Segmentation for Object-Based Image Analysis (OBIA): A review of algorithms and challenges from remote sensing perspective

Mohammad D. Hossain, Dongmei Chen
Laboratory of Geographic Information and Spatial Analysis, Department of Geography and Planning, Queen’s University, Kingston, ON K7L 3N6, Canada

A R T I C L E   I N F O

Keywords:
OBIA
Remote sensing
High spatial resolution
Image segmentation
Geographic object

A B S T R A C T

Image segmentation is a critical and important step in (GEographic) Object-Based Image Analysis (GEObIA or OBIA). The final feature extraction and classification in OBIA is highly dependent on the quality of image segmentation. Segmentation has been used in remote sensing image processing since the advent of the Landsat-1 satellite. However, after the launch of the high-resolution IKONOS satellite in 1999, the paradigm of image analysis moved from pixel-based to object-based. As a result, the purpose of segmentation has been changed from helping pixel labeling to object identification. Although several articles have reviewed segmentation algorithms, it is unclear if some segmentation algorithms are generally more suited for (GE)OBIA than others. This article has conducted an extensive state-of-the-art survey on OBIA techniques, discussed different segmentation techniques and their applicability to OBIA. Conceptual details of those techniques are explained along with the strengths and weaknesses. The available tools and software packages for segmentation are also summarized. The key challenge in image segmentation is to select optimal parameters and algorithms that can general image objects matching with the meaningful geographic objects. Recent research indicates an apparent movement towards the improvement of segmentation algorithms, aiming at more accurate, automated, and computationally efficient techniques.

1. Introduction

Remote sensing technology has been widely used to extract land cover/use information efficiently as it has the ability to obtain data for a large area repeatedly (Pu and Landry, 2012). Images captured by earlier remote sensing sensors such as AVHRR, MSS, TM usually had pixels bigger than ground features, requiring sub-pixels or per-pixel analysis for features mapping (Blaschke, 2010). However, after the launch of the IKONOS (IK) satellite in 1999, the spatial resolution of images increased significantly (further improved in QuickBird (QB), WorldView-1 (WV-1), WorldView-2 (WV-2), WorldView-3 (WV-3), WorldView4 (WV-4), other recent sensors and UAVs). Pixel-based methods used for moderate and low-resolution imagery fail to utilize the spatial variation of different land covers in the high-resolution images (Campbell and Wynne, 2011) as these methods do not consider neighboring pixels which are the part of the same land cover. Consequently, (GEographic) Object-Based Image Analysis (GEObIA or OBIA) has emerged as an effective way of analyzing high spatial resolution images (Blaschke, 2010).

OBIA is an alternative to a pixel-based method with basic analysis unit as image objects instead of individual pixels (Castilla and Hay, 2008; Blaschke, 2010). This method intends to bypass the problem of artificial square cells as used in per-pixel method (Fisher, 1997; Burnett and Blaschke, 2003; Blaschke, 2010) by grouping a number of pixels into shapes with a meaningful representation of the objects. The aim of OBIA is to address more complex classes that are defined by spatial and hierarchical relationships within and during the classification process (Lang, 2008). OBIA is usually composed of two main phases: (1) image segmentation, and (2) feature extraction and classification. The most basic and critical step is image segmentation (Blaschke et al., 2008; Cheng et al., 2001; Zhang, 1997) and the accuracy of following object-based feature extraction and classification mainly depends on the quality of image segmentation (Mountrakis et al., 2011; Su and Zhang, 2017). Image segmentation is defined as a method of dividing an image into homogeneous regions (Pal and Pal, 1993). These regions represent land covers such as buildings, trees, water bodies, and grasslands which are known as image object in GEObIA (Costa et al., 2018; Heumann, 2011).

Image segmentation has been utilized differently in different fields (Kerfoot and Bresler, 1999; Pham et al., 2000) such as computer vision, medical imaging, and range imaging. Many methods applied to remote sensing imageries are imported from other fields as the underlying
principle is the same (Dey et al., 2010). Many segmentation algorithms have been developed and used in different fields. In remote sensing, hundreds of papers involving image segmentation have been published (Zhang, 2006) although many algorithms are not applicable to the object-based model (Davis and Wang, 2003). It is essential to identify algorithms available for object-based segmentation along with their pros and cons to ascertain the efficacy of individual algorithm.

Several previous papers have reviewed and discussed different segmentation techniques such as Haralick and Shapiro, 1985; Pal and Pal, 1993; Schiewe, 2002; Carlee et al., 2005; Shankar, 2007; Dey et al., 2010, 2013; Blaschke, 2010; Chen et al., 2018a. The oldest two reviews were mostly concerned with pixel-based segmentation techniques such as thresholding, Markov Random Field (MRF), neural network, and fuzzy clustering for segmenting moderate resolution images. The next four reviews have focused on the applicability of previously developed segmentation algorithms on high-resolution satellite images and the underlying concepts of segmentation methods. The review paper published in 2010 (Blaschke, 2010) went through the development of GEOBIA, application of multiresolution segmentation (MRS), algorithms developed for OBIA and their future trends. The one published in 2013 (Dey et al., 2013) tried to identify rules for selecting segmentation techniques from current algorithms for urban land cover mapping and provide some insight on possible methods that can be used for the same purpose. However, image segmentation is domain specific (Zouagui et al., 2004) and still a subject of ongoing research despite currently available numerous techniques. The latest one (Chen et al., 2018a) reviewed the emerging trends, and future opportunities of GEOBIA and image segmentation was only a small part of it. None of the previous review papers provided a particular focus and complete picture on object-based segmentation algorithms used in the remote sensing community.

This paper reviews the current object-based segmentation algorithms and has tried to identify their pros and cons. Peer-reviewed journals published between 1999 (when IK satellite launched) and 2018 have been searched using websites such as Google Scholar, Web of Knowledge and Scopus (Elsevier) with the keywords of segmentation, remote sensing and object-based. This review found 290 articles that matched with the keywords and the majority of them are published in the following top remote sensing journals: Photogrammetric Engineering & Remote sensing, ISPRS Journal of Photogrammetry and Remote Sensing, Remote Sensing of Environment, International Journal of Remote Sensing, International Journal of Applied Earth Observation and Geoinformation, IEEE Journal of Selected Topics on Applied Earth Observations and Remote Sensing, IEEE Transactions on GeoScience and Remote Sensing, Remote Sensing, Remote Sensing Letters and Pattern Recognition. Half of the articles are not included in previous review papers. Apart from those articles, some earlier articles and articles from other journals have also been cited for readers who are interested in understanding the fundamental and mathematical formulation of algorithms. This review is different from previous ones with the particular attention on object-based segmentation. It is also not limited to urban land cover mapping only, and it tries to indicate the strengths and weaknesses of each method. The rest of the paper is broadly organized as follows. Section 2 presents a brief discussion on GEOBIA and OBIA. Section 3 presents different segmentation algorithms used in object-based image analysis including edge- and region-based, hybrid methods, and semantic techniques. The section that follows describes the challenges in segmentation methods. Finally, it provides a summary of current issues.

2. OBIA and GEOBIA

Methods for low-resolution image processing are based on the classification of individual pixels (Blaschke et al., 2014). In low-resolution images, individual pixel contains one or even multiple land cover classes. By contrast, the intra-class spectral variability is significant in the high-resolution images (Blaschke et al., 2004). As a result, pixel-based algorithms are failing to provide better accuracy in high-resolution image analysis (Blaschke et al., 2004; Pu et al., 2011; Tehrany et al., 2014). In Geographic Information Science (GIScience), the single land cover is represented as an object and further analysis is conducted based on objects instead of pixels. Object-Based Image Analysis (OBIA) has been defined as “a sub-discipline of GIScience devoted to partitioning remote sensing (RS) imagery into meaningful image-objects and assessing their characteristics through spatial, spectral and temporal scale” (Hay and Castilla, 2006). The primary purpose of OBIA is to provide a method for analyzing high-spatial resolution imagery by using spectral, spatial, textural and topological characteristics (Lang, 2008). OBIA incorporate both geographic information (GI) and remote sensing. Image analysis is also done in other disciplines, such as computer vision, material science or biomedical imaging. Blaschke et al. (2004) have introduced the term ‘GEographic Object-Based Image Analysis (GEOBIA)’ to indicate image analysis performed by remote sensing scientists, GIS specialists, and environmental disciplines. Surveys such as that conducted by Blaschke (2010) have identified 145 peer-reviewed journal paper relevant to GEOBIA. However, the literature search carried out by Blaschke et al. (2014) reported over 600 relevant journal articles on the same issue which indicates that numbers have quadrupled over four years. This article also undertakes a brief literature survey using websites such as Google Scholar, Web of Knowledge and Scopus (Elsevier) with the keywords of OBIA, GEOBIA, segmentation, remote sensing, object-based, object-oriented, per-pixel, and other various spelling alternatives. As demonstrated in Fig. 1, the number of articles increased significantly at the same pace. There are two substantial reasons for this: availability of high spatial resolution remote sensing images and software (both commercial and open source) for implementing GEOBIA. In GEOBIA, it is assumed that image objects produced by segmentation can be explicitly linked to the geographic objects of interest (Shackelford and Davis, 2003; Zhou et al., 2007). Thus, segmentation is the key to the GEOBIA (Lizarazo and Elsner, 2011).

3. Segmentation

As mentioned in the Introduction section, the objective of segmentation is to partition an image into a set of disjointed regions that are different according to specific properties such as texture, color, shape, size and gray level (Lucchese and Mitray, 2001). Mathematically, segmentation can be defined as follows (Cheng et al., 2001): P() is the homogeneity criteria, R is the entire image and (Ri) will be a segment of R if: (1) Ri ⊆ R (2) R = ∪R(i), (3) R(i) ∩ R(j) = Ø (4) P(R(i) ∪ R(j)) = False when i ≠ j and Ri and Rj are neighbors. Earlier literature have categorized segmentation as (a) Pixel-based (Mardia and Hainsworth, 1988) (b) Edge-based (Perona and Malik, 1990) (c) Region-based (Beveridge et al., 1989) and (d) Hybrid method (Haris et al., 1998) based on object identification method. Pixel-based methods consist of image thresholding and segmentation in the feature space (Schiewe, 2002). In this case, each spatially continuous unit needs to be assigned a unique label. However, this method is not suitable for OBIA (Wang et al., 2015c) thus it does not warrant further discussion. Apart from the earlier classification, segmentation algorithms are also classified based on hierarchy (Guindon, 1997), object extraction method (Maxwell and Zhang, 2006), object representation method (Rosenfield and Davis, 1979), the homogeneity criteria (Baatz and Schäpe, 2000). Besides, Zhang (1997) classified segmentation algorithms as the boundary- and region-based approach based on discontinuity and similarity of object areas. This review is following earlier classification system based on how segments are generated and the subsequent sections provide details of those methods.

3.1. Edge-based segmentation

Edge-based techniques (Haralick, 1981; Ikonomopoulos, 1982;
Kundu and Pal, 1986; Lu and Jain, 1989; Zhou et al., 1989) first identify edges and then close them by using contouring algorithms (Cao et al., 2016). It assumes that between edges, the pixel properties change abruptly (Shih and Cheng, 2004). From this perspective, edges are regarded as boundaries between objects and located where changes occur (Martin et al., 2004). There are many algorithms available for the identification of the object edges that capture the geometrical and physical characteristics of image objects. A variety of edge detectors has been used for different purposes. Jain et al. (1995) divided edge detection into three steps: filtering, enhancement, and detection. Different filtering methods have been proposed (Castilla et al., 2008; Chen et al., 2006; Chen et al., 1999; Kerem and Ulusoy, 2013) to produce minimum blurring and displacement of edges. Enhancement accentuates pixels where there is a significant change in local intensity values (Fosgate et al., 1997). Enhanced data are used for detecting real edges. Many hard-coded operators (Andrey and Tarroux, 1998; Canny, 1987; Deriche, 1990; Đuríkovič et al., 1995; Farid and Simoncelli, 1997; He et al., 2014; Kass et al., 1988; Leymarie and Levine, 1993; Marr and Hildreth, 1980; Mayunga et al., 2007; Meer and Georgescu, 2001; Peng et al., 2005; Prewitt, 1970; Robinson, 1977) are employed for edge detection. Apart from those, scholars’ also implemented soft computing methods such as fuzzy-based approach (Trivedi and Bezdek, 1986), genetic algorithm-based approach (Natowicz et al., 1995), and neural network-based approach (Manjunath et al., 1990). Nevertheless, all operators create broken edges and miss some essential edges (Jevtic et al., 2009). The performance of each operator is evaluated based on the false edge, missing edge, edge angle, distance from the true edge and distortion (Lacherez and Mitray, 2001).

After the identification of edges, the next step is transforming the edges into closed boundaries. This step often involves the exclusion of edges that are produced by noise, the connecting of gaps at places where no edge is detected and decisions to join those edge parts that make up a single object. Multiple edge-linking methods (Jain et al., 1994; Liu et al., 2008; Lu and Chen, 2008) have been suggested to compensate edges that are not entirely linked. Scholars also utilize Hough transform (Ballard, 1981; Kiryati and Eldar, 1991) to search the ideal edges that best fit the partial edges. However, this method works well for simple parametric shapes (Maintz, 2005). Researchers also utilized neighborhood search (Ghita and Whelan, 2002) to find a candidate to link the edge pixels. Although many algorithms have been proposed to identify edges and connect them to generate objects, still identifying perfect edges to create image objects is a challenge.

Another edge-based algorithm based on mathematical morphology is the Watershed Transformation (WT) (Vincent and Soille, 1991). Mathematical morphology is efficient than traditional edge detection algorithms (Kaur and Garg, 2011). Thus, WT becomes most popular edge-based segmentation methods in remote sensing community. Watershed (Meyer and Beucher, 1990; Svoboda et al., 2007) simulates real-life flooding approach (Munoz et al., 2003) and transforms the image into a gradient and identifies objects with a topographical surface (Mezaris et al., 2004). As WT identifies segment boundary first and then creates the object, we consider it as an edge based method which is supported by earlier research (Carleer et al., 2005; Dey et al., 2010; Yang et al., 2017) as well. However, De Smet and Pries (2000) and Meineel et al. (2004) indicated that the WT has the properties of both edge detection and region growing techniques. Performance of watershed segmentation largely depends on the algorithm executed to compute the gradient. Typical gradient operator produces an over-segmented result in watershed segmentation due to noise or texture pattern (Zuva et al., 2011). Although multiple algorithms (Chen et al., 2008, 2006; Sun and He, 2008; Tarabalka et al., 2008; Wang, 1997; Weickert, 2001) proposed to generate the gradient image in order to increase the performance of the WT, still watershed generate an over-segmented image. Nevertheless, among the edge-based techniques, the watershed is widely used for natural image segmentation. Table 1 demonstrates some algorithms used for watershed segmentation. Only few studies found in literature where a solely edge-based method was used for object-based segmentation in remotely sensed images.

### 3.2. Region-based segmentation

The edge-based methods try to explore the object boundaries first and then determine the object itself by filling them in (Guindon, 1997). The region-based methods (Davis et al., 1975; Ohta et al., 1980; Pal and Pal, 1987; Pong et al., 1984) take the opposite approach. Those start from the inside of an object and then expand outward until meeting the object boundaries (Zhang, 2006). Theoretically, edge-based and region-based are different representations of the same object. However, the
### Table 1
Summary of watershed-based segmentation techniques.

<table>
<thead>
<tr>
<th>Sub-Class</th>
<th>Procedure</th>
<th>Performance and advantages</th>
<th>Limitations</th>
<th>Test data (sensor, spatial resolution, bands) and application</th>
<th>Previous literatures</th>
</tr>
</thead>
</table>
| Edge-constrained           | Edge-constrained watershed segmentation and edge allocation were used to generate sub-object primitives | • Could able to reduce false edges caused by noise and formed boundary-closed segments.  
• Provided high segmentation accuracy and boundary precision.                                                                                                      | • Not suitable for strongly textured images.                                                                                                                                                               | • Aerial, SPOT and ALOS  
• 0.3 m, 10 m and 2.5 m respectively  
• RGB, RGB + NIR and panchromatic (PAN) respectively                                                                                                                                  | Wang and Li (2014)  
Others: Wang et al. (2015b)                                                                                                                                                    |
| Marker                     | An Edge Embedded Marker-based Watershed (EEMW) algorithm was utilized for marker extraction and pixel labeling | • Successfully segmented small objects and provided accurate boundary.  
• Integrated edge information into segmentation.                                                                                                                                                                      | • Parameters were chosen manually.                                                                                                                                                                           | • Target not specified  
• QB  
• No details for resolution, and bands  
• Target road, green belt, playground, farm, lake, and house                                                                                                                        | Li et al. (2010a)  
Others: Gaetano et al. (2012), Jiao et al. (2010), Turker and Sumer (2008), Li et al. (2012), Gaetano et al. (2015)  
Wang and Wang (2016)                                                                                                                                       |
| Edge-embedded              | The result of the Canny edge detection was embedded into watershed segmentation | • Provided initial segments.  
• The region boundary achieved by this method was highly consistent with actual boundaries.  
• Improved region shape analysis and spatial relationship reasoning.                                                                                                          | • Broken and burrs roads were in the results.  
• The further processing required to improve accuracy.                                                                                                                                             | • ALOS, GF-1  
• 10 m and 8 m respectively  
• RGB + NIR  
• Target road                                                                                           | Mykonas et al. (2015)                                                                                                                                             |
| Grey Level Co-occurrence Matrix (GLCM) | In addition to the gradient, GLCM was also added as a separate band in watershed segmentation | • Provided different accuracy for road and farmland.  
• Over-segmentation reduced for farmland.  
• Utilized texture of the image object.                                                                                                                                                | • A single scale used for segmentation.  
• Over-segmentation still exists.                                                                                                                                                                       | • QB  
• 2.44 m multispectral (MLS)  
• RGB  
• Target road and agricultural land                                                                                                                                     | Wang et al. (2005)                                                                                                                                       |
| Classification-based       | Watershed segmentation utilized an inverted probability map as an input for flooding | • Seed selection was the key.  
• Can generate object efficiently.  
• Implemented an adaptive object recognition framework.                                                                                                                                                                  | • Applicable only on binary classification.                                                                                                                                                            | • Aerial image  
• No details provided for resolution and bands  
• Target frozen oil, sand ore, mixed vegetation                                                                                              | Levner and Zhang (2007)  
Others: Derivaux et al., (2010)                                                                                                                                             |
| Hierarchical               | Multilevel hierarchical segmentation was created using the gradient generated from the multichannel morphological technique | • Performed better than single-level watershed segmentation.  
• Reduced miss-segmentation errors.                                                                                                                                                                                   | • Applied only to map impervious surfaces.                                                                                                                                                              | • QB and IK  
• 2.44 m (MLS) and 0.61 m (PAN) for QB, 4 m (MLS) and 1 m (PAN) for IK  
• RGB + NIR + PAN  
• Target impervious surface                                                                                                                                         | Li et al. (2011)  
Others: Najman and Schmitt (1996)                                                                                                                                             |
region-based approaches may generate radically different results than edge-based approaches (Kavzoglu and Tonbul, 2017; Mainitz, 2005). Region-based methods assume that neighboring pixels within the same region have similar values (Tremeau and Borel, 1997). Region-based segmentation methods have two basic operations: merging and splitting (Fan et al., 2001). The basic approach to region-based image segmentation is (Bins et al., 1996): (a) obtain an initial (over or under) segmentation of the image, (b) merge or split those adjacent segments that are similar (or dissimilar) and (c) go to the previous step until no segments that should be merged or split remain.

3.2.1. Region growing/merging

Region growing (Adams and Bischof, 1994) is the most popular and simple algorithms for region-based segmentation. There are two main issues in region merging/growing segmentation, selection of seed region and similarity (Lucchese and Mitray, 2001). Seeded region growing (Deng and Manjunath, 2001) has two internal pixel order dependencies that create different resulting segments (Mehnert and Jackway, 1997). If multiple pixels have the same difference measure to their neighboring regions, then first order dependency happen (Shih and Cheng, 2005). The second-order dependency arises when one pixel has the same variation measure in several regions. Seed selection increase computational cost and execution time (Freixenet et al., 2002). To overcome the seeding problem, a single-seeded region growing technique was proposed by Verma et al. (2011), Mirghasemi et al. (2013) utilized Particle Swarm Optimization (PSO) to solve the localization problem. Wang et al. (2010) implemented the K-means clustering algorithm to generate seed in Region-based Image Segmentation Algorithm (RISA). Zhang et al. (2014a) proposed a hybrid region merging (HRM) method to segment high-resolution remote sensing images. They combined global-oriented and local-oriented region merging strategies into an integrated framework. By contrast, Byun et al. (2011) presented an approach based on modified seeded region growing and region merging utilizing a block-based seed selection method. Researchers are still in search of the better approach that can serve without seed (Wu et al., 2015) or can be unbiased by neighbors even though seeded (Fan et al., 2005).

After the selection of seeds, the region then grows by adding adjacent pixels that are similar, according to a specific homogeneity criterion, increasing the size of the region gradually. Therefore, the homogeneity criterion is the critical function of determining whether a pixel belongs to the growing region or not (Nock and Nielsen, 2004). The decision to merge is based only on the contrast between the current pixel and the region. Falah et al. (1994) and Xiaohan et al. (1992) implemented a homogeneity criterion containing the value of the modulus of the gradient of the pixel and the weighted sum of the contrast between the region and the pixel. Only pixels having low gradient values (below a certain threshold) were aggregated to the growing region in each iteration. In those cases, the spatial resolution of the image played a significant role.

Any object that is smaller than the spatial resolution of the image cannot be detected in segmentation process. However, if objects are bigger than spatial resolution, then it will be fragmented onto pixels. In contrast to the region growing techniques, region-merging techniques start from an initial region. Multi-Resolution Segmentation (MRS) arose from this idea (Hay et al., 2003). The Fractal Net Evolution Approach (FNEA) is a multisolution approach developed by Baatz and Schäpe (2000) and implemented in many research (Gao et al., 2017; Johnson, 2013; Kalantar et al., 2017; Li et al., 2016; Mallinis et al., 2008; Mathieu et al., 2007; Srivastava et al., 2015; Wang and Aldred, 2011; Yang et al., 2015c). The FNEA is a region merging hierarchical segmentation and starts with 1-pixel (Blaschke et al., 2004). Each coarse level acquires input from a finer level. If an object is recognized at the finer level, then it repeats its representation at each coarse level. It utilizes pairwise merging to form bigger objects. This procedure is also known as pairwise data clustering (Blaschke et al., 2008). Instead of global criteria, it uses local criteria and can integrate spectral, shape, texture, size, spatial, prior knowledge and contextual properties of the image objects.

Apart from the MRS, researchers also implemented other region merging approaches such as Mean-Shift (MS) (Comaniciu and Meer, 2002), Hierarchical Stepwise Optimization (HSO) (Beaulieu and Goldberg, 1989), and Recursive Hierarchical Segmentation (RHSeg) (Tilton et al., 2012). The MS is a clustering algorithm with non-parametric density estimation. It segments the image by grouping all pixels that are closer in the spatial and spectral domain and then connecting the corresponding convergence points. Based on this aspect, this review considered MS as a region-based approach. HSWO is also a clustering method that starts with individual data point and sequentially reduces number of clusters by merging. It utilizes a heap data structure to determine best merge. Hierarchical Segmentation (HSeg) (Tilton et al., 2010) is the improvement of HSWO in the merging process. It incorporates spectrally similar but spatially disjoint regions in the merging step. This process requires excessive computation time. To overcome this issue, RHSeg divides the image into subsections and apply HSeg on each subsection. Finally, it recombines the segmentation results from the subsections. This aggregation method may lead to processing window artifacts. Therefore, RHSeg contains an additional step to eliminate the artifacts.

3.2.2. Region splitting and merging

As global measures (used in region merging) caused imbalanced performance (Yang et al., 2017), researchers implemented split and merge (Cheesasuvit et al., 1986; Chen and Pavlidis, 1979; Chen et al., 1991; Horowitz and Pavlidis, 1976; Suk and Chung, 1983) as local measures for segmentation process in order to get a better result (Liu and Sclaroff, 2001). The splitting process starts from the entire image then based on the criterion for inhomogeneity (grey values, texture, internal edges, or various other criteria) split the image into segments (Blaschke et al., 2004). The split and merge method combines a bottom-up approach with a top-down approach (Guindon, 1997). Bottom-up approach generates an object by combining/merging pixels while the top-down approach shifts from splitting the entire image into image objects based on heterogeneity criteria (Benz et al., 2004). In this case, the initial seed is merely the entire image (Ohlander et al., 1978). If the seed is not homogeneous, the splitting method divides the seed into four squared subregions. Those subregions act as a seed in the next level and continue until all subregions become homogeneous (Martin et al., 2004). Kelkar and Gupta (2008) introduced improved quadtree method for the split-and-merge segmentation. Manousakas et al. (1998) implemented principles of simulated annealing and boundary elimination to improve the quality of traditional split and merge algorithms. However, the fusion of two segments is upfront, the splitting of a segment requires proper sub-segments. The primary disadvantage of region splitting is that the resulting image tends to mimic the data structure used to represent the image and comes out too square (Cheng et al., 2001). The region merging approach is often combined with region splitting to merge the similar regions for creating a homogeneous region as large as possible. Alshehhi and Marpu (2017) executed hierarchical merging and splitting image segmentation based on color, and shape features for road extraction from urban area images. Table 2 summarized different region-based segmentation techniques. As indicated in the table, most algorithms are facing difficulty in defining appropriate parameters. In addition, though some of them generated promising results, they have not been evaluated by applying to different image setting. Another issue is that in many cases segmentation results were compared with segments generated by eCognition (using FNEA) even though its results depend on user-defined parameters.
Table 2
Summary of different region-based segmentation techniques.

<table>
<thead>
<tr>
<th>Algorithms Sub-Class</th>
<th>Procedure</th>
<th>Performance and advantages</th>
<th>Limitations</th>
<th>Test data (sensor, spatial resolution, bands) and application</th>
<th>Previous literatures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region growing/ Seeded</td>
<td>The seed was identified based on geometry using Stroke Width Transformation (SWT). A Gaussian Mixture Model (GMM) was implemented to differentiate background and non-background. Finally, a Convex Active Contour (CAC) model was operated to merge road seeds and identify whole road segments</td>
<td>• Work well in identifying roads irrespective of shape, width, direction or intensity variation. • Lower running time compared to other algorithms implemented for the same purpose.</td>
<td>• Further investigation is required to identify the applicability of the proposed method in images where there are land covers with similar geometric and radiometric characteristics, such as long narrow turbid water canal.</td>
<td>• UAV</td>
<td>Zhou et al. (2016)  Others: Ispahiani et al. (2010), Zirkovic (2004), Deriche et al. (2017) Others for segmenting roads: Mohammadzadeh and Zoj (2010), Sun and Messinger (2013)</td>
</tr>
<tr>
<td>Mixture Model (GMM)</td>
<td>The seed was generated from the gradient image. Spectral-morphological characteristics of a pixel were considered as a criterion of homogeneity</td>
<td>• Performed fine in segmenting building and other fabricated structures in urban and suburban areas. • Utilized both spectral and spatial information of images. • Automated seed selection and merging process.</td>
<td></td>
<td></td>
<td>Epshtein et al., (2010), Zivkovic (2004), Deriche et al. (2017)</td>
</tr>
<tr>
<td>Mean Shift</td>
<td>Horizontal and vertical semivariogram was used to identify spatial bandwidth (window size). Acquired bandwidth was applied on mean-shift-based multiscale segmentation</td>
<td>• In addition to merging criteria (MC), it provided importance to merging order (MO) as well. • Merging priority was estimated based on the inter-segment heterogeneity and intra-segment homogeneity.</td>
<td>• Manual selection of parameters is required.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region Adjacency Graph (RAG)</td>
<td>The local best region-growing strategy was used to create initial segments. RAG was created based on initial segments. Edge strength was used as a merging criterion. In order to produce multi-scale segments, a local best region merging process was applied</td>
<td>• Work well in identifying roads irrespective of shape, width, direction or intensity variation. • Lower running time compared to other algorithms implemented for the same purpose.</td>
<td></td>
<td></td>
<td>Pesaresi and Benediktsson (2001), Bellens et al. (2008), Barrile and Bilotta (2016), Ling et al. (2012), Yuan et al. (2014), Wang et al. (2015c)</td>
</tr>
<tr>
<td>Mean Shift Horizontal and vertical semivariogram</td>
<td>Overlapping tiles were generated to apply stabilized mean-shift filtering algorithm. Each tile was processed independently using the connected-component algorithm. Unique labeling conducted by shifting values from one tile to another</td>
<td>• The proposed method could identify the appropriate scale. • Utilized global and local structure of the objects in multiple scales. • Combined geostatistics and pattern recognition.</td>
<td>• Applied only on PAN images.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Shift Spatial filtering</td>
<td>Stabilized mean-shift horizontal and vertical semivariogram was applied to stabilize the mean shift algorithm.</td>
<td>• Intrinsic performance evaluation was not compared among different algorithms. • Arbitrary parameters were utilized during the segmentation.</td>
<td></td>
<td></td>
<td>Michel et al. (2015) Other tile-wise: Tilton (2010), Michel et al. (2012), Banerjee et al. (2012), Körting et al. (2013), Tzotsos and Argialas (2006), Xing et al. (2014) Others using mean shift: Wang et al. (2015a), Wang et al. (2012)</td>
</tr>
<tr>
<td>Mean Shift Spatial filtering</td>
<td>The local best region-growing strategy was used to create initial segments. RAG was created based on initial segments. Edge strength was used as a merging criterion. In order to produce multi-scale segments, a local best region merging process was applied</td>
<td>• Initial over segments ensure inclusion of all object boundary. • RAG indicated the relationship between the neighboring segments. • Edge strength ensures merging the right segments.</td>
<td>• Though the value range and physical meaning are different, the sum of standard deviation and compactness were applied as the merging criterion. • The scale was predefined.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Sub-Class</th>
<th>Procedure</th>
<th>Performance and advantages</th>
<th>Limitations</th>
<th>Test data (sensor, spatial resolution, bands) and application</th>
<th>Previous literatures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical sorting</td>
<td></td>
<td>Initial segments were generated using statistical region merging and minimum heterogeneity rule was utilized for object merging</td>
<td>• Utilized spectral, spatial, scale, and shape of image objects.</td>
<td>• Need multiple user inputs.</td>
<td>• QB</td>
<td>Li et al. (2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Eliminated small redundant objects.</td>
<td>• Success depends on sort function and merging predicate.</td>
<td>• 2.44 m (MLS) and 0.61 m (PAN)</td>
<td>Others: Huang et al. (2014), Nielsen and Nock (2003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Generated results in vector and raster format.</td>
<td></td>
<td>RGB + NIR + PAN</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Target road, highway, grass, and buildings.</td>
<td></td>
<td>Target road, playground, water, and artificial targets</td>
<td></td>
</tr>
<tr>
<td>Multi features</td>
<td></td>
<td>The scale was selected based on local area homogeneity index J-value. Multi-scale boundaries constraint strategy was used for multi-scale segmentation. Merging was done based on multi-features</td>
<td>• The sequence of multi-scale J-images can adequately reflect homogeneity of spectral distribution of local region.</td>
<td>• Tested only on a subset of the image.</td>
<td>• QB</td>
<td>Wang et al. (2016)</td>
</tr>
<tr>
<td>Structural constraints</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.4 m spatial resolution</td>
<td>Others: Cano-García and Alonso-Sarría (2015)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Structural constraints such as parallel straight-line neighborhood, perpendicular straight-line neighborhood, and parallel straight-line zone were incorporated in the merging process</td>
<td>• Utilized straight-line boundaries featured for roads and buildings.</td>
<td></td>
<td>0.3 m for Aerial and 2.5 m for ALOS</td>
<td></td>
</tr>
<tr>
<td>Scale-space filtering</td>
<td></td>
<td>Anisotropic morphological leveling was used for removing noise and complexity available in high-resolution images. MSEG (Tzotsos and Argialas, 2006) was employed for generating primitive image objects</td>
<td>• The segmentation algorithm was applied to simplified images.</td>
<td>• Applicable to man-made structures only.</td>
<td>RGBA for aerial and PAN for ALOS</td>
<td>Wang et al. (2017a,b,c)</td>
</tr>
<tr>
<td>Split and merge</td>
<td></td>
<td>Initial segments were generated based on spectral heterogeneity. Merging was done based on spectral variance difference and edge penalty</td>
<td>• Provided better results in over-, under- and well-segmentation rate.</td>
<td>• Accuracy was assessed for classification results only.</td>
<td>Target buildings and roads</td>
<td>Others: Wang et al. (2017b,c), Wang et al. (2017a,b), Wang et al. (2015a,b,c), Wang and Li (2014) Wang and Wang (2016)</td>
</tr>
<tr>
<td>Spectral variance difference</td>
<td></td>
<td></td>
<td>• Constrained spectral variance difference can limit the influence of the segment size.</td>
<td></td>
<td>Aerial, SAR and Landsat</td>
<td></td>
</tr>
<tr>
<td>Hierarchical multiple Markov chain</td>
<td></td>
<td>In the splitting phase, segments were created based on spatial and spectral properties of objects. Using these segments as a unit, merging was done based on texture</td>
<td>• Identified complex textured areas efficiently.</td>
<td>• Five parameters must be set manually to implement the algorithm.</td>
<td>WV-2, aerial and RapidEye image</td>
<td>Others: Kavzoglu et al. (2016)</td>
</tr>
<tr>
<td>Hierarchical Split Merge</td>
<td></td>
<td>Band ratio was employed as a region description. Entropy was used as a texture measurement. Fuzzy logic based similarity measure was used for merging</td>
<td>• Provided a sequence of nested segmentation maps.</td>
<td>• Scale needs to be set small initially which cause over-segmentation.</td>
<td>0.6 m (pan sharpen), 1 m and 5 m respectively</td>
<td></td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
<td></td>
<td>• Removed non-continuous regions.</td>
<td>• Required additional steps compared to MRS.</td>
<td>RGB + NIR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Agglomerative merging process reduced undesired segments.</td>
<td></td>
<td>Target farmland, road, buildings, river, and reservoir</td>
<td>edge penalty: Zhang et al. (2014a,b,c)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Others:  Deng et al. (2013), Lucieer et al. (2005), Hu et al. (2005), Miao et al. (2015), Chen et al. (2015), Ghebreno et al. (2009), Wuest and Zhang (2009), Ojala and Pietikäinen (1999), Chen and Chen (2002)
3.3. Hybrid method (HM)

To overcome the limitation of both edge- and region-based method, scholars integrated the results of edge- and region-based method and are expected to provide better segmentation results (Al-Hujazi and Sood, 1991; Fan et al., 2001; Moigne et al., 1995). As discussed earlier, edge-based methods are precise in detecting edges, however, facing problem in generating closed segments. By contrast, region-based methods create closed regions, however, resulting in imprecise segment boundaries (Wang and Li, 2014). As a result, a recent trend in image segmentation is to execute an HM (Gaetano et al., 2015; Li et al., 2014), in which the initial segments are first outlined using edge-based methods, then merged using region-based methods. Such HMs utilize both boundary pixels to outline the initial segments and the interior pixels to merge similar segments (Zhang et al., 2014b). Mueller et al. (2004) combined edge and region-based techniques to extract large man-made objects such as agricultural fields. In the first part, they extracted shape information. The edge map offers an additional criterion in decisions. In the second part, they used this information to control region growing algorithm. On the other hand, Gambotto (1993) suggested using edge information to stop the growing process. Li et al. (2010c) recommended texture clustering which was executed as constraints in HSWO. Merging was conducted based on the region adjacency and neighbor graph.

Many region-merging methods employ a single global parameter to control the iterative process of merging segments as it gives the user control over under- and over-segmentation. Even so, the same threshold is used for all segments regardless of their homogeneity with other segments. Instead of using a global threshold, Johnson and Xie (2011) and Chen et al. (2014) have used local measures to identify segments that are under and over segmented at the selected optimal scale parameter and further refined them by appropriate splitting and merging. This local refinement strategy is efficient in improving segmentation quality because it eliminates under and over-segmentation problems (Yang et al., 2016). However, the further splitting and merging steps face a challenge when executing local refinement in an operational context (Yang et al., 2016). In addition, those methods considered heterogeneity between adjacent segments as the merging criteria. Both homogeneities within the segments and heterogeneities between the segments should be considered equally. Wang et al. (2018c) proposed an HM considering the objective of heterogeneity and relative homogeneity during the merging process. Table 3 compiled different hybrid segmentation methods. Most of the studies started from the edge-based method which create an over-segmented image. Then the region-based method was conducted to merge similar segments based on either homogeneity or heterogeneity. In the merging process, there are two main issues, merging criteria and merging order. Variance, area weighted variance, Moran’s I, spectral angle, F measure, spectral and geometric properties were used as the merging criteria. To identify the adjacent relationship, many studies utilized RAG, and nearest neighbor graph. As illustrated in the table, when combining both the edge- and region-based methods, issues of the individual algorithm such as over-segmentation, seed selection, generate initial region for merging, under-segmentation are compensated by other(s). Though HMs provide some promising results, their implementations are troublesome.

3.4. Semantic methods

Machine learning (ML) has proven successful for many applications in recent years that affect remote sensing arena as well. ML algorithms are “approximators” which learn from the training data and act accordingly. Unlike unsupervised methods (such as region growing), ML-based semantic segmentation algorithms such as Markov Random Field (MRF) (Farag et al., 2005; Krishnamachari and Chellapapp, 1997; Poggi et al., 2005; Tran et al., 2005; Tupin and Roux, 2005), Bayesian Network (Bouman and Shapiro, 1994; D’Elia et al., 2003; Zhang and Ji, 2010), Neural Network (NN) (Kurnaz et al., 2005; Villmann et al., 2003), active Support Vector Machine (aSVM) (Mitra et al., 2004), weighted aggregation (Du et al., 2016), and Deep Convolution Neural Network (DCNN) (Audebert et al., 2016) are supervised approaches. In semantic segmentation method, each pixel is allotted a class label of its enfolding object.

Among the semantic algorithms, MRF accounts for a large percentage (Geman and Graffigne, 1986; Melas and Wilson, 2002; Zhuowen and Zhu, 2002). MRF is a probabilistic method that seizes the contextual limits within the neighboring pixels. Feng et al. (2010) used split-and-merge techniques to segregate the main problem to a series of sub-problems. Tree-structured graph cut, hierarchical graph cut and net-structured graph cut was used to obtain labeling accuracy and spatial coherence. The proposed method was computationally efficient and well performed in terms of robustness to noise and soft boundary preservation. In order to obtain proper segments from noisy images with the complex and macro-texture pattern, Zheng et al. (2013) integrated the MRF model with Multi-Region Resolution (MRR) segmentation. They applied the proposed method to QB, SPOT-5, and Synthetic Aperture Radar (SAR) images. The proposed method segmented images into three broad land cover types such as only farmland, woodland, and urban area. Researches also utilized edge penalty function (Yu and Clausi, 2008), discrete wavelet transform (Jung et al., 2005), multiscale approach (Moser and Serpico, 2008) and region-based strategies (Moser et al., 2013) with MRF to segment high-resolution images. Recently among the semantic methods, DCNN has been used in researches (Kemker et al., 2018; Nigam et al., 2018; Zhang et al., 2016) as it has the capability to treat data as a nested model. In those methods, raw images are used as an input and pixels generate object when it passes through multiple layers.

Though semantic algorithms showed encouraging results, they are facing many challenges. For instance, in high-resolution images, it is difficult to define suitable features with semantic meaning due to high texture (Zhang et al., 2014b). In addition, the scale and hierarchy available on those images make it difficult in determining semantic rules (Burnett and Blaschke, 2003) which can differentiate objects in different scales. Apart from that, different image objects may have similar spectral value (such as water and shadow) which create ‘semantic gap’ (Wang et al., 2013). Furthermore, semantic algorithms also suffer computational burden to extract structural information (Yang et al., 2008), require a vast amount of training data and a significant number of parameters for tuning (Chen et al., 2018a).

3.5. Available software/tools

Though hundreds of algorithms have been proposed for segmentation, only a few of them have been implemented and are available as a tool/software. Among them, eCognition is the popular and widely used segmentation software. According to Blaschke (2010), 50–55% OBIA articles employed eCognition. Its success prompted other commercial software developers such as Hexagon Geospatial, Harris Geospatial Solutions, ESRI, and PCI Geomatics to develop their tools. Apart from those, another set of tools such as EDISON, SCRIM, and GeoSegment developed in an academic environment, and others are open source tools (Table 4) such as SAGA, GRASS GIS developed by other developers.

4. Discussions

4.1. Pros and Cons of different algorithms

Advantages of edge-based segmentation are that algorithms are less complicated compared to region-based segmentation (Felzenzwalb and Huttenlocher, 2004), works fine in images with the decent contrast between object and background (Kaganami and Beiji, 2009; Tsai et al., 2003), computationally efficient (Lin et al., 2003) and can correspond...
well with the object edges (Chen et al., 2018b). Problems in edge-based segmentation are that algorithms do not function well on images with smooth transitions and low contrast, sensitive to noise (Iannizzotto and Vita, 2000). Due to poor performance in the detection of textured objects (Yu et al., 2006), edge-based segmentation has not been applied widely in high-resolution images. In addition, if it misses part of the boundary, then disjointed edges permit merging of dissimilar regions (Kermad and Chehdi, 2002). Furthermore, the multi-spectral image makes the edge detection process more complicated (Li et al., 2010b) due to the inconsistent location of edges in the multiple bands. Finally, edge-based methods rely on local data and thus misses essential contextual information at larger scales (Gaetano et al., 2015) which is the key in object-based image analysis. However, the edge-based method can be used to support region-based techniques (Sappa, 2006).

The region-based methods generate spatially and/or spectrally homogeneous segments based on the defined properties. In addition, region-based methods can produce segments at multi-scales. For instance, the shape of a segment at one scale level can be used as a variable at another level (Wang et al., 2010). Also, those methods allow users to choose multiple criteria at the same time. Furthermore, users have the freedom to select the seed point and merging criteria. Finally, those methods are less sensitive to noise when compared with edge-
4.2.2. Segmentation from low-level pixel grouping

Due to the size and nature (artificial entity) of pixels, in many cases they do match with the size of the object. It would be efficient if segmentation start from the perceptual meaningful entities (Csillik, 2017). The idea of superpixels (which is low-level grouping of pixels) came from this concept. In contrast to pixels, superpixels carry more information and follow the natural image boundary (Zhengqin and Jiansheng, 2015). In addition, superpixels have an intermediate scale between pixels and objects (Achanta et al., 2012), reduce noise and outliers (Shi and Wang, 2014), and can speed up the subsequent process. Achanta et al. (2012) categorized superpixels algorithms into graph and gradient ascent methods. Multiple algorithms have been proposed to generate superpixels such as superpixels lattice (Moore et al., 2008), turbopixels (Levinshtein et al., 2009), quick shift (Vedaldi and Soatto, 2008) and simple linear iterative clustering (SLIC) (Achanta et al., 2012). Stefanski et al. (2013) tested the performance of superpixels contour in remote sensing image segmentation to optimize segmentation parameters and improve classification accuracy. They reported that it is easy to handle as only two parameters to deal with and can optimize parameters selection process. Csillik (2017) proposed a segmentation workflow where famous MRS algorithm started from superpixels instead of individual pixel. They stated that the proposed workflow significantly reduced processing time and provided better accuracy.

Instead of using a single scale for the entire image, Fonseca-Luengo et al. (2014) offered a hierarchical multiscale segmentation using superpixels (SLIC) which allowed users to detect objects at different scales. It can provide a realistic local optical scale (Gonzalo-Martín et al., 2016) and a better understanding of land cover and its objects. Yin and Yang (2017) compared superpixels with sub-pixels to map urban green space and concluded that superpixels provided higher accuracy than sub-pixels. Apart from above researches, recent trends are to incorporate superpixels with other models such as probabilistic fusion model (Zhang et al., 2015d), probability density function (Liu et al., 2015).

### Table 4

<table>
<thead>
<tr>
<th>Tool/Software</th>
<th>Reference/Developer</th>
<th>Website</th>
<th>Algorithm</th>
<th>Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>InterSeg</td>
<td>Happ et al. (2016)</td>
<td><a href="http://www.lvc.ele.puc-rio.br/wp/?cat=41">http://www.lvc.ele.puc-rio.br/wp/?cat=41</a></td>
<td>Region-based</td>
<td>Available upon request</td>
</tr>
<tr>
<td>BerkeleyImgSeg</td>
<td>Clinton et al. (2010)</td>
<td><a href="http://www.imgseg.com/">http://www.imgseg.com/</a></td>
<td>Region-based</td>
<td>Commercial</td>
</tr>
<tr>
<td>IMAGINE Spatial Modeler</td>
<td>Hexagon Geospatial</td>
<td><a href="http://community.hexagongeospatial.com/IMAGINE-Spatial-Modeler/tkb-p/eTSpatialModeler">http://community.hexagongeospatial.com/IMAGINE-Spatial-Modeler/tkb-p/eTSpatialModeler</a></td>
<td>Edge-based</td>
<td>Commercial</td>
</tr>
<tr>
<td>ENVI Feature Extraction</td>
<td>Harris Geospatial Solutions</td>
<td><a href="http://www.harrisgeospatial.com/docs/routines-164.html">http://www.harrisgeospatial.com/docs/routines-164.html</a></td>
<td>Edge-based</td>
<td>Commercial</td>
</tr>
<tr>
<td>IDRISI GIS Tool</td>
<td>Clark Labs</td>
<td><a href="http://clarklabs.org/terrain/idrisi-gis/">http://clarklabs.org/terrain/idrisi-gis/</a></td>
<td>Region-and edge-based</td>
<td>Freeware</td>
</tr>
<tr>
<td>Object Analyst</td>
<td>PCI Geomatics</td>
<td><a href="http://www.pcigeomatics.com/geomatica-help/concepts/focus_c/oa_intro.html">http://www.pcigeomatics.com/geomatica-help/concepts/focus_c/oa_intro.html</a></td>
<td>Region-based</td>
<td>Commercial</td>
</tr>
<tr>
<td>SCRM</td>
<td>Castilla et al. (2008)</td>
<td><a href="http://www.castlink.ca/scrm/scrm">http://www.castlink.ca/scrm/scrm</a></td>
<td>Region-and edge-based</td>
<td>Freeware</td>
</tr>
<tr>
<td>GeoSegment</td>
<td>Chen (2018)</td>
<td><a href="http://130.15.95.215/lagisa/">http://130.15.95.215/lagisa/</a></td>
<td>Region-based</td>
<td>Online tool, available upon registration</td>
</tr>
<tr>
<td>Optimization Algorithm/Approach</td>
<td>Segmentation Algorithm</td>
<td>Cost Function/Criteria</td>
<td>Performance and Advantages</td>
<td>Limitations</td>
</tr>
<tr>
<td>---------------------------------------------------------------------</td>
<td>------------------------</td>
<td>-------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Genetic algorithm (GA) for optimizing segmentation parameters</td>
<td>MRS</td>
<td>Area, number of segments and weight coefficients.</td>
<td>Classification accuracy improved around 5% if compared to support vector machine.</td>
<td>Get rid of manually selecting parameters.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The proposed method can be tested by segmenting other man-made features.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The hybrid method requires longer processing time compared to differential evolution.</td>
<td></td>
</tr>
<tr>
<td>Hybrid metaheuristics approach for parameter tuning</td>
<td>MRS</td>
<td>Precision and recall</td>
<td>Compared seven strategies (single, population and hybrid metaheuristics) regarding speed and proper solution.</td>
<td></td>
</tr>
<tr>
<td>An adaptive approach for scale selection</td>
<td>MRS</td>
<td>Thematic maps</td>
<td>Utilized the inherent features of segmented objects and prior knowledge to improve segmentation performance.</td>
<td>Many parameters were selected based on trial and error basis.</td>
</tr>
<tr>
<td>Spectral and spatial statistics for scale selection</td>
<td>Mean Shift</td>
<td>Local variance (LV)</td>
<td>Pre-estimated the optimal scale before segmentation from the spatial statistics.</td>
<td>Guaranteed high homogeneity and high heterogeneity within and between segments.</td>
</tr>
<tr>
<td>Spectral measures for scale selection</td>
<td>MRS</td>
<td>SA</td>
<td>Utilized all spectral bands.</td>
<td>Did not consider intra-segment homogeneity and inter-segment heterogeneity.</td>
</tr>
<tr>
<td>Regression tree model for generalizable scale parameters</td>
<td>MRS</td>
<td>Meta-analysis</td>
<td>Narrow down the range of suitable scale parameters.</td>
<td>Could not identify exact scale for different land use.</td>
</tr>
<tr>
<td>Classification driven approach for scale selection</td>
<td>MRS</td>
<td>GLCM</td>
<td>Identified best segmentation scale for sub-decimeter resolution UAV images.</td>
<td>Target was only vegetation and bare land.</td>
</tr>
<tr>
<td>Spatial autocorrelation for scale selection</td>
<td>MRS</td>
<td>Rate of Change (ROC) of Moran’s I (MI)</td>
<td>The global score method excluded under-segmented scale.</td>
<td>MI considered spatial distribution of segments.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Identifying optimal scale from ROC-MI curve still challenging.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
et al., 2017), adaptive region merging (Ko and Ding, 2016), purpose dependent grouping (Maboudi et al., 2016a), multiscale and multi-feature normalized cut (Zhong et al., 2016), minimum spanning tree (Wang et al., 2017c) and binary merge tree (Su and Zhang, 2018) to identify optimal scale and parameters as well as to minimize under- and over-segmentation problem.

4.2.3. Multiscale segmentation

In high-resolution images, an individual object is modeled by many pixels. Pixels within an individual object tend to display high spectral autocorrelation. Even so, image objects exhibits an intrinsic scale, hierarchical structure and are composed of structurally associated parts. As a result, Modifiable Areal Unit Problem (MAUP) (Marceau, 1999) is frequent in remote sensing images. Image objects can be treated differently in different scales. Therefore, multiscale segmentation is an important issue for GEOBIA as a single scale is not suitable to represent different image objects. However, there is no single optimal scale (Hay et al., 2003), thus scholars are trying to identify scales that are specific to the dominant image objects within a scene.

Among the segmentation algorithms, MRS has been widely used in the literature. The primary challenge in MRS is selecting appropriate parameters as geographical objects varied in size, shape, and texture (Ma et al., 2015; Teodor and Araujo, 2016). Among the parameters, scale plays a vital role in MRS. Selection of object-based scale in segmentation is the key to GEOBIA because a wrong scale will lead to either over- or under-segmentation (Ming et al., 2012). In order to determine optimal scale, a trial-and-error method is commonly executed in remote sensing (Eisank et al., 2014; Ninsawat and Hossain, 2016; Radoux and Defourny, 2007; Zhang et al., 2014c). However, the trial-and-error method is time-consuming and impractical for many applications (Im et al., 2014; Ma et al., 2017). Multi-scale segmentation algorithms utilize user-defined scale in different ways. For instance, in the FNEA scale determines the average size of the object whereas in WT it defines the sampling window size, valley and catchment area threshold (Ming et al., 2015). Finding optimal scale is troublesome due to several issues such as the implicit relationship between scale and image data, the intricate link between segmentation results on different scales (Ming et al., 2012).

4.2.4. Optimization of segmentation parameters

Apart from scale, texture also can increase segmentation accuracy (Kim et al., 2011). Parameters optimization is a topic of research for decades, and the recent trend is to employ automatic, optimal parameter determination procedure (Chen et al., 2018a). Optimal parameters will enhance intra-segment homogeneity, inter-segment heterogeneity (Yang et al., 2015b), and classification accuracy (Gao et al., 2011). A collaborative approach (integration of thematic maps generated from the classification method to MRS) was implemented by Troya-Galvis et al. (2016) to develop a generic segmentation procedure. Saba et al. (2016) introduced an automatic image segmentation method by using genetic algorithm optimization with a new cost function. Furthermore, Esch et al. (2008) utilized fuzzy logic and iterative optimization respectively to identify optimal parameters. As indicated in Table 5, multiple approaches have been proposed in the literature to identify optimal parameters for MRS and mean shift. In addition to those, some scholars have integrated MRS with other models. For example, Li et al. (2008) proposed MRS by using Statistical Region Merging (SRM) and Minimum Heterogeneity Rule (MHR). SRM was utilized for initial segmentation and MHR for merging objects. Similarly, Gu et al. (2018) integrated graph-based segmentation with MRS where initial segments were generated by using graph theory, and merging was done by FNEA. Yang et al. (2015b) introduced a new energy function to quantify the relationship between image objects and their neighbors. Chen et al. (2012b) prescribed a soft image segmentation model based on multiresolution and probability of pixel merging at the top level. Nevertheless, defining appropriate segmentation parameters even for a single image is a significant challenge (Chen et al., 2018a).

4.2.5. Evaluation of segmentation results

Optimal segmentation parameters selections methods intended to select parameters by post evaluation. Parameters are selected based on supervised and unsupervised method. Supervised method select parameters based on the similarity between the corresponding trial-and-error results and the reference data (Ghosh and Joshi, 2014; Wang et al., 2018a, b). Similarity can be based on area overlap (Clinton et al., 2010; Yang et al., 2015a; Zhang et al., 2015a), correctly matched objects numbers (Liu et al., 2012; Marpu et al., 2010), object location (Montaghi et al., 2013), spectral discrepancy (Anders et al., 2011), border fitness (Albrecht, 2010; Neubert et al., 2008), or combination of these (Witharana and Civo, 2014; Zhang et al., 2015b). Unsupervised methods compare the resultant segments with the good segmentation (based on intra-segment homogeneity and inter-segment heterogeneity) (Drağut et al., 2014; Zhang et al., 2017b) and used in remote sensing as there is no true ground-truth segmentation of an image against which the output of an algorithm can be compared. Both methods are facing difficulty in either generating reference objects or defining criteria that can quantify intra-segment homogeneity and inter-segment heterogeneity between objects.

Shapes of objects extracted from the segmentation are used by classification algorithms to extract patterns for object labeling. They are also used to assist in quantifying spectral statistics of each object. Apart from the shape, the location of objects is also essential for geospatial analysis. Traditional pixel-based accuracy assessment methods are incapable of calculating object shape and location (Clinton et al., 2010). Several area- and shape-based goodness measures have been proposed in the literature (Clinton et al., 2010; Zhang, 1996) to judge segmentation results. However, those measures use predefined objects from an image as the training objects. This process makes the segmentation accuracy assessment method somehow subjective as the training objects depend on human judgment.

4.2.6. Image-objects vs geo-objects

Image-objects do not exist independently within digital images. Segmentation is the primary unit of GEOBIA and aims to identify image-objects based on discreteness, coherency, and contrast (Castilla and Hay, 2008). By contrast, a geographic object (geo-object) refers to an object having certain minimum size on or near the earth surface, with many permanent properties and differs from its surroundings based on specific properties. When segmentation algorithms can generate geo-objects, then segments are termed as meaningful image-objects. However, achieving meaningful image-objects is challenging due to complicated radiometric and semantic relationship between image- and geo-objects and hierarchical details of objects. Thus, human interpretation of meaningful image-objects varies from segmentation algorithms. This conceptual gap is termed as over- and under-segmentation in segmentation results.

4.3. Future directions

Segmentation is the key component of GEOBIA by reducing image complexity, making image content understandable and producing meaningful image objects (Lang, 2008). Pixels are the basic unit of a raster image and usually square shaped. As pixels are not natural entities, they do not match with the image content. By contrast, the hexagon is eligible to represent earth surface more efficiently (Sahr et al., 2003). Based on this concept, Hofmann and Tieke (2014) proposed hexagonal cell-based MRS approach. For the testing purpose, they utilized WV-2 images, and their target was segmenting a soccer stadium. This method provided better results when compared to the square cell-based MRS especially in segmenting linear and round shaped features. Another way of dealing square blocks is a low-level grouping of pixels which is more natural and efficient to work with
The idea of superpixels (Ren and Malik, 2003) also comes from the concept of low-level pixel grouping. The same evolution trend is expected in the future.

Geo-objects refer to the spatial entities formed by numerous elements distributed within a geographic area. In high-resolution remote sensing images, geo-objects are distributed in a number of pixels. Thus, an object that appears homogeneous in one scale may become heterogeneous at another scale. Although a significant amount of research devoted to identifying optimal scale for segmentation, still the question of “what is the optimum (or range at least) segmentation scale for different image objects within a scene?” roaming in remote sensing arena. The scale has a multi-dimensional nature (Malenovsky et al., 2007), complex hierarchy, and variability (Wu and Li, 2009). Studies have utilized geographical variance (Moellerling and Tobler, 1972), wavelet transform (Percival, 1995), local variance (Woodcock and Strahler, 1987) for measuring spatial structure. However, Nijland et al. (2009) identified that there is no spatial scale appropriate for identifying and analyzing various urban features. What’s more, different segmentation algorithms treat scale in a different way. We may expect the advancement in dealing with the segmentation scale for generating meaningful image objects would make rapid progress.

In addition to scale, segmentation results vary based on homogeneity or heterogeneity criteria. Many algorithms have utilized textual parameters to developed rules for homogeneity and heterogeneity. As a bit change of homogeneity or heterogeneity leads to different segmentation results, Hay and Castilla (2008) termed segmentation as an “ill-posed problem.” Thus, a substantial amount of research dedicated to segmentation parameters optimization (as shown in Table 5). Nevertheless, more advanced techniques focusing on developing a methodology of parameters optimization that is applicable in any context is deemed necessary. In addition, creating an object-based segmentation dataset similar to “The Berkeley Segmentation Dataset and Benchmark” (available at https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/; which is not specific to remote sensing images) has become essential to compare the performance of different segmentation algorithms and parameters in different image settings.

Although the recent trend is applying the hybrid segmentation methods, these algorithms are facing problems due to implementation complexity, and lack of available software and tools. More hybrid segmentation techniques along with the tools will evolve in future. Another issue in these cases is applying algorithms on large dataset due to the computational power of existing computer hardware. Several studies have implemented tile-wise segmentation and showed promising results. However, these methods have several questions to answer, dealing objects in the tile edges, removing processing window artifacts and dynamic selection of parameters based on the geo-objects in each tile are among the most prominent. Another interesting topic in segmentation is the evaluation of segmentation results. Object-based supervised and unsupervised methods are facing problem in creating reference objects and defining criteria, more research will focus on this topic.

GEOBIA method based on the idea of one-to-one mapping between segments and image objects. Thus, classification algorithms require perfect segments to provide accurate prediction. Based on the above discussions, it can be concluded that it is challenging to find an algorithm that can generate perfect segments. In addition, many objects are composed of non-homogeneous regions in high-resolution images such as the roof of a house (often composed of light and dark regions) which not likely to be segmented together. To resolve this issue, Troya-Galvis et al. (2018) proposed a method to modify initial segments based on the trained classifier. Their work is an extension of Collaborative segmentation-classification (CoSC) approach (Troya-Galvis et al., 2016). Although the proposed method is time intensive and requires sufficient training data, it showed promising results. Furthermore, a scale self-adaptive segmentation method based on exponential sampling scale and weighted LV was proposed by Wang et al. (2018a,b) for historic land use/land cover (LULC) change database updating. In this case, the segmentation process was guided by historical LULC boundaries. By contrast, Golinkoff (2013) used objects area to guide the segmentation process.

5. Summary

With the advancement of remote sensing technology, high spatial resolution remote sensing images have been used widely in different fields due to their rich spectral and textural information. GEOBIA has evolved to analyze those high-resolution images. As a critical component of GEOBIA process, image segmentation algorithm has been a hotspot recently. Though many algorithms have been proposed, all algorithms have some pros and cons. For instance, edge-based algorithms are easy to implement, but they are missing the contextual information. The region-based method generates better results compared to the edge-based method, however, finding appropriate seeds and other parameters is the real challenge in that case. To resolve the seeding problems, superpixels algorithms are introduced in remote sensing image segmentation. Another recent trend is to execute a hybrid method, although those algorithms are complicated and no software package available in the market to implement. Researches still trying to identify algorithms (with optimal parameters) for the segmentation process which can accurately identify individual image objects.

Segmentation influences classification accuracy. However, using optimal parameters for segmentation algorithms is not the only solution for achieving higher accuracy in OBIA. Different parameter combinations would lead to similar classification results. Recently some researches have collaborated segmentation with classification in high-resolution image analysis (Heumann, 2011; Wang and Aldred, 2011; Cuillik, 2017; Guo and Du 2017; Hadavand et al. 2017). In those cases, the segmentation does not have to be perfect. The classification process may include object generating steps which will assist in overcoming the over-segmentation problem and building the complex objects.

Declaration of interest
None.

Funding
This research is supported by Canada Natural Science and Engineering Research Council (NSERC) Discovery grant.

References

Performance Evaluation. Springer.


