# Applied Remote Sensing

2222222

RemoteSensing.SPIEDigitalLibrary.org

## Urban area detection from highspatial resolution remote sensing imagery using Markov random fieldbased region growing

Chen Zheng Leiguang Wang Hui Zhao Xiaohui Chen



### Urban area detection from high-spatial resolution remote sensing imagery using Markov random field-based region growing

Chen Zheng,<sup>a</sup> Leiguang Wang,<sup>b,\*</sup> Hui Zhao,<sup>c</sup> and Xiaohui Chen<sup>d</sup>

 <sup>a</sup>Henan University, School of Mathematics and Information Sciences, Jinming Road, Kaifeng 475000, China
 <sup>b</sup>Southwest Forestry University, School of Forestry, Panlong District, Kunming 650224, China
 <sup>c</sup>Henan University, School of Software, Jinming Road, Kaifeng 475000, China
 <sup>d</sup>Henan University, Library, Jinming Road, Kaifeng 475000, China

Abstract. Dynamically changing urban areas require periodic automatic monitoring, but urban areas include various objects and different objects show diverse appearances. This makes it difficult to effectively detect urban areas. A region-growing method using the Markov random field (MRF) model is proposed for urban detection. It consists of three modules. First, it provides an automatic urban seed objects extraction approach by designing three features with respect to urban characteristics. Second, the method uses an object-based MRF to model the spatial relationship between urban seed objects and surrounding objects. Third, a MRF-based region-growing criterion is proposed to detect urban areas based on seed points and spatial constraints. The strength of the proposed method lies in two aspects. One is that automatic selection of seed points is presented instead of manual selection. The other one is that the region-growing technique, instead of probabilistic inference, is used to solve the MRF optimization problem. Experiments on aerial images and SPOT5 images demonstrate that our method provides a better performance compared with the region-growing method, the classical and object-based MRF methods, or some other state-of-art methods. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10 .1117/1.JRS.8.083566]

**Keywords:** remote sensing images; granularity information; Markov random field; region growing; urban area detection.

Paper 14092 received Feb. 11, 2014; revised manuscript received Jun. 27, 2014; accepted for publication Jul. 24, 2014; published online Aug. 15, 2014.

#### **1** Introduction

In recent years, urban detection has become more and more crucial for many applications. It helps government agencies and urban region planners in updating the geographic information system and forming plans. Moreover, due to an enormous number of human activities, the scope of urban areas quickly changes from time to time. Considering the conflict between the need for periodically detecting urban areas and the high-human cost, many approaches had been proposed to automatically detect urban areas from remote sensing images.<sup>1–8</sup> However, an urban area is an abstract semantic object. It is a comprehensive region including several subobjects such as buildings, roads, trees, water bodies, grass spaces, etc. This means that classical spectral-based recognition methods cannot be simply transferred to extract urban areas. Hence, besides spectral value, features that are more effective are needed for urban detection. Since urban scenes usually have a unique texture with respect to natural scenes, texture analysis becomes one main approach for urban monitoring.<sup>9–11</sup> However, the texture pattern of urban scenes is not consistent in all kinds of areas. Methods of texture analysis may suffer from a lack of robustness. In order to answer this problem, several methods have been studied.

<sup>\*</sup>Address all correspondence to: Leiguang Wang, E-mail:wlgbain@gmail.com

For instance, Benediktsson et al.<sup>1</sup> adopted morphological transformations to extract features of urban areas and classify them using a neural network. Weizman and Goldberger<sup>12</sup> built a visual dictionary to learn the urban visual words and then detected the urban regions based on the dictionary. Sirmacek and Ünsalan<sup>13</sup> employed the local feature points extracted by the Gabor filter to vote for the candidate urban areas. Furthermore, Kajimoto and Susaki<sup>14</sup> and Liu et al.<sup>15</sup> extracted the urban areas from polarimetric SAR images using the polarization orientation angle and only positive samples, respectively. However, algorithms may have less transferability with respect to different urban characteristics, as no single-feature descriptor is available for all kinds of the urban objects.

On the contrary, some subobjects that consist of a typical urban pattern can be well detected according to their own characteristics. For instance, man-made objects, such as buildings<sup>6,7,16</sup> and roads,<sup>17-19</sup> usually have compact shapes. In contrast, spectral features are important for detecting natural objects, e.g., vegetations<sup>20,21</sup> and water bodies.<sup>21</sup> Hence, an alternative way of urban detection is to first detect some urban subobjects and then extract the entire urban area based on the extracted subobjects. The region-based classification is a widely used approach to detect certain land cover objects.<sup>22-25</sup> However, different urban areas may consist of different subobjects. Meanwhile, some subobjects, such as trees and water bodies, may appear in both urban areas and the nonurban areas. This phenomenon makes the region-based urban detection methods challenging, even though each urban subobject can be accurately classified. As urban objects are spatially adjacent, one possible way to answer this problem is to take the spatial information of objects into account. The Markov random field (MRF)<sup>26</sup> model provides a statistical way to model spatial contextual information, and it has been extended to the region level for image classification. <sup>23–25</sup> For example, Wu et al.<sup>23</sup> used some rectangular regions as the initial objects and then classified the polarimetric SAR images using the Wishart MRF. However, the accuracy of classification is still limited when the rectangular region is located on the edge of some objects. Zhang et al.<sup>24</sup> improved this method by using a mean shift to obtain the finer initial regions. Wang and Zhang<sup>25</sup> used the Gaussian distribution to recognize images instead of the Wishart distribution. Although these MRF-based classification approaches usually obtained remarkable results, they assumed that each land class obeyed a certain probability distribution, e.g., the Wishart or Gaussian distribution. Nevertheless, the assumption about the probability distribution does not hold in the case of detecting urban areas, as urban areas are often represented as complex regions with various subobjects. Using the probabilistic inference of the MRF model in terms of common probability distributions cannot appropriately detect urban areas.

Motivated by this observation, this paper proposes an MRF-based region-growing method to extract urban areas. Our main contributions include two aspects. First, the proposed method introduces a new MRF-based region-growing criterion to overcome the limitation of the traditional probabilistic inference way of the MRF model. The method retains the advantages of the MRF model in the description of the regional spatial constraints. Both the spatial constraints and the characteristic of urban areas are considered to design a region-growing criterion. Second, an automatic seed objects extraction method is proposed for the MRF-based region growing. The method automatically extracts three features to describe the spectral and granularity information and uses these three features to detect buildings and their shadows as seed points. Our method provides an unsupervised way to detect urban areas, which makes it possible to capture the correlations among various urban objects by combining the benefits of region growing and the MRF model.

The rest of this paper is organized as follows. Section 2 introduces the method for initializing seeds, and Sec. 3 presents the details of the MRF-based region-growing method. Section 4 discusses the results obtained by applying our method on remote sensing images. Finally, Sec. 5 draws a conclusion.

#### 2 Selection of Seed Points

The selection of seed points is a fundamental step for a region-growing algorithm. The main concept of the selection of seed points is grounded in the observation that the buildings are

located in every corner of the city and are often adjacent to shadow areas. Hence, we extract them and their shadows as seed points in this section. In order to appropriately detect seed points, we will first explore three features  $F^1$ ,  $F^2$  and  $F^3$ . The details are given in the following sections.

#### 2.1 Extract the Pixel-Level Spectral Value F<sup>1</sup>

Because buildings usually show a bright appearance in an image and their shadows are dark, a spectral value  $F^1$  is used to describe this feature. Namely, for a given image  $Y = (Y^1, Y^2, ..., Y^P)$ , each spectral channel  $Y^t$   $(1 \le t \le P)$  is defined on an  $M \times N$  rectangular lattice S, i.e.,  $S = \{s | s = (i, j), 1 \le i \le M, 1 \le j \le N\}$  and  $Y^t = (y_s^t)_{M \times N}$ . Then, spectral value  $F^1 = (f_s^1)_{M \times N}$  is defined as  $f_s^1 = \prod_{t=1}^P y_s^t$ , which can describe the spectral value of each pixel s on different channels.

#### 2.2 Extract the Region-Level Spectral Variance F<sup>2</sup>

Different urban objects have various appearances, so their spectral variance should be relatively large. Hence, we design a region-level spectral variance  $F^2$  to capture this feature. First, the initial objects are obtained using a mean shift method,<sup>27</sup> which constructs a probability density to reflect the underlying distribution of points in some feature space and to map each point to the mode of the density which is closest to it. Then, the given image Y is divided into an oversegmented region set  $\mathbf{R}$ , i.e.,  $\mathbf{R} = \{R_1, R_2, \ldots, R_k\}$ . Each  $R_i$  of  $\mathbf{R}$  denotes an over-segmented region  $(i = 1, 2, \ldots, k), R_i \cap R_j = \emptyset$   $(i \neq j)$ , and k is the number of these regions. With the region set  $\mathbf{R}$ , we can further define the neighborhood system  $N = \{N_i | i = 1, 2, \ldots, k\}$  to describe the spatial context of regions. Here, each  $N_i$  denotes the set of regions neighboring  $R_i$ . Let  $M(R_i)$  be the mean value of pixels in  $R_i$ , and the local spectral variance between region  $R_i$  and its adjacent regions can be calculated as follows:

$$V(R_i) = \frac{1}{|N_i|} \left\{ [M(R_i) - \mu_i]^2 + \sum_{j \in N_i} [M(R_j) - \mu_i]^2 \right\},\tag{1}$$

where  $\mu_i = (1 + |N_i|)^{-1} [M(R_i) + \sum_{j \in N_i} M(R_j)]$ , and  $|N_i|$  is the number of regions in  $N_i$ .

In Eq. (1), every region has the same impact on  $V(R_i)$ . Intuitively, it may be preferable to determine the impacts in Eq. (1) using an adaptive way. Hence, the equation for  $V(R_i)$  is revised as

$$V(R_i) = \frac{1}{|N_i|} \left\{ [M(R_i) - \mu_i^*]^2 + \sum_{j \in N_i} [M^*(R_j, R_i) - \mu_i^*]^2 \right\}.$$
 (2)

In Eq. (2),  $M^*(R_i, R_i)$  is defined as follows:

$$M^*(R_j, R_i) = \begin{cases} M(R_j) & \text{if } \alpha > p \\ \alpha M(R_j) + (1 - \alpha)M(R_i) & \text{if } \alpha \le p \end{cases}$$

where  $\alpha = |R_j| \cdot |R_i|^{-1}$  and  $\mu_i^* = (1 + |N_i|)^{-1} [M(R_i) + \sum_{j \in N_i} M^*(R_j, R_i)]$ , region size  $|R_i|$  is the number of pixels in region  $R_i$ . In this revised equation, the impact of  $R_j$  will be reduced when the ratio of  $|R_j|$  to  $|R_i|$  is less than p. That is to say, the effect of each region is affected by its region size. Here, p is a threshold with which to measure the ratio between the sizes of two regions. Since the relationship of region sizes among different objects is relatively stable, we empirically set p as 0.3 in this paper.

Based on the  $V(R_i)$ ,  $F^2 = (f_s^2)_{M \times N}$  is defined as  $f_s^2 = V[R(s)]^{1/2}$  to reflect the spectral variance among regions. Here, R(s) is the region to which pixel s belongs.

#### **2.3** Extract the Granularity Information $F^3$

Urban areas have more different types of objects and more complicated appearances than nonurban areas. Therefore, in the over-segmented region set  $\mathbf{R} = \{R_1, R_2, \dots, R_k\}$ , objects of urban areas usually have smaller region sizes than objects of nonurban areas. In other words, the granularity of urban areas is finer than that of nonurban areas. Hence, we employ the region size and the spatial relationship among regions to define  $F^3 = (f_s^3)_{M \times N}$ , i.e.,

$$f_s^3 = (P[R(s)] - P[R(s)] \cdot \log\{P[R(s)]\}) + \frac{1}{|N_{R(s)}|} \sum_{j \in N_{R(s)}} \{P(R_j) - P(R_j) \cdot \log[P(R_j)]\}, \quad (3)$$

where  $P[R(s)] = |R(s)| \cdot (M \cdot N)^{-1}$ . In the above equation,  $P[R(s)] - P[R(s)] \cdot \log\{P[R(s)]\}$  is used to reflect the region size of R(s) and  $\sum_{j \in N_{R(s)}} \{P(R_j) - P(R_j) \cdot \log[P(R_j)]\}$  is used to describe the context information of regions. Note that  $f(x) = x - x \cdot \log(x)$  is a monotonically increasing convex function when  $x \in [0,1]$ . Hence, the monotonicity of f(x) can make  $P[R(s)] - P[R(s)] \cdot \log\{P[R(s)]\}$  to indicate the region size. What is more, if we assume that  $|N_{R(s)}|$  is fixed, the convexity of f(x) can make  $\sum_{j \in N_{R(s)}} \{p(R_j) - p(R_j) \cdot \log[p(R_j)]\}$  take a small value when the sizes of regions neighboring R(s) are close. It will lead to a consistent result with a smooth region size, which is suitable for capturing the granularity information since the granularities of regions are usually similar for one certain object.

An example to illustrate these features is shown in Fig. 1, where Figs. 1(b), 1(d), and 1(e) are features  $F^1$ ,  $F^2$  and  $F^3$  extracted from Fig. 1(a). From this example, one can see that the buildings in Fig. 1(b) are bright, which denotes a high  $F^1$  value, and their shadows are of the low  $F^1$  value. Similarly, the spectral variance  $F^2$  of urban areas is larger than that of others areas, and urban areas have a small granularity  $F^3$  value. Based on these features, we design  $E^1 = (e_s^1)_{M \times N}$ ,  $E^2 = (e_s^2)_{M \times N}$ ,  $E^3 = (e_s^3)_{M \times N}$ , and  $E^4 = (e_s^4)_{M \times N}$  to describe the buildings' spectral values, dark shadows' spectral values, regional spectral variance, and granularity information, respectively. They are

$$e_s^1 = \begin{cases} 1 & f_s^1 > F_{\gamma}^1 \\ 0 & \text{otherwise} \end{cases}, \quad e_s^2 = \begin{cases} 1 & f_s^1 \\ 0 & \text{otherwise} \end{cases},$$
$$e_s^3 = \begin{cases} 1 & f_s^2 > F_{\lambda}^2 \\ 0 & \text{otherwise} \end{cases}, \quad e_s^4 = \begin{cases} 1 & f_s^3 \\ 0 & \text{otherwise} \end{cases},$$

where  $F_{\gamma}^1$ ,  $F_{1-\gamma}^1$ ,  $F_{\lambda}^2$ , and  $F_{\pi}^3$  denote the  $\gamma$ ,  $1 - \gamma$ ,  $\lambda$ , and  $\pi$  fractile of  $F^1$ ,  $F^2$  and  $F^3$ , respectively.

 $\gamma, \lambda$ , and  $\pi$  are the key parameters for the selection of seed points. The parameter  $\gamma$  is used to make  $E^1$  capture the spectral feature of buildings. Since buildings usually take a high spectral value, they are often expressed as the tail of the histogram of  $F^1$ . Hence,  $\gamma$  is set to a high value to get the tail of the histogram of  $F^1$ , such as Fig. 1(f). Correspondingly,  $E^2$  uses  $1 - \gamma$  to obtain the first peak of the histogram of  $F^1$ , which describes the dark shadows with a low  $F^1$  value.



**Fig. 1** (a) Original aerial image. (b) Pixel-level spectral value. (c) Initial over-segmented region set  $\mathbf{R}$ . (d) Region-level spectral variance. (e) Granularity information. (f) Histogram of  $\mathbf{F}^1$  and  $\gamma$  (g) Histogram of  $\mathbf{F}^2$  and  $\lambda$ . (h) Histogram of  $\mathbf{F}^3$  and  $\pi$ . (i) Seed points extracted based on (b), (d), and (e).

For the same reason,  $E^3$  and  $E^4$  are set with a high  $\lambda$  value and low  $\pi$  value to catch the tail of the histogram of  $F^2$  and the first peak of the histogram of  $F^3$ , respectively. These can extract buildings' spectral variance and granularity features. An illustration of setting  $\gamma$ ,  $\lambda$ , and  $\pi$  is shown in Figs. 1(f)–1(h).

Then, by sequentially combining  $E^1$ ,  $E^2$ ,  $E^3$ , and  $E^4$ , seed points can be obtained. Namely, we first use  $D^1 = (d_s^1)_{M \times N}$  to get pixels belonging to buildings and adjoining the shadows, or pixels belonging to shadows and adjoining the buildings. This is defined as

$$d_s^1 = \begin{cases} 1 & \text{if } e_s^1 = 1 \& \sum_{t \in w(s,r)} e_s^2 \ge l \text{ or } e_s^2 = 1 \& \sum_{t \in w(s,r)} e_s^1 \ge l \\ 0 & \text{otherwise} \end{cases},$$

where the local square window w(s, r) is centered at site s and its radius is r. Then, we further consider the information of  $E^3$  and  $E^4$  by defining  $D^2 = (d_s^2)_{M \times N}$  and  $D^3 = (d_s^3)_{M \times N}$  as

$$d_{s}^{2} = \begin{cases} 1 & \text{if } d_{s}^{1} = 1 \& \sum_{t \in w(s,r)} e_{s}^{3} \ge l \text{ or } e_{s}^{3} = 1 \& \sum_{t \in w(s,r)} d_{s}^{1} \ge l \\ 0 & \text{otherwise} \end{cases}, \\ d_{s}^{3} = \begin{cases} 1 & \text{if } d_{s}^{2} = 1 \& \sum_{t \in w(s,r)} e_{s}^{4} \ge l \text{ or } e_{s}^{4} = 1 \& \sum_{t \in w(s,r)} d_{s}^{2} \ge l \\ 0 & \text{otherwise} \end{cases}.$$

At last, seed points will be selected as the set  $D = \{s | d_s^3 = 1, s \in S\}$ .

For r and l, these seed points are used to determine whether a local window w(s, r) simultaneously contains pixels from  $E^1$ ,  $E^2$ ,  $E^3$ , and  $E^4$  and whether pixels of each kind are not less than l. Because a building is spatially adjacent to its shadow, they can be effectively detected together using a relative small patch of the given image. Hence, by setting r to 2 for  $D^1$ ,  $D^2$ , and  $D^3$ , we use the local window w(s, r = 2) as the small patch to select seed points in the following. At the same time, if there are buildings and their shadows in the small patch, there will be at least one pixel labeled 1 in the patch for each  $e_s^i$ , i = 1, 2, 3, 4. Therefore, l is set to 1. It means that only a pixel which simultaneously possesses or neighbors  $E^1$ ,  $E^2$ ,  $E^3$ , and  $E^4$  within the small local window w(s, 2) can be chosen as the seed point. An example is shown in Fig. 1(i). Note that one pixel would show different sizes of the Earth's surface in remote sensing images with various spatial resolutions, which may affect the setting of parameter r. Namely, r can be set to 1 for the low-spatial resolution remote sensing images.

#### 3 MRF-Based Region Growing

Based on extracted seed points, a MRF-based region-growing criterion is proposed in this section. First, the MRF model is briefly reviewed. Then, the proposed criterion for urban detection is introduced.

#### 3.1 MRF Model

Let  $X = \{X_{R_i} | R_i \in \mathbf{R}\}$  be the label random field defined on the over-segmented region set R. We use 1 to flag urban areas and 0 to flag nonurban areas, and each random variable  $X_{R_i}$  takes a value of 1 or 0 to represent the label of region  $R_i$  it belongs to. If  $\mathbf{x} = \{x_{R_i} | R_i \in \mathbf{R}\}$  denotes the realization of X, the optimal realization  $\hat{\mathbf{x}}$  can be obtained by maximizing the posterior probability, i.e.,

$$\hat{x} = \operatorname*{argmax}_{x} P(X|Y) = \operatorname*{argmax}_{x} P(Y|X) \cdot P(X). \tag{4}$$

The energy form of Eq. (4) is

$$\hat{\boldsymbol{x}} = \operatorname{argmin}\{-\log[P(\boldsymbol{Y}|\boldsymbol{X})] - \log[P(\boldsymbol{X})]\}.$$
(5)

In Eq. (5), the likelihood function P(Y|X) is used to describe image features. In this paper, we assume that all  $Y_{R_i}$  of Y are independent given labels. That is

$$P(Y|X) = \prod_{R_i \in R} P(Y_{R_i}|X_{R_i}).$$

The distribution of random field P(X) is assumed to be of the Markovianity property, i.e.,

$$P(X) = \prod_{R_i \in \mathcal{R}} P(X_{R_i} | X_{R_j}, j \in N_i).$$

Therefore, Eq. (5) can be rewritten as

$$\hat{x} = \underset{x}{\operatorname{argmin}} \bigg\{ \sum_{R_i \in R} [-\log P(Y_{R_i} | X_{R_i}) - \log P(X_{R_i})] \bigg\}.$$
(6)

Due to the complexity caused by interactions among labels, it is difficult to find the solution of the MRF model. Hence, the local optimal solution  $\hat{x} = (\hat{x}_{R_i})$  can be obtained as follows:

$$\hat{x}_{R_i} = \underset{x_{R_i}}{\operatorname{argmin}} [-\log P(Y_{R_i}|X_{R_i}) - \log P(X_{R_i})] = \underset{x_{R_i}}{\operatorname{argmin}} [E_f(R_i) + E_l(R_i)], \quad (7)$$

where the likelihood energy  $E_f(R_i)$  is the cost of the observation of  $R_i$ , and the label energy  $E_l(R_i)$  is the cost of the label of  $R_i$ .

#### 3.2 MRF-Based Region Growing

In this section, an MRF-based region-growing criterion is introduced to find the optimal realization  $\hat{x}$ . To minimize the total energy of the MRF model, the proposed method will iteratively merge adjacent regions that could decrease the total energy. Namely, for neighboring regions  $R_i$ and  $R_t$ , the total changed energy  $E(R_i, R_t)$  is first calculated these two regions are merged. Based on Eq. (7),  $E(R_i, R_t)$  equals the sum of the changed likelihood energy  $E_f(R_i, R_t)$  and the changed label energy  $E_l(R_i, R_t)$ , i.e.,

$$E(R_{i}, R_{t}) = E_{f}(R_{i}, R_{t}) + E_{l}(R_{i}, R_{t}).$$
(8)

Here

$$E_f(R_i, R_t) = E_f(R_i \cup R_t) - E_f(R_i) - E_f(R_t)$$
  
=  $|R_i| \cdot [M(R_i) - M(R_i \cup R_t)]^2 + |R_t| \cdot [M(R_t) - M(R_i \cup R_t)]^2,$  (9)

where  $|R_i| \cdot [M(R_i) - M(R_i \cup R_t)]^2$  and  $|R_t| \cdot [M(R_t) - M(R_i \cup R_t)]^2$  can reflect the change of the observations in region  $R_i$  and  $R_t$ , respectively. The changed label energy of  $R_i$  is defined as

$$E_l(R_i, R_t) = -\sum_{j \in N_i} \beta |R_i| V_l(R_t, R_j), \qquad (10)$$

where the pair-clique potential

$$V_l(R_t, R_j) = \begin{cases} 1 & \text{if } x_{R_t} = x_{R_j} \\ 0 & \text{otherwise} \end{cases}.$$

 $E_l(R_i, R_i)$  uses  $|R_i|$  to consider all changed label energies for each pixel in  $R_i$  and its neighbors when  $x_{R_i}$  is relabeled as  $x_{R_i}$ . Then, by merging region  $R_i$  and its neighboring region that can minimize the total changed energy, a MRF-based region-growing approach can realize urban detection step by step. The details of the rule of region growing are given in Algorithm 1.

The proposed criterion is different from traditional region-growing methods, as it does not begin from seed points but from nonseed points. We only consider the nonurban regions labeled 0 and their region sizes are less than the threshold. For each selected region  $R_i$ , the energy values are calculated between  $R_i$  and its neighbor regions, respectively. Then,  $R_i$  is merged with the Zheng et al.: Urban area detection from high-spatial resolution remote sensing imagery...

#### Algorithm 1

Input: the observed image.

Output: urban detection result.

1) Set a threshold T.

2) If there exists a region  $R_i$  satisfying  $|R_i| < T$  and  $x_{R_i} = 0$ , select  $R_i$  and go to step 3; else, stop.

3) For  $R_i$  and its neighbor region  $R_t$ , based on Eqs. (8–10), calculate the total changed energy  $E(R_i, R_t)$ .

 $R_t, t \in N_i$ 

4) Find the region  $R_i^*$  that has the minimum energy value, i.e.,  $R_i^* = \operatorname{argmin} E(R_i, R_t)$ .

Merge  $R_i$  and  $R_i^*$  as a new region labeled  $x_{R_i^*}$ , then go to step 2.

one neighbor region that has the minimum energy value. Hence,  $R_i$  merged with an urban region will lead to a larger urban region; in contrast,  $R_i$  merged with a nonurban region will result a new nonurban region. Therefore, the rule of our approach is a competition rule of region growing for both urban and nonurban regions.

Urban areas can be extracted using the region-growing criterion. Namely, urban areas are first initialized using the label field  $x = \{x_{R_i} | R_i \in R\}$  based on seed points D, i.e., set  $x_{R_i} = 1$  if  $R_i \cap D \neq \emptyset$ ; or else, set  $x_{R_i} = 0$ . Then, by increasing the thresholds, the growing criterion gradually updates the urban areas. Note that different sun angles may affect the shadow length and direction, but it does not change the spatial topological relationship between buildings and their shadows. Hence, the proposed method is robust for effective detection of varying urban areas contained in different remote sensing images.

#### 3.3 Parameter Setting

There are two parameters in the MRF-based region-growing criterion, i.e.,  $\beta$  and T. The potential parameter  $\beta$  is used to balance the influence between  $E_f(R_i, R_t)$  and  $E_l(R_i, R_t)$ . A high  $\beta$  value emphasizes  $E_l(R_i, R_t)$  and leads to results with large homogeneous objects. On the contrary, a low  $\beta$  value emphasizes  $E_f(R_i, R_t)$  and is suitable for getting results with many details. Hence,  $\beta$  should select different values for various applications. However, as the relationship between urban and nonurban areas is quite stable,  $\beta$  is fixed and is empirically set as 0.05 for simplifying the parameter setting.

The threshold *T* is used to control the process of region growing. By gradually increasing *T*, small regions labeled nonurban are merged into larger urban regions or nonurban regions, then urban areas are extracted. In practice, we used T = 25 as the initial threshold and doubled the threshold each time. The final termination threshold was determined by the change of the spectral variance. The assumptions supporting this threshold selection are that urban areas consist of various subobjects and their spectral variance should be large; if the nonurban areas are wrongly recognized as urban areas, an abrupt change of the spectral variance should be observed. Here, we use CR(i, i + 1) to show the change rate of spectral variances, i.e.,

$$\operatorname{CR}(i, i+1) = \frac{\operatorname{Std}T(i+1) - \operatorname{Std}T(i)}{\operatorname{Std}T(i)},\tag{11}$$

where  $\operatorname{Std} T(i)$  denotes the standard deviation of detected urban areas with termination threshold T = T(i). Then, we can take the inflection point of  $\operatorname{CR}(i, i + 1)$  as the final termination threshold, after which  $\operatorname{CR}(i, i + 1)$  will abruptly decrease. An example is shown in Fig. 2, where we use T = [25, 50, 100, 200, 400, 800, 1600] as the candidates of termination thresholds. Some extracted urban areas are illustrated in Figs. 2(a)-2(g).  $\operatorname{Std} T(i)$  with different Ts is calculated and given in Fig. 2(h), where the corresponding  $\operatorname{CR}(i, i + 1)$ s are also shown in Fig. 2(i). As  $\operatorname{CR}(200,400)$  is an inflection point, we take T = 400 as the final termination threshold for this experiment and Fig. 2(e) shows the corresponding detection result.

Zheng et al.: Urban area detection from high-spatial resolution remote sensing imagery...



**Fig. 2** Example of parameter *T*: (a) urban area with T = 25; (b) urban area with T = 50; (c) urban area with T = 100; (d) urban area with T = 200; (e) urban area with T = 400; (f) urban area with T = 800; (g) urban area with T = 1600; (h) StdT(i); and (i) CR(*i*, *i* + 1).

#### **4 Experiments**

The MRF-based region-growing method provides an unsupervised way for the monitoring of urban areas. With the aim of fully evaluating the performance of the proposed method, experiments and comparisons were carried on two groups of images, i.e., aerial images (Sec. 4.1) and SPOT5 images (Sec. 4.2).

#### 4.1 Experiments of Aerial Images

In this experiment, three aerial images, as shown in Fig. 3, are used to test our method and other urban extraction methods. These aerial images were acquired in 2009 and are located in Taizhou City, China. The three images have the same size of  $500 \times 500$ , and the spatial resolution is 0.4 m. The test images contain plane agriculture fields and small villages, where urban objects show various spectral appearances and some nonurban objects are similar to seed points in terms of spectral characteristics. This makes urban detection challenging. Moreover, the following competitive methods are also considered for comparison:

- 1. The traditional region-growing method:<sup>28</sup> it detects urban areas without employing the MRF model.
- 2. The classical MRF model:<sup>29</sup> it uses the generated probabilistic model at the pixel level to obtain results.
- 3. The object-based MRF (OMRF) model:<sup>25</sup> it extends the MRF model from the pixel level to the object level for capturing the macrotexture pattern of a given image; this uses

Zheng et al.: Urban area detection from high-spatial resolution remote sensing imagery...



**Fig. 3** Experiments of aerial images: (a1)–(c1) aerial images; (a2)–(c2) traditional region growing; (a3)–(c3) Markov random field (MRF); (a4)–(c4) object-based MRF (OMRF); (a5)–(c5) support vector machine (SVM); (a6)–(c6) object-based SVM; and (a7)–(c7) MRF-based region growing.

initial over-segmented regions to build the region adjacency graph (RAG) and defines the MRF model on the RAG to realize the segmentation.

- 4. The two-class support vector machine (SVM):<sup>30</sup> it is provided by ENVI software, which is a commonly used classification approach with training data.
- 5. The object-based SVM:<sup>22</sup> it extracts the regional features from a hierarchical tree of the scene and obtains a classification using the SVM classifier.

For the sake of fairness, we chose the same seed points to train the urban areas for the traditional region-growing method and the two SVM methods and deliberately selected samples to train the nonurban areas for these SVM methods as well. We also tuned the parameters of these methods to get their optimal performances. For the traditional region-growing method, we chose the threshold parameter following the instructions in the literature.<sup>28</sup> For the two-class SVM, we set the radial basis function as the kernel type, the gamma in kernel function as 0.33, and the penalty function as 100, respectively. For the object-based SVM, we use 0.1% as the ratio of training samples based on the literature.<sup>22</sup> Therefore, the comparison can demonstrate the difference between our model and other state-of-the-art methods.

Experimental results of aerial images are shown in Fig. 3. Here, the caption of Fig. 3 consists of two parts, where the first part using the alphabetical order denotes different test images and the second part using the number order denotes different detection methods. Detected urban objects are represented as yellow masks over the test images. From the comparative test, one can see that the proposed method exhibits a remarkable improvement for urban detection. Namely, the traditional region-growing method, as shown in Figs.  $3(a_2)-3(c_2)$ , still has huge misclassifications which belong to different object categories and have similarity spectral appearances. The main reason is that the traditional region-growing method only uses the spectral features which do not consider the spatial constraint. By employing the spatial context information, the classical MRF model has less misclassification of nonurban areas. However, this pixel-level generate model can just recognize the parts of the urban areas with similar appearances, since it cannot model the complex and macropatterns by incorporating the long-range interactions. It also wrongly labels some urban objects as nonurban, such as the roofs of buildings and vegetation. The OMRF model utilizes the regions to describe the macrospatial constraints and improves the classical MRF model, e.g., Figs. 3(a4) and 3(c4). However, the OMRF model usually leaves the characteristic of urban areas out of consideration, which may lead to some undesirable results such as Fig. 3(b4). The SVM method trains data to obtain urban areas. Although it can effectively recognize buildings, urban vegetation objects are sometimes classified as nonurban areas because of the lack of spatial information. The object-based SVM improves the pixel-based SVM and gets results that are more consistent by considering the object semantic information with regional features. Nevertheless, it still cannot sufficiently use spatial information whose results have some misclassifications. Compared with these methods, our MRF-based region-growing method first considers the urban characteristics when we select seed points, then employs the MRF defined on the region level to capture regional spatial constraints, and finally proposes a corresponding region-growing criterion that utilizes these features to detect urban areas. Hence, our method demonstrates a better performance than the other methods.

Experimental results are quantitatively evaluated by the overall accuracy (OA) and kappa coefficient  $\kappa$ . OA and  $\kappa$  are the two indicators that measure the degree of similarity between two images.<sup>31</sup> If  $P_{ij}$  is the proportion of subjects that were assigned to the *i*'th class by the first image and the *j*'th class by the second image and denotes  $P_{i\bullet} = \sum_{j=1}^{k} P_{ij}$  and  $P_{\bullet j} = \sum_{i=1}^{k} P_{ij}$ , then

$$OA = \sum_{i=1}^{k} P_{ii}, \text{ and } \kappa = \frac{\sum_{i=1}^{k} P_{ii} - \sum_{i=1}^{k} P_{i\bullet} P_{\bullet i}}{1 - \sum_{i=1}^{k} P_{i\bullet} P_{\bullet i}}.$$

The OA and  $\kappa$  of aerial images are given in Table 1.

From these quantitative indexes, we know that MRF-based region growing can enhance both the OA and kappa for each experimental image. This also shows that our method extracts a better scope of urban areas than do the other methods. In particular, when the topographic features are Zheng et al.: Urban area detection from high-spatial resolution remote sensing imagery...

	Fig. 3(a)		Fig.	Fig. 3(b)		Fig. 3(c)	
	к	OA	κ	OA	κ	OA	
Traditional region growing	0.379	0.591	0.398	0.604	0.489	0.648	
Classical MRF	0.778	0.913	0.460	0.684	0.615	0.758	
OMRF	0.883	0.953	0.663	0.803	0.770	0.863	
Two-class SVM	0.806	0.923	0.617	0.770	0.683	0.796	
Object-based SVM	0.911	0.966	0.740	0.850	0.832	0.905	
MRF-based region growing	0.914	0.967	0.902	0.952	0.886	0.938	

Table 1 Comparison of results.

Note: For each column, the bold value denotes the best index among all the indexes in this column.



Fig. 4 Kappa and overall accuracy (OA) of experiments of aerial images: (a) kappa and (b) OA.



Fig. 5 Experiments of SPOT5 images: (a and d) Original SPOT5 image; (b and e) ground truth (red); and (c and f) results of MRF-based region growing (yellow).

complex, the enhancement of indices is obvious. For clarity, the quantitative indices of Table 1 are illustrated in Fig. 4.

#### 4.2 Experiments of SPOT5 Images

The effectiveness of the proposed method is further tested in this section. Two SPOT 5 remote sensing images, as shown in Fig. 5, are employed for the next experiment. These test images are located on the Pingshuo area of China. Both sizes are  $438 \times 438$ . These test images mainly consist of three object types, i.e., urban areas, cultivated land, and woodland. Among them, urban green space and woodland and urban building and cultivated land have similar spectral appearances, respectively. This phenomenon increases the difficulty of urban detection.

Experiments of SOPT5 images are illustrated in Fig. 5. Compared with the ground truth, the MRF-based region-growing method performs well and the results are close to the ground truth. This demonstrates that our model can effectively extract urban areas from different datasets.

#### **5** Conclusions

To summarize, we proposed an unsupervised urban detection method by unifying the regiongrowing method and the MRF model. It first uses the granularity information and spectral features to automatically extract some typical urban objects as the seed points, which can be treated as the skeleton for the urban areas. Then, the MRF is employed to model the spatial relationships between urban seed points and other urban objects. At last, the region-growing criterion uses these relationships to recognize urban nonseed objects, which will lead to consistent results. The main novelty of the method the automatic extraction of urban seed points and the detection of urban areas using a region-growing criterion under the regional MRF-based spatial constraints. The effectiveness of the proposed method is validated by experimental results obtained from various high-spatial resolution remote sensing images. Compared to a traditional region-growing method, the classical and object-based MRF models, and the common and object-based SVM, our method can provide more precise and more meaningful results, which verifies that our method is suitable to detect urban areas. However, this method is only proper for urban detection. If it is used to extract other terrestrial objects, then one has to design a new seed extraction method and modify the region-growing criterion.

For the method presented, the potential parameter  $\beta$  need to be empirically set. If this parameter can be estimated in an adaptive way, then it will improve the current method.

#### Acknowledgments

The authors are very grateful to the editor and the anonymous referees for comments and suggestions, which led to the present improved version of the manuscript. This work is supported jointly by the National Natural Science Foundation of China, under Grants 41301470, 41001286, 41101425, and 41001251, and the basic research funds for the provincial universities. The authors would like to thank Associate Prof. Tiancan Mei, Wuhan University, China, for kindly providing aerial images.

#### References

- J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.* 41(9), 1940–1949 (2003).
- 2. D. Lu et al., "Detection of urban expansion in an urban-rural landscape with multitemporal QuickBird images," *J. Appl. Remote Sens.* **4**(1), 041880 (2010).
- 3. P. Gamba, M. Aldrighi, and M. Stasolla, "Robust extraction of urban area extents in HR and VHR SAR images," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **4**(1), 27–34 (2011).

- X. Huang, L. Zhang, and P. Li, "Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery," *IEEE Geosci. Remote Sens. Lett.* 4(2), 260–264 (2007).
- C. Corbane et al., "Comparative study on the performance of multiparameter SAR data for operational urban areas extraction using textural features," *IEEE Geosci. Remote Sens. Lett.* 6(4), 728–732 (2009).
- P. Gamba, B. Houshmand, and M. Saccani, "Detection and extraction of buildings from interferometric SAR data," *IEEE Trans. Geosci. Remote Sens.* 38(1), 611–617 (2000).
- B. Sirmacek and C. Ünsalan, "Urban-area and building detection using SIFT keypoints and graph theory," *IEEE Trans. Geosci. Remote Sens.* 47(4), 1156–1167 (2009).
- 8. C. Chen and L. Chang, "Rapid change detection of land use in urban regions with the aid of pseudo-variant features," *J. Appl. Remote Sens.* **6**(1), 063574 (2012).
- P. C. Smits and A. Annoni, "Updating land-cover maps by using texture information from very high-resolution space-borne imagery," *IEEE Trans. Geosci. Remote Sens.* 37(3), 1244– 1254 (1999).
- S. Yu, M. Berthod, and G. Giraudon, "Toward robust analysis of satellite images using map information—application to urban area detection," *IEEE Trans. Geosci. Remote Sens.* 37(4), 1925–1939 (1999).
- 11. G. Rellier et al., "Texture feature analysis using a Gauss-Markov model in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.* **42**(7), 1543–1551 (2004).
- L. Weizman and J. Goldberger, "Urban-area segmentation using visual words," *IEEE Geosci. Remote Sens. Lett.* 6(3), 388–392 (2009).
- 13. B. Sirmacek and C. Ünsalan, "Urban area detection using local feature points and spatial voting," *IEEE Geosci. Remote Sens. Lett.* 7(1), 146–150 (2010).
- 14. M. Kajimoto and J. Susaki, "Urban-area extraction from polarimetric SAR images using polarization orientation angle," *IEEE Geosci. Remote Sens. Lett.* **10**(2), 337–341 (2013).
- Y. Liu et al., "Urban area extraction from polarimetric SAR imagery using only positive samples," in *ICSP Proc.*, pp. 2332–2335, IEEE (2010).
- A. Thiele et al., "Building recognition from multi-aspect high-resolution in SAR data in urban areas," *IEEE Trans. Geosci. Remote Sens.* 45(11), 3583–3593 (2007).
- 17. S. Hinz and A. Baumgartner, "Automatic extraction of urban road networks from multiview aerial imagery," *ISPRS J. Photogramm. Remote Sens.* **58**(1–2), 83–98 (2003).
- Y. He, H. Wang, and B. Zhang, "Color-based road detection in urban traffic scenes," *IEEE Trans. Intell. Transp. Syst.* 5(4), 309–318 (2004).
- 19. J. Hu et al., "Road network extraction and intersection detection from aerial images by tracking road footprints," *IEEE Trans. Geosci. Remote Sens.* **45**(12), 4144–4157 (2007).
- T. Jan, L. Tobia, and H. Patrick, "Urban vegetation classification: benefits of multitemporal Rapid Eye satellite data," *Remote Sens. Environ.* 136(9), 66–75 (2013).
- I. Sebari and D. He, "Automatic fuzzy object-based analysis of VHSR images for urban objects extraction," *ISPRS J. Photogramm. Remote Sens.* 79(5), 171–184 (2013).
- 22. L. Wang et al., "Adaptive regional feature extraction for very high spatial resolution image classification," *J. Appl. Remote Sens.* **6**(1), 061708 (2012).
- Y. Wu et al., "Region-based classification of polarimetric SAR images using Wishart MRF," IEEE Geosci. Remote Sens. Lett. 5(4), 668–672 (2008).
- B. Zhang et al., "Region-based classification by combining MS segmentation and MRF for POLSAR images," J. Syst. Eng. Electron. 24(3), 400–409 (2013).
- X. Wang and X. Zhang, "A new localized superpixel Markov field for image segmentation," in *Proc. IEEE Conf. Multimedia and Expo*, pp. 642–645, IEEE (2009).
- S. Z. Li, Markov Random Field Modeling in Computer Vision, 3rd ed., Springer-Verlag, New York (2009).
- D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* 24(5), 603–619 (2002).
- R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using MATLAB*, Pearson Prentice Hall, Upper Saddle River, New Jersey (2003).
- 29. J. Besag, "On the statistical analysis of dirty pictures," J. R. Stat. Soc. B 48(3), 259–302 (1986).

- 30. C. Cortes and V. Vapnik, *Support-Vector Networks, Machine Learning*, Springer-Verlag, New York (1995).
- 31. R. Unnikrishnan and M. Hebert, "Measure of similarity," in *Seventh IEEE Workshop on Application of Computer Vision*, pp. 394–394 (2005).

**Chen Zheng** is currently an assistant professor at the School of Mathematics and Information Sciences, Henan University. He received his BS degree in mathematics (information sciences) from Henan University in 2007 and his MS and PhD degrees in statistics and image processing of remote sensing from Wuhan University, in 2009 and 2012, respectively. His current research interests include various topics in remote sensing and image processing.

**Leiguang Wang** received his PhD degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS) from Wuhan University in 2009. Since 2012, he has been an associate professor with Southwest Forestry University, Kunming, China. He is the author of more than 10 articles. His research interests include remote sensing image segmentation and pattern recognition.

**Hui Zhao** received his MS degree in the School of Mathematics and Information Sciences from Henan University in 2004. He is currently an associate professor with Henan University, Kaifeng, China. His current research interests include digital image analysis and recognition.

Xiaohui Chen received her MS degree in the School of Mathematics and Statistics from South-Central University for Nationalities in 2011. She is currently with Henan University, Kaifeng, China. Her current research interests include digital image analysis and remote sensing images segmentation.