Double-Variance Measures: A Potential Approach to Parameter Optimization of Remote Sensing Image Segmentation

Yongji Wang, Zhihui Tian, Qingwen Qi, and Jun Wang

***Abstract*—The unsupervised segmentation evaluation (USE) method has been commonly used for remote sensing segmentation parameter (SP) determinations to produce good segmentation results, due to its objectiveness and high efficiency. Existing studies have used different criteria to measure homogeneity and heterogeneity and have used certain combination strategies to form overall evaluations. However, different criteria have unique statistical characteristics. The differentiated statistical characteristics maintained in homogeneity and heterogeneity calculations may result in inherent instability in the USE results, leading to unsuitable SP selections. Moreover, few studies have focused on the simultaneous determination of a single optimal SP and multiple optimal SPs. In this article, double-variance (DV) measures were proposed for recognizing more suitable SPs. Then, two combination strategies, F-measure and local peak (LP), were applied to test the potential of using DV measures to determine a single SP and multiple SPs, respectively. The multiresolution segmentation algorithm and Gaofen-1 data were used to test the proposed method. The comparative results indicated that the DV is a more promising internal homogeneity and external heterogeneity metric for segmentation evaluation and optimal SP determination compared to conventional methods. The F-measure-based DV method could produce better overall goodness of segmentation for differently sized natural geo-objects, compared with the competing methods. The LP-based DV method could obtain multiple optimal scales that produced better segments for the identification of small, natural geo-objects to large, semantic geo-objects, compared to the competitive methods.**

***Index Terms*—Double-variance (DV), geographic-object-based image analysis (GEOBIA), image segmentation, parameter optimization, unsupervised evaluation.**

I. Introduction

THE detailed information on geographical objects provided by high-resolution satellite imagery raises a new challenge for traditional pixel-based image analysis [1], [2], since the phenomenon of capturing the same geo-object with different spectra and different geo-objects with the same spectrum occurs in high-resolution satellite imagery. To address this problem, geographic-object-based image analysis (GEOBIA) has been developed to yield better accuracy [3], [4] because of its lower sensitivity to strong spectral variability within a geo-object [5]. In addition, GEOBIA can generate textural features and shape concepts of geo-objects to improve the accuracy of image analysis [6]–[8].

Image segmentation is a key issue in GEOBIA since it can provide objects to GEOBIA by partitioning remote sensing images into meaningful groups of pixels [9], and the objects are considered the minimum units of GEOBIA. Therefore, many scholars have paid attention to obtaining good objects for GEOBIA and have developed segmentation methods, such as the multiresolution segmentation (MRS) [10], mean-shift segmentation [11], machine learning methods [12]–[14], and hybrid segmentation [8], [15]–[19]. However, defining appropriate segmentation parameters (SPs) for good segmentation results is a major challenge [20], [21], since the aforementioned algorithms almost entirely use certain user-defined parameters to control the segmentation quality [22]. In most cases, inappropriate SPs will result in bad segmentation, which will cause low accuracy in the subsequent image analysis [23]. Hence, segmentation quality evaluation is a prerequisite for obtaining appropriate SPs and fine segmentation results.

Segmentation quality evaluation methods can be grouped into the subjective trial-and-error method and objective criterion-based method. The trial-and-error method allows the specialist to obtain appropriate SPs for fine segmentation by visually comparing multiple segmentation results [24]. However, this method is time- and labor-consuming and is inevitably affected by subjectivity. The criterion-based method includes the supervised and unsupervised evaluation methods, in which the difference lies in whether or not reference polygons are needed. The former allows the specialist to obtain fine segmentation by computing the dissimilarity between the segmentation results and the reference polygons [25]–[28]. However, generating reference polygons is subjective and time-consuming [29]. Inappropriately generated reference polygons will cause false evaluation results. The latter method allows the specialist to obtain more efficient and less subjective segmentation evaluation results without needing reference polygons [25], [30]–[34]. At the moment, the unsupervised evaluation method can achieve automation in selecting appropriate SPs [31]. Thus, much effort

Manuscript received November 23, 2020; revised January 13, 2021; accepted January 22, 2021. Date of publication January 26, 2021; date of current version February 12, 2021. *(Corresponding author: Yongji Wang.)*

Yongji Wang and Zhihui Tian are with the School of Geoscience and Technology, Zhengzhou University, Zhengzhou 450001, China (e-mail: wangyongji@zzu.edu.cn; iezhtian@zzu.edu.cn).

Qingwen Qi is with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China, and also with the University of Chinese Academy of Sciences, Beijing 100864, China (e-mail: qiqw@igsnrr.ac.cn).

Jun Wang is with the Shandong University of Science and Technology, Qingdao 266510, China (e-mail: wj811811@163.com).

Digital Object Identifier 10.1109/JSTARS.2021.3054638

has been put into unsupervised segmentation evaluation (USE) methods in recent years.

The USE method commonly supports the idea that good segmentation should be homogeneous in the intrasegment area and heterogeneous in the intersegment area [32], [40], [41]. Some mathematical criteria have been proposed and applied to compute intrasegment homogeneity and intersegment heterogeneity in existing studies [2], [24]–[26], [31], [32], [34]–[37]. Nevertheless, the consideration of the impact of different mathematical criteria applied to intrasegment homogeneity and intersegment heterogeneity calculations based on segmentation assessment results has rarely been reported in the existing literature. Most methods use different mathematical criteria to measure intrasegment homogeneity and intersegment heterogeneity, such as the T and D metrics [35], area-weighted variance (WV), and Moran's I metrics [32], [37]. Although these metrics have been demonstrated to be effective for evaluating segmentation quality and determining the appropriate SPs for good segmentation results, the impact of the characteristics of different mathematical criteria on segmentation evaluation results has remained obscure until now. Different mathematical criteria have unique statistical characteristics. If using different mathematical criteria to indicate the intrasegment homogeneity and intersegment heterogeneity, the differentiated statistical characteristics between the criteria may result in inherent instability in the segmentation evaluation results, leading to unsuitable SP selections and poor goodness of segmentation results. Therefore, it is interesting to study whether segmentation evaluation accuracy would be improved if the mathematical criteria with the same statistical characteristics are used to calculate both homogeneity and heterogeneity.

Some specific combination strategies, applied to integrate the intrasegment homogeneity and intersegment heterogeneity into a comprehensive indicator, are required for the segmentation quality evaluation and suitable SP selection. Early studies mainly focused on examining combination strategies for single optimal SP determinations [2], [32], [33], [37]–[39]. However, it is impossible for every geo-object to be segmented well in the segmentation of high-resolution satellite imagery obtained by a single optimal SP. Thus, some studies have paid attention to combination strategies for determining multiple optimal SPs to segment high-resolution satellite imagery for more accurate image analyses [2], [37], [40], [41]. We need to decide to recognize single or multiple optimal SPs according to the specific application requirements. If we want to produce one overall good segmentation result, we can tradeoff the small geo-objects and the large geo-objects well by determining a single SP. If we want to recognize differently sized geo-objects well, multiple SPs are required.

In this article, a new USE method that incorporates double-variance (DV) measures for homogeneity and heterogeneity calculations was developed to obtain more reasonable SPs. Then, two integrated strategies were applied to combine the intrasegment homogeneity and intersegment heterogeneity into one comprehensive indicator for the overall segmentation quality assessment, one of which could achieve the determination of a single optimal SP and the other could achieve multiple optimal SPs.

II. Methods

A schematic of the application of the proposed new USE method for remote sensing image analysis is displayed in Fig. 1. First, we used the MRS algorithm [42] to produce a set of segmentation results by setting different SPs. Then, we calculated the area WV and area-weighted relative variance (WRV) to indicate the intrasegment homogeneity and intersegment heterogeneity, respectively. Finally, two combination strategies were used to assess the segmentation quality and determine the appropriate SPs, in which the F-measure strategy and local peak (LP) strategy determined the single and multiple optimal SPs, respectively. More details of the proposed method are described in Fig. 1 and the following sections.

A. Intrasegment Homogeneity and Intersegment Heterogeneity of Segments Measured by the Area WV and Area WRV Methods

A commonly accepted opinion in most studies incorporating the USE method considers that good segmentation should have good internal homogeneity and external heterogeneity [32]. Using some mathematical criteria to indicate what is universally accepted as a satisfactory segmentation result has become popular [29]. In this study, we used a similar approach to assess segmentation quality and developed DV measures (i.e., WV and WRV) while maintaining the same statistical characteristics to indicate internal homogeneity and external heterogeneity.

1) Intrasegment Homogeneity Indicator: WV: We use WV as the intrasegment homogeneity indicator in this article. Good intrasegment homogeneity is indicated by a low WV value. The calculation of WV can be divided into the following two steps.

First, the variance of each segment in a given segmentation result is calculated as the mean across all bands of an image, as follows:

$$v_i = \frac{\sum_{\mathrm{j}=1}^{m} v_{ij}}{m} \tag{1}$$

where m is the number of bands contained in an image, v_{ij} is the variance in band j, and v_i is the variance averaged across all bands for segment i.

Second, the internal homogeneity is indicated by the area WV and then defined as follows:

$$\mathrm{WV} = \frac{\sum_{i=1}^{n} a_i \cdot v_i}{\sum_{i=1}^{n} a_i} \tag{2}$$

where a_i represents the number of pixels in segment i and n represents the number of segments in a segmentation.

The approach takes full advantage of the band information contained in an image and the area information contained in each segment, making the calculation of the internal homogeneity more objective.

2) Intersegment Heterogeneity Indicator: WRV: Since the variance can be used to quantify the spectral difference among pixels within a segment and indicate internal homogeneity, it is possible to apply it to indicate the external heterogeneity by quantifying the spectral difference among segments. We develop the WRV measure as the intersegment heterogeneity indicator in this study. In contrast with the intrasegment homogeneity

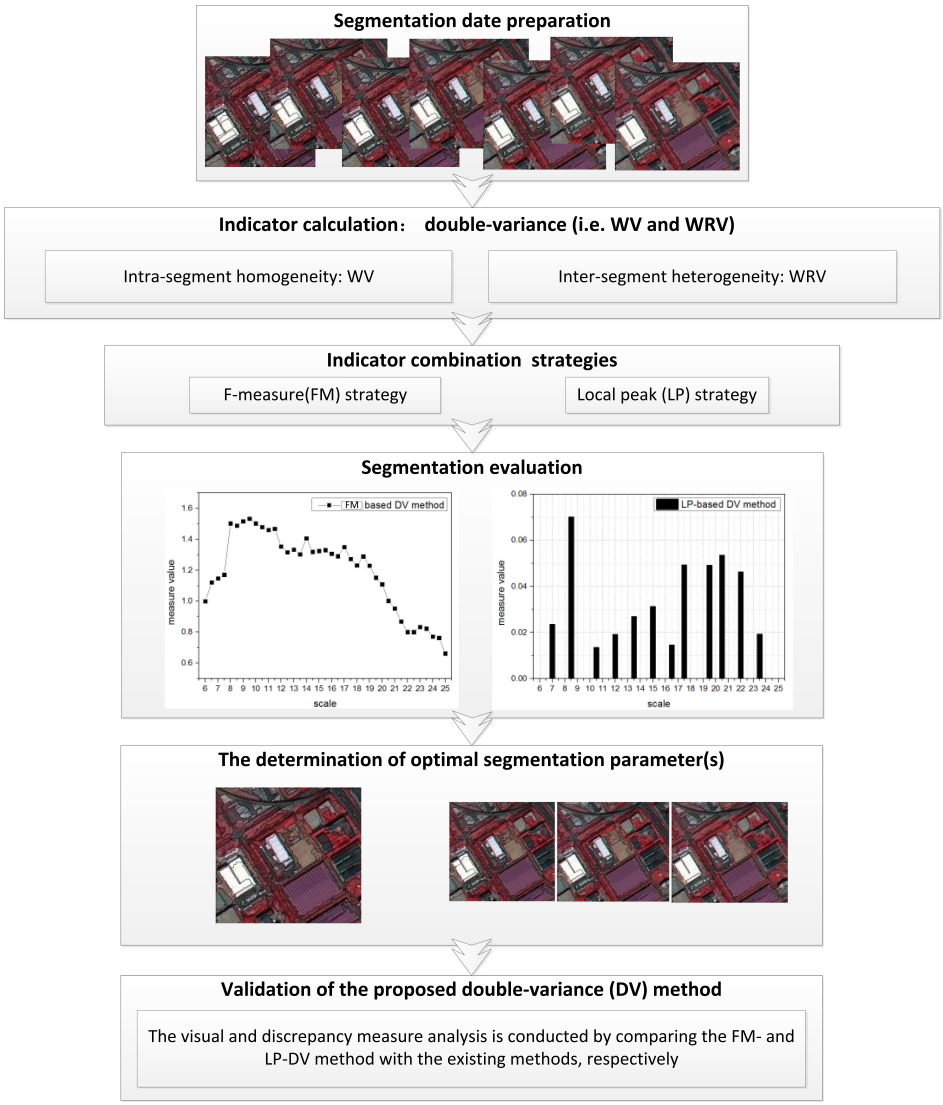


Fig. 1. Flowchart of the proposed new USE method using DV measures for remote sensing images.

indicator, good intersegment heterogeneity is indicated by a high WRV value. The calculation of WRV can be partitioned into the following five steps.

1) The region adjacency graph [43], [[44] is applied to reveal the relationship between every segment and its neighboring segments for subsequent external heterogeneity calculations.
2) The variance between each segment and one of its neighboring segments in a given segmentation result is computed using the following definition:

$$v_{ik} = \frac{(\mu_i - \mu_{ik})^2 + (\mu_k - \mu_{ik})^2}{2} \tag{3}$$

where μ_i and μ_k are the mean of segments *i* and *k*, respectively, μ_{ik} is the mean among segments *i* and *k*, and v_{ik} is the variance among segment *i* and its neighboring segment *k*.
3) The product of the common border between each segment and one of its neighbors and the area of that neighbor is calculated as follows:

$$Q_{ik} = l_{ik} \cdot a_k \tag{4}$$

where l_{ik} is the common border between segment *i* and neighboring segment *k*, a_k is the area of segment *k*, and Q_{ik} is their product.
4) The relative variance (RV) between each segment and all of its neighboring segments is calculated using the following formula:

$$\mathrm{RV}_i = \frac{\sum_{k=1}^{\Omega} Q_{ik} \cdot v_{ik}}{\sum_{k=1}^{\Omega} Q_{ik}} \tag{5}$$

where Ω represents all of the neighboring segments of segment *i*. The approach considers the contribution of the common border between each segment and one of its neighbors and the area of the neighbor in the formula above, reducing the inherent instability in calculating the

variance between each segment and all of its neighboring segments.

5) The external heterogeneity is indicated by the area WRV averaged across all bands as follows:

$$\overline{\mathrm{RV}_i} = \frac{\sum_{\mathrm{j}=1}^{m} \mathrm{RV}_{ij}}{m} \tag{6}$$

$$\mathrm{WRV} = \frac{\sum_{i=1}^{n} a_i \cdot \overline{\mathrm{RV}_i}}{\sum_{i=1}^{n} a_i} \tag{7}$$

where RV_{ij} represents the RV between each segment and all of its neighboring segments in band *j*.

B. Determination of Appropriate SPs Using Two Integrated Strategies

It is difficult to determine the appropriate SPs only according to the two calculations of WV and WRV. To address the problem, using a specific combination strategy to integrate the WV and WRV measures to capture the tradeoff between them for the determination of the appropriate SPs is critical. Different combination strategies can achieve different purposes for the determination of the appropriate SPs, such as the determination of single or multiple optimal SPs. Accordingly, this article selected two combination strategies (F-measure and LP) for the determination of the single and multiple optimal SPs.

1) Single Optimal SP Determination: F-Measure: The first combination strategy is the F-measure strategy [1], [37], which helps determine a single optimal SP. To allow the intrasegment homogeneity and intersegment heterogeneity measures to be treated equally before implementing the F-measure strategy, the WV and WRV values are normalized by the following formula:

$$\mathrm{WV}_{\mathrm{norm}} = \frac{\mathrm{WV}_{\max} - \mathrm{WV}}{\mathrm{WV}_{\max} - \mathrm{WV}_{\min}} \tag{8}$$

$$\mathrm{WRV}_{\mathrm{norm}} = \frac{\mathrm{WRV} - \mathrm{WRV}_{\min}}{\mathrm{WRV}_{\max} - \mathrm{WRV}_{\min}} \tag{9}$$

where WV and WRV are the values obtained by the WV and WRV calculations performed with a set of segmentation results, respectively. High $\mathrm{WV}_{\mathrm{norm}}$ and $\mathrm{WRV}_{\mathrm{norm}}$ values indicate good intrasegment homogeneity and intersegment heterogeneity.

Then, the normalized WV and WRV values are combined into a global value by the F-measure, which can be used to determine a single optimal parameter indicated by the highest value. The F-measure is defined as follows:

$$F - \mathrm{measure} = \frac{2 \cdot \mathrm{WV}_{\mathrm{norm}} \cdot \mathrm{WRV}_{\mathrm{norm}}}{\mathrm{WV}_{\mathrm{norm}} + \mathrm{WRV}_{\mathrm{norm}}}. \tag{10}$$

2) Multiple Optimal SP Determination: LP: The second combination strategy is the LP strategy [2], which can determine multiple optimal SPs.

First, the ratio of the WV and WRV values is computed with the purpose of simultaneous consideration in both internal homogeneity and external heterogeneity, as follows:

$$\mathrm{RATIO} = \frac{\mathrm{WV}}{\mathrm{WRV}}. \tag{11}$$

When producing a segment that is essentially in agreement with the corresponding geo-objects, the segment boundaries will be retained in the segmentation results with some subsequent SPs, whereas the internal homogeneity and external heterogeneity of that segment will remain the same [33]. Thus, the trend of the RATIO value will become rather flat after the optimal SP is reached. To find the change in a certain SP level, the derivative of RATIO was computed using the following formula:

$$\dot{\mathrm{RATIO}}(l) = \frac{d\mathrm{RATIO}(l)}{dl} = \frac{\mathrm{RATIO}(l) - \mathrm{RATIO}(l - \Delta l)}{\Delta l} \tag{12}$$

where $\mathrm{RATIO}(\mathrm{l})$ is the ratio of WV and WRV at SP l and Δl is the increment of the SP.

According to the aforementioned statement, $\dot{\mathrm{RATIO}}(l)$ is generally greater than $\dot{\mathrm{RATIO}}(l - \Delta l)$ and $\dot{\mathrm{RATIO}}(l + \Delta l)$ if SP l is considered appropriate for obtaining good segments. The difference in the derivative of RATIO between the scale parameter l and its neighboring scales can be calculated as follows.

$$\begin{aligned}\mathrm{Diff}_l = &(\dot{\mathrm{RATIO}}(l) - \dot{\mathrm{RATIO}}(l - \Delta l)) \\ &+ (\dot{\mathrm{RATIO}}(l) - \dot{\mathrm{RATIO}}(l + \Delta l)).\end{aligned} \tag{13}$$

A greater Diff_l indicates a more suitable scale parameter l that can be used to produce good segments. A histogram of Diff_l is then generated, and the set of the greatest relative Diff_l values are regarded as the LPs. The scales at the LPs are recognized as the appropriate SPs in this article. Since differently sized geo-objects occur in the same image, it is possible to identify some LPs for segmentation.

C. Validation of the Proposed Method Using Discrepancy Measures

To test the proposed method, the F-measure-based DV method was visually compared with the overall goodness (OG) [37] and Zhang (Z) methods [35], and the LP-based DV method was in visually compared with the spectral angle (SA) [2] and energy function (E) methods [31]. The F-measure-based DV method and the OG method have the same combination strategy but different heterogeneity measures, and the LP-based DV method, the SA method, and the E method have the same combination strategy but different homogeneity and heterogeneity measures. The homogeneity and heterogeneity measures and combination strategy are obviously different between the F-measure-based DV method and the Z method. Note that the SP that produced the highest OG value or lowest Z value was considered the single optimal SP. The SPs that produced the greatest relative E values were regarded as candidates for the multiple optimal SPs.

Then, to further validate the proposed method, some discrepancy measures were used to assess the segmentation results quantitatively. In this article, the following discrepancy measures were calculated: the quality rate (QR) [45], oversegmentation (OS), undersegmentation (US), Euclidean distance (ED) of OS and US [27], and the precision (P), recall (R), and F-measure (F) of P and R [1]. Note that all the discrepancy measures range from 0 to 1, and lower values of QR, OS, US, and ED and higher values of P, R, and F indicate less dissimilarity between the segmentation results and reference polygons. The F-measure

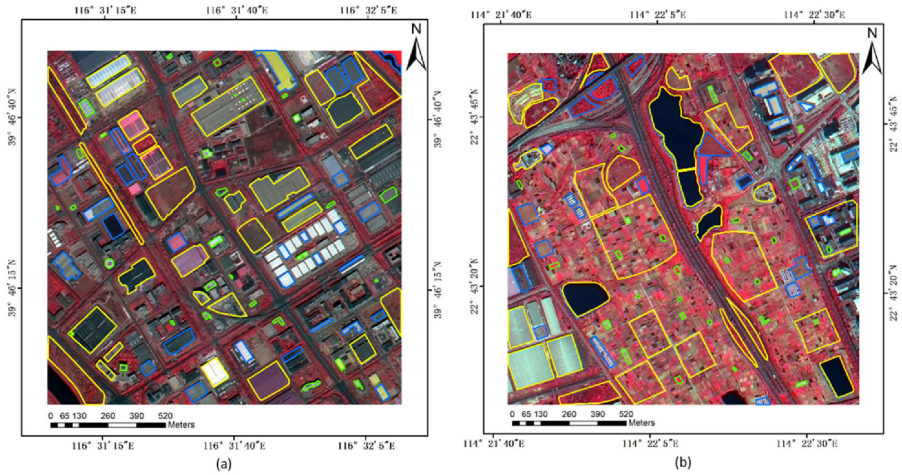


Fig. 2. Study images of (a) T1, an urban area located in Beijing, China and (b) T2, a farmland area located in Shenzhen, China, both of which were acquired by the GF-1 sensor. These images are presented with a combination of near-infrared, red, and green bands (green polygons: the reference of the fine-scale level; blue polygons: the reference of the medium-scale level; and yellow polygons: the reference of the coarse-scale level).

is similar to that seen in (10), but $\mathrm{WV}_{\mathrm{norm}}$ and $\mathrm{WV}_{\mathrm{norm}}$ are replaced by P and R. More details on these discrepancy measures can be seen in [1], [27], and [45].

III. Results

A. Study Sites and Images

This study selected two remote sensing applications with which to validate the SP optimization method, including urban geo-object identification and agricultural division. Both images were acquired from the Gaofen-1 (GF-1) satellite and differed in geo-object type and city. The GF-1 imagery contains a four-band multispectral image of 8-m resolution with blue, green, red, and infrared bands and a panchromatic image of 2-m resolution. Then, a multispectral image with a resolution of 2 m was produced using the nearest-neighbor diffusion-based pansharpening method [46].

Efficient segmentation is a key issue for geo-object identification in high-resolution satellite imagery, such as building recognition [47] and impervious surface extraction [48]. For this application, we selected a typical urban area located in Beijing, China and the GF-1 image used was acquired on May 8th, 2016 [see Fig. 2(a)]. Agricultural division is indispensable for efficient and precise agricultural management [49]. In this experiment, a GF-1 image that contains a farmland area located in Shenzhen, China, was acquired on October 16th, 2015 [see Fig. 2(b)]. Two subsets of 1.6×1.6 km each were clipped from the two acquired GF-1 images (see Fig. 2). The combined use of urban and rural images ensures the effectiveness and objectivity of the proposed DV method.

B. Analysis of DV Measures (WV and WRV)

For this study, the GF-1 images of urban and rural areas were segmented by the MRS algorithm embedded in the eCognition software, in which the shape and compactness parameters were fixed to 0.05 and 0.1, respectively, and the scale parameter increased from 6 to 25 at an interval of 0.5 to produce a set of segmentation results. Note that we chose the shape and compactness parameters according to visual analysis and according to the guideline that the segments matched the geo-objects as far as possible; this article mainly studied the parameter optimization of the scale parameters. In addition, 6 was determined as the starting scale that was least oversegmented and 25 was selected as the ending scale that was clearly undersegmented.

To validate the effectiveness of WV and WRV, the two measures were computed for two sets of segmentation results of urban and rural images, as plotted in Fig. 3. In the case of oversegmentation, the WV and WRV values were relatively low, indicating that the segmentation was internally homogeneous but not externally heterogeneous. Moreover, in the case of undersegmentation, the WV and WRV values were relatively high, indicating an inverse situation to that of oversegmentation. As the scale parameter increased, the WV and WRV values had increasing trends, validating the potential of the proposed DV method in measuring the internal homogeneity and external heterogeneity. However, it is interesting to observe that the WRV values had more variation along the changing trend with the coarsening of the scale parameter than did the WV values. The reason for this may be that when producing a segment essentially in agreement with the corresponding geo-objects, the segment boundaries will be retained in the segmentation results with some subsequent scale parameters, and the external heterogeneity of this segment will remain the same [33]. However, some other segments will be wrongly merged into geo-objects, resulting in the WRV decreasing with some subsequent scale parameters. This phenomenon indicated that WRV is more sensitive to undersegmentation than WV is and provides a theoretical possibility for selecting multiple appropriate scale parameters for subsequent image analysis. Thus, we concluded that the DV (WV and WRV) method is a promising internal homogeneity and external heterogeneity metric for segmentation evaluations and optimal scale parameter determinations.

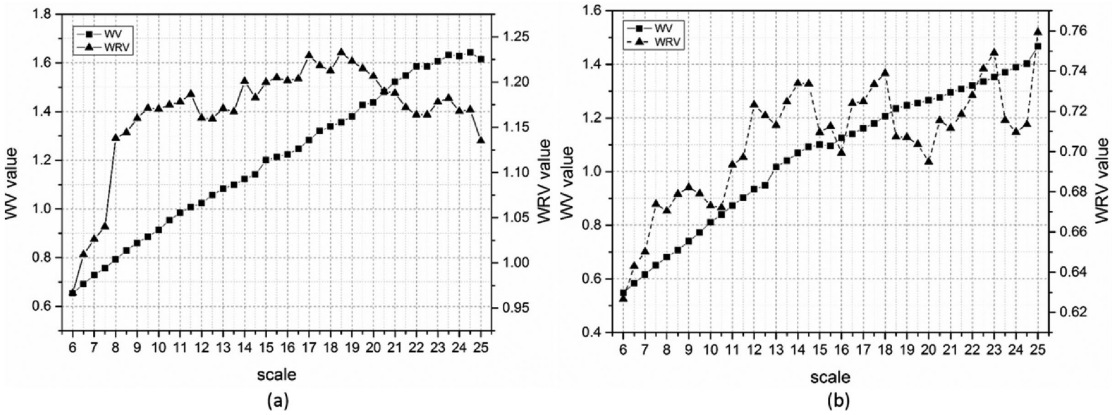


Fig. 3. WV and WRV values calculated for two set of segmentation results of (a) T1 and (b) T2.

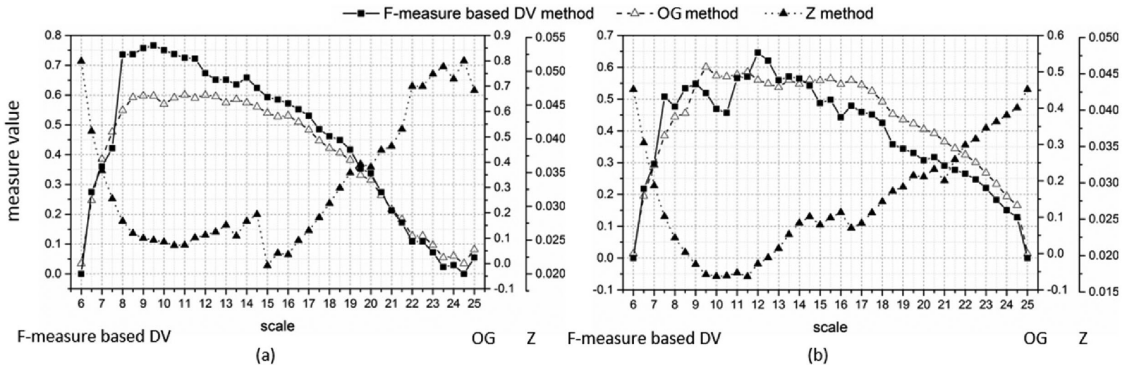


Fig. 4. Evaluation results of (a) T1 and (b) T2 obtained using the F-measure-based DV method, OG method, and Z method.

C. Comparative Analysis of the F-Measure-Based DV Method

1) Single Optimal SP Determination: The results of the F-measure-based DV method, OG method, and Z method are displayed in Fig. 4. As indicated by the highest values of the F-measure-based DV and OG and the lowest value of Z, the appropriate scale parameters obtained using the three single optimal SP selection methods were set at 9.5, 11, and 15 for the urban image and 12, 9.5, and 11.5 for the rural image, respectively.

To demonstrate which of these calculated parameters is best, visual and quantitative comparisons were conducted. For this experiment, 60 reference polygons (green and blue polygons) for each test image were generated to calculate the discrepancy metrics mentioned in Section II-C (see Fig. 2).

2) Comparative Experimental Results of the Urban Image (T1): Three subsets of the segmentation results of T1, obtained by the three scale parameters 9.5, 11, and 15, are presented in Fig. 5. For the first subset, the segmentation result obtained with a scale parameter of 9.5 allowed the differently sized buildings to be segmented well, whereas they were undersegmented to varying degrees in the segmentation results obtained with the scale parameters of 11 and 15. For the second subset, small geo-objects could be distinguished well in the segmentation results obtained with scale parameters of 9.5 and 11, but some small geo-objects were wrongly merged into other geo-objects in the segmentation result with a scale parameter of 15. For the third subset, the large building segmentations in the three segmentation results were similar. However, the tree cluster and

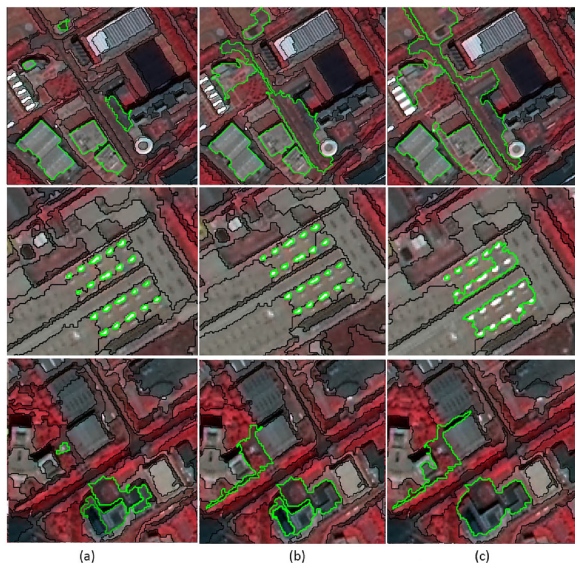


Fig. 5. Subsets of segmentation results of T1 obtained by the three scale parameters of 9.5, 11, and 15. (a) F-measure-based DV result: 9.5. (b) OG result: 11. (c) Z result: 15.

TABLE I
DISCREPANCY MEASURES OF THE SEGMENTATION RESULTS OBTAINED BY THE THREE SCALE PARAMETERS OF 9.5, 11, AND 15 FOR THE T1 IMAGE

Method	Scale	Discrepancy metrics						
		QR	OS	US	ED	P	R	F
F-measure based DV	9.5	0.3281	0.16	0.1916	0.1765	0.7544	0.4764	0.292
OG	11	0.3511	0.1347	0.2428	0.1963	0.7869	0.4293	0.2778
Z	15	0.3738	0.0811	0.3115	0.2276	0.8629	0.3353	0.2451

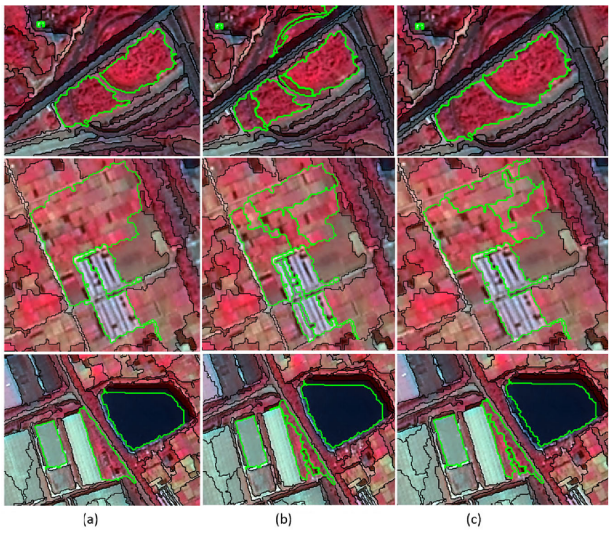


Fig. 6. Subsets of segmentation results of T2 obtained by the three scale parameters of 12, 9.5, and 11.5. (a) F-measure-based DV result: 12. (b) OG result: 9.5. (c) Z result: 11.5.

one small building were confused in the segmentation results obtained with the scale parameters of 11 and 15. Moreover, some medium-sized buildings were slightly and severely under-segmented in the segmentation results obtained with the scale parameters of 11 and 15, respectively. Overall, the single optimal scale parameter obtained by the F-measure-based DV method performed better than the other two methods in allowing good segmentation.

Then, seven discrepancy metrics, QR, OS, US, ED, P, R, and F, were computed for quantitative evaluations of the segmentation results obtained by the three scale parameters of 9.5, 11, and 15 (see Table I). As discussed in Section II-C, low QR, OS, US, and ED values and high P, R, and F values indicate good segmentation. The lowest QR and ED values and highest F value occurred in the segmentation result obtained by the F-measure-based DV method, indicating that the proposed method had a better overall performance in allowing good segmentation.

3) Comparative Experimental Results of the Rural Image (T2): For the rural image (T2), visual interpretation was first implemented, and subsets of the segmentation results at the scale parameters of 12, 9.5, and 11.5 are shown in Fig. 6. For the first subset, the segmentation results obtained with scale parameters of 11.5 and 12 displayed segmenting of the tree clusters well, but the road could not be identified in these two results. The situation was inverse in the segmentation result obtained with a scale parameter of 9.5. For the second subset, similar performances in segmenting the cultivation hothouse were presented in the segmentation results obtained with the scale parameters of 11.5 and 12, which was better than that obtained with a scale parameter of 9.5. However, varying degrees of oversegmentation of the farmland were found in the segmentation results with scale parameters of 9.5 and 11.5. For the third subset, the single optimal scale parameters obtained by the three methods produced similar segmentations of the buildings and a small pond. However, some tree clusters were oversegmented in the segmentation results obtained with the scale parameters of 9.5 and 11.5.

Similarly, the seven discrepancy measures were also calculated to quantitatively analyze the segmentation results obtained by the three scale parameters of 12, 9.5, and 11.5 (see Table II). The result showed a similar conclusion as Table I, indicating that the F-measure-based DV method had the most potential to obtain the single optimal scale parameter for producing the highest OG of segmentation. Although the proposed method had a better overall performance in determining the single optimal SP and producing good segmentation, the fact cannot be ignored that it is impossible to use a single optimal SP for segmenting every geo-object well in a segmentation. Thus, multiple optimal SP determination becomes critical in satisfying some remote sensing applications for different purposes. This article developed an LP-based DV method for the determination of multiple appropriate SPs, and detailed experimental results are presented in the next section.

D. Comparative Analysis of the LP-Based DV Method

1) Multiple Optimal SP Determination: The results of the LP-based DV method, E method, and SA method are shown in Fig. 7. It was interesting to observe that when applied to the urban image, the method produced the same number of measure values (i.e., 12) as were obtained for the rural image using the LP-based DV method, whereas the distribution had an obvious difference [see Fig. 5(a) and (d)]. However, a similar phenomenon was not observed in urban and rural results using the E method and SA method [see Fig. 5(b) and (e), Fig. 5(c) and (f)]. For the urban result obtained by the LP-based DV method, four scale parameters 8.5, 15, 17.5, and 20.5 had the greatest relative measure values and were thus the most likely candidates of the multiple optimal scale parameters [see Fig. 5(a)]. However, 7.5, 14.5, and 19 were considered the most likely multiple optimal scale parameters using the E method [see Fig. 5(b)]. The parameters 12.5, 15, and 21 were the likely candidates of the multiple optimal scale parameters as indicated by the result obtained by the SA method [see Fig. 5(c)]. For the rural image, the multiple optimal scale parameters obtained by the proposed method were 8, 12.5, 16, 18.5, and 23.5 [see Fig. 5(d)], whereas 7, 10, 14.5, 18.5, and 23.5 were regarded as the appropriate scale parameters when using the E method [see Fig. 5(e)]. The SA method determined

TABLE II
DISCREPANCY MEASURES OF THE SEGMENTATION RESULTS OBTAINED BY THE THREE SCALE PARAMETERS OF 12, 9.5, AND 11.5 FOR T2 IMAGE

Method	Scale	Discrepancy metrics						
		QR	OS	US	ED	P	R	F
F-measure based DV	12	0.3991	0.1163	0.3088	0.2333	0.8431	0.3264	0.2353
OG	9.5	0.4055	0.174	0.29	0.2391	0.8291	0.326	0.234
Z	11.5	0.4113	0.1293	0.3135	0.2398	0.8347	0.3261	0.2345

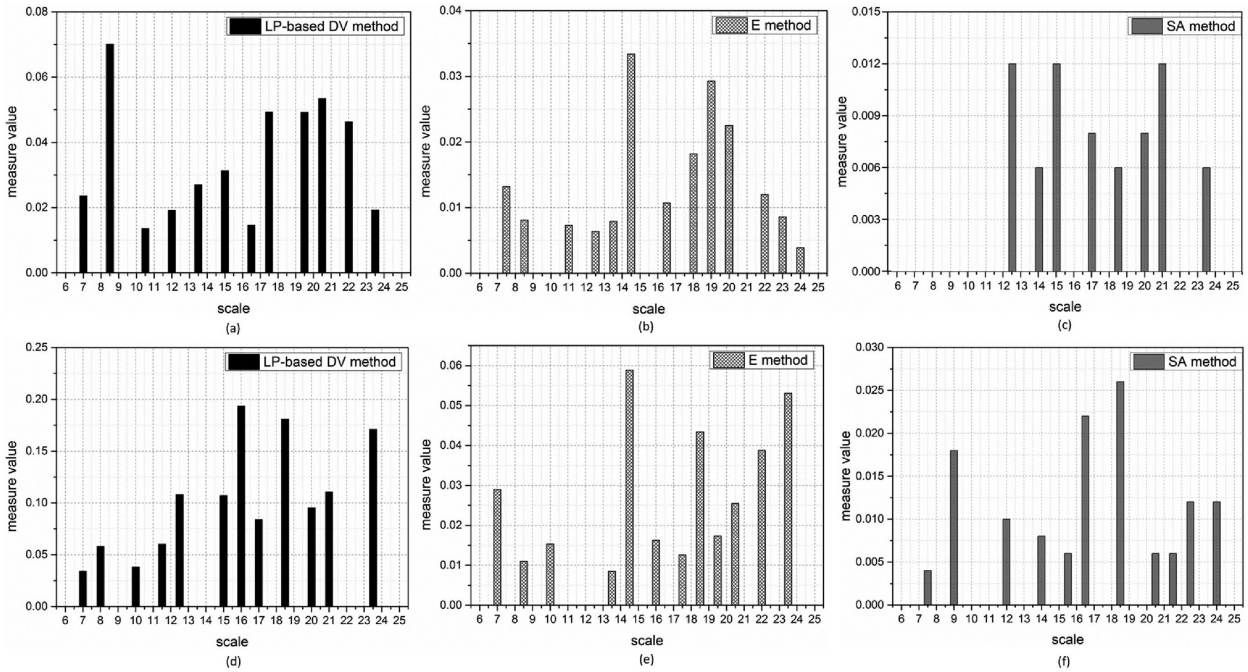


Fig. 7. Evaluation results using the LP-based DV method, E method, and SA method. (a) T1—LP-based DV result. (b) T1—E result. (c) T1—SA result. (d) T2—LP-based DV result. (e) T2—E method. (f) T2—SA method.

9, 18.5, and 22.5 as suitable scale parameters [see Fig. 5(f)]. In addition, an interesting phenomenon was observed: the single optimal scale parameters determined by the F-measure-based DV method, OG, and Z method were almost always among the intervals of the multiple optimal scale parameters obtained by the LP-based DV method, E method, and SA method.

To validate the proposed method, the scale parameters ranging from 6–25 were divided into three levels: the fine-scale level (6–12), medium-scale level (12–18), and coarse-scale level (18–25). Then, 30 reference polygons were generated from each level at each test image to calculate the discrepancy metrics mentioned in Section II-C (see Fig. 2). At the fine-scale level, the land is covered in small geo-objects, such as small buildings, flower beds, and farm fields. At the medium-scale level, the land is covered in residential buildings, workshops, grasslands, and tree clusters. At the coarse-scale level, the land is mainly covered in semantic geo-objects, including tree belts, industrial areas, residential areas, and farming areas.

2) Comparative Experimental Results of the Urban Image (T1): For the urban image (T1), the scale parameters of 8.5, 15, 17.5, and 20.5 obtained by the LP-based DV method, those of 7.5, 14.5, and 19 obtained by the E method, and those of 12.5, 15, and 21 obtained by the SA method were divided into three levels, as follows: the fine-scale level, including 7.5 and 8.5; the medium-scale level, including 12.5, 14.5, 15, and 17.5; and the coarse-scale level, including 19, 20.5, and 21. This article selected some subsets from each level to compare the segmentation results obtained by the three methods (see Fig. 8). At the fine-scale level [see Fig. 8(a)–(c)], for geo-objects, including small buildings, the segmentation results at the scale parameters of 8.5 and 7.5 were similar. However, the factories showed more oversegmentation in the results obtained at the scale parameter of 7.5. At the medium-scale level, similar segmentation results for the workshops were shown at the scale parameters of 14.5 and 15, and these results were better than those obtained at the scale parameters of 12.5 and 17. However, the grassland was oversegmented in that at the scale parameters of 12.5 and 14.5 [see Fig. 8(d)]. Among the three segments, the factories were delineated better in the segmentation obtained at the larger scale [see Fig. 8(e)]. At the coarse-scale level, for the residential area as a geo-object, the segmentation obtained at the scale parameter of 20.5 was better than that obtained at the scale parameter of 19 [see Fig. 8(f)]. For large geo-objects, such as industrial areas, segmentation at the scale parameter of 20.5 was better for fitting the semantic objects [see Fig. 8(g)]. The visual inspection indicated that the LP-based DV method had the most potential to allow for good segmentation in the three scale levels.

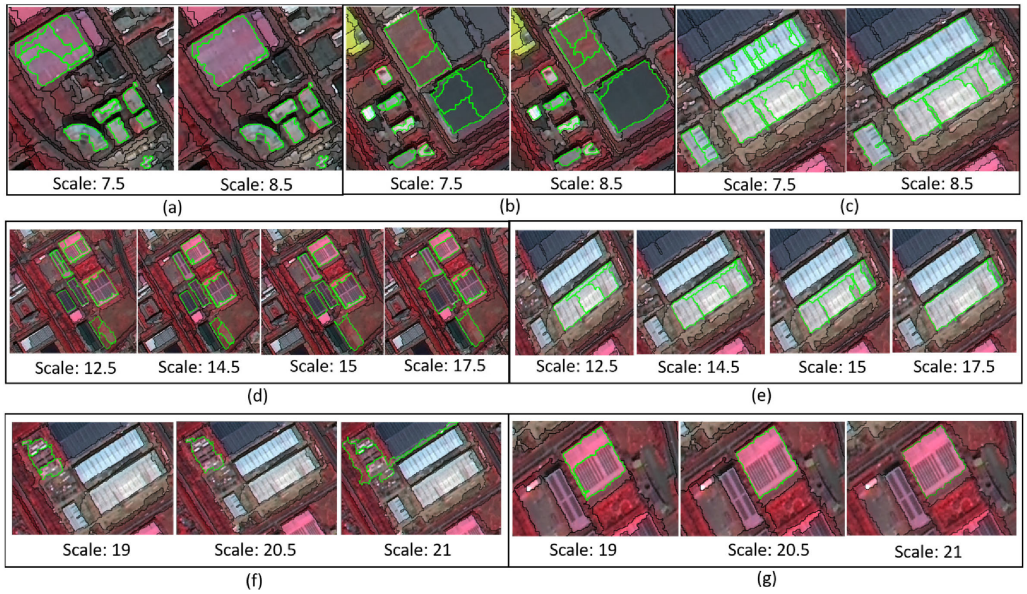


Fig. 8. Subsets of the segmentation results of T1 obtained by the LP-based DV method, E method, and SA method. (a)–(c) Three subsets of the fine level. (d) and (e) Two subsets of the medium level. (f) and (g) Two subsets of the coarse level.

TABLE III
DISCREPANCY MEASURES OF THE SEGMENTATION RESULTS OBTAINED BY THE LP-BASED DV METHOD, E METHOD, AND SA METHOD AT THE THREE LEVELS FOR THE T1 IMAGE

Level	Method	Scale	Discrepancy metrics						
			QR	OS	US	ED	P	R	F
fine-scale	LP-based DV	8.5	0.3137	0.1279	0.2095	0.1736	0.8713	0.5573	0.3399
	E	7.5	0.3562	0.1938	0.1954	0.1946	0.7191	0.5925	0.3249
	SA	-	-	-	-	-	-	-	-
medium-scale	LP-based DV	15	0.2153	0.1016	0.1283	0.1157	0.8529	0.3861	0.2658
		17.5	0.2676	0.076	0.2055	0.1549	0.9167	0.327	0.241
	E	14.5	0.2236	0.104	0.1343	0.1201	0.849	0.3774	0.2613
	SA	12.5	0.2512	0.1522	0.1176	0.136	0.8014	0.4158	0.2738
		15	0.2153	0.1016	0.1283	0.1157	0.8529	0.3861	0.2658
coarse-scale	LP-based DV	20.5	0.313	0.1888	0.1616	0.1757	0.7893	0.3479	0.2415
	E	19	0.3168	0.2016	0.1551	0.1798	0.779	0.3657	0.2489
	SA	21	0.3133	0.1885	0.1636	0.1765	0.8024	0.3369	0.2373

The seven discrepancy measures were computed at each scale level to further validate the LP-based DV method (see Table III). As indicated by the lowest QR and ED values, the scale parameters of 8.5, 15, and 20.5 obtained by the LP-based method could be used to produce segmentation results at each scale level that achieved the best match between the reference polygons and corresponding segments. However, the SA method produced the highest F value at the medium-scale level, and the E method obtained the highest F value at the coarse-scale level. The observed changes in the OS and P values at the fine-scale level were larger than those at the medium-scale and coarse-scale levels, indicating that the image was more sensitive to oversegmentation at the fine-scale level since the fine-scale level mainly focused on the recognition of small geo-objects. The observed changes in the US and R values at the medium-scale level were the largest among the three levels, indicating the most sensitivity to undersegmentation at the medium-scale level. A possible explanation for this result is that the medium-scale level is the transition phase between natural geo-objects and semantic geo-objects. A smaller scale parameter produced segmentation that delineated natural geo-objects well, whereas a larger scale parameter resulted in segmentation that identified semantic geo-objects [see Fig. 8(c) and (f)].

3) Comparative Experimental Results of the Rural Image (T2): For the rural image (T2), similarly, the scale parameters of 8, 12.5, 16, 18.5, and 23.5 obtained by the LP-based DV method, those of 7, 10, 14.5, 18.5, and 23.5 obtained by the E method, and those of 9, 18.5, and 22.5 obtained by the SA method were also grouped into three scale levels: the fine-scale level, including 7, 8, 9, and 10; the medium-scale level, including 12.5, 14.5, and 16; and the coarse-scale level, including 18.5, 22.5, and 23.5. Then, selected samples of the segmentation results obtained at the three scale levels are shown in Fig. 9. At the fine-scale level, segments obtained from the two scale parameters of 8 and 7 were similar and were better than those obtained from the scale parameters of 9 and 10 for individual paddy fields [see Fig. 9(a)]. The small buildings were segmented well in the segmentation results obtained by the four scale parameters, whereas the scale

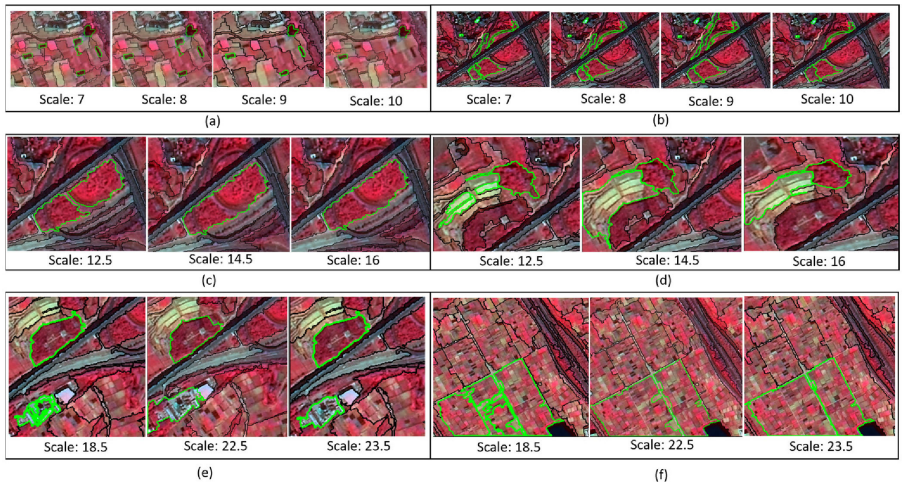


Fig. 9. Subsets of the segmentation results of T2 obtained by the LP-based DV method, E method, and SA method. (a) and (b) Two subsets of the fine level. (c) and (d) Two subsets of the medium level. (e) and (f) Two subsets of the coarse level.

TABLE IV
DISCREPANCY MEASURES OF THE SEGMENTATION RESULTS OBTAINED BY THE LP-BASED DV METHOD, E METHOD, AND SA METHOD AT THE THREE LEVELS FOR THE T2 IMAGE

Level	Method	Scale	Discrepancy metrics						
			QR	OS	US	ED	P	R	F
fine-scale	LP-based DV	8	0.2972	0.0995	0.2242	0.1734	0.8918	0.63	0.3692
	E	7	0.2632	0.1107	0.178	0.1482	0.8779	0.6626	0.3776
		10	0.4361	0.0822	0.3785	0.2739	0.9012	0.5661	0.3477
	SA	9	0.413	0.0873	0.3523	0.2567	0.8976	0.594	0.3575
medium-scale	LP-based DV	12.5	0.3388	0.1559	0.2216	0.1916	0.8477	0.4485	0.2933
		16	0.3744	0.063	0.3319	0.2389	0.9516	0.3685	0.2657
	E	14.5	0.3569	0.0936	0.2914	0.2164	0.9027	0.385	0.2699
	SA	-	-	-	-	-	-	-	-
coarse-scale	LP-based DV	18.5	0.4041	0.2041	0.2339	0.2195	0.6874	0.364	0.238
		23.5	0.3536	0.1239	0.2641	0.2063	0.8643	0.3405	0.2443
	E	18.5	0.4041	0.2041	0.2339	0.2195	0.6874	0.364	0.238
		23.5	0.3536	0.1239	0.2641	0.2063	0.8643	0.3405	0.2443
	SA	18.5	0.4041	0.2041	0.2339	0.2195	0.6874	0.364	0.238
		22.5	0.3545	0.1248	0.2638	0.2064	0.8343	0.3453	0.2442

parameters of 8 and 7 oversegmented the tree clusters and that of 10 undersegmented the road [see Fig. 9(b)]. At the medium-scale level, the building rooftops and tree clusters were recognized well at the scale parameter of 12.5, whereas undersegmentation was recognized at the other two scale parameters [see Fig. 9(c) and (d)]. However, if the group of building rooftops was considered a geo-object, the segments obtained at the scale parameters of 14.5 and 16 would be less oversegmented [see Fig. 9(d)]. At the coarse-scale level, the small building was delineated well because of the strong spectral contrast to the surroundings [see Fig. 9(e)]. The segmentation obtained at the scale parameter of 23.5 recognized the residential area better than that obtained at the scale parameters of 18.5 and 22.5 [see Fig. 9(e)]. The farming areas were identified well in the segmentation results at the scale parameter of 23.5 [see Fig. 9(f)]. The results indicated that the coarse-scale parameters obtained by the LP-based DV method had the most potential for segmenting semantic geo-objects and small geo-objects with strong spectral contrast to neighboring geo-objects.

Similar to those calculated for the urban image, the discrepancy measures calculated at each level are presented in Table IV and were used to quantitatively analyze the segmentation results obtained by the LP-based DV method, E method, and SA method for the rural image. The differences in QR, ED, and F indicated that the segmentation results at the scale parameters of 7, 12.5, and 23.5 had the highest segmentation quality at the fine-scale, medium-scale, and coarse-scale levels, respectively. The LP-based DV method performed better than the E method at the medium-scale and coarse-scale levels, but performed more poorly at the fine-scale level. However, the optimal scales obtained by the two methods were very close at the fine level (scale: 8 versus 7), and the performances shown in Fig. 9(a) and (b) were similar. The LP-based DV method had better performance than the SA method at every level.

IV. DISCUSSION

This article developed DV measures for recognizing more reasonable SPs to produce segments that best fit the geo-objects. The primary contribution of this work is the area WRV model, which is used jointly with the area WV metric. In the WRV model, the process is as follows: first, the variance between

each segment and one of its neighboring segments is calculated, and the contributions of the common border between each segment and one of its neighbors and the area of the neighbor are considered to form the RV of each segment. Then, the contribution of the area of each segment is considered to calculate the external heterogeneity. The process of calculating WRV is the same as the process of calculating WV, ensuring the same statistical characteristics regarding homogeneity and heterogeneity. In addition, the considerations of the contribution of the common border between each segment and one of its neighbors and the area of that neighbor is likely to add inherent stability in the WRV calculation, thus determining a more reasonable SP. The comparative experiment showed that the proposed method had a better overall performance than the competing methods, demonstrating the potential of DV-based measures in improving the accuracy of segmentation evaluations.

This article adopted two combination strategies to determine the single optimal parameter and a set of useful scale parameters. The F-measure-based DV method is suitable for obtaining good segmentation in which there is a good compromise for recognizing differently sized natural geo-objects. The LP-based DV method is more likely to be used to obtain good segments of small, natural geo-objects to large, semantic geo-objects. Compared with the F-measure-based DV method, the LP-based method could help produce segments, including semantic geo-objects. However, LP-based strategy, when used for multiple optimal scale parameter identification, leads to a set of scale parameters that do not include the scale determined by using the F-measure strategy. The possible reason for this result is that the F-measure is a global approach that considers the OG of segmentation to determine the optimal scale parameter. In contrast, LP is a local approach that finds the change in a certain SP level to recognize a set of useful scale parameters. It is unlikely that the LP method will identify the same scale parameter as the F-measure method due to the difference between the methods in determining the optimal scale parameter idea. In addition, combining the segmentation results obtained by the different optimal SPs was the main difficulty for subsequent image analysis. Thus, future work will focus on the approach of producing an ultimate segmentation result in which the best-delineated geo-objects are chosen from the corresponding segmentation obtained by the multiple optimal SPs.

Accuracy and efficiency are two critical factors for selecting which combination strategy should be used to determine the optimal SPs, and these two factors must be carefully weighted to satisfy different types of applications [50]. The massive number of different sizes and spectral features of geo-objects that occur in high-resolution satellite imagery make it impossible that every geo-object is segmented well in a given segmentation obtained by a certain optimal SP. From this perspective, a group of optimal SPs, rather than a single parameter, is required to obtain good segments for the identification of small, natural geo-objects to large, semantic geo-objects. Thus, accuracy is considered the most important factor, and the LP-based DV method is thus a good alternative to conventional methods. However, if we want to classify a specific type of natural geo-object, such as a building, the F-measure-based DV method may be a more suitable selection. The large factories were oversegmented in the segmentation results obtained using the fine- and medium-scale parameters [see Fig. 8(c) and (e)], whereas the small and medium buildings were delineated well with these parameters [see Fig. 8(a), (b), and (d)]. The classification accuracy would not be reduced if this segmentation result, obtained by using fine- or medium-scale parameters, was used to classify the natural geo-objects of the buildings. From this perspective, the consideration of efficiency should be given more weight than that of accuracy. The F-measure-based DV method is a better alternative than the LP-based DV method since it provides a good compromise for segmenting differently sized natural geo-objects. It is not appropriate to say which of the two methods is better since they each have their unique application advantages. It is instead a matter for researchers to choose suitable segmentation evaluation methods according to different application purposes.

The rule of selecting validating reference polygons is different between the F-measure-based DV method and the LP-based DV method. The geo-objects recognized by the single optimal scale parameter are natural geo-objects in the urban and rural images, and differently sized natural geo-objects occur in the segmentation results obtained with scales at the fine-scale and medium-scale levels. Thus, the union of reference polygons at the fine-scale and medium-scale levels was used for validating the segmentation results of the F-measure-based DV method. The geo-objects identified by the multiple optimal scales vary from small, natural geo-objects to large, semantic geo-objects. Thus, the reference polygons at the fine-scale, medium-scale, and coarse-scale levels need to be used to validate the segmentation results of the LP-based DV method. However, different people have different opinions on the definition of semantic geo-objects. It may be slightly subjective to generate the reference polygons of semantic geo-objects in this study. However, we have determined the reference polygons of semantic geo-objects from the perspective of common knowledge as much as possible.

The visual and quantitative results (see Figs. 5 and 6 and Tables I and II) demonstrated that the F-measure-based DV method has more potential to obtain a single optimal SP than other methods, and this optimal parameter could produce good segmentation in which there is a good compromise for differently sized natural geo-objects. However, the performance of the LP-based DV method was not as good as that of the F-measure-based DV method, especially in the rural image. The quantitative results obtained at the fine-scale level, shown in Table IV, showed that the E method performed better than the proposed method, as indicated by its lowest QR and ED values and highest F value. Even so, the proposed method still showed good performance at the fine-scale level in the rural image. This result revealed that no single measure indicator performed the best at any scale level. In the future, the measurement indicator will be further studied and improved to obtain more objective SPs. It was also observed that the SA method could not be used to determine the optimal scale at the fine-scale level in the urban image (see Table III), or at the medium-scale level in the rural image (see Table IV). The possible reason for this result may be

that the SA method only considers homogeneity in determining multiple optimal scales.

In addition, the WRV measure showed different changing trends between the urban and rural images (see Fig. 3). There was a trend of first increasing and then decreasing in the urban image, whereas the trend was increasing in the rural image. The possible reason for this result may be that the spectral contrast between the geo-objects and surrounding geo-objects in the rural image is lower than that in the urban image, reducing the advantage of the WRV measure in the heterogeneity calculations. The different changes in the OS and P and the US and R indicators shown in Tables III and IV also demonstrated this opinion. The large changes observed in the OS and P at the fine-scale level and those in the US and R at the medium-scale level indicated the high spectral contrast present in the urban image (see Table III). The similar changes observed in the OS and P and the US and R at the three scale levels indicated the low spectral contrast present in the rural image (see Table IV).

V. Conclusion

In this article, a USE method that uses DV measures to calculate homogeneity and heterogeneity was proposed for recognizing more a reasonable SP to produce segments that best fit the geo-objects. Then, the F-measure and LP combination strategies were used to determine single and multiple optimal scale parameters. To validate the proposed method, the MRS algorithm was applied to segment GF-1 images, and a series of segmentation results at a certain range of scale parameters was obtained. The comparative results indicated that the DV (WV and WRV) method is a more promising internal homogeneity and external heterogeneity metric for segmentation evaluations and optimal scale parameter determinations. The F-measure-based DV method could produce better OG of segmentation for differently sized natural geo-objects compared with the OG and Z methods. The LP-based DV method could obtain multiple optimal scales that produced better segments for the identification of small, natural geo-objects to large, semantic geo-objects compared to the E and SA methods. In the future, the further improvement of the DV measures and the approach of combining the segmentation results obtained by multiple optimal scales into one ultimate segmentation result will be our focus.

Acknowledgment

The authors would like to thank the editor and anonymous reviewers for contributing their comments to improve this article.

References

[1] X. L. Zhang, X. Z. Feng, P. F. Xiao, G. J. He, and L. J. Zhu, "Segmentation quality evaluation using region-based precision and recall measures for remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 102, pp. 73–84, Apr. 2015.

[2] J. Yang, P. J. Li, and Y. H. He, "A multi-band approach to unsupervised scale parameter selection for multi-scale image segmentation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 94, pp. 13–24, Aug. 2014.

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

[4] M. D. Hossain and D. Chen, "Segmentation for object-based image analysis (OBIA): A review of algorithms and challenges from remote sensing perspective," *ISPRS J. Photogrammetry Remote Sens.*, vol. 150, pp. 115–134, Apr. 2019.

[5] K. Johansen, L. A. Arroyo, S. Phinn, and C. Witte, "Comparison of geo-object based and pixel-based change detection of riparian environments using high spatial resolution multi-spectral imagery," *Photogrammetric Eng. Remote Sens.*, vol. 76, no. 2, pp. 123–136, Feb. 2010.

[6] G. J. Hay and G. Castilla, *Geographic Object-Based Image Analysis (GEOBIA): A New Name for a New Discipline*. Berlin, Germany: Springer, 2008, pp. 75–89.

[7] C. Cleve, M. Kelly, F. R. Kearns, and M. Morltz, "Classification of the wildland-urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography," *Comput. Environ. Urban Syst.*, vol. 32, no. 4, pp. 317–326, Jul. 2008.

[8] X. L. Zhang, P. F. Xiao, X. Z. Feng, J. G. Wang, and Z. Wang, "Hybrid region merging method for segmentation of high-resolution remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 98, pp. 19–28, Dec. 2014.

[9] A. Pekkarinen, "A method for the segmentation of very high spatial resolution images of forested landscapes," *Int. J. Remote Sens.*, vol. 23, no. 14, pp. 2817–2836, Jul. 20, 2002.

[10] P. N. Happ, R. S. Ferreira, C. Bentes, G. A. O. P. Costa, and R. Q. Feitosa, "Multiresolution segmentation: A parallel approach for high resolution image segmentation in multicore architectures," in *Proc. Int. Conf. Geograph. Object-Based Image Anal.*, 2010, vol. 38-4-C7.

[11] 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.

[12] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[13] N. Wang, F. Chen, B. Yu, and Y. C. Qin, "Segmentation of large-scale remotely sensed images on a Spark platform: A strategy for handling massive image tiles with the MapReduce model," *ISPRS J. Photogrammetry Remote Sens.*, vol. 162, pp. 137–147, Apr. 2020.

[14] T. F. Su, T. X. Liu, S. W. Zhang, Z. Y. Qu, and R. P. Li, "Machine learning-assisted region merging for remote sensing image segmentation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 168, pp. 89–123, Oct. 2020.

[15] Y. J. Wang, Q. Y. Meng, Q. W. Qi, J. Yang, and Y. Liu, "Region merging considering within- and between-segment heterogeneity: An improved hybrid remote-sensing image segmentation method," *Remote Sens.*, vol. 10, no. 5, May 2018, Art. no. 781.

[16] J. Yang, Y. H. He, and J. Caspersen, "Region merging using local spectral angle thresholds: A more accurate method for hybrid segmentation of remote sensing images," *Remote Sens. Environ.*, vol. 190, pp. 137–148, Mar. 2017.

[17] Z. W. Hu, Q. Zhang, Q. Zou, Q. Q. Li, and G. F. Wu, "Stepwise evolution analysis of the region-merging segmentation for scale parameterization," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2461–2472, Jul. 2018.

[18] T. F. Su and S. W. Zhang, "Multi-scale segmentation method based on binary merge tree and class label information," *IEEE Access*, vol. 6, pp. 17801–17816, 2018.

[19] Y. J. Wang, Q. W. Qi, L. L. Jiang, and L. Ying, "Hybrid remote sensing image segmentation considering intrasegment homogeneity and intersegment heterogeneity," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 22–26, Jan. 2020.

[20] D. Ming, J. Li, J. Wang, and M. Zhang, "Scale parameter selection by spatial statistics for GeOBIA: Using mean-shift based multi-scale segmentation as an example," *ISPRS J. Photogrammetry Remote Sens.*, vol. 106, pp. 28–41, 2015.

[21] J. Goncalves, I. Pocas, B. Marcos, C. A. Mucher, and J. P. Honrado, "SegOptim—A new R package for optimizing object-based image analyses of high-spatial resolution remotely-sensed data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 76, pp. 218–230, Apr. 2019.

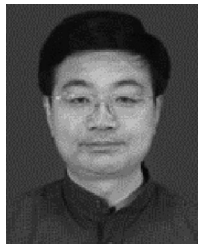
[22] Y. Liu *et al.*, "Discrepancy measures for selecting optimal combination of parameter values in object-based image analysis," *ISPRS J. Photogrammetry Remote Sens.*, vol. 68, pp. 144–156, Mar. 2012.

[23] D. P. Ming, T. Y. Ci, H. Y. Cai, L. X. Li, C. Qiao, and J. Y. Du, "Semivariogram-based spatial bandwidth selection for remote sensing image segmentation with mean-shift algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 5, pp. 813–817, Sep. 2012.

[24] B. Johnson and Z. X. Xie, "Unsupervised image segmentation evaluation and refinement using a multi-scale approach," *ISPRS J. Photogrammetry Remote Sens.*, vol. 66, no. 4, pp. 473–483, Jul. 2011.
[25] S. Chabrier, B. Emile, C. Rosenberger, and H. Laurent, "Unsupervised performance evaluation of image segmentation," *EURASIP J. Appl. Signal Process.*, vol. 2006, 2006, Art. no. 096306.
[26] T. F. Su and S. W. Zhang, "Local and global evaluation for remote sensing image segmentation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 130, pp. 256–276, Aug. 2017.
[27] N. Clinton, A. Holt, J. Scarborough, L. Yan, and P. Gong, "Accuracy assessment measures for object-based image segmentation goodness," *Photogrammetric Eng. Remote Sens.*, vol. 76, no. 3, pp. 289–299, Mar. 2010.
[28] J. Yang, Y. H. He, J. Caspersen, and T. Jones, "A discrepancy measure for segmentation evaluation from the perspective of object recognition," *ISPRS J. Photogrammetry Remote Sens.*, vol. 101, pp. 186–192, Mar. 2015.
[29] H. Zhang, J. E. Fritts, and S. A. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," *Comput. Vis. Image Understanding*, vol. 110, no. 2, pp. 260–280, May 2008.
[30] H. J. Tong, T. Maxwell, Y. Zhang, and V. Dey, "A supervised and fuzzy-based approach to determine optimal multi-resolution image segmentation parameters," *Photogrammetric Eng. Remote Sens.*, vol. 78, no. 10, pp. 1029–1044, Oct. 2012.
[31] J. Yang, Y. H. He, and Q. H. Weng, "An automated method to parameterize segmentation scale by enhancing intrasegment homogeneity and intersegment heterogeneity," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1282–1286, Jun. 2015.
[32] G. M. Espindola, G. Camara, 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.
[33] L. Dragut, D. Tiede, and S. R. Levick, "ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data," *Int. J. Geograph. Inf. Sci.*, vol. 24, no. 6, pp. 859–871, 2010.
[34] T. F. Su, "Unsupervised evaluation-based region merging for high resolution remote sensing image segmentation," *GISci. Remote Sens.*, vol. 56, no. 6, pp. 811–842, Aug. 2019.
[35] X. L. Zhang, P. F. Xiao, and X. Z. Feng, "An unsupervised evaluation method for remotely sensed imagery segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 2, pp. 156–160, Mar. 2012.
[36] Y. Wang, Q. Qi, and Y. Liu, "Unsupervised segmentation evaluation using area-weighted variance and Jeffries-Matusita distance for remote sensing images," *Remote Sens.*, vol. 10, no. 8, 2018, Art. no. 1193.
[37] B. A. Johnson, M. Bragais, I. Endo, D. B. Magcale-Macandog, and P. B. M. Macandog, "Image segmentation parameter optimization considering within- and between-segment heterogeneity at multiple scale levels: Test case for mapping residential areas using Landsat imagery," *ISPRS Int. J. Geo-Inf.*, vol. 4, no. 4, pp. 2292–2305, Dec. 2015.
[38] Y. Gao, J. F. Mas, N. Kerle, and J. A. N. Pacheco, "Optimal region growing segmentation and its effect on classification accuracy," *Int. J. Remote Sens.*, vol. 32, no. 13, pp. 3747–3763, 2011.
[39] F. Canovas-Garcia and F. Alonso-Sarria, "A local approach to optimize the scale parameter in multiresolution segmentation for multispectral imagery," *Geocarto Int.*, vol. 30, no. 8, pp. 937–961, Sep. 2015.
[40] T. R. Martha, N. Kerle, C. J. van Westen, V. Jetten, and K. V. Kumar, "Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4928–4943, Dec. 2011.
[41] A. M. Vamsee, P. Kamala, T. R. Martha, K. V. Kumar, G. J. Sankar, and E. Amminedu, "A tool assessing optimal multi-scale image segmentation," *J. Indian Soc. Remote Sens.*, vol. 46, no. 1, pp. 31–41, Jan. 2018.
[42] M. Baatz and A. Schäpe, "Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation," in *Proc. Beiträge zum AGIT-Symp.*, 2000, pp. 12–23.
[43] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.
[44] K. Haris, S. N. Efstratiadis, N. Maglaveras, and A. K. Katsaggelos, "Hybrid image segmentation using watersheds and fast region merging," *IEEE Trans. Image Process.*, vol. 7, no. 12, pp. 1684–1699, Dec. 1998.
[45] U. W. Ulrike Sturm, "Further investigations on segmentation quality assessment for remote sensing applications," in *ISPRS Hannover Workshop 2009 High-Resolution Earth Imaging Geospatial Inf.*, 2009.
[46] W. H. Sun, B. Chen, and D. W. Messinger, "Nearest-neighbor diffusion-based pan-sharpening algorithm for spectral images," *Opt. Eng.*, vol. 53, no. 1, Jan. 2014, Art. no. 013107.
[47] C. Benedek, X. Descombes, and J. Zerubia, "Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 33–50, Jan. 2012.
[48] Y. Y. Zhou and Y. Q. Wang, "Extraction of impervious, surface areas from high spatial resolution imagery by multiple agent segmentation and classification," *Photogrammetric Eng. Remote Sens.*, vol. 74, no. 7, pp. 857–868, Jul. 2008.
[49] L. Yan and D. P. Roy, "Conterminous united states crop field size quantification from multi-temporal Landsat data," *Remote Sens. Environ.*, vol. 172, pp. 67–86, Jan. 2016.
[50] J. K. Udupa *et al.*, "A framework for evaluating image segmentation algorithms," *Comput. Med. Imag. Graph.*, vol. 30, no. 2, pp. 75–87, Mar. 2006.

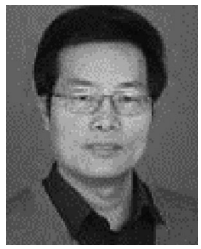
Yongji Wang received the Ph.D. degree in cartography and geographical information system from the State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, in 2020.

He is currently a Lecturer with the School of Geoscience and Technology, Zhengzhou University, Zhengzhou, China. His study interests include remote sensing image processing and GIS spatial analysis.



Zhihui Tian received the Ph.D. degree in cartography and geographical information system from Wuhan University, Wuhan, China.

He is currently a Professor with the School of Geoscience and Technology, Zhengzhou University, Zhengzhou, China. His study interests include remote sensing image processing, smart city engineering, and high-performance geographic computing.



Qingwen Qi received the Ph.D. degree in cartography and geographical information system from the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, in 1996.

He is currently a Professor with the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences. His research interests include methodology of geoinformation sciences, analysis and application of geographic information, and cartographic generalization automation.

Jun Wang is currently working toward the Ph.D. degree in cartography and geographic information engineering from the Shandong University of Science and Technology, Qingdao, China.

Since 2018, he has been a Visiting Doctor with the Institute of Geographic Science and Resources, Chinese Academy of Sciences, Beijing, China. His research interests include computer theory and technology, the theory of legal geography, geographic information science and methods, the theory and technology of agricultural informatization, and remote sensing image segmentation technology.