A Bilevel Contextual MRF Model for Supervised Classification of High Spatial Resolution Remote Sensing Images

Yu Shen, Jianyu Chen, *Member, IEEE*, Liang Xiao, *Member, IEEE*, and Delu Pan

***Abstract*—Markov random field (MRF) based methods have been widely used in high spatial resolution (HSR) image classification. However, many existing MRF-based methods put more emphasis on pixel level contexts while less on superpixel level contextual information. To cope with this issue, this article presents a novel bilevel contextual MRF framework, named BLC-MRF, for HSR imagery classification. Specifically, pixel and superpixel level dependence are incorporated into the proposed MRF model to fully exploit spectral–spatial contextual information and preserve object boundaries in HSR images. In BLC-MRF, a pixel level MRF model is first performed and then cascaded as an input of a superpixel level MRF. In superpixel level, unary and pairwise potential terms are constructed by using the superpixel probability estimation method and spectral histogram distance, respectively. At last, a contextual MRF model is conducted and the final classification map can be computed by using α-expansion algorithm. The benefits of BLC-MRF are twofold: first, the pixel and superpixel level contextual information can be exploited under MRF framework to preserve object boundaries for improving the classification performance, and, second, the algorithm can provide promising results with a small number of training samples. Experimental results on three HSR datasets demonstrate that the proposed approach outperforms several state-of-the-art methods in terms of the classification performance.**

***Index Terms*—Classification, high spatial resolution (HSR), Markov random field (MRF), superpixel map.**

Manuscript received April 25, 2019; revised September 25, 2019; accepted October 24, 2019. Date of publication December 2, 2019; date of current version February 4, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFC1400903, in part by the National Major Research Plan of China under Grant 2016YFF0103604, in part by the National Natural Science Foundation of China under Grants U1609202, 41501466, 61571230, and 61871226, in part by the Jiangsu Provincial Natural Science Foundation of China under Grant BK20161500, in part by the Fundamental Research Funds for the Central Universities under Grant 30918011104, and in part by the Jiangsu Provincial Social Developing Project under Grant BE2018727. *(Corresponding author: Yu Shen.)*

Y. Shen is with the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China, and also with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: shenyu0305@126.com).

J. Chen is with the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China (e-mail: chenjianyu@sio.org.cn).

L. Xiao is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: xiaoliang@njust.edu.cn).

D. Pan is with the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China (e-mail: pandelu@sio.org.cn).

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

Digital Object Identifier 10.1109/JSTARS.2019.2950946

I. Introduction

With the rapid development of satellite imaging technologies, a number of high spatial resolution (HSR) remotely sensed images are being acquired. Therefore, automatic understanding of HSR images has been a notably hot topic of research in recent years. One of the important topics is HSR image classification, which aims at classifying an HSR remote sensing image into a thematic map. Due to the rich geometric and detailed information, the classification of HSR image plays an important role in various areas such as damage assessment of environment accidents, vegetation detection, and urban planning [1], [2].

Over the past decades, a large number of approaches have been developed for HSR image classification. Among these approaches, many pixel level approaches, such as support vector machine (SVM) [3], multinomial logistic regression (MLR) [4], and extreme learning machine (ELM) [5], have been proven effective in HSR image classification under small-size training samples. Pixel-based approaches usually exploit discriminant spectral information to distinguish varying objects of remotely sensed images. However, due to the existence of noise and mixed spectral pixels in HSR remote sensing images, pixel-based classifiers can result in outliers or errors in thematic maps. To overcome these drawbacks, on the one hand, kernel tricks are proposed to improve the linear separability of data. For example, composite kernel SVM (SVMCK) [6] and ELM (ELMCK) [7] are considered for remote sensing image classification and provide better performance than traditional classifiers. On the other hand, more and more researchers propose to exploit spectral–spatial contextual information, which assumes that the nearest neighboring pixels of an HSR image share similar spectral features and consist the same land cover types.

There are five main approaches to integrate spectral–spatial contextual information [8], [9]: 1) filtering-based method, 2) relearning method, 3) deep learning based method, 4) object-based image analysis (OBIA), and 5) random fields. For the filtering-based method, many different types of effective filters, such as Gabor filter [10], [11], bilateral filer [12], wavelet transformation [13], and local binary pattern (LBP) [14], [15] have been developed for remotely sensed image classification in success. Relearning-based method considers the frequency and spatial distribution of image classes and then class distribution is learned in an iterative manner [8]. More recently, deep learning based supervised classification models have attracted increasing attention. Deep learning based models have the ability to learn HSR image features for classification with a deep neural network [16]. For example, in [17], convolutional neural network (CNN) is used for extracting spatial feature. CNN is a kind of deep neural network that can learn convolutional filters automatically. In [17], the HSR images are set as small patches (e.g., $5 \times 5, 7 \times 7$) for the deep network training by back propagation. Recently, the CNN extensions residual network (ResNet) is also applied in HSR image classification. The ResNet uses fundamental CNN structures as residual blocks to facilitate learning a deeper network. In [18], a spectral–spatial ResNet (SSRN) is proposed for supervised remote sensing image classification. Specifically, SSRN uses residual blocks to connect the convolutional layers from both spectral and spatial dimension. Therefore, an improvement in the classification performance can be ensured in SSRN. However, deep learning based methods only show their potential with an enough amount of training samples [16].

With the spatial resolution of remote sensing images increasing, the OBIA (or GEOBIA for geospatial object based image analysis) has attracted the attention of many researchers [19], [20]. The object-based classification framework has the capacity to integrate segmentation and classification algorithms to improve the classification performance [21]. In this framework, segmentation algorithms, such as watershed algorithms [22] and mean shift [23], are usually used to divide spatial regions into nonoverlapping homogeneous objects. After that, the segmentation objects (superpixels) are considered as basic analysis units to generate spectral–spatial features of the original images. By incorporating spatial information of homogeneous regions, object-based methods are believed to help preserve object boundaries and suppress the effect of classification errors or outliers that often appear in pixel-based classification methods [24]. For example, in [25] and [26], superpixel level sparse representation classification methods are developed to jointly integrate spectral and spatial information for remote sensing image classification. However, it is difficult to obtain an accurate segmentation map. Once a superpixel is estimated wrongly, pixels in that superpixel will all be classified by mistake.

Random field methods, including Markov random field (MRF) and conditional random field (CRF), are advanced statistic modeling tools that can effectively integrate spatial contextual information into image processing under Bayesian inferring framework [27], [28]. Specifically, both MRF and CRF based methods are able to incorporate labeled data and observed spatial contextual feature into an integrated framework. The CRF is a type of MRF whose clique potential is conditioned on input features [29], [30].

Both group of random fields have been widely used in remote sensing image processing. It has been proven that MRF methods has the ability to improve the classification performance with spatial priors. For remote sensing image classification problem, the MRF is usually defined to model the spatial dependence in the local neighboring pixels (four- or eight-neighborhood system)[31]–[33]. For instance, in [34], the SVM and MRF are integrated as a spectral–spatial classification model for remote sensing image classification. Some other remote sensing images features, such as three-dimensional wavelet [35], nonlocal spatial information [36] and co-occurrence matrix [37] are also incorporated into MRF models to improve the classification performance. Unfortunately, the spatial prior of MRF can also lead to an oversmoothed phenomenon [37]. Due to the spatial smoothness property of MRF, this phenomenon usually occurs in the object boundaries where the values of pixels change drastically. Therefore, in order to preserve the object boundaries, some researchers start to focus on introducing object-based methods into MRF classification process. For instance, in [38] and [39], researchers applied superpixel-based majority voting as a postprocessing of MRF classification. However, these MRFs still only consider superpixels as region constraints and the interactions between superpixels are not fully exploited. If a superpixel is estimated wrongly, pixels in the superpixel will be affected by mistake. To deal with this problem, one solution is to use higher order potentials in MRF [40]. However, the use of higher order potentials is limited due to the complexity of models and difficulty of efficient inference [41]. Another solution is to apply superpixel level MRF methods. In superpixel level MRF, the spatial dependence is modeled in adjacent superpixels. Just like pixel level MRF model, the class label of a superpixel is not only decided by itself, but also affected by its neighboring superpixels. Some superpixel level MRF models are also used in remote sensing image problems. For example, in [42], pixel level and superpixel level MRF work in parallel and generate a decision fusion result for remote sensing image classification.

In this article, to incorporate spectral–spatial contextual information and boundary preserving, a novel bilevel framework integrating pixel and superpixel level MRF, named BLC-MRF, is proposed for remote sensing image classification. The overview of our proposed method is displayed in Fig. 1. In pixel level, the MRF model focuses on local spectral dependence and the neighborhood of each central pixel. After that, a segmentation algorithm is used to generate a superpixel map. We then construct the superpixel level MRF by integrating pixel level outputs and superpixel spatial dependence. The final classification result is obtained by solving the bilevel MRF optimization problem.The main contributions of this article are summarized as follows.

1) We propose a bilevel contextual MRF framework for HSR imagery classification. Through the integration of pixel level and superpixel level MRF, a coarse-to-fine mechanism is established to make full use of spectral–spatial contextual information and produce promising results with a small number of training samples.

2) A spectral histogram distance as feature similarity descriptors of adjacent superpixels is presented to create a superpixel level MRF minimizing model, in which new unary and pairwise potential terms are proposed. By minimizing the superpixel level MRF model, adjacent homogeneous superpixels merging process can be created to refine misclassified superpixels. Moreover, the object boundaries can be preserved as superpixels can help suppress outliers in the boundary areas.

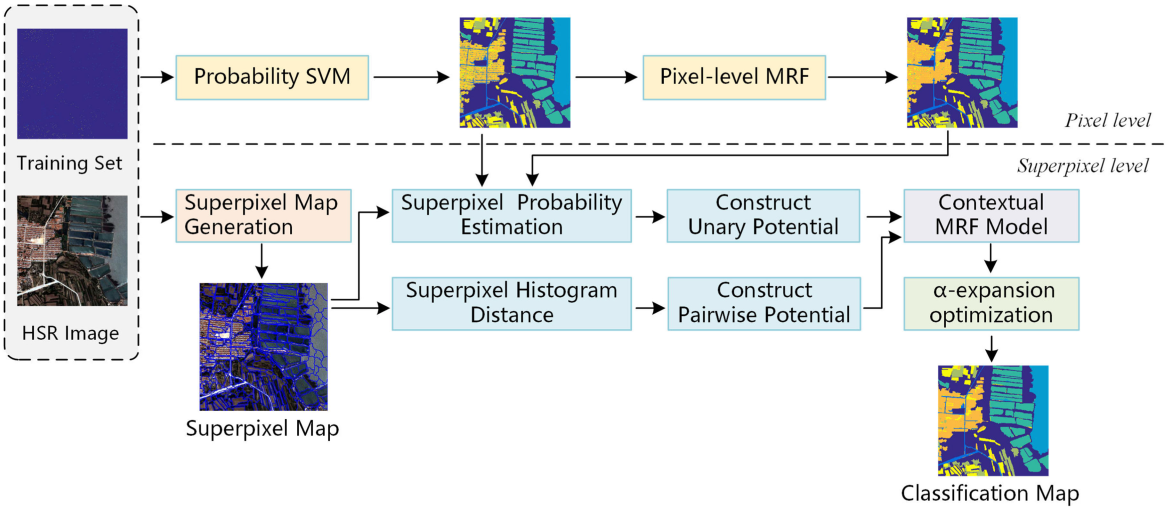


Fig. 1. Flowchart of the proposed BLC-MRF classification framework.

Experimental results on three HSR remote sensing image datasets (two multispectral images and a hyperspectral image) demonstrate the efficiency of our proposed framework. In comparison with some state-of-the-art classifiers, our proposed method shows a promising classification performance when the training data are limited (0.3% to 0.8% on three datasets).

The remainder of this article is organized as follows. Section II briefly describes the pixel level MRF model. The proposed classification framework (BLC-MRF) is presented in Section III. Section IV provides the parameter sensitivity analysis and experimental results. Section V gives the conclusion and future work.

II. Pixel Level MRF Model

Let us first define some notations used throughout this article. Let $X = \{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N\} \in \mathbb{R}^{B\times N}$ be an HSR image with N pixels and B dimensions (spectral bands), and each x_i represents the observed spectral vector. Let $\mathrm{y} = \{\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_N\}$ denote the corresponding labels of X, and each y_i takes a value from label set $\mathcal{K} = \{1, \ldots, K\}$. $\mathcal{K}$ represents the label set with K class labels.

As described in Section I, the MRF model is an advanced statistic modeling tool that directly models the posterior probability of labels, which has been proven effective in integrating spectral–spatial information of remote sensing images [27]. The joint probability of MRF can be expressed with a Gibbs random field, which can be written as

$$P(\mathrm{y}) = (1/z)\exp(-E) \tag{1}$$

where z is a normalization factor and E is known as the energy function. Given the observed image data, the aim of the classification task is to find the optimal estimation of $\hat{\mathrm{y}}$, which can be formulated as

$$\hat{\mathrm{y}} = \arg\max_{\mathrm{y}} P(\mathrm{y}|\mathrm{x}) = \arg\min_{\mathrm{y}} E(\mathrm{y}|\mathrm{x}). \tag{2}$$

According to the Bayesian maximum *a posterior* (MAP) framework, finding the maximization of the probability $P(\mathrm{y}|\mathrm{x})$ is equivalent to minimizing the energy function $E(\mathrm{y}|\mathrm{x})$, which can be represented as the sum of two potential terms

$$E(\mathrm{y}|\mathrm{x}) = \sum_{i\in N}\psi_i(\mathrm{y}_i, \mathrm{x}) + \alpha\sum_{i\in N, j\in V_i}\psi_{ij}(\mathrm{y}_i, \mathrm{y}_j, \mathrm{x}) \tag{3}$$

where $\psi_i(\mathrm{y}_i, \mathrm{x})$ denotes the unary potential term, while $\psi_{ij}(\mathrm{y}_i, \mathrm{y}_j, \mathrm{x})$ is the pairwise potential term. V_i represents the neighboring pixels of i and α is the tuning parameter of the pairwise term.

In the unary potential term, probability classifiers are usually employed to reflect the relationship between pixels and corresponding class labels. Based on the observed feature vector, a probability classifier can be used to compute the possibilities of each pixel belonging to a certain class independently. As a widely used probability classifier, the probabilistic SVM [43] is applied here in this article. In this end, the unary potential can be denoted as

$$\psi_i(\mathrm{y}_i, \mathrm{x}) = -\ln(P(\mathrm{y}_i = k|\mathrm{x})) \tag{4}$$

where $\ln(P(\mathrm{y}_i = k|\mathrm{x}))$ is the probability of sample x belonging to class k, which can be computed by probabilistic SVM.

The pairwise potential term is usually used to integrate spectral–spatial information into the MRF model. Since the neighboring pixels have a higher probability of belonging to the same land cover types, the pairwise potential term reflects the conspicuous spatial correlation in each neighboring system [44], which is defined as

$$\psi_{ij}(\mathrm{y}_i, \mathrm{y}_j, \mathrm{x}) = \begin{cases} 0, & \text{if } \ i = j \\ 1 + \theta_a(\exp(-\theta_b||\mathrm{x}_i - \mathrm{x}_j||^2)), & \text{else} \end{cases} \tag{5}$$

where θ_a is a constant smooth parameter and θ_b represents the mean square difference between the neighboring pixels.

From (2)–(4), it is obvious that the MRF model is able to integrate spatial contextual information in class labels and

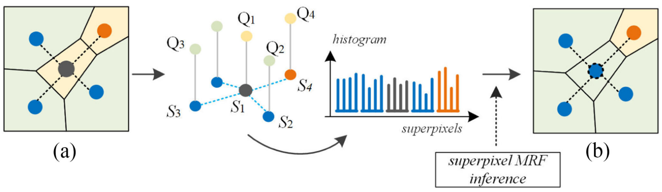


Fig. 2. Graphical example of the superpixel level MRF model. S_i represents a superpixel and Q_i represents the corresponding class label. The light green and light yellow color represent two different land cover areas. The misclassified black node in (a) is corrected in (b) after superpixel level MRF.

formulate a complete expression. Therefore, the pixel level MRF model has been successfully applied in various image processing fields. However, with the spatial resolution of remote sensing images increasing, it is difficult to effectively suppress the spectral noise and outliers by simply using a pixel level MRF model. Therefore, in this article, to further improve the HSR image classification performance, a pixel and superpixel integrated MRF model (BLC-MRF) is proposed for HSR remote sensing images classification.

III. Proposed Classification Framework

In this article, to further exploit the spectral–spatial information, a bilevel MRF classification algorithm (BLC-MRF) is proposed, as illustrated in Fig. 1. In the pixel level, a pixel-wise classifier is applied on HSR imagery to obtain an initial classification result. Then, MRF is applied on initial result and the observed image data, aiming to produce the first level classification outputs. The probabilistic SVM is utilized in pixel level MRF to generate initial classification map.

In the superpixel level of BLC-MRF framework, as shown in Fig. 2, superpixels play the role of basic units in classification process. A segmentation algorithm named ERS [45] is used to generate superpixel map. After that, superpixels are classified by an improved probability estimation method. In this step, the pairwise potential term is designed to integrate spectral–spatial features of superpixels. At last, BLC-MRF is performed to obtain the final result and the corresponding classification map.

A. Superpixel Map Generation

A variety of segmentation algorithms can be used for superpixel generation, such as SLIC [46] and mean shift [23]. In this article, ERS algorithm is used for remote sensing image segmentation. ERS algorithm has good accuracy and boundary recall properties with a high time efficiency. This method considers image segmentation as a clustering problem whose object function includes two components: 1) the entropy rate; 2) the balance term on the cluster distribution. The ERS algorithm for HSR segmentation can be briefly described in the following.

1) Map an HSR image to a graph $G = (V, E)$, where the vertices represent the pixels and edge weights denote the pairwise similarities. ERS algorithm recasts image segmentation as a graph cut problem. The goal is to find a subset of edges such that the resulting graph contains exactly M clusters (superpixels). The objective function is formulated as
$$\max_{A} H(A) + \lambda B(A) \quad \text{s.t.} A \subseteq E \tag{6}$$
where $H(A)$ is the entropy rate term and is acquired by the random walk on the constructed graph. $B(A)$ is the balance term used for encouraging superpixels with similar sizes. The λ is the weight of the balance term.
2) Maximize objective function through a greedy algorithm [47]. This algorithm starts with an empty set (a fully disconnected graph with $A = \emptyset$) and sequentially adds edges to the set. At each iteration, it adds edges that yields the largest gain. The iteration stops when the number of connected subgraphs reaches a preset value (superpixel number M).
3) Generate a superpixel map. The pixels in each subgraph are extracted and form a superpixel. The subgraphs obtained in previous steps are represented as superpixels of the HSR image.

At last, an nonoverlapping superpixel map is obtained from the original HSR image, which will be used for superpixel level contextual MRF model construction.

B. Superpixel Level MRF Classification

In superpixel level MRF classification step, as shown in Fig. 2, superpixels act as basic variables. It can be assumed that a superpixel map $S = \{S_1, S_2, \ldots, S_M\}$ with M superpixels has been obtained after superpixel map generation. Corresponding to the superpixels S, the superpixel class label is denoted by $Q = \{Q_1, Q_2, \ldots, Q_M\}$, where Q_m takes the value from label set $\mathcal{K}$. The superpixel MRF classification task is to find an optimal estimation $\hat{Q}$ that maximizes the posterior, which is formulated as
$$\hat{Q} = \arg\max_{Q} P(Q|S) = \arg\min_{Q} E(S). \tag{7}$$
Under MAP framework, finding the maximum probability $P(Q|S)$ is equal to finding the minimization of energy function $E(S)$. Therefore, this energy function can be represented as a sum of two potential terms
$$E(S) = \sum_{m \in M} \psi_m(Q_m, S) + \beta \sum_{m \in M, r \in U_m} \psi_{mr}(Q_m, Q_r, S) \tag{8}$$
where $\psi_m(Q_m, S)$ denotes the unary potential term and $\psi_{mr}(Q_m, Q_r, S)$ is the pairwise potential term. U_m is the local neighboring superpixels of superpixel m and β is the tuning parameter of the pairwise term.

After the establishment of the superpixel level MRF model, the rest problem is to formulate the unary and pairwise potential terms.

1) Unary potential term: Unary potential describes the relationship between a single superpixel and its corresponding class label, which usually relies on local spectral features. Specifically, the unary potential term can be defined as
$$\psi_m(Q_m, S) = -\ln(P(Q_m = k|S)) \tag{9}$$

where $P(Q_m = k|S)$ is the class probability of superpixel S belonging to class k.

After superpixel map generation, the class probability of each superpixel is determined based on pixel level classification result. In this step, an improved superpixel probability estimation strategy is proposed by integrating probabilistic SVM and MRF classification results of first level. Let $S_m = \{\mathrm{x}_1^m, \mathrm{x}_2^m, \ldots, \mathrm{x}_{n^m}^m\}$ be a superpixel with n^m pixels. For each superpixel S_m in an HSR image, let n_k^m be the number of pixels belonging to class k after first-level MRF classification and let $P_k(\mathrm{x}_i^m)$ denote the probability of pixel x_i^m belonging to class k. Then, the class probability of superpixel S_m can be defined as

$$P(Q_m = k|S_m) = \frac{1}{Z}\left(n_k^m + \sum_{x_i^m \in S_m} P_k(\mathrm{x}_i^m)\right), k = 1, \ldots, K. \tag{10}$$

The Z is the normalization factor and can be calculated as $Z = \sum_{k=1}^{K}(n_k^m + \sum_{\mathrm{x}_i^m} P_k(\mathrm{x}_i^m))$.

2) Pairwise potential term: Since the neighboring pixels in homogeneous regions usually share the same class label, the pairwise potential term models a smooth prior to integrate the neighboring superpixels relations. The typical smooth prior can be written as

$$\psi_{mr}(Q_m, Q_r, S) = \begin{cases} 0, & \text{if} \quad m = r \\ g_{mr}(S), & \text{else} \end{cases} \tag{11}$$

where $g_{mr}(S)$ is the smooth term related to the superpixels S. The smooth term is designed to measure the difference between adjacent superpixels.

To describe the dependence among superpixels, we need to represent these superpixels with some feature descriptors. A superpixel can be described in many aspects, such as spectrum, texture, shape, and size. Among these descriptors, spectrum histogram is an effective feature descriptor to represent the superpixel spectral feature and has been widely used in pattern recognition [48].

At first, we uniformly quantize and divide each spectral band into T levels for an HSR image. The histogram of each superpixel is then calculated in the feature space of T^B bins. Let Hist_{Q_m} be the normalized histogram of a superpixel Q_m and the similarity between superpixels Q_m and Q_r can be defined by Bhattacharyya coefficient [49] as

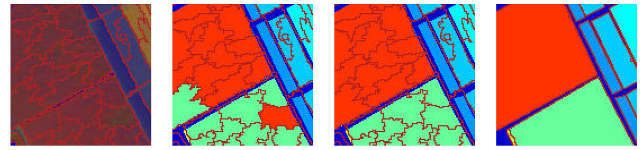
$$\rho(Q_m, Q_r, S) = \sum_{u=1}^{T^B} \sqrt{\mathrm{Hist}_{Q_m}^u \cdot \mathrm{Hist}_{Q_r}^u} \tag{12}$$

where the superscript u represents the uth bin of feature space. The Bhattacharyya coefficient ρ actually reflects the perceptual similarity between two superpixels. The higher the ρ between Q_m and Q_r is, the higher the similarity between them is.

By using the Bhattacharyya coefficient between two adjacent superpixels, the term of g_{mr} can be defined as

$$g_{mr}(S) = 1 - \rho(Q_m, Q_r, S). \tag{13}$$

The main purpose of the superpixel level MRF is to build an adjacent superpixel level refining and merging process. By minimizing the MRF energy, the misclassified superpixels in a homogeneous area would be rectified through their adjacent superpixels in the MRF model. For a superpixel S_i with the corresponding label $Q_i = k$, which is assigned by the pixel level MRF model, we form its adjacent superpixel $\bar{S}_i = \{S_i^j\}, j = 1, 2, \ldots, n$. Then, the superpixel similarity $g_{ij}(S) = 1 - \sum_{u=1}^{T^B} \sqrt{\mathrm{Hist}_{Q_i}^u \cdot \mathrm{Hist}_{Q_j}^u}$ between S_i and the adjacent superpixels in $\bar{S}_i$ can be calculated with (12) and (13). In a homogeneous area with low value of spectral histogram difference for each adjacent superpixel pair, when a superpixel S_i is misclassified, the value of $g_{ij}(S)$ will be large, and thus, the MRF has larger energy.



(a) (b) (c) (d)

Fig. 3. Process of superpixel level MRF optimization over a part of a remote sensing image. (a) Generated superpixels. (b) Classification results after superpixel probability estimation. (c) Classification results after superpixel level MRF. (d) Final output.

In the iterative procedure, the correct label information of surrounding superpixels will be propagated to the current misclassified superpixel with the MRF energy decreasing. In this end, when the superpixel MRF achieves the minimum energy [see (8)], the misclassified superpixels in this homogeneous area will be assigned a correct label, and the adjacent superpixels will be merged to a connected area with the same correct labels. In Fig. 3, we show an example for the process of superpixel level MRF optimization over a part of a remote sensing image.

C. Inference MRF Model by α-Expansion Algorithm

The solution to the pixel level and superpixel level MRF model is an NP-hard problem [50], thus, it is time demanding to find the perfect solution. Fortunately, this kind of problem can be inferred by a variety of optimization algorithms, such as the iterated conditional model [51], particle swarm optimization [52], and graph cuts [53]. Among these methods, a graph cuts algorithm has been proven fast in a global energy minimizing problem. In our proposed model, a graph cuts based α-expansion algorithm [54] is employed for the bilevel MRF model optimizing.

To find an optimal solution of the MRF model, a special local search method is designed in α-expansion algorithm to segment all α and non-α values with graph cuts [54]. The algorithm is efficient in iterating through each possible label for α until it converges. Owing to the property of α-expansion algorithm, the inference of the MRF model can be effectively optimized.

The previous sections gave a description of key steps modeling MRF classification process, including superpixel map generation and definition of potential terms in the MRF model. The whole BLC-MRF classification framework is described in Algorithm 1.

TABLE I
CLASSIFICATION ACCURACY (%), OA (%), AND KAPPA COEFFICIENT FOR QINGDAO AREA

class	Test	Train	SVMCK [6]	LBP-DF [14]	MFASR [55]	SVM-OO [39]	SVMMRF [34]	SuperMRF [29]	SMLR-WMRF [31]	DPSCRF [41]	CNN [17]	SSRN [18]	BLC-MRF
1	3687	120	88.69	86.18	67.08	90.58	81.17	92.17	94.37	93.30	95.49	93.79	82.83
2	9592	120	90.65	84.76	78.54	89.36	91.25	94.19	92.98	92.84	96.60	96.44	93.97
3	30565	120	97.30	95.38	99.57	96.99	95.85	99.14	96.76	97.22	99.36	99.19	98.20
4	17913	120	99.20	98.61	97.51	99.50	99.19	99.38	99.31	99.41	99.76	99.75	99.41
5	6107	120	96.07	88.32	99.13	99.11	98.61	99.62	99.36	99.12	95.18	97.48	99.20
6	22930	120	67.78	84.63	95.68	83.64	83.20	83.98	89.79	90.00	80.19	86.35	91.05
OA			89.09	91.33	94.57	93.19	92.42	94.58	95.18	95.33	93.87	95.43	**95.65**
κ			0.860	0.888	0.929	0.912	0.902	0.930	0.938	0.940	0.921	0.941	**0.943**

The best results of OA and Kappa are highlighted in bold.

Algorithm 1: BLC-MRF Framework for HSR Image Classification.

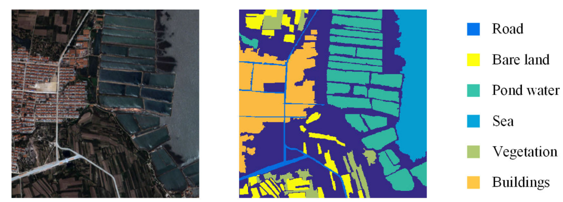
Input: Raw HSR image with pixels $X = \{x_1, x_2, \ldots, x_N\} \in \mathbb{R}^{B \times N}$, where B is the dimension and x denotes the pixels. Training samples I with class number K.
Output: Classified thematic map.
1: **Begin.**
2: Pixel level MRF classification. Computing the initial classification result $P(x)$ and further obtain the pixel level classification result by using the pixel level MRF model.
3: Superpixel map generation by using ERS algorithm. Superpixels are denoted as $S = \{S_1, S_2, \ldots, S_M\}$ with M superpixels.
4: Superpixel level MRF classification. Using (10) to get superpixel class probability. Using (11) (12) and (13) to calculate the adjacent superpixels' similarity. Using (8) to model superpixel level classification.
5: Using α-expansion algorithm to obtain final classification result.
6: **End.**

IV. EXPERIMENTAL RESULTS

In this section, to evaluate the classification performance of BLC-MRF, we apply proposed method over three HSR image datasets, including two HSR multispectral images[1] and one HSR hyperspectral image. The three test images are described as follows.

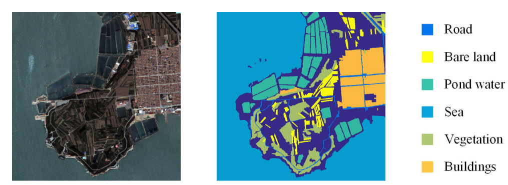
The first experimental image was collected by QuickBird2 sensor in 2012 from Qingdao area in Shandong Province, China. The image has four spectral bands and a spatial resolution of 2.4 m with a size of 400 × 400 pixels. The ground truth contains six land cover classes and a total of 91 514 labeled pixels with 720 training samples, as shown in Table I. Fig. 4 gives an overview of this image and the corresponding ground truth map.

The second experiment dataset was also collected by QuickBird2 sensor in 2012 over Jiaozhou Bay in Shandong Province, China. This image has the same basic information as first dataset. The size of the image is 800 × 800 with 469 001 labeled pixels. Fig. 5 presents the false-color image, as well as the corresponding ground truth map. Six land cover types are considered, as detailed in Table II.



(a) (b)

Fig. 4. Qingdao area dataset. (a) RGB false-color image. (b) Ground truth map.



(a) (b)

Fig. 5. Jiaozhou Bay dataset. (a) RGB false-color image. (b) Ground truth map.

The third HSR image Salinas Valley was collected by AVRIS sensor over Salinas Valley, California with a spatial resolution of 3.7 m by pixel. This image has 224 spectral bands with a size of 512 × 217 pixels. The number of bands is reduced to 200 after 24 water absorption bands are removed. As detailed in Table III, this dataset contains 54 129 labeled pixels and 720 training samples in 16 land cover classes. The overview image and ground truth are displayed in Fig. 6.

The purpose of the experiment is to evaluate the classification performance of BLC-MRF over three HSR datasets in comparison with other classifiers. The classifiers used are described as follows.

1) *pixel level classifier:* SVMCK classifier [6];
2) *filtering-based classifiers:* LBP with decision fusion classifier (LBP-DF) [14], multiple-feature-based adaptive spare representation (MFASR) [55];
3) *object-based classifiers:* SVMMRF with object-oriented voting (SVM-OO) [39];
4) *MRF-based classifiers:* SVM classifier with MRF (SVMMRF)[34], superpixel level MRF classifier (SuperMRF) [29], sparse MLR with weighted MRF

[1] https://www.researchgate.net/profile/Liang_Xiao10/publications.

TABLE II
CLASSIFICATION ACCURACY (%), OA (%), AND KAPPA COEFFICIENT FOR JIAOZHOU BAY

class	Train	Test	SVMCK [6]	LBP-DF [14]	MFASR [55]	SVM-OO [39]	SVMMRF [34]	SuperMRF [29]	SMLR-WMRF [31]	DPSCRF [41]	CNN [17]	SSRN [18]	BLC-MRF
1	11642	250	59.54	71.72	34.86	83.22	66.89	60.57	81.06	81.20	88.52	91.19	84.78
2	23517	250	89.78	84.43	62.66	91.28	79.25	92.85	83.07	91.26	95.36	94.52	91.05
3	52021	250	94.97	93.35	93.93	96.82	96.72	94.87	97.61	97.32	94.81	98.33	97.48
4	283193	250	98.71	99.20	99.91	99.69	98.71	99.38	99.66	99.81	99.51	99.52	99.78
5	44250	250	92.10	76.35	98.12	96.79	88.74	98.13	96.07	92.91	91.23	93.61	96.90
6	52878	250	75.60	84.50	93.25	78.18	71.20	87.06	84.85	87.40	77.27	84.77	85.54
OA			93.63	93.30	94.78	95.83	92.67	96.07	96.10	96.57	95.19	96.69	**96.83**
κ			0.894	0.888	0.912	0.930	0.878	0.934	0.935	0.943	0.920	0.944	**0.947**

The best results of OA and Kappa are highlighted in bold.

TABLE III
CLASSIFICATION ACCURACY (%), OA (%), AND KAPPA COEFFICIENT FOR SALINAS VALLEY

class	Train	Test	SVMCK [6]	LBP-DF [14]	MFASR [55]	SVM-OO [39]	SVMMRF [34]	SuperMRF [29]	SMLR-WMRF [31]	DPSCRF [41]	CNN [17]	SSRN [18]	BLC-MRF
1	1979	30	99.13	99.86	100.00	100.00	98.64	98.18	100.00	100.00	100.00	100.00	100.00
2	3696	30	98.75	99.84	99.57	100.00	99.25	95.70	100.00	100.00	99.97	100.00	100.00
3	1946	30	96.80	99.30	99.08	100.00	99.84	100.00	100.00	100.00	100.00	100.00	100.00
4	1364	30	99.09	98.94	99.86	99.93	99.33	99.56	99.93	99.93	100.00	99.85	99.92
5	2648	30	97.81	96.18	98.72	99.28	98.03	94.07	99.28	99.28	100.00	100.00	99.23
6	3929	30	99.25	95.97	99.72	99.95	99.09	95.78	99.95	99.95	100.00	100.00	99.95
7	3949	30	99.10	97.37	99.61	99.58	98.69	96.68	99.58	99.58	100.00	100.00	99.54
8	11241	30	73.18	93.29	89.11	90.24	72.39	96.36	89.66	91.00	85.70	94.12	99.38
9	6173	30	97.56	99.25	99.82	99.97	98.78	98.90	99.87	99.97	99.97	100.00	100.00
10	3248	30	89.68	96.99	96.25	91.82	85.63	99.20	98.59	98.68	99.03	98.58	97.80
11	1038	30	98.31	99.45	97.83	100.00	95.30	98.65	100.00	100.00	99.90	100.00	96.73
12	1897	30	99.82	96.54	100.00	100.00	99.96	94.94	100.00	100.00	100.00	100.00	100.00
13	886	30	98.77	96.56	95.93	97.86	99.05	99.44	98.20	98.10	100.00	100.00	97.82
14	1040	30	96.66	98.85	96.32	98.69	93.04	98.94	98.66	97.62	100.00	100.00	98.57
15	7238	30	73.16	95.88	85.30	94.28	71.93	93.99	97.13	98.87	92.22	95.41	99.42
16	1777	30	97.72	100.00	89.29	100.00	97.93	100.00	100.00	100.00	99.83	100.00	99.07
OA			89.17	96.83	94.79	96.56	88.73	96.82	97.22	97.74	95.86	98.13	**99.43**
κ			0.880	0.965	0.942	0.962	0.875	0.965	0.969	0.975	0.9539	0.978	**0.993**

The best results of OA and Kappa are highlighted in bold.

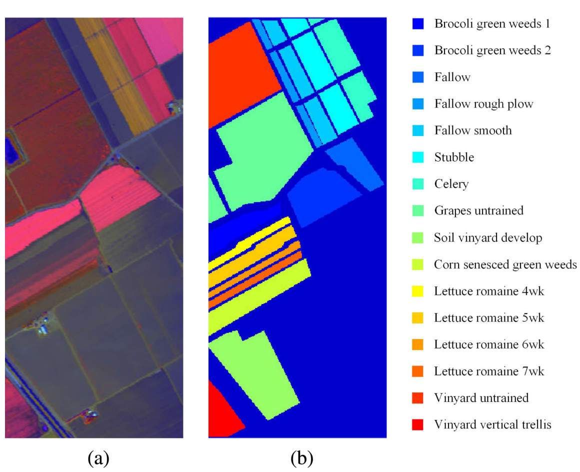


Fig. 6. Salinas Valley dataset. (a) RGB false-color image. (b) Ground truth map.

(SMLR-WMRF) [31], detail preserving smoothing classifier based on CRF (DPSCRF) [41];

5) *deep learning methods:* CNN [17] and SSRN[18].

The CNN model contains three convolutional layers with a $5 \times 5 \times 128$ convolutional kernels. The SSRN contains five convolutional layers with $1 \times 1 \times 24$ and $3 \times 3 \times 128$ convolutional kernels for spectral and spatial domain, respectively. In CNN and SSRN, the pixel samples have the size of $7 \times 7 \times B$ for training and testing. The classification results are obtained after 200 epochs. All relevant parameters of these algorithms are set based on the reference papers.

A. Experimental Setup

In our proposed BLC-MRF, there are three vital parameters, including tuning parameter α and β, and superpixel number M. In our experiments, α, β, and *M* are selected according to cross validation. The parameter α in (2) controls the strength of pixel level spatial contextual information and β controls the superpixel level contextual strength. M controls the number of superpixels of each HSR image after segmentation. The selection of the three parameters will be discussed in next experiments. In pixel level, the SVMMRF is applied for initial classification. The smooth parameter θ_a on three datasets are set to 1.2, 1.5, and 0.9, respectively. The bin number T is set to 8. All the experiments are conducted on a computer with 3.4 GHz CPU and 16 GB random access memory.

The classification results are assessed by the overall accuracy (OA), and kappa coefficient (κ). OA is computed by the ratio between the correctly classified test pixels and the total number of test pixels. The κ coefficient is computed by weighting

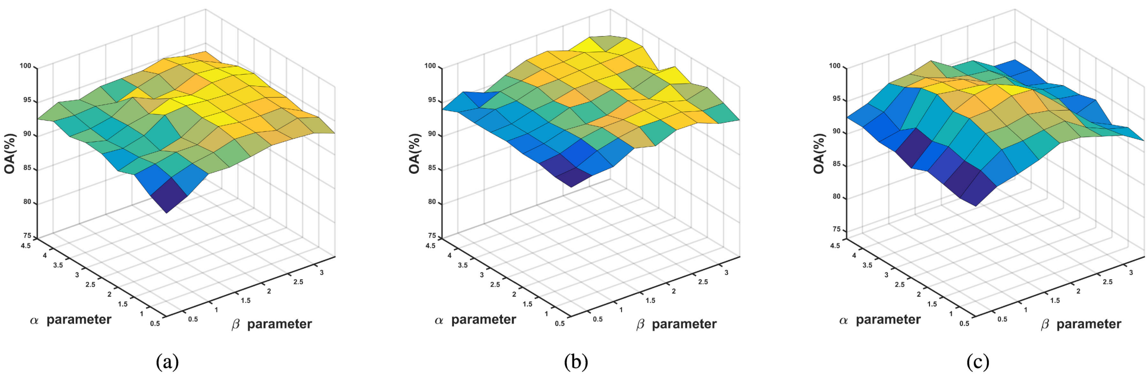


Fig. 7. Sensitive analysis of parameter α and β for proposed method. (a) Qingdao area: 120 training pixels per class. (b) Jiaozhou Bay: 250 training pixels per class. (c) Salinas: 30 training pixels per class.

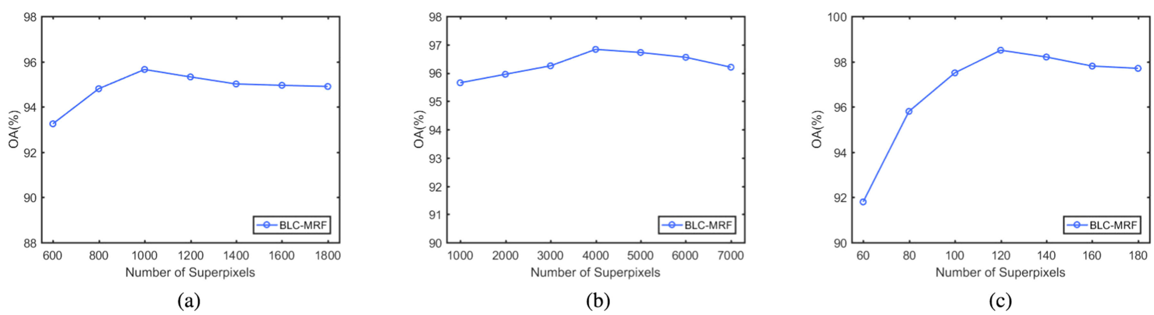


Fig. 8. Classification accuracy from different number of superpixels. (a) Qingdao area. (b) Jiaozhou Bay. (c) Salinas Valley.

the measurement accuracies. The F index [56] is a summary measurement that reflects the correspondence between boundary map. In our experiment, F index is in a range of [0,1]. A higher F index value indicates a better correspondence map with the ground truth map. To ensure the fairness of experiments, every experiment is conducted ten times with different training samples.

B. Sensitivity Analysis of Parameters

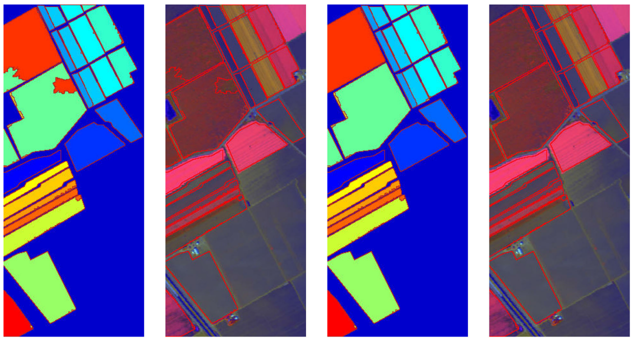
To assess the effect of the tuning parameters α and β of the proposed method, a sensitivity analysis of parameter α and β is conducted on three HSR images, as shown in Fig. 7.

As displayed in Fig. 7, the sensitivity analysis is conducted with parameters α and β varying in the preset range. Here, α is chosen from the range of [0.5, 1, 1.5, . . . , 4.5] and β is varying from 0.2 to 3.4. Two observations can be made from Fig. 7. The first is that the OA of Salinas Valley changes significantly when parameters α and β are varying. In contrast, the OA of Qingdao area and Jiaozhou Bay changes slowly. This is mainly because the Salinas Valley dataset has complicated land cover types and a relatively small number of superpixels, making Salinas Valley sensitive to parameter changes. The second observation is that the OA value changes significantly when parameter β is varying, meanwhile, OA varies slowly when α changes. This is because that the superpixel level MRF model plays an important role in the whole algorithm and the output of pixel level is a basic part in the superpixel level. In this experiment, according to Fig. 7, to achieve the optimal classification performance, α and β are, respectively, set to 3 and 2.6 for Qingdao area, 2.5 and 3.4 for Jiaozhou Bay and 2.5 and 1.4 for Salinas Valley.

The influence of the number of superpixels *M* on classification accuracy is also investigated. As shown in Fig. 8, we can see the OA values obtained from BLC-MRF on three HSR datasets. It is obvious that the classification accuracies have a climbing trend when the number of superpixel is increasing. After reaching the highest point, the OA starts to drop slowly. This phenomenon is more evident on Salinas Valley dataset. From Fig. 8, it can be seen that the proposed method achieves the satisfactory classification performance when the superpixel number is around 1000 for Qingdao area, 4000 for Jiaozhou Bay and 120 for Salinas Valley, respectively.

C. Analysis of the Integration of Superpixel Level MRF

In this section, to validate the necessity of integration superpixel level MRF, we first compare the results of BLC-MRF with and without the superpixel level MRF processing. In the BLC-MRF without superpixel MRF circumstance, the superpixels are classified only with the probability estimation. Fig. 9 shows the outputs of BLC-MRF with and without superpixel level MRF. The visual improvements in the classification performance can be seen in Fig. 9. Through the integration of superpixel level MRF, some misclassified superpixels can be corrected. Meanwhile, the object boundaries can be preserved as superpixels can help suppress outliers in the boundary areas.



(a) (b) (c) (d)

Fig. 9. Comparison with the results of BLC-MRF with and without superpixel level MRF on Salinas Valley dataset. The red lines draw the object boundaries of different land cover areas. (a) Classification result of BLC-MRF without superpixel level MRF. (b) Boundary map of BLC-MRF without superpixel level MRF on RGB false-color image. (c) Classification result of BLC-MRF. (d) Boundary map of BLC-MRF on RGB false-color image.

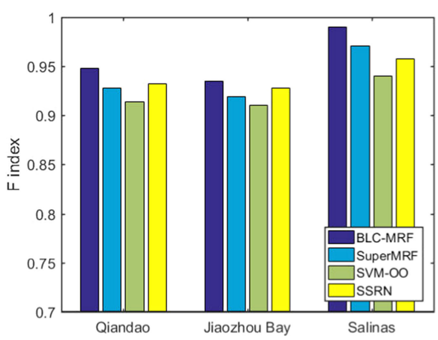


Fig. 10. F index on three HSR datasets.

To further validate the preserving boundaries ability of superpixel level MRF, we compare the output of the proposed BLC-MRF and some superpixel integrated methods (SVM-OO, SuperMRF) and SSRN. We calculate the F index on three HSR images, as shown in Fig. 10. As we can see, the proposed method exhibits the best performance with respect to the considered F index. In Salinas dataset, the BLC-MRF achieves the highest F index value (up to 0.99), indicating that the object boundaries of Salinas Valley is well preserved. Through the integration of superpixel level MRF, we can obtain more accurate classification performance and complete objects.

Apart from the boundary preserving ability, we also compare the energy minimization process of superpixel levevl MRF in BLC-MRF with the SuperMRF method [29]. In SuperMRF, the unary energy term is constructed with majority voting strategy and the pairwise term is constructed by measuring the difference of mean spectrum of superpixels. The α-expansion optimization method is also applied in SuperMRF. Therefore, it is reasonable to record and compare the energy minimization process of both superpixel level MRF models. Figs. 11 and 12 display the energy minimization curves of both superpixel level MRF methods. As we can see, the proposed BLC-MRF achieves a lower energy cost, indicating that the constructed energy terms in BLC-MRF are suitable for improving the HSR image classification performance.

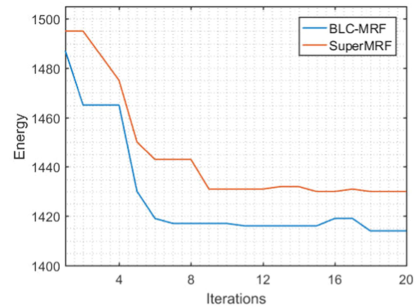
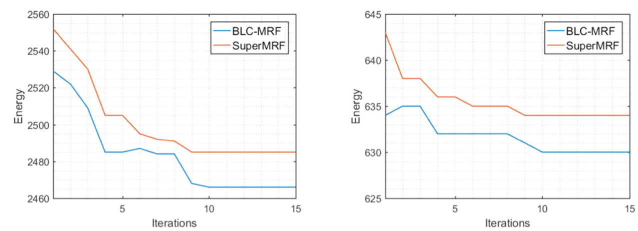


Fig. 11. Energy minimization curve of superpixel level MRF of BLC-MRF and superMRF over Qingdao area.



(a) (b)

Fig. 12. Energy minimization curve of superpixel level MRF of BLC-MRF and SuperMRF over Jiaozhou Bay and Salinas Valley.

D. Classification Results

Fig. 13 depicts the OA results as a function of the number of training pixels on three datasets. For comparison, we also show the OA of some other classifiers in experiment. From Fig. 13, we can find that OA of most classification methods shows a climbing trend with the training samples increasing. For the classifiers integrated MRF models (SuperMRF, SMLR-WMRF, DSPMRF, and proposed method), the OA is higher than other classifiers in most cases, indicating the importance of integrating spectral–spatial information.

In this experiment, we also compare the classification performance of BLC-MRF with two deep learning based methods (CNN and SSRN). As shown in Fig. 13, both methods (CNN and SSRN) show a significant increase in OA with the number of training samples increasing. However, as a large number of neurons need to be trained, a deep neural network is usually difficult to construct in the case of a very small-size of training samples [16]. In contrast, owing to the property of superpixels and MRF, the classification mistakes can be suppressed when few samples are available. Therefore, from Fig. 13, it can be seen that the OA of CNN and SSRN is slightly lower than BLC-MRF when the training pixels are very few (only about 0.3% to 0.8%). On the other hand, with the number of training samples increasing, the deep learning based methods start to show their efficiency. In Fig. 13(a) and (b), we can observe that the OA of SSRN begins to exceed BLC-MRF when the number of training samples is larger than 180 and 300. In general, our proposed method achieves fine classification accuracy on three data sets, which indicates the effectiveness of our proposed bilevel MRF framework.

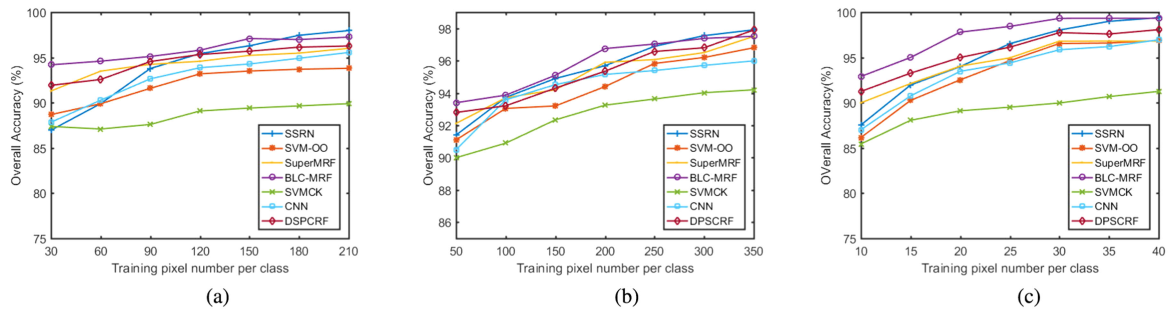


Fig. 13. Impact of the number of training samples. (a) Qingdao area. (b) Jiaozhou Bay. (c) Salinas Valley.

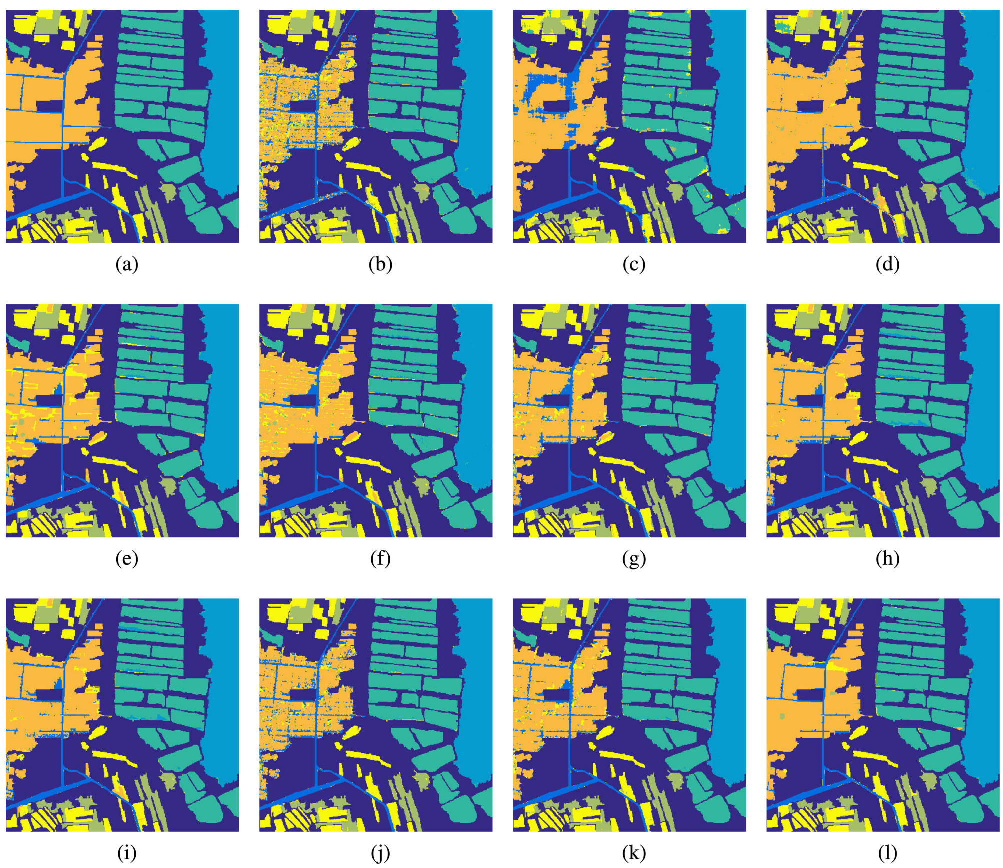


Fig. 14. Classification results on Qingdao area dataset. (a) Ground truth. (b) SVMCK. (c) LBP-DF. (d) MFASR. (e) SVM-OO. (f) SVMMRF. (g) SuperMRF. (h) SMLR-WMRF. (i) DPSCRF. (j) CNN. (k) SSRN. (l) Proposed BLC-MRF.

Tables I to III show the classification accuracy per class, OA and kappa coefficient using different classification methods. To evaluate the classification performance on a small number of training samples, the number of training data on three datasets are set to 120, 260, and 30 per class, respectively. The proportion of training sample on three datasets are 0.78%, 0.32%, and 0.88%. From Table I to III, it can be seen that classifiers integrating spatial contextual information show a great improvement over traditional pixelwise classifiers in classification accuracy. In particular, methods that integrate superpixel processing (SVM-OO, SuperMRF, and BLC-MRF) obtain better OA than traditional classifiers, indicating that superpixel generation has a positive effect on HSR image classification. It can also be observed that most methods obtain promising accuracies (up to 100%) on Salinas Valley dataset even though the training samples is only 30 per class. In contrast, the accuracies of methods on the other two datasets are close and lower than that on Salinas Valley. Finally, it can be seen that the proposed method achieves the highest accuracy.

Figs. 14–16 illustrate the reference map and the classification maps of the classifiers listed in Tables I–III. The number of training samples is the same as Tables I–III. From the classification maps, we can observe that the pixelwise classifier generated maps (SVMCK) are strongly affected by spectral noise, which leads to a poor classification performance. For the filtering-based classifiers (MFASR and LBP-DF), the

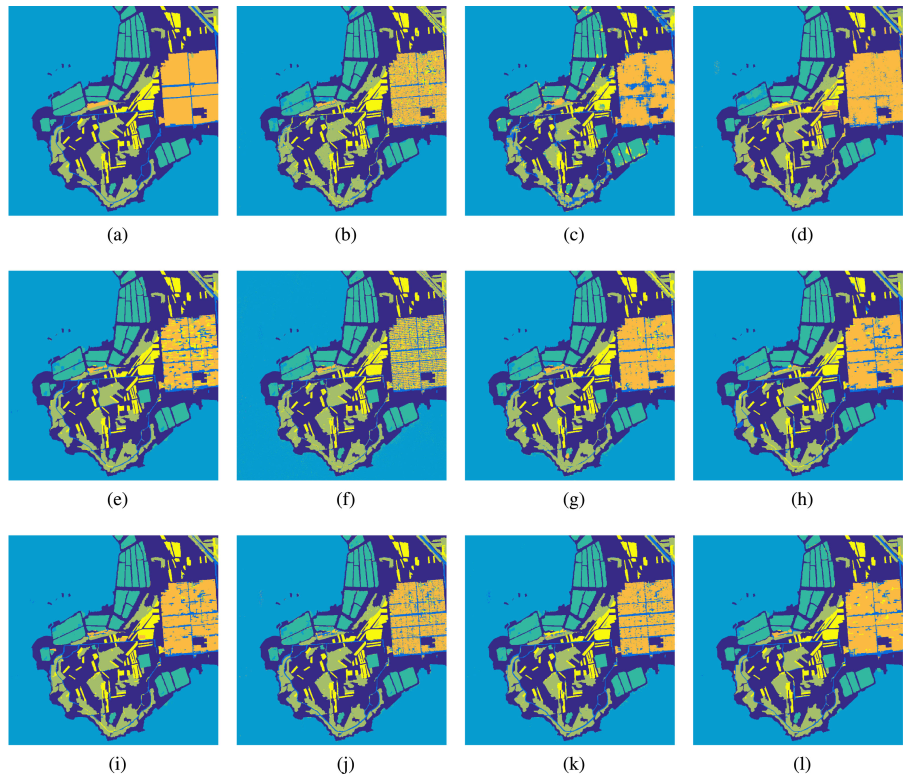


Fig. 15. Classification results on Jiaozhou Bay dataset. (a) Ground truth. (b) SVMCK. (c) LBP-DF. (d) MFASR. (e) SVM-OO. (f) SVMMRF. (g) SuperMRF. (h) SMLR-WMRF. (i) DPSCRF. (j) CNN. (k) SSRN. (l) Proposed BLC-MRF.

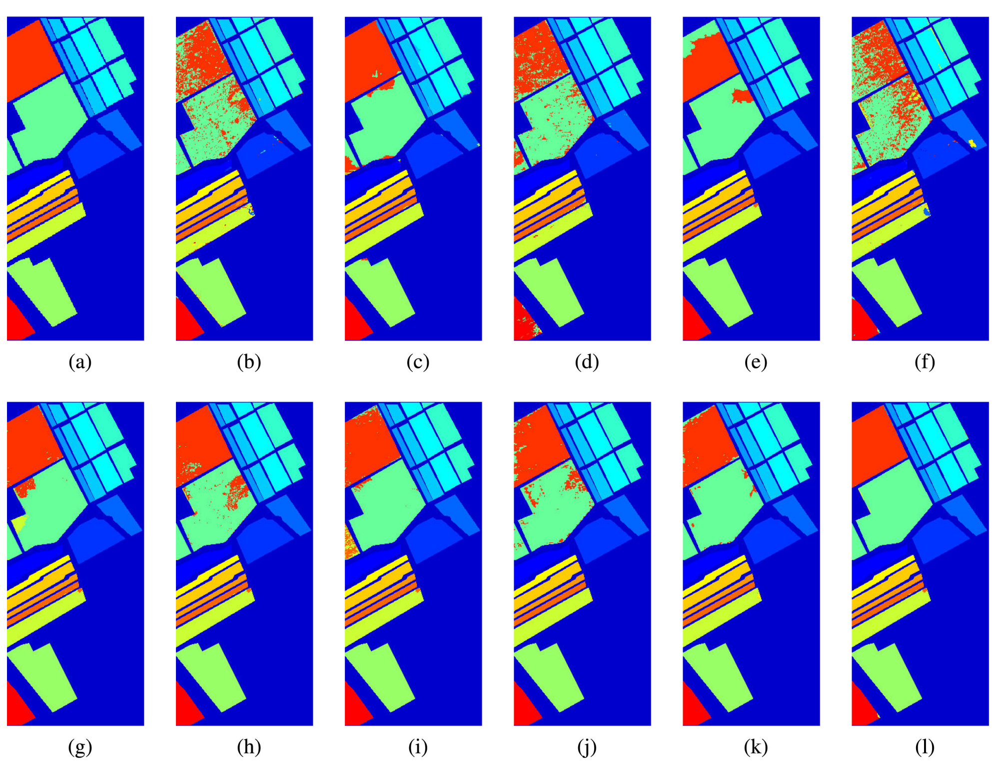


Fig. 16. Classification results on Salinas Valley dataset. a) Ground truth. (b) SVMCK. (c) LBP-DF. (d) MFASR. (e) SVM-OO. (f) SVMMRF. (g) SuperMRF. (h) SMLR-WMRF. (i) DPSCRF. (j) CNN. (k) SSRN. (l) Proposed BLC-MRF.

classification maps are improved significantly. For the classifiers integrated MRF models (SuperMRF, SMLR-WMRF, DSPCRF, and proposed method), these methods have a good effect on edges while the classification performance is poor at small regions. CNN and SSRN use the deep convolutional network to combine spectral and spatial information and show a good performance on HSR image classification, Moreover, it is obvious that the obtained map of the proposed method achieves best classification performance. It is worth mentioning that the proportion of training samples of Jiaozhou Bay dataset is the smallest (around 0.3%). Even though the number of training samples is small, the classification performance of BLC-MRF is still satisfactory.

V. Conclusion

In this article, a novel bilevel contextual MRF classification framework, named BLC-MRF, is proposed for HSR image classification. In BLC-MRF, two kinds of contextual cues, including pixel and superpixel level contextual information, are incorporated into the MRF model. In the pixel level, SVMMRF is applied to obtain initial pixel level classification result. In the superpixel level, ERS algorithm is first used for superpixel map generation. Then, two potential terms are constructed for the contextual MRF model. On the one hand, an improved superpixel probability estimation is proposed for unary potential computation. On the other hand, we propose to utilize superpixel spectral histogram distance to establish a new pairwise potential term. The final classification result is achieved by an MRF optimization algorithm. By exploiting the pixel and superpixel level contextual information, noise and outliers are suppressed and object boundaries are preserved; thus, the positive impact of BLC-MRF framework can be ensured.

We have illustrated the superiority of our proposed method on three HSR datasets in comparison with some other classification approaches. The experimental results demonstrate that our proposed method also provides promising classification performance with a small number of training samples. In the future work, we will focus on integrating superpixel and machine learning models to further exploit HSR image feature for classification.

References

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

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

[4] B. Krishnapuram, L. Carin, M. A. T. Figueiredo, and A. J. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 957–968, Jun. 2005.

[5] M. Pal, A. E. Maxwell, and T. A. Warner, "Kernel-based extreme learning machine for remote-sensing image classification," *Remote Sens. Lett.*, vol. 4, no. 9, pp. 853–862, Jun. 2013.

[6] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[7] Y. Zhou, J. Peng, and C. L. P. Chen, "Extreme learning machine with composite kernels for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2351–2360, Jun. 2015.

[8] X. Huang, Q. Lu, L. Zhang, and A. Plaza, "New postprocessing methods for remote sensing image classification: A systematic study," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7140–7159, Nov. 2014.

[9] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018.

[10] L. He, J. Li, A. Plaza, and Y. Li, "Discriminative low-rank gabor filtering for spectral-spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1381–1395, Mar. 2017.

[11] W. Li and Q. Du, "Gabor-filtering-based nearest regularized subspace for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1012–1022, Apr. 2014.

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

[13] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.

[14] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.

[15] S. Jia, B. Deng, J. Zhu, X. Jia, and Q. Li, "Local binary pattern-based hyperspectral image classification with superpixel guidance," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 749–759, Feb. 2018.

[16] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[17] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, Jan. 2017.

[18] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.

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

[20] J. Guo, X. Zhou, J. Li, A. Plaza, and S. Prasad, "Superpixel-based active learning and online feature importance learning for hyperspectral image analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 1, pp. 347–359, Jan. 2017.

[21] O. Csillik, "Fast segmentation and classification of very high resolution remote sensing data using SLIC superpixels," *Remote Sens.*, vol. 9, no. 3, p. 243, Mar. 2017.

[22] 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.

[23] 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.

[24] B. Cui, X. Xie, X. Ma, G. Ren, and Y. Ma, "Superpixel-based extended random walker for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3233–3243, Jun. 2018.

[25] J. Li, H. Zhang, and L. Zhang, "Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5338–5351, Oct. 2015.

[26] S. Li, T. Lu, L. Fang, X. Jia, and J. A. Benediktsson, "Probabilistic fusion of pixel-level and superpixel-level hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7416–7430, Dec. 2016.

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

[28] M. Golipour, H. Ghassemian, and F. Mirzapour, "Integrating hierarchical segmentation maps with MRF prior for classification of hyperspectral images in a Bayesian framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 805–816, Feb. 2016.

[29] U. B. Gewali and S. T. Monteiro, "A tutorial on modelling and inference in undirected graphical models for hyperspectral image analysis," *Int. J. Remote Sens.*, vol. 39, no. 20, pp. 7104–7143, Oct. 2018.

[30] 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.

[31] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral-spatial hyperspectral image classification with weighted Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1490–1503, Mar. 2015.

[32] P. Chen, J. D. B. Nelson, and J.-Y. Tourneret, "Toward a sparse Bayesian Markov random field approach to hyperspectral unmixing and classification," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 426–438, Jan. 2017.

[33] I. Hedhli, G. Moser, S. B. Serpico, and J. Zerubia, "Classification of multisensor and multiresolution remote sensing images through hierarchical Markov random fields," *IEEE Geosci. Remote Sens. Lett.*, pp. 1–5, 2017.

[34] 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.

[35] X. Cao, L. Xu, D. Meng, Q. Zhao, and Z. Xu, "Integration of 3-dimensional discrete wavelet transform and Markov random field for hyperspectral image classification," *Neurocomputing*, vol. 226, pp. 90–100, Feb. 2017.

[36] X. Zhang, Z. Gao, L. Jiao, and H. Zhou, "Multifeature hyperspectral image classification with local and nonlocal spatial information via Markov random field in semantic space," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1409–1424, Mar. 2018.

[37] L. Wang, X. Huang, C. Zheng, and Y. Zhang, "A Markov random field integrating spectral dissimilarity and class co-occurrence dependency for remote sensing image classification optimization," *ISPRS J. Photogramm. Remote Sens.*, vol. 128, pp. 223–239, Jun. 2017.

[38] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

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

[40] T. Zhang, W. Yan, J. Li, and J. Chen, "Multiclass labeling of very high resolution remote sensing imagery by enforcing nonlocal shared constraints in multilevel conditional random fields model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 2854–2867, Jul. 2016.

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

[42] D. Tuia, M. Volpi, and G. Moser, "Decision fusion with multiple spatial supports by conditional random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3277–3289, Jun. 2018.

[43] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.

[44] A. H. S. 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.

[45] M. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Colorado Springs, CO, USA, Jun. 2011, pp. 2097–2104.

[46] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[47] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions-I," *Math. Program.*, vol. 14, no. 1, pp. 265–294, Dec. 1978.

[48] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Interactive image segmentation by maximal similarity based region merging," *Pattern Recognit.*, vol. 43, no. 2, pp. 445–456, Feb. 2010.

[49] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. COM-15, no. 1, pp. 52–60, Feb. 1967.

[50] R. C. Dubes and A. K. Jain, "Random field models in image analysis," *J. Appl. Statist.*, vol. 20, no. 5/6, pp. 121–154, Apr. 1993.

[51] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, vol. 48, no. 3, pp. 259–302, 1986.

[52] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th Int. Symp. Micro Mach. Hum. Sci.*, Nagoya, Japan, Oct. 1995, pp. 39–43.

[53] 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.

[54] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[55] L. Fang, C. Wang, S. Li, and J. A. Benediktsson, "Hyperspectral image classification via multiple-feature-based adaptive sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1646–1657, Jul. 2017.

[56] J. Pont-Tuset and F. Marques, "Supervised evaluation of image segmentation and object proposal techniques," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1465–1478, Jul. 2016.

Yu Shen received the B.E. degree in computer science and technology in 2015 from the Nanjing University of Science and Technology (NJUST), Nanjing, China, where he is currently working toward the Ph.D. degree in computer science with the School of Computer Science and Engineering.

His research interests include the areas of remote sensing image processing, computer vision, and machine learning.

Jianyu Chen (M'12) received the M.Sc. degree in environmental science and the Ph.D. degree in earth science from Zhejiang University, Hangzhou, China, in 1998 and 2004, respectively.

He is currently a Professor of remote sensing with the State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China. He has been a Postdoctoral Researcher with the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China. He has also been a Visiting Professor with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada.

His research interests include object-based image analysis, remote sensing of coastal environment, and geographic information system.

Liang Xiao (M'11) received the B.S. degree in applied mathematics and the Ph.D. degree in computer science from the Nanjing University of Science and Technology (NJUST), Nanjing, China, in 1999 and 2004, respectively. From 2006 to 2008, he was a Postdoctoral Research Fellow with Pattern Recognition Laboratory, NJUST. From 2009 to 2010, he was a Postdoctoral Fellow with Rensselaer Polytechnic Institute, Troy, NY, USA. Since 2013, he has been the Deputy Director of Jiangsu Key Laboratory of Spectral Imaging Intelligent Perception, Jiangsu, China. Since 2014, he has been the second Director of the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, NJUST, where he is currently a Professor with School of Computer Science and Engineering. His fields of interest include remote sensing image processing, hyperspectral remote sensing, image modeling and computer vision, machine learning, and pattern recognition.

Delu Pan received the B.S. degree in physics from the Nanjing University of Science and Technology (NJUST), Nanjing, China, in 1968.

He is currently a Senior Scientist (expert of marine remote sensing) with the Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, China. He is also an Academician of the Chinese Academy of Engineering, Beijing, China, and a Professor with Zhejiang University, Hangzhou, China.

Mr. Pan is the Vice President of the Chinese Society of Oceanography and has been a member of the International Ocean Color Coordinating Group since 2000. He is the Editor-in-Chief of *Acta Oceanologica Sinica* (a journal of the Chinese Society of Oceanography).