ISPRS Journal of Photogrammetry and Remote Sensing 128 (2017) 223–239

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

A Markov random field integrating spectral dissimilarity and class co-occurrence dependency for remote sensing image classification optimization

Leiguang Wang [a,d], Xin Huang [b,*], Chen Zheng [c,d], Yun Zhang [d]

[a] Southwest Forestry University, School of Forestry, China
[b] Wuhan University, School of Remote Sensing and Information Engineering, China
[c] Henan University, School of Mathematics and Statistics, China
[d] University of New Brunswick, Department of Geodesy and Geomatics Engineering, Canada

ARTICLE INFO

Article history:
Received 26 August 2016
Received in revised form 17 March 2017
Accepted 29 March 2017
Available online 18 April 2017

Keywords:
Remote sensing image classification
Spatial regularization
Markov Random Fields (MRFs)
Spatial energy function
Spectral dissimilarity
Class co-occurrence dependency

ABSTRACT

This paper develops a novel Markov Random Field (MRF) model for edge-preserving spatial regularization of classification maps. MRF methods based on the uniform smoothness lead to oversmoothed solutions. In contrast, MRF methods which take care of local spectral or gradient discontinuities, lead to unexpected object particles around boundaries. To solve these key problems, our developed MRF method first establishes a spatial energy function integrating local spectral dissimilarity to smooth the initial classification map while preserving object boundaries. Second, a new anisotropic spatial energy function integrating the class co-occurrence dependency is constructed to regularize pixels around object boundaries. The effectiveness of the method is tested using a series of remote sensing data sets. The obtained results indicate that the method can avoid oversmoothing and significantly improve the classification accuracy with regards to traditional MRF classification models and some other state-of-the-art methods.

1. Introduction

Remote sensing image classification, which aims to classify a remotely sensed image into a thematic map, is a very active research field. As more and more images with higher spatial resolution became available, advanced classification methods utilize not only the spectral, but also the spatial properties in order to improve classification accuracy. In this context, a large collection of spectral-spatial classification methods (Benediktsson et al., 2005; Blaschke, 2010; Chen et al., 2016; Huang et al., 2014; Li et al., 2013; Zhang and Jia, 2012; Zhang et al., 2006) have been proposed.

Representative spatial features include pixel shape index (Zhang et al., 2006), extended morphological profiles (Benediktsson et al., 2005) and extended morphological attribute profiles (Dalla Mura et al., 2011), among many others (Cheng and Han, 2016; Fauvel et al., 2013). Another group of spatial-spectral methods is known as object-based image analysis (OBIA) (Blaschke, 2010; Walter, 2004; Zheng et al., 2013), which utilizes segments as basic units for extracting features. OBIA can suppress the salt-and pepper noise that is often observed from pixelwise classification results. Besides these hand-crafted spatial features, features learnt automatically from input images, which are known as deep learning-based methods (Chen et al., 2016, 2014; Hinton and Salakhutdinov, 2006), became popular recently. These methods, typically convolutional neural network (Chen et al., 2016), by simulating the processing of the primate visual system through a deep hierarchy, can extract a series of low- and high-level features. Both hand-crafted and learnt features make efforts on integrating spatial information at feature extraction stage or during the classification stage.

Recently, a number of works have developed strategies to integrate the spatial information at the postprocessing stage, such as relearning (Huang et al., 2014), object-based method (Büschenfeld and Ostermann, 2012), filtering based method (Kang et al., 2014), and Markov random fields (MRFs) (Aghighi et al., 2014; Schindler, 2012; Tarabalka et al., 2010). These methods generally rely on the common assumption that neighboring pixels tend to belong to the same class.

Filtering methods impose a kernel on an initial label image using a sliding window, and then assign each center pixel to an

* Corresponding author.
E-mail address: huang_whu@163.com (X. Huang).

http://dx.doi.org/10.1016/j.isprsjprs.2017.03.020
0924-2716/

output value obtained according to the existing gray and/or label values in the window. For instance, Kang et al. (2014) proposed to use the guided filter to achieve edge-preserving smoothing of the probabilistic map of an initial label image. Unfortunately, the overall high accuracies are essentially achieved at the cost of oversmoothed objected boundaries. Object-based methods (Büschenfeld and Ostermann, 2012) conduct (weighted) majority voting over each image object to determine the resulting class to be assigned to the object. During the voting process, the initial probabilities of labels and the distance from the current pixel to the corresponding object border are considered. However, the effectiveness of object-based voting is also influenced by the performance of segmentation algorithms. More recently, Huang et al. (2014) have presented a new concept of relearning to smooth the initial classification result. The goal is achieved by iteratively updating the initial result according to the frequency and spatial arrangement of the class labels.

Some MRFs (Aghighi et al., 2014; Li et al., 2012; Moser et al., 2013; Tarabalka et al., 2010) can also be seen as postprocessing methods, as they utilize neighbor label information to produce smoothing effects on initial classification results. Under the MAP-MRF framework, this optimization is formulated as the minimization of the class posterior probability, which is equivalent to minimizing an energy function comprising the feature and label models (Li, 2009). The feature model is related to features used in the classification, and is often initialized by the output of a spectral feature-based pixelwise classification (Schindler, 2012; Tarabalka et al., 2010). Meanwhile, the label model is related to the spatial prior of classes, which is formulated as a MRF (Moser et al., 2013). In contrast, without modeling the feature and label models individually, alternative MRF methods can directly express the class posterior probability as a MRF (Li, 2009; Zhang and Jia, 2012). This group of MRFs is also named as conditional random fields (CRFs).

Therefore, both groups belong to the random field model assumed to exhibit the Markov property, and use the probability function to model the spatial interactions between image sites. Both of them assumes that the class labels and/or feature levels in a neighborhood of the image lattice do not change abruptly. It has been demonstrated that, even with this simple spatial prior, MRF methods can perform quite well in terms of improving classification accuracy (Huang et al., 2014; Schindler, 2012; Zhong and Wang, 2010). However, this generic smoothness prior also lead to oversmoothed solutions when effective, i.e., the class boundaries do not align with real object boundaries (Schindler, 2012; Tarabalka et al., 2010; Zhang and Jia, 2012). The main reason is that the uniform smoothness assumption is often violated at the image boundaries, where abrupt changes of pixel values occur. Therefore, several works have established more complex spatial *a priori* models involving local discontinuities, such as the derivative magnitude (Tarabalka et al., 2010; Yu and Clausi, 2008) or the spectral difference (Moser et al., 2013). These models aim to suppress the smoothness effect when the value of the term becomes larger (often with high probability exactly corresponding to, or near real boundaries). In this way, the models can effectively preserve edges.

Unfortunately, as shown in Figs. 8(a) and 12 (g), these state-of-the-art MRF models involving local discontinuities, still suffer from unexpected and isolated class labels around object boundaries, where a salt-and-pepper noise effect can be appreciated. These pixels located around boundaries have distinct spectral presentations with surrounding pixels. According to the MRF models, with a high possibility, they are labelled as the class with the most similar spectral property, rather than the spatial neighboring classes. Whereas, in order to properly consider spatial dependency among different land classes and obtain better visual inspection, we have strong motivations to divide these pixels with different spectral properties into the surrounding classes. For instance, in Fig. 8, pixels located between different crop types are expected to be recognized as one of the adjacent types, rather than some other types with similar spectral properties but long spatial distance.

From previous literature review, we can find that most of postprocessing optimization methods perform quiet well in homogenous regions. A key aspect when utilizing optimizations is to design a proper spatial model, which can deal with the features and labels around object boundaries. An expected spatial regularization method should refine an already classified map, smooth labels in homogenous regions, meanwhile, align the boundaries among different labels with real object boundaries.

In this context, this paper encodes spatial *a priori* assumptions involving both spectral dissimilarity and class co-occurrence dependency into two spatial energy functions, and results in a new MRF method with two-step spatial regularization. The first spatial energy function integrating local spectral dissimilarity is to smooth the initial classification map while preserving object boundaries. The second spatial energy function integrating the class spatial dependency is constructed to further regularize pixels around object boundaries. It is also our main contribution in this research.

The rest of the paper is organized as follows. Section 2 presents background on MRF-based methods intended to achieve spatial smoothness. The proposed MRF method with two-step spatial regularization is presented in Section 3. Experimental results and discussions, analyzing the influence of different spectral dissimilarity metrics and a detailed parameter sensitivity assessment, are presented in Section 4. Comparisons with some other state-of-the-art methods are conducted in Section 5. Conclusions and hints at plausible future research lines are given in Section 6.

2. Background on MRF-based methods to achieve spatial smoothness

2.1. Notations and problem formulation

Let S denote the set of sites over which a remote sensing image $\boldsymbol{I}$ is defined, and let $\Omega = \{1, 2, \ldots, k\}$ denote a set of labels, being k the number of labels. Both the observation random field $\mathbf{Y}$ and the label random field $\mathbf{X}$ are defined on S. The observed image $\mathbf{y} = \{y_i | i \in S\}$ is a realization of the observation random field $\mathbf{Y}$. A label image $\mathbf{x} = \{x_i | i \in S, x_i \in \Omega\}$ is a realization of $\mathbf{X}$, in which each x_i takes a value from Ω denoting the class to which the site i belongs.

The spatial regularization task performed by a pixelwise classification map is formulated as finding an optimal estimation $\widehat{\mathrm{x}}$ that maximizes the posterior $P(\mathbf{X}|\mathrm{y})$ given the observed image y. According to the Bayesian rule and the log-linear property, finding the maximum *a posteriori* (MAP) solution $\widehat{\mathrm{x}}$ of $P(\mathbf{X}|\mathrm{y})$ (also called optimal configuration) is equivalent to minimizing the following two-part energy function (Li, 2009):

$$\widehat{\mathrm{x}} = \arg\max_x (p(\mathbf{y}|\mathbf{x})P(\mathbf{x})) = \operatorname{argmin}_x (E_f + E_l) \tag{1}$$

with

$$E_f = -\log(p(\mathbf{y}|\mathbf{x})), \tag{2}$$

and

$$E_l = -\log(P(\mathbf{x})). \tag{3}$$

With the aforementioned formulation in mind, two issues need to be addressed: (1) how to analytically represent the feature model E_f and the label model E_l, and (2) how to find a solution for the objective function (1).

It is worth noting that "*feature model*" is referred to as "*spectral model*" or "*spectral energy function*" in many works (refer to Tarabalka et al. (2010) and Aghighi et al. (2014)), since the *features* considered in most feature models are only related to the *spectral* properties of remote sensing images. Meanwhile, "*label model*" is also referred to as "*spatial context model*" or "*spatial model*", as this model deals with the spatial relationships happening in the label image. In this paper, we use these terms simultaneously to express the same meaning.

2.2. Analytical representation of feature and label models

In the feature model E_f, each feature vector $y_i \in \mathbf{y}$ is often assumed to be independent given the label x_i, thus $p(\mathbf{y}|\mathbf{x})$ turns to:

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i \in S} p(y_i|x_i) \tag{4}$$

where $p(y_i|x_i)$ can be assigned directly to the estimated class probabilities for each pixel site by the probabilistic support vector machine (SVM) (Moser et al., 2013; Tarabalka et al., 2010; Wu et al., 2004):

$$p(y_i|x_i) = l_{i,x_i} \text{ with } x_i \in \Omega, \tag{5}$$

where l_{i,x_i} denotes the probability of the pixel i belonging to a given class x_i. By substituting (4) and (5) into (2), E_f turns to:

$$E_f = -\sum_{i \in S} \log(l_{i,x_i}). \tag{6}$$

In the label model E_l, the joint probability $P(\mathbf{x})$ is assumed to exhibit the Markovianity property in the MRF model (Li, 2009), i.e.,

$$P(\mathbf{x}) = \prod_{i \in S} P(x_i|x_j, j \in N_i), \tag{7}$$

where N_i denotes the set of neighboring sites, typically defined as a second order neighboring system (8-neighborhood connectivity). Moreover, $P(x_i|x_j, j \in N_i)$ can be modeled by a pair-wise multilevel logistic model (MLL) (Derin and Elliott, 1987), which is analytically depicted as:

$$P(x_i|x_j, j \in N_i) = \frac{1}{Z_l} \exp\left(-\sum_{j \in N_i} V(x_i, x_j)\right), \tag{8}$$

with the clique potential function $V(x_i, x_j)$:

$$V(x_i, x_j) = \begin{cases} \beta & \text{if } x_i \neq x_j \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

where Z_l is a normalization factor and β, called the potential or smoothing parameter, is a positive number reflecting the pair-site interaction between two neighboring sites. Thus, by substituting (7)–(9) into (3), E_l turns to

$$E_l = \beta \sum_{i \in S} \sum_{j \in N_i} [1 - \delta(x_i, x_j)], \tag{10}$$

where $\delta(x_i, x_j)$ is the Kronecker delta function. Any violation of the smoothness condition, i.e., $\forall x_i \neq x_j$ with $j \in N_i$, incurs a positive penalty β. As the expected configurations are those with lower E_l values, the MLL model (10) favors a smooth $\mathbf{x}$. By introducing (10) and (6) into (1), the analytical objective function of MRF is thus written as:

$$\widehat{x} = \underset{\{x_i, i \in S\}}{\operatorname{argmin}} \sum_{i \in S} \left\{ -\log(l_{i,x_i}) + \beta \sum_{j \in N_i} [1 - \delta(x_i, x_j)] \right\} \tag{11}$$

According to the objective function (11), the local neighborhood composed of different classes incurs a penalty given by the positive number β. Because the more probable solutions are those with lower values, the optimal configuration favors smoothness over the whole image lattice. Therefore, this classic MRF model is quite effective in creating homogenous regions, but also leads to over-smoothed object boundaries (Schindler, 2012; Yu and Clausi, 2008). For this reason, a spatial prior model, which can better align with real object boundaries and fit object details, still needs to be defined. This is the first problem that we address in Section 3.

2.3. Smoothing parameter selection and optimization methods

The solutions of the objective functions (11) and (14) can be obtained by various combinatorial optimization techniques, such as iterated conditional mode (ICM) (Besag, 1986), belief propagation (Li, 2009), and graph cuts (Boykov and Kolmogorov, 2004; Boykov et al., 2001). ICM is a steepest local energy descent method, which usually exhibits short runtimes. However, due to the greedy nature of the method, its performance largely depends on the initial estimate (Besag, 1986). In contrast, due to the global optimization property and an acceptable polynomial time complexity with respect to the number of sites, the α-expansion graph cut (EGC) algorithm (Boykov and Kolmogorov, 2004) has drawn increasing attention in the remote sensing community.

Before inferring the labels by conducting an EGC or ICM optimization method for the objective functions, another parameter that needs to be estimated is the smoothing parameter β. This parameter controls to what extent the spatial context model is involved in the optimization process. Tarabalka et al. (2010) reported that $\beta \in [2, 4]$ is preferable for a MRF model with gradient information. In our spatial regularization task, based on the pixel-wise probability estimates for the individual classes, an automatic search process is designed to determine the parameter β. This is the second problem we would like to address in the next section.

3. Proposed method

This section is organized as follows. Firstly, we describe a group of spatial models considering a set of different spectral dissimilarity metrics. Secondly, we establish an anisotropic spatial model by taking into account the class co-occurrence dependency. Finally, we address the related smoothing parameter selection problem and summarize the whole process.

3.1. Spatial model integrating spectral dissimilarity

As discussed in Section 2, the classic MRF model tends to over-smooth near-boundary pixels. For the purpose of edge preservation, a natural extension is to integrate local statistics, such as pair-wise edge strength or spectral dissimilarity information, into the spatial model E_s. By introducing a spatial adaptive interactive parameter $f(\Delta_{i,j})$ into the potential function (9), $V(x_i, x_j)$ is updated as follows:

$$V(x_i, x_j) = \begin{cases} \beta f(\Delta_{i,j}) & \text{if } x_i \neq x_j \\ 0 & \text{otherwise} \end{cases}. \tag{12}$$

where $f(\Delta_{i,j})$ is a monotonically decreasing function with respect to $\Delta_{i,j}$ (the dissimilarity between the neighboring pixel sites i and j). And analogous to (10), E_l turns to:

$$E_l = \beta \sum_{i \in S} \sum_{j \in N_i} f(\Delta_{i,j})[1 - \delta(x_i, x_j)]. \tag{13}$$

Correspondingly, a new objective function can be obtained:

$$\widehat{x} = \underset{\{x_i, i \in S\}}{\operatorname{argmin}} \sum_{i \in S} \{-\log(l_{i,x_i}) + \beta \sum_{j \in N_i} f(\Delta_{i,j})[1 - \delta(x_i, x_j)]\}, \tag{14}$$

Generally, $f(\Delta_{i,j})$ is defined as an exponential function (Yu and Clausi, 2008). Here, we define $f(\Delta_{i,j})$ as:

$$f(\Delta_{i,j}) = \exp(-|\Delta_{i,j}|). \tag{15}$$

Furthermore, in order to reach an analytical formation of the new objective function (14), the dissimilarity measurement $\Delta_{i,j}$ should be also defined. Instead of using gradient-based metrics, we chose spectral-based metrics to define $\Delta_{i,j}$ since the gradient of gray values may not correctly model real boundaries, especially in highly-textured regions (Martin et al., 2004; Wang et al., 2014). Considering the highly varying spectral characteristics of remote sensing images, we examine four spectral metrics for this purpose, including the spectral angle mapper (SAM), spectral information divergence (SID) (Chang, 2000), the combination of SAM and SID (SAM-SID) (Du et al., 2004), and a normalized Euclidean distance (NED) (Robila and Gershman, 2005) which are respectively defined as follows:

$$\Delta_{i,j}^{\mathrm{SAM}} = \cos^{-1}\left(\frac{\langle y_i, y_j\rangle}{\|y_i\|_2\|y_j\|_2}\right) \tag{16}$$

$$\Delta_{i,j}^{\mathrm{SID}} = \sum_{b=1}^{B}\widehat{y}_{i,b}\log\frac{\widehat{y}_{i,b}}{\widehat{y}_{j,b}} + \sum_{b=1}^{B}\widehat{y}_{j,b}\log\frac{\widehat{y}_{j,b}}{\widehat{y}_{i,b}} \tag{17}$$

$$\Delta_{i,j}^{\mathrm{SAM-SID}} = \Delta_{i,j}^{\mathrm{SID}} \cdot \sin(\Delta_{i,j}^{\mathrm{SAM}}) \tag{18}$$

$$\Delta_{i,j}^{\mathrm{NED}} = \sqrt{\sum_{b=1}^{B}\left(\frac{y_{i,b}}{\bar{y}_b} - \frac{y_{j,b}}{\bar{y}_b}\right)^2} \tag{19}$$

where y_i is a B-dimensional spectral vector with B elements, $y_{i,b}$ ($b = 1,\ldots,B$), $\bar{y}_b$ is the mean value of y_b, and $\widehat{y}_{i,b}$ is the normalized value by $y_{i,b}$ over the summation of all elements of y_i.

By substituting (16)–(19) into (15) respectively, and further into (14), a group of new objective functions are obtained. The solution of each objective function will serve as the initial state of the second MRF model.

3.2. Spatial model integrating class co-occurrence dependency

Integrating local dissimilarity information into a spatial model is effective for preserving edges, but it also creates unexpected object particles around boundaries. In order to solve this problem, we propose a second spatial model that integrates the class co-occurrence dependency. In this way, a greater penalty is applied to class pairs with weak spatial relation and a smaller penalty to class pairs with strong spatial relation.

In detail, x_i represents the label of a generic site i; $i+1$ is the site next to i along direction d; the pair-site interaction between the site i and the site $i+1$ along direction d can be expressed with a class/label co-occurrence probability $g_d(x_i, x_{i+1})$. As the set of pair-site cliques was defined over a second-order neighborhood system (8 – neighborhood system), the direction d should be any of 8 directions, that is, $d \in \{1, 2, \ldots, 8\}$ is set. Subsequently, the classic potential function (9) can be rewritten as a direction-dependent potential function:

$$V(x_i, x_j) = \begin{cases} \beta[1 - g_d(x_i, x_{i+1})] & \text{if } x_i \neq x_{i+1} \text{ and } d \in \{1,2,\ldots,8\} \\ 0 & \text{otherwise} \end{cases}. \tag{20}$$

Correspondingly, a new objective function can be updated from (11) as follows:

$$\widehat{x} = \underset{\{x_i, i\in S\}}{\operatorname{argmin}} \sum_{i\in S}\left\{-\log(l_{i,x_i}) + \beta \sum_{d\in\{1,2,\ldots,8\}}[1 - g_d(x_i, x_{i+1})][1 - \delta(x_i, x_{i+1})]\right\}, \tag{21}$$

According to (20), for a given pair of sites along direction d, if the two labels show strong spatial dependency, i.e., $g_d(x_i, x_{i+1})$ is large, a small penalty is applied to the spatial model through $[1 - g_d(x_i, x_{i+1})]$. As the more likely configurations for (21) are those with lower function values, a new class assignment accompanying a stronger spatial relation will be accepted with a higher probability. In this way, only those label pairs with high co-occurrence dependency are preserved, and the output of the first model is further regularized.

After establishing the model (21), an immediate question is how to estimate the probability $g_d(m, n)$ with $\forall m, n \in \Omega$ (for simplicity, x_i, x_{i+1} are re-denoted as m, n respectively.). The probability $g_d(m, n)$ is naturally linked to the frequency ratio of a given classification map and estimated as

$$g_d(m, n) = \frac{|S_{(m\to n)_d}|}{|S_m|} \quad d = 1, 2, \ldots, 8, \tag{22}$$

where S_m is the set of pixels that are assigned with the label m and $|S_m|$ is the number of the pixels; $(m \to n)_d$ denotes a co-occurring pair with the label m and n along the direction d, and $S_{(m\to n)_d}$ is the set of these pairs. In practice, $|S_{(m\to n)_d}|$ can be computed by following the similar routine adopted by the well-known gray level co-occurrence matrix (GLCM) (Haralick et al., 1973). Consequently, for any direction $d = \{1, 2, \ldots, 8\}$, the class co-occurrence probability is associated with a $k \times k$ matrix $G_d = \{g_d(m, n), \forall m, n \in \Omega\}$, where k is the number of class.

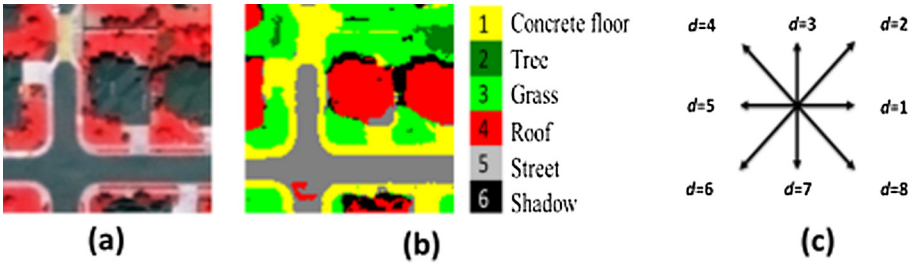
Fig. 1 clarifies the previous idea on a simple (urban) image. Fig. 1(b) lists 6 classes contained in (a). According to the 8 direction configurations in (c), one of the 8 probability matrixes with $G_{d=3}$ is shown in (d). This matrix is composed by two kinds of elements: the main diagonal elements and non-diagonal elements. The main diagonal elements are related to the intra-class dependency. The non-diagonal elements, representing the inter-class possibility, denote the strength of spatial interactions. Taking $g_3(4, 6)$ in (d) as an example, this entry has a significantly higher value (0.0612) than other off-diagonal elements in the same row, which means that the roof-shadow class pair (along the direction $d = 3$) has a tighter interaction as compared to the interactions between roof and other classes.

It is apparent, if the accuracy of co-occurrence probabilities is not satisfactory, the defined spatial context model (20) may lead to a poor spatial optimization performance. Therefore, in order to let the model to play a positive role, we initialize the co-occurrence probabilities according to the output map of the first model and update the estimation iteratively during the second optimization process. This is also the main reason why we design a two-step MRF method, rather than integrate the spectral and spatial factors in a single spatial model.

We should also mention that the employment of the class co-occurrence matrix is not a fresh idea. For instance, Huang et al. (2014) utilize the primitive co-occurrence matrix to re-learn the feature space of initial classification; the class co-occurrence matrix has also been employed to describe multiple Markov chains (Scarpa et al., 2009); Aghighi et al. (2014) define block and global co-occurrence matrixes for estimating the smoothing parameter of an edge-preserving MRF model. Our contribution is that the matrix is integrated into the spatial part of the second energy function, and aimed to optimizing noisy boundaries of initial classification map.

3.3. Selection of smoothing parameter

Before finding solutions of the two objective functions (14) and (21), the values of the parameter β in the two functions are required to be known. We only estimate the optimal value for the first MRF model (14) and directly set the parameter β in the second function (21) as equivalent to the value in the first one



	1	2	3	4	5	6
1	0.8754	0.0004	0.0560	0.0013	0.0573	0.0009
2	0.0000	0.7969	0.1724	0.0000	0.0077	0.0230
3	0.0397	0.0134	0.8971	0.0138	0.0027	0.0134
4	0.0037	0.0005	0.0005	0.9190	0.0120	0.0612
5	0.0604	0.0000	0.0018	0.0200	0.9124	0.0014
6	0.0188	0.0188	0.1770	0.0697	0.0207	0.6817

(d)

Fig. 1. An example of the calculation of the co-occurrence probability matrix. (a) Urban area image; (b) classification map and corresponding legends; (c) predefined neighboring system; (d) one of the co-occurrence probability matrixes: G_3, and each element $g(m, n)$ represents the frequency ratio related to the class m and n with the spatial configuration $d = 3$.

(14). This is because the spectral energy models applied to the two objective functions are the same, and the second function is initialized by the output of the first one.

Adopting the average accuracy (AA) as the evaluation metric, the optimal value of β is derived by a set of trials based on *reliable* samples. We define an initial classification result on pixel i is *reliable*, if the confidence of this pixel, which is the ratio of the first-largest possibility to the second-largest class possibility, is great than to 2. All samples fulfilling this condition consist of a *reliable* reference data set. The detailed selection process is described as follows:

Step (1): Define an initial search sequence for each image: $R = [2^{-2},\ 2^{-1},\ 2^{-0}, \ldots, 2^{6}]$; Specify β in (14) with each value in R in sequence; obtain a series of optimized classification maps by resorting to a combinatorial optimization method, e.g., the α-expansion graph cut (EGC) algorithm (Boykov and Kolmogorov, 2004).
Step (2): Calculate the AA value for each map based on the *reliable* data set, and set β associated to the highest value (denoted as β_*);
Step (3): Denote the second element before β_* in R as β_{*-2}; update R with the vector of ten linearly equally spaced points between β_{*-2} and β_*; repeat steps (1)–(2) once, and then output β_* and the optimized thematic map associated with this optimal value.

3.4. Overall algorithm

With the observed image $\boldsymbol{I}$, the initial classification map $\mathbf{x}$, and the estimated class probabilities l_{i,x_i} ($x_i \in \Omega$ and $i \in S$), the whole process is described as follows:

(1) Initialize the objective function (14). Specify the initial classification map $\mathbf{x}$ as the *a priori* information $\mathbf{x}^{(0)}$; specify the output of the probabilistic SVM classifier l_{i,x_i} ($x_i \in \Omega$ and $i \in S$) as the initial class probability $l_{i,x_i}^{(0)}$; calculate various spectral dissimilarities $\Delta_{i,j}$ according to (16)–(19), respectively, in order to obtain $f(\Delta_{i,j})$ in (15) for each pixel in the remote sensing image.
(2) Solve the objective functions (14) with different β values by the α-expansion graph cut (EGC) algorithm (Boykov and Kolmogorov, 2004) respectively, determine β_* as the one corresponding to the highest AA value and output the associated classification map $\mathbf{x}^{(t_{\max 1})}$. Section 3.3 described this process in detail.
(3) Initialize the objective function (21) based on the outputs of the objective function (14). Specify the classification map $\mathbf{x}^{(t_{\max 1})}$ obtained from the previous MRF model as the *a priori* information $\mathbf{x}^{(0)}$; specify the output of the probabilistic SVM classifier as the initial class probabilities $l_{i,x_i}^{(0)}$; calculate the initial multi-directional co-occurrence probabilities $g_d^{(0)}(m,n)$ ($\forall m, n \in \Omega$ and d = 1,…,8) based on $\mathbf{x}^{(t_{\max 1})}$ according to (22), and set $\beta = \beta_*$.
(4) Set $t = 0$.
(5) Update $\mathbf{x}^{(t)}$ into $\mathbf{x}^{(t+1)}$ in an ICM manner (Besag, 1986). Sequentially update each $x_i^{(t)}$ into $x_i^{(t+1)}$, and obtain a new $\mathbf{x}^{(t+1)} = \{x_i^{(t+1)} \mid i \in S, x_i^{(t+1)} \in \Omega\}$, where $x_i^{(t+1)} \in \mathbf{x}^{(t+1)}$ is the one that produces the minimum energy for (21), that is,

$$x_i^{(t+1)} = \arg\min_{x \in \{1,2,\ldots,k\}} \left\{ -\log\left(l_{i,x}^{(t)}\right) + \beta_* \sum_{d=\{1,2,\ldots 8\}} \left[1 - g_d^{(t)}\left(x, x_j^{(t)}\right)\right] \times \left[1 - \delta\left(x, x_j^{(t)}\right)\right] \right\} \quad (23)$$

(6) Renew the co-occurrence probabilities $g_d^{(t+1)}(m,n)$ with $\forall m$, $n \in \Omega$ and $d = \{1, 2, \ldots, 8\}$ based on $\mathbf{x}^{(t+1)}$.
(7) If $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)}$ or a maximum number of iterations (in our case, 20) is reached, output $\mathbf{x}^{(t)}$, else set $t = t + 1$ and go to step 5.

For illustrative purposes, a generic flowchart of the proposed method is also presented in Fig. 2, where the indexes of key equations are included. Starting with an initial classification map, the first optimization step is intended to preserve edges while smoothing the raw classification map. The second step is intended to further optimize the output of the first step, by focusing on the pixels located at object boundaries. In the first stage, a heuristic search method according to the confidence of the classification results is integrated for determining proper smoothing parameters.

Note that we choose EGC to solve the first objective function (14), whereas we solve the second objective function (21) by ICM based on computational concerns. With similar optimization results, ICM is much faster than EGC when applied to the second function. Another issue is the convergence of the ICM based iterations. During the ICM iterative process, the probability $p(\mathbf{y}|\mathbf{x})$ is non-decreasing within boundary [0,1]. As a monotonous bounded sequence necessarily converges, the convergence of the iteration process can be guaranteed in this case.

4. Data sets and experimental setting

The experimental data sets, setting and metrics used for the evaluation are introduced in this section.

4.1. Data sets

We used hyperspectral and multispectral images to evaluate the effectiveness of the proposed MRF method. A summary of the considered data sets is reported in Table 1.

The two popular hyperspectral images, *University* and *Salinas*, are illustrated in Fig. 3. The *University* image was obtained from the Reflective Optics System Imaging Spectrometer (ROSIS), acquired over the University of Pavia, Italy. This image includes 9 classes and contains a total of 42,776 reference pixels. The *Salinas* image was obtained over an agricultural area by the Airborne Visible Infra-Red Imaging Spectrometer (AVIRIS). After discarding 20 water absorption bands, totally 204 bands are available for interpretation. The image contains 16 classes and 54,129 pixels are available for validation. Both images are widely used data sets for testing hyperspectral classification algorithms.

Two multispectral images are also used in our experiments and displayed in Fig. 4. The *Hainan* image was acquired by the WorldView-2 satellite, covering a rural area in the Hainan Province in China. The *Wuhan* image was acquired over the city of Wuhan in China by the ZY-3 satellite, which contains very dense buildings. Both images have 7 classes of interest including roads, grass, buildings, soil, shadow, trees and water body.

In order to produce the initial classification maps, hand-crafted or automatically learned features (e.g., from a CNN model) could be used. However, as previewed in the introduction part, the pixel-wise classifications, which involves of spatial features, may provide less-noisy and high-accuracy initial maps. In order to highlight the efficiency of the proposed method, we just adopted basic spectral features. i.e., all spectral bands of the hyperspectral/multispectral images and probabilistic SVM classifier for the pixelwise classification. Following the works (Wu et al., 2004, Chang and Lin, 2011), the optimal parameters for these classifiers (implemented using a radial basis function (RBF) kernel) were obtained by five-fold cross validation. Randomly selected samples from the reference maps (50 pixels per-class for the hyperspectral images and 100

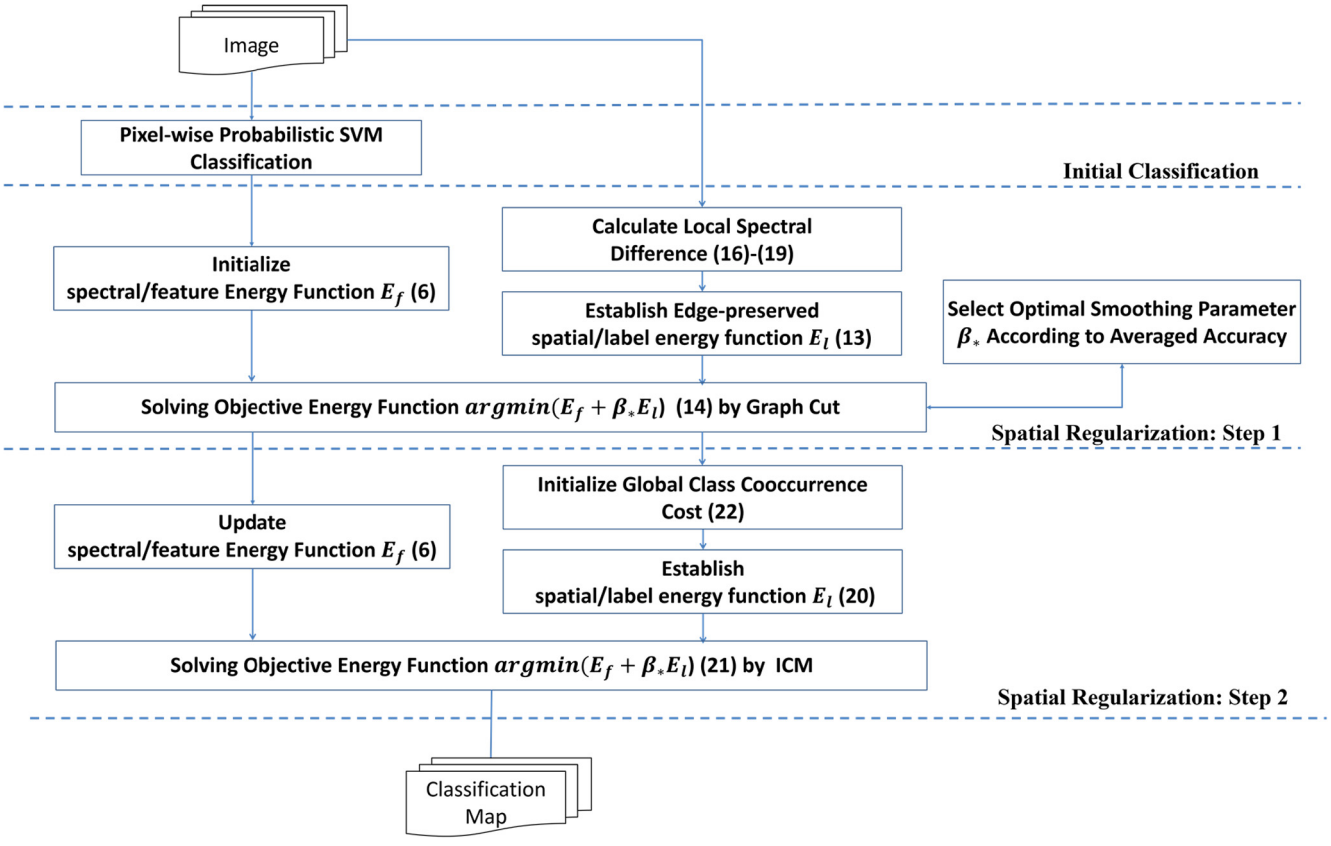


Fig. 2. Flowchart of the proposed MRF-based optimization method.

Table 1
Experimental data sets used in this study.

	Name	Sensor Name	Size	Bands	Resolution (m)	Area
Hyperspectral	University	ROSIS	610 × 340	103	1.3	Rural
	Salinas	AVIRIS	512 × 217	224	3.7	Agriculture
Multispectral	Hainan	WorldView-2	600 × 520	8	2.0	Rural
	Wuhan	ZY-3	651 × 499	4	5.8	Dense Urban

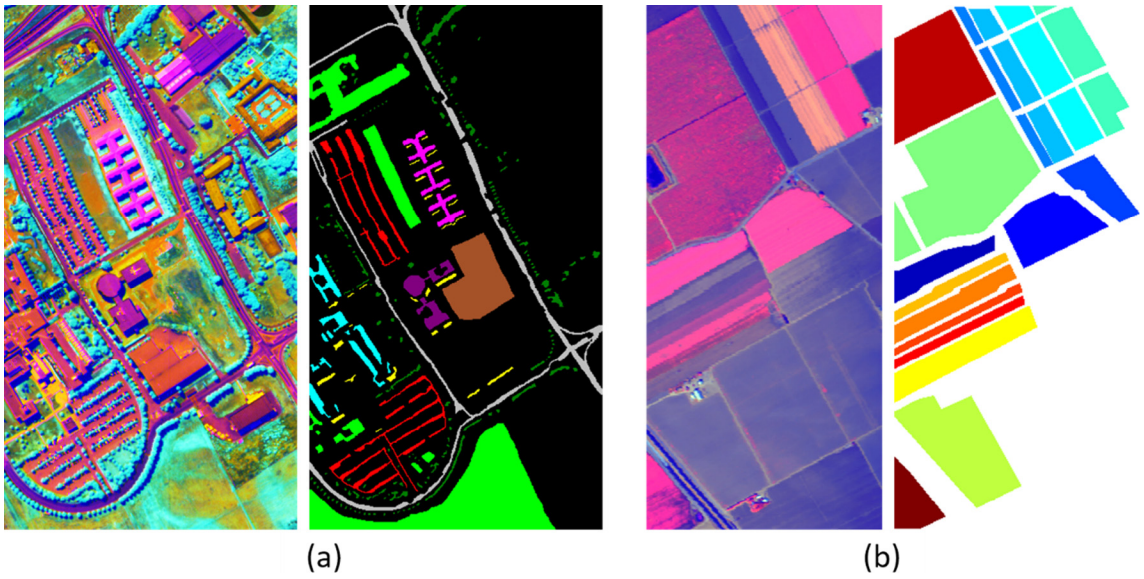


Fig. 3. Two hyperspectral data sets considered in experiments: *University* (a) and *Salinas* (b). The false color images shown in (a) and (b) are generated using the first three principal components as red, green and blue components; the second columns of each sub-figures show the ground truth pixels for each image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

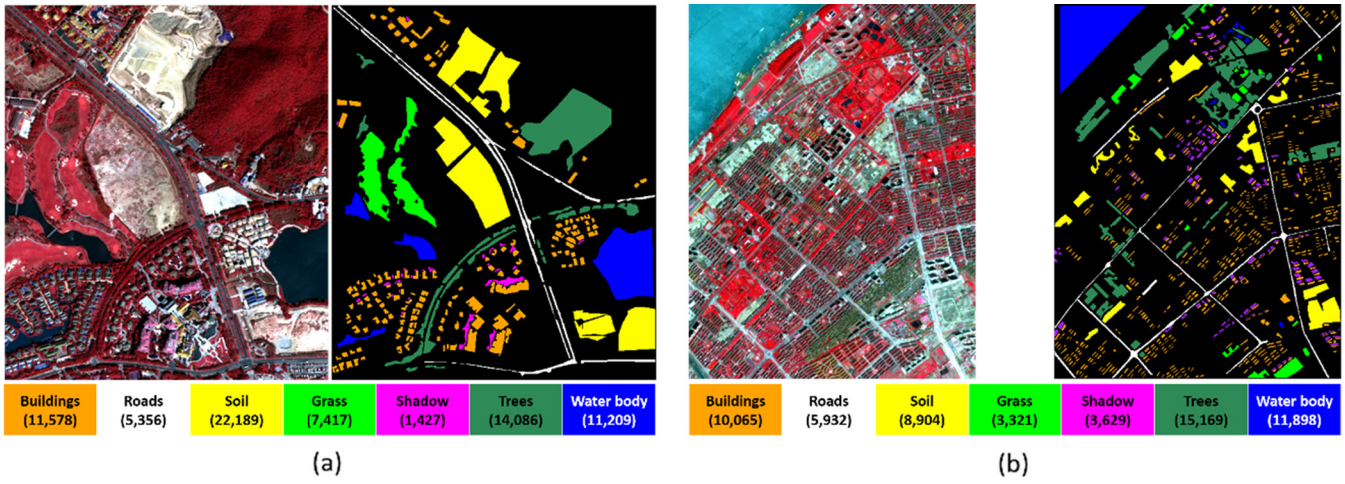


Fig. 4. Two multispectral data sets considered in experiments: WorldView-2 *Hainan* (a) and ZY-3 *Wuhan* (b). Each sub-figure illustrates a false color image with near-infrared, red and green bands, a ground truth map and class legends with the number of validation samples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

pixels per-class for the multispectral images) were utilized for training.

4.2. Considered evaluation metrics

Commonly used accuracy metrics include Kappa, overall accuracy (OA) and average accuracy (AA). As these metrics conduct pixel-based evaluation on limited reference samples, the evaluation is biased and not sufficiently on estimating the performance of algorithms around boundary areas (Rutzinger et al., 2009). Therefore, we considered three extra metrics including the Moran's *I* spatial autocorrelation index (MI) (Espindola et al., 2006), the empirical evaluation of segmentation (EES) (Borsotti et al., 1998) and the match cost between two boundary maps (MC) (Martin et al., 2001) for comprehensive assessment. MI and EES are sample-independent metrics and estimate object regions from different views; MC computes the correspondence of two boundary maps.

In detail, MI is defined as follows:

$$\mathrm{MI} = \frac{R}{\sum_{i=1}^{R}\sum_{j=1}^{R} w_{i,j}} \frac{\sum_{i=1}^{R}\sum_{j=1}^{R} w_{i,j}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{j=1}^{R}(y_i - \bar{y})^2} \tag{24}$$

where R is the total number of regions; y_i and $\bar{y}$ are the mean value of the region i and the whole image respectively, $w_{i,j}$ is a measure of spatial adjacency of regions, where $w_{i,j}$ is given a value of 1 if two regions are neighbors, and 0 otherwise. Thus, on average, MI denotes the difference between each region and its neighbors.

EES is defined as follows:

$$\mathrm{EES} = \frac{1}{10,000(N \times M)} \sqrt{R} \times \sum_{i=1}^{R} \left[\frac{e_i^2}{1 + \log A_i} + \left(\frac{R(A_i)}{A_i} \right)^2 \right] \tag{25}$$

where $N \times M$ is the size of image, R is the number of regions, $R(A_i)$ is the number of regions having an area equal to A_i and e_i represent the average color error of the ith region, which is defined as the sum of the Euclidean distances between each spectral vectors of the pixels of region i and the average vector of region i. The sum in EES consists of two terms, the first term is high for non-homogeneous regions (high e_i) and the second may be high for small regions (as the number of regions of area A_i may be large). Therefore, a small EES value is preferred, as it indicate a segmentation contain large and homogeneous regions. MI and EES together provides a comprehensive on closed regions derived from classification maps. MI assess the inter-region heterogeneity. In contrast, EES assess the intra-region homogeneity.

MC compute the minimum-cost correspondence between boundary maps. The cost of corresponding two pixels is equal to their Euclidean distance, and two pixels beyond the predefined distance threshold cannot be corresponded. Therefore, a small MC value is preferred, as it indicate a better correspondence with reference maps. In our cases, the distance threshold is set as a fraction (=0.01) of image diagonal and the reference boundary maps are manually sketched by experts. Here, we define a pixel as a *boundary* pixel if a class change between the current pixel and its neighboring pixels is detected.

5. Experimental results and discussion

5.1. General results obtained by the proposed MRF models

In order to achieve a comprehensive evaluation of the tested methods, a series of different MRF models were carried out based on the four considered data sets. The initial results from the probabilistic SVM are named as Raw; Classic-MRF denotes the results of the classic MRF (11), which is also implemented with the proposed parameter selection method and solved by the graph cut; SAM-MRF, SID-MRF, SAMSID-MRF and NED-MRF denote the proposed spectral-class spatial model (21), but with 4 different spectral difference metrics [shown in (16)–(19)].

Based on the randomly selected training sets, each method was repeated 30 times in order to ensure statistical significance. Both the mean and standard deviation values of OA from the four test images are reported in Table 2.

The McNemar's test (Foody, 2004) was also applied to each pair of results obtained from two different methods, where the Z-Score is applied to validate whether the differences between two methods are statistically significant. At a general 5% significance level, the significant difference between methods is accepted if $|Z| > 1.96$. More specifically, $Z > 1.96$ or $Z < -1.96$ indicates that the second method is more or less accurate, respectively, and $|Z| < 1.96$ indicates the two methods have no significant difference. For instance, as shown in Table 3, "SAM-MRF vs. NED-MRF" results in "3+, 1N", indicating that the NED-MRF method obtained three instances of positive significance and one instance of no significance in the four considered data sets when compared with SAM-MRF.

As shown by Table 2, compared with the raw classification results, all methods achieve significant improvements. For example, the increments in terms of OA are around 12% (*University*), 10% (*Salinas*), 4% (*Hainan*) and 4% (*Wuhan*), respectively. This observation is also supported by the results of the McNemar's test shown in Table 3, since all MRF methods obtains "4+" with respect to the raw results. Meanwhile, the proposed four methods are also significantly better than the classic MRF method, as all the methods have achieved "4+" over the Classic-MRF.

Among all the four proposed models, the NED-MRF model achieved the highest OA in most cases. When considering all related statistics shown in Table 3, the difference between each pair of methods selected from the four proposed ones are statistically significant, as the statistics are dominated by "+" or "−". Therefore, a descending order on the performance of the four models can be determined in terms of the statistics: NED-MRF > SAM-MRF > SID-MRF > SAMSID-MRF.

Some of the classification maps obtained for the *Salinas* data set are presented in Fig. 5. This data set is special, as a lot of "1+" and "1N" values were obtained for this data set as listed in Table 3. It is clear in Fig. 5 that the "salt-and-pepper" problem in the raw classification maps can be substantially reduced by all MRF-based methods. Furthermore, as illustrated in the local patches designated as 1–4 in Fig. 5, the differences among the compared methods is obvious and mainly occurs around objects edges. In this regard, the classic MRF method [see Fig. 5(d)], which is based on the smoothness assumption, tends to heavily oversmooth small structures. In contrast, the proposed methods [see Fig. 5(e)–(h)] show the advantages of constraining oversmoothness and preserving edges. Moreover, the NED-MRF method shows superior visualization performance than other tested methods. According to both quantitative and qualitative assessments, the proposed MRF methods (particularly the NED-MRF) can achieve a good tradeoff between oversmoothness and spatial regularization.

We have also tested the computational time using a Windows 7 laptop with an Intel I7 2.7-GHz CPU using 16-GB memory. The averaged processing times of the four proposed methods in this computing architecture are reported in Table 4. As can be seen, the NED-MRF method always results in the shortest processing time, i.e., 67.54, 60.21, 89.97, and 265.81 s for *University*, *Salinas*, *Hainan*, and *Wuhan* images, respectively. When considering the two optimization steps, the first one required much more computational time than the second one. This step is quite time-consuming, as it involves an exhaustive search process for selecting an optimal smoothing parameter.

5.2. Analysis of the integration of class co-occurrence dependency

In this subsection, by comparing the output of the first and second stages of the proposed framework, we validate the necessity of integrating class co-occurrence dependency. For a given data set, each considered method was repeated 30 times with different initial training samples.

We calculated and compared the four metrics values (i.e., OA, EES, MI and MC), resulting in four sub-figures for each data set. As similar conclusions can be drawn for all data sets, Fig. 6 just presents the two results obtained from *University* and *Hainan*. The first sub-figure shown in Fig. 6(a) consists of 4 pairs of OA curves obtained from *University*; each pair is composed of two results, obtained after integrating the class co-occurrence dependency and without such integration. For instance, NED vs. NED-MRF,

Table 2
Overall accuracy of the MRF-based algorithms proposed in this study for the different considered remote sensing data sets. Starting from randomly selected training sets, each algorithm is repeated 30 times. Both the mean and standard deviation of the OA and AA values are reported in the form of "mean ± standard deviation".

	University	*Salinas*	*Hainan*	*Wuhan*
Raw	84.5 ± 2.01	88.7 ± 1.08	93.2 ± 0.67	88.5 ± 0.71
Classic-MRF	89.0 ± 2.06	91.9 ± 1.12	94.4 ± 0.67	90.1 ± 0.66
SAM-MRF	96.7 ± 1.79	98.1 ± 1.56	97.7 ± 0.74	93.1 ± 0.54
SID-MRF	96.2 ± 1.85	97.9 ± .149	97.6 ± 0.78	93.1 ± 0.49
SAMSID-MRF	96.1 ± 1.97	97.8 ± 1.47	97.5 ± 0.83	93.1 ± 0.52
NED-MRF	98.2 ± 1.41	99.2 ± 0.67	97.9 ± 0.74	93.1 ± 0.47

Table 3
Summary of McNemar's test conducted for the classification maps provided by 6 different methods using 4 data sets, respectively, where + , n and – denote positive, no and negative statistical significance, respectively.

	Classic-MRF	SAM-MRF	SID-MRF	SAMSID-MRF	NED-MRF
Raw	4+	4+	4+	4+	4+
Classic-MRF	–	4+	4+	4+	4+
SAM-MRF	–	–	4–	4–	3+, 1N
SID-MRF	–	–	–	3–, 1+	4+
SAMSID-MRF	–	–	–	–	4+

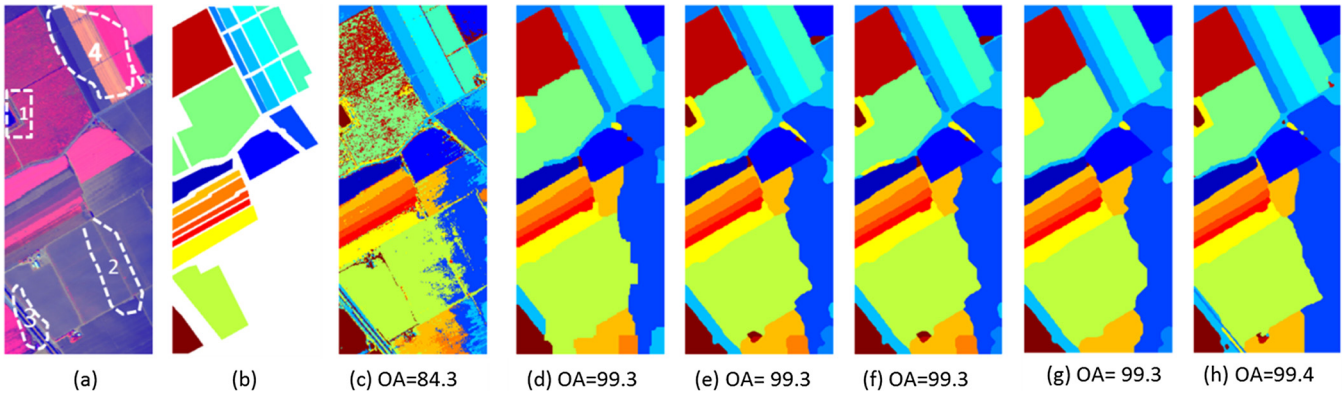


Fig. 5. Classification results for the *Salinas* data set obtained by different MRF-based methods. (a) Three-band false color composite image, (b) validation data, (c) raw classification result obtained by the probabilistic SVM, (d) classic MRF method, (e) SAM-MRF method, (f) SID-MRF method, (g) SAMSID-MRF method, and (h) NED-MRF method. The noise (pepper-and-salt effect) can be appreciated in the raw image (c), while a visible gap among maps with similar OA values and superior performance of the proposed methods (e–h) in object boundaries can be appreciated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Processing time (seconds) for the different MRF-based methods when applied to different data sets.

Method		*University*	*Salinas*	*Hainan*	*Wuhan*
NED-MRF	Step 1	43.85	38.18	61.92	186.12
	Step 2	20.59	15.32	24.94	69.18
	Total	67.54	60.21	89.97	265.81
SAM-MRF	Step 1	116.69	80.17	79.55	205.76
	Step 2	16.54	15.09	24.95	77.73
	Total	144.43	102.83	107.60	293.88
SID-MRF	Step 1	86.24	85.87	85.22	227.13
	Step 2	22.17	15.11	26.20	73.73
	Total	111.52	112.03	114.51	311.40
SAMSID-MRF	Step 1	120.28	84.29	97.30	220.31
	Step 2	28.33	26.07	28.10	77.80
	Total	153.83	113.07	128.49	307.82

NED denotes the MRF method only with spectral dissimilarity in terms of the normalized Euclidean distance (NED), while NED-MRF denotes the same method implemented with both NED and class co-occurrence dependency.

As it can be seen in Fig. 6, in most cases the NED-MRF and NED exhibit the best performances with respect to the four considered metrics. Meanwhile, it is also apparent that the integration of the class-occurrence possibilities leads to constant increments over different data sets and experimental settings.

Due to the lack of validation samples, especially samples around object boundaries, the increments in terms of OA seem not significant and are as low as 1%. When considering the MI, and EES met-

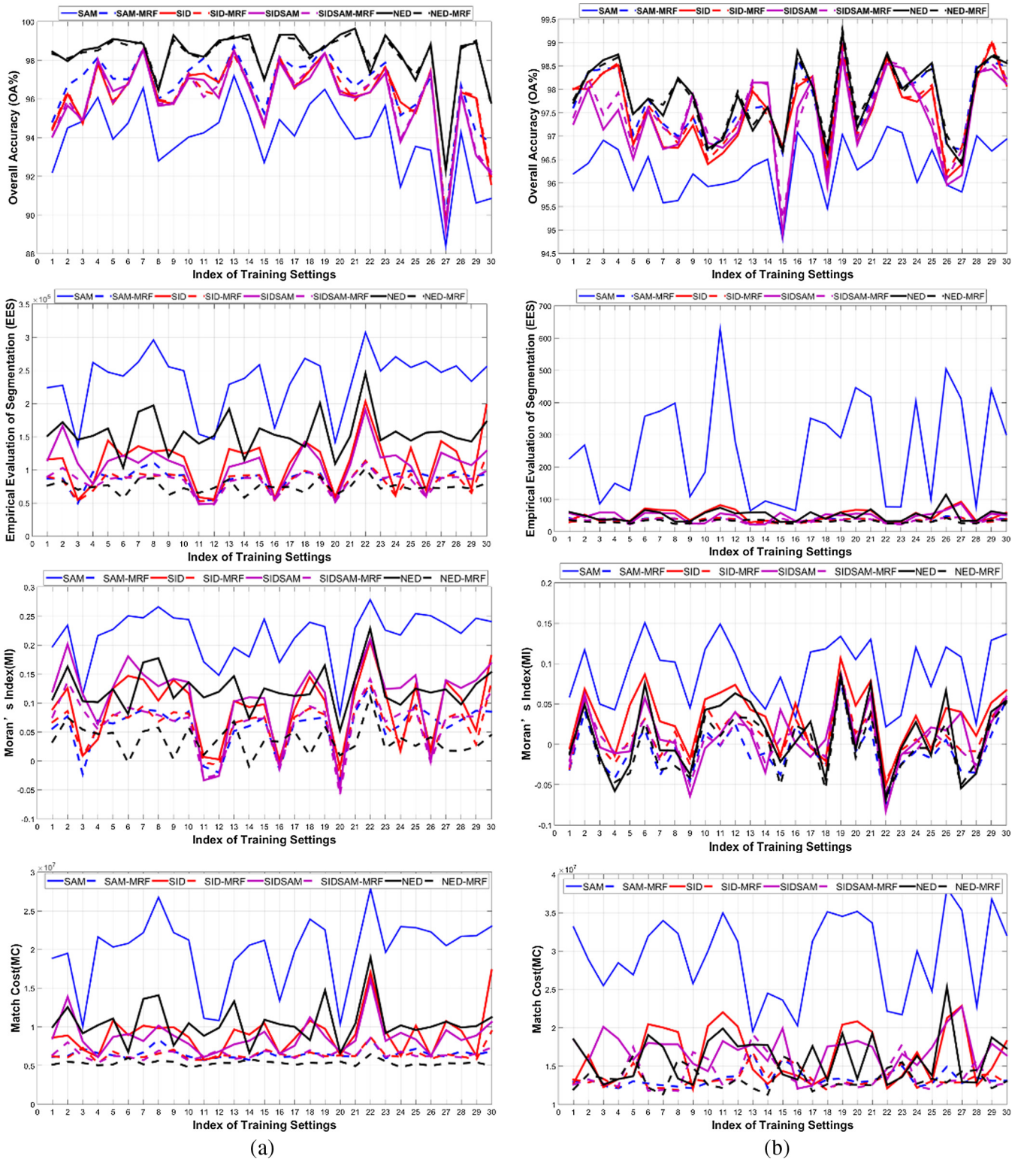


Fig. 6. Comparison of different pairs of MRF-based methods according to OA, EES, MI and MC: (a) *University*, and (b) *Hainan*. The solid line and the dotted line with the same color respectively denote a method considering only spectral dissimilarity and the corresponding method integrating both spectral dissimilarity and class-occurrence dependency. Expect OA, small values of three other metrics are preferred. Noteworthy details include that the class co-occurrence dependency leads to an improvement of the four metrics in most cases, and the stable performance of the proposed methods across different data sets and experimental settings. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

rics, as illustrated in the second to and third columns of Fig. 6, the integration of class co-occurrence dependency leads to an obvious decrease of values in most cases. Small values of MI indicate large inter-region heterogeneity, thereby exhibiting clear boundaries between regions; small values of EES indicate that the classification map is composed of large and homogenous regions. Therefore, both MI and EES values indicate that the integration can improve classification results.

Furthermore, MC clearly indicates the improvement around boundaries, as the integration of class co-occurrence dependency greatly reduce the match cost (better correspondence between detected and ideal boundaries) for all cases. Meanwhile, as shown in Figs. 7 and 8, the actual visual improvements between each pair of maps is also quite clear. Through the integration of spatial dependency, more accurate location of boundaries and more compact objects can be obtained, the noisy pixels among different land objects are largely suppressed.

Both visual and quantitative assessments from regions and boundaries suggest that a spatial model only considering local spectral information is often sensitive to gray value discontinuities and confirm the need to integrate class co-occurrence dependency in the MRF module.

5.3. Evaluation of the selection strategy of the β parameter

This section focuses on analyzing the proposed strategy for selecting the value of β parameter. When an MRF method is applied to a given data set, the optimal β value is selected from a search set. We exhaustively tested every candidate value in the set. Fig. 9 shows the *reliable* reference samples obtained according to the ratio of SVM possibilities; Table 5 reports the β values corresponding to the highest AA and Kappa, and the selected β.

As it can be seen in Fig. 9, the reference samples consist of a representative subset of the original image. When considering the selected β values with regards to their counterpart values related to the optimal average accuracy (A_{opt}) and Kappa (K_{opt}), the three values are identical or very close in most cases. Large differences [2, 0.5, 0.44] and [4, 1, 0.92] were observed from *University* with the SAM-MRF method, and for *Hainan* with the NED-MRF method. We report their producer's accuracies in Table 6 for further investigation.

It is clear from Table 6 that a large β value is often preferred in order to model some artificial land-cover classes, such as road, building, asphalt and painted sheet. On the contrary, for the recognition of natural land classes such as trees, grass, shadow and gravel, a relatively small value is preferred. Artificial land classes

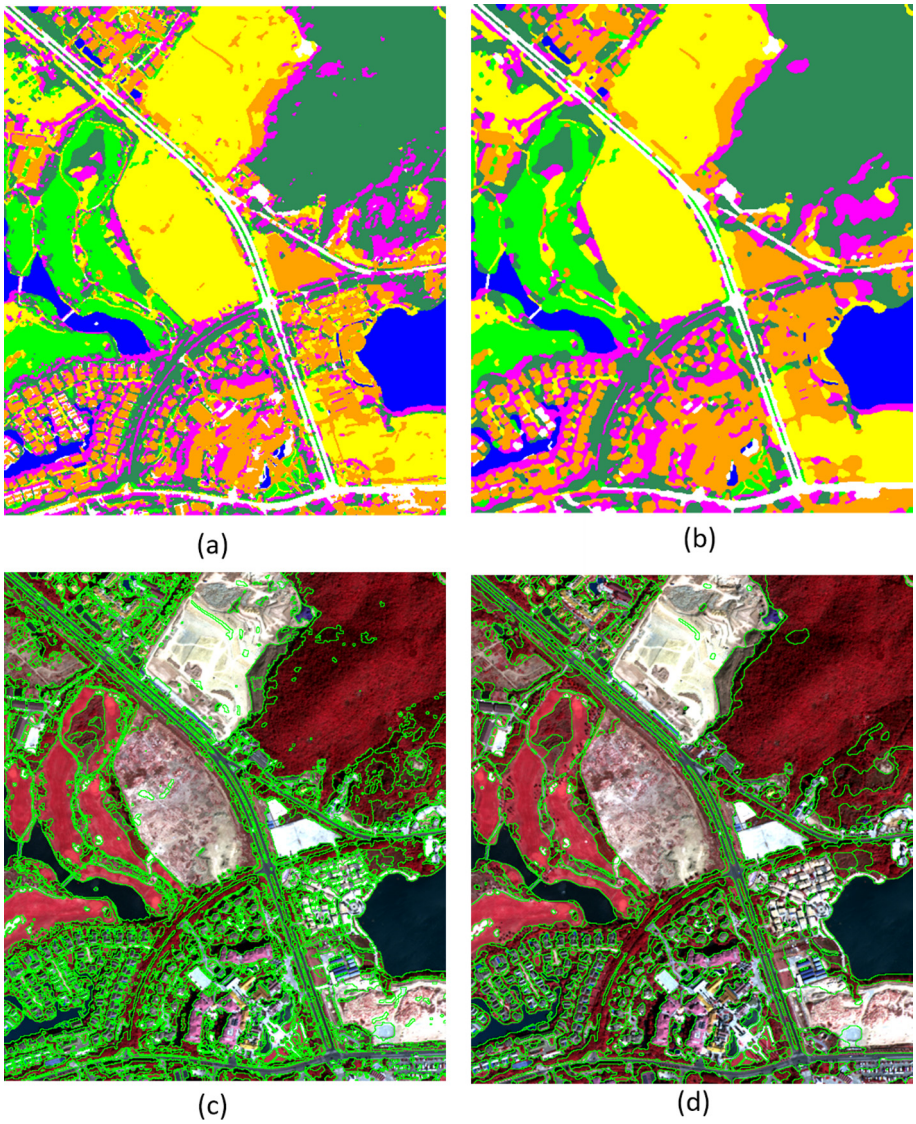


Fig. 7. Comparison of the results obtained with and without integrating class co-occurrence dependency using the *Hainan* image. (a) SAM (MRF method only with the spectral dissimilarity of SAM), which results in OA = 96.7%, MI = 0.0410, EES = 148.0, and MI = 3.31E07; (b) SAM-MRF (MRF method with both SAM and class co-occurrence dependency), which results in OA = 98.5%, MI = −0.0418, EES = 34.0, and MI = 1.29E07. (c) The overlay of original bands and the object boundaries responding to (a). (d) The overlay of original bands and the object boundaries responding to (b). Noteworthy details include the visual gap between (a) and (b), and the fact that SAM-MRF performs constantly better than SAM with respect to the four considered metrics.

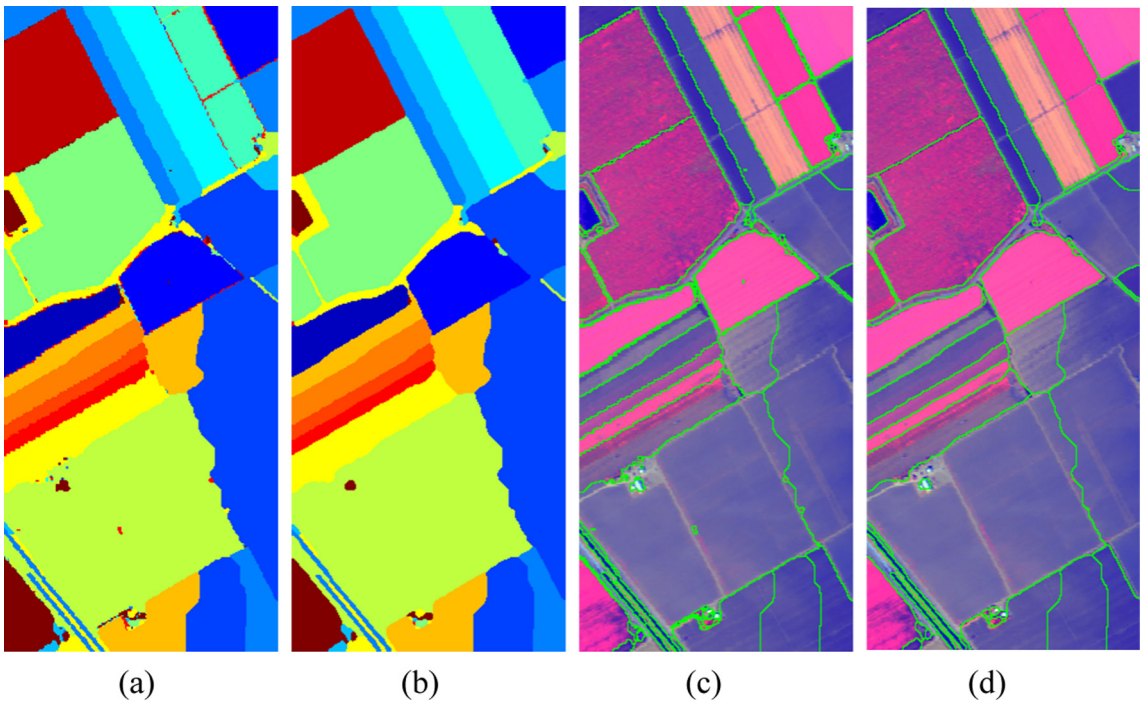


Fig. 8. Comparison of the results obtained with and without integrating class co-occurrence dependency using the *Salinas* image. (a) NED (MRF method only with the spectral dissimilarity of NED); (b) NED-MRF (MRF method with both SAM and class co-occurrence dependency); (c) the overlay of original bands and the object boundaries responding to (a); (d) the overlay of original bands and the object boundaries responding to (b). NED-MRF performances better around boundaries.

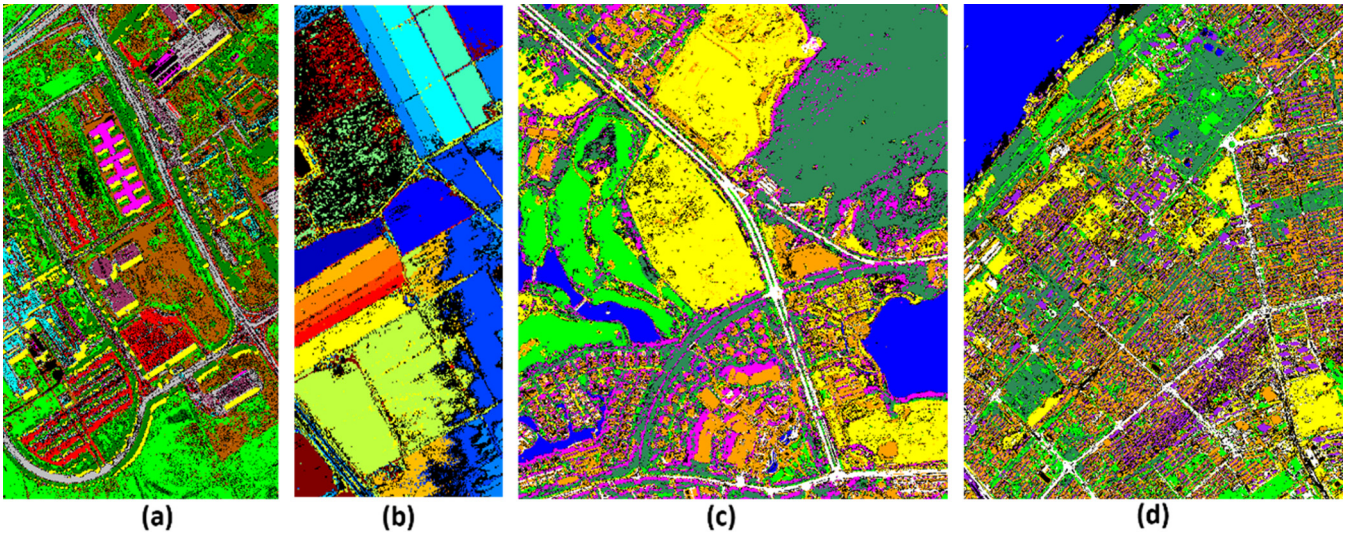


Fig. 9. *Reliable* samples obtained according to the ratio of class possibilities and used for selecting proper smoothing parameters: (a) *University*, (b) *Salinas*, (c) *Hainan*, and (d) *Wuhan*. Black color in each figures denotes the background.

Table 5
Smoothing parameter β values corresponding to the highest kappa (denoted as K_{opt}), the highest average accuracy (denoted A_{opt}) and the selected smoothing parameters according to the proposed methods.

	SAM-MRF			SID-MRF			SIDSIN-MRF			NED-MRF		
	K_{opt}	A_{opt}	Selected	K_{opt}	A_{opt}	Selected	K_{opt}	A_{opt}	Selected	K_{opt}	A_{opt}	Selected
University	2	0.5	0.44	2	0.92	0.92	1	0.5	0.5	3.67	3.67	3.67
Salina	3.67	3.67	3.67	4	3.67	3.67	4	3.67	3.67	14.67	14.67	14.67
Hainan	1.83	1.67	1.33	1.83	1.5	1.5	2	1.83	1.83	4	1	0.92
Wuhan	0.5	0.25	0.25	0.5	0.25	0.25	0.5	0.25	0.25	0.5	0.25	0.25

Table 6
Evaluation of the influence of β values on the classification quality: (a) *Hainan* image, NED-MRF method, and (b) *University* image, SAM-MRF method. Three β values are associated with the optimal Kappa, the optimal average accuracy, and the proposed selected method respectively. The highest values of every class and every metric are in bold.

Hainan	NED-MRF			
Class Name	# sample	β = 4	β = 1	β = 0.92
(a) Results for the Hainan image				
Road	5357	**89.70%**	85.41%	85.14%
Grass	7417	98.39%	99.05%	**99.05%**
Building	11,578	**100.00%**	99.94%	99.92%
Soil	22,189	**99.87%**	99.77%	99.85%
Shadow	1427	89.21%	**92.99%**	92.43%
Tree	14,086	99.01%	**99.72%**	99.64%
Water body	11,209	**100.00%**	**100.00%**	**100.00%**
Kappa	–	**0.9734**	0.9680	0.9671
OA	–	**97.84%**	97.39%	97.32%
AA	–	96.60%	**96.70%**	96.57%
MI	–	0.0112	0.0056	**−0.0363**
EES	–	47.39	45.87	**23.10**
University	SAM-MRF			
Class Name	# sample	β = 2	β = 0.5	β = 0.44
(b) Results for the University image				
Trees	3064	89.95%	**92.79%**	92.56%
Asphalt	6631	**96.79%**	96.18%	93.94%
Bitumen	1330	**99.62%**	99.40%	99.25%
Gravel	2099	84.99%	**89.04%**	88.90%
Sheet	1345	**99.93%**	**99.93%**	**99.93%**
Shadow	947	98.31%	**99.89%**	**99.89%**
Bricks	3682	**99.78%**	99.19%	99.24%
Meadows	18,649	**96.51%**	91.96%	91.95%
Bare Soil	5029	**100.00%**	**100.00%**	**100.00%**
Kappa	–	**0.9540**	0.9334	0.9287
OA	–	**96.45%**	94.75%	94.38%
AA	–	96.21%	**96.49%**	96.18%
MI	–	0.1459	0.1501	**−0.0442**
EES	–	124544.76	117030.96	**46814.40**

often exhibit compact shape and a relatively large size, whereas natural classes have unpredictable shapes and fragmental appearances. When considering the overall performance, the selected β obtained slightly lower values in terms of OA, AA and Kappa. In turn, when considering two simple-independent metrics such as MI and EES, the proposed method achieves the best segmentation performance.

According to the previous analysis, we can draw the following conclusions: the proposed selection method can select an effective parameter for the proposed MRF-based methods; combined with the proposed MRF-based methods, the optimized maps can achieve a good tradeoff among classes with various sizes and geometric structures, preserving the shapes of small-size objects.

5.4. Comparisons among postprocessing methods

In this section, the proposed NED-MRF method is compared with several state-of-the-art post-classification optimization methods, including edge-preserved filtering (EPF) (Kang et al., 2014), a recently proposed relearning primitive co-occurrence matrix (Relearning) (Huang et al., 2014), weighted majority vote (WMV) (Büschenfeld and Ostermann, 2012), multinomial logistic regression with subspace projection (MLR-SP) (Li et al., 2012) and the MRF-based spatial regularization method with gradient information (Gradient-MRF) (Tarabalka et al., 2010). Generally, these methods can be divided into two categories: local and global. EPF and WMV are labelled as local filtering-based methods, as they only consider spatial-contextual information in a local neighborhood. Spatial models of the MLR-SP, Gradient-MRF and NED-MRF methods were developed from the classic MLL model (Derin and Elliott, 1987), and the methods aimed to maximize the posterior over the entire random field. Therefore, these three methods belong to the global regularization category. R-PCM can also be seen as a global optimization method, as it involves a gradual re-learning process of class frequency and spatial arrangement.

In order to achieve a fair comparison, all methods start from the same initial SVM probabilistic output and are assessed based on the same test samples. The detailed experimental setting of initial classifications can be found in the previous Section 4.1. Optimal parameters for these baseline methods were predefined or tuned following the original works. The experimental settings are summarized as follows. For the EPF, the guide images are the first three bands for the multispectral images and the first three principal components for the hyperspectral images, respectively; the regularization parameter ε and the local window radius w are set as [0.01, 4] for *University*, *Hainan* and *Wuhan* images, and as [0.01, 8] for the *Salinas* image. For the WMV, a spatial bandwidth $H_S = 5$, a spectral (range) bandwidth $H_r = 5$ and a minimum region size M = 20 are considered. For the Relearning method, we consider a window radius $w = \{3, 4, 5\}$. For the MLR-SP and Gradient-MRF methods, the optimal smoothing parameter β is selected in the range {0.5, 1, 1.5, 2, 3, 4}, which corresponds to the highest OA. We should emphasize that the only predefined parameter in our proposed method is the initial search range of the smoothing parameter, and it was set to the same value across all experiments.

As shown in Table 7, all methods employed improve the raw classification to different degrees. The increments of OA achieved by the NED-MRF method are 14.4% for the *University* data set, 10.8% for the *Salinas* data set, 4.2% for the *Hainan* data set, and 3.8% for the *Wuhan* data set, respectively.

When comparing the 6 postprocessing algorithms, the proposed method NED-MRF yields the best performance with respect to

Table 7
Quantitative classification/segmentation evaluation of the EPF (Kang et al., 2014), WMV (Büschenfeld and Ostermann, 2012), Relearning (Huang et al., 2014), MLR-SP (Li et al., 2012), Gradient-MRF (Tarabalka et al., 2010) and NED-MRF methods based using the four data sets with OA, KAPPA, AA, MI, EES and MC metrics. The best results are in bold and the second best results are underlined.

Image Name	Metric	Raw	Local method		Global method			
			EPF	WMV	R-PCM	MLR-SP	Gradient-MRF	NED-MRF
University	OA	84.26%	95.74%	97.61%	93.28%	91.01%	95.02%	**98.67%**
	Kappa	81.05%	94.53%	96.88%	91.46%	88.06%	93.65%	**98.24%**
	AA	88.38%	96.80%	97.75%	94.79%	89.46%	96.09%	**98.13%**
	MI	0.3965	0.0964	**0.0344**	0.3358	0.1703	0.1964	0.0767
	EES	6.02E+05	8.83E+04	**7.72E+04**	4.04E+05	1.79E+05	1.40E+05	8.07E+04
	MC	–	6.26E+06	9.69E+06	3.70E+07	9.85E+06	1.08E+07	**4.73E+06**
Salina	OA	88.50%	97.97%	97.62%	97.50%	95.70%	98.88%	**99.42%**
	Kappa	87.57%	97.76%	97.37%	97.23%	95.23%	98.76%	**99.36%**
	AA	94.33%	98.95%	98.30%	98.27%	95.25%	99.10%	**99.31%**
	MI	0.5194	0.1645	0.1256	0.4190	0.1734	0.3865	**0.0109**
	EES	4.07E+05	5.57E+04	5.51E+04	2.43E+05	5.77E+04	1.03E+05	**3.28E+04**
	MC	–	6.96E+05	6.81E+05	9.13E+05	6.94E+05	7.08E+05	**6.67E+05**
Hainan	OA	93.62%	97.19%	96.45%	97.76%	95.63%	88.48%	**97.88%**
	Kappa	92.28%	96.56%	95.66%	97.25%	94.67%	85.82%	**97.39%**
	AA	93.42%	96.83%	96.16%	**97.73%**	94.88%	89.04%	96.44%
	MI	0.2286	−0.0068	0.0805	0.0766	0.0033	0.0125	**−0.0362**
	EES	1.36E+04	7.03E+01	1.12E+02	3.27E+02	3.56E+01	4.87E+01	**2.28E+01**
	MC	–	2.30E+07	2.04E+07	3.06E+07	**1.57E+07**	2.76E+07	1.67E+07
Wuhan	OA	89.25%	93.27%	92.81%	**94.69%**	91.90%	85.25%	93.02%
	Kappa	87.40%	91.99%	91.46%	**93.68%**	90.43%	82.07%	91.71%
	AA	88.99%	91.98%	91.82%	**94.48%**	91.35%	81.17%	92.28%
	MI	0.3045	0.0355	0.1485	0.1655	**0.0337**	0.1310	0.1375
	EES	2.92E+04	**5.05E+02**	1.04E+03	2.47E+03	6.53E+02	6.49E+02	1.30E+03
	MC	–	1.74E+07	2.52E+07	5.05E+07	**1.38E+07**	1.86E+07	3.05E+07

most metrics. However, especially, in terms of OA, AA and Kappa, the established improvements seem not significant. This observation should also be attributed to the lack of the reference samples that are located around the object boundaries. However, according to MI, EES and MC, we still can conclude that, the proposed method is superior to the compared methods, especially in boundary areas.

Moreover, the whole maps and local details are also illustrated for further visual inspection [see Figs. 10–12]. In general, the visual gaps among different algorithms are quite obvious. The MRF-based methods exhibit superior performance with regards to the local-based regularization methods, such as EPF and WMV.

When considering local-based regularization methods, the EPF and WMV methods can retain small objects very well, whereas some land-cover classes with large size cannot be identified completely, which can be attributed to the fixed size of the filter kernel when sliding over the image plane. Small objects can benefit from small kernels, but large objects may need large kernels to properly characterize spatial context. Therefore, the main weakness of the

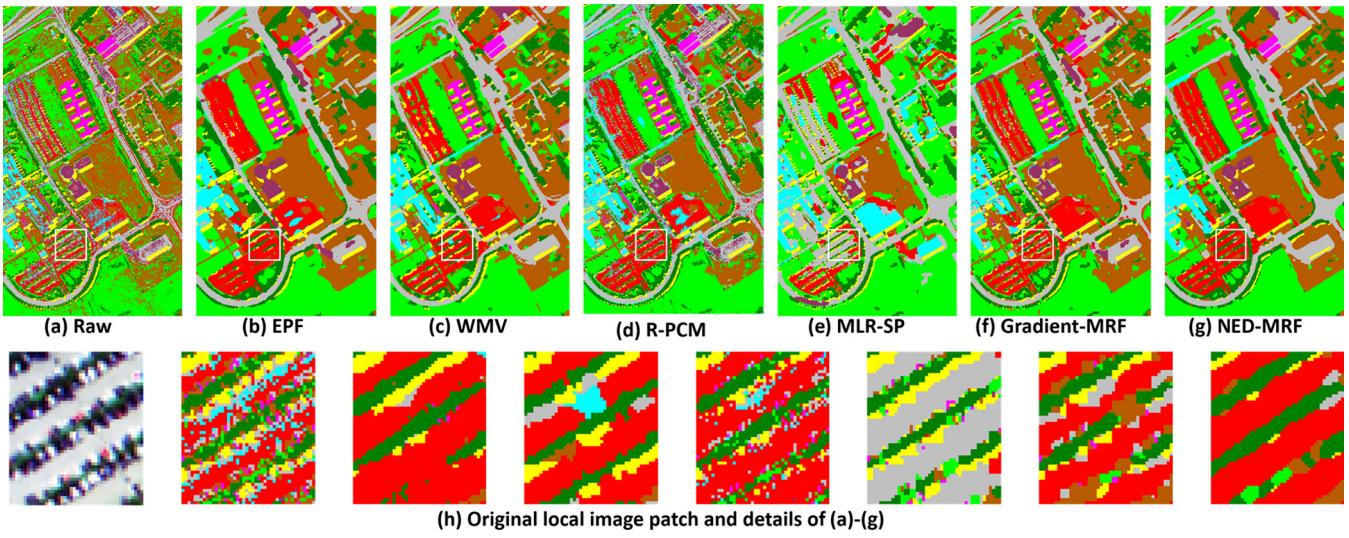


Fig. 10. Classification maps for different optimization algorithms with the *University* data set. The corresponding quantitative statistics are presented in Table 7. Noteworthy details include: visual improvements of all algorithms over the initial map (a), geometrically less correct class boundaries in local optimization methods (b) and (c), the oversmoothness of the classic spatial model (e), noisy pixels distributed along class boundaries in (d) and (f), and visibly superior performance of NED-MRF (g).

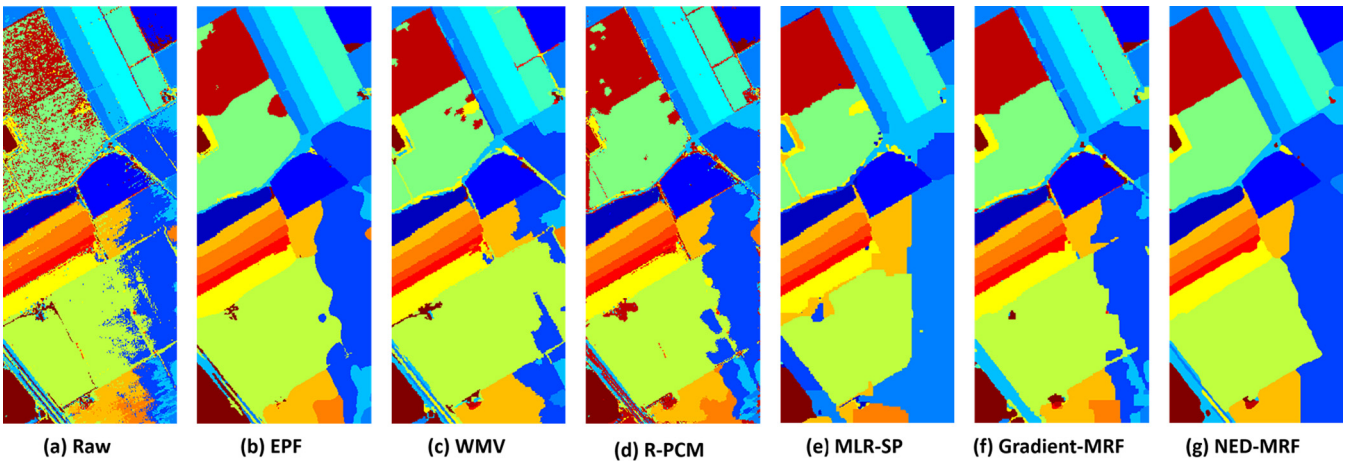


Fig. 11. Classification maps for different optimization algorithms with the *Salinas* data set. The corresponding quantitative statistics are presented in Table 7.

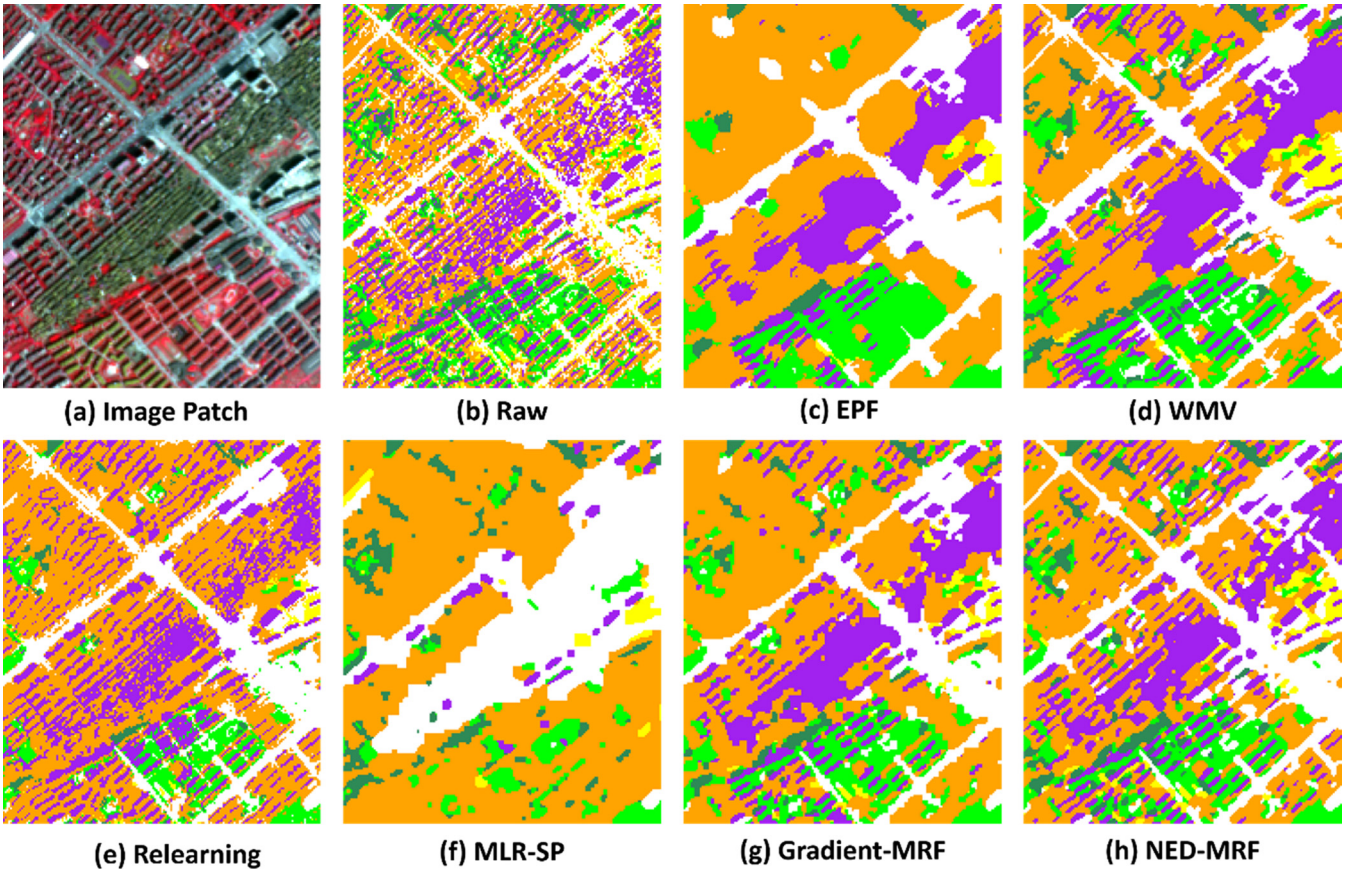


Fig. 12. (a) A selected patch of the *Wuhan* image represented as a three-band false color composite. (b) Initial classification result. (c)–(h) Outputs of different optimization algorithms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

local filtering-based methods is that a single scale parameter (i.e., the kernel size in EPF, and the minimum of region size in WMV) cannot preserve land-cover objects with different sizes.

The global optimization methods perform differently near the object boundaries. As shown in Fig. 10(h), "cars" pixels are mixed with "tree" pixels. As "car" is an uninterested land class, we hope it can be filtered, meanwhile, the boundaries of the "tree" class are still well preserved. This goal only can be achieved by the proposed NED-MRF method. A similar conclusion also can be drawn from Figs. 11 and 12. As for MLR-SP [see Fig. 12(e)], overall high accuracies are achieved at the cost of introducing some distortions of object boundaries. the MLR-SP method embeds an isotropic spatial model. This model forcedly labelled neighboring pixels with the same class labels, even for the pixels located close to the boundaries. On the contrary, aiming at edge preservation, the Gradient-MRF method integrated gradient information into the spatial energy function. As shown in Figs. 10(f) and 11(f), the output results benefit from this integration, but still suffered from the presence of noisy pixels near the boundaries. This problem is attributed to the inadequate exploration of spatial context infor-

mation: the pair-site interaction of the Gradient-MRF method only incorporates the gradient information in a small neighborhood (typically, 4- or 8-connected neighborhoods), rather than spatial relationships among different land classes. The proposed model solves this problem by incorporating global class co-occurrence dependency into the spatial energy function. As shown in Fig. 10 (g), the NED-MRF method yields the most complete object characterizations, while effectively smoothing noisy class labels around object boundaries.

The *Wuhan* data set, which contains dense buildings with a slightly lower spatial resolution (5.8 m), is presented as a special case in Table 7, where the relearning method achieved the highest OA, Kappa and AA. To further explore this case, patches of the results, as well as the image, are illustrated in Fig. 12. As can be seen, different optimization methods are effective for smoothing the raw map to different extents. However, EPF, WMV, Gradient-MRF and MLR-SP produced oversmoothed shadows and narrow roads when achieving their own highest accuracies. In contrast, the relearning method generated a quite noisy classification map and a pepper-and-salt effect still can be observed. This limitation may be caused by the absence of spatial correlation description between a classified pixel and a specific surrounding pixel. The method sums up multidirectional spatial correlation matrices into one matrix, and tries to iteratively relearn this matrix. As this matrix does not carry directional information, the relearning process cannot guarantee that a new class assignment on neighboring pixels obeys a proper spatial relationship. The best results according to visual inspection are produced by NED-MRF. Although this method exhibited a slightly lower overall accuracy, the structures and shape of small and narrow objects (e.g., shadows and stand-alone buildings) are well preserved.

Another interesting observation from Fig. 12 and Fig. 11 is that, although the raw maps are quite noisy, there are no large misclassified regions. As the proposed method only considered interrelationships between sites maintained by a 8-neighborhood system, if such large misclassified regions exist, the proposed method still can produce smoothed regions, but cannot reconstitute their true class labels.

Based on the above assessment, we can reach the following conclusions: the smoothness prior (i.e., the class labels in a neighborhood of the image lattice generally do not change abruptly) is a basic premise to optimize the classification results. When considering different optimization schemes, the global methods, e.g., MRF-based methods, can outperform the local methods such as EPF and WMV. Although a simple spatial prior model in the MRF-based methods (i.e., MLL) already yielded satisfactory enhancements in terms of overall accuracy, more sophisticated models capable of dealing with spectral discontinuities and class co-occurrences are still required in order to avoid oversmoothing. The new spatial model integrating both spectral dissimilarity and class co-occurrence dependency provided by this exhibits superior performance in the task of regularizing pixels and avoiding oversmoothness.

6. Conclusions and future lines

Markov random fields (MRFs) provide a flexible theoretical framework for of optimizing noisy classification maps in remote sensing imagery. The classic MRF model assumes uniform smoothness over the whole image lattice. This assumption is often violated at the image areas exhibiting discontinuities, e.g., pixels around object boundaries. With the ultimate goal of adopting the smoothness prior while preserving discontinuities properly, this paper proposes a novel two-step MRF method for remote sensing image post-classification. By integrating local spectral dissimilarity, the first step of the proposed method can preserve discontinuities and characterize the spatial relationships among different classes. By taking the class co-occurrence dependency into account, the objective of the second step is to smooth the noisy pixels around object boundaries. The proposed method is tested using four hyperspectral/multispectral data sets. The experimental results allowed us to reach the following conclusions:

(1) Among the four considered spectral metrics, the normalized Euclidean distance (NED) achieves the best performance in most cases. The results are quite robust against different landscape scenes and sensor types.
(2) A initial pixelwise classification is expected to provide a map with small misclassified regions. Large misclassified regions can still be smoothed, but cannot be corrected into the true labels, as the proposed method is built on a small second-order neighborhood system.
(3) In general, global optimization methods are superior to local optimization methods. Among all global methods, the proposed NED-MRF method achieves the best performance in most cases. Unlike other MRF-based methods that exhibit oversmoothing or noisy boundaries while achieving the highest classification accuracies, the proposed method can not only properly smooth the initial classification and eliminate the salt-and-pepper effect, but also significantly improve the classification accuracy.

The previous observations suggest the effectiveness of the proposed method in the task of post-classification of remotely sensed images. In the future, we will pursue efficient implementations of the proposed method using high performance computing architectures. We will also provide additional comparisons using different types of remote sensing images.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grants 41571372, 91338111 and 41301470, by the China National Science Fund for Excellent Young Scholars under Grant 41522110, by the Foundation for the Author of National Excellent Doctoral Dissertation of PR China under Grant 201348, and in part by the Canada Research Chairs Program. The authors wish to thank Prof. Paolo Gamba, University of Pavia, Italy, for kindly providing the ROSIS data set. The authors also would like to thank the editors and the reviewers for their insightful comments and suggestions.

References

Aghighi, H., Trinder, J., Tarabalka, Y., Samsung, L., 2014. Dynamic block-based parameter estimation for MRF classification of high-resolution images. IEEE Geosci. Remote Sens. Lett. 11, 1687–1691.
Büschenfeld, T., Ostermann, J., 2012. Edge preserving land cover classification refinement using mean shift segmentation. In: The 4th GEOBIA, Rio De Janeiro – Brazil, pp. 242–247.
Benediktsson, J.A., Palmason, J.A., Sveinsson, J.R., 2005. Classification of hyperspectral data from urban areas based on extended morphological profiles. IEEE Trans. Geosci. Remote Sens. 43, 480–491.
Besag, J., 1986. On the statistical analysis of dirty pictures. J. Roy. Stat. Soc.: Ser. B (Methodol.) 48, 259–302.
Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogram. Remote Sens. 65, 2–16.
Borsotti, M., Campadelli, P., Schettini, R., 1998. Quantitative evaluation of color image segmentation results. Pattern Recogn. Lett. 19, 741–747.
Boykov, Y., Kolmogorov, V., 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans. Pattern Anal. Mach. Intell. 26, 1124–1137.
Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23, 1222–1239.

Chang, C.-I., 2000. An information theoretic-based approach to spectral variability, similarity and discriminability for hyperspectral image analysis. IEEE Trans. Inf. Theory 46, 1927–1932.
Chen, Y., Jiang, H., Li, C., Jia, X., 2016. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 54, 6232–6251.
Chen, Y., Lin, Z., Zhao, X., Wang, G., Gu, Y., 2014. Deep learning-based classification of hyperspectral data. IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens. 7, 2094–2107.
Cheng, G., Han, J., 2016. A survey on object detection in optical remote sensing images. ISPRS J. Photogram. Remote Sens. 117, 11–28.
Dalla Mura, M., Villa, A., Benediktsson, J.A., Chanussot, J., Bruzzone, L., 2011. Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis. IEEE Geosci. Remote Sens. Lett. 8, 542–546.
Derin, H., Elliott, H., 1987. Modeling and segmentation of noisy and textured images using gibbs random fields. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-9, 39–55.
Du, Y., Chang, C.-I., Ren, H., Chang, C.-C., Jensen, J.O., D'Amico, F.M., 2004. New hyperspectral discrimination measure for spectral characterization. Opt. Eng. 43, 1777–1786.
Espindola, G.M., Camara, G., Reis, I.A., Bins, L.S., Monteiro, A.M., 2006. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. Int. J. Remote Sens. 27, 3035–3040.
Fauvel, M., Tarabalka, Y., Benediktsson, J.A., Chanussot, J., Tilton, J.C., 2013. Advances in spectral-spatial classification of hyperspectral images. Proc. IEEE 101, 652–675.
Foody, G.M., 2004. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. Photogram. Eng. Remote Sens. 70, 627–633.
Haralick, R.M., Shanmugam, K., Dinstein, I.H., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. SMC-3, 610–621.
Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507.
Huang, X., Lu, Q., Zhang, L., Plaza, A., 2014. New postprocessing methods for remote sensing image classification: a systematic study. IEEE Trans. Geosci. Remote Sens. 52, 7140–7159.
Kang, X., Li, S., Benediktsson, J.A., 2014. Spectral-spatial hyperspectral image classification with edge-preserving filtering. IEEE Trans. Geosci. Remote Sens. 52, 2666–2677.
Li, J., Bioucas-Dias, J.M., Plaza, A., 2012. Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields. IEEE Trans. Geosci. Remote Sens. 50, 809–823.
Li, J., Reddy Marpu, P., Plaza, A., Bioucas-Dias, J.M., Atli Benediktsson, J., 2013. Generalized composite kernel framework for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 51, 4816–4829.
Li, S.Z., 2009. Markov Random Field Modeling in Image Analysis. Springer-Verlag London Limited.
Martin, D., Fowlkes, C., Tal, D., Malik, J., 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings. Eighth IEEE International Conference on Computer Vision, 2001. ICCV 2001, pp. 416–423.
Martin, D.R., Fowlkes, C.C., Malik, J., 2004. Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Trans. Pattern Anal. Mach. Intell. 26, 530–549.
Moser, G., Serpico, S.B., Benediktsson, J.A., 2013. Land-cover mapping by Markov modeling of spatial-contextual information in very-high-resolution remote sensing images. Proc. IEEE 101, 631–651.
Robila, S.A., Gershman, A., 2005. Spectral matching accuracy in processing hyperspectral data. International Symposium on Signals, Circuits and Systems (ISSCS 2005), vol. 161, pp. 163–166.
Rutzinger, M., Rottensteiner, F., Pfeifer, N., 2009. A comparison of evaluation techniques for building extraction from airborne laser scanning. IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens. 2, 11–20.
Scarpa, G., Gaetano, R., Haindl, M., Zerubia, J., 2009. Hierarchical multiple Markov chain model for unsupervised texture segmentation. IEEE Trans. Image Process. 18, 1830–1843.
Schindler, K., 2012. An overview and comparison of smooth labeling methods for land-cover classification. IEEE Trans. Geosci. Remote Sens. 50, 4534–4545.
Tarabalka, Y., Fauvel, M., Chanussot, J., Benediktsson, J.A., 2010. SVM-and MRF-based method for accurate classification of hyperspectral images. IEEE Geosci. Remote Sens. Lett. 7, 736–740.
Walter, V., 2004. Object-based classification of remote sensing data for change detection. ISPRS J. Photogram. Remote Sens. 58, 225–238.
Wang, L., Liu, G., Dai, Q., 2014. Optimization of segmentation algorithms through mean-shift filtering preprocessing. IEEE Geosci. Remote Sens. Lett. 11, 622–626.
Wu, T.-F., Lin, C.-J., Weng, R.C., 2004. Probability estimates for multi-class classification by pairwise coupling. J. Mach. Learn. Res. 5, 975–1005.
Yu, Q., Clausi, D.A., 2008. IRGS: image segmentation using edge penalties and region growing. IEEE Trans. Pattern Anal. Mach. Intell. 30, 2126–2139.
Zhang, G., Jia, X., 2012. Simplified conditional random fields with class boundary constraint for spectral-spatial based remote sensing image classification. IEEE Geosci. Remote Sens. Lett. 9, 856–860.
Zhang, L., Huang, X., Huang, B., Li, P., 2006. A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery. IEEE Trans. Geosci. Remote Sens. 44, 2950–2961.
Zheng, C., Wang, L., Chen, R., Chen, X., 2013. Image segmentation using multiregion-resolution MRF model. IEEE Geosci. Remote Sens. Lett. 10, 816–820.
Zhong, P., Wang, R., 2010. Learning conditional random fields for classification of hyperspectral images. IEEE Trans. Image Process. 19, 1890–1907.