International Journal of Applied Earth Observation and Geoinformation 15 (2012) 7–15

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

Accuracy assessment of contextual classification results for vegetation mapping

Guy Thoonen[a,*], Koen Hufkens[b], Jeroen Vanden Borre[c], Toon Spanhove[c], Paul Scheunders[a]

[a] IBBT, Vision Lab, University of Antwerp, Universiteitsplein 1, B-2610 Antwerpen, Belgium
[b] Department of Geography & Environment, Boston University, 675 Commonwealth Avenue, Boston, MA 02215, USA
[c] Research Institute for Nature and Forest (INBO), Kliniekstraat 25, B-1070 Brussel, Belgium

ARTICLE INFO

Article history:
Received 3 November 2010
Accepted 18 May 2011

Keywords:
Contextual classification
Natura 2000
Accuracy assessment
Confusion matrix

ABSTRACT

A new procedure for quantitatively assessing the geometric accuracy of thematic maps, obtained from classifying hyperspectral remote sensing data, is presented. More specifically, the methodology is aimed at the comparison between results from any of the currently popular contextual classification strategies. The proposed procedure characterises the shapes of all objects in a classified image by defining an appropriate reference and a new quality measure. The results from the proposed procedure are represented in an intuitive way, by means of an error matrix, analogous to the confusion matrix used in traditional thematic accuracy representation. A suitable application for the methodology is vegetation mapping, where lots of closely related and spatially connected land cover types are to be distinguished. Consequently, the procedure is tested on a heathland vegetation mapping problem, related to Natura 2000 habitat monitoring. Object-based mapping and Markov Random Field classification results are compared, showing that the selected Markov Random Fields approach is more suitable for the fine-scale problem at hand, which is confirmed by the proposed procedure.

1. Introduction

In the past decades, classification of hyperspectral imagery has become a major research topic in the remote sensing field. Over the years, a multitude of algorithms has been developed and applied (Lu and Weng, 2007). In addition, especially considering the strong increase in spatial resolution, researchers have acknowledged the benefits of including contextual information in the classification process. Indeed, instead of only considering isolated pixels, experiments have shown that also taking into account information from neighbourhood relationships potentially leads to higher classification accuracies (Dell'Acqua et al., 2004). Furthermore, the results are often smoother, show a clearer delineation of the boundaries between classes, and, as a result, are much easier to interpret.

Among the many implementations for including contextual information, three major strategies can be identified:

- Including morphological or texture information in the feature vectors together with the spectral information (Dalla Mura et al., 2010; Murray et al., 2010).
- Incorporating contextual information as prior information in a Bayesian framework and modelling dependencies between neighbouring pixels as Markov Random Fields (MRF) (Jackson and Landgrebe, 2002; Tso and Olsen, 2005).
- Performing classification after first segmenting an image, referred to as Object-based Image Analysis (OBIA) (Zhou et al., 2009; Yu et al., 2006).

Regardless of the strategy selected, it is desirable to assess the accuracy of the classification results. Traditionally, this is done by comparing classification results with reference data collected on the ground and, subsequently, representing the agreement between both sets of data in an error matrix or confusion matrix (Congalton and Green, 1999). A popular statistical measure for the agreement of two classification results, derived from the confusion matrix, is the Kappa coefficient (Bishop et al., 1977). Unfortunately, collection of data in the field is very expensive and time-consuming. Consequently, ground reference data usually consist of a limited number of isolated points, containing little information on the transitions between classes (Foody, 2002). As a result, these measures only provide information on the accuracy of the labels, called thematic accuracy.

Therefore, in addition, visual inspection is often used in practice to assess the geometric quality of a thematic map, i.e. the accuracy of the class transitions. Unfortunately, this procedure is subjective and does not deliver quantitative evidence. Other procedures, which originate from the assessment of segmentation maps,

* Corresponding author. Tel.: +32 3 265 2444; fax: +32 3 265 2245.
E-mail addresses: guy.thoonen@ua.ac.be (G. Thoonen), khufkens@bu.edu (K. Hufkens), jeroen.vandenborre@inbo.be (J. Vanden Borre), toon.spanhove@inbo.be (T. Spanhove), paul.scheunders@ua.ac.be (P. Scheunders).

0303-2434/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jag.2011.05.013

involve quantitative evaluation of manually delineated reference objects, which are selected based on the high resolution property of the aerial or satellite imagery (Esch et al., 2008; Li et al., 2010; Möller et al., 2007).

Recently, Persello and Bruzzone (2010) have introduced a protocol for the accuracy assessment of classification results of very high resolution images. This protocol adequately separates thematic accuracy assessment from geometric accuracy assessment, in which the former separately measures the contribution of homogeneous and transition regions. The latter, on the other hand, is determined by selecting a number of reference objects by photo-interpretation and using a set of measures to characterise the differences between the classification map and these reference objects. The set of measures include oversegmentation, undersegmentation, edge location, fragmentation error and shape error. Ultimately, thematic and geometric measures can be combined in a global measure. It should be noted that the delineated reference objects do not need to contain a single class, nor should all classes be included in all the reference objects, only those classes for which the geometric properties are relevant. For further details, the reader is referred to Persello and Bruzzone (2010).

However, in certain applications, like for instance vegetation monitoring, it can be very difficult to visually delineate areas on aerial or satellite images at a scale that is appropriate to the detail of the land cover types under study. Often, the resulting objects will be too coarse and contain a multitude of smaller objects of many detailed classes, that cannot be perceived by eye. For instance, it is possible to manually delineate the transitions between a heathland and a forest area, transitions that are likely to be correctly detected by many classification techniques. Within the resulting heathland object, however, many land cover types (e.g. heather bushes, bare sand, moss patches, . . .) are spatially interwoven at a very fine scale, most probably leading to high fragmentation and oversegmentation measures, while shape and edge correspondence measures are likely to be good, no matter which classification strategy is used. Nevertheless, the spatial arrangements of these small objects in the maps products can be very different, depending on the classification methodology.

Still, knowledge of the constituent land cover types of a habitat object is often desirable in vegetation monitoring, e.g. to assess and keep track of the quality of a given habitat area. Moreover, numerous studies have linked patch characteristics with biodiversity in general and, more specifically, with vegetation diversity, species distribution and abundance, seed dispersal, erosion and micro-meteorological properties (Favier et al., 2004; Andrén, 1994). Unfortunately, in field mapping, it is simply too complex and time-consuming to map all the individual land cover patches. Therefore, field mappers delineate and label objects at a higher level, using, e.g. predefined codes for mosaics, and estimate cover of each of the subtypes in percentage (Bunce et al., 2008; JNCC, 2007). Cover estimation is, however, known to be highly influenced by between-observer variation, and can only be improved through sustained training effort (Gallegos Torell and Glimskär, 2009). Moreover, this methodology does not provide location-information for each of the land cover types in the object, leaving thus very little reliable data for monitoring of changes. Nevertheless, considering the above, patches are often considered a key factor in the analysis of the landscape mosaic. Consequently, there is a need to accurately classify vegetation patches and their intricate relationship to each other, without the interference from classification noise. While remote sensing has the potential to provide very fine-scaled mapping (Xie et al., 2008), measures to assess their performance are still lacking, especially considering the geometric properties of small, but meaningful objects.

To close this gap, this paper presents a method that characterises the shapes of all objects, i.e. patches of connected pixels of the same class, in the image. The method is directly aimed at assessing the accuracy of the classification results of any of the aforementioned contextual classification strategies. To this end, a new, relative reference is defined, that allows comparison of contextual classification strategies that use the same spectral classifier. The effects of using contextual information are characterised by introducing measures for the change in edge shapes between objects with respect to the reference. Moreover, the results are presented in a way that is closely related to the representation of the confusion matrix, well-known from its application in thematic accuracy assessment. Note that the proposed reference is not absolute, hence it is not intended to be used as an alternative to the references and methods described above. Rather, it serves as a valuable addition to these other techniques, because it considers important elements that these techniques cannot perceive.

Section 2 provides some background on the application the method is intended to be used with. Section 3 describes the details of the methodology. Section 4 shows the results of our experiments and discusses these results. Finally, the conclusions of this paper are presented in Section 5.

2. Background

The context in which the method is applied, is a Belgian multidisciplinary project on habitat status monitoring, called Habistat (Haest et al., 2000). Elements such as habitat loss, climate change and invasive alien species are important causes of the current biodiversity crisis. The goal of the Habitat directive of the European Union is to protect rare or endangered habitats or species. One of the specific measures is the foundation of the Natura 2000 network (EEC, 1992), an ecological network of protected areas, spread over the whole continent. Nature conservation for these areas is the responsibility of the European member states, each of which needs to take appropriate measures to bring and maintain the sites on their respective territory in a good conservation status. Moreover, the member states are committed to report on the status of the Natura 2000 sites, habitats and species on a regular basis. This way, it is possible to keep track of the trends and take conservation measures whenever appropriate.

The goal of the Habistat project is to investigate the use of remote sensing as a tool to aid in Natura 2000 habitat monitoring. However, habitats are usually not homogeneous vegetation patches of a single or a few dominant species. Instead, they show a high variety in facies at different scale levels. At a large scale, the facies of the same habitat may differ between regions as a result of climatic or soil conditions. But also at a very fine scale, most habitats are in fact intricate mixtures of different land cover types. Within the project, a framework has been developed to deal with the complete trajectory, from breaking down habitats into land cover types, to reconstructing habitat types and their conservation status from land cover classification results (Haest et al., 2000). As a consequence, land cover classification is a key step within the framework, as is its accuracy assessment. Within the project, the framework has been developed for Western European heathlands.

The amount of detail required to determine the habitat types, and additionally, their status, leads to a high number of land cover classes. For instance, several age classes of Calluna-dominated heathland (young, adult, old, . . .) are to be recognised. While these classes are spatially intertwined at a very fine scale and, hence, they will most likely benefit from using contextual classification techniques, distinguishing them by visual interpretation of the image is non trivial. This situation leads to the example from Section 1, where only coarse objects with respect to the detail of the land cover types can be delineated, while the land cover types form small, yet important, objects in their own right, which makes this