Toward Evaluating Multiscale Segmentations of High Spatial Resolution Remote Sensing Images

Xueliang Zhang, Pengfeng Xiao, *Member, IEEE*, Xuezhi Feng, Li Feng, and Nan Ye

***Abstract*—Object-based analysis of high spatial resolution remote sensing images addresses the matter of multiscale segmentation. However, existing segmentation evaluation methods mainly focus on single-scale segmentation. In this paper, we examine the issue of supervised multiscale segmentation evaluation and propose two discrepancy measures to determine the manner in which geographic objects are delineated by multiscale segmentations. A QuickBird scene in Hangzhou, China, is used to conduct the evaluation. The results reveal the effectiveness of the proposed measures, in terms of method comparison and parameter optimization, for multiscale segmentation of high spatial resolution images. Moreover, meaningful indications for selecting suitable multiple segmentation scales are presented. The proposed measures are applicable to performance evaluation and parameter optimization for multiscale segmentation algorithms.**

***Index Terms*—High spatial resolution remote sensing, image segmentation, multiscale segmentation accuracy, object-based image analysis (OBIA), scale selection, supervised evaluation.**

I. Introduction

RECENT years have witnessed a proliferation of commercially available high spatial resolution remote sensing images. Technological advances in sensor spatial resolution have allowed remote sensing to be used in detailed studies of the earth's surface [1] and have promoted a set of new applications with great geometric mapping accuracy and a high level of thematic detail [2]. However, improvement in spatial resolution technology has increased the internal spectral variability of land cover classes.

Pixel-based classification is often confused with increased spectral variability and frequently neglects spatial elements within high spatial resolution images, therefore decreasing the potential accuracy of classification [3], [4]. In order to exploit fully the information content of high spatial resolution images, object-based image analysis (OBIA) [1], [5] has become the principal method and proven to be a new and evolving paradigm [6]. OBIA mainly involves segmentation and classification (or recognition) steps. Segmentation is the process by which an image is partitioned into a set of spatially contiguous image objects composed of a group of pixels with homogeneity or semantic significance. Classification involves labeling image objects as geographical objects. Instead of analyzing the spectral behavior of individual pixels, OBIA groups neighboring pixels into objects. This process produces a meaningful representation of object shape and causes spatial information, particularly contextual information, to become explicit. Furthermore, OBIA is insensitive to spectral variation, which aids in resolving the problem of salt-and-pepper effects [7]–[10].

An important feature of OBIA is its ability to conduct multiscale analysis of one image by starting with objects at multiple segmentation scales [5], [11]–[15]. Several studies have demonstrated a preference of using OBIA with multiscale segmentations in lieu of a single segmentation scale [16], [17]. The following requirements for multiscale segmentation in OBIA are given. First, multiscale approaches are required for monitoring, modeling, and managing complex landscapes [18], [19]. In this case, an image is partitioned into objects, similar to the manner in which humans conceptually organize a landscape. Second, a high spatial resolution scene is often composed of objects of different sizes, shapes, and spatial locations, which creates difficulties for bottom-up segmentation methods to describe various objects in a single scale by means of low-level image features [7], [20], [21]. Based on multiscale segmentations, significant objects can be represented and analyzed at their respective optimal representation scales [11], [22]. Third, objects in high spatial resolution images can be described at different semantic levels, when applications using different cartographic or mapping scales are employed. For example, individual trees, a group of trees, and a forest can be delineated at fine or coarse levels. In this paper, a multiscale segmentation evaluation method is proposed to address the first two concerns and evaluates how accurately the various objects are represented by multiscale segmentations.

OBIA is sensitive to segmentation scale but has difficulty in selecting segmentation scale parameters [1]. We can compensate oversegmentation by a successive classification process. However, undersegmentation decreases classification accuracy [23]. Therefore, several studies have proposed using small

Manuscript received May 30, 2014; revised September 22, 2014; accepted December 7, 2014. This work was supported by the National Basic Research Program of China under Grant 2011CB952001, by the Qinglan Project of Jiangsu Province, and by the Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

X. Zhang, P. Xiao, X. Feng, and L. Feng are with the Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology; the Key Laboratory for Satellite Mapping Technology and Applications of State Administration of Surveying, Mapping and Geoinformation of China; and the Department of Geographic Information Science, Nanjing University, Nanjing 210023, China (e-mail: xiaopf@nju.edu.cn).

N. Ye was with the Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology; the Key Laboratory for Satellite Mapping Technology and Applications of State Administration of Surveying, Mapping and Geoinformation of China; and the Department of Geographic Information Science, Nanjing University, Nanjing 210023, China. He is now with the Department of Civil Engineering, Monash University, Clayton, Vic. 3800, Australia (e-mail: xiaopf@nju.edu.cn).

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

Digital Object Identifier 10.1109/TGRS.2014.2381632

primitive objects as minimum classification units [24], [25]. Nevertheless, the quest for an "optimal" segmentation scale has led to the development of numerous evaluation methods. In OBIA, evaluating segmentation results is critical to selecting suitable segmentation approaches and parameters [26].

In addition to visual assessment [27], [28], several supervised and unsupervised methods have been developed for segmentation evaluation. A supervised evaluation method [29]–[39] compares segmented results with the reference by region overlapping, boundary distance, or shape difference. This method is apparently effective but is, in fact, time consuming when creating the reference. An unsupervised evaluation method [40]–[50] defines quality measures for conditions to be satisfied by an "ideal" segmentation. This evaluation method thus characterizes segmentation algorithms by computing goodness measures based on segmented results without the reference. This method is objective but has the added difficulty of having to design effective indicators. Another system-level strategy attempts to include segmentation within object-based classification, in which the segmentation scale with the highest classification accuracy is considered optimal [51], [52]. Although several properties intrinsic to a segmentation algorithm but are independent of applications may not be revealed by this strategy [53], it shows that object-based classification is sensitive to segmentation scales and that different segmentation scales can result in apparent classification differences.

Most proposed evaluation methods emphasize evaluating single-scale segmentation results. As previously described, various objects in a high spatial resolution image can be represented by different segmentation scales. Therefore, evaluating multiscale segmentation quality becomes necessary. Recently, multiscale segmentation optimization has garnered much attention in the study of OBIA. For a system-level evaluation, Johnson and Xie [17] identified multiple segmentation scales by classification results based on multiscale segmentations. For an unsupervised evaluation, other studies generated a single optimal segmentation by fusing multiscale segmentations according to local-oriented unsupervised evaluation [11], [47], [54]. However, the range of involved scales was found to be limited. By contrast, multiple segmentation scales were selected according to a change in homogeneity [43], [55], [56]. The proposed method in this study provides an elegant means of evaluating the effectiveness of the selected scales for this unsupervised strategy. For a supervised evaluation, multiscale segmentation optimization has been achieved by using the single-scale evaluation measure based on different sets of reference objects. For example, some studies have provided reference objects and suitable segmentation scales for different land cover classes [10], [22]. Drăguţ *et al.* evaluated the effectiveness of multiple segmentation scales by providing several sets of reference objects at different scales [55]. The difficulty of this strategy is preparing appropriate sets of reference objects that reflect changes of scales. Specifically, Trias-Sanz *et al.* have proposed an edge-based evaluation method to optimize features for multiscale segmentations [57]. However, discussion on parameter optimization and segmentation scale selection is lacking.

The objective of this study is to develop the multiscale segmentation evaluation method. We demonstrate the applicability of the proposed measures on parameter optimization and segmentation scale selection. Furthermore, we try to address questions related to selecting suitable multiscale segmentations for representing various objects in high spatial resolution images.

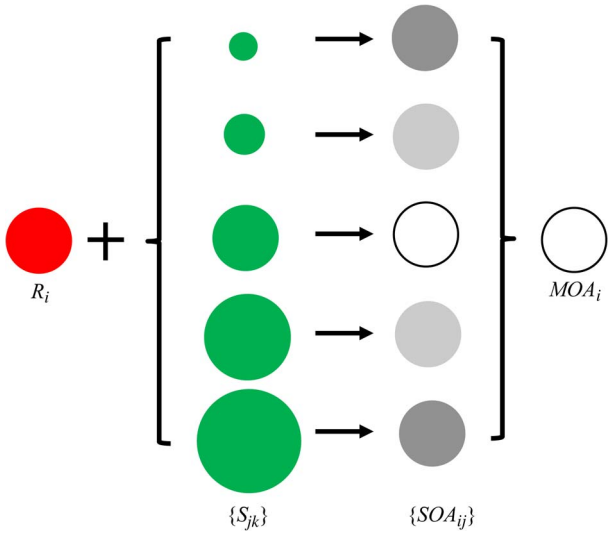

Fig. 1. Example of calculating the MOA_i of a reference object R_i, by comparing it with a set of multiscale segments $\{S_{jk}\}$ at its corresponding location. White circles for $\{SOA_{ij}\}$ and MOA_i represent a value of 1, and gray circles represent values less than 1.

II. Methodology

A. Evaluating Multiscale Segmentations at Object Level

We conduct a multiscale segmentation evaluation in order to reveal the manner in which various objects are represented by segments at multiple scales. To assess the accuracy of multiscale segmentations $\{S_j\}$ at the object level, a multiscale object accuracy (MOA) measure is proposed, by comparing $\{S_j\}$ with the reference R. The MOA measure is defined on the basis of each reference object, by examining how well it is expressed by one of the multiscale segmented regions at the corresponding location. Thus, we calculate and summarize the MOA_i values of all reference objects to obtain the MOA value for the entire image. An example of an MOA_i calculation is shown in Fig. 1.

First, we define the single-scale object accuracy (SOA_{ij}) measure for a reference object R_i, by comparing R_i with a set of overlapped segmented regions $\{S_{jk}, k = 1, 2, \ldots, n_s\}$ at scale j, where n_s is the number of segmented regions in S_j. SOA_{ij} is calculated according to the following:

$$SOA_{ij} = \max\left\{\frac{2|R_i \cap S_{jk}|}{|R_i| + |S_{jk}|}, k = 1, 2, \ldots, n_s\right\} \quad (1)$$

where $|\cdot|$ denotes the number of pixels in a region.

The MOA_i of R_i is then defined as the maximal SOA_{ij} value across n scales, such that

$$MOA_i = \max\{SOA_{ij}, j = 1, 2, \ldots, n\}. \quad (2)$$

Finally, the MOA value of multiscale segmentations $\{S_j\}$ is the weighted summation of n_r reference objects, yielding the

following equation:

$$MOA = \sum_{i=1}^{n_r} MOA_i |R_i|/|R|. \quad (3)$$

These definitions show that MOA is a generic framework for multiscale segmentation evaluation because SOA_{ij} can be defined as different single-scale accuracy measures. However, a single-scale measure should be calculated by considering each object separately. An effective multiscale evaluation can then directly reflect the manner in which various objects are represented by segments at different scales.

The SOA_{ij} defined in (1) causes the MOA value to range from 0 to 1. In an ideal case, when the segmented regions are identical to the reference objects, the MOA value is 1. If the segmentation quality is low, the MOA value decreases in direct proportion. Moreover, the SOA_{ij} measure is sensitive to both over- and undersegmentation. Given the reference segment R_i, if the segmented region S_{jk} is oversegmented, the numerator of (1) is significantly smaller than the denominator. By contrast, an undersegmented S_{jk} can lead to a larger denominator than the numerator in (1). Both of these make the SOA_{ij} value decrease. We note that the single-scale measure should be sensitive to both over- and undersegmentation for each object. Otherwise, the over- or undersegmented segments can achieve high values for the multiscale evaluation measure.

According to the definition given in (2), the MOA value increases or remains unchanged with each increment in the number of scales in $\{S_j\}$. The MOA measure not only can indicate the overall accuracy of multiscale segmentations but also can reveal specific properties. For example, we can know the segmentation accuracy of each reference object in (2), from which an accuracy map can be produced. This map is shown in Figs. 7 and 9. In addition, we can determine that each reference object is best represented by a corresponding segment across the scales, according to S_{ij} in (2) and as shown in Figs. 12–14. Revealing statistics can then be generated, such as the number of reference objects that are best represented by the segmentation at each scale, as shown in Table II.

B. Evaluating Multiscale Segmentations at Pixel Level

The measure of bidirectional consistency error (BCE^*), which evaluates a segmentation by multiple references [58], is extended to assess the accuracy of multiscale segmentations $\{S_j\}$, by comparing multiple segmentations with a single reference R at the pixel level. A sample calculation is provided in Fig. 2.

First, a measure A is defined to determine the degree to which the reference R and a segmentation result S_j agree at pixel p_m. That is

$$A(R, S_j, p_m) = 1 - \frac{|B(R, p_m) \backslash B(S_j, p_m)|}{|B(R, p_m)|} \quad (4)$$

$$A(S_j, R, p_m) = 1 - \frac{|B(S_j, p_m) \backslash B(R, p_m)|}{|B(S_j, p_m)|} \quad (5)$$

where $B(S_j, p_m)$ and $B(R, p_m)$ denote the segmented region in S_j and the reference object in R that contains the point p_m,

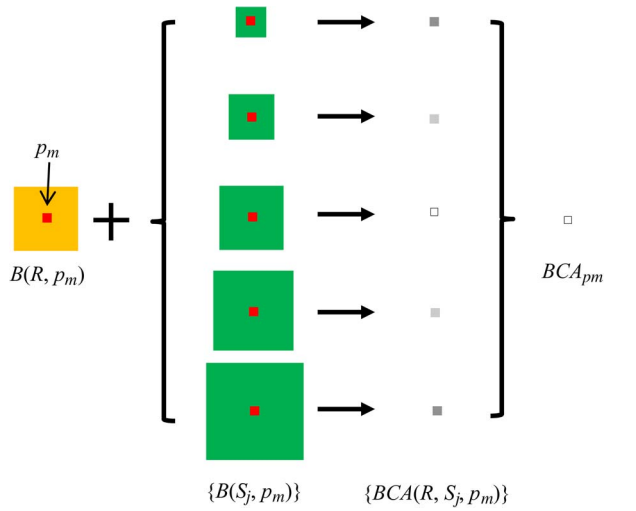

Fig. 2. Example of calculating the bidirectional consistency accuracy (BCA) for a pixel p_m within the reference object $B(R, p_m)$, by comparing $B(R, p_m)$ with a set of segmented regions $\{B(S_j, p_m)\}$ that contain p_m. The white squares for $\{BCA(R, S_j, p_m)\}$ and BCA_{p_m} represent the value of 1, and the gray squares represent values less than 1.

respectively, and $\cdot\backslash\cdot$ indicates the set difference. For example, $B(R, p_m)\backslash B(S_j, p_m)$ represents the set of pixels that belongs to $B(R, p_m)$ but not to $B(S_j, p_m)$. The $A(R, S_j, p_m)$ measure is sensitive to oversegmentation, and $A(S_j, R, p_m)$ is sensitive to undersegmentation. Given the reference object $B(R, p_m)$, if $B(S_j, p_m)$ becomes increasingly oversegmented, the value of $A(R, S_j, p_m)$ continues to decrease to 0. However, the $A(S_j, R, p_m)$ value does not change as it increases to 1. By contrast, if $B(S_j, p_m)$ becomes increasingly undersegmented, the value of $A(S_j, R, p_m)$ continues to decrease to 0, but the $A(R, S_j, p_m)$ value does not change after it increases to 1. Therefore, to maintain the sensitivity to both over- and undersegmentation, the measure of bidirectional consistency accuracy $BCA(R, S_j, p_m)$ or $BCA(S_j, R, p_m)$ of p_m at segmentation scale j is defined as

$$\begin{aligned} BCA(R, S_j, p_m) &= BCA(S_j, R, p_m) \\ &= \min\{A(S_j, R, p_m), A(R, S_j, p_m)\}. \end{aligned} \quad (6)$$

Then, the bidirectional consistency accuracy at pixel p_m (BCA_{p_m}) across n segmentation scales is defined as

$$BCA_{p_m} = \max\{BCA(R, S_j, p_m), j = 1, 2, \ldots, n\}. \quad (7)$$

Finally, the BCA of the entire image is calculated as the sum of BCA_{p_m} at each pixel. That is

$$BCA = \frac{1}{N} \sum_{m=1}^{N} BCA_{p_m} \quad (8)$$

where N is the number of pixels in the image. The BCA value is between 0 and 1 according to the definitions, and a large BCA value indicates high multiscale segmentation quality. Moreover, because the accuracy of each pixel across multiple scales is given as the highest value in (7), the MOA value increases if segmentation scales are added to the set of segmentations. In addition to indicating the overall segmentation

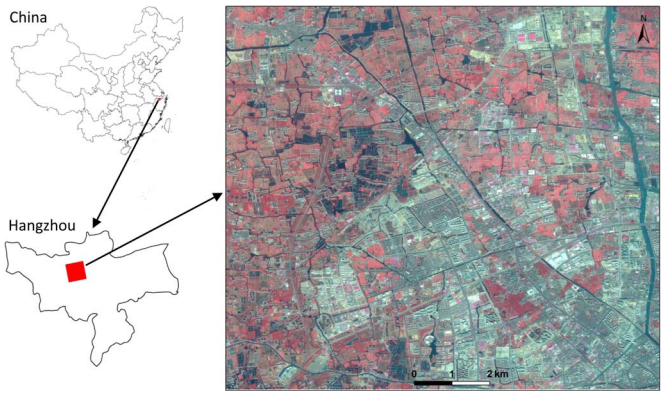

Fig. 3. Location of the study area and the QuickBird image shown with near-infrared, red, and green bands.

accuracy by BCA in (8), the multiscale segmentation accuracy of each pixel can be calculated according to (7), which is visualized as an accuracy map shown in Figs. 7 and 9.

III. Experiments

A. Data and Experimental Setup

The study area is a section of Hangzhou City, Zhejiang Province, China (see Fig. 3). The city is located at approximately 30.3° N and 120.2° E and is part of the Yangtze River Delta Economic Zone, i.e., China's largest core economic region. A QuickBird scene was acquired for use in this study on March 2, 2008, and is 107 km^2 in size. The spatial resolution is 0.6 m after pansharpening using the method proposed by Zhang [59]. The area includes both urban (e.g., commercial, industrial, and residential) areas and rural geographical objects (e.g., farmland, trees, and open water), thus providing a diversity of land cover classes.

We used five subsets from the QuickBird scene as test images, which we labeled T1–T5 and as shown in Fig. 4. These images represent the landscapes of an urban area, a farmland, a residential zone, a water area, and an industrial zone, respectively. The sizes are 658 × 504, 474 × 489, 793 × 623, 538 × 546, and 996 × 550 pixels for T1–T5, respectively. The reference objects for T1–T5 are created by a specialist in remote sensing, who outlined the boundary of each semantic object, rather than the homogeneous regions, because the definition of homogeneity is more arbitrary than that of a semantic object [58]. These manual delineations were reviewed by a second operator to catch any obvious errors. Finally, we identified 165, 156, 554, 107, and 292 reference objects in T1–T5, respectively.

In the experiment, we first determine the effectiveness of the proposed measures on evaluating single-scale segmentation by comparing them with existing discrepancy measures. Since the multiscale evaluation measures are extended from single-scale segmentation evaluation, if they cannot accurately reveal single-scale segmentation quality, they are considered ineffective in evaluating multiscale segmentations. We then illustrate the effectiveness of the proposed measures on optimizing parameters for multiscale segmentations. Using the multiresolution segmentation method (MRS) [60] embedded in the eCognition Developer 8 as an example, we employ the proposed measures to optimize its shape parameter. In Section III-D, the measures are applied to compare three region-based multiscale segmentation algorithms, including the MRS, the fast hierarchical segmentation method (FHS) [61], and the mean-shift-based clustering method (MS) [62]. In addition to showing the effectiveness of the proposed measures in selecting the multiscale segmentation method, we reveal the relationships between the performances of single-scale and multiscale segmentations. Finally, the MOA measure is used to reveal the effectiveness of different combinations of segmentation scales on representing various objects, which indicates the manner in which multiple segmentation scales are selected appropriately.

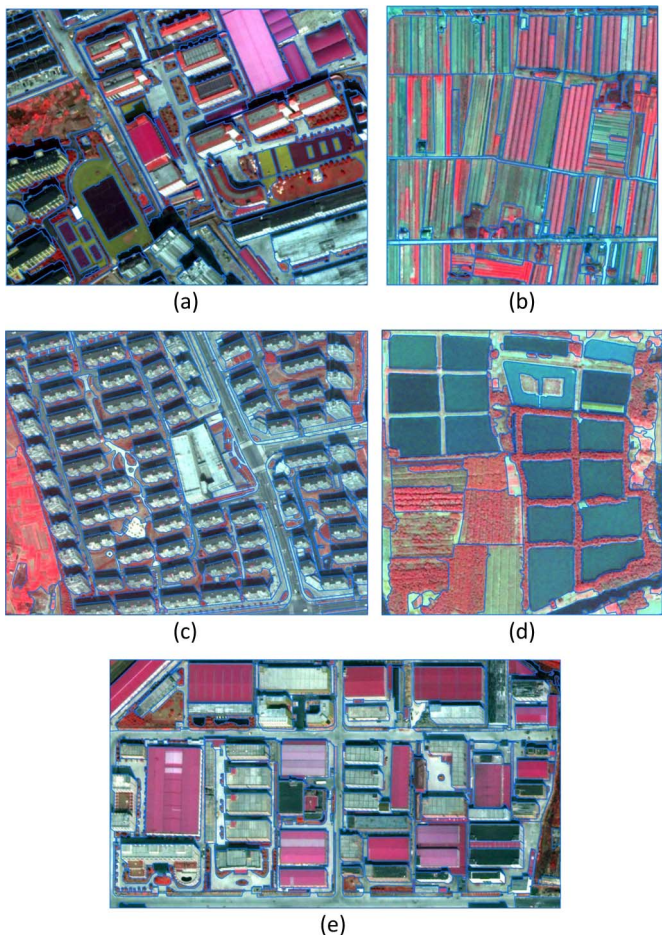

Fig. 4. References of test images: (a) T1, (b) T2, (c) T3, (d) T4, and (e) T5. Blue lines indicate manually delineated object boundaries.

B. Accuracy Assessment for Single-Scale Segmentation

The proposed measures MOA and BCA can be used to evaluate single-scale segmentation, when the number of scales n is equal to 1 in (2) and (7), respectively. If the measures cannot reveal single-scale segmentation quality, their performance in evaluating multiscale segmentations must be ineffective. We used three single-scale evaluation measures for comparison, including the symmetric partition distance (D_{sym}) [63], the adjusted Rand index (ARI) [64], and the discrepancy measure E proposed by Carleer *et al.* [65].

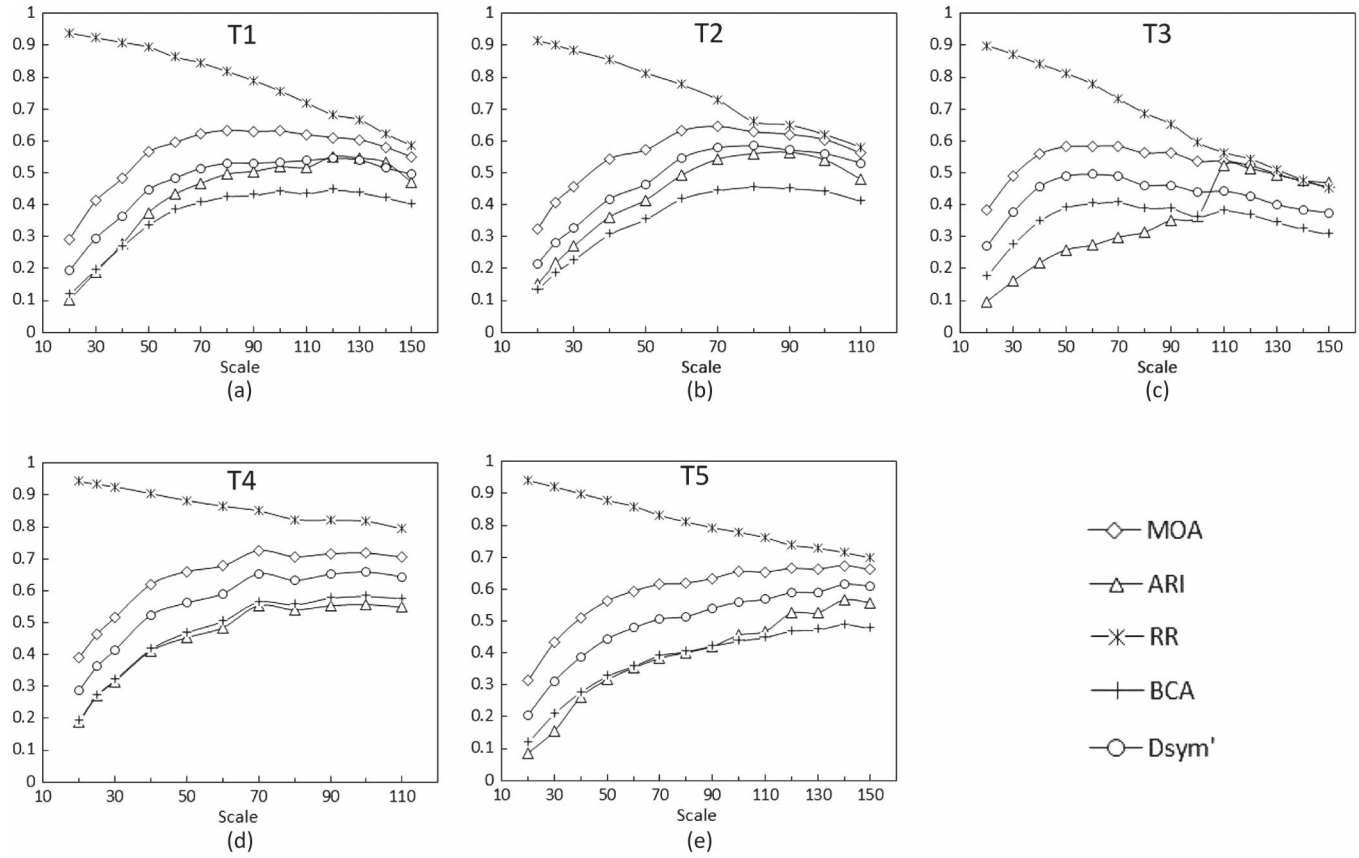

Fig. 5. Evaluation results of single-scale segmentations produced by the MRS method using accuracy measures MOA, BCA, ARI, RR, and D'_{sym}. Test images T1–T5 are shown in (a)–(e), respectively.

D_{sym} is defined as the minimal number of pixels that must be removed from both the reference segmentation (R) and the segmentation result (S) in order for the remaining pixels in R and S to be identical. The measure E is defined as the ratio of the number of improperly segmented pixels to the number of pixels in the image. The ARI assesses the level of agreement of two partitions based on the comparison of object triples, i.e., the manner in which three distinct objects are delineated by the two partitions. In this paper, we change D_{sym} to D'_{sym} and E to a rightly segmented ratio (RR), according to the formulas $D'_{\mathrm{sym}} = 1 - D_{\mathrm{sym}}$ and $RR = 1 - E$, respectively. The maximal value for the three measures is 1, and a large value indicates high segmentation quality. The RR measure is tolerant to oversegmentation, such that it continues to increase when the segmentation becomes increasingly oversegmented.

The segmentation results produced by the MRS method [60] are evaluated. MRS is a region-growing method controlled by scale, shape, and compactness parameters. The shape and compactness parameters are set at a constant of 0.5, but different scale parameters are set for each test images. The scale parameters of T1, T3, and T5 are increased from 20 to 150 in 10-unit steps, and those of T2 and T4 are set as 20, 25, and from 30 to 110 in 10-unit steps. According to visual analysis, the small scale parameters produce oversegmented results, and large scale parameters result in undersegmentation. We calculated the evaluation measures of MOA, BCA, D'_{sym}, RR, and ARI for each segmentation scale and plotted the results in Fig. 5 to show the quality change in conjunction with segmentation scales.

The changes of MOA, BCA, D'_{sym}, and ARI are similar when the values increase starting from the finest scale and decrease after reaching the highest value. The best segmentation scales individually indicated by the four measures are nearly the same, except that the best scale for ARI is slightly coarser than that of the MOA, BCA, and D'_{sym} measures for T1–T3. As expected, the RR values continue to decrease from fine to coarse scales. However, it is interesting to note that, if the decreasing rate of RR is slow, the values of the other four measures decrease less significantly at coarse scales, as for T4 and T5. The comparisons show that the proposed measures MOA and BCA can reveal the quality of single-scale segmentations and are sensitive to both over- and undersegmentation, which is the necessary condition to achieve effective performance during multiscale segmentation evaluation.

C. Optimization of Multiscale Segmentation Parameters

The MRS method is used to show the effectiveness of the proposed measures in optimizing parameters for multiscale segmentations. We set the compactness parameter of MRS as 0.5, due to its lack of sensitivity [66], and we focus on optimizing the shape parameter. Five sets of multiscale segmentations for each test image are produced by setting the shape parameters to 0.1, 0.3, 0.5, 0.7, and 0.9, respectively. For shape 0.5, the

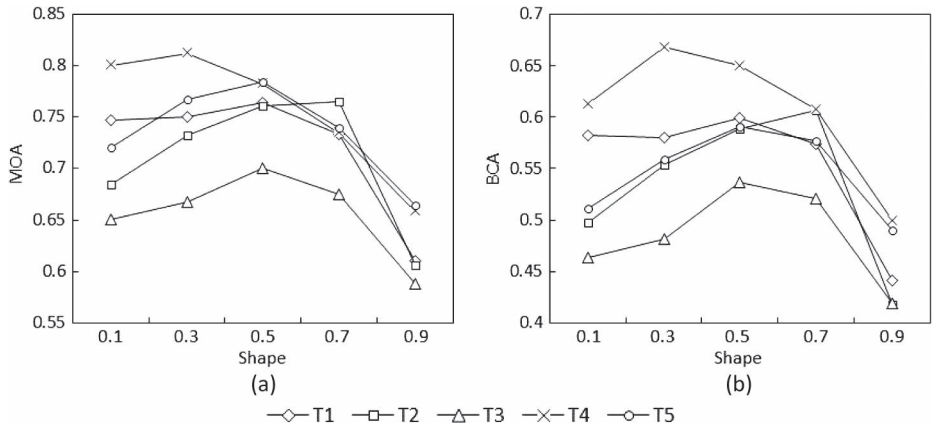

Fig. 6. Multiscale segmentation accuracy of MRS results when setting different shape parameters for test images T1–T5: (a) MOA measure; (b) BCA measure.

scale parameters are set to the same as those in Section III-B. For the remaining shape parameters, the scale parameters are set to achieve similar multiscale segmentations as those of shape 0.5. The MOA and BCA measures of each set of multiscale segmentations are then calculated to show the differences caused by shape parameters in Fig. 6.

Although the absolute MOA and BCA values are different, the qualitative changes of the two measures as the shape parameter increases are similar. Both measures indicate that shape 0.9 achieves the worst performance for all test images and that the performance of shape 0.1 is the next worst. This illustrates the necessity of integrating spectral and shape features. The best shape parameter is 0.7 for T2, 0.3 for T4, and 0.5 for T1, T3, and T5, which parallels the characteristics of the test images. The farmlands in T2 have regular shape, so that the relatively large shape parameter is of high quality. The objects in T4 are mainly ponds and trees displaying significant spectral contrast, so that the small shape parameter performs more effectively than does a large shape parameter. The shape and spectral features of objects in T1, T3, and T5 are complex, so that shape 0.5 achieves the best performance.

To illustrate the differences caused by shape parameters, multiscale segmentation accuracy maps of T5 are presented in Fig. 7. The MOA accuracy map clearly shows the segmentation quality of each reference object. It is clear that most objects in the MOA accuracy map of shape 0.5 are brighter than those in shapes 0.1 and 0.9, indicating that shape 0.5 performs more effectively than the other shapes. In the BCA accuracy map, reference objects are broken, and the segmentation quality is represented at the pixel level, which suggests the heterogeneous parts in reference objects that decrease segmentation accuracy. For example, when comparing the accuracy maps of shape 0.5 with those of shape 0.9, we show that the pixels near object boundaries have lower values in the BCA accuracy map of shape 0.9, which reflects the effect of neglecting spectral information when setting a large shape parameter.

To optimize parameters for other multiscale segmentation algorithms, a similar procedure as used in this example can be applied, which is to produce sets of multiscale segmentations by establishing different parameters. Multiscale segmentation evaluation results, including values and accuracy maps, can then be used to determine suitable parameters.

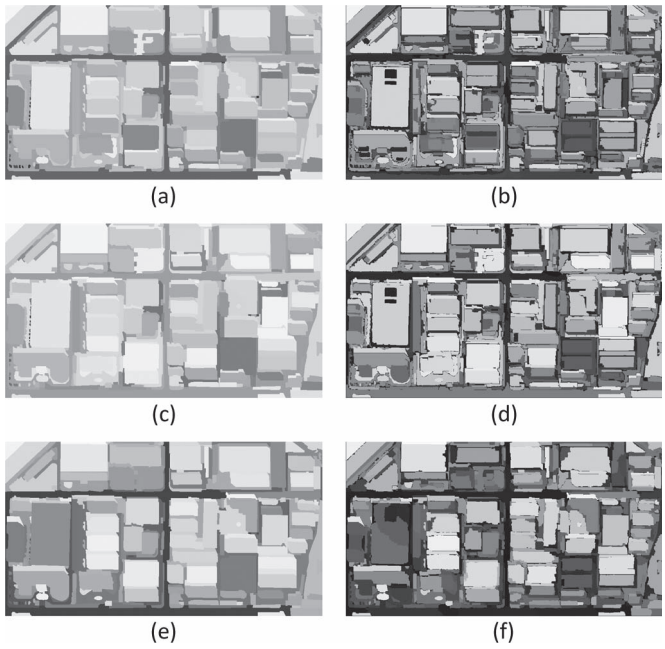

Fig. 7. Multiscale segmentation accuracy maps of MRS results for test image T5: (a) MOA, shape 0.1; (b) BCA, shape 0.1; (c) MOA, shape 0.5; (d) BCA, shape 0.5; (e) MOA, shape 0.9; (f) BCA, shape 0.9. The white areas represent segmentation accuracy with a value of 1, and the black areas represent a value of 0.

D. Comparison of Multiscale Segmentation Methods

Here, we use the proposed measures to compare segmentation methods MRS, MS, and FHS. This process allows us to demonstrate the effectiveness of selecting an appropriate multiscale segmentation method. (For more details on segmentation methods, please refer to references [60]–[62].) The scale parameters of MRS are set identical to those in Section III-B, and the shape parameter is set at a constant of 0.5. The scale parameter of the FHS method is the number of regions. We then produce multiscale segmentations of FHS that have the same region number as that of MRS. In addition, the spectral weight of FHS is set to 0.5. The MS method has three parameters (h_s, h_r, M) to control the segmentation scale, where h_s and h_r are bandwidth parameters of the spatial and range domain, and M is the minimum region area. Determining the proper scale parameters for the MS method [67], [68] is difficult. In this

TABLE I
MULTISCALE SEGMENTATION ACCURACY OF MRS, MS, AND FHS METHODS BASED ON MOA AND BCA MEASURES

Segmentation method	T1		T2		T3		T4		T5	
	MOA	*BCA*	*MOA*	*BCA*	*MOA*	*BCA*	*MOA*	*BCA*	*MOA*	*BCA*
MRS	0.764	0.599	0.782	0.650	**0.701**	0.537	**0.761**	**0.589**	**0.784**	**0.590**
FHS	**0.778**	**0.619**	0.799	0.660	0.684	0.502	0.757	0.587	0.783	0.589
MS	0.746	0.612	**0.825**	**0.723**	0.697	**0.544**	0.723	0.562	0.760	0.586

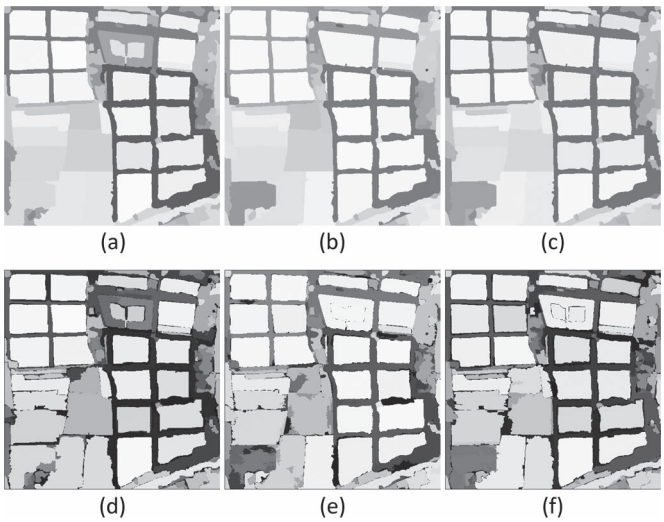

Fig. 8. Multiscale segmentation accuracy maps of (a) and (d) MRS, (b) and (e) MS, and (c) and (f) FHS results for test image T4. The top row is associated with the MOA measure, and the bottom row is associated with the BCA measure. The white areas represent a segmentation accuracy of 1, and the black areas represent an accuracy of 0.

paper, the h_s and h_r parameters are set from 7 to 20 at an interval of 1 for T1, T3, and T5, and from 7 to 27 at an interval of 2 for T2 and T4. These settings are necessary because the spectral contrast among objects in T2 and T4 is larger than those of the other images. The parameter M is set differently to ensure the segmentation scales of MS approximate those of MRS.

Multiscale segmentation accuracies indicated by MOA and BCA for the three segmentation methods are shown in Table I. The two measures are consistent in revealing the segmentation quality for each segmentation method. In most cases, the MRS method can achieve the best multiscale segmentation performance. The FHS method performs similarly to MRS, except with respect to test image T3. However, the performance of the MS method is different with MRS and FHS. For test images T2 and T3, the MS method can achieve the best performance. However, for T1, T4, and T5, it has the worst performance among all methods. The causes will be illustrated in the following, by combining with single-scale segmentation quality. Finally, the multiscale segmentation evaluation results reveal that no segmentation method can achieve the best performance for all test images.

The MOA and BCA accuracy maps of T4 are shown in Fig. 8 to illustrate the differences caused by the three segmentation methods. According to Table I, for T4, the overall multiscale segmentation accuracy of MRS is the highest, and the accuracy of FHS is slightly lower than that of MRS but significantly higher than that of MS. In the accuracy maps, it is apparent that not all the objects are better segmented by MRS than by FHS and MS. The segmentation accuracy of certain objects produced by MRS, such as the trees that surround pools of water, is even lower than that produced by MS and FHS. This illustrates that although one method achieves the better overall multiscale segmentation performance than the others, it is not necessarily the best method for each object. The findings may inspire future studies to integrate different segmentation methods, in order to represent various objects more accurately in high spatial resolution images.

To further illustrate the relationship between single-scale and multiscale segmentation performance, single-scale accuracies calculated by MOA and BCA for the three segmentation methods are plotted in Fig. 9. In most cases, the FHS method achieves the highest single-scale segmentation accuracy, followed by the MS method. However, this observation is inconsistent with the multiscale segmentation accuracy indicated in Table I, in which the MRS method achieves the best performance in most cases. This inconsistency shows that no direct relationship between the best single-scale performance and the best multiscale segmentation performance exists.

As previously discussed, the multiscale segmentation performances of MRS and FHS are similar, whereas MS performs differently. Accordingly, compared with that of MS, the single-scale accuracy curves of MRS and FHS in Fig. 9 are relatively similar, as the segmentation scale coarsens. However, the single-scale segmentation accuracy of MS is often higher than that of MRS and FHS at fine scales, but it decreases considerably at coarse scales. This shows that MS can perform better at fine scales but worse at coarse scales than can MRS and FHS. In addition, this may explain why the multiscale segmentation accuracy of MS is high for test images T2 and T3 but significantly low for the other images compared with the accuracy of MRS and FHS. The reference objects in T2 and T3 (see Fig. 4) are relatively small. Therefore, a high performance of MS at fine scales contributes to high multiscale segmentation accuracy. By contrast, the reference objects in test images T1, T4, and T5 are relatively large. Thus, a poor performance of MS at coarse scales contributes to reduced multiscale segmentation accuracy compared with that of MRS and FHS. Based on this analysis, we can determine that the single-scale segmentation qualities for all segmentation scales are related to multiscale segmentation accuracy. However, the relation is qualitatively limited and cannot be described quantitatively at present.

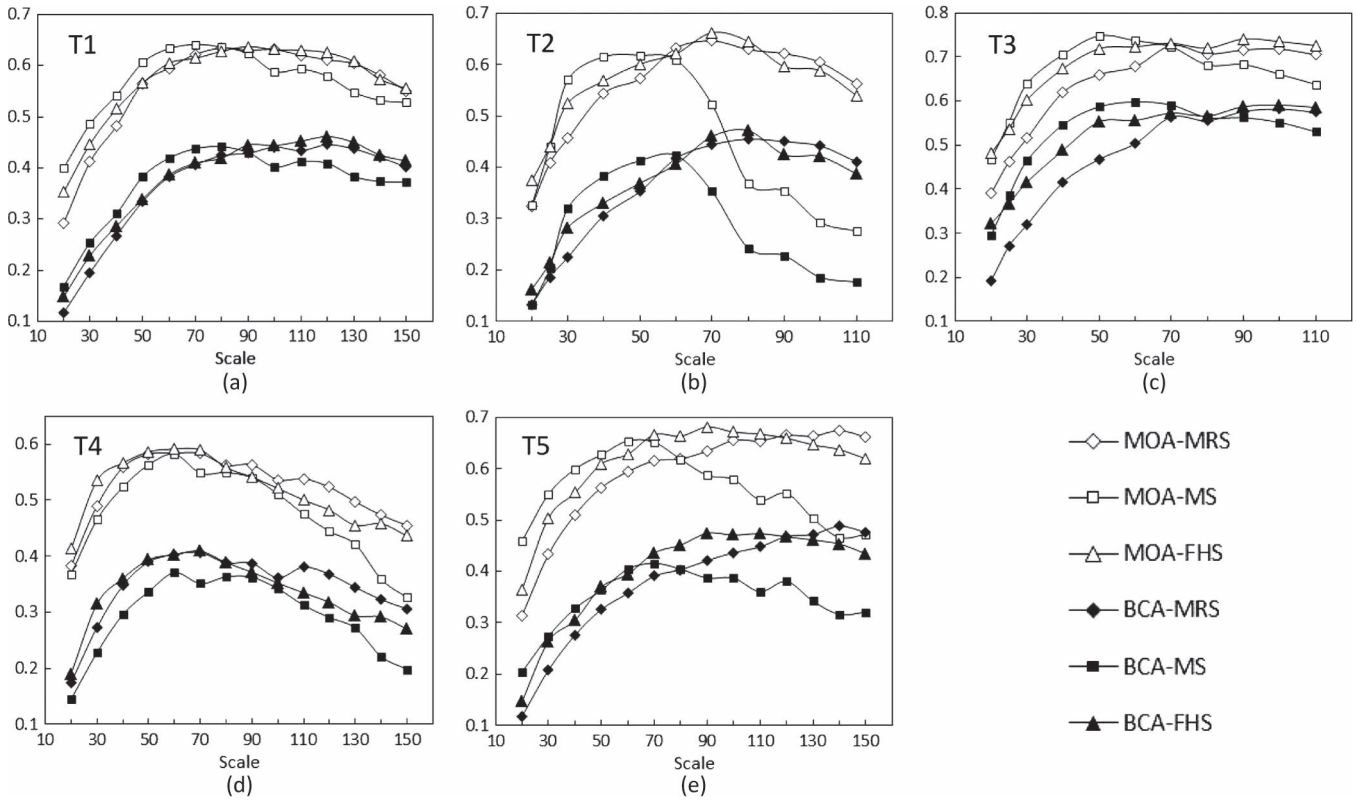

Fig. 9. Accuracy of single-scale MRS, MS, and FHS segmentation results based on MOA and BCA measures.

E. Selection of Multiple Segmentation Scales

Here, the effectiveness of combinations of different segmentation scales in representing various objects are analyzed, giving indications on how to select suitable multiple segmentation scales for high spatial resolution images. In Sections III-C and D, we describe that multiscale evaluation is performed on a set of segmentations having more than ten scales. However, the following questions arise: 1) Are so many segmentation scales necessary to represent the various objects? 2) How many scales are sufficient and how to choose these scales appropriately? Multiscale segmentations produced by the MRS method (i.e., the same as those described in Section III-B) are used to examine these questions during an evaluation of the MOA measure.

First, we examine the effect of the number of segmentation scales in multiscale segmentation accuracy. Changes in the MOA accuracy as the number of scales increases are plotted in Fig. 10. When selecting m scales from n scales $(m < n)$, the number of combinations of m scales is $c(n, m)$, which is equal to $n!/((n-m)!m!)$. For each m, we generate all combinations and calculate the MOA accuracy of each combination. The maximal, minimal, and mean MOA values and the standard deviation of the MOA values in the set of combinations of m segmentation scales are calculated. Fig. 10 shows that, when the number of scales increases, the maximal, minimal, and mean MOA values of the corresponding set of combinations all increase. However, this rate decreases with each increment in the number of scales. Generally, when the number of scales is three or four, the maximal MOA accuracy achieves nearly the highest for all test images. This shows that three or four segmentation scales are sufficient to represent the various objects. However, the standard deviation and changing range of MOA values for each set of combinations decrease in direct proportion to the increment in the number of scales. This shows that, when the number of selected scales is small, the variations of multiscale segmentation accuracy within different combinations are significant. In this case, selecting proper multiple segmentation scales is difficult.

We then illustrate the manner in which to select proper multiple segmentation scales, particularly a small number of scales. The intervals between selected neighboring scales are considered the main factor and analyzed. In an ideal case, all neighboring scales have the same interval. For example, if we select three from six scales in intervals of two, the set of combinations of three scales is limited to {{1,3,5},{2,4,6}}. Given the number of selected scales, the effectiveness of scale intervals on multiscale segmentation performance can be shown by evaluating different sets of scale combinations at different intervals. The evaluation results are plotted in Fig. 11, for each given number of scales. When the number of selected scales is two, the mean MOA values increase in direct proportion to the increment of scale intervals and then decrease when the scale intervals is large for T1–T3. For T4 and T5, the mean MOA values do not decrease significantly when the scale intervals become large. When selecting two segmentation scales having a large interval, one scale should be fine and the other coarse. Since the objects in T4 and T5 are relatively large, coarse-scale segmentation contributes to high multiscale segmentation accuracy. By contrast, coarse-scale segmentation decreases the multiscale segmentation accuracy of T1–T3 because the objects

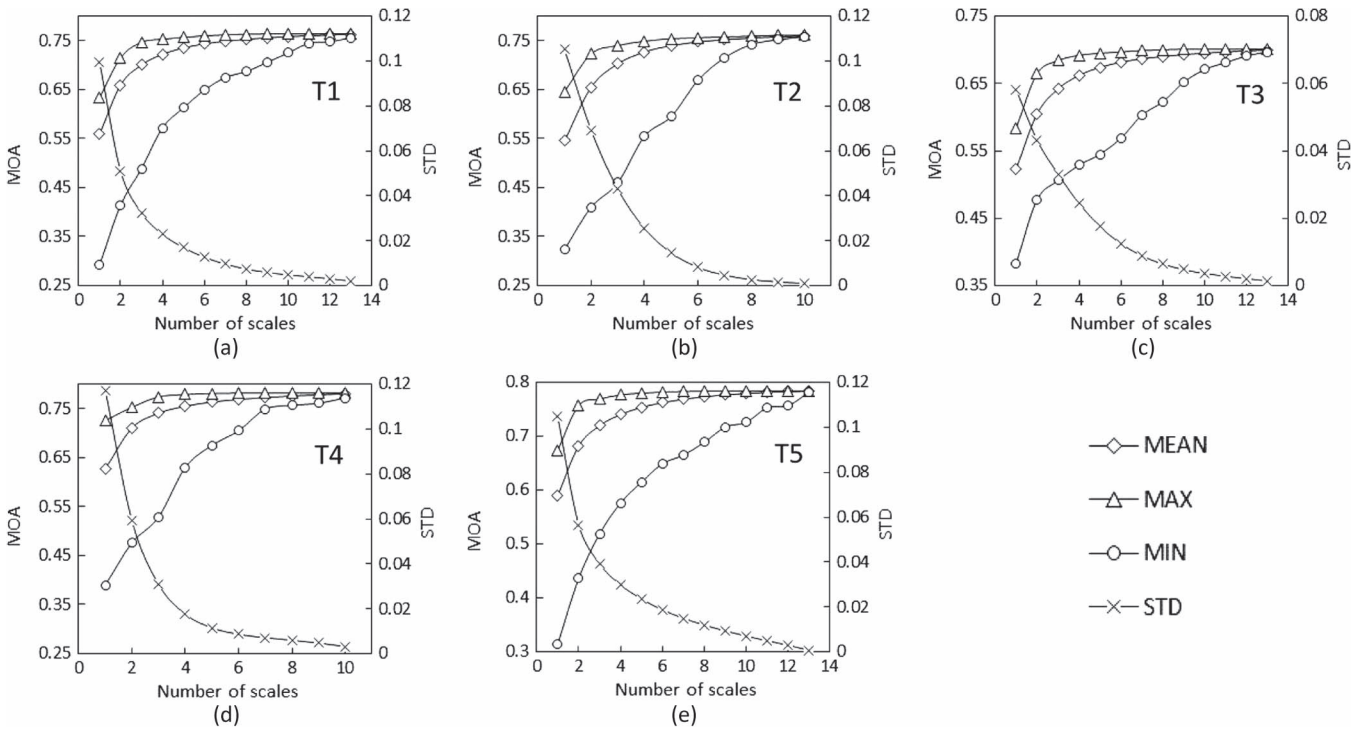

Fig. 10. Changes in the maximum, mean, minimum, and standard deviation of MOA values for sets of MRS multiscale segmentations with increasing numbers of scales. Given the number of scales, each set includes all possible combinations of segmentation scales.

in these images are small. If the number of the selected scales is large, the allowed scale intervals in the ideal case are limited. However, we observe that the mean MOA values continue to increase in direct proportion to increments in the scale intervals, when the number of selected scales is large. The results show that we should select multiple segmentations at suitably large scale intervals to represent the objects in high spatial resolution images.

The optimal combinations of three segmentation scales are used to further illustrate the manner in which we select multiple segmentation scales. Statistics of the three optimal segmentation scales for each test image are presented in Table II. Scale parameter intervals are extensive. Progressing from fine to coarse scales, the number of segments in the three segmentations is significantly larger than, similar to, and significantly smaller than that in the reference segmentation. This shows that the optimal combinations of segmentations simultaneously include the scales of apparent oversegmentation, a similar scale to that of the reference, and apparent undersegmentation. The number of expressed reference segments by the three scales decreases from fine to coarse, but the area of the expressed objects at each scale does not decrease in direct proportion. This shows that fine-scale segmentation can represent large numbers of small objects, whereas coarse-scale segmentation can describe several large objects.

Finally, to present visually Table II statistics related to the optimal combination of segmentation scales, the three optimal segmentation scales and segments that best express reference objects at each scale are shown in Figs. 12–14. The three segmentation scales are oversegmented, medium, and undersegmented, respectively. Small reference objects are represented best by fine-scale segmentation. Medium reference objects are delineated best by the segmentation with a similar scale as that of the reference. Finally, coarse-scale segmentation distinguishes large reference objects more effectively than do other scales.

TABLE II
STATISTICS RELATED TO THE OPTIMAL COMBINATION OF THREE SEGMENTATION SCALES INDICATED BY THE MOA MEASURE

Test image	Scale parameters	Number of segments	Number of expressed objects	Area proportion of the expressed objects
T1	50/100/150	435/152/70	120/35/10	0.362/0.365/0.273
T2	30/60/100	549/169/65	86/54/16	0.168/0.478/0.354
T3	30/70/150	1306/374/125	406/134/14	0.317/0.462/0.221
T4	40/70/110	266/108/59	70/20/17	0.187/0.435/0.378
T5	30/60/140	1341/481/138	162/81/49	0.149/0.284/0.567

IV. DISCUSSION

In this paper, we conduct multiscale segmentation evaluation on a single reference. Although high-level information is retained in the reference by delineating the boundaries of semantic objects rather than those of homogeneous regions, variations caused by different interpretations should be addressed. Uncertainties encountered when delineating references is a topic that has been studied by Albrecht [69], with respect to remote sensing images, and by Martin [58], with respect to natural images. Albrecht explained that the disagreement between interpreters was not related to the entire area of objects indicating thematic uncertainty, but rather to the margins along the object boundaries. Moreover, Albrecht separated the different interpretations

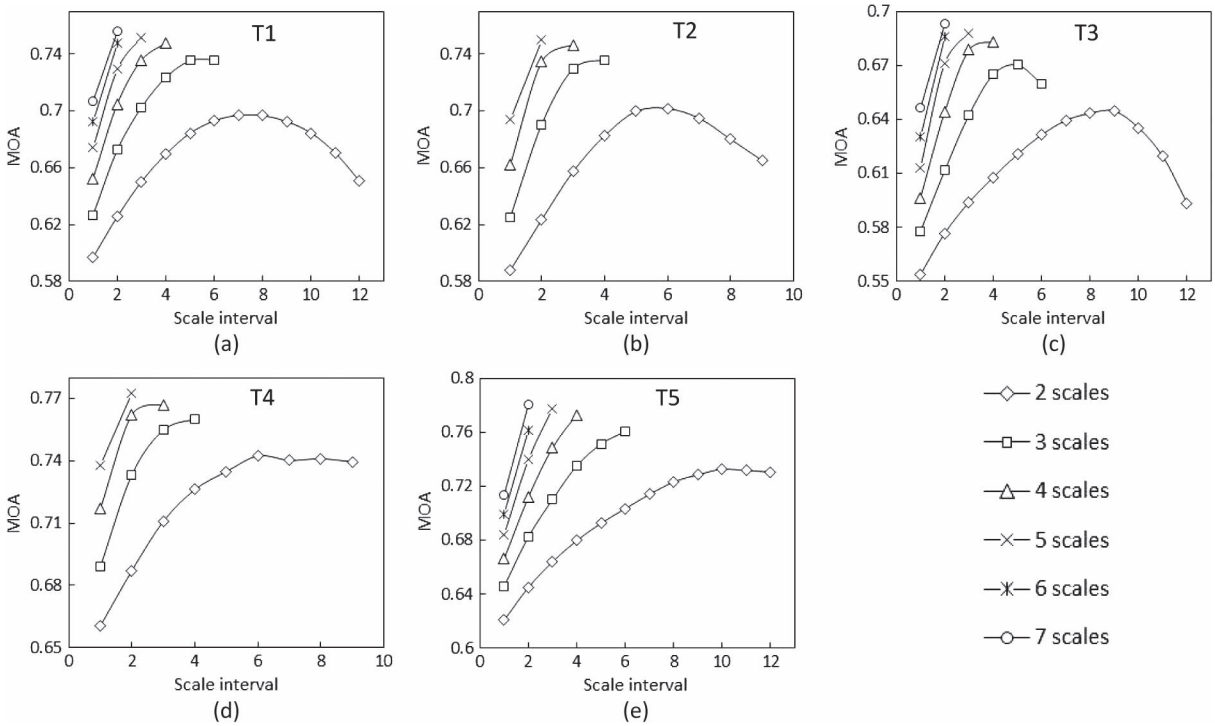

Fig. 11. Changes in multiscale segmentation accuracy as scale intervals increase.

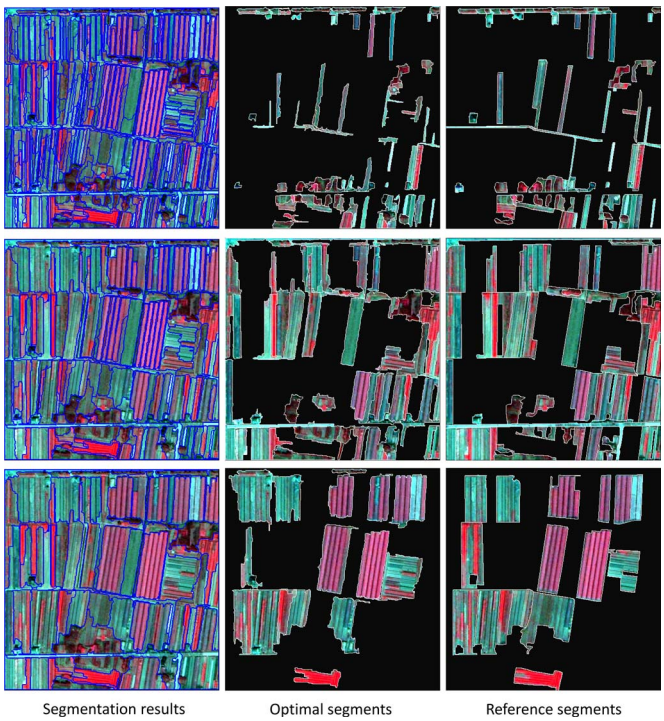

Fig. 12. (Left) Optimal combination of three segmentation scales produced by the MRS method for test image T2. (Middle) Optimal segments at each scale. (Right) Expressed reference segments.

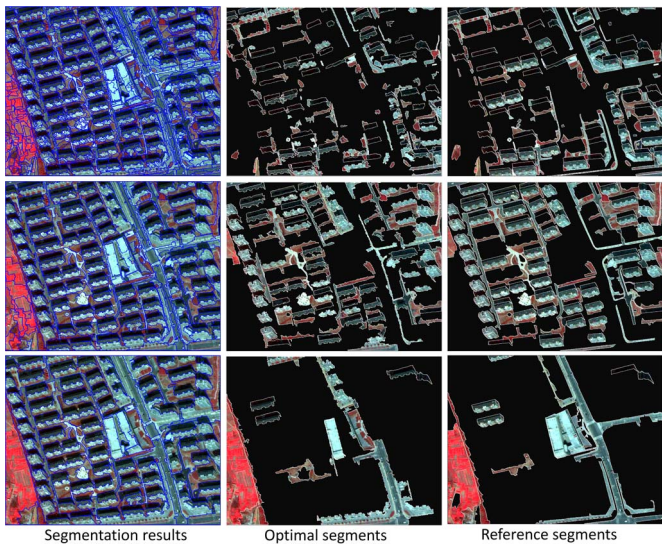

Fig. 13. (Left) Optimal combination of three segmentation scales produced by the MRS method for test image T3. (Middle) Optimal segments at each scale. (Right) Expressed reference segments.

into detailed and generalized groups. This is similar to the research of Martin, who determined that the references produced by different interpreters are consistent except in the case of variations of oversegmenting or undersegmenting certain objects. Both researchers discovered that the boundary deviation of reference objects is not significant. To address boundary deviation, Albrecht produced a medium interpretation as the reference. However, unfortunately, the effect of boundary deviation on segmentation evaluation results was not analyzed, which we believe would be a meaningful guideline for reference delineation. On the other hand, some researchers conducted segmentation evaluation for natural images based on multiple references [53], [58]. In the future, extending the proposed measures to evaluate single-scale or multiscale segmentations of high spatial resolution images using multiple references would be beneficial to this field of research. Regarding the evaluation

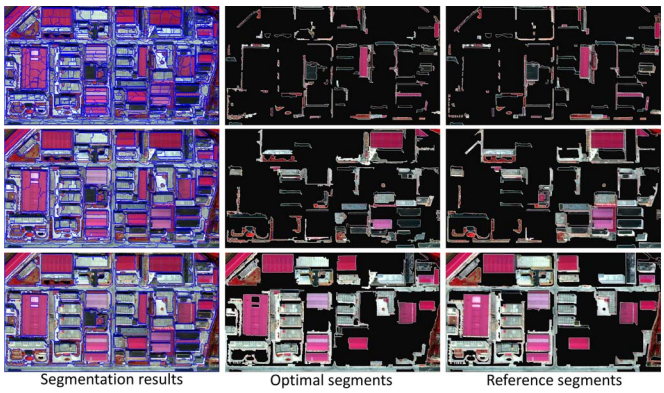

Fig. 14. (Left) Optimal combination of three segmentation scales produced by the MRS method for test image T5. (Middle) Optimal segments at each scale. (Right) Expressed reference segments.

of single-scale segmentation using multiple references, the MOA and BCA measures can be applied directly by exchanging the roles of the reference and segmentation. Specifically, by preparing multiple references with different abstract levels, we can evaluate the manner in which the multiscale segmentation method can describe objects at different semantic levels (e.g., a single tree, group of trees, or an entire forest) in high spatial resolution images.

The proposed multiscale segmentation evaluation measures are first conducted on each object or pixel and then combined to evaluate the performance of the whole partitions. In Section II-A, we argued that multiscale segmentation evaluation should be performed on each unit of object or pixel first across multiple scales. The combination of units can then indicate the manner in which the objects are described by multiple segmentations, as shown in Table I. By contrast, if we first evaluate all segments at each segmentation scale and then combine the single-scale evaluation results across the scales, the multiscale segmentation performance is not clear, as shown in Fig. 8. Table I reveals the multiscale segmentation quality directly, but Fig. 8 does not. Although a qualitative relationship exists between Fig. 8 and Table I based on our analysis in Section III-D, this relationship cannot be quantitatively described at present. A critical concept of the proposed measures is to characterize the delineation quality of each evaluation unit (i.e., an object or pixel) based on corresponding multiple segmentations. In this paper, the maximal operator is used for evaluation, as in (2) and (7), which is designed to show the semantic coherence by exploring the one-to-one relation between the segments and the reference objects. However, other fusing strategies can be applied according to requirements of specific applications, such as the weighted summation strategy [57].

During some evaluation tasks, we evaluate segmentations based on selected objects or land cover classes. For example, if an application extracts farmland from an image, the segmentation evaluation tends to be performed on specific farmland objects. In these cases, the proposed measures can be applied by combining the selected elements rather than all elements in the image.

The evaluation results show that the proposed measures can assess multiscale segmentation accuracy and indicate the manner in which multiple segmentation scales can be selected. The proposed measures can manage various combinations of multiple segmentation scales. Therefore, applications for parameter optimization using multiscale segmentation methods can be expanded. For example, in this study, we optimize the shape and scale parameters of the MRS method separately. However, they can be optimized concurrently to produce multiple segmentations that more effectively represent various objects. Moreover, the proposed measures can help to evaluate the effectiveness of methods for automatic selection of suitable multiscale segmentations [43], [55], [56].

The evaluation results in Figs. 12–14 show that objects with significant spectral heterogeneity and complex shape remain poorly represented by single segments. This reflects the limitations of bottom-up segmentation methods that use only low-level image features and may require the integration of prior knowledge into remote sensing image segmentation procedures [20], [21].

V. Conclusion

In this paper, we examined the problem of evaluating multiscale segmentation performance. Two discrepancy measures are proposed to assess multiscale segmentation accuracy: the MOA measure at the object level and the BCA measure at the pixel level. In addition to indicating the overall accuracy of multiscale segmentations, the proposed measures can reveal accuracy from local insight according to accuracy maps. Moreover, the MOA measure illustrates a general multiscale evaluation framework at the object level, in which other indicators of object accuracy can be embedded. The proposed methods are suitable for multiscale segmentation evaluation.

We used five test images from a QuickBird scene in Hangzhou, China, to conduct the evaluation. The effectiveness of the proposed measures is demonstrated in the case of single-scale evaluation by comparing them with existing measures, which is the necessary condition to achieve effective performance during multiscale segmentation evaluation. Then, the proposed measures are successfully applied to parameter optimization and method comparison for multiscale segmentation of high spatial resolution images. The relationships between the performances of single-scale and multiscale segmentations were analyzed and reveal that the single-scale segmentation qualities of a set of multiscale segmentations are qualitatively related to multiscale segmentation performance.

Finally, the proposed measures were used to analyze the effect of different combinations of multiple segmentation scales in representing objects. Meaningful findings related to the selection of suitable multiscale segmentation scales include the following: 1) Multiscale segmentation accuracy continues to improve if additional segmentation scales are evaluated. However, three or four scales used with the test images can approximate the best performance in terms of accuracy. 2) High multiscale segmentation accuracy can be achieved by setting suitably large scale intervals. 3) Multiple segmentation scales should include oversegmented, medium, and undersegmented scales concurrently to represent various objects in high spatial resolution images.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments.

REFERENCES

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

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

[3] T. Blaschke and J. Strobl, "What's wrong with pixels? Some recent developments interfacing remote sensing and GIS," *GIS-Zeitschrift für Geoinformationssysteme*, vol. 14, no. 6, pp. 12–17, 2001.

[4] G. J. Hay and G. Castilla, "Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline," *In Object-Based Image Analysis*, in Berlin, Germany Springer-Verlag, 2008, pp. 75–89.

[5] U. C. Benz, P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen, "Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information," *ISPRS J. Photogramm. Remote Sens.*, vol. 58, no. 3/4, pp. 239–258, Jan. 2004.

[6] T. Blaschke *et al.*, "Geographic object-based image analysis—Towards a new paradigm," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 180–191, Jan. 2014.

[7] J. Schiewe, L. Tufte, and M. Ehlers, "Potential and problems of multiscale segmentation methods in remote sensing," *GIS-Zeitschrift für Geoinformationssysteme*, vol. 6, no. 1, pp. 34–39, 2001.

[8] Q. Yu *et al.*, "Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 7, pp. 799–811, Jul. 2006.

[9] G. Yan, J. F. Mas, B. H. P. Maathuis, Z. Xiangmin, and P. M. Van Dijk, "Comparison of pixel-based and object-oriented image classification approaches—A case study in a coal fire area, Wuda, Inner Mongolia, China," *Int. J. Remote Sens.*, vol. 27, no. 18, pp. 4039–4055, Sep. 2006.

[10] S. W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng, "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery," *Remote Sens. Environ.*, vol. 115, no. 5, pp. 1145–1161, May 2011.

[11] H. G. Akçay and S. Aksoy, "Automatic detection of geospatial objects using multiple hierarchical segmentations," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2097–2111, Jul. 2008.

[12] D. G. Blumberg and G. Zhu, "Using a hierarchical multi-resolution mechanism for the classification and semantic extraction of landuse maps for Beer-Sheva, Israel," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 3273–3289, Aug. 2007.

[13] M. S. Chubey, S. E. Franklin, and M. A. Wulder, "Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 4, pp. 383–394, Apr. 2006.

[14] G. J. Hay, G. Castilla, M. A. Wulder, and J. R. Ruiz, "An automated object-based approach for the multiscale image segmentation of forest scenes," *Int. J. Appl. Earth. Observ.*, vol. 7, no. 4, pp. 339–359, Dec. 2005.

[15] G. Mallinis, N. Koutsias, M. Tsakiri-Strati, and M. Karteris, "Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site," *ISPRS J. Photogramm. Remote Sens.*, vol. 63, no. 2, pp. 237–250, Mar. 2008.

[16] M. Kim, T. A. Warner, M. Madden, and D. S. Atkinson, "Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: Scale, texture and image objects," *Int. J. Remote Sens.*, vol. 32, no. 10, pp. 2825–2850, May 2011.

[17] B. Johnson and Z. Xie, "Classifying a high resolution image of an urban area using super-object information," *ISPRS J. Photogramm. Remote Sens.*, vol. 83, pp. 40–49, Sep. 2013.

[18] C. Burnett and T. Blaschke, "A multi-scale segmentation/object relationship modeling methodology for landscape analysis," *Ecol. Model.*, vol. 168, no. 3, pp. 233–249, Oct. 2003.

[19] G. J. Hay, T. Blaschke, D. J. Marceau, and A. Bouchard, "A comparison of three image-object methods for the multiscale analysis of landscape structure," *ISPRS J. Photogramm. Remote Sens.*, vol. 57, no. 5/6, pp. 327–345, Apr. 2003.

[20] E. Borenstein and S. Ullman, "Combined top-down/bottom-up segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 12, pp. 2109–2125, Dec. 2008.

[21] Z. Tu, X. Chen, A. L. Yuille, and S. C. Zhu, "Image parsing: Unifying segmentation, detection, and recognition," *Int. J. Comput. Vis.*, vol. 63, no. 2, pp. 113–140, Feb. 2005.

[22] N. S. Anders, A. Seijmonsbergen, and W. Bouten, "Segmentation optimization and stratified object-based analysis for semi-automated geomorphological mapping," *Remote Sens. Environ.*, vol. 115, no. 12, pp. 2976–2985, Dec. 2011.

[23] J. Schiewe, "Segmentation of high-resolution remotely sensed data—Concepts, applications and problems," in *Int. Archives Photogramm., Remote Sens. Spatial Inf. Sci.*, 2002, vol. 34, pp. 380–385.

[24] I. Dronova, P. Gong, and L. Wang, "Object-based analysis and change detection of major wetland cover types and their classification uncertainty during the low water period at Poyang Lake, China," *Remote Sens. Environ.*, vol. 115, no. 12, pp. 3220–3236, Dec. 2011.

[25] R. Raši *et al.*, "An automated approach for segmenting and classifying a large sample of multi-date Landsat imagery for pan-tropical forest monitoring," *Remote Sens. Environ.*, vol. 115, no. 12, pp. 3659–3669, Dec. 2011.

[26] M. Neubert, H. Herold, and G. Meinel, "Assessing image segmentation quality—Concepts, methods and application," in *Object-Based Image Analysis*, Berlin, Germany: Springer-Verlag, pp. 769–784, 2008.

[27] Y. J. Zhang, "A survey on evaluation methods for image segmentation," *Pattern Recognit.*, vol. 29, no. 8, pp. 1335–1346, Aug. 1996.

[28] G. Meinel and M. Neubert, "A comparison of segmentation programs for high resolution remote sensing data," in *Proc. Int. Archives Photogramm., Remote Sens.*, 2004, vol. 35, pp. 1097–1105.

[29] N. Clinton, A. Holt, J. Scarborough, Y. Li, and P. Gong, "Accuracy assessment measures for object-based image segmentation goodness," *Photogramm. Eng. Remote Sens.*, vol. 76, no. 3, pp. 289–299, Mar. 2010.

[30] F. J. Estrada and A. D. Jepson, "Benchmarking image segmentation algorithms," *Int. J. Comput. Vis.*, vol. 85, no. 2, pp. 167–181, May 2009.

[31] Y. Liu *et al.*, "Discrepancy measures for selecting optimal combination of parameter values in object-based image analysis," *ISPRS J. Photogramm. Remote Sens.*, vol. 68, pp. 144–156, Mar. 2012.

[32] A. Lucieer, "Uncertainties in segmentation and their visualization," Ph.D. dissertation, Utrecht Univ. Int. Inst. Geo-Inf. Sci. Earth Observ., Enschede, The Netherlands, 2004.

[33] P. R. Marpu, M. Neubert, H. Herold, and I. Niemeyer, "Enhanced evaluation of image segmentation results," *J. Spatial Sci.*, vol. 55, no. 1, pp. 55–68, Jun. 2010.

[34] M. Möller, L. Lymburner, and M. Volk, "The comparison index: A tool for assessing the accuracy of image segmentation," *Int. J. Appl. Earth. Observ.*, vol. 9, no. 3, pp. 311–321, Aug. 2007.

[35] C. Persello and L. Bruzzone, "A novel protocol for accuracy assessment in classification of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1232–1244, Mar. 2010.

[36] A. Räsänen, A. Rusanen, M. Kuitunen, and A. Lensu, "What makes segmentation good? A case study in boreal forest habitat mapping," *Int. J. Remote Sens.*, vol. 34, no. 23, pp. 8603–8627, Oct. 2013.

[37] C. Witharana, D. L. Civco, and T. H. Meyer, "Evaluation of data fusion and image segmentation in earth observation based rapid mapping workflows," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 1–18, Jan. 2014.

[38] Q. Zhan, M. Molenaar, K. Tempfli, and W. Shi, "Quality assessment for geo-spatial objects derived from remotely sensed data," *Int. J. Remote Sens.*, vol. 26, no. 14, pp. 2953–2974, Jul. 2005.

[39] Y. J. Zhang, "Evaluation and comparison of different segmentation algorithms," *Pattern Recognit. Lett.*, vol. 18, no. 10, pp. 963–974, Oct. 1997.

[40] S. Chabrier, B. Emile, C. Rosenberger, and H. Laurent, "Unsupervised performance evaluation of image segmentation," *EURASIP J. Appl. Signal Process.*, vol. 2006, no. 1, pp. 096306-1–096306-12, Mar. 2006.

[41] J. Chen, J. Li, D. Pan, Q. Zhu, and Z. Mao, "Edge-guided multiscale segmentation of satellite multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4513–4520, Nov. 2012.

[42] P. Corcoran, A. Winstanley, and P. Mooney, "Segmentation performance evaluation for object-based remotely sensed image analysis," *Int. J. Remote Sens.*, vol. 31, no. 3, pp. 617–645, Feb. 2010.

[43] L. Drăguţ, D. Tiede, and S. R. Levick, "ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data," *Int. J. Geogr. Inf. Sci.*, vol. 24, no. 6, pp. 859–871, Mar. 2010.

[44] G. M. Espindola, G. Câmara, I. A. Reis, L. S. Bins, and A. M. Monteiro, "Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation," *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 3035–3040, Jul. 2006.

[45] D. Faur, I. Gavat, and M. Datcu, "Salient remote sensing image segmentation based on rate-distortion measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 855–859, Oct. 2009.

[46] Y. Gao, J. F. Mas, N. Kerle, and J. A. Navarrete Pacheco, "Optimal region growing segmentation and its effect on classification accuracy," *Int. J. Remote Sens.*, vol. 32, no. 13, pp. 3747–3763, Jul. 2011.

[47] B. Johnson and Z. Xie, "Unsupervised image segmentation evaluation and refinement using a multi-scale approach," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 4, pp. 473–483, Jul. 2011.
[48] A. Stein and K. De Beurs, "Complexity metrics to quantify semantic accuracy in segmented Landsat images," *Int. J. Remote Sens.*, vol. 26, no. 14, pp. 2937–2951, Jul. 2005.
[49] H. Zhang, J. E. Fritts, and S. A. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," *Comput. Vis. Image Understand.*, vol. 110, no. 2, pp. 260–280, May. 2008.
[50] X. Zhang, P. Xiao, and X. Feng, "An unsupervised evaluation method for remotely sensed imagery segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 156–160, Mar. 2012.
[51] L. Dronova *et al.*, "Landscape analysis of wetland plant functional types: The effects of image segmentation scale, vegetation classes and classification methods," *Remote Sens. Environ.*, vol. 127, pp. 357–369, Dec. 2012.
[52] A. S. Laliberte and A. Rango, "Texture and scale in object-based analysis of subdecimeter resolution unmanned aerial vehicle (UAV) imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 761–770, Mar. 2009.
[53] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 929–944, Jun. 2007.
[54] T. Esch, M. Thiel, M. Bock, A. Roth, and S. Dech, "Improvement of image segmentation accuracy based on multiscale optimization procedure," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 463–467, Jul. 2008.
[55] L. Drăguţ, O. Csillik, C. Eisank, and D. Tiede, "Automated parameterisation for multi-scale image segmentation on multiple layers," *ISPRS J. Photogramm. Remote Sens.*, vol. 88, pp. 119–127, Feb. 2014.
[56] J. Yang, P. Li, and Y. He, "A multi-band approach to unsupervised scale parameter selection for multi-scale image segmentation," *ISPRS J. Photogramm. Remote Sens.*, vol. 94, pp. 13–24, Apr. 2014.
[57] R. Trias-Sanz, G. Stamon, and J. Louchet, "Using colour, texture, and hierarchical segmentation for high-resolution remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 63, no. 2, pp. 156–168, Mar. 2008.
[58] D. R. Martin, "An empirical approach to grouping and segmentation," Ph.D. dissertation, Comput. Sci. Div., Univ. California, Berkeley, Berkeley, CA, USA, Rep. UCB/CSD-3-1268, Aug. 2003.
[59] Y. Zhang, "Problems in the fusion of commercial high-resolution satellite images as well as Landsat 7 images and initial solutions," *Int. Archives Photogramm., Remote Sens.*, vol. 34, pt. 4, Jul. 2002.
[60] M. Baatz and A. Schäpe, "Multiresolution segmentation—An optimization approach for high quality multi-scale image segmentation," in *Angewandte Geographische Informations-Verarbeitung XII*, J., Strobl and T., Blaschke and G. Griesebner Eds. Karlsruhe, Germany: Wichmann Verlag, 2000, pp. 12–23.
[61] X. Zhang, P. Xiao, and X. Feng, "Fast hierarchical segmentation of high-resolution remote sensing image with adaptive edge penalty," *Photogramm. Eng. Remote Sens.*, vol. 80, no. 1, pp. 71–80, Jan. 2014.
[62] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
[63] J. S. Cardoso and L. Corte-Real, "Toward a generic evaluation of image segmentation," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1773–1782, Nov. 2005.
[64] L. Hubert and P. Arabie "Comparing partitions," *J. Classification*, vol. 2, no. 1, pp. 193–218, Dec. 1985.
[65] A. P. Carleer, O. Debeir and E. Wolff, "Assessment of very high spatial resolution satellite image segmentations," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 11, pp. 1285–1294, Nov. 2005.
[66] C. Witharana and D. L. Civco, "Optimizing multi-resolution segmentation scale using empirical methods: Exploring the sensitivity of the supervised discrepancy measure Euclidean distance 2 (ED2)," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 108–121, Jan. 2014.
[67] D. Comaniciu, "An algorithm for data-driven bandwidth selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 281–288, Feb. 2003.
[68] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.
[69] F. Albrecht, "Uncertainty in image interpretation as reference for accuracy assessment in object-based image analysis," in *Proc. 9th Int. Symp. Spatial Accuracy Assess. Nat. Resour. Environ. Sci.*, Leicester, U.K., 2010, pp. 13–16.

Xueliang Zhang was born in Wangcheng, Hunan, China, in 1987. He received the B.S. degree in geographical information system in 2010 from Nanjing University, Nanjing, China, where he is currently working toward the Ph.D. degree in cartography and geographical information system.

He is the author of six articles published in international journals. His research interests include remote sensing image processing and high-resolution remote sensing image segmentation.

Pengfeng Xiao (M'13) was born in Ningxiang, Hunan, China, in 1979. He received the B.S. degree in land resource management from the Hunan Normal University, Changsha, China, in 2002 and the Ph.D. degree in cartography and geographical information system from Nanjing University, Nanjing, China, in 2007.

From 2007 to 2009, he was a Lecturer with the Department of Geographic Information Science, Nanjing University, where he has been an Associate Professor since 2010. He is the author of two books and more than 50 articles. His research interests include remote sensing image processing, land use and land cover change, and remote sensing of snow cover.

Dr. Xiao was a recipient of the Geographical Society of China (GSC) Prize for the Best Paper by Young Author in 2006, the Chinese Professional in Geographic Information Systems (CPGIS) Best Student Paper Award in 2007, and the International Society for Photogrammetry and Remote Sensing (ISPRS) Prize for the Best Paper by Young Author in 2008.

Xuezhi Feng was born in Dingxi, Gansu, China, in 1953. He received the B.S. degree in cartography from Nanjing University, Nanjing, China, in 1979.

From 1980 to 1981, he was a Lecturer with the Department of Geography, Nanjing University. He then joined the Lanzhou Institute of Glaciology and Cryopedology, Chinese Academy of Sciences, Lanzhou, China, where he was an Assistant Professor from 1981 to 1990, an Associate Professor from 1990 to 1994, and a Professor from 1994 to 1997. Since 1997, he has been a Professor with the Department of Geographic Information Science, Nanjing University. He is the author of four books and more than 200 articles. His research interests include remote sensing of snow cover, snowmelt water resource, and remote sensing image processing.

Prof. Feng is an Executive Council Member of the China Association of Remote Sensing Application (CARSA) and an Honorary Council Member of the Associate on Environment Remote Sensing of China (AERSC).

Li Feng received the B.S. degree in computer science and technology from the Nanjing Normal University, Nanjing, China, in 2004, the M.S. degree in cartography and geographical information system from Nanjing University, Nanjing, China, in 2007, and the Ph.D. degree in civil and environmental engineering from the University of Stuttgart, Stuttgart, Germany, in 2012.

She is currently a Postdoctoral Fellow with the Department of Geographic Information Science, Nanjing University. Her research interests include geographic information systems and remote sensing and their application to the land use/cover change and urban development.

Nan Ye received the B.E. degree in hydraulic and hydropower engineering from Tsinghua University, Beijing, China, in 2006 and the Ph.D. degree in civil engineering from Monash University, Melbourne, Vic., Australia, in 2014.

In 2013, he was a Research Assistant with the Department of Geographic Information Science, Nanjing University, Nanjing, China. He is currently a Research Fellow with Monash University. His research interests include retrieval of soil moisture and snow properties from microwave observations.