ELSEVIER

International Journal of Applied Earth Observation and Geoinformation 7 (2005) 339–359

INTERNATIONAL JOURNAL OF APPLIED EARTH OBSERVATION AND GEOINFORMATION

www.elsevier.com/locate/jag

An automated object-based approach for the multiscale image segmentation of forest scenes

Geoffrey J. Hay [a,*], Guillermo Castilla [b], Michael A. Wulder [c], Jose R. Ruiz [b]

[a] *Department of Geography, University of Calgary, Earth Sciences 356, 2500 University Drive N.W. Calgary, Alta., Canada T2N 1N4*
[b] *IDR – Sección de Teledetección y SIG, Campus Universitario, S/N, 02071 Albacete, Spain*
[c] *Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, 506 West Burnside Road, Victoria, BC, Canada V8Z 1M5*

Received 25 October 2004; accepted 24 June 2005

Abstract

Over the last decade the analysis of Earth observation data has evolved from what were predominantly per-pixel multispectral-based approaches, to the development and application of multiscale object-based methods. To empower users with these emerging object-based approaches, methods need to be intuitive, easy to use, require little user intervention, and provide results closely matching those generated by human interpreters. In an attempt to facilitate this, we present *multiscale object-specific segmentation* (MOSS) as an integrative object-based approach for automatically delineating image-objects (i.e., segments) at multiple scales from a high-spatial resolution remotely sensed forest scene. We further illustrate that these segments cognitively correspond to individual tree crowns, ranging up to forest stands, and describe how such a tool may be used in computer-assisted forest inventory mapping. MOSS is composed of three primary components: *object-specific analysis* (OSA), *object-specific upscaling* (OSU), and a new segmentation algorithm referred to as *size constrained region merging* (SCRM). The rationale for integrating these methods is that the first two provide the third with object-size parameters that otherwise would need to be specified by a user. Analysis is performed on an IKONOS-2 panchromatic image that represents a highly fragmented forested landscape in the Sooke Watershed on southern Vancouver Island, BC, Canada.

Keywords: Object-specific analysis (OSA); Object-specific upscaling (OSU); Size constrained region merging (SCRM); Object-based; Image-objects; Multiscale segmentation; High-resolution; Computer-assisted forest mapping

1. Introduction

Over the last decade there has been a noticeable shift in the analysis of Earth observation (EO) data, from what has been predominantly 30 years of per-pixel multispectral-based approaches (Asner et al., 2003), towards the development and application of multiscale object-based methods (Blaschke et al., 2000; Hay et al., 2001, 2003; Castilla, 2003; Blaschke et al., 2004). *Multiscale* denotes the multiple spatial dimensions at which entities, patterns and processes can be observed and measured. *Object-based* refers to the discretization and attribution of such entities. A key driver in this object-based shift has been the dramatic increase in commercially available high resolution digital remote sensing imagery that is characterized by spatial resolutions 5.0 m and finer (Wulder, 1998; Wulder et al., 2004). An important property of these data is that

* Corresponding author. Tel.: +1 403 220 5584; fax: +1 403 282 6561.
E-mail addresses: gjhay@ucalgary.ca, ghay@sympatico.ca (G.J. Hay).

0303-2434/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jag.2005.06.005

for many applications, they are also H-resolution (Hay et al., 1996). H-resolution or *H-res* refers to a scene model where image-objects are composed of many individual pixels (Strahler et al., 1986). *Image-objects* are groups of connected pixels having similar digital numbers (DN) that visually represent objects in an image that may correspond to real-world entities. Thus, when a human interpreter views a high-resolution image, she typically sees it filled with groups of pixels with meaning in the real world (Schneider and Steinwender, 1999). A key characteristic of image-objects lies in the contextual (i.e., neighbourhood) information imparted by pixel groups, which is highly dependent upon the spatial resolution at which they are viewed. Additional motives towards the development of object-based approaches include:

- Recognition that individual pixels are not true geographical objects. Instead, they are cells of an arbitrarily imposed grid whose boundaries lack any real counterpart, and whose content (DN) represents an integration of energy from outside their spatial (pixel) boundary (Fisher, 1997; Atkinson, 2004).
- An appreciation that humans do not view a world composed of individual pixels, but as a continuum of discrete objects, whose size, shape, spatial arrangement and context change(s) depending upon the scale(s) at which they are assessed (Marceau, 1999; Marceau and Hay, 1999).
- Numerous object-oriented tools, techniques, methods and software already exist within the Computer Science (Graham et al., 2001) and Geographic Information Science communities (Longley et al., 2001). When combined with high spatial resolution imagery, and the notion of cognitively intuitive discrete image-objects, new and powerful opportunities exist for the object-oriented data-management, -mining, -querying and -modelling of remote sensing data with richer semantics, and greater integration with vector-based geographic information systems (GIS), than pixel-based approaches can provide (Schiewe et al., 2001).
- Significant advances in powerful low-cost computing, increased access to sophisticated statistical, programming, and image-analysis software, post September 11, 2001 security issues, and the daily generation of terra-bytes of EO data, have provided the impetus for new (automated and semi-automated) tools to analyze and mine such voluminous data, opportunities to develop new geo-information markets from the results, and the promise to meet increasingly sophisticated user demands.
- Our planet, and the ecosystems it sustains are *complex systems* composed of a large number of spatially heterogeneous components that interact in a non-linear way and exhibit emergence, self-organization and adaptive properties through time (Waldrop, 1992; Prigogine, 1997; Kay and Regier, 2000; Wu and Marceau, 2002). By their very nature, such systems necessitate a multiscale approach in their analysis, monitoring, modelling and management (Hay et al., 1997, 2002). Furthermore, existing ecological theories such as the Hierarchical Patch Dynamics Paradigm (HPDP) (Wu, 1999) provide conceptual frameworks for guiding and explaining the hierarchical and multiscale structure of landscapes that can be operationalized within an object-based approach (Hay et al., 2001, 2003; Castilla, 2003).
- Object-based approaches represent viable solutions to the *modifiable areal unit problem* (MAUP). Earth-observation data represent the primary data source for large area environmental mapping and updating; however, they also represent a special case of the MAUP (Marceau et al., 1994). The MAUP originates from the use of arbitrarily defined and modifiable spatial units used to acquire data over a geographical area (Openshaw, 1984). Because these data do not explicitly correspond to geographical entities, but rather are an aggregation of the content of the spatial units, the value of the analysis results based upon them may not possess any validity independently of the units that are used. One way to overcome the MAUP is to focus the analysis on meaningful geographical entities (i.e., image-objects) rather than arbitrary defined spatial units (such as individual pixels) (Fotheringham and Wong, 1991; Hay et al., 2001).

In general, *object-based approaches* refer to image-processing techniques that when applied either result in the segmentation (i.e., partitioning) of an image into discrete non-overlapping units based on specific criteria, or are applied to define specific multiscale characteristics—from which segmentation may then be based (Hay et al., 2002; Hall and Hay, 2003). After segmentation, attributes can be assigned allowing for class designation (Schneider and Steinwender, 1999), and topological relationships between segments (Burnett and Blaschke, 2003). *Segments* refer to segmented image-objects. Blaschke et al. (2004) provide an overview of numerous segmentation techniques used in remote sensing, though in theory, any image segmentation algorithm could be used to partition a remote sensing scene (Zhang, 1996). However, the real

challenge is to define appropriate segmentation parameters (typically based on spectral homogeneity, size, or both) for the varying sized, shaped, and spatially distributed image-objects composing a scene, so that segments can be generated that satisfy user requirements. Simply because a scene can be segmented, does not imply that the resulting segments are visually meaningful, or correspond to entities of ecological or management interests. This issue is exemplified by the fact that the meaning(s) and dimensions of segments change depending upon their scale of analysis. Furthermore, no single (spatial) scale is optimal for characterizing the multitude of different scene components, hence, the need for *multiscale object-based methods* (Hay et al., 1997; Baatz and Schape, 2000; Hay and Marceau, 2004).

A necessary condition for the acceptance and adoption of object-based approaches is the development of image-analysis methods that closely match (and ideally exceed) the results of those produced by the innate multiscale object-processing capabilities of human interpreters. For instance, in the domain of biomedical imaging, the automated detection and delineation of image-objects (i.e., organs and tumours) is already considered operational (Dynapix, 2004). In Earth observation, a number of promising methods have also been explored including object-specific forest-texture mapping (Hay and Niemann, 1994; Hay et al., 1996), individual tree crown classification (Gougeon, 1995), object-specific multiscale analysis and upscaling (Hay et al., 1997, 2001; Hay and Marceau, 2004), object-based change detection (Hall and Hay, 2003), object-based lidar models for forest inventory update (Wulder and Seemann, 2003), multiscale object-based land-cover mapping (Hall et al., 2004), scale-space sedimentary analysis (Stewart et al., 2004), multiscale watershed analysis (Steinhardt and Volk, 2003), and tree-crown-based forest stand delineation (Leckie et al., 2003). However, the majority of these (and other studies prior to year 2000) are typically conducted with software developed in-house for research purposes.

In 2000, eCognition became the first (non-research focused) commercially available software for multiresolution segmentation and object-oriented fuzzy-rule classification, specially suited for high-resolution imagery (Burnett and Blaschke, 2003; Van der Sande et al., 2003; Benz et al., 2004). Segmentation follows a proprietary bottom-up region merging technique (Baatz and Schape, 2000) starting with one-pixel objects, which are iteratively merged into larger objects based primarily on a user defined *scale parameter*. In previous research, Blaschke and Hay (2001) attempted to use pre-defined object-specific spatial measures to automate, and guide the selection of this scale parameter, but were largely unsuccessful, as there is no recognizable relationship between the scale parameter in *eCognition* (which is unitless) and spatial measures (i.e., area) specific to the image-objects composing a scene. Consequently, users have to find useful segmentation levels on a trial and error basis (Blaschke and Hay, 2001). While *eCognition* does provide a commercially available toolset that incorporates interesting segmentation, topological and semantic components, the setting of an important 'scale parameter' largely responsible for object segmentation that is not intuitively linked to a specific spatial scale or object size, nor to an associated ecological framework, represents serious limitations for some users (Hay et al., 2003).

The primary objective of this paper is to present *multiscale object-specific segmentation* (MOSS) as an integrative object-based approach for *automatically* segmenting meaningful forest-objects at multiple scales from a high-spatial resolution EO scene. Segmentation is based on spatial measures explicitly related to the varying sized, shaped, and spatially distributed *image-objects* that compose a scene rather than on arbitrarily defined scale parameters.

MOSS is an integrative three-part methodology (Section 3.1) that builds upon prior research (Section 3.2) by incorporating components of *object-specific analysis* (OSA; Section 3.2.1) and *object-specific upscaling* (OSU; Section 3.2.2), with a new segmentation sequence referred to as *size constrained region merging* (SCRM; Section 3.3). This sequence includes a new image smoothing algorithm (Section 3.4) and edge detection method (Section 3.5), a watershed transform (Section 3.6), and new region-merging (Section 3.7) and vectorization components (Section 3.8). Once presented, we show how object-specific histogram information may be used to automatically define dominant image classes (Section 4.1), and illustrate how MOSS can be used to automatically delineate *image-objects* at multiple scales (Section 4.2) that correspond to a range of objects from individual tree crowns, to forest stands (Section 4.3). We then discuss how such a tool could be used for computer-assisted forest inventory mapping (Section 5.1), and describe how SCRM parameters can be extended with a simple model, to allow for larger area analysis (Section 5.2). We then summarize our findings (Section 6).

We note that previous research using OSA and OSU on multispectral IKONOS-2 (4 m) data (Hall et al., 2004), and (*user-defined*) SCRM on multispectral Thematic Mapper (30 m) data (Castilla, 2003, 2004;

Castilla et al., 2004) have shown favourable results. However, in this communication, we present these three components as an integrated automated approach and restrict our analysis to a single channel H-res panchromatic IKONOS-2 scene, which is considered analogous to the digitized air photos used for Canadian forest inventory purposes.

2. Data and field site

While much of the information needed to generate forest inventory datasets in Canada are still derived from ortho-correct panchromatic airborne imagery, the use of high-resolution satellite imagery is increasingly available, and becoming more commonly used for large area forest inventories. In this research, analysis is performed on a 1000 × 1000 pixel sub-image extracted from an IKONOS-2 panchromatic scene (1.0 m spatial resolution), that is representative of the digitized air photos used for Canadian forest inventories.

This 100 ha sub-image of Rithet Creek (acquired on July 17, 2000) represents a highly fragmented forested landscape located in the Sooke Watershed, on southern Vancouver Island, BC, Canada (Fig. 1). In this area, the very dry maritime Coastal Western Hemlock biogeoclimatic subzone dominates, though a small component of moist maritime Coastal Douglas-Fir subzone also exists. Due to its longevity, and the area history of repeated fire and windthrow disturbances, successional processes on any site unit rarely proceed to a climax forest; as a consequence Coastal Douglas-Fir [*Pseudotsuga menziesii* (Mirb.) Franco var. *menziesii*] is the dominant seral tree species (GVWD, 1993).

Although detailed forest inventory maps (GVWD, 1993) for this sub-area distinguishes between 28 different forest polygons, three principal stand types are readily visible in the image, each of which illustrates the dominant seral tree species (Coastal Douglas-Fir) (Fig. 2). Throughout this paper, they will be referred to as *mature*, *young*, and *immature*. The mature stands are

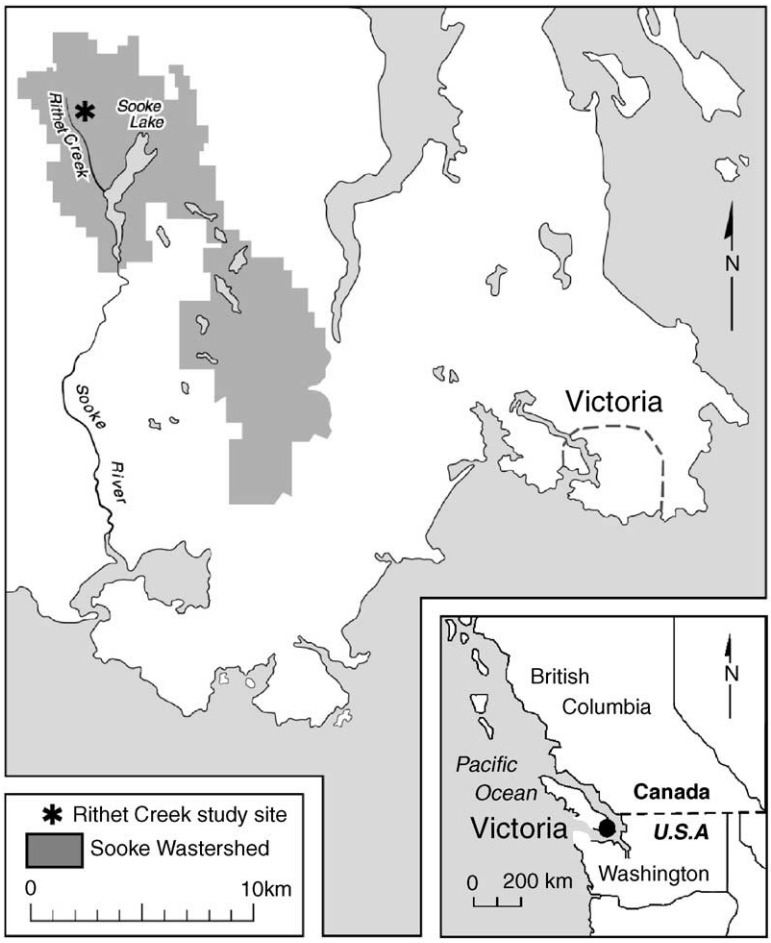

Fig. 1. Scene map.

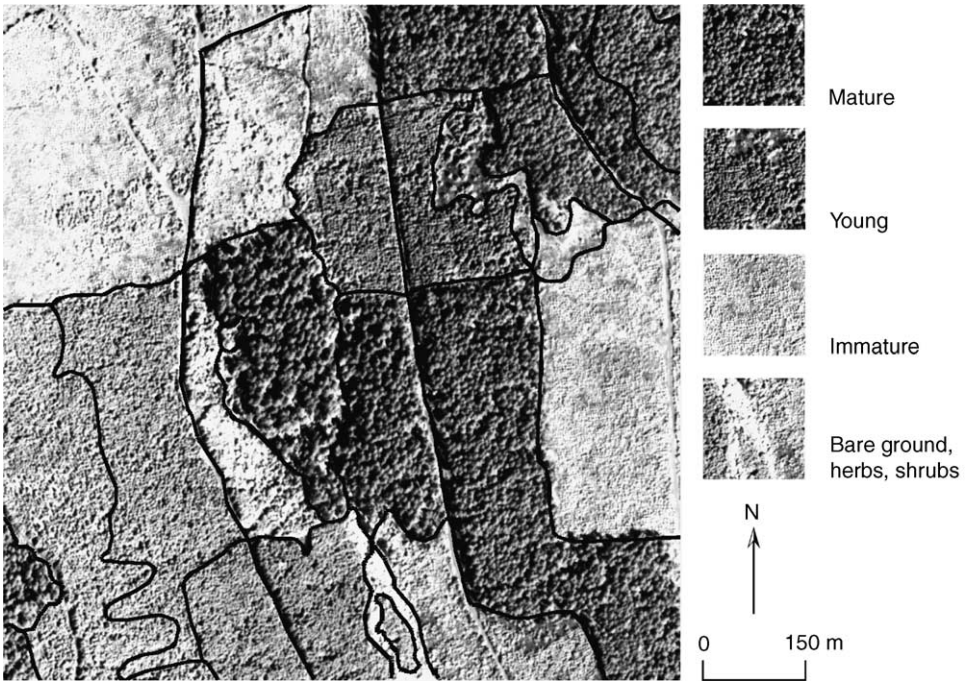

Fig. 2. IKONOS study site, forest polygons overlays, and four primary cover types.

represented by a dark coarse image-texture, and are located at image center, and top- and bottom-right image quadrants. They consist of trees approximately 141–250 years old, 55–65 m tall, with a crown diameter of 9–15 m and a crown-closure of 56–65%. Young stands (visually displayed in medium coarse toned textures are located north of image center and mid-lower left) range in age from 35 to 40 years, with heights of 19–28 m, a crown diameter of 5–7 m and a crown closure of 76–85%. The immature stands (represented by relatively smooth toned textures) are located principally at the top left and mid-right portions of the image. They range in age from 12 to 19 years, with heights of 0–10 m, a crown diameter of 2–5 m and a crown closure of 0–45%. There is also a mixed class that includes bare ground (bright tones), and herbaceous and shrub layers (in smooth light grey tones/textures). Three main gravel roads (linear features) vertically segment the scene, with one road (in the upper left image quadrant) bisecting into two smaller 'Y'-shaped roads.

3. Methods

3.1. MOSS: an overview

Multiscale object-specific segmentation represents an integration of three methods for automatically segmenting a remote sensing image into a meaningful object-based hierarchical representation that corresponds cognitively to visual interpretation. These methods are: object-specific analysis, object-specific upscaling, and size constrained region merging.[1] In general terms, OSA is a multiscale approach that employs varying sized adaptive kernels to automatically define unique spatial measures specific to the individual image-objects composing a remote sensing scene (Hay et al., 1997, 2001, 2003). This results in the generation of three related images, the *variance*, *area*, and *mean* images. Object-specific spatial measures from the area image are then used in a weighting function to automatically upscale (OSU) the mean image to a coarser spatial resolution (Fig. 3). OSA and OSU are then iteratively applied to the newly generated upscaled (mean) images, resulting in a multiscale hierarchy of variance, area, and mean images. Spatial statistics are automatically extracted from each of the area images and used in SCRM to automatically segment the original image into a hierarchy of topologically discrete multiscale segments that correspond to objects typically

[1] Unless explicitly stated, Dr. G.J. Hay wrote all object-specific code in IDL v 5.2-6.0 (interactive data language) and all SCRM code was written by Dr. G. Castilla in IDL 6.0 and ENVI 4.0 (environment for visualizing images, http://www.rsinc.com/idl).

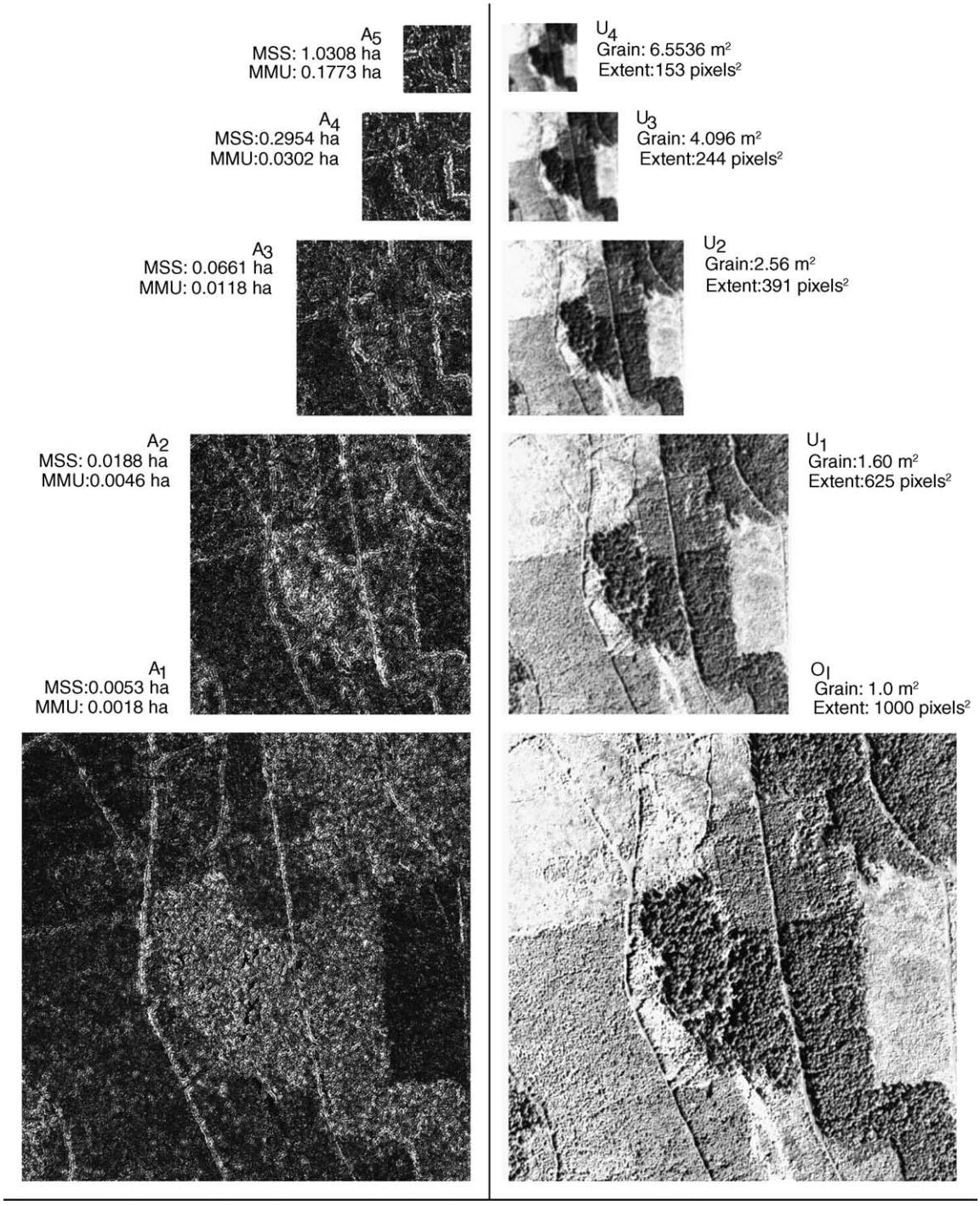

Fig. 3. Hierarchy of area (A_{1-5}) and upscale images (U_{1-4}) compared to the original image (O_I).

defined using visual interpretation. SCRM is a segmentation sequence comprised of image smoothing, a gradient watershed transform and region merging. The following sections provide a more detailed explanation on all of these procedures.

3.2. Object-specific analysis and object-specific upscaling: an overview

The underlying goal behind OSA/OSU is to provide an object-based mechanism (model) that allows for

spatially dominant image-objects to evolve through scale in an attempt to understand process morphology from the resulting multiscale patterns. As such, an underlying premise of OSA is that all pixels within an image are considered high spatial resolution samples (Strahler et al., 1986) of the image-objects they model, even though, both high- and low-spatial resolution (L-res) samples exist in a single image. *L-res* refers to instances where pixels have a larger spatial extent than the realworld-objects they model; thus, a single pixel represents an integration of many smaller image-objects. Recall that *H-res* refers to situations where pixels have a smaller spatial extent than realworld-objects; thus, a single image-object is composed of many individual pixels. Therefore, in OSA, individual pixels (the fundamental image primitive) are used to define the larger image-object(s) they are a part of. A brief overview of OSA and OSU methodology follows; however, detailed accounts have previously been defined (see Hay et al., 2001; Hall and Hay, 2003; Hay and Marceau, 2004).

3.2.1. OSA

Hay et al. (1997) noted that when plotting the variance of digital values generated by sampling image-objects within increasingly larger kernels, the resulting plots produced curves with distinct breaks, or 'thresholds' in variance as the analyzing kernel contacted the image-object's edges. Evaluation of the shape of these variance curves (Hay et al., 1997, 2001, 2003) enabled the creation of a set of robust heuristics that define the spatial extent (i.e., kernel area) where an individual pixel is spectrally related to the (larger) image-object it is a part of. This object-specific spatial measure is critical for further analysis. Essentially, it can be considered as a measure of the maximum extent of the spatial autocorrelation between a pixel and its neighbours.

The primary OSA heuristic is composed of three different percentage values, each of which represents the difference in variance defined between two concurrent kernels (one smaller than the other) over a specific range of kernel sizes. If the difference in variance between the two kernels is less than or equal to the heuristic threshold value, processing is stopped. The corresponding *mean*, *variance*, and *area* values are then recorded for the pixel under analysis within the specified kernel. This dynamic process is then applied to the remaining pixels within the original image (O_I), resulting in the generation of corresponding variance (V_I), area (A_I), and mean (M_I) images. These three images are referred to as the first *image-set* and this type of adaptive-kernel processing is referred to as *object-specific analysis*. These object-specific data can then be used to guide an appropriate segmentation algorithm as outlined in this paper (Section 3.3) or as previously defined using *marker controlled watershed segmentation* (Hall and Hay, 2003; Hall et al., 2004; Hay and Marceau, 2004). The *variance image* is essentially a *gradient* or *edge* image. The *area image* defines the spatial influence, i.e., the kernel size or number of pixels 'spectrally related' to the pixel under analysis. The *mean image* is composed of an average of the H-res pixels that constitute part of individual objects assessed at their respective scales. In each of the area images illustrated in Fig. 3, dark tones indicate that a small object-specific kernel was defined, while bright values indicate that a large kernel was assessed.

3.2.2. OSU

To evaluate how image-objects evolve through scale, the first mean image is resampled to a coarser resolution (i.e., grain) by applying an object-specific weighted algorithm (OSU), and the OSA process is iterated on the new upscaled image (U_1) (Fig. 3). The upscale pixel size can either be user defined, or automatically defined based on a one-fourth resampling heuristic (Slater, 1980) in relation to the smallest analyzing kernel and the spatial resolution of the image (Hay et al., 2001). When defined automatically, this results in an upscaling heuristic of 1.6 (see Hay and Marceau, 2004). That is, each pixel in the first upscale image has a grain equal to 1.6 pixels in the original image (O_I). Thus, the extent of the new upscale image is obtained by dividing the length of the original image (i.e., 1000 pixels) by 1.6, resulting in 625 pixels (in both *x*- and *y*-directions) (Fig. 3)

For upscaling to occur, the upscale kernel is used as a mask to generate a weighted area value for each pixel as follows. Beginning at the origin, the upscale kernel is overlaid on the corresponding A_I, and the DN of each area pixel (within the mask) is divided by the sum of all area pixels in the mask. This generates a fractional area weight that sums to one. Each area weight (in the mask) is multiplied by its corresponding original grey value, and then summed. This summed value represents the new area weighted upscale value that corresponds to the original pixels in the upscale mask. The non-overlapping upscale kernel is then applied to the remaining image resulting in a new upscale image (U_1). To determine the upscale spatial resolution, this process is iterated by dividing the length of the newly upscaled mean image (U_1) by 1.6. Thus, at the next iteration, the image size is $(625/1.6) = 391$ pixels^2 with a spatial

resolution of (1000/391) = 2.56 pixels^2. For more detail on OSU see Hay and Marceau (2004). In the illustrated example (Fig. 3), we present four upscale images and the five corresponding area images (A_{1-5}), as beyond this number, the images become too small for visual assessment.

3.3. Size constrained region merging: an overview

Size constrained region merging (Castilla, 2003) represents a segmentation sequence that may be summarized as follows. A geocoded input image (Fig. 4a) is filtered with *gradient inverse weighted edge preserving smoothing* (GIWEPS; Castilla, 2003) to remove superfluous gradient minima represented by coarse image-textures. The output of this newly developed filter (GIWEPS) is an almost piecewise constant image, in which each uniform region represents the area of influence of a gradient minimum (Fig. 4b). A gradient magnitude image (i.e., an edge image) is then computed (Fig. 4c), and the output image is searched for local minima. The area of influence of each minimum is contoured and labelled with a watershed algorithm (Fig. 4d). Then the resulting regions are iteratively merged by increasing dissimilarity until they exceed the size [expressed in working pixel units (WPS)] of the minimum mapping unit (MMU), and are close to the mean segment size (MSS) specified from OSA/OSU statistics. Finally, a labelled image, composed of merged 'watersheds' that cogni-

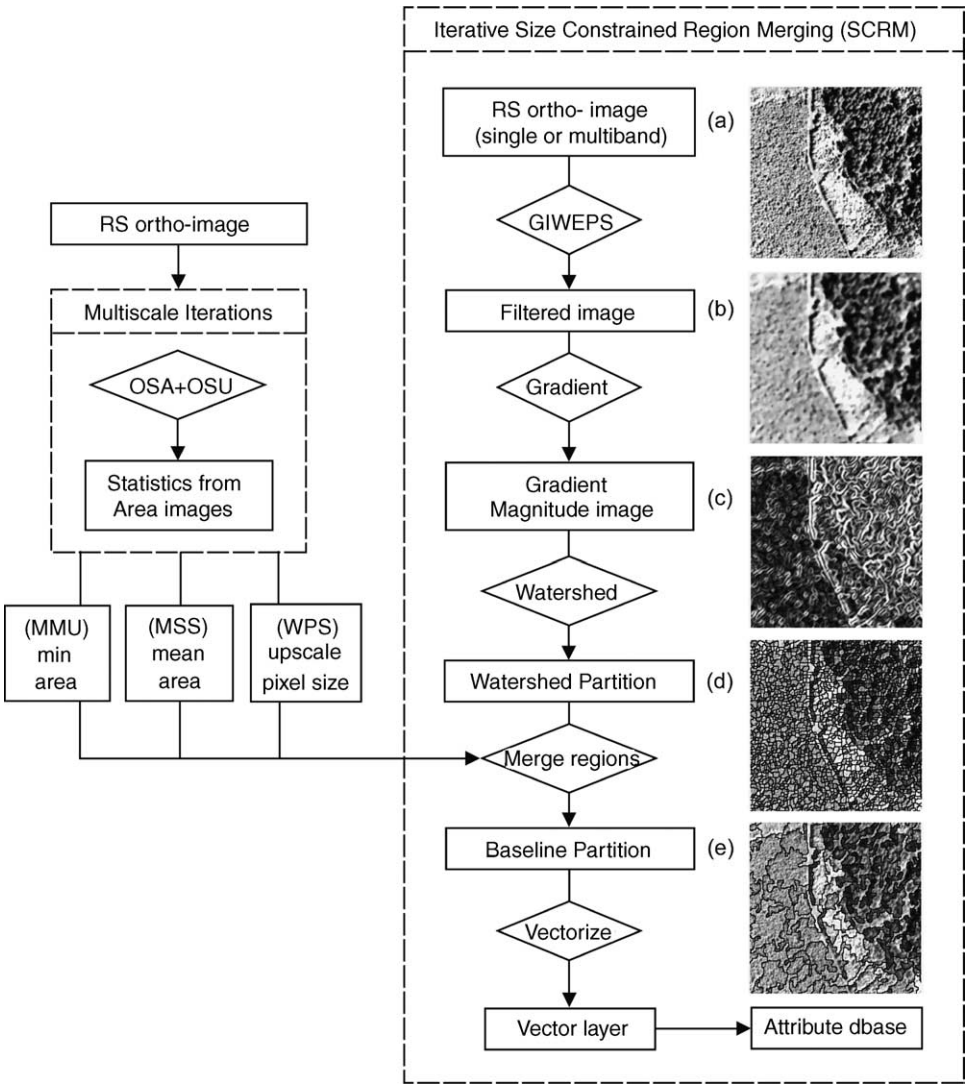

Fig. 4. Integrated OSA/OSU and SCRM—workflow.

tively represent individual image-objects, is converted into a vector layer (Fig. 4e), and attributes are optionally compiled.

3.3.1. SCRM background

The SCRM algorithm builds upon a number of ideas from previous research. The first is the stepwise optimisation algorithm of Beaulieu and Goldberg (1989). It begins by considering single pixels as the initial segments. At each iteration, the two adjacent segments that show the highest degree of fitting are merged. Thus, the candidate pair merged at each pass is the one that produces the least increment in heterogeneity. Segments are merged gradually in this way until there is no candidate pair below a user-defined threshold. The final partition is considered *optimal* regarding the minimization of the heterogeneity criterion, but the procedure is slow as it allows only one merge per pass (a strategy known as *global mutual best fitting*). Based on this work, Woodcock and Harward (1987) introduced a faster algorithm that allowed multiple merges per pass and included size constraints. As these authors noted, the global threshold alone led to inadequate results, since there was often a great disparity in size of the output regions. Areas marked by coarse texture will consist of many small regions (often individual pixels), whereas smooth uniform areas are segmented into large regions. Therefore, they supplemented their algorithm with size constraints that prevented excessive growing in smooth areas and forced the development of segments exceeding the minimum required size in areas with high local variance.

Another way of tackling the uneven growth of segments between areas of smooth and coarse texture is to enable multiple merges per pass, by distributing the candidate pairs to be merged at each iteration as far as possible away from each other in the image. This is the strategy followed by the segmentation algorithm embedded in eCognition (Baatz and Schape, 2000). In this way, it achieves a uniform growth of segments throughout the image, so that the final segments all have a similar size. Since a conservative (small) threshold permits fewer merges than a greater one, the mean size of segments will grow with the value of the threshold.

A major pitfall of these three methods is that they start the merging sequence with individual pixels. Apart from being computationally expensive, pixels are artificial units, without an explicit geographic reality (Fisher, 1997). A way to tackle this incongruity is to start the region merging process with *blobs* instead of single pixels (Castilla, 2003). Blobs are tiny homogeneous regions, darker, brighter or of different hue than their surroundings (Marr, 1982; Lindeberg, 1994). Therefore, just as pixels are the building blocks of an image, here, *blobs* are the perceptual primitive counterparts (Castilla, 2003). An effective method to contour blobs, assuming they correspond to catchment basins of gradient minima, is the watershed transform (Vincent and Soille, 1991), the principal morphological segmentation method. We note, a variation referred to as *marker-controlled watershed segmentation* (MCS; Meyer and Beucher, 1990) – where segmentation occurs only at user defined (i.e., marked) locations – has been successfully used for object-based multispectral land cover segmentation and classification (Hall and Hay, 2003; Hall et al., 2004). However, MCS is not used in this study since SCRM replaces the need of choosing the markers with the filtering of the input image.

The use of a segmentation sequence consisting of image smoothing and/or gradient magnitude simplification, watershed transform, and region merging, has also been reported previously (Ji and Park, 1996; Fjørtoft et al., 1998; Haris et al., 1998; Weickert, 1998). However, none of these studies were related to land cover mapping. Furthermore, except for the watershed transform, the algorithms proposed here are new, and our method is conceptually consistent with object-oriented analysis, an asset that many segmentation algorithms lack.

3.3.2. Applying SCRM

In order to use SCRM (Fig. 4), three spatial parameters (two of which are based on object-size) must be specified:

(i) the working pixel size (WPS);
(ii) the minimum mapping unit (MMU);
(iii) the mean segment size (MSS).

These can be defined either by a user, or automatically extracted from the area images generated at each OSA/OSU iteration (see Section 4.2 for details). WPS represents the spatial resolution (i.e., grain) of the image, and is defined as the length of the side of the original pixel. It also corresponds to the desired spatial accuracy of boundaries to be generated, since the mean distance between vertices in the output vector layer is roughly double this length. MMU represents the minimum size required for segments, which is given in original pixel (areal) units. From an object-specific perspective, MMU is represented by the minimum DN of the associated area image(s). MSS corresponds to the

desired average size of output segments. This can be defined by finding the mean of the corresponding area images. Once specified, the next step involves filtering the input image (Fig. 4a).

3.4. Image smoothing

An intrinsic characteristic of remotely sensed images (at any spatial resolution) is *image-texture*, the local variation of DNs within an image-object, and between different image-objects resulting from the relationship between sensor spatial resolution, and the size, shape, spatial configuration, and spectral characteristics of the scene components being viewed (Hay and Niemann, 1994; Hay et al., 1996). If a gradient magnitude image (i.e., an 'edge' image) is computed without *pre-conditioning* of the input image (O_I), the computation will result in an intricate structure full of edges and local minima, especially in areas with coarse image-texture. Therefore, the structure of the image has to be simplified so that gradient minima resulting from coarse image-textures are removed. Coarse texture can be interpreted as being caused by objects with a recurrent pattern of distribution that cannot be resolved by the sensor. Therefore, they should be grouped into larger image-objects that can only be delineated after having suppressed the effects of the smaller objects. We note that the suppression of coarse texture in order to obtain a first delineation of the scene does not preclude the ability to define a specific texture measure at a latter stage, as segment statistics are not computed from the DNs of the smoothed image, but from the original one. Ideally, simplification should be performed only upon the unresolved elements of the scene that generate image-texture, leaving untouched the elements corresponding to edges. Such a process is commonly called edge preserving smoothing (EPS) (Abramson and Schowengerdt, 1993).

Unlike conventional smoothing techniques such as simple averaging or Gaussian filtering, EPS filters adapt to the structure of the image, so that the local operator is different at each position. A simple example is the *median filter*. In GIWEPS, the new digital number of a given pixel is the weighted mean of the DNs of its eight neighbours. The weight of each neighbour is proportional to its similarity to that pixel, and the degree of proportionality is governed by a diffusivity parameter. The filter is applied iteratively, up to a point where change between consecutive output images is negligible (Fig. 4b). We note that output images beyond this point still resemble the original image over thousands of iterations (Castilla, 2003).

3.5. Gradient magnitude image

If one considers a given grey-level image as a digital elevation model (DEM), then the gradient magnitude image is the slope map that corresponds to that DEM. Thus, at each pixel of a grey-level image, the gradient magnitude is the slope of the steepest descent line crossing that pixel. The dissimilarity measure used here is the Euclidean distance between points (i.e., pixel signatures) in the feature space (Castilla, 2004). Thus, gradient minima are those pixels whose gradient is lower than that of their eight neighbours in the gradient magnitude image (Fig. 4c). In the unusual case of plateaus (regions with equally low-valued pixels), there are no proper minima, and the centroid of the plateau is selected as a local minimum representing the region. If we consider gradient minima as *perceptual attractors*, then their basins of attraction are the primal regions building the spatial structure of the image. By adopting this analogy, the area of influence of each gradient minimum may be viewed as a blob in the filtered image. Though beyond the scope of this paper to further explore, this analogy is the foundation of *semiophysic* image analysis (Thom, 1988), as it enables a correspondence between structural–functional units in an image and the imaged landscape (see Castilla, 2003 for details).

3.6. Watershed partition

The topographic concept of *watershed* was first introduced to the field of image analysis by Beucher and Lantuéjoul (1979), and implemented into an efficient algorithm by Vincent and Soille (1991). The basic idea is to consider the gradient magnitude image (Section 3.5) not as a slope map but as a DEM in itself, with the goal being to find the drainage divides, or watersheds, of that virtual territory. Watersheds define a network of ridges that enclose the dales, or catchment basins, where each drop of rain would drain. If the gradient magnitude image (Fig. 4c) were visualized as a DEM representing a lunar-scape, full of craters (dark areas) with ridges (bright areas) of different heights, then each crater would correspond to a blob in the filtered image.

Conceptually, the applied watershed transform simulates a gradual immersion of a surface. Imagine that crater bottoms (i.e., gradient minima) are springs where pressurized underground water upwells. As this process occurs, water will begin to flood areas adjacent to each spring. Further suppose that the flow at each spring is such that the altitude of the water plane of the submersed areas is the same for the entire image-scape

(hence the analogy with immersion rather than flooding). Now, in places where the water flowing from two different crater bottoms would merge, we build a dam of 1-pixel thickness, slightly taller than the highest crater of the scene. When the latter is completely submersed, we stop the immersion. The resulting dams are the watersheds of the pseudo-topographic scene, which in turn define a complete *partition* of the image. In the output partition, watershed pixels are set to zero (Fig. 4d), whereas non-zero pixels have as DN the numeric label of the segment to which they belong.

3.7. Region merging

In this step, the regions of the watershed partition (i.e., blobs) are aggregated into segments until all regions in the partition are larger than the specified minimum size (MMU), and the merging sequence is such that the homogeneity of the resulting regions is maximal given the size constraint (MMU and MSS). We note that many region-merging algorithms include a size constraint (Woodcock and Harward, 1987; Hagner, 1990; Baraldi and Parmiggiani, 1994); however, in these examples thresholds are set on a *dissimilarity* measure, which is typically a non-intuitive input parameter based on *spectral* magnitudes that can only be defined meaningfully after numerous iterations. The input parameters defined by SCRM are directly related to the size (i.e., area) of the image-objects in the scene as defined either by a user, or by iterative OSA/OSU.

In SCRM, the signature of a segment is the mean value of the pixels belonging to it. Therefore, the signature of a new segment is the weighted (by size) mean of the signatures of the two merged segments. In this way, segment signatures are computed from the original image only once, at the beginning of the merging procedure. The same can be said about the *adjacency table* (an array returning the list of neighbours of any given segment), which is first computed from the watershed partition and then updated using Boolean algebra. From this adjacency table (AT) and the *signature list* (SL), the identification of the most similar neighbour (MSN) to each segment is trivial (Castilla, 2004).

The utility of table-based updating of both signatures and adjacency is that it is very fast, and it enables the use of the global mutual best fitting strategy (see Section 3.3.1), which is computationally slow. Consequently, in each iteration only a candidate pair – for which the dissimilarity criterion is minimal – is merged. Next, the AT, SL, and MSN arrays are updated, and a new iteration proceeds. This process continues until the *sum* of the number of segments currently larger than MMU ($N_{\text{area}>\text{MMU}}$), *plus* the result of dividing by MSS the area currently occupied by segments smaller than MMU ($S_{\text{area}<\text{MMU}}$), is *less than* the result of *dividing* the area of the image (I_{area}) by MSS:

$$N_{\text{area}>\text{MMU}} + \frac{S_{\text{area}<\text{MMU}}}{\text{MSS}} < \frac{I_{\text{area}}}{\text{MSS}} \quad (1)$$

Eq. (1) represents a partial stop criterion that guarantees that the final mean size of segments (MSS) will be close to the desired size. Once met, the segment size is taken into account in the merging procedure, so that the best fitting pair is allowed to merge only if at least one of both segments is smaller than the MMU. In this way, homogeneous regions are formed first, and then dissimilar regions smaller than MMU are progressively incorporated to the former until all segments are larger than MMU.

The actual merging of two segments involves replacing the label of one with the label of the other in the final label list (FLL). FLL is an array of length equal to the number of segments (i.e., blobs) in the initial watershed partition. At any point during the merging process, there is a link that keeps track of blobs composing each segment, so that FLL can be easily updated. Once the merging is completed, a new raster layer (referred to as the *baseline image*) is created from the watershed partition by replacing the DN of pixels inside each blob with the new label registered in the corresponding position of FLL. Finally, watershed pixels lying in the interior of final segments are filled with the numeric label of the corresponding segment.

3.8. Vectorization

Once the baseline image has been generated, an optional last step is to convert it into a vector layer. The centres of boundary pixels (zero-valued) are considered the initial vertices forming the arcs. Note that this is analogous to considering boundary pixels as a transition zone between patches that can be represented by its medial axis. The nodes (i.e., junctions between arcs) are identified, so that each polygon is defined by the set of arcs bounding the corresponding segment. In order to give a smooth appearance to arcs (Fig. 4e), a spline interpolation is applied to the centroid of each of three consecutive vertices. Once the image is vectorized, a database can then be generated with associated attribute information (i.e., radiometric and spatial) for each segment. These vectors can also optionally be used to generate associated attribute raster layers. The resulting

(vector) layers may then be used in a GIS as an efficient method for automatically defining training and or sampling 'regions of interest' for supervised classifications, for object-based analysis and or classification (Devereux et al., 2004), and as a guiding template for computer-assisted photointerpretation. This last use is discussed further in Section 5.1.

4. Results

4.1. Object-specific histograms and scene structure

The evolution of image-objects captured in $U_{1–4}$ histograms, reveal a shift through scale from single, to multi-modal peaks, which in turn visually correspond to dominant object classes in the landscape. Thus, without a priori information, it may be possible to use the multi-modal histogram information generated from the family of object-specific upscale images ($U_{1–4}$) to automatically define 'meaningful' image classes that correspond to the visually dominant image-objects in the scene (Fig. 5). To illustrate that this form of histogram information is only available with the use of object-specific analysis, we evaluated three non-object-based resampling strategies in order to examine the shape of their resulting histograms through scale.

Evaluation was conducted by upscaling (i.e., resampling) the original image to four different spatial resolutions using nearest neighbour, bilinear, and *bicubic* interpolation that spatially corresponded to the hierarchy of images with the same x and y dimensions as $U_{1–4}$ (Fig. 3). Resampling was conducted in Matlab using the *imresize* function.[2] We note that visually similar histogram results were obtained using each of the three resampling algorithms for each set of four upscale images; however, the images resulting from bicubic interpolation were perceptually closer to those in Fig. 3. In addition, due to the strong similarity in histogram shape between each of the histograms generated with bicubic resampling, we have only selected $O_{I–bicubic\,(U_4)}$ for a comparison with the four object-specific histograms ($U_{1–4}$) illustrated in Fig. 5. $O_{I–bicubic\,(U_4)}$ results from applying bicubic resampling to the original image (O_I) with the same x and y dimensions as U_4. Though the number of samples between O_I and $O_{I–bicubic\,(U_4)}$ is very different, the general shape of their single mode histograms is visually similar, which is in stark difference to the multi-mode histogram of U_4 (Fig. 5).

Although numerous forest polygons and four general forest classes exist in the initial scene (Fig. 2), these classes are all represented in O_I and $O_{I–bicubic\,(U_4)}$ by a single-mode near Gaussian distribution. Due to the shape of these curves we conclude that it would be difficult if not impossible to confidently extract unique class specific information from O_I or $O_{I–bicubic\,(U_4)}$, based exclusively on the shape of their histograms. In contrast, at least four principle classes could be automatically isolated from $U_{3–4}$ images using classic signal processing methods e.g., defining the zero crossings of the histogram curves, or using simple histogram thresholding (i.e., density slicing). We note that DNs around the large peak at 100 visually represent the mature class while those near 160 represent the young class—especially those located in the lower left quadrant of the image (Fig. 2). DNs clustering around 180 represent the immature class, prevalent along the middle right edge of the image as a long vertical rectangle, and those near 220 represent the very bright values corresponding to bare areas in the upper left quadrant of the image.

4.2. Automatically defining WPS, MMU, and MSS from multiscale area images

Since object-specific analysis and upscaling allow spatially dominant image-objects to evolve through scale, object size parameters can be generated for use in SCRM by defining the minimum and average statistics of all pixels in each area image in relation to their corresponding spatial resolution (Fig. 3). Recall that each area DN is an object-specific measure that defines the spatial extent (i.e., the area) of an image-object it (the pixel) is a part of. When these *size* statistics are defined, Table 1 shows that at a spatial resolution or *working pixel size* (WPS) of 1 m^2, the *minimum mapping unit* (MMU) for A_1 corresponds to segments with a minimum area of 0.0018 ha (18 pixels × 1 m^2) and a *mean segment size* (MSS) of 0.0053 ha (53 pixels × 1 m^2). For A_2, the WPS is 1.6 m^2, the MMU is 0.0046 ha, and the MSS is 0.0188 ha. Corresponding WPS, MMU, and MSS values are also defined in Table 1 for $A_{3–5}$. As expected, the general trend of these results shows that larger image-objects exist (in each scene) as spatial resolution coarsens, as evident by the increasing MMU and MSS values.

If the frequency histogram (i.e., DN frequency over DN value) of an Area image were normally distributed,

[2] *Imresize* automatically applies a low-pass filter to the image before interpolation. This filtering reduces the effect of Moiré patterns—ripple patterns that result from aliasing during resampling (MATLAB 6.1, Release 12.1).

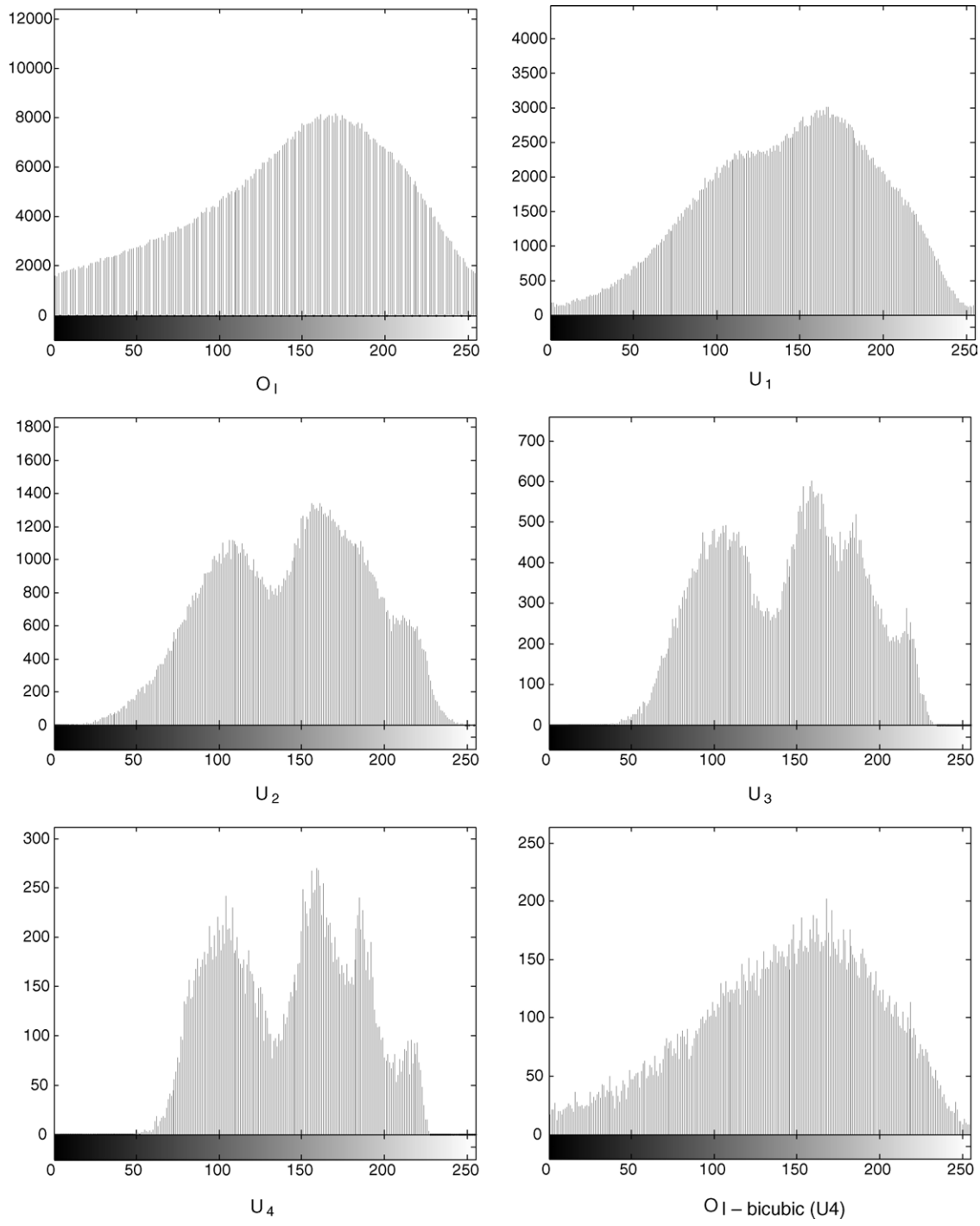

Fig. 5. A comparison of histograms showing how class structure changes through scales with object-specific analysis and upscaling: original image (O_I), upscale images ($U_{1–4}$), and a bicubic resampling of O_I to the same spatial resolution as U_4 ($O_{I-\text{bicubic}(U_4)}$).

the image mean (which is equivalent to the MSS), along with the mode and median would each represent a cumulative distribution of 50% of the total DNs composing the image. That is, the resulting segments would spatially represent 50% of the different sized image-objects constituting the scene at this scale. However, when the frequency histograms of $A_{1–5}$ are assessed in relation to their MSS, we find that the MSS

Table 1
SCRM statistics automatically extracted from area (A_{1-5}) images.

Area images	x, y (p)	WPS (m^2)	MMU (ha)	MSS (ha)	Cumulative %
A_1	1000	1	0.00018	0.0053	66.67
A_2	625	1.6	0.0046	0.0188	56.07
A_3	391	2.56	0.0118	0.0661	73.19
A_4	244	4.096	0.0302	0.2954	88.09
A_5	153	6.554	0.0773	1.0308	94.46

ha, hectares; m, meters; p, pixel units.

Table 2
Vector layer results from five iterations of multiscale SCRM

Vector layers	# Watershed segments	# Merged segments	Merging time (s)	Mean size (ha)	Processing time (s)
VL_1	43969	13265	119	0.007538	161.828
VL_2	20072	2960	27	0.033783	42.922
VL_3	8205	610	6	0.164249	13.687
VL_4	3111	137	1	0.729086	6.109
VL_5	1089	39	>1	2.57797	4.437

for A_1 represents a cumulative value of nearly 67% of the entire image (Table 1, *Cumulative %*). Consequently, this MSS spatially represents nearly 67% of all possible image-objects composing the scene at this WPS. The MSS for A_2, A_3, A_4, and A_5, respectively, represents 56%, 73%, 88%, and 94% of all different sized image-objects that populate each corresponding image (U_{1-4} in Fig. 3). Thus, in A_5, only 6% of the image-objects composing the scene are larger than the segments defined by the MSS of 1.0308 ha (Table 1). This suggests that automatically defined multiscale MSS values constitute highly representative spatial samples (of all possible image-objects composing a scene), regardless of the scale being assessed.

4.3. Automatic generation of five hierarchical layers of multiscale segments

When the WPS, MMU, and MSS information from Table 1, is applied iteratively in the SCRM workflow (Section 3.3), they result in five separate layers of multiscale vectors (VL_{1-5}), which we hypothesize represent the hierarchical evolution of dominant scene structure through scale. Each layer is composed of a different number of segments of different size, shape, and spatial location, with an MSS greater than that specified (compare *mean size* in Table 2, with *MSS* in Table 1).

By applying the A_1 parameters defined automatically by OSA/OSU (Table 1), 43,969 watershed segments were generated in the first SCRM iteration. These were merged to produce a vector layer (VL_1) composed of 13,265 segments with a mean size of 0.0075 ha. In the second SCRM iteration, use of A_2 parameters resulted in 20,072 watershed segments that were merged into 2960 segments comprising vector layer 2 (VL_2). The resulting mean segment size was 0.0337 ha. This procedure was repeated with A_{3-5} parameters resulting in 610, 137, and 39 merged segments for VL_{3-5}, respectively. Total processing time for these three layers ranged from 13.6 to 4.4 s[3] (Table 2).

Due to the large number of total segments generated, it is not feasible to present all possible layers in a single figure. Therefore, in Fig. 6 we present all (39) segments from VL_5 (bold lines) and all (137) segments from VL_4 (thin lines) overlaid on the original image. The white square in the upper right image-quadrant represents the location of sub-images displayed in Fig. 7, where the corresponding hierarchical segments from VL_{1-4} are shown in greater detail.

It is important to recall that specified MSS values represent the average size of the segments to be generated; however, the eventual segment size will vary depending on the spectral and spatial characteristics of the image-objects composing the scene in relation to the MMU and WPS sizes specified. Simply because an average segment size is specified, does not guarantee

[3] Processing was conducted using a Pentium IV, 2.53 GHz workstation, with 1.5 GB of RAM.

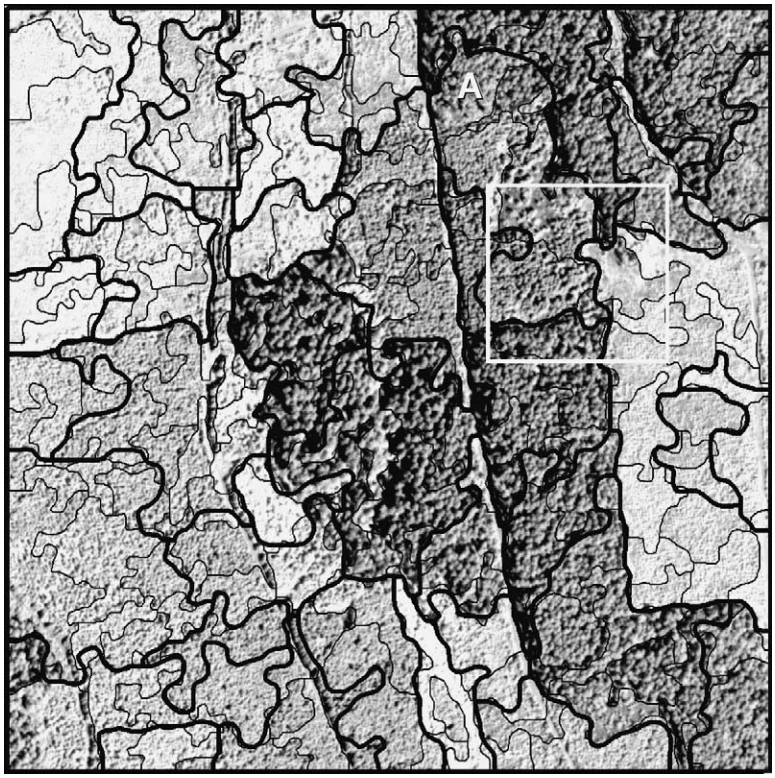

Fig. 6. Segment overlays from vector layer 5 (bold lines) MSS = 1.0308 ha and vector layer 4 (thin lines) MSS = 0.2954 ha.

that image-objects of corresponding size(s) actually exist in the image. This may result when a scene is composed of only very large and very small image-objects, which produce an average object-size that is not physically present in the landscape. For example, in VL_5, the 39 segments have a mean size of 2.57 ha even though the specified MSS was 1.0308 ha, while in VL_4, the 137 segments have a mean size of 0.729 ha, even though the specified MSS was 0.2954 ha (compare results in Tables 1 and 2).

5. Discussion

An important characteristic of MOSS is that it automatically provides a multiscale overview of the dominant structure existing in the scene, based on the scene components themselves, without any user intervention. Visual inspection of vector layers $VL_{1–5}$ (Figs. 6 and 7) reveal details ranging from individual tree crowns, crown shadows, and canopy gaps, to larger vegetation units that correspond to sub-dominant, and dominant stand components. In this section, we illustrate and discuss the potential of MOSS for computer-assisted mapping, and describe how with a simple model it could be extended to evaluate larger landscape components than initially assessed with OSA and OSU.

5.1. Can MOSS provide a computer-assisted solution for forest inventory mapping and information generation?

Manual aerial photo interpretation is the most common approach for generating management forest inventories, of which stand delineation and subsequent attribution are critical components. However, interpretation accuracy, consistency, and timeliness are recurring concerns (Hall, 2003). For example, in 1:10,000 to 1:20,000 scale photography, accuracies are generally 70–85% for main species in a stand, but can be lower (Leckie et al., 2003). In addition, inventory data volumes are often enormous, yet manual interpretation typically completes only 5–15 photos per day. This represents approximately one photo per hour, or 400 new hectares and 10–15 stands per hour (Leckie et al., 1998; data of Leckie and Gillis, 1995). As a result, the photointerpretation process is both time consuming and represents a significant cost component of inventory

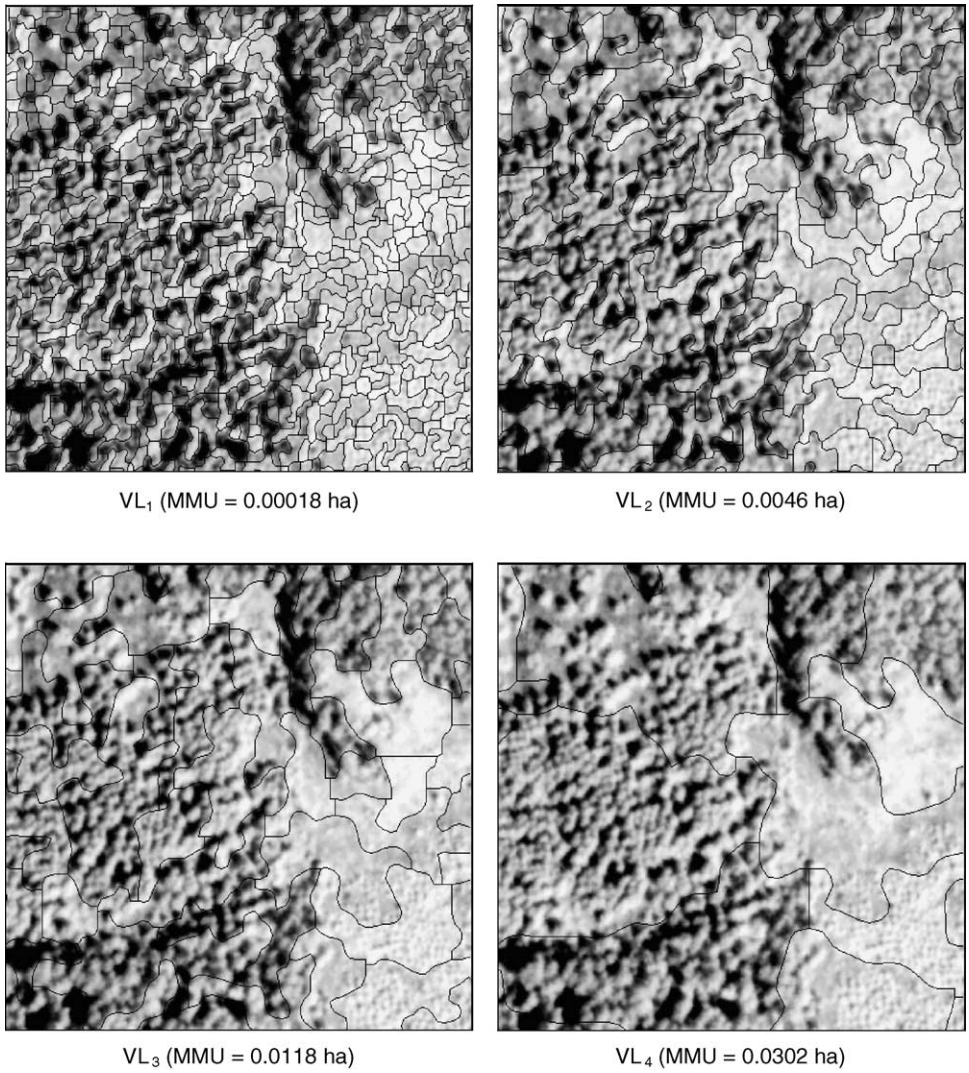

Fig. 7. Sub-images of vector layer overlays ($VL_{1–4}$), showing changing segment detail through scale.

mapping at around 25% (Leckie and Gillis, 1995). Furthermore, new parameters are becoming increasingly important such as stems/ha, stand gap size distribution, number of snags or large trees, crown size distribution, and other environmental strata-based forest conditions (i.e., nesting habitat, change detection, and damage assessments) (Gougeon and Leckie, 2001); which can additionally burden the interpretation process (Hall, 2003). As a result, forest managers are increasingly considering computer-assisted interpretation aids to forest inventory mapping. However, such techniques must fit current infrastructure using 1:10,000 to 1:20,000 scale digitized photography and or high spatial resolution satellite imagery. They must be simple to apply, not require sophisticated or expensive equipment, nor require inordinate fine-tuning or trial and error by the interpreter (Leckie et al., 1998), and above all, they must produce meaningful results. If these requirements can be met, then the incremental inclusion of automated processing to the increasingly digital photo interpretation process will create opportunities to increase interpretation accuracies and efficiencies per unit area, provide new information products, and lower costs.

Forest managers use interpretations of air-photos and high-spatial resolution imagery to provide a generalization of complex forest ecosystems that enable logical groupings (e.g., age, height, crown closure, etc.) upon which management scenarios may be based and implemented. Typically, photo interpreters operate by

nitially delineating the most contrasting units, with a second pass to delineate more subtle units, or units worthy of more intensive characterization. Attribution then proceeds once all spatial units have been delineated (Leckie and Gillis, 1995). Based on this approach, MOSS could be integrated within this interpretation framework by providing the following capabilities:

1. aid interpreters in making object delineation decisions;
2. provide assistance with determination of sub-unit and super-unit delineations.

Since object-based vector layers can be pre-generated, a near real-time overview of the scene structure at multiple scales will be available to the analyst. Once overlaid on the original image, the analyst can immediately begin to visually evaluate and verify boundaries, and their (automatically defined spectral/spatial) attributes. For example, in Fig. 6, visual inspection shows that the VL_5 segments model the larger stand components in the scene with good agreement. However, there are places [see A (upper-middle) in Fig. 6] where components of different vegetation cover (i.e., Mature and Young) have been merged within the same (bold) segment. From the perspective of a guiding template for computer-assisted interpretation, if this merge or portions of it (i.e., the arcs composing this polygon) were not desired, a user could simply overlay finer segments, such as those from $VL_{1–4}$ (Fig. 7) to evaluate the smaller polygons the larger segment is composed of. Thus, an interpreter need not digitise new arcs, she merely selects and erase (e.g., with mouse clicks) the arcs (of different scales) considered most relevant. For example, finer segments from VL_4 (located directly below A in Fig. 6) reveal forest structure that visually appears more accurate than the corresponding Forest polygon boundary defined in Fig. 2. With a simple mouse click these detailed segments can be added to the working vector layer, and components of the undesired (bold) polygon, can simply be deleted. However, this first implies that the segments correspond to features of interest for the analyst.

5.1.1. Additional forest applications

Since DN heterogeneity is lower within a segment, than between segments, MOSS scenes could also be used as part of a stratified sampling program. In such cases, users need a means to determine what greater structures exist in the landscape, and the level of variability within these units. For example, sampling of Lidar data, over a range of forest structural types, may be guided by such an approach. The generated hierarchy of multiscale segments could also find application in individual tree and canopy gap analysis, baseline mapping, polygon updating, change detection, object-based classifications, and for verifying the accuracy of existing forest polygons.

5.2. Can we extend MOSS results to provide information about broader scales than those initially evaluated?

In the GVWD, forest information is generated for stands 2.0 ha and greater, which is in the typical range of the minimum polygon sizes used in Canada (Gillis and Leckie, 1993). Stands smaller than this are generally aggregated into larger neighbouring units. However, due to the size of the original image (O_I), only five iterations of OSA/OSU were applied (Section 3.2.2), resulting in a maximum defined MSS value of 1.0308 ha for A_5 (Table 1). While we recognize that the mean segment size generated from this 1.0308 ha value was 2.57 ha (see Table 2 VL_5, *mean size*), it caused us to consider if there is a way to extend the current spatial information to coarser scales to facilitate a more general perception of the overall dominant scene structure, while also providing a mechanism to define layer(s) of interest with a MSS closer to the 2.0 ha requirements of the GVWD? In addition, are there *natural* scene measures that can be used to delineate coarse scale segments based on the fine scale spatial structure of the scene? By 'natural', we mean the values are not arbitrarily defined by an analyst, but are instead derived from the scene components themselves.

One solution is to use the previously defined WPS, MMU, and MSS values resulting from $A_{1–5}$, to model (i.e., forecast) coarser scale SCRM parameters from which new vector layers can be generated. In this example, forecasting was done using a regression approach with either an exponential or third order-polynomial model for two additional scales (at half-step intervals). Model selection was based on the highest R^2-value and how well the model visually fit the input data.

5.2.1. Defining WPS_{model}

Forecasting WPS values is relatively trivial as each incremented value results from upscaling with a resampling heuristic of 1.6 (see Section 3.2.2). Thus, based on the five WPS values in Table 1, the spatial resolution of the next two model area images (A_6 and

Table 3
Coarse scale model results from forecasting WPS, MMU, and MSS for two half-periods

Vector model	WPS (m^2)	MMU (ha)	MSS (ha)	# Watershed segments	# Merged segments	Mean size (ha)	Processing time (s)
VL_6	8.201	0.1210	1.8827	667	18	5.560	4.078
VL_7	10.42	0.1953	3.3963	409	12	8.333	3.815

ha, hectares; m, meters; p, pixel units; s, seconds.

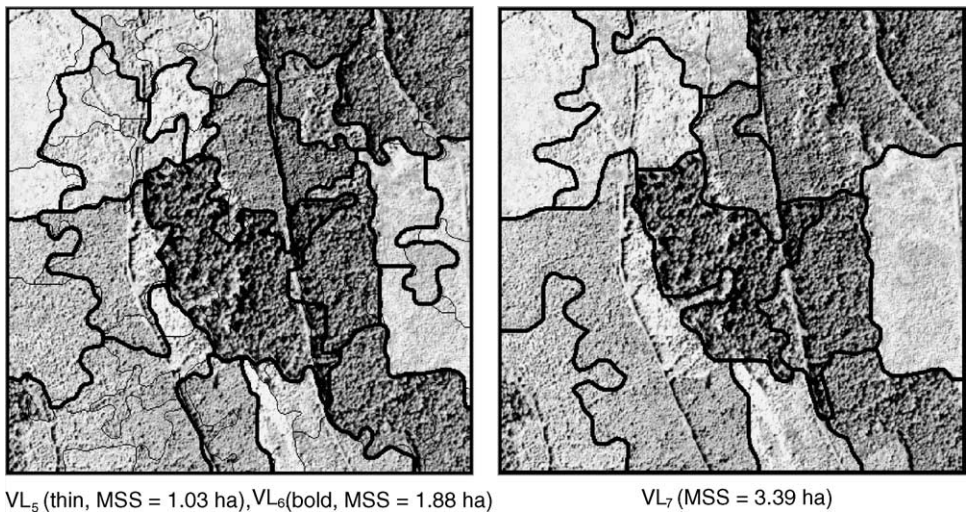

Fig. 8. Two additional hierarchical vectors layers ($VL_{6,7}$) generated from model (MSS, MMU, and WPS) values.

A_7) will be equal to 10.417 m^2 (6.554 m × 1.6 m) and 16.667 m^2 (10.417 m × 1.6 m). We also note that these two results are exactly predicted with an exponential model[4] of the same five values, forecast for two periods ($R^2 = 1$). The efficacy of this model is that it also allows values to be defined between the two periods (i.e., in half-steps).

5.2.2. Defining MMU_{model}

Based on the scene structure at each scale increment, the minimum value(s) defined in A_{1-5} have been constant at 18 pixels (for each image), even though the minimum kernel size in OSA is composed of 5 pixels. Thus, the forecast MMU values for the next two scales are 0.1953 ha (18 pixels × 10.417 m^2) and 0.500 ha (18 pixels × 16.667 m^2).

5.2.3. Defining MSS_{model}

When the MSS values from Table 1 are forecast for two periods then plotted and fitted with a third order polynomial trend-line[5] the resulting MSS values for A_6 and A_7 are 3.3963 ha and 10.47 ha, respectively, for each period (R^2 is 0.9994). These values are larger than the 2.0 ha requirement for the GVWD. However, by using a half-period forecast we are able to define a new MSS of 1.8827 ha that fits between the previously defined A_5 MSS value of 1.038 ha, and the new A_6 value of 3.3963 ha. This new value along with its corresponding WPS of 8.201 m^2 (defined from the half-step location in the WPS exponential model, Section 5.2.1) and MMU of 0.1210 ha (18 pixels × 8.201 m^2) enables the generation of vector layer 6 (VL_6), while the WPS, MMU, and MSS values defined for A_6 result in the generation of vector layer 7 (VL_7; Table 3). Those values defined from A_7 are not considered in this example as they exceed the initial 2.0 ha criteria.

To simplify viewing, we have illustrated the new layers in two associated graphics. In Fig. 8, the left image displays an overlay of VL_6 (bold) with finer scale vectors from VL_5 (thin). The right image shows the coarser sized vectors from VL_7 that represent the upper-most layer in

[4] $y = 0.625e^{0.47x}$ where x can be substituted with the values 6 and 7, which represent A_6 and A_7.

[5] The exponential model produced an $R^2 = 0.9893$.

this vector-layer triad. The mean size of segments in these layers range from 2.57 ha for VL_5, 5.56 ha for VL_6 to 8.33 ha for VL_7 (Table 3). By generating the initial SCRM spatial parameters from OSA/OSU then using the results to model coarser segments, an interpreter will be able to generate a hierarchical range of vector layers which provide a larger area perspective of 'naturally occurring' dominant scene structures.

6. Conclusion

Over the last decade there has been increasing interest in the object-based analysis of EO data. To support this trend, image-processing tools are required that can take advantage of the *H-res* image-object information inherent to high-resolution imagery. In addition, for widespread adoption, such tools need to be easy to use, fit within existing image-processing and ecological frameworks, require minimal user 'tweaking' or 'trial and error', have a sound theoretical underpinning, and most importantly provide results that closely match or exceed those generated by a human interpreter. In an effort to meet these conditions, we have presented *multiscale object-specific segmentation* as an integrative object-based approach for automatically delineating cognitively meaningful multiscale image-objects from high-resolution EO data. Segmentation is based on automatically extracted spatial measures that are explicitly related to the varying sized, shaped, and spatially distributed *image-objects* that compose a scene.

When applied to a high-resolution forest image we show that MOSS can be used to automatically delineate a range of objects that correspond from individual tree crowns to forest stands, and described how such a tool could be used for computer-assisted forest inventory mapping. We also discuss how SCRM parameters can be extended with a simple model, to allow for larger scale analysis. Presentation of the theoretical and scientific workings of MOSS is intended to provide users with an understanding of the segmentation process. Using an application from forest inventory illustrates how integrating a segmentation procedure within an existing information generation framework may increase the consistency of results, and indicates how cost savings may be realized through an incremental inclusion of automated processes to an increasingly digital interpretation environment.

Acknowledgements

This research has been funded by an *Natural Sciences and Engineering Research Council of Canada* (NSERC) and a *Groupe de recherche en écologie forestière interuniversitaire* (GREFI) post-doctoral fellowship awarded to Dr. Geoffrey Hay. We appreciate the generous use of computing resources and informative discussions provided by Dr. Danielle Marceau, Director of the Geocomputing lab at the University of Montreal. The Operators of the Greater Victoria Watershed District Office are thanked, for providing forest inventory data and for enabling research access in the watershed area. We are also grateful to the Institute for Regional Development of the University of Castilla La Mancha (IDR-UCLM) for providing software licensing to Dr. Castilla, and for comments from two anonymous reviewers who have helped to improve this manuscript.

References

Abramson, S.B., Schowengerdt, R.A., 1993. Evaluation of edge-preserving smoothing filters for digital image mapping. ISPRS J. Photogrammetry Remote Sensing 48 (23), 2–17.

Asner, G.P., Hicke, J.A., Lobell, D.B., 2003. Per-pixel analysis of forest structure. Vegetation indices, spectral mixture analysis and canopy reflectance modeling. In: Wulder, M.A., Franklin, S.E. (Eds.), Remote Sensing of Forest Environments. Concepts and Case Studies. Kluwer Academic Publishers, pp. 209–254.

Atkinson, P.M., 2004. Resolution manipulation and sub-pixel mapping. In: de Jong, S.M., van der Meer, F.D. (Eds.), Remote Sensing and Digital Image Analysis. Including the Spatial Domain. Book Series: Remote Sensing and Digital Image Processing, vol. 5Kluwer Academic Publishers, Dordrecht, pp. 51–70 (Chapter 4).

Baatz, M., Schape, A., 2000. In: Strobl, J., Blaschke, T. (Eds.), Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation, Angewandte Geographische Informationsverarbeitung, vol. 12. Wichmann-Verlag, Heidelberg, pp. 12–23.

Baraldi, A., Parmiggiani, F., 1994. A Nagao-Matsuyama approach to high-resolution satellite image classification. IEEE Trans. Geosci. Remote Sensing 32 (4), 749–758.

Beaulieu, J.M., Goldberg, M., 1989. Hierarchy in picture segmentation: a stepwise optimisation approach. IEEE Trans. Pattern Anal. Mach. Intell. 11, 150–163.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J. Photogrammetry Remote Sensing 58, 239–258.

Beucher, S., Lantuéjoul, C., 1979. Use of watersheds in contour detection. In: International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation, Rennes, France, 17–21 September. CCETT/IRISA Report no. 132, pp. 2.1–2.12.

Blaschke, T., Lang, S., Lorup, E., Strobl, J., Zeil, P., 2000. In: Cremers, A., Greve, K. (Eds.), Object-Oriented Image Processing in an Integrated GIS/Remote Sensing Environment and Perspectives for Environmental Applications, Environmental Information for Planning, Politics and the Public, vol. 2. Verlag, Marburg, pp. 555–570.

Blaschke, T., Hay, G., 2001. Object-oriented image analysis and scale-space: theory and methods for modeling and evaluating

multi-scale landscape structure. Int. Arch. Photogrammetry Remote Sensing 34, 22–29.

Blaschke, T., Burnett, C., Pekkarinen, A., 2004. Image segmentation methods for object-based analysis and classification. In: de Jong, S.M., van der Meer, F.D. (Eds.), Remote Sensing and Digital Image Analysis. Including the Spatial Domain. Book Series: Remote Sensing and Digital Image Processing, vol. 5Kluwer Academic Publishers, Dordrecht, pp. 211–236 (Chapter 12).

Burnett, C., Blaschke, T., 2003. A multi-scale segmentation/object relationship modelling methodolgy for landscape analysis. Ecol. Model. 168, 233–249.

Castilla, G., 2003. Object-oriented analysis of remote sensing images for land cover mapping: conceptual foundations and a segmentation method to derive a baseline partition for classification. Ph.D. Thesis, Polytechnic University of Madrid. http://www.montes.upm.es/Servicios/biblioteca/tesis/GCastillaTD_Montes.pdf.

Castilla, G., 2004. Size-constrained region merging: a new tool to derive basic land-cover units from remote sensing imagery. In: Theory and Applications of Knowledge Driven Image Information Mining, with Focus on Earth Observation. ESA SP-553 (CDROM).

Castilla, G., Lobo, A., Solana, J., 2004. Size-constrained region merging (SCRM): a new segmentation method to derive a baseline partition for object-oriented classification. In: Proceedings of SPIE, vol. 5239. pp. 472–482.

Devereux, B.J., Amable, G.S., Costa Posada, C., 2004. An efficient image segmentation algorithm for landscape analysis. Int. J. Appl. Earth Observ. Geoinf. 6, 47–61.

Dynapix, 2004. Esperantotm—Biomedical Imaging Software Platform. Image Analysis Solutions. http://www.dynapix-intelligence.com/dynapix_products_esperanto_en.html (White paper).

Fisher, P., 1997. The pixel: a snare or a delusion. Int. J. Remote Sensing 18, 679–685.

Fjørtoft, R., Lopès, A., Marthon, P., Cubero, E., 1998. An optimal multiedge detector for SAR image segmentation. IEEE Trans. Geosci. Remote Sensing 36 (3), 793–802.

Fotheringham, A.S., Wong, D.W.S., 1991. The modifiable areal unit problem in multivariate statistical analysis. Environ. Plann. A 23, 1025–1044.

Gillis, M., Leckie, D. 1993. Forest Inventory Mapping Procedures Across Canada. Information Report PI-X-114. Forestry Canada, Petawawa National Forestry Institute, 79 pp.

Gougeon, F.A., 1995. A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. Can. J. Remote Sensing 24 (3), 274–284.

Gougeon, F.A., Leckie, D.G., 2001. Individual tree crown image analysis—a step towards precision forestry. In: Proceedings of the First International Precision Forestry Symposium. Seattle, Washington, USA, June 17–20. University of Washington, Seattle (CDROM).

Graham, I., O'Callaghan, A., Wills, A.C., 2001. Object Oriented Methods. Principles and Practice, third ed. Addison-Wesley, 832 pp.

GVWD, 1993. Greater Victoria Water District Watershed Management Forest Cover Classification. Sooke Lake Watershed. Scale 1:10000. Prepared by Hugh Hamilton Ltd., Vancouver, BC.

Hagner, O., 1990. Computer-aided forest stand delineation and inventory based on satellite remote sensing, Proceedings of the SNS/IUFRO Workshop on the Usability of Remote Sensing for Forest Inventory and Planning, vol. 94–105. Swedish University of Agricultural Sciences, Remote Sensing Laboratory, Report 4, Umeå.

Hall, O., Hay, G.J., 2003. A multiscale object-specific approach to digital change detection. Int. J. Appl. Earth Observ. Geoinf. 4/4, 311–327.

Hall, O., Hay, G.J., Bouchard, A., Marceau, D.J., 2004. Detecting dominant landscape objects through multiple scales: an integration of object-specific methods and watershed segmentation. Landscape Ecol. 19 (1), 59–76.

Hall, R.J., 2003. The roles of aerial photographs in forestry remote sensing image analysis. In: Wulder, M.A., Franklin, S.E. (Eds.), Remote Sensing of Forest Environments. Concepts and Case Studies. Kluwer Academic Publishers, pp. 47–75.

Haris, K., Efstratiadis, S.N., Katsaggelos, A.K., 1998. Hybrid Image Segmentation Using Watersheds and Fast Region Merging. IEEE Trans. Image Process. 7 (12), 1684–1699.

Hay, G.J., Niemann, K.O., 1994. Visualizing 3-D texture: a three dimensional structural approach to model forest texture. Can. J. Remote Sensing 20 (2), 90–101.

Hay, G.J., Niemann, K.O., McLean, G.F., 1996. An object-specific image-texture analysis of H-resolution forest imagery. Remote Sensing Environ. 55, 108–122.

Hay, G.J., Niemann, K.O., Goodenough, D.G., 1997. Spatial thresholds, image-objects and upscaling: a multiscale evaluation. Remote Sensing Environ. 62, 1–19.

Hay, G.J., Marceau, D.J., Bouchard, A., Dubé, P., 2001. A multiscale framework for landscape analysis: object-specific upscaling. Landscape Ecol. 16, 471–490.

Hay, G.J., Dubé, P., Bouchard, A., Marceau, D.J., 2002. A scale-space primer for exploring and quantifying complex landscapes. Ecol. Model. 153 (1–2), 27–49.

Hay, G.J., Blaschke, T., Marceau, D.J., Bouchard, A., 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. Photogrammetric Eng. Remote Sensing 57, 327–345.

Hay, G.J., Marceau, D.J., 2004. Multiscale object-specific analysis (MOSA): an integrative approach for multiscale landscape analysis. In: de Jong, S.M., van der Meer, F.D. (Eds.), Remote Sensing and Digital Image Analysis. Including the Spatial Domain. Book Series: Remote Sensing and Digital Image Processing, vol. 5Kluwer Academic Publishers, Dordrecht, pp. 71–92 (Chapter 3).

Ji, S., Park, H.W., 1996. Image segmentation of color image based on region coherency. In: Kaiser, P.K., Boynton, R.M. (Eds.), Proceedings of ICIP 98', vol. 1.Human Color Vision. Optical Society of America, Washington, DC, pp. 80–83.

Kay, J., Regier, H., 2000. Uncertainty, complexity, and ecological integrity: insights from an ecosystem approach. In: Crabbe, P., Holland, A., Ryszkowski, L., Westra, L. (Eds.), Implementing Ecological Integrity: Restoring Regional and Global Environmental and Human Health, Kluwer, NATO Science Series, Environmental Security. pp. 121–156.

Leckie, D.G., Gillis, M.D., 1995. Forest inventory in Canada with emphasis on map production. Forestry Chronicle 71 (1), 74–88.

Leckie, D.G., Gillis, M.D., Gougeon, F., Lodin, M., Wakelin, J., Yuan, X., 1998. Computer-assisted photointerpretation aids to forest inventory mapping: some possible approaches. In: Hill, D.A., Leckie, D.G. (Eds.), Proceedings of the International Forum: Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry, Victoria, BC, February 10–12. Canadian Forest Service, Pacific Forestry Center, Victoria, BC, pp. 335–343.

Leckie, D.G., Gougeon, F.A., Walsworth, N., Paradine, D., 2003. Stand delineation and composition estimation using semi-auto-

mated individual tree crown analysis. Remote Sensing Environ. 85, 355–369.

Lindeberg, T., 1994. Scale-space theory: a basic tool for analyzing structures at different scales. J. Appl. Stat. 21 (2), 225–270.

Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W., 2001. Geographic Information Systems and Science. John Wiley and Sons, p. 454.

Marceau, D.J., Howarth, P.J., Gratton, D.J., 1994. Remote sensing and the measurement of geographical entities in a forested environment. Part 1. The scale and spatial aggregation problem. Remote Sensing Environ. 49 (2), 93–104.

Marceau, D.J., 1999. The scale issue in the social and natural sciences. Can. J. Remote Sensing 25 (4), 347–356.

Marceau, D.J., Hay, G.J., 1999. Remote sensing contributions to the scale issue. Can. J. Remote Sensing 25 (4), 357–366.

Marr, D., 1982. Vision. A Computational Investigation into the Human Representation and Processing of Visual Information W.H. Freeman and Company, p. 396.

Meyer, F., Beucher, S., 1990. Morphological segmentation. J. Visual Commun. Image Represent. 1 (1), 21–46 Academic Press.

Openshaw, S., 1984. The Modifiable Areal Unit Problem, Concepts and Techniques in Modern Geography (CATMOG), no. 38 , 40 pp.

Prigogine, I., 1997. The End of Certainty: Time, Chaos and the New Laws of Nature. Free Press, New York, 240 pp.

Schiewe, J., Tufte, L., Ehlers, M., 2001. Potential and problems of multi-scale segmentation methods in remote sensing. Geographische Informationssysteme 6 (34), 39.

Schneider, W., Steinwender, J., 1999. Landcover mapping by interrelated segmentation and classification of satellite images, Int. Arch. Photogrammetry Remote Sensing, 32, Part 7-4-3 W6, Valladolid.

Slater, P.N., 1980. Remote Sensing: Optics and Optical Systems. Addison-Wesley, 575 pp.

Steinhardt, U., Volk, M., 2003. Meso-scale landscape analysis based on landscape balance investigations: problems and hierarchical approaches for their resolution. Ecol. Model. 168, 251–265.

Stewart, S.A., Hay, G.J., Rosin, P.L., Wynn, T.J., 2004. Multiscale structure in sedimentary basins. J. Basin Res. 16, 183–197.

Strahler, A.H., Woodcock, C.C., Smith, J.A., 1986. On the nature of models in remote sensing. Remote Sensing Environ. 20 (2), 121–139.

Thom, R., 1988. Esquisse d'une sémiophysique. InterEditions, Paris.

Van der Sande, C.J., de Jong, S.M., de Roo, A.P.J., 2003. A segmentation and classification approach of IKONOS-2 imagery for land cover mapping to assist flood risk and flood damage assessment. Int. J. Appl. Earth Observ. Geoinf. 4, 217–229.

Vincent, L., Soille, P., 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans. Pattern Anal. Mach. Intell. 13 (6), 583–598.

Waldrop, M.M., 1992. Complexity. The emerging science at the edge of order and chaos Simon and Schuster, 380 pp.

Weickert, J., 1998. In: Levi, P., Ahlers, R.J., May, F., Schanz, M. (Eds.), Fast Segmentation Methods Based on Partial Differential Equations and the Watershed Transformation. Mustererkennung. Springer, Berlin.

Woodcock, C.E., Harward, V.J., 1987. Nested-hierarchical scene models and image segmentation. Int. J. Remote Sensing 13 (16), 3167–3187.

Wu, J., 1999. Hierarchy and scaling: extrapolating information along a scaling ladder. Can. J. Remote Sensing 25 (4), 367–380.

Wu, J., Marceau, D.J., 2002. Modelling complex ecological systems: an introduction. Ecol. Model. 153 (1–2), 1–6.

Wulder, M.A., 1998. Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. Prog. Phys. Geogr. 22 (4), 449–476.

Wulder, M.A., Seemann, D., 2003. Forest inventory height update thorough the integration of lidar data with segmented landsat imagery. Can. J. Remote Sensing 29 (5), 536–543.

Wulder, M.A., Hall, R.J., Coops, N.C., Franklin, S.E., 2004. High spatial resolution remotely sensed data for ecosystem characterization. BioScience 54 (6).

Zhang, Y.J., 1996. A survey on evaluation methods for image segmentation. Pattern Recognit. 29 (8), 1335–1346.