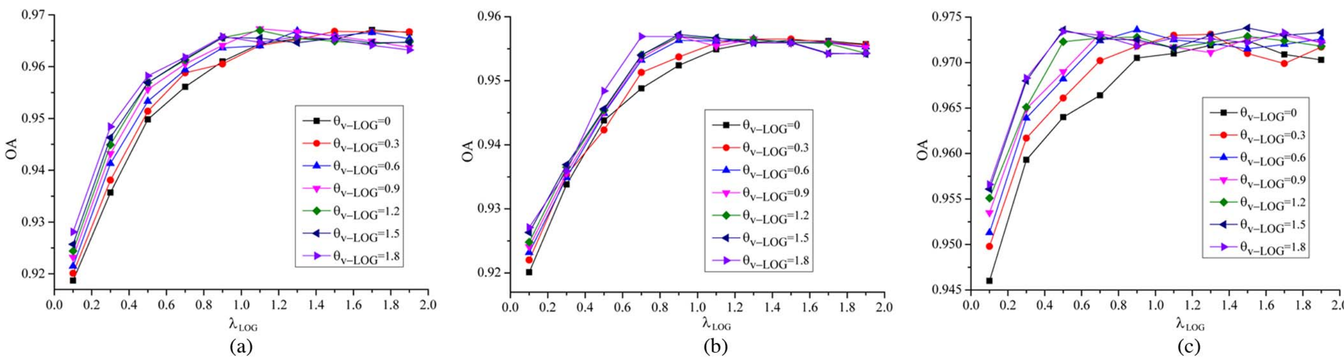


Fig. 14. Sensitivity analysis for the parameters of CRF-LOG in the CRF + OO classification fusion framework, with the three data sets. (a) Fancun QuickBird image. (b) Wuhan IKONOS image. (c) Washington DC HYDICE image.

As shown in Fig. 12, with the gradual increase in the size of S, the classification accuracy of CRF + OO first increases for all the data sets. The accuracy then remains roughly stable for the Fancun data set but slightly decreases for the Wuhan and Washington DC data sets. The increase in the parameter S means an increase in the smoothing effect of CRF-LOG in a defined small region; thus, the capability of removing the salt-and-pepper classification noise becomes increasingly strong, which leads to the initial increase in the classification accuracy. However, after the parameter reaches a certain value, the spatial smoothing effect can become too much, and the result of CRF-LOG tends to be oversmoothed. However, the fusion result is not susceptible to the classification of CRF-LOG in large regions due to the use of a majority voting strategy, so that there may be an appropriate stable result in the CRF + OO experiments, which leads to the remaining unchanged or slight decrease in the classification accuracies.

B. Sensitivity Analysis for the CRF Parameters

In order to study the sensitivity for the parameters of CRF-QG in the classification fusion framework, the other parameters are set to be constant. In the additional experiments for the Fancun, Wuhan, and Washington DC data sets, λ and θ_v of CRF-LOG and the size parameter are set to be 1.1/1.2/25, 0.9/1.5/5, and 0.5/1.5/20, respectively. In addition, parameter λ of CRF-QG, i.e., λ_{QG}, is set from 10 to 190 with an interval of 20, and parameter θ_v of CRF-QG, i.e., $\theta_{v\text{-QG}}$, is selected from 0 to 1.8 with an interval of 0.3 for the three data sets. The relationship between the parameters λ and θ_v of CRF-QG and the classification accuracy (OA) is reported in Fig. 13(a)–(c).

As shown in Fig. 13(a)–(c), the classification accuracies of CRF + OO first increase and then remain approximately stable, on the whole, with the increase in parameter λ_{QG} when keeping $\theta_{v\text{-QG}}$ unchanged. The increase in parameter λ_{QG} means an increase in the spatial effect; thus, the capability of using the neighborhood spatial information to remove the salt-pepper classification noise becomes increasingly strong, which leads to the initial increase in the classification accuracy. However, after parameter λ_{QG} reaches a certain value, the spatial smoothing effect may become too great. However, since the quasi-gamma unary potential strengthening the spectral information trades off the smoothing information and the spectral information, there may be an appropriate stable result in the CRF + OO experiments. For parameter $\theta_{v\text{-QG}}$, a fine-tuning function for the classification accuracy is found when keeping λ_{QG} unchanged.

When studying the impact of the parameters of CRF-LOG in the CRF + OO framework on the classification accuracy, we set the other parameters of CRF + OO to be constant. Another set of experiments is performed for the Fancun, Wuhan, and Washington DC data sets, where λ and θ_v of CRF-QG and the size parameter are set to be 190/2.1/25, 160/2.1/5, and 13/1.8/20, respectively. In addition, parameter λ of CRF-LOG, i.e., λ_{LOG}, is set from 0.1 to 1.9 with an interval of 0.2, and parameter θ_v of CRF-LOG, i.e., $\theta_{v\text{-LOG}}$, is selected from 0 to 1.8 with an interval of 0.3 for the three data sets. The relationship between λ_{LOG}, $\theta_{v\text{-LOG}}$, and the classification accuracy (OA) is reported in Fig. 14(a)–(c).

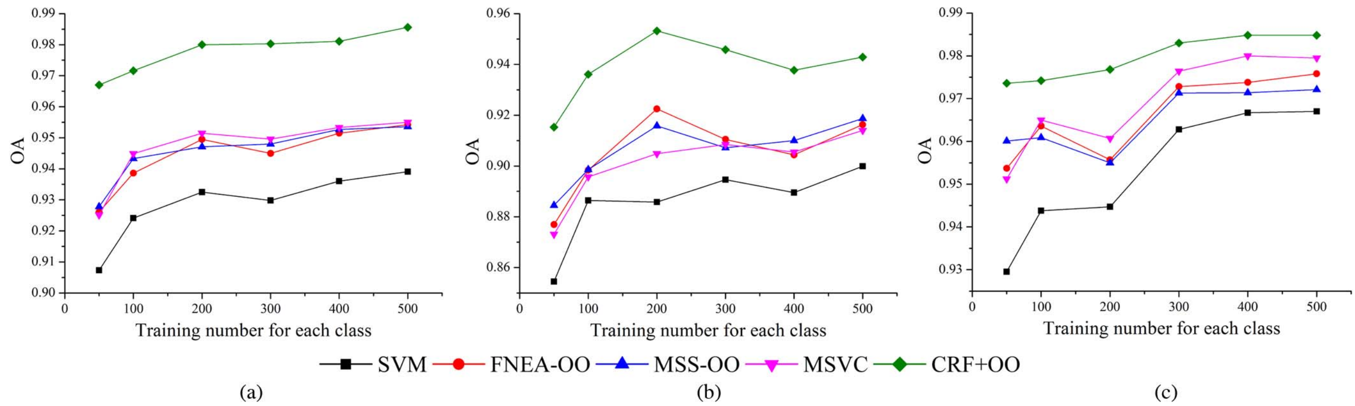


Fig. 15. Sensitivity analysis for the number of training samples, with the three data sets. (a) Fancun QuickBird image. (b) Wuhan IKONOS image. (c) Washington DC HYDICE image.

From Fig. 14(a)–(c), we can draw a similar conclusion, in that the classification accuracies of CRF + OO first increase and then remain approximately stable, on the whole, as parameter λ_{LOG} increases when keeping $\theta_{v\text{-LOG}}$ unchanged. The reason is again that the strengthening of the spatial effect results in the classification accuracy increasing at first with the gradual increase in parameter λ_{LOG}. After parameter λ_{LOG} reaches a certain value, the spatial smoothing effect becomes dominant, resulting in an oversmooth classification performance for CRF-LOG. However, since the classification result of CRF-LOG is first combined to produce the segmentation map and is then utilized in the classification fusion scheme at the object level for CRF + OO, the smoothing effect has only a limited negative influence on the final fusion classification, which results in stable classification accuracy. When analyzing the effect on the classification accuracy of parameter $\theta_{v\text{-LOG}}$, a fine-tuning function for the classification accuracy is found when keeping λ_{LOG} unchanged.

C. Sensitivity Analysis for the Training Set Size

To analyze the sensitivity to the number of training samples, the training numbers for each class are varied between 50, 100, 200, 300, 400, and 500 for the Fancun, Wuhan, and Washington DC data sets. The classification result with the highest OA accuracy is selected for all the classification algorithms at each training number. It should be noted that the training samples are randomly selected from the overall ground truth, and the remaining samples are used to evaluate the classification accuracies.

As shown in Fig. 15(a)–(c), as the number of training samples increases, all the classification algorithms show a similar trend. The experiments confirm that the classification accuracies of the methods incorporating spatial contextual information (MSS-OO, FNEA-OO, MSVC, and CRF + OO) are all better than SVM. In addition, the two object-oriented classification approaches (MSS-OO and FNEA-OO) have a similar ability for HSR image classification. The classification accuracy of MSVC is slightly higher than the two object-oriented classification approaches with the Fancun and Washington DC data sets, whereas MSVC shows a slightly poorer performance than the object-oriented classification methods for the Wuhan data set. Since CRF + OO uses a classification fusion scheme at the object level to integrate the classification results of CRF, it has the best performance of all the approaches, with all three data sets.

VI. Conclusion

In this paper, a hybrid object-oriented CRF classification framework (CRF + OO), which integrates object-oriented classification with pairwise CRF classification, has been developed to address the segmentation scale choice problem for object-oriented classification and the oversmoothing phenomenon of CRF for HSR remote sensing imagery. One reason for the oversmoothing in pairwise CRF is that the pairwise spatial term is dominant when compared with the unary spectral potential. Therefore, the quasi-gamma unary potential is proposed to restrain the oversmooth effect of pairwise CRF, which strengthens the spectral information, to trade off the spatial smoothing effect and the spectral effect. In order to make the model more robust and adaptive to HSR imagery, a classification fusion scheme at the object level is put into use to achieve homogeneous regions in the final classification result. The objects are obtained by a connected-component labeling algorithm, combining the CRF classification results. This segmentation map using the supervised classification result always has an approximately optimal scale for each land cover type. Therefore, the final classification result is achieved by a classification fusion scheme integrating the classification information at the object level. Three real data experiments using three types of HSR images from QuickBird, IKONOS, and HYDICE demonstrate the effectiveness of the proposed algorithm, compared with other state-of-the-art classification algorithms, and they confirm that the CRF + OO algorithm has a competitive quantitative and qualitative performance for HSR image classification.

Another important reason for the oversmoothing in pairwise CRF is that the limited neighborhood restricts the representational power of the model, meaning that it is unable to capture the rich statistics of HSR images. In our future work, high-order random fields will be studied to model these complex statistics.

ACKNOWLEDGMENT

The authors would like to thank the Editor, the Associate Editor, and the anonymous reviewers for their helpful comments and suggestions. The authors would also like to thank Beijing Panorama Space Technology Co., Ltd., for providing the free QuickBird image and the research group supervised by D. Landgrebe of Purdue University, West Lafayette, IN, USA, for providing the free downloads of the HYDICE 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. 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.

[3] G. M. Foody, "Status of land cover classification accuracy assessment," *Remote Sens. Environ.*, vol. 80, no. 1, pp. 185–201, Apr. 2002.

[4] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1267–1279, Oct. 2010.

[5] 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.

[6] S. Moustakidis, G. Mallinis, N. Koutsias, J. B. Theocharis, and V. Petridis, "SVM-based fuzzy decision trees for classification of high spatial resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 149–169, Jan. 2012.

[7] 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.

[8] F. Del Frate, F. Pacifici, G. Schiavon, and C. Solimini, "Use of neural networks for automatic classification from high-resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 800–809, Apr. 2007.

[9] 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.

[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] 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.

[12] J. Li and D. C. Tao, "Exponential family factors for Bayesian factor analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 6, pp. 964–976, Jun. 2013.

[13] J. Li and D. Tao, "On preserving original variables in Bayesian PCA with application to image analysis," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4830–4843, Dec. 2012.

[14] J. Li and D. Tao, "Simple exponential family PCA," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 3, pp. 485–497, Mar. 2013.

[15] 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.

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

[17] 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.

[18] M. Baatz and A. Schäpe, "Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation," in *Angewandte Geographische Informationsverarbeitung XII*, A. Strobl, A. Blaschke, and A. Griesebner, Eds. Heidelberg, Germany: Wichmann-Verlag, 2000, pp. 12–13.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral-spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.

[24] 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.

[25] M. Kim, M. Madden, and T. A. Warner, "Forest type mapping using object-specific texture measures from multispectral IKONOS imagery: Segmentation quality and image classification issues," *Photogramm. Eng. Remote Sens.*, vol. 75, no. 7, pp. 819–829, Jul. 2009.

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

[27] S. Y. Chen, H. Y. Tong, and C. Cattani, "Markov models for image labeling," *Math. Probl. Eng.*, vol. 2012, pp. 814356-1–814356-18, 2012.

[28] 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.

[29] 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.

[30] 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.

[31] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th ICML*, 2001, pp. 282–289.

[32] 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. Eur. Conf. Comput. Vis.*, 2006, pp. 1–15.

[33] 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.

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

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

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

[37] 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.

[38] 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.

[39] 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.

[40] 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.

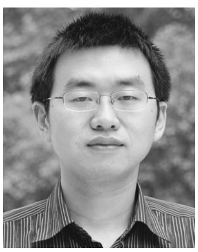
[41] L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr, "Associative hierarchical CRFs for object class image segmentation," in *Proc. Int. Conf. Comput. Vis.*, 2009, pp. 739–746.

[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, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "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.

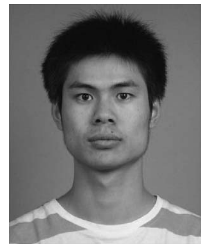
[45] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27, Apr. 2011.
[46] J. Platt, “Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods,” *Adv. Large Margin Classifiers*, vol. 10, no. 3, pp. 61–74, 1999.
[47] Y. Y. Boykov and M. P. Jolly, “Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images,” in *Proc. Int. Conf. Comput. Vis.*, 2001, pp. 105–112.
[48] 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.
[49] L. Shapiro and G. Stockman, *Computer Vision*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.
[50] K. Wu, E. Otoo, and K. Suzuki, “Optimizing two-pass connected-component labeling algorithms,” *Pattern Anal. Appl.*, vol. 12, no. 2, pp. 117–135, Jun. 2009.



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.

Since 2007, he has been with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, where he is currently a Professor. He has authored/coauthored over 60 research papers, including more than 25 peer-reviewed articles in international journals such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING; the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS; and *Pattern Recognition*. His research interests include multi- and hyperspectral remote sensing data processing, high-resolution image processing and scene analysis, and computational intelligence.

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



Ji Zhao received the B.S. degree in surveying from 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.

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 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 with the 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 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 has authored/coauthored over 300 research papers and is a 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 and an Executive Member of the China Society of Image and Graphics. He regularly serves as a Cochair for the series SPIE Conferences on Multispectral Image Processing and Pattern Recognition, Conference on Asia Remote Sensing, and many other conferences. He is an Executive Member (Board of Governor) of the Chinese National Committee for the International Geosphere-Biosphere Programme. He also serves as an Associate Editor of *International Journal of Ambient Computing and Intelligence*, *International Journal of Image and Graphics*, *International Journal of Digital Multimedia Broadcasting*, *Geo-spatial Information Science*, *Journal of Remote Sensing*, and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He edits several conference proceedings, issues, and geoinformatics symposiums.