Remote Sensing of Environment 205 (2018) 338–351

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Review

Supervised methods of image segmentation accuracy assessment in land cover mapping

Hugo Costa*, Giles M. Foody, Doreen S. Boyd

School of Geography, University of Nottingham, Nottingham NG7 2RD, UK

ARTICLE INFO

Keywords:
OBIA
GEOBIA
Empirical goodness methods
Quality
Classification

ABSTRACT

Land cover mapping via image classification is sometimes realized through object-based image analysis. Objects are typically constructed by partitioning imagery into spatially contiguous groups of pixels through image segmentation and used as the basic spatial unit of analysis. As it is typically desirable to know the accuracy with which the objects have been delimited prior to undertaking the classification, numerous methods have been used for accuracy assessment. This paper reviews the state-of-the-art of image segmentation accuracy assessment in land cover mapping applications. First the literature published in three major remote sensing journals during 2014–2015 is reviewed to provide an overview of the field. This revealed that qualitative assessment based on visual interpretation was a widely-used method, but a range of quantitative approaches is available. In particular, the empirical discrepancy or supervised methods that use reference data for assessment are thoroughly reviewed as they were the most frequently used approach in the literature surveyed. Supervised methods are grouped into two main categories, geometric and non-geometric, and are translated here to a common notation which enables them to be coherently and unambiguously described. Some key considerations on method selection for land cover mapping applications are provided, and some research needs are discussed.

1. Introduction

Land cover mapping is a very common application of remote sensing and has been increasingly conducted through object-based image analysis (Blaschke, 2010). Object-based image analysis has been described as an advantageous alternative to conventional per-pixel image classification, and adopted in a diverse range of studies (Bradley, 2014; Feizizadeh et al., 2017; Matikainen et al., 2017; Strasser and Lang, 2015).

Objects are typically discrete and mutually exclusive groups of neighbouring pixels and used as the basic spatial unit of analysis. Objects may be delimited or obtained via a range of sources (e.g. cadastral data), but typically are constructed through an image segmentation analysis, and thus often called segments. In this paper the terms “object” and “segment” are used synonymously. Image segmentation is performed by algorithms with the purpose of constructing objects corresponding to geographical features distinguishable in the remotely sensed data, which may be useful for applications such as land cover mapping.

Constructing objects poses a set of challenges. For example, it is necessary to select a segmentation algorithm from the numerous options available, but comparative studies (e.g. Basaeed et al., 2016; Neubert et al., 2008) are uncommon. Also each of the segmentation algorithms is typically able to produce a vast number of outputs depending on the parameter settings used. Selecting the most appropriate segmentation is, therefore, difficult.

Multiple methods have been proposed to assess the accuracy of an image segmentation and are normally grouped in two main categories: empirical discrepancy and empirical goodness methods, also commonly referred to as supervised and unsupervised methods respectively (Zhang, 1996). Most of the supervised methods essentially compare a segmentation output to a reference data set and measure the similarity or discrepancy between the two representations (e.g. overlapping area) (Clinton et al., 2010). Unsupervised methods measure some desirable properties of the segmentation outputs (e.g. object's spectral homogeneity), thus measuring their quality (Zhang et al., 2008).

There is no standard approach for image segmentation accuracy assessment, and some studies have compared accuracy assessment methods. Supervised and unsupervised methods are normally compared separately. For example, with regard to supervised methods, Clinton et al. (2010), Räsänen et al. (2013), and Whiteside et al. (2014) compared dozens of methods, all of them focused on some geometric property of the objects, such as positional accuracy relative to the reference data. These and other studies highlight the differences and

* Corresponding author.
E-mail address: hugoagcosta@gmail.com (H. Costa).

https://doi.org/10.1016/j.rse.2017.11.024
Received 27 December 2016; Received in revised form 20 November 2017; Accepted 29 November 2017
0034-4257/

similarities obtained from the methods compared so the reader gains a perspective of the field. However, many other supervised methods have been proposed yet are barely compared against previous counterparts; these tend to be newly proposed methods (e.g. Costa et al., 2015; Liu and Xia, 2010; Marpu et al., 2010; Su and Zhang, 2017). Furthermore, the methods are often described using a notation suitable for the specific case under discussion, which makes the cross-comparison of methods difficult.

Studies like Clinton et al. (2010) are valuable in reviewing the field of image segmentation accuracy assessment, but they often focus on the geometry of the objects evaluated and ignore that a supervised but non-geometric approach may be followed (e.g. Wang et al., 2004). Moreover, supervised methods are typically compared within a specific study case without discussion of further and important issues, such as the suitability of the methods as a function of context. As image segmentation is increasingly used in a wide range of applications, the behaviour and utility of specific methods is expected to vary in each case. Thus, selecting a method to assess the accuracy of image segmentation may be based on an incomplete understanding of the available options and ultimately problematic.

This paper reviews the state-of-the-art of image segmentation accuracy assessment in land cover mapping applications. The literature published in three major remote sensing journals in 2014–2015 is reviewed to provide an overview of the field, namely the methods used and their popularity. In particular, the supervised methods are thoroughly reviewed as they are widely used. A comprehensive description of which supervised methods are available is presented with the aim of providing a basis on which the remote sensing community may consider and select a suitable method for particular applications. A discussion on which methods should be used is provided, and research needs are highlighted.

2. Background

Image objects are typically expected to delimit features of the Earth's surface such as land cover patches that are remotely sensed using an air/spaceborne imaging system. Image segmentation cannot, however, deliver results exactly according to the desired outcome for multiple reasons, such as unsuitable definition of segmentation algorithm parameter settings, and insufficient spectral and spatial resolution of the data. Thus, image segmentation error is common, namely under- and over-segmentation. Under-segmentation error occurs when image segmentation fails to define individual objects to represent different contiguous land cover classes, thus constructing a single object that may contain more than one land cover class. On the contrary, over-segmentation error occurs when unnecessary boundaries are delimited, and thus multiple contiguous objects, potentially of the same land cover class, are formed.

Segmentation errors have been traditionally identified through visual inspection, but it has some drawbacks, especially when assessing large areas and comparing numerous segmentation outputs. Specifically, visual interpretation is time consuming, subjective, and the results produced by the same or different operators may not be reproducible (Van Coillie et al., 2014; Lang et al., 2010). As a result, objective and quantitative methods for the assessment of image segmentation accuracy may be necessary and have become more popular in recent years.

The literature published during 2014–2015 in three remote sensing journals was reviewed to provide an overview of the state-of-the-art of image segmentation accuracy assessment. The journals were *Remote Sensing of Environment*, *ISPRS Journal of Photogrammetry and Remote Sensing*, and *Remote Sensing Letters*. These journals were selected to represent the variety of current publication outlets in the field. Historically, the former journal has had the greatest impact factor among the remote sensing journals. The second journal has been particularly active in publishing papers on object-based image analysis. The latter journal is a relatively young journal dedicated to rapid publications. The papers that included specific terms (namely "obia", "geobia", "object-based", and "object-oriented") in the title, abstract, and key words were retained for analysis. A total of 55 out of 67 papers that matched the search terms were identified as relevant, each describing techniques for constructing objects which were used as the basic spatial unit in land cover mapping applications.

These 55 papers were analysed, and it was noticeable that 17 papers (30.9%) do not document if or how the accuracy of the image segmentation outputs was assessed. This shows that image segmentation accuracy assessment is often overlooked as an important component of an image segmentation analysis protocol. It is speculated that visual interpretation was used in most of the cases that provide no information accuracy, as having used no sophisticated method may reduce any motivation for documenting the topic. The remaining 38 papers explicitly described the methods used, and often more than one method was adopted. Visual interpretation was widely used, with 15 papers (25.3% of the total of papers) describing that the qualitative appearance of the segmentations influenced the assessment of the results (e.g. Qi et al., 2015). Details were typically not given, such as the time dedicated to visual interpretation and number of interpreters.

When a quantitative alternative to subjective visual interpretation was explicitly adopted, the methods used varied widely. A rudimentary strategy of assessing the accuracy of image segmentations, and used in five papers (9.1%), was to use simple descriptive statistics, such as the average of some attributes of the objects like area, to get an impression of the segmentation output. The statistics were used in a supervised or unsupervised fashion. In the former situation, the statistics were compared to the statistics of a reference data set depicting desired polygonal shapes, and small differences were regarded as indicative of large segmentation accuracy (e.g. Liu et al., 2015). When no reference data were used (i.e. unsupervised fashion), the statistics identified the image segmentation from the set obtained with the most desirable properties, such as a target mean size (i.e. area) of the objects (Hultquist et al., 2014). Although descriptive statistics can measure some quantitative properties of an image segmentation, they provide a very limited sense of the accuracy of the objects, for example in the spatial domain, and here they are not regarded as a true accuracy assessment method. The latter are typically more evolved and normally grouped into supervised and unsupervised methods.

Supervised methods were found in 21 (38.2%) of the papers reviewed (e.g. Zhang et al., 2014). Although there was no dominant method, the Area Fit Index (Lucieer and Stein, 2002) and Euclidean distance 2 (Liu et al., 2012) were the supervised methods that were most used with three appearances each (Belgiu and Drăguţ, 2014; Drăguţ et al., 2014; Witharana et al., 2014; Witharana and Civco, 2014; Yang et al., 2014). Many of the other methods identified were used only once (e.g. Carleer et al., 2005). These and other supervised methods are, however, thoroughly described in the next section. Unsupervised methods were applied in 13 (23.6%) of the papers surveyed (e.g. Robson et al., 2015). The unsupervised method most used in the literature reviewed was the Estimation of Scale Parameter (ESP or ESP2) tool (Drăguţ et al., 2010, 2014) available in the popular eCognition software. The segmentation algorithms available in this software were used in most of the papers surveyed (36 papers, 65.5%) to construct image objects.

Object-based image analysis has received much attention and acceptance (Blaschke et al., 2014; Dronova, 2015), but the accuracy assessment of image segmentation, which is a central stage of the analysis, appears to be in a relatively early stage of maturation. Although procedures for image segmentation accuracy assessments have not been standardized, a more harmonized approach is desirable. Using subjective visual interpretation may be acceptable and suitable for some applications; the reasons are seldom explained in the literature. Among the quantitative methods proposed for image segmentation accuracy assessment, supervised approaches seem to be the most frequently