High-Resolution Remote Sensing Image Classification Using Associative Hierarchical CRF Considering Segmentation Quality

Yun Yang, Alfred Stein, Valentyn. A. Tolpekin, and Yang Zhang

***Abstract*— This letter proposes an associative hierarchical conditional random field (AHCRF) model to improve the classification accuracy of high-resolution remote sensing images. It considers segmentation quality of superpixels, avoids a time-consuming selection of optimal scale parameters, and alleviates the problem of classification accuracy sensitive to undersegmentation errors that is present in traditional object-oriented classification methods. The model is built on a graph hierarchy, including the pixel layer as a base layer and multiple superpixel layers derived from a mean shift presegmentation. It extracts clustered features of pixels for superpixels at each layer and then defines the potentials of the AHCRF model. We suggest a weighted version of the interlayer potential using the size of a superpixel as a measure to reflect segmentation quality. In this way, erroneously labeled pixels of a superpixel are penalized. Experiments are presented using a part of the downsampled Vaihingen data from the ISPRS benchmark data set. Results confirm that our model shows more than 80% overall classification accuracy and is superior to the original AHCRF model and comparable to other models. It also alleviates the choosing of suitable segmentation parameters.**

***Index Terms*— Hierarchical systems, higher order statistics, image classification, image segmentation, remote sensing.**

I. Introduction

LAND cover or land use classification from high spatial resolution (HSR) remote sensing imagery acquired from Quickbird, Worldview satellites and so on has been addressed in the past. A variety of methods have been proposed to classify reliably and accurately. Common examples are object-oriented methods as in [1] combining a statistical classifier with multiscale image analysis, deep learning [2], and the hierarchical probabilistic topic model [3]. Their characteristic is their capability to model multiscale object features and semantic information, thus reducing classification errors.

Compared with these context-based models, conditional random field (CRF) [4] models have unique advantages in modeling a complex spatial context for image labeling. They have been extensively applied to object detection for land cover/use classification using remote sensing images from various sources [5]–[7]. For HSR imagery, those context-based models could be adopted for object detection and recognition.

Previous research [5]–[7] mainly focused on the second-order CRF model, with a limited ability in dealing with the spatial context. To compensate for this, hierarchical CRF models have been proposed to deal with longer range spatial dependence of variables for image labeling as in [8]–[12]. Here, we mention the robust P^N higher order potential [13], the harmony potential [14], the co-occurrence potential [15], and the patch-matching potential [16]. Part of them have been extended to object extraction and classification mainly from remote sensing imagery with HSR as in [17]–[19], showing a potential application of hierarchical CRF in remote sensing [20], whereas the associative hierarchical CRF (AHCRF) model [21], [22], as an extension of the robust P^N potential and associative Markov networks, is of interest. It is able to model object features from multiple spatial scales and to express the interlayer relationship of objects. In addition, it is flexible as compared with those non-CRF models and it is more straightforward to obtain a solution.

Although Senthilnath *et al.* [23] suggested a novel hierarchical clustering technique capable of automatically determining the optimal value of the number of clusters, our goal is to seek a solution that can correct the class labeling errors from incorrect clustering or segmentation caused by improper parameters. The innovation is to improve the classification accuracy by defining an interlayer higher order potential with respect to segmentation quality of superpixels.

II. Materials and Methods

A. Basic Structure of the AHCRF Framework

Ladický *et al.* [22] presented the AHCRF model as a general framework that can be described as a Gibbs energy $E(\mathbf{x})$

$$E(\mathbf{x}) = \sum_{i \in V} \psi_i(x_i) + \sum_{i,j \in V, j \in N_i} \psi_{ij}(x_i, x_j) + \min_{\mathbf{x}^{(1)}} E^{(1)}(\mathbf{x}, \mathbf{x}^{(1)}). \quad (1)$$

Manuscript received September 7, 2016; revised March 14, 2017, July 8, 2017, and November 6, 2017; accepted January 8, 2018. Date of publication March 14, 2018; date of current version April 20, 2018. This work was supported in part by the Project of the National Natural Science Fund under Grant 41301386, in part by the Open Project of the Key Laboratory of Degraded and Unused Land Consolidation Engineering, Ministry of Land and Resources of China, under Grant SXDJ2017-3, and in part by the Project of the National Natural Science Fund under Grant 41501498. *(Corresponding author: Yun Yang.)*

Y. Yang is with the College of Geology Engineering and Geomatics, Chang'an University, Xi'an 710054, China, and also with the Key Laboratory of Degraded and Unused Land Consolidation Engineering, Ministry of Land and Resources, Xi'an 710016, China (e-mail: yangyunbox@163.com).

A. Stein and V. A. Tolpekin are with the Faculty of Geo-Information Science and Earth Observation, University of Twente, 7514 AE Enschede, The Netherlands (e-mail: a.stein@utwente.nl; v.a.tolpekin@utwente.nl).

Y. Zhang is with the Land Engineering Construction Group, Xi'an 710075, China (e-mail: yangzh@163.com).

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

Digital Object Identifier 10.1109/LGRS.2018.2804345

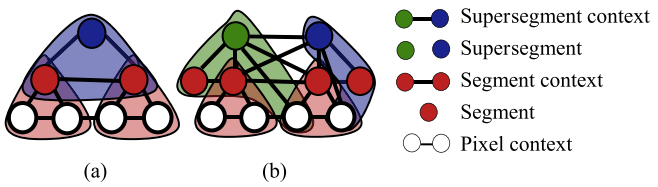

Fig. 1. Diagram of the AHCRF model. (a) Nested segmentation. (b) Nonnested segmentation.

In (1), the set V contains all pixels, with each pixel representing a node of a probability graph model in an image. The index $i \in V$ indicates a pixel site and j indicates a pixel site in a neighborhood N_i, $N_i \subset V$ of i. The symbols x_i and x_j denote the label variable at pixel sites i and j, with $x_i \in L$ where L is a label set. The notation $\mathbf{x}$ refers to the pixel label state of the base layer, and $\mathbf{x}^{(1)}$ is the label state of the first superpixel layer in the image. The term $\min_{\mathbf{x}^{(1)}} E^{(1)}(\mathbf{x}, \mathbf{x}^{(1)})$ indicates minimization of the Gibbs energy $E^{(1)}(\mathbf{x}, \mathbf{x}^{(1)})$ in order to reach the optimal $\mathbf{x}^{(1)}$, where $E^{(1)}(\mathbf{x},\mathbf{x}^{(1)})$ is an interlayer higher order potential between the base layer and the first superpixel layer (here $\mathbf{x}^{(0)} = \mathbf{x}$). It can be defined as a sum of a pairwise term related to variables of two adjacent superpixels c and d in the nth layer, an interlayer term related to superpixel c in the nth layer and its children in the $(n-1)$th layer, and a recursive term $\min_{\mathbf{x}^{(n+1)}} E^{(n+1)}(\mathbf{x}^{(n)}, \mathbf{x}^{(n+1)})$ related to variables from superpixel c in the nth layer and its parent in the $(n+1)$th layer. For more details, see [22].

This general framework can realize an association at different object levels, e.g., at the pixel, segment, supersegment, and scene levels. It is consistent with human cognition and image understanding for land cover or land use classification from HSR remote sensing imagery. To further develop object-oriented classification technology of an HSR image, we propose an AHCRF model that considers superpixel segmentation quality on the basis of [21] and [22]. It aims to improve classification suffering from undersegmented superpixel errors caused by traditional object-based methods.

B. Hierarchical Graph Construction

To deal with the large variety of geo-objects with different spatial scales, a remote sensing image is represented hierarchically by means of multiscale segmentation, producing a set of superpixel (i.e., segment or supersegment). Superpixels in different layers are either nonnested (here means intersecting) or nested.

We adopted nonnested segmentation to yield multiple superpixel layers as it is favorable to label consistency of segments [24]. Thus, a given image is represented with a pixel layer as the base layer and multiple superpixel layers as the first layer, the second layer, and so on, leading to a hierarchical graph [see Fig. 1(b)].

For feature extraction from the pixel layer, spectral and texture features were included, using the local binary pattern (LBP) and Texton operators. Clustering of the feature vectors was performed in order to quantify dense features and then to form sparse clustered features, thus defining unary potentials over segments and supersegments. For the superpixel layers, statistical features were derived from those.

C. Definition of the AHCRF Model

According to the AHCRF framework [22], the interlayer higher order term $\psi_c^p(\mathbf{x}_c^{(n-1)}, x_c^{(n)})$ can be decomposed into an unary potential with respect to superpixel c in the nth layer and an interlayer potential with respect to superpixel c in the nth layer and its children in the $(n-1)$th layer. We now continue with redefinition of the higher order term.

The unary potential function $\varphi_c(x_c^{(n)})$ is defined as

$$\varphi_c\left(x_c^{(n)}\right) = \begin{cases} \gamma_c^{\max} & \text{if } x_c^{(n)} \in l_F \\ \gamma_c^{l} & \text{if } x_c^{(n)} \in L \end{cases} \tag{2}$$

where $x_c^{(n)}$ is the label variable of superpixel c of the nth layer that takes a value from the extended label set $L^E = L \cup \{l_F\}$, where l_F is a free label different from a given label $l (l \in L)$. The function associates the cost γ_c^l with $x_c^{(n)}$ taking a label l in L and the maximum cost $\gamma_c^{\max}$ with $x_c^{(n)}$ taking the free label l_F. Therefore, the superpixel c in the nth layer has no dominant label and satisfies $\gamma_c^l \leq \gamma_c^{\max}$. Usually γ_c^l is defined as a probability function of classifiers like a decision tree [25], a random forest (RF) [26], or a support vector machine (SVM) [27], [28]. In our model, we selected RF instead of SVM because of its excellent classification performance [29] for HSR imagery.

We further propose a weight function of RF probability using segmentation quality based upon the size of the superpixel. Thus, γ_c^l is defined as a probability weighed by the segmentation quality of each superpixel

$$\gamma_c^l = G(c) \min(-\log P(x_c = l|\mathbf{y}), \alpha) \tag{3}$$

where $P(x_c = l|\mathbf{y})$ is the posterior probability of x_c having the label l, given an observed variable $\mathbf{y}$. The weight $G(c)$ as a measure of superpixel segmentation quality is defined as

$$G(c) = (|c|)^n \tag{4}$$

where the size of the superpixel is taken as a measure of the segmentation quality. With this choice, a superpixel becomes more heterogeneous and the segmentation quality deteriorates if a superpixel contains more pixels. This will cause a larger value of the unary potential, that is, a larger penalty will be imposed upon superpixel c with label $l \in L$. This quality measure is simple but effective. Although other measures to evaluate superpixel segmentation quality like the variance [13] can be adopted, they are less effective for HSR imagery in this letter. Our model performed well for $n = 2$ in the following experiments. In (2), the maximal cost $\gamma_c^{\max}$ was defined as

$$\gamma_c^{\max} = G(c)\alpha \tag{5}$$

where α is a parameter that we set as $\alpha = -\log 0.1$. Clearly, $\gamma_c^{\max}$ varies as the value of $G(c)$ changes.

The interlayer potential function $\varphi_c(\mathbf{x}_c^{(n-1)}, x_c^{(n)})$ with respect to superpixel c of the nth layer and its children of the $(n-1)$th layer can be formulated as a piecewise function referred to in [22], in which the cost k_c^l, as a penalization of each inconsistent superpixel c for arbitrary label l, is closely related to the definition of $G(c)$ in (4), that of γ_c^l as in (3), and that of $\gamma_c^{\max}$ as in (5).

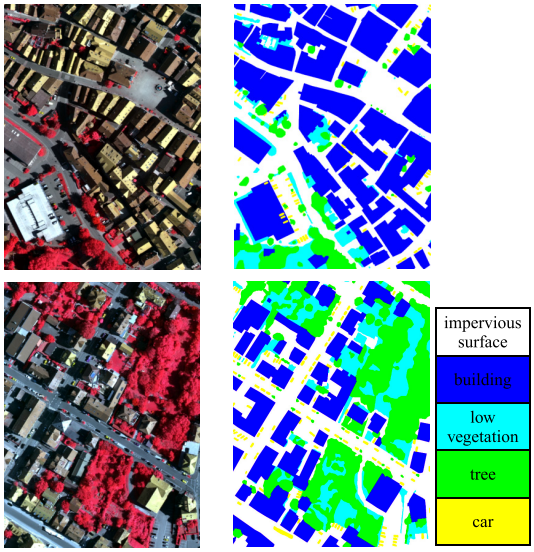

Fig. 2. Two test images (the first column) and their ground truth (the second column).

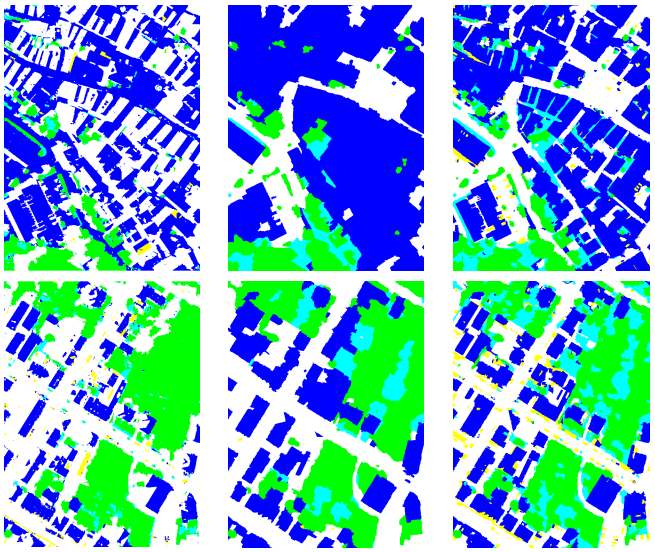

Fig. 3. Labeled results of the first test image (top row) and the second test image (bottom row) using the superpixel-based non-CRF method (the first column), the original AHCRF model (the second column), and our model (the third column).

The proposed definition of segmentation quality is critical to the unary potential function $\varphi_c(x_c^{(n)})$ and the interlayer potential function $\varphi_c(\mathbf{x}_c^{(n-1)}, x_c^{(n)})$. Also it is favorable to penalize inhomogeneous superpixels that are valued at an arbitrary label $l \in L$. Finally, the unary potential function $\psi_i(x_i)$ is defined over the pixel i using a probability RF classifier, whereas the pairwise potential function $\psi_{ij}(x_i, x_j)$ is defined over adjacent pixels i, j and the pairwise function $\psi_{c,d}(x_c^{(n)}, x_d^{(n)})$ over the adjacent superpixels c and d in the nth layer [22].

The redefined AHCRF model was implemented using our improved version of Automatic Labeling Environment (ALE) code[1] for HSR image classification.

III. Experiments and Analysis

A. Experiments

We tested the procedures on the airborne image data set of Vaihingen, Germany, for 2-D semantic labeling, published by ISPRS WG III/4. The area covers an urban scene of a small village with many detached buildings and small multistory buildings. The data set consists of very high resolution orthophotos with corresponding digital surface models (DSMs). The original images were of a 9-cm resolution with the near infrared, red and green bands. In the experiment, a subset of the original data set, including the former four images, was downsampled to a 0.3-m resolution in order to reduce the computation time. Each image of the downsampled subset has more than 444 000 pixels. The five classes were defined as impervious surfaces, building, low vegetation, tree, and car. An RGB show of two test images and ground truth are in Fig. 2.

First, a multiscale segmentation using the mean shift algorithm [30] was performed to yield superpixel layers. In this way, an automatic implementation of our method was obtained by setting the segmentation parameters' size of spatial window (*hs*), size of color window (*hr*), and minimum region (*MinR*). We used parameter values equal to (3.5, 5.5, 10), (5.5, 3.5, 10), and (5.5, 5.5, 10) for the three superpixel layers of the test images, respectively. Here, the first value refers to the planar distance between points, the second to the Euclidian distance in the LUV color space, and the third denotes a minimum of the number of pixels among all superpixels in image segmentation.

Second, for extracting features, we realized that spectral features include three spectral channels. Texture features derived from the LBP operator with an 11-size window and the Texton operator with a scale value equal to 0.7 were incorporated. These dense pixel features were clustered into 50 clusters using a standard K-means algorithm, where the values of the parameters were chosen using trial and error, taking the time-cost and classification accuracy into consideration.

Four images from Vaihingen data set with 0.3-m resolution were randomly divided into two training images and two test images. All the pixels from the two training images as training samples were used to train the parameters of our model. All the pixels from the two test images were labeled using our model. Labeled results of both test images with our model and two others (i.e., M3 and M6 in the following) are shown in Fig. 3.

B. Evaluation and Analysis

Six representative models were chosen to compare with the proposed AHCRF model (see Table I). These models are M1: the pixel-based RF model, M2: the pixel-based RF-CRF model, M3: a superpixel-based non-CRF method, M4: the superpixel-based CRF model implemented using STAIR Library, M5: the robust P^N HoCRF model [13], and M6: the original AHCRF model. The last two are hierarchical CRF models closely related to our model (called as M7).

All models except M3 and M4 were implemented under the same condition. They were tested using the same training samples and evaluated using ground truth data. For M3,

[1] http://www.inf.ethz.ch/personal/ladickyl/

TABLE I

COMPARISONS OF CLASSIFICATION PERFORMANCE AMONG THE SEVEN MODELS

Models	M1	M2	M3	M4	M5	M6	M7
OA/%	60.67	62.28	63.41	69.70	76.17	77.88	81.59
AA/%	43.95	43.23	49.45	50.62	72.27	58.38	72.82
Time/s	392.1	419.7	280.2	250.5	459.3	475.7	549.3

Here, OA is the overall accuracy $\sum_{i\in L} N_{ii} / \sum_{i,j\in L} N_{ij}$, AA is the average accuracy $\sum_{i\in L}(N_{ii} / |L| \sum_{j\in L} N_{ij})$, N_{ii} is the number of pixels of label i labeled i, N_{ij} is the number of pixels of label i labeled j, and $|L|$ is the total of the labels, defined as [21]. Time is a total of each steps from an input of images to the output of classification results.

TABLE II

COMPARISON OF RECALL OF INDIVIDUAL CLASSES AMONG THE SEVEN MODELS

Class	impervious surface	building	low vegetation	tree	car
M1	53.3	71.0	9.5	85.8	0.1
M2	44.8	87.8	12.4	71.1	0.0
M3	73.3	59.5	13.8	85.9	14.7
M4	58.0	85.4	10.2	92.8	6.7
M5	76.5	90.0	60.1	88.4	46.5
M6	68.2	92.7	38.0	92.3	0.7
M7	78.7	87.1	61.2	89.2	47.9

the edge-based watershed segmentation algorithm [31] at scale 50 with the iteratively merging of adjacent segments based on a full λ-schedule algorithm with level 50 followed by maximum likelihood classification based on incorporation of spectral, spatial, and texture features was carried out. Trials at five different but representative segmentation scales were done. The optimal classification accuracy is listed in Table I. For M4, a mean shift algorithm with parameters hs $= 4$, hr $= 6$, and MinR $= 10$ yielded the superpixels.

Fig. 3 and also Table I show that our proposed model is better suited to correctly discriminate details of objects like cars from their background, e.g., roads, as compared with the original AHCRF model. This is because our model can partly alleviate incorrectly classified pixels caused by undersegmentation error of objects. In addition, DSM data were incorporated into the model that showed an increase of the overall accuracy to 82.96%.

Table I shows that the AHCRF models are superior to traditional pixel-based CRF and a two-layer higher order CRF model. More importantly, our proposed AHCRF model with varying weights considering segmentation quality of superpixels performed better than the original AHCRF model, showing a 3.71% improvement in overall accuracy. Our model, however, has the highest time-cost due to more time-consuming image segmentation steps.

Considering the similarity and relevance of our model to some popular classification methods, the M3–M6 methods/models were compared with our model. The comparison of recall [21] (obtained as $N_{ii}/\Sigma_{j\in L} N_{ij}$) for individual class is presented in Table II (in percentage terms).

From Table II, we observe that the seven models performed best for larger objects like buildings and trees, but they were less efficient for small objects like cars leading to lower overall accuracy and average accuracy, especially for M1 and M2. The explanation mainly is that less training samples were chosen for this class by using automatic sample selection. Further experiments showed that both the individual accuracy of the class car and the average accuracy would greatly improve if more training samples containing the class car were chosen, making the model more time-consuming. Additionally, cars are small objects and they are easily undersegmented and merged into roads or buildings. However, Table II also shows that individual efficiency of the class car improved using M5 and M7. Furthermore, our proposed model shows a better improvement in the recall of the class car than M5 and others.

TABLE III

COMPARISONS OF OUR MODEL WITH DIFFERENT SUPERPIXEL LAYERS

Number of layers	1	2	3	4
OA/%	62.28	80.34	81.59	80.70
AA/%	43.23	65.38	72.82	69.87
Time/s	454.7	542.0	550.3	694.8

TABLE IV

COMPARISONS AT DIFFERENT SEGMENTATION PARAMETERS FOR OUR MODEL

Segmentation parameters	I	II	III	IV
OA/%	81.59	77.90	75.28	80.71
AA/%	72.82	67.46	60.74	63.84
Time/s	550.3	441.9	505.9	866.5

Here, the parameters (*hs, hr, MinR*) are set as follows: I: (3.5, 5.5, 10), (5.5, 3.5, 10), (5.5, 5.5, 10); II: (3.5, 5.5, 10), (3.5, 5.5, 20), (3.5, 5.5, 30); III: (3.5, 5.5, 10), (3.5, 7.5, 10), (3.5, 11.5, 10), IV: (7.5, 5.5, 10), (5.5, 7.5, 10), (7.5, 7.5, 10)

To analyze the effect of the number of layers on classification results, Table III compares the classification accuracy and time-cost at different cases with one to four superpixel layers in our model.

Table III shows that the highest accuracy occurs for the proposed AHCRF model with three superpixel layers. Clearly, use of more superpixel layers does not lead to a higher classification accuracy. The time-cost also increases with the number of superpixel layers.

To analyze the effect of segmentation parameters on classification, Table IV compares the classification accuracy and time-cost using our model by setting four different groups of segmentation parameters of the mean shift algorithm according to the nested relationship between current superpixel and its children in a multiscale segmentation. As stated in [21], nonnested segmentation is better suited to enhance the label consistency of those pixels actually belonging to the same category in a superpixel. Those pixels, therefore, have a higher probability to be labeled as the same class.

Table IV leads to the following conclusions. The AHCRF model is superior in classification to the CRF model without higher order potentials and to the traditional pixel-based CRF model under the same test conditions. This is irrespective of the group of parameters used for segmentation. Classification accuracy, however, is different. Usually, a higher accuracy is achieved if the hierarchy graph is built upon superpixels

in a nonnested segmentation way like the use of parameter group I or IV instead of a nested segmentation way like parameter group II or III (see Table IV). Otherwise, the classification accuracy decreases mainly because of less overlapping pixels between superpixels using parameter group II or III in a nested segmentation way.

Additionally, we find that the time-cost for image segmentation with parameter sets I and IV is usually larger than that with parameter sets II and III (see Table IV). For this reason, an implementation of superpixel segmentation in a nested segmentation way is only possible with a region growing segmentation based upon its children.

IV. Conclusion

In this letter, we presented a specific AHCRF model with an innovative interlayer higher order potential based upon the size of superpixels under the original AHCRF framework. Its aim was to improve land cover classification from high-resolution remote sensing imagery. Experiments showed that our model improves the classification accuracy as compared with the original AHCRF model. Its main merit is that it addresses the undersegmentation error, which usually heavily reduces classification accuracy, as compared with traditional object-oriented classification methods for remote sensing imagery with an HSR. Thus, users are no longer required to put much care on choosing the parameters for image segmentation. In addition, our AHCRF model has a higher classification accuracy if the nonnested image segmentation way is adopted. The reason is that overlapping superpixels at different layers strengthens the label consistency of neighboring pixels.

Our AHCRF model with the proposed segmentation quality measure did not completely avoid the classification error caused by undersegmentation of superpixels. Further improving these aspects still needs to be done.

Acknowledgment

The authors would like to thank the anonymous reviewers and the editors for their constructive feedback.

References

[1] Z. Lv, H. He, J. A. Benediktsson, and H. Huang, "A generalized image scene decomposition-based system for supervised classification of very high resolution remote sensing imagery," *Remote Sens.*, vol. 8, no. 10, pp. 814–831, 2016.

[2] M. Längkvist, A. Kiselev, M. Alirezaie, and A. Loutfi, "Classification and segmentation of satellite orthoimagery using convolutional neural networks," *Remote Sens.*, vol. 8, no. 4, pp. 329–349, 2016.

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

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

[5] T. Hoberg, F. Rottensteiner, R. Q. Feitosa, and C. Heipke, "Conditional random fields for multitemporal and multiscale classification of optical satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 2, pp. 659–673, Feb. 2015.

[6] J. Zhao, Y. Zhong, H. Shu, and L. Zhang, "High-resolution image classification integrating spectral-spatial-location cues by conditional random fields," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4033–4045, Sep. 2016.

[7] J. D. Wegne, U. Soergel, and B. Rosenhahn, "Segment-based building detection with conditional random fields," in *Proc. JURSE*, Apr. 2011, pp. 205–208.

[8] M. Y. Yang and W. Förstner, "A hierarchical conditional random field model for labeling and classifying images of man-made scenes," in *Proc. IEEE/ISPRS Workshop Comput. Vis. Remote Sens. Environ.*, Nov. 2011, pp. 196–203.

[9] L.-L. Wang and N. H. C. Yung, "Improved hierarchical conditional random field model for object segmentation," *Mach. Vis. Appl.*, vol. 26, nos. 7–8, pp. 1027–1043, 2015.

[10] Z. Zhang, M. Y. Yang, and M. Zhou, "Multi-source multi-scale hierarchical conditional random field model for remote sensing image classification," *ISPRS Ann. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 2, no. 3, pp. 293–300, 2015.

[11] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2587–2600, Sep. 2006.

[12] X. Su, C. He, Q. Feng, X. Deng, and H. Sun, "A supervised classification method based on conditional random fields with multiscale region connection calculus model for SAR image," *IEEE Trans. Geosci. Remote Sens.*, vol. 8, no. 3, pp. 497–501, May 2011.

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

[14] J. M. Gonfaus, X. Boix, J. van de Weijer, A. D. Bagdanov, J. Serrat, and J. Gonzàlez, "Harmony potentials for joint classification and segmentation," in *Proc. CVPR*, Jun. 2010, pp. 3280–3287.

[15] L. Ladický, C. Russell, P. Kohli, and P. H. S. Torr, "Graph cut based inference with co-occurrence statistics," in *Proc. ECCV*, 2010, pp. 239–253.

[16] S. Gould, "Multiclass pixel labeling with non-local matching constraints," in *Proc. CVPR*, Jun. 2012, pp. 2783–2790.

[17] 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.

[18] J. A. Montoya-Zegarra, J. D. Wegner, L. Ladický, and K. Schindler, "Semantic segmentation of aerial images in urban areas with class-specific higher-order cliques," *ISPRS Ann. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 2, no. 3, pp. 127–133, 2015.

[19] E. Li, J. Femiani, S. Xu, X. Zhang, and P. Wonka, "Robust rooftop extraction from visible band images using higher order CRF," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4483–4495, Aug. 2015.

[20] J. Chen *et al.*, "Information from imagery: ISPRS scientific vision and research agenda," *ISPRS J. Photogram. Remote Sens.*, vol. 115, pp. 3–21, May 2016.

[21] L. Ladický, C. Russell, P. Kohli, and P. H. S. Torr, "Associative hierarchical CRFs for object class image segmentation," in *Proc. ICCV*, Sep./Oct. 2009, pp. 739–746.

[22] L. Ladický, C. Russell, P. Kohli, and P. H. S. Torr, "Associative hierarchical random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1056–1077, Jun. 2014.

[23] J. Senthilnath, D. Kumar, J. A. Benediktsson, and X. Zhang, "A novel hierarchical clustering technique based on splitting and merging," *Int. J. Image Data Fusion*, vol. 7, no. 1, pp. 19–41, 2016.

[24] C. Pantofaru, C. Schmid, and M. Hebert, "Object recognition by integrating multiple image segmentations," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 481–494.

[25] J. R. Quinlan, "Simplifying decision trees," *Int. J. Hum.-Comput. Studies*, vol. 51, no. 2, pp. 497–510, 1999.

[26] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[27] 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.

[28] Z. He, Y. Shen, M. Zhang, Q. Wang, Y. Wang, and R. Yu, "Spectral-spatial hyperspectral image classification via SVM and superpixel segmentation," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2014, pp. 422–427.

[29] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogram. Remote Sens.*, vol. 114, pp. 24–31, Apr. 2016.

[30] 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.

[31] X. Y. Jin, "Segmentation-based image processing system," U.S. Patent 8 260 048, Sep. 4, 2012.