applied optics

Context guided belief propagation for remote sensing image classification

Tiancan Mei,[1,*,†] Le An,[2,†] and Bir Bhanu[3]

[1]School of Electronic Information Wuhan University, Wuhan 430072, China
[2]BRIC, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA
[3]Center for Research in Intelligent Systems, University of California, Riverside, California 92521, USA
*Corresponding author: mtc@whu.edu.cn

Received 15 December 2014; revised 28 February 2015; accepted 12 March 2015; posted 12 March 2015 (Doc. ID 229361); published 9 April 2015

We propose a context guided belief propagation (BP) algorithm to perform high spatial resolution multispectral imagery (HSRMI) classification efficiently utilizing superpixel representation. One important characteristic of HSRMI is that different land cover objects possess a similar spectral property. This property is exploited to speed up the standard BP (SBP) in the classification process. Specifically, we leverage this property of HSRMI as context information to guide messages passing in SBP. Furthermore, the spectral and structural features extracted at the superpixel level are fed into a Markov random field framework to address the challenge of low interclass variation in HSRMI classification by minimizing the discrete energy through context guided BP (CBP). Experiments show that the proposed CBP is significantly faster than the SBP while retaining similar performance as compared with SBP. Compared to the baseline methods, higher classification accuracy is achieved by the proposed CBP when the context information is used with both spectral and structural features.

OCIS codes: (100.2960) Image analysis; (100.3008) Image recognition, algorithms and filters; (100.4993) Pattern recognition, Baysian processors.

http://dx.doi.org/10.1364/AO.54.003372

1. INTRODUCTION

High spatial resolution multispectral imagery (HSRMI) is widely used in remote sensing applications due to the abundant information it contains and the easy access to the data with low cost. HSRMI classification performance is of crucial importance for HSRMI applications, such as land cover mapping and disaster relief.

Traditionally, image classification is executed at the pixel level. The pixel-based classification methods assume high interclass variation and low intraclass variation among the pixels. However, this assumption only holds for coarse or medium spatial resolution multispectral imagery (MSRMI). As only the spectral feature of a pixel is used to represent MSRMI, most of the previous work focuses on effective classifiers, such as maximum likelihood (ML) [1], support vector machine (SVM) [2], and minimum error classifier [3]. Since there are distinct spectral characteristics among land cover classes, methods at the pixel level may work well for MSRMI. However, the emergence of HSRMI has invalidated the assumption which is suitable for MSRMI. Hence, methods at the pixel level usually fail to perform well on HSRMI data.

Nowadays, it is widely accepted that spatial information and interaction between neighboring pixels will help to address the challenge of HSRMI classification [4]. To discriminate similar land cover classes, many feature descriptors that consider spatial information have been proposed, such as texture, shape, and structural feature [5–8]. Aiming at modeling the image data and inferring the class label of each pixel, classification frameworks have been well studied [9–11]. The graphical-model-based framework is a probabilistic model that represents the structure of joint distribution in a compact way [12]. As a special case of the undirected graphical model, Markov random fields (MRF) has been widely used for image classification. At the early stage, MRF was defined at the pixel level [13]. Since the neighborhood relationship among pixels is regular on the 2D lattice, the pixel-based MRF methods could conveniently model the spatial context information to classify an image [14]. The limitation of pixel-wise MRF is that it usually discards the spatial information. Inspired by the observation that pixels with similar properties often share the same class label, the MRF is extended to operate at superpixel level [15,16]. The advantage of these methods is that higher order features based on all pixels composing a superpixel can be computed and used for classification [17]. The efficient inference algorithm is a critical part of MRF-based classification framework to make the graphical model computation tractable. Belief propagation (BP) [18] is one of the most popular inference algorithms used in the MRF framework. One limitation of the standard BP is

that its complexity is a quadratic function of the number of classes. Despite substantial adoption and adaptation of BP in the remote sensing field, how to incorporate the characteristic of HSRMI in the BP to reduce the computational complexity has rarely been investigated.

In this paper, we learn the characteristic of HSRMI automatically and use it to reduce the complexity as compared with standard BP for HSRMI classification. Specifically, we utilize the context information in HSRMI to guide message propagation. BP performs inference on MRF models by passing local messages to neighboring nodes. For *standard BP* (SBP), all possible labels which one node may take are considered when computing the messages that it passes to the neighboring nodes. In fact, only a part of the possible labels can provide meaningful information to its neighboring nodes. As a result, passing messages with meaningful labels will speed up the BP algorithm by reducing the number of labels involved in messages passing, while retaining the functionality of SBP. This messages passing scheme with selected class labels is referred to as *context guided BP* (CBP). To tackle the challenge of low interclass variation of HSRMI, a sequential classification scheme is proposed. In the *first* step, SVM is used to classify the input image using spectral information only. The output of SVM and unsupervised segmentation results are used to reduce the number of labels in messages passing for each node. In the *second* step, both of the spectral and structural features are fed into CBP to classify the input image.

The rest of this paper is organized as follows. The proposed CBP for MRF-based image classification is explained in detail in Section 2. The experimental results on remote sensing images are shown in Section 3. Finally, Section 4 concludes this paper.

2. METHODOLOGY

The proposed method consists of three main parts, as shown in Fig. 1. *First*, the superpixels and the initial classification results are obtained by unsupervised clustering and supervised classification, respectively. *Second*, the characteristic of HSRMI is learned from the initial segmentation and classification results by analyzing the co-occurrence of class labels in each superpixel. *Third*, the CBP is used to estimate the optimized solution of the MRF model. In the end the classification result is given by the optimized solution of the MRF model. For clarity and completeness of our method, this section first gives a brief introduction of the MRF-based image classification framework with the SBP explained in detail, before the proposed CBP is presented.

A. MRF Framework for Classification

Let S denote a set of lattice points, random field $X \triangleq \{X_s, s \in S\}$ is defined on S with a neighborhood system ϵ, each variable in X is associated with a site s and takes value from a discrete set $L = \{1, \ldots, M\}$, where M is the number of possible class labels, and $x \triangleq \{x_s, s \in S\}$ is a realization of X, which takes value from $\Omega = L^M$. Any possible label assignment to X is called a labeling. The neighborhood system ϵ is the set of edges connecting variables in the random field. Image data $y \triangleq \{y_s, s \in S\}$ is assumed to be a realization of the random field $Y \triangleq \{Y_s, s \in S\}$. The probability of assigning a specific label to X is referred to as $P(X = x)$, in which X is assumed to possess the Markovian property. Given an observed image y, the posterior distribution over the label field follows Gibbs distribution and can be written in the following form, according to Hammersley–Clifford theorem [19]:

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

where Z is a normalizing constant and $\psi_c(X_c)$ is a potential function defined over the variables $X_c = \{x_s, s \in c\}$ that constitute cliques c. A clique c is a set of variables X_c which are conditionally dependent on each other. C is the set of all cliques. To simplify, we drop the notation of random X and Y, then the posterior energy is simplified as

$$E(x|y) = -\log P(x|y) - \log Z = \sum_{c\in C}\psi_c(X_c). \quad (2)$$

The labeling that minimizes the posterior energy corresponds to maximum *a posteriori* (MAP) estimation of the labeling given by

$$x^* = \arg\min_{x\in\Omega} E(x|y). \quad (3)$$

The potential functions commonly used for image classification are defined on unary clique and pairwise clique. As a result, the energy function can be written as

$$E(x|y) = \sum_{i\in S}\phi(x_i) + \sum_{(i,j)\in\epsilon}\phi(x_i, x_j). \quad (4)$$

The unary potential $\phi(x_i)$ is commonly defined as the negative log likelihood of a label being assigned to site i, and it represents the contribution of observed data to the posterior energy. The pairwise potential $\phi(x_i, x_j)$ often takes the form of a Potts model [17] defined as

$$\phi(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j, \\ \beta & \text{otherwise,} \end{cases} \quad (5)$$

which imposes the smooth constraint on the posterior energy. β is the parameter used to adjust the penalization when neighboring sites take different labels. For pixel-wise MRF, each element in S denotes a pixel and ϵ denotes all the edges that connect pixel i and j. For superpixel-wise MRF, each element s in S represents a superpixel, and ϵ is a set of all the edges that connect superpixel i and j.

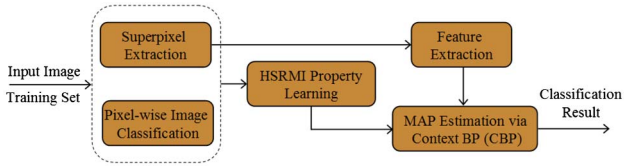


Fig. 1. Framework of the proposed method.

B. Standard BP

BP is one of the most widely used algorithms to conduct inference on MRF graph for producing highly accurate results in practice. Although exact inference can be done only for acyclic graph, BP has been successfully applied to loopy graph in different domains [20]. SBP solves the minimization problem by iteratively passing messages between neighboring nodes in

the graph. In each iteration, each node sends a message to its neighboring nodes, while receiving messages from neighboring nodes simultaneously. The iteration terminates when all messages converge. Each message is a vector of dimension determined by the number of labels. Messages can be intuitively interpreted as how likely a node thinks its neighboring node takes a certain label. Let the message sent from node i to neighboring node j be denoted as m_{ij}, it is updated at each iteration in the following way:

$$m_{ij}(x_j) = \min_{x_i \in L} \left(\phi(x_i) + \phi(x_i, x_j) + \sum_{k \in N(i) \setminus j} m_{ki}(x_i) \right), \quad (6)$$

where $\phi(x_i)$ and $\phi(x_i, x_j)$ are unary and pairwise potential functions, respectively. $N(i) \setminus j$ is the set of all neighboring nodes of i except node j. The updating rule indicates that node i should traverse all possible labels to determine the message m_{ij}. Meanwhile, it should receive the messages from its neighbors except node j before passing messages to the neighboring node j. When messages passing converges, the belief of each node is given as

$$b(x_i) = \phi(x_i) + \sum_{k \in N(i)} m_{ki}(x_i). \quad (7)$$

The belief $b(x_i)$ is the approximated posterior probability of node i being assigned label x_i. The label x_i^* of node i is selected by the following minimization:

$$x_i^* = \arg\min_{x_i \in L} b(x_i). \quad (8)$$

The max-product (min-sum in negative log form) BP is used in this paper in order to obtain the MAP estimation. More details about BP can be found in [18].

C. Context Guided BP

In SBP, node i should traverse all possible labels to pass messages to its neighboring node j. In practice, only part of all the labels of node i provide useful information that is necessary to be sent to node j. For example, if we know that node i is more likely to take label *water* or *shadow* prior to passing messages to its neighbors, it may be unnecessary to traverse all labels except *water* and *shadow*. Based on this observation and motivated by previous work of Komodakis and Tziritas [21], we propose a method, referred to as CPB, to pass messages with select labels only by leveraging the context information in HSRMI to efficiently classify HSRMI data.

It is well known that land cover objects in HSRMI are highly related to each other and there are various kinds of interactions between them [22]. The low spectral variation across land cover classes is one example. In our method, we utilize this characteristic to prune the labels used in messages passing to speed up SBP. In SBP, irrelevant labels in messages passing will not only increase the computation complexity but may also introduce irrelevant information, thus hindering the convergence. The difference between our work and the method in [21] lies in the selection of labels for messages passing. The method in [21] chooses label x_i to pass messages if belief $b(x_i)$ exceeds a certain threshold. If the threshold of belief is set too high, almost all labels will be considered to pass messages. Otherwise, if it is set too low, labels that are relevant to this node may be excluded. Instead, we choose labels to pass messages by utilizing the context information of HSRMI, which is independent of belief that one node has.

The basic idea of CBP is that land cover objects sharing similar property can be grouped into a label subspace, from which we choose labels to pass messages, instead of using all possible labels. This can be explained by Fig. 2. The classification results using only spectral information are shown in Fig. 2(b). We can see a lot of misclassifications in the results due to different classes having similar spectral properties. For example, *water* and *shadow* as pointed out by white arrows in Fig. 2(a) possess similar spectral properties. As a result, a large number of *water* pixels are misclassified as *shadow* as shown in the corresponding area marked by the white boundary in Fig. 2(b). Further, we can observe that *water* and *shadow* coexist in one superpixel marked by the white boundary in the lower right part of Fig. 2(b).

Since each superpixel is a strong predictor of spectral consistency, object classes coexisting in a superpixel should have low spectral variation across them. As a result, if one of those coexisting object classes is observed in a superpixel, it strongly indicates that other coexisting objects may exist in the same superpixel. By analyzing object co-occurrence in all superpixels,

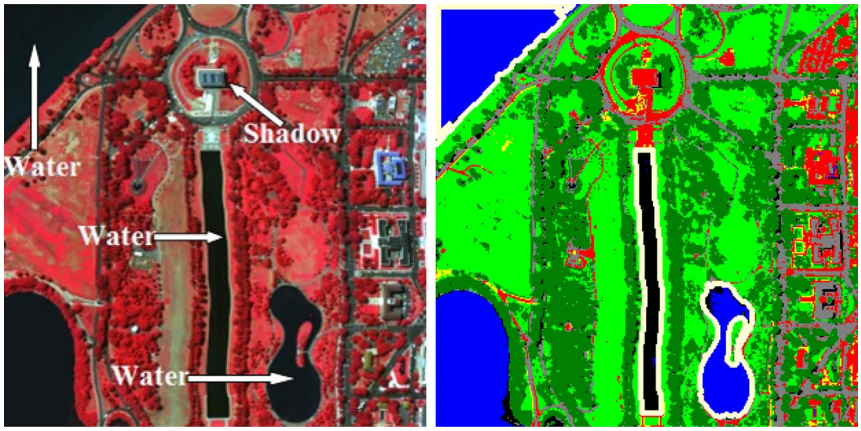


Fig. 2. Example of land cover with similar spectral properties being misclassified. (a) Original image. (b) Classified and segmented result. Note that the water part in the middle is misclassified as shadow.

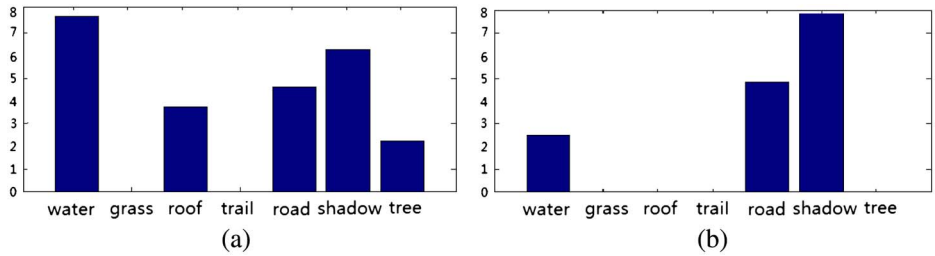


Fig. 3. Histogram of objects in a superpixel. (a) Majority of pixels are classified correctly. (b) Majority of pixels are misclassified. (The counts are in logarithmic scale).

we can find out what objects share similar properties. Thus, if a pixel is initially classified as label l_i, we should only consider those labels that share similar properties with l_i to pass messages to its neighbors. The label pruning principle also applies to nodes that receive messages from their neighbors. When the initial classification results are obtained and label subspace L_i for each label is determined by the property of HSRMI, the message passing method is modified as

$$m_{ij}(x_j) = \min_{x_i \in L_i} \left(\phi(x_i) + \phi(x_i, x_j) + \sum_{k \in N(i) \setminus jj} m_{ki}(x_i) \right). \quad \textbf{(9)}$$

In the same way, the label of a node i which minimizes the belief $b(x_i)$ is modified as

$$x_i^* = \arg\min_{x_i \in L_i} b(x_i). \quad \textbf{(10)}$$

Let M_i and M_j be the number of elements in label subspace L_i and L_j, respectively. Since M_i and M_j are smaller than M, the complexity of modified messages passing method reduces from $O(M^2)$ to $O(M_i \times M_j)$, compared to Eq. (6). L_i and L_j are determined by the initial classification results of node i and node j. In the following, we discuss how to get the label subspace for each label in L by using the context information of HSRMI.

After initial classification and segmentation, we can learn the context of HSRMI by analyzing the coexistence frequency of classes in each superpixel. For example, most of the pixels inside a superpixel pointed out by the white arrow on the top left in Fig. 2(a) are assigned correctly to class *water* in the initial classification. The classes histogram of this superpixel is shown in Fig. 3(a). As observed from the histogram, the first and the second largest number of pixels inside are assigned to *water* and *shadow*, respectively. This suggests that *water* and *shadow* share similar spectral characteristics inside a superpixel with consistent spectral properties. In the middle of Fig. 2(a), most of the pixels in the superpixel pointed out by the white arrow are misclassified as *shadow*. The classes histogram of this superpixel is shown in Fig. 3(b). While most pixels inside this superpixel are misclassified, we can learn from the histogram that *water* and *shadow* share similar spectral characteristics in the same way. Regardless of the classification accuracy in a superpixel, the frequency histogram of classes in that superpixel can provide information about what classes share similar spectral properties. By introducing a class frequency histogram, we can represent the context of HSRMI in the form of co-occurrence probability of different classes in all superpixels.

The flow chart to determine the label subspace in five major stages is shown in Fig. 4. In the first stage, the input image is segmented into superpixels using the mean-shift algorithm [23]. In the second stage, we classify the image using pixel spectral information with the help of training samples. In the third stage, the occurrence frequency of each object class in a superpixel is obtained by counting the number of pixels assigned to certain object classes. In the fourth stage, a co-occurrence table with M rows and M columns is constructed based on object class co-occurrence frequency in each superpixel. Finally, the

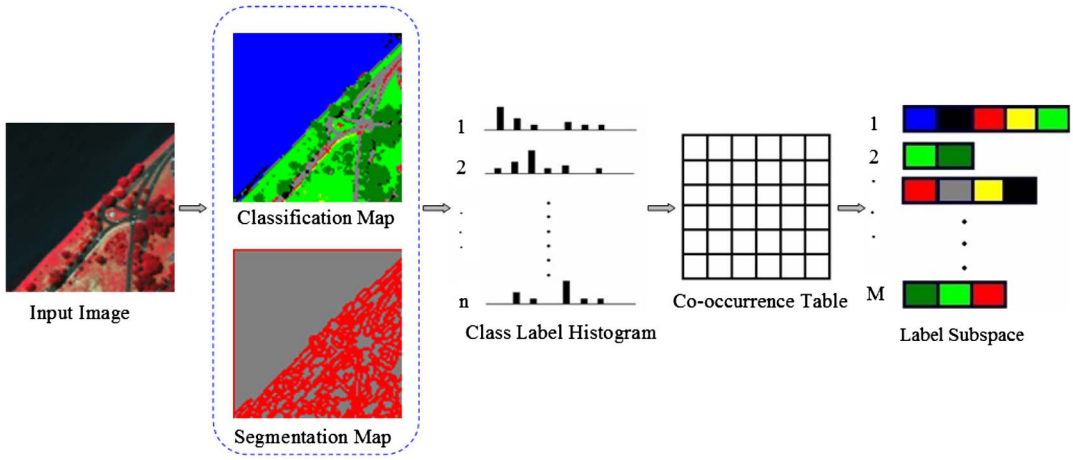


Fig. 4. Flow chart of using HSRMI property to determine label subspace for each label L_i.

jth label subspace is formed by choosing labels from the jth row of normalized co-occurrence table whose co-occurrence probability exceeds a certain threshold. The number of label subspaces is equal to the number of land cover classes. The pseudocode that constructs the label co-occurrence table is given in Algorithm 1. $T(L_i, L_j)$ represents the frequency that label (L_i, L_j) coexist in all superpixels. L_1 is the label that the most of pixels in superpixel s_i take, L_2 is the label that the jth largest number of pixels in superpixel s_i take.

Algorithm 1: *Label Subspace Construction*

```
Input: initial segmentation and classification results;
Output: co-occurrence table;
for each superpixel s_i in image do
    Compute the label histogram h_i of s_i;
    Sort the histogram h_i in descending order;
    Save the label index vector in array Lindex;
    L_1 ← the label that most of the pixels in s_i take;
    for j = 2 to M do
        L_2 ← j^th element of Lindex;
        T(L_1, L_2) ← T(L_1, L_2) + h_i(j);
    end for
end for
```

In Fig. 2, *water* and *shadow* possess similar spectral characteristics, it is high likely that they will appear in one superpixel simultaneously. This indicates that the value of T(*water*, *shadow*) in the co-occurrence table is likely to be high and then they will be grouped into the same subspace. If a pixel is classified as label l by pixel-wise classification, CBP will choose a label from the lth subspace to pass messages. After the label subspace is formed, the proposed CBP is executed as described in Algorithm 2.

Algorithm 2: *Context Guided BP*

```
Input: image I for classification and training sample
Output: classification result;
Perform mean-shift segmentation;
Perform SVM classification;
Construct label co-occurrence table;
while not converge do
    Pass message m_ij for all edge (i, j) ∈ ϵ;
end while
Compute the belief of each node;
Assign a label to each node that minimize its belief;
```

The CBP is implemented on MRF graph with superpixels as its nodes. The region adjacency graph (RAG) is built to represent the MRF graph. To ensure that belief of each node approximates the posterior probability, messages between each pairwise nodes should pass in both directions. Each node is visited in the order as saved in RAG to send messages in the forward pass, and then this procedure is executed in the inverse order. Each node will be visited only once in either forward or backward message passing.

The unary potential function of each node is given as the SVM scores and we use the Potts model to define the pairwise potential. For each superpixel, we extract the mean and variance over the spectral values of pixels in the superpixel as well as the determinant of covariance over the pixel spectral feature to represent spectral property. The normalized area of the superpixel is applied as a spatial feature to represent the size of land cover objects, which is defined as $S_i \backslash (Hi \times Wi)$, where S_i is the area of a superpixel, and H_i and W_i are height and width of input image, respectively. The normalized area helps to disambiguate objects that share similar spectral characteristics. Note that the goal of our method is to improve the computation efficiency of the BP algorithm. Thus, selecting and utilizing more discriminative spatial features is beyond the scope of this paper. The superpixel feature vector of the training sample is used to train the SVM model, then the score of each superpixel is produced by the SVM model. Since the training sample is given at the pixel level, various object classes may be contained in one superpixel. The class label of a superpixel is assigned by a majority voting rule, i.e., the label with the most number of pixels contained in this superpixel is assigned to this superpixel. The same rule applies in order to determine which label subspace is used for label pruning.

3. EXPERIMENTAL RESULTS

Qualitative and quantitative evaluation of the proposed CPB are presented in this section. For quantitative measures, classification accuracy of each class, overall accuracy, and the kappa coefficient [24] are reported.

The baseline method of SVM-based pixel-wise classification is implemented using a library for support vector machines (LIBSVM) [25]. Two SVM parameters are needed to be set, which are penalty parameter C and RBF kernel parameter γ. We use LIBSVM for training and testing the SVM classifier. Specifically, a grid search with exponentially growing sequences of C and γ is performed using a 5-fold cross-validation strategy. The training data are randomly divided into five subsets of same size and in each run four subsets are used as training data and the other subset is used for testing.

The threshold of co-occurrence is set to select classes that are relevant to a certain class initially assigned to a node in the graph. As a data-dependent parameter, it is difficult to determine the threshold theoretically. In our work, the threshold is empirically set. To determine this threshold, first, each co-occurrence table is sorted in descending order based on co-occurrence values. Then, the minimum co-occurrence of the first one third classes is selected as the threshold. We observed in the experiments that if more than one third classes were considered to pass message, the classification accuracy did not improve since irrelevant classes provided little meaningful information, while increased computation time. On the other hand, when less than one third classes were considered in message passing, the classification accuracy deteriorated as some of the relevant classes were discarded in prior before message passing.

To ensure that the number of classes in each label subspace is approximately one third of the total number of classes in the input image, different thresholds are empirically set for each label. To encourage that only one object class is contained in one superpixel, parameters of mean-shift segmentation are selected to generate oversegmentation results. The proposed CBP is compared with SBP in terms of both accuracy and

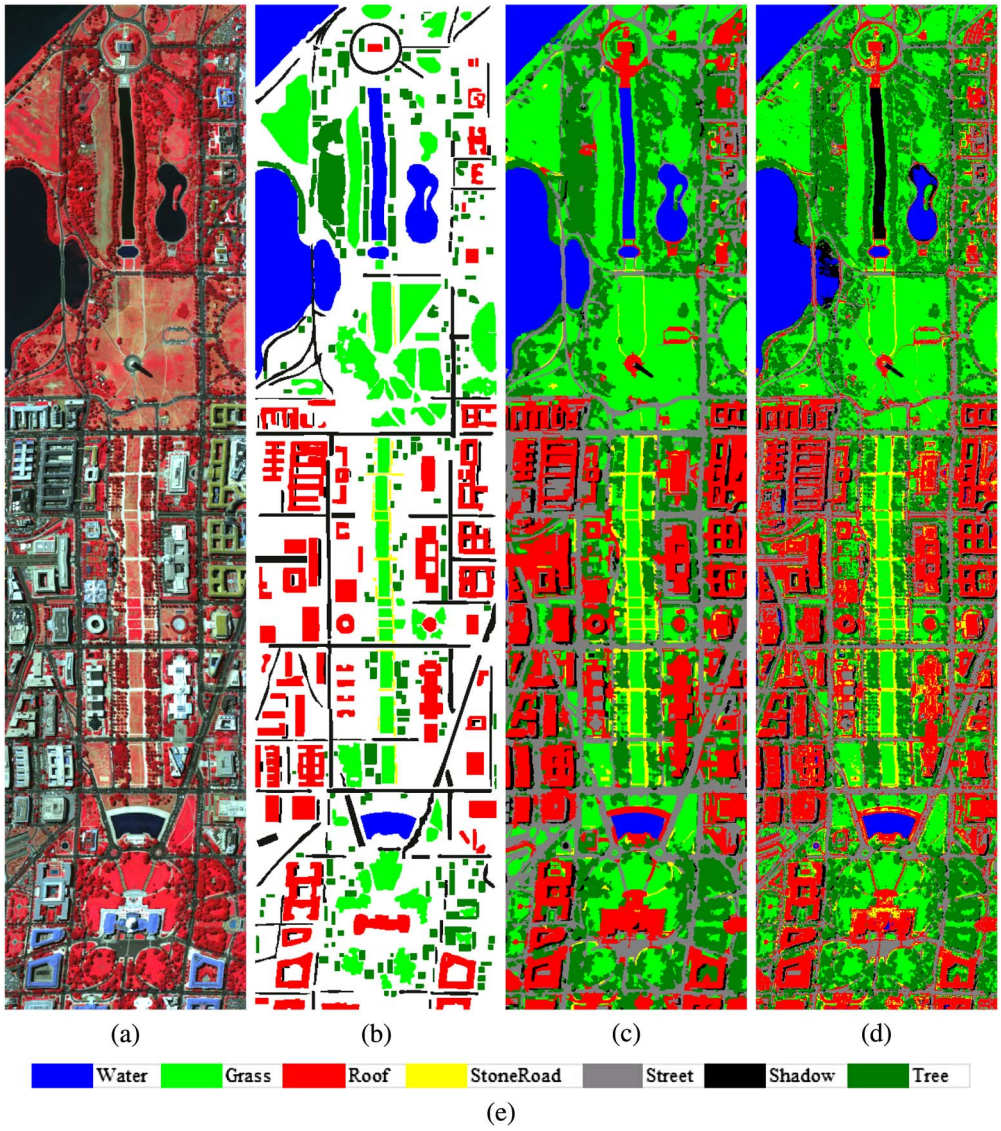


Fig. 5. Washington, DC image data. (a) Original image. (b) Ground-truth land cover labels. (c) Classification results with CBP. (d) Classification result with SVM. (e) Color coding for different classes. Best viewed in color.

efficiency with detailed analysis. In addition, we also compare our approach with belief-based label pruning BP (BBP) [21].

A. Image Data

Three images are used in the experiment. The first image is from the well-known hyperspectral digital imagery collection experiment (HYDICE)—an airborne hyperspectral image with a size of 307 × 1280, acquired over *Washington, DC* [Fig. 5(a)]. Figure 5(b) shows the corresponding ground-truth land cover labels. The number of training and test samples available per class is listed in Table 1. Since band selection is beyond the scope of this paper, we choose to use bands 63, 52, and 36 as red, green, and blue channels, respectively, to compose a pseudocolor image. The second image in the experiment is a QuickBird (QB) image with a 2.4 m spatial resolution (512 × 512), acquired from *Wuhan* in China [Fig. 6(a)]. Figure 6(b) shows the corresponding ground-truth land cover labels. This image contains nine classes (Table 2). The third image is also a QuickBird (QB) image with a 2.4 m spatial resolution (512 × 512), acquired from *Colorado* in the U.S. [Fig. 7(a)]. Figure 7(b) shows the corresponding ground-truth land cover labels. In this image five land cover classes are present (Table 3).

Table 1. Number of Samples in Washington, DC

Class Name	Training Sample	Test Sample
Water	638	29,610
Grass	1517	39,648
Roof	3394	40,829
Trail	542	1783
Road	2086	24,397
Shadow	722	2355
Tree	1622	24,457

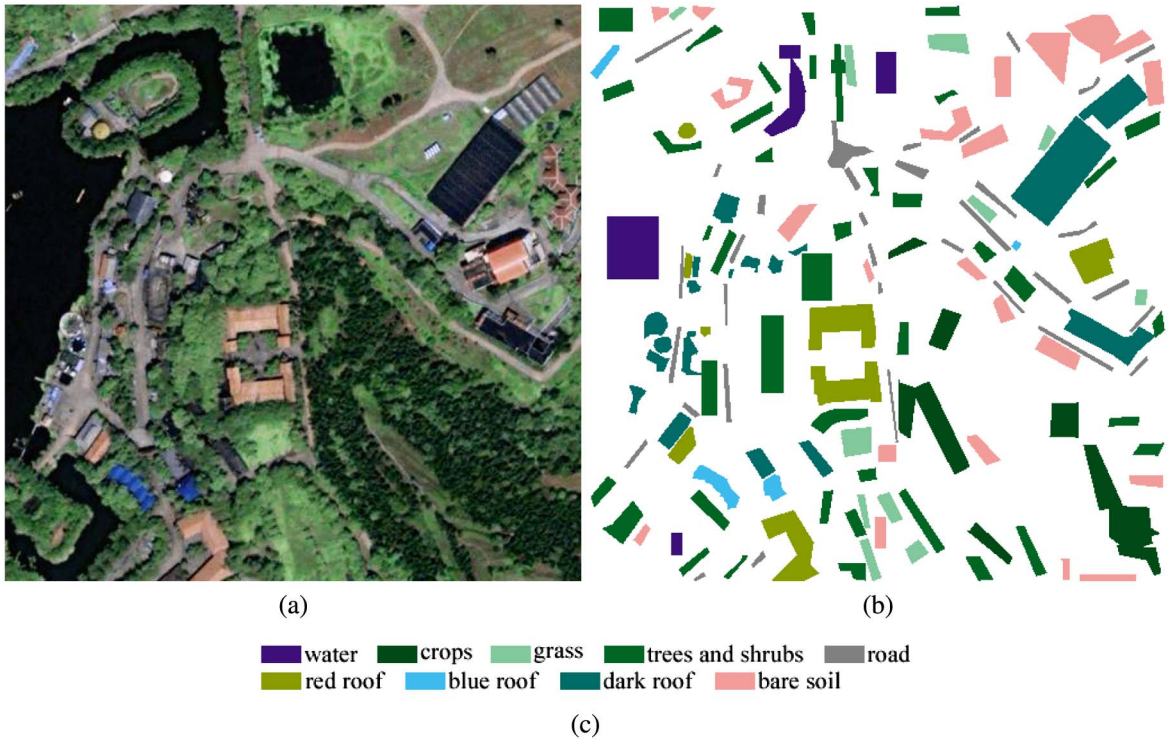


Fig. 6. Wuhan image data. (a) Original image. (b) Ground-truth land cover labels. (c) Color coding for different classes. Best viewed in color.

B. Comparison with SVM and SBP

The classification results of *Washington, DC* using CBP and SVM are shown in Figs. 5(c) and 5(d), respectively. Since only the spectral feature is used for SVM classification, objects that share similar spectral characteristics are prone to misclassification.

We choose two labels with the first two largest co-occurrence probabilities from each row of the co-occurrence table to form the label subspace. The classification result of *Washington, DC* using CBP is almost the same as that of SBP, as shown in Table 4. It demonstrates that CBP is able to maintain the classification performance of SBP even though the number of class labels used for messages passing is reduced by incorporating context information. The classification accuracy for each class and the overall accuracy (Table 4) also support that CBP can achieve almost equal classification accuracy as compared to SBP. The proposed context-based label selection can eliminate the redundancy of the original label set to speed up the SBP. This can be verified from Table 7. The running time of CBP decreases by 256 ms, and it is about one order of magnitude faster than SBP. The classification result that combines both spectral and structural features is shown in Fig. 5(c). Compared to Fig. 5(d) where a large number of *water* pixels are misclassified as *shadow*, the misclassified pixels are assigned to correct labels by incorporating structural features in MRF. The accuracy of *water* increases from 75.63% to 93.11%. Classification accuracies of other classes have also been improved, especially for those classes with accuracy less than 80%. The improvement can be attributed to the encoding of relationship between neighboring superpixels and the incorporation of the structural feature. Note that the spectral variation of *roof* is so drastic that some *roof* cannot be separated from *road* even though structural features are considered. Most of *roof* with dark color are misclassified as *road* due to its similarity with *road* and the structural feature is not discriminative enough to separate *roof* from *road*. For better discrimination, more robust structural features can be incorporated and it is beyond the scope of this paper.

Table 2. Number of Samples in Wuhan

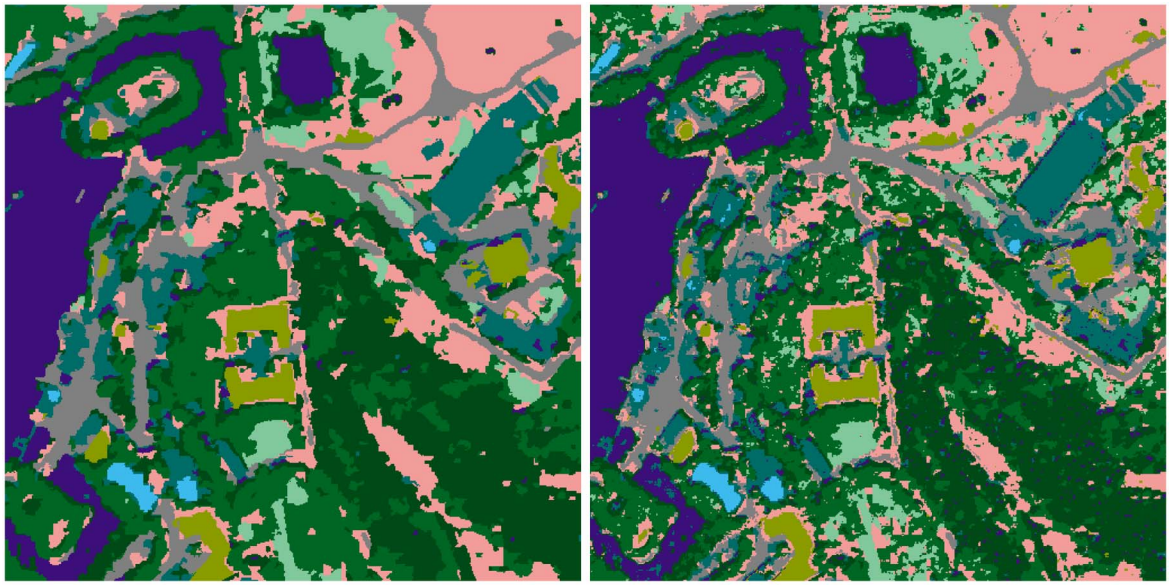
Class Name	Training Sample	Test Sample
Water	1393	4644
Tree	2028	6761
Grass	949	3166
Crop	4181	13,939
Red roof	1979	6598
Blue roof	363	1211
Dark roof	3109	10,366
Bare soil	2701	9006
Road	1399	4665

Table 3. Number of Samples in Colorado

Class Name	Training Sample	Test Sample
Cement	800	28,150
Tree	800	58,863
Grass	800	88,481
Building	800	39,361
Road	800	38,073

Table 4. Comparison of the Classification Results on the *Washington, DC* Image

	Accuracy (%)								
Methods	Water	Grass	Roof	Trail	Road	Shadow	Tree	OA (%)	kappa (%)
CBP	**93.12**	94.43	**88.36**	**94.11**	**84.14**	**92.52**	92.64	**90.82**	**88.58**
SBP	93.04	**94.55**	88.30	**94.11**	83.80	84.89	**93.06**	90.74	88.48
SVM	75.63	92.52	82.67	89.00	82.67	78.25	89.85	83.73	80.82



(a) (b)

Fig. 7. Colorado image data. (a) Original image. (b) Ground-truth land cover labels. (c) Color coding for different classes. Best viewed in color.

(a) (b)

(c)

Fig. 8. Classification results on Wuhan image data. (a) Results by CBP. (b) Results by SVM. Best viewed in color.

Table 5. Comparison of the Classification Results on the *Wuhan* Image

	Accuracy (%)										
Methods	Water	Tree	Grass	Crop	Red roof	Blue roof	Dark roof	Bare soil	Road	OA (%)	kappa (%)
CBP	**99.76**	**97.60**	82.63	85.04	91.32	**92.32**	87.14	**87.41**	**89.58**	89.35	87.59
SBP	**99.76**	**97.60**	82.98	**85.24**	**91.97**	87.45	**88.15**	**87.41**	**89.58**	**89.56**	**87.83**
SVM	99.18	91.05	**91.82**	76.81	86.84	92.15	80.89	83.49	88.6	84.92	82.53

(a) (b)

Fig. 9. Classification results on Colorado image data. (a) Results by CBP. (b) Results by SVM. Best viewed in color.

Table 6. Comparison of the Classification Results on the *Colorado* Image

	Accuracy (%)						
Methods	Cement	Tree	Grass	Roof	Road	OA (%)	kappa (%)
CBP	75.80	**73.51**	**70.06**	**86.07**	**94.23**	**77.68**	**73.27**
SBP	**80.09**	71.19	68.33	81.67	91.16	77.17	72.71
SVM	77.14	71.81	63.53	79.76	88.41	73.14	65.00

The classification results of image *Wuhan* using CBP and SVM are shown in Fig. 8. We choose three labels with the first three largest co-occurrence probabilities from each row of the co-occurrence table to form the label subspace. The running time for *Wuhan* is shown in Table 7. The computation time of CBP decreases by 221 ms, which is significantly faster than SBP. Although *Wuhan* has different characteristics (e.g., less manmade objects such as house and road) as compared to *Washington, DC*, the same conclusion can be drawn from Table 5 that the classification accuracy of CBP is almost the same as SBP, and both CBP and SBP are superior to the SVM baseline. Not surprisingly, there are some misclassified pixels. This misclassification is mainly due to the limitations in structural features and can be addressed by proposing more descriptive features, which is beyond the scope of this paper.

The classification results of image *Colorado* using CBP and SVM is shown in Fig. 9. We choose two labels with the first two largest co-occurrence probabilities from each row of the co-occurrence table to form the label subspace.The running time for *Colorado* is shown in Table 7. The computation time of CBP decreases by 37 ms, about five times faster than SBP. In image *Colorado*, class *tree* highly interweaves with class *grass* and these two classes share similar spectral characteristics. Some *tree* samples are misclassified as *grass* and vice versa. Therefore, the classification accuracy of these two classes is relatively low. The classification accuracy of CBP is very similar to SBP, and both are superior to SVM (Table 6).

Table 7. Comparison of Computation Time (in ms)

Method	Washington, DC	Wuhan	Colorado
CBP	**34**	**42**	**9**
SBP	290	263	46

Table 8. Comparison of CBP and BBP on the *Washington, DC* image

	Accuracy (%)									
Methods	Water	Grass	Roof	Trail	Road	Shadow	Tree	OA (%)	kappa (%)	Times (ms)
CBP	**93.12**	94.43	**88.36**	94.11	**84.14**	**92.52**	92.64	**90.82**	**88.58**	**34**
BBP	83.65	**94.69**	87.89	**94.50**	83.81	91.64	**93.32**	89.09	86.47	270

Table 9. Comparison of CBP and BBP on the *Wuhan* Image

	Accuracy (%)											
Methods	Water	Tree	Grass	Crop	Red Roof	Blue Roof	Dark roof	Bare soil	Road	OA (%)	kappa (%)	Times (ms)
CBP	**99.76**	**97.60**	82.63	85.04	91.32	92.32	87.14	**87.41**	**89.58**	89.35	87.59	**42**
BBP	**99.76**	**97.60**	**82.98**	**85.23**	**92.30**	**93.23**	**88.15**	**87.41**	**89.58**	**89.71**	**88.01**	159

Table 10. Comparison of CBP and BBP on the *Colorado* Image

	Accuracy (%)							
Methods	Cement	Tree	Grass	Roof	Road	OA (%)	kappa (%)	Times (ms)
CBP	75.80	73.51	**70.06**	**86.07**	**94.23**	**77.68**	**73.27**	**9**
BBP	**76.59**	**73.67**	68.77	85.50	94.21	77.23	72.28	35

In Table 7, we compare the running time of performing CBP and SBP on three image data. Our method is implemented in C++ without code optimization on a laptop with an Intel core i7 2.4 GHz CPU and 8 GB RAM. The results show that the proposed CBP is significantly faster than the SBP on different images while maintaining the capacity of SBP. The reason is that the computation complexity of CBP is reduced from $O(n \times M^2)$ to $O(n \times N^2)$, where M is the number of classes in the test image, N is the average number of classes in each label subspace, and n is the number of superpixels. Due to the fact that the number of classes in each label subspace is smaller than M, N is smaller than M. As a result, our method can be significantly faster than the original SBP algorithm since the time complexity is quadratic on M and N. It is worth noting that we only list the running time of performing messages passing in Table 7. Since the other components of the classification framework of these two methods are the same, the running time of these parts are not taken into account.

In summary, the experiment results show that the proposed CBP can significantly speed up the SBP while achieving similar accuracy. In addition, the classification results using structural features are more accurate compared to the classification results using only spectral information.

C. Comparison with Belief-Based Label Pruning BP

In this section, we compare the performance of the proposed CBP with belief-based label pruning BP (BBP) [21]. When passing messages between neighboring nodes, BBP does not consider those labels x_i when the belief $b(x_i)$ is higher than a threshold. For fair comparison, we apply BBP with different thresholds and report the highest classification results of BBP as compared to CBP. The comparison results of three image data sets are shown in Tables 8–10. For all three images, we can see that performance of the proposed method is almost the same with BBP in terms of overall accuracy, kappa coefficient, and accuracy of each class. This validates that context obtained by analyzing the spectral similarity in each superpixel can represent the relationship between different classes, and this relationship can be used to guide messages passing in BP. Although the accuracy of CBP is not superior to BBP, the running time of CBP is much less than BBP. The reason that BBP is slower than CBP is that the number of labels used for messages passing is determined by belief of that node. To get higher classification accuracy, the threshold of belief is usually set higher. Consequently, the number of labels selected by belief threshold is larger than that of CBP, which results in a slower computation as compared to CBP.

4. CONCLUSIONS

In this paper, a context guided BP (CBP) is proposed for HSRMI classification at the superpixel level. As contrast to standard BP (SBP) that passes messages between neighboring nodes by considering all possible class labels, the proposed CBP reduces the number of labels involved in messages passing by considering spectral similarity between land cover types as context information for improved efficiency. Experiments on three different kinds of HSRMI image data showed that the proposed CBP is much faster than SBP while retaining the accuracy of SBP. As compared to the baseline method classifying at the pixel level, the proposed method achieved higher classification accuracy. This is due to the fact that superpixel-based feature descriptors are more discriminative than pixel-based spectral descriptors. Future work includes the introduction of superpixel-based spectral similarity context, which help to form a hierarchical MRF and better capture the local context

information of HSRMI. In addition, we would like to identify and utilize more discriminative spatial features to further improve the classification accuracy.

†Le An and Tiancan Mei contributed equally to this work and are both first authors.

China Scholarship Council (201208420525); National Natural Science Foundation of China (NSFC) (40971219); Surveying, Mapping and Geoinformation Research Fund (210300001).

REFERENCES

1. J. Paola and R. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification," IEEE Trans. Geosci. Remote Sens. **33**, 981–996 (1995).
2. S. Moustakidis, G. Mallinis, N. Koutsias, J. Theocharis, and V. Petridis, "SVM-based fuzzy decision trees for classification of high spatial resolution remote sensing images," IEEE Trans. Geosci. Remote Sens. **50**, 149–169 (2012).
3. J. T. Guillen-Bonilla, E. Kurmyshev, and E. González, "Algorithm for training the minimum error one-class classifier of images," Appl. Opt. **47**, 541–547 (2008).
4. N. Longbotham, C. Chaapel, L. Bleiler, C. Padwick, W. Emery, and F. Pacifici, "Very high resolution multiangle urban classification analysis," IEEE Trans. Geosci. Remote Sens. **50**, 1155–1170 (2012).
5. S. Newsam, L. Wang, S. Bhagavathy, and B. S. Manjunath, "Using texture to analyze and manage large collections of remote sensed image and video data," Appl. Opt. **43**, 210–217 (2004).
6. J. Yuan, D. Wang, and R. Li, "Remote sensing image segmentation by combining spectral and texture features," IEEE Trans. Geosci. Remote Sens. **52**, 16–24 (2014).
7. H. Oriot and A. Michel, "Building extraction from stereoscopic aerial images," Appl. Opt. **43**, 218–226 (2004).
8. B. Bhanu, P. Symosek, and S. Das, "Analysis of terrain using multispectral images," Pattern Recogn. **30**, 197–215 (1997).
9. C. Benedek and T. Sziranyi, "Change detection in optical aerial images by a multilayer conditional mixed Markov model," IEEE Trans. Geosci. Remote Sens. **47**, 3416–3430 (2009).
10. L. Yi, G. Zhang, and Z. Wu, "A scale-synthesis method for high spatial resolution remote sensing image segmentation," IEEE Trans. Geosci. Remote Sens. **50**, 4062–4070 (2012).
11. L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high spatial resolution images," IEEE Trans. Geosci. Remote Sens. **44**, 2587–2600 (2006).
12. C. Galleguillos and S. Belongie, "Context based object categorization: a critical survey," Comput. Vis. Image Understanding **114**, 712–722 (2010).
13. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6**, 721–741 (1984).
14. G. Poggi, G. Scarpa, and J. Zerubia, "Supervised segmentation of remote sensing images based on a tree-structured MRF model," IEEE Trans. Geosci. Remote Sens. **43**, 1901–1911 (2005).
15. A. Katartzis, I. Vanhamel, and H. Sahli, "A hierarchical Markovian model for multiscale region-based classification of vector-valued images," IEEE Trans. Geosci. Remote Sens. **43**, 548–558 (2005).
16. T. Blaschke, "Object based image analysis for remote sensing," ISPRS J. Photogr. Remote Sens. **65**, 2–16 (2010).
17. P. Kohli, L. Ladick, and P. Torr, "Robust higher order potentials for enforcing label consistency," Int. J. Comput. Vis. **82**, 302–324 (2009).
18. P. Felzenszwalb and D. Huttenlocher, "Efficient belief propagation for early vision," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004)* (IEEE, 2004), Vol. **1**, pp. I–261–I–268.
19. J. Besag, "Spatial interaction and the statistical analysis of lattice systems," J. R. Stat. Soc. Ser. B. Methodol. **36**, 192–236 (1974).
20. J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Generalized belief propagation," in *Annual Conference on Neural Information Processing Systems (NIPS)* (MIT, 2000), Vol. **13**, pp. 689–695.
21. N. Komodakis and G. Tziritas, "Image completion using efficient belief propagation via priority scheduling and dynamic pruning," IEEE Trans. Image Process. **16**, 2649–2661 (2007).
22. M. Dalla Mura, J. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," IEEE Trans. Geosci. Remote Sens. **48**, 3747–3762 (2010).
23. D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," IEEE Trans. Pattern Anal. Mach. Intell. **24**, 603–619 (2002).
24. R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," Remote Sens. Environ. **37**, 35–46 (1991).
25. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," ACM Trans. Intell. Syst. Technol. **2**, 27 (2011).