ISPRS Journal of Photogrammetry and Remote Sensing 101 (2015) 186–192

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

A discrepancy measure for segmentation evaluation from the perspective of object recognition

Jian Yang [a,b,*], Yuhong He [b], John Caspersen [c], Trevor Jones [d]

[a] Department of Geography, University of Toronto, 100 St. George Street, Toronto, ON M5S 3G3, Canada
[b] Department of Geography, University of Toronto Mississauga, 3359 Mississauga Rd North, Mississauga, ON L5L 1C6, Canada
[c] Faculty of Forestry, University of Toronto, 33 Willcocks Street, Toronto, ON M5S 3B3, Canada
[d] Forest Research and Monitoring Section, Ontario Ministry of Natural Resources and Forestry, 1235 Queen Street East, Sault Ste Marie, ON P6A 2E5, Canada

ARTICLE INFO

Article history:
Received 22 May 2014
Received in revised form 13 December 2014
Accepted 18 December 2014

Keywords:
GEOBIA
Segmentation evaluation
Discrepancy measurement
Object recognition
Corresponding segment
SEI

ABSTRACT

Within the framework of geographic object-based image analysis (GEOBIA), segmentation evaluation is one of the most important components and thus plays a critical role in controlling the quality of GEOBIA workflow. Among a variety of segmentation evaluation methods and criteria, discrepancy measurement is believed to be the most useful and is therefore one of the most commonly employed techniques in many applications. Existing measures have largely ignored the importance of object recognition in segmentation evaluation. In this study, a new discrepancy measure of segmentation evaluation index (SEI) redefines the corresponding segment using a two-sided 50% overlap instead of one-sided 50% overlap that has been commonly used. The effectiveness of SEI is further investigated using the schematic segmentation cases and remote sensing images. Results demonstrate that the proposed SEI outperforms the other two existing discrepancy measures, Euclidean Distance 2 (ED2) and Euclidean Distance 3 (ED3), both in terms of object recognition accuracy and identification of detailed segmentation differences.

1. Introduction

Image segmentation is the first and most critical step for geographic object-based image analysis (GEOBIA). It is a process of partitioning the entire image into a number of non-overlapping segments for subsequent object recognition, image classification, or information extraction. In the past decades, a variety of popular segmentation algorithms have been proposed, such as watershed segmentation (Li et al., 2010; Li and Xiao, 2007; Vincent and Soille, 1991; Yang et al., 2014a), mean-shift segmentation (Comaniciu and Meer, 2002), and region-merging segmentation (Baatz and Schäpe, 2000; Benz et al., 2004). Nevertheless, it still remains challenging and problematic to integrate a comprehensive framework for segmentation evaluation (Ryherd and Woodcock, 1996; Shandley et al., 1996). Compared to accuracy assessment of image classification that views individual pixel as the evaluation unit (e.g., error matrix), the quality of image segmentation needs to be measured from the perspective of object recognition; that is, how well segments match real geo-objects.

In general, the methods of segmentation evaluation include visual inspection, quantitative evaluation, and indirect evaluation (e.g., classification accuracy) (Li et al., 2011). Although it might be subjective and qualitative, visual inspection is still a commonly-used evaluation method (Pesaresi and Benediktsson, 2001; Zhang et al., 2008). On the other hand, object-based image classification accuracy can imply the quality of image segmentation, but it is not a direct indication of the segmentation accuracy (Kim et al., 2009; Zhang et al., 2005). Despite the popularity of visual inspection and indirect evaluation, quantitative evaluation is receiving more attention because of the increased amount of information that can be produced using segmentation evaluation (e.g., over-segmentation and under-segmentation).

Quantitative evaluation methods can be categorized into three types: analytical, empirical goodness, and empirical discrepancy, among which empirical discrepancy has been proven the most effective and widely used (Zhang, 1996). To assess the accuracy of image segmentation, the direct empirical discrepancy measurement quantifies the dissimilarity between a reference polygon and a corresponding segment. Several such discrepancy measures have been proposed (Carleer et al., 2005; Clinton et al., 2010; Lucieer and Stein, 2002; Möller et al., 2007; Weidner, 2008; Zhan et al., 2005). These area-based measures, later categorized as geometric

* Corresponding author at: Department of Geography, University of Toronto, 100 St. George Street, Toronto, ON M5S 3G3, Canada. Tel.: +1 416 978 3375.
E-mail address: jiangeo.yang@mail.utoronto.ca (J. Yang).

http://dx.doi.org/10.1016/j.isprsjprs.2014.12.015
0924-2716/

discrepancy (Liu et al., 2012), are able to characterise the non-overlapping area of a reference polygon and a corresponding segment (Fig. 1a), however, they require that a reference polygon only have one corresponding segment. In reality, it is very common for a reference polygon to have several corresponding segments (Fig. 1b). Thus, the prerequisite of a one-to-one correspondence between polygons and segments is almost impossible to satisfy when using the geometric discrepancy measures. The drawback in the geometric discrepancy measures has been recently addressed using newly proposed indices which consider arithmetic discrepancy (Liu et al., 2012). Arithmetic discrepancy is defined as the number of corresponding segments for any reference polygon (Liu et al., 2012).

In general, the discrepancy measures that are proposed by previous studies label a candidate segment as the corresponding segment of a reference polygon when the overlapping area is over 50% of **EITHER** the reference polygon **OR** the candidate segment (Clinton et al., 2010; Liu et al., 2012; Yang et al., 2014b). However, the use of one-sided 50% overlap in the discrepancy measures would not work in many segmentation cases, since segmentation evaluation ignores object recognition. Using two segmentation results for a polygon as an example (Fig. 1b and c), all segments (dotted line) in the two cases can be considered as corresponding segments for the reference polygon (full line), because the overlapping area is over 50% (in fact 100%) of each candidate segment and, the one-sided 50% overlap prerequisite for the discrepancy measures is thus satisfied. Consequently, the two segmentation results for the polygon are exactly the same in terms of geometric discrepancy (i.e. the non-overlapping areas are the same) and arithmetic discrepancy (i.e. three segments corresponding to one reference polygon are the same in both cases). However, it is obvious that the candidate segments in Fig. 1c reflect a higher segmentation quality than those in Fig. 1b, because the largest segment in Fig. 1c can better recognize the reference polygon. This example clearly demonstrated that the current discrepancy measures cannot efficiently recognize objects as a result of only utilizing one-sided 50% overlap. To differentiate these two cases, a new discrepancy (hereafter defined as object-recognized discrepancy) is needed to identify whether or not a corresponding segment is able to properly recognize a reference object. Differing from geometric and arithmetic discrepancy, the new discrepancy will incorporate the object recognition information for a correctly recognized object through the prerequisite that the overlapping area between a reference polygon and the candidate segment has to be more than 50% of **BOTH** the reference polygon **AND** the candidate segment (Lamar et al., 2005).

As far as we know, no study has yet incorporated the information of object recognition into the discrepancy measures, in spite of the fact that object recognition is a most important criterion when measuring the discrepancy between a reference object and a corresponding segment. It is thus of great necessity to identify if the reference object is correctly identified before a detailed discrepancy measurement is calculated. In order to address this knowledge gap, we propose a new discrepancy measure to evaluate the quality of image segmentation which takes into account geometric discrepancy, arithmetic discrepancy, and object-recognized discrepancy. Schematic segmentation cases and remote sensing images are used to examine the performance of the proposed index in comparison with the other two existing discrepancy measures (Liu et al., 2012; Yang et al., 2014b).

2. Discrepancy measures

Since object recognition is one of the most important objectives of image segmentation, this study adopted the two-sided 50% overlap as the prerequisite to identify the corresponding segment for a reference polygon. Under this context, the reference polygon without any corresponding segment will be considered as an omitted or missing object in the process of segmentation evaluation.

Several discrepancy measures that have recently been proposed based on the one-sided 50% overlap of corresponding segments include Potential Segmentation Error (PSE), Number-of-Segments Ratio (NSR) and Euclidian Distance 2 (ED2) (Liu et al. (2012).

$$\begin{aligned} \mathrm{PSE} &= \frac{\mathrm{area}(S_{1S} - R)}{\mathrm{area}(R)} \\ \mathrm{NSR} &= \frac{|m - v|}{m} \\ \mathrm{ED2} &= \sqrt{\mathrm{PSE}^2 + \mathrm{NSR}^2} \end{aligned} \tag{1}$$

where R and S_{1S} are the datasets of reference polygons and one-sided 50% overlap of corresponding segments while m and v are the numbers of reference polygons and corresponding segments. A PSE value of zero implies that there is no under-segmentation, and a NSR value of zero means that there is an optimal one-to-one relationship between the reference polygons and corresponding segments.

Although ED2 considers both geometric and arithmetic matches, the difficulty of PSE and NSR normalization can result in the exaggeration of over-segmentation at finer scales when NSR overwhelms PSE (Yang et al., 2014b). In addition, the compensation effect caused by the coexistence of one-to-many over-segmentation and many-to-one under-segmentation can result in invalid NSR. For instance, since the one-sided 50% overlap fails to ensure that there is at least a candidate segment corresponding to any one reference polygon, the over-segmentation of other reference objects can compensate this effect on the NSR even if there is an omitted or missing reference object. Thus, Yang et al. (2014b) developed the local metrics of OverSegmentation 2 (OS2), UnderSegmentation 2 (US2), and Euclidian Distance 3 (ED3).

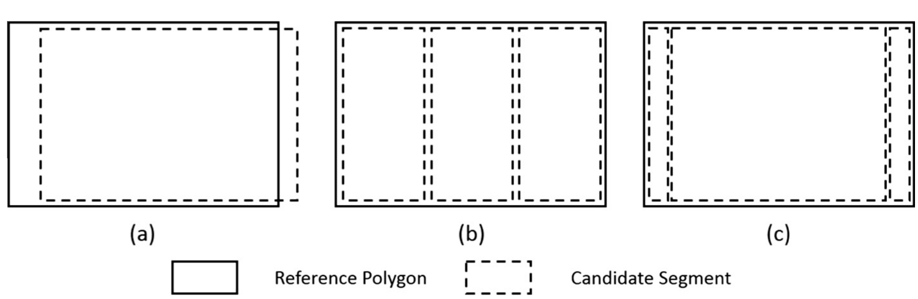

Fig. 1. An illustration of discrepancies between a reference polygon and candidate segments, including geometric discrepancy (a), arithmetic discrepancy (b), and object-recognized discrepancy (c).