Object-based contextual image classification built on image segmentation

Thomas Blaschke
Department of Geography & Geoinformatics
University of Salzburg
Salzburg, Austria
Thomas.blaschke@sbg.ac.at

Abstract— The continuously improving spatial resolution of remote sensing sensors sets new demand for applications utilizing this information. The need for the more efficient extraction of information from high resolution RS imagery and the seamless integration of this information into Geographic Information System (GIS) databases is driving geo-information theory and methodology into new territory. As the dimension of the ground instantaneous field of view (GIFOV), or pixel size, decreases many more fine landscape features can be readily delineated, at least visually. The challenge has been to produce proven man-machine methods that externalize and improve on human interpretation skills. Some of the most promising results come from the adoption of image segmentation algorithms and the development of so-called object-based classification methodologies. This paper builds on a discussion of different approaches to image segmentation techniques and demonstrates through several applications how segmentation and object-based methods improve on pixel-based image analysis/classification methods. In contrast to pixel-based procedure, image objects can carry many more attributes than only spectral information. In this paper, I address the concepts of object-based image processing, and present an approach that integrates the concept of object-based processing into the image classification process. Object-based processing not only considers contextual information but also information about the shape of and the spatial relations between the image regions.

Keywords - multiscale; segmentation; object-based; classification, context classifier

I. INTRODUCTION

During the last decade, the development of data acquisition techniques has fostered the use of numerical methods to process the large volumes of data that are collected in remote sensing platforms. Increased spatial information may be valuable in a variety of situations. For instance, the recent range of hyperspectral data, high resolution data, and SAR data (e.g. ASTER, MODIS, QUICKBIRD, IKONOS, KOMPSAT-2, ENVISAT, ALOS, and so on) will provide detailed attribute information. For the remotely sensed data classification, many classifiers based on the spectral analysis of individual pixels have been proposed and significant progress has been achieved. However, these approaches have their limitations since most remote sensing image classification techniques are based on per-pixel procedures [7]. They analyze pixels mainly using multivariate statistics. So the current results still cannot compare with human photo-interpreters [33].

An obvious new requirement in multiple image applications is the need for image mosaicing to create a seamless, large-area database that would cover whole regions of similar ecology and physiography [27]. With the aim of achieving higher accuracy, integration of additional data has been investigated such as multi-source data fusion approach and contextual classification models [29, 6, 40, 39, 35]. Nevertheless, human photo-interpreters also implicitly use structural knowledge in the manual classification process. They do not only consider contextual information but also information about the shape of and the spatial relations between the image regions. This type of process has hardly been utilized in previous research. Considering this condition, a completely different type of information analysis in images and other kind of geographic information is slowly gaining more attraction, namely object-based image analysis. As it will be laid out in the following chapter these approaches are not mature but very promising. The main advantages are expected from satellite image classification using expert structural knowledge and object-based image classification approach [4, 7].
In this paper, I discuss contextual image classification approaches and present an approach that combines approach that integrates the concept of object-based processing into the image classification process. First, I will introduce the concept of object-based image analysis, and compare this concept with traditional pixel-based image analysis. The core concept of object-based image analysis is that the important information necessary to interpret an image is not represented in single pixels, but in meaningful image objects and their mutual relationship. I will demonstrate how segmentation and object-based methods improve on traditional pixel-based image analysis/classification methods and I will position this approach against alternatives and demonstrate some example results from several studies.

II. REVIEW OF EXISTING METHODOLOGIES

A. Per-pixel methods

Traditional per-pixel classification methods have provided adequate results in many applications [17, 18], but it has been argued that many such techniques do not make full use of available spatial information [7], despite incorporation of pixel-based texture measurements such as those based on convolution filters [28] and on the grey-level co-occurrence matrix [22]. For instance, many forest features of interest, such as cut blocks and seismic lines, can be visually identified through associations with other classes or by the presence of straight lines. Per-pixel classification cannot conveniently incorporate this knowledge [7, 16, 10].

For a homogeneous feature to be detected, its size generally has to be significantly larger than the resolution cell. If the feature is smaller than this, it may not be detectable as the average brightness of all features in that resolution cell will be convolved, although sub-pixel algorithms and spectral unmixing techniques try to split up the proportion of the signal from different known surfaces characteristics.

The general critique is that traditional methods for analysis of remotely sensed image data is the classification of pixels based on pixels in the same land cover class being close in spectral feature space. This does not hold true for complex environments and their respective classifications. In addition, the pixel-centered view is usually uni-scale in methodology, exploring the pixels of only one scale of imagery and of only one scale within the image [7, 11].

In most cases, information important for the understanding of an image is not represented in single pixels but in meaningful image objects and their mutual relations. Prerequisite for the successful automation of image interpretation are therefore procedures for image object extraction which are able to dissect images into sets of useful image objects. This is usually done using image segmentation techniques.

B. Image segmentation

Segmentation is not new [30, 22], but only a few of the existing approaches are widely available in commercial software packages and lead to qualitatively convincing results while being robust and operational. One reason is that the segmentation of an image into a given number of regions is a problem with a huge number of possible solutions and it is opposed to the so called Modifiable Area Unit Problem (MAUP) [25]. The high degrees of freedom must be reduced to a few which are satisfying the given requirements. Additionally, segmentation needs to address a certain scale: does the application require information about single bushes or trees or about land cover units such as orchards or mires? Most segmentation approaches don’t allow the user to specify a certain scale of consideration and a level of detail or generalization, accordingly. Some region growing algorithms use thresholds which usually embrace the summed variation of all bands used.

![Image Segmentation](image segregation of a small subset of an infrared orthophoto of the city of Salzburg, Austria.)

There are two ways of achieving the segmentation: 1) prior knowledge (e.g., physical units such as soil types); or 2) automatic division of the space based on some property of the underlying statistics, e.g. the variogram. Segmentation based on prior knowledge has been implemented for many years. For example, [5] segmented a remotely sensed image based on vector polygon data and used variograms estimated within segments to aid per-field classification. Data on some prior 'stratification' are generally incorporated into the kriging
process via stratified kriging which is nothing more than kriging applied to the strata independently [19]. Either way, image segmentation aims for the delineation of relatively homogeneous areas.

C. Context information and texture

Dissecting an image is one way to incorporate context information of the pixels under consideration. Increased classification accuracy can be obtained if it is possible to avoid relying exclusively on the per-pixel spectral response pattern and instead employ different image data. As stated above, the strong motivation to develop techniques for the extraction of image objects stems from the fact that most image data exhibit characteristic texture which is neglected in common pixel-based classifications. "The two most obvious types of image data that are underutilized in current classification procedures are image context and image texture" ([18], p.91). Context relies on a summary description of the relationships among pixels or, more frequently, among classes. The premise is that a pixel’s most probable classification, when viewed in isolation, may change when viewed in some context [21, 6, 13, 40].

Figure 3. Pixel classification with and without incorporating spatial context.

Image texture is a quantification of the spatial variation of image tone values that defies precise definition because of its perceptual character [24]. Insight into how texture might be analyzed by computer has focused on the structural and statistical properties of textures [23, 20]; subsequently, much effort has been expended on optimizing satellite image texture measures for the land cover mapping application [34, 37].

In principle, texture and other relevant information can be analyzed from the immediate neighborhood of the pixel and results can be assigned to the central pixel. Examples of this are moving window filters which can be implemented with help of convolution masks [28]. If texture is integrated in remote sensing applications, it is usually done by creating additional image layer through moving window technology where the resulting values represent the variety, variance or range of neighboring pixels of a focal pixel, e.g. in a 5 by 5 pixel kernel. Context examples which are typically expressed through moving window techniques: Variance in tone and color, equality of reflectance values of neighboring pixels, variety of values, range, directionality or lack of it.

The texture of an image can be defined in terms of its smoothness or its roughness. One field of image processing in which the quantification of texture plays a crucial role is that of industrial/machine vision. These systems are used to assess the quality of products by measuring the texture of their surface. Most methods are based on the statistical properties of an image as well as the spectral or Fourier characteristics of airborne data, radar or VHR-satellite data which are playing an increasing role in remote sensing. But how to include neighborhood information across several spectral bands for a pixel-based analysis? Several research groups tried to do this by using pre-defined boundaries (‘per-parcel classification’ or ‘per-field classification’, see [3]). Besides methodological questions one also has to ask, what to do in case there are no boundaries readily available or exactly those boundaries should be updated. Other solutions to include neighborhood in the image analysis are based on filters or kernel techniques, ranging from simple moving window implementations [31, 20] to recent sophisticated methodologies. Recently, [35] introduced an iterative statistical approach to fuse spectral information with spatial and temporal contextual information for the classification of multitemporal and/or multisensor remote-sensing images. Their experimental results show, in terms of classification accuracy, the improvement that can be reached by exploiting the contextual information. Reference [29] proposed a cascade spatio-temporal contextual classifier that involves Markov random models (MRFs) to model spatial correlation. This global joint classification scheme provides a solution in terms of the Bayes rule for minimum error. However, most of these techniques or methodologies, respectively, show huge computational load and are hardly transferable.

D. Object-oriented approaches

In object oriented analysis shape and context are clumped into a fourth attribute, that defining Fellowship [10]; ‘to which pixel population does this pixel belong’ [14]. It has been suggested that classifying remotely sensed images in pixel wise fashion (using only the first three pixel attributes) is a special case of the super-set of object-based classification methodologies which utilize all four [38, 14]. This is not illogical and [10] propose a continuum of classification methods incorporating more or less fellowship information.

To extract objects of interest, the statistical analysis of pixels exclusively based on their spectral statistic is not sufficient. As laid out in several recent publications [7, 15, 16], the advent of higher resolution image data increased the need for more efficient methods more than ever. Generally, for high resolution data, segmentation as a pre-classification step is preferred over pixel based classification because the resulting division of space tends to involve fewer and more compact subregions. However, [1] have implemented classification-based approaches within a geostatistical framework. Figure 4 illustrates the main difference when building objects prior to the classification process. Some degree of generalization is applied in this phase. For many applications, e.g. land cover mapping, generalization is intrinsically required to produce tangible target objects of the same class which are relatively homogenous according to the class definition.
To summarize, the main advantage of a segmentation algorithm is that spatial proximity of like classes is promoted and objects consisting of several pixels result. Consequently, the working hypothesis for this paper is, that segmentation approaches are generally more suitable for high resolution data, where pixels tend to be spatially clumped.

III. THE MULTISCALE IMAGE SEGMENTATION / OBJECT RELATIONSHIP MODELING APPROACH

Although image segmentation is not new [22] and many different image segmentation algorithms exist, a surprisingly small number of methods are embedded in widely available software packages. Recently, [36] compared six different available software packages. They reported significant differences in the results and concluded that some ‘unsupervised’ segmentation approaches (without a-priori definition of the range of target scales) leads to unconvincing results. Driven by the dissemination of a commercial software package called ‘eCognition’ (www.definiens-imaging.com) this multiscale segmentation approach is becoming popular mainly in Europe but increasingly worldwide [16]. Recently, [11] developed a five step methodology for landscape analysis built on this multiscale segmentation and called it multiscale segmentation / object relationship modeling (MSS/ORM). This methodology is used for guidance in this section and for the examples applications.

A. Multiscale segmentation

The multiscale consideration of landscape pattern gains much attraction recently but the realization in practice becomes difficult and data-intensive. Only for some small areas, field surveys and mapping at different scale is possible. One step forward to support this approach is a nested multi-scale image processing of the same data sources. The multi-scale segmentation searches for the gradient of flux zones between and within patches: areas where the varying strengths of interactions between ‘holons’ [11] produce surfaces. Multi-scale segmentation equates to searching for changes in image object heterogeneity/homogeneity.

A model of the relationships between the segmented image objects is built. Some object relationships are automatically derived. Using an object-oriented software environment, the resulting different object scales have to be logically connected. In this setting each object ‘knows’ its intrinsic relation to its super object (is within) and its sub objects (are contained) as well as the relations to the neighboring objects at the same scale. For instance, the characteristics of level –1 objects (such as mean spectral values, spectral value heterogeneity, and sub-object density, shape and distribution) can be automatically calculated and stored in the description of each level 0 object. Other relationships are semantic, requiring the knowledge of the expert on the landscape in question. This relationship model information is stored in the system through a variety of mechanisms, for example as attributes in GIS vector objects or in a proprietary object-orientated database format.
the multiscale segmentation / object relationship modeling approach of [7]. Note that at each level, objects intrinsically ‘know’ their size and shape, their neighbors, respective border lengths and so on. Vertically, objects ‘know’ their super-objects in which they are embedded and the number, sizes, shapes, spectral characteristics etc. of the subobjects which they include.

B. Classification

Object-based classification starts with the crucial initial step of grouping neighboring pixels into meaningful areas, especially for the end-user. The segmentation and object (topology) generation must be set according to the resolution and the scale of the expected objects. The spatial context plays a modest role in pixel based analysis. Filter operations, which are an important part of the pixel based spatial analysis, have the limitation of their window size. In object analysis, this limitation does not exist. The spatial context can be described in terms of topologic relations of neighboring objects.

Land cover and individual vegetation cover types are usually considered within the context of a classification hierarchy invoked by a conceptual model of vegetation as a geographic phenomenon (gradients or patches, mapped as fields or entities, on the basis of vegetation attributes alone, or vegetation and environmental attributes) [18]. An example of a hierarchical land cover classification system is the Anderson Land Use and Land Cover Classification System [2] comprised of four Levels (I, II, III, IV) designed for use with a variety of remotely sensed data. The system assumes that no one ideal classification of land use and land cover can be developed, but flexible classes and an open-ended structure can be used to accommodate many of the different uses that such classification maps are intended to serve [41]. This classification logic can be well represented by an object-oriented or object-based classification scheme. Some results of an object-based classification are shown in figure 8.

IV. DISCUSSION

In the past, traditional per-pixel classifiers often misclassified low density residential neighborhoods as forest due to the extensive canopy cover and the small spatial signature of individual or small clusters of houses. By classifying “neighborhoods” in a large-segment level, and “forest” or “impervious” in a small-segment level within the “neighborhood” larger segments, classes such as turf-and-tree and residential could be identified [9, 12, 16, 32, 36].
It was stated that defining the segmentation levels is a crucial step in the methodology in order to 'hit' appropriate levels of detail and subsequently the objects of interest. While many unsupervised segmentation algorithms don’t require user decisions the multi-scale approach necessitates the definition of levels. But how to define the image objects? What should be the rule of thumb for a segmentation or a pre-segmentation of image primitives which can then build up the corresponding objects? Only very recently, several approaches in image analysis and pattern recognition are exploited to generate hypothesis for the segmentation rules as an alternative to knowledge-based segmentation [8, 32, 26].

The new data available necessitate improved techniques to fully utilize the potential resulting from the combination of high-resolution imagery and the variety of medium-resolution multi-spectral imagery widely available. High-resolution panchromatic images exhibit variety within a so far ‘homogeneous’ area (a pixel in a medium-resolution image). This ‘within-object heterogeneity’ (for anthropogenic objects) or ‘within-patch heterogeneity’ (for natural or near-natural features) can be exploited explicitly beyond texture algorithms by bridging remote sensing, image analysis and GIS functionality based on objects. Several successful applications exist: [9] utilize this approach for the analysis of bush and shrub encroachment in Germany, while [12] apply sub-object analysis to different mire systems in Europe. The understanding of local heterogeneity in a panchromatic image has a strong effect on the standard deviation value in the same ‘window’ area of the image. Nevertheless, they can be simultaneously used to classify areas, where spectral values in multispectral bands are less important compared to local texture.

TABLE I. EVALUATION OF THE SEGMENTATION-BASED CLASSIFICATION

<table>
<thead>
<tr>
<th>Method</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmentation process as a technical process (on different scales),</td>
<td>Challenging, no standards</td>
</tr>
<tr>
<td>Building relations to neighbours sub-, superobjects</td>
<td>Easy using GIS-functionality, intrinsic process in-oo-software environment</td>
</tr>
<tr>
<td>Building relations to sub- and superobjects with coarser and finer resolution, respectively.</td>
<td>Only for multi-scale segmentation</td>
</tr>
<tr>
<td>Data fusion (different sensor, resolution, raster, vector data)</td>
<td>High potential</td>
</tr>
<tr>
<td>Re-usable semantic models (classes)</td>
<td>High potential</td>
</tr>
<tr>
<td>Feature analysis (spectral, shape, neighbourhood relation features)</td>
<td>One of the main strengths</td>
</tr>
<tr>
<td>Usability, initial learning curve</td>
<td>Similar</td>
</tr>
<tr>
<td>Defining scale parameter for segmentation</td>
<td>Difficult, most problematic issue</td>
</tr>
<tr>
<td>Automation, repeatability, transferability</td>
<td>High potential, first empirical tests, e.g. [16]</td>
</tr>
<tr>
<td>Performance</td>
<td>Problematic for large images</td>
</tr>
<tr>
<td>Accuracy assessment</td>
<td>Unsolved (see discussion)</td>
</tr>
</tbody>
</table>

One advantage of the MSS/ORM approach is the potential to transfer the classification rule system to other scenes. While this was claimed earlier as hypotheses [14, 7] it could recently empirically verified by [16]. These authors demonstrated that a rule set developed for a subset of a Landsat ETM image was applied to another subset without significant loss in the resulting accuracy. The rule system is applied using a protocol and so it can be rerun without extensive preparation. It also facilitates some adaptation of the rules where appropriate, while not requiring complete restructuring of the classification protocol where it is not necessary.

The approach used provides certain options to assess the fuzzy membership classification values for each object and their respective stabilities. The fuzzy logic that forms the formal basis for classification assigns a membership value to each object between zero (totally ambiguous) and one (unambiguous) for each potential class. Broadly speaking, they include using logical operators to combine functions generated for each input feature (such as an image spectral band). The functions are determined by user input or automatically generated maximum and minimum values, and by slope and curve shape in relation to the selected feature. This allows to assess ‘classification stability’, which is defined as the difference between the fuzzy score of the best class (highest fuzzy membership) and that of the second-best class. But for the real accuracy assessment – the comparison with ground truth – still many problems remain. The comparison with the landcover data sets can’t be based on pixel statistics anymore. Rather, technically speaking, polygons have to be compared against polygons through various GIS overlay techniques taking into account the varying sizes and different degrees of overlap and misclassification, respectively.

Context oriented, object-based image analysis and classification attains much attraction recently. It is demonstrated that this promises a high potential within integrated GIS/RS image analysis. Comprehensive studies using multi-sensor data sets which explore the ‘behavior’ (stability and consistency of the image objects and their respective classification results due to different data situations) are urgently required. Effective image analysis techniques depend on analyst knowledge at some stage in the procedure. Object-based classification is designed to maximally tap expert knowledge and allow its efficient incorporation into image analysis while maintaining a record of how it was applied [16]. The MSS/ORM approach seems to be able incorporate expert information and external knowledge. While none of the various pixel-based classification methods seems to really satisfy all the needs for the production of reliable, robust and accurate information similar to objects identified by a human interpreter, this approach seems to be a step forward from land cover to land use classifications.

ACKNOWLEDGMENT

Cooperation with Definiens AG, Munich (www.definiens-imaging.com) is gratefully acknowledged. I thank the Landscape Analysis and Resource Management Research Group LARG for support and feedback.
REFERENCES


