Size-constrained Region Merging (SCRM): An Automated Delineation Tool for Assisted Photointerpretation

Guillermo Castilla, Geoffrey J. Hay, and Jose R. Ruiz-Gallardo

Abstract

The manual delineation of vegetation patches or forest stands is a costly and crucial stage in any land-cover mapping project or forest inventory based upon photointerpretation. Recent computer techniques have eased the task of the interpreter; however, a good deal of craftsmanship is still required in the delineation. In an effort to contribute to the automation of this process, we introduce Size-Constrained Region Merging (SCRM), a recently implemented software tool that provides the interpreter with an initial template of the to-be-mapped area that may reduce the manual digitization portion of the interpretation. In essence, SCRM transforms an ortho-rectified aerial or satellite image (single or multi-channel) into a polygon vector layer that resembles the work of a human interpreter, whom with no a priori knowledge of the scene, was given the task of partitioning the image into a number of homogeneous polygons all exceeding a minimum size. We provide background information on SCRM foundations and workflow, and illustrate its application on three different types of satellite images.

Introduction

At the beginning of this century, the advent of very high resolution civilian remote sensing (RS) satellites, such as Ikonos and QuickBird brought into crisis traditional pixel-based image analysis. Pixel-based analysis assumes that different land-cover classes behave as distinct surfaces made of recurrent elements that yield a particular joint reflectance profile (i.e., spectral signature) when observed at a suitable scale. In addition, it assumes that the area over which a single measurement is made (i.e., a pixel) is large enough to include a sufficient number of elements producing a typical class signature (Woodcock and Strahler, 1987). Thus, an implicit premise of pixel-based methods is that "the spatial resolution of the image is coarser than the spatial resolution of the classification" (Goodchild, 1994). The latter may be formally defined as "the minimum size of the circle, expressed by its diameter, over which the surroundings of a geographic point have to be observed in order to define the label at that point" (Castilla, 2003). Clearly, the above assumption does not hold true for most high-resolution imagery and land-cover classes. For example, imagine a natural landscape that has been partitioned into a set of small, square plots in order to analyze it. Imagine further that the plots have opaque walls, so that an observer standing on any plot cannot see outside. Let us now allot the side of each plot a length of 2.5 m, equivalent to the pixel size of a QuickBird multi-spectral image. If the above assumption is valid, there should be enough evidence within each plot so as to correctly classify it. To test this, imagine that we randomly place a hypothetical observer inside one of these plots in an area covered by sparse mixed woodland. Upon analysis, it would be highly unlikely for the observer to identify this piece of terrain as belonging to *sparse mixed woodland*, no matter the actual position of that plot.

This scenario and other limitations of pixel-based methods applied to high-resolution imagery (Schiewe *et al.*, 2001; Hay *et al.*, 2005) have increased the emphasis on object-oriented approaches (OOA) (Blaschke *et al.*, 2000 and 2004). OOA use objects in addition to classes in order to model the landscape. Within OOA, an object represents an individual, unit, or entity, either real or abstract, with a well-defined role in the problem domain (Booch, 1991). Any given object is an instance of some particular class. Conversely, a class is a set of objects that share a common structure and a common behavior. The class relations among objects are represented in *kind-of* hierarchies (*taxonomies*) that provide *inheritance*, and structural relations among objects are represented as *part-of* hierarchies (*partonomies*) that provide *encapsulation* (information hiding) (Booch, 1991). The OOA is especially suited for implementing a hierarchical patch model of the landscape (Woodcock and Harward, 1992; Wu and Loucks, 1995), where a *patch* is a landscape object, i.e., a discrete spatial unit differing from its surroundings in structure or function. Patches within a given hierarchic level may be *composed of* several patches from a lower level, while at the same time

Guillermo Castilla is with the Department of Geography, University of Calgary, Calgary, AB, T2N 1N4, Canada, and formerly with the Instituto de Desarrollo Regional, Universidad de Castila La Mancha, Campus Universitario, 02071 Albacete, Spain (gcastill@ucalgary.ca).

Geoffrey J. Hay is with the Department of Geography, University of Calgary, 2500 University Dr. NW, Calgary, AB T2N 1N4, Canada.

Jose R. Ruiz-Gallardo is with the Instituto de Desarrollo Regional, Universidad de Castila La Mancha, Campus Universitario, 02071 Albacete, Spain.

Photogrammetric Engineering & Remote Sensing
Vol. 74, No. 4, April 2008, pp. 409–419.

0099-1112/08/7404–0409/$3.00/0

they may be *part of* a larger patch at a higher level. Hence, a patch may be an entity ranging from the area covered by an isolated tree to an island continent. Additionally, patches from different levels may overlap in size; however, the mean size of patches within each level must increase steadily with the level. Since there are no clear differences in size between borderline instances from adjacent levels, an arbitrary size threshold must be established for a given area to qualify as a patch of a certain level, e.g., contiguous forested areas less than one hectare are not forests but forest stands. Such a threshold is analogous to the minimum mapping unit (MMU) used in cartography to impose the desired level of generalization over a territory. Once a hierarchical model is defined by a set of size constraints, the link between landscape and image can be realized on the premise that *image objects* at a given resolution coincide with landscape objects at some hierarchical level (Castilla, 2003; Hay *et al.*, 2001). Image objects are delimited regions of the image that are internally coherent and different from their surroundings. According to this premise, they may represent structural-functional landscape units at some level of generalization.

Image objects may be obtained manually, through visual interpretation (i.e., photointerpretation), or using some semi-automated, object-oriented image classification method. For decades, photointerpretation has been, and to a good extent is still the method of choice for producing fine-scale forest and land-cover maps. However, more recently, innovation in the RS/GIS market with eCognition® software (Definiens, 2005) has provided users with the first commercial tools for GEographic Object-Based Image Analysis (GEOBIA, Hay and Castilla 2008). The first step in GEOBIA is image segmentation, the partitioning of the image into a set of jointly exhaustive, mutually disjoint regions that are more uniform within themselves than when compared to adjacent regions. These regions are subsequently used as basic units to form classified objects. Unlike typical pixel-based classifications, an object-based classification takes into account not only conventional features such as spectral signatures, but additional features that cannot be derived for individual pixels (e.g., size and shape features), and most importantly, relational features between regions. With such capabilities, GEOBIA has the potential to supersede not only conventional pixel-based methods, but also photointerpretation. These prospects are beginning to be recognized by leading companies in the RS image analysis sector, who progressively are including segmentation modules in their recent product releases. However, GEOBIA is still in its infancy.

Image understanding is a complex cognitive process for which we may still lack key concepts. In particular, most image segmentation methods have been developed heuristically without a deeper examination of the semantic implications of the segmentation process. Therefore, to ensure a strong foundation, research in this field needs to focus on a deeper understanding of the relationship between image objects and landscape objects, rather than exclusively in the development of new techniques tailored for specific applications. Our work is partly devoted to fill this conceptual gap.

Since significant research remains until fully automated image interpretation is achieved, we suggest that the general approach to GEOBIA should be a pragmatic one. As such, the short-term goal, rather than trying to replace human interpreters, would be to support them in generating more timely, consistent and accurate products (Leckie *et al.*, 1998). This is relevant as on-screen digitization is the mainstay of today's photointerpretation. Furthermore, many consulting companies customarily use proprietary software or COTS (commercial off the shelf products) that create topology while digitizing, and display on-demand ancillary information to ease label allocation. Notwithstanding these advances, interpretation accuracy, consistency, and speed/cost are recurring concerns (Hall, 2003). Therefore, new and or better tools are required that produce incremental improvements in these areas. They need not provide final solutions or 100 percent correct results; they simply need to be tools that are useful and that can be easily corrected when they go awry. Specifically, they must be simple to apply, not require expensive equipment, not substantially alter the mapping workflow, nor involve inordinate fine-tuning by the interpreter (Leckie *et al.*, 1998). In order to facilitate these requirements, we introduce Size-Constrained Region Merging (SCRM). Essentially, SCRM transforms an ortho-rectified aerial or satellite image (single or multi-channel), into a polygon vector layer. This layer resembles the work of a human interpreter, whom without *a priori* knowledge of the scene was given the task of partitioning the image into a specific number of relatively homogeneous polygons all exceeding a minimum size. This layer may then be used as an initial template in the task of the interpreter, who just needs to aggregate (and sometimes correct) pre-delineated regions by simple drag-and-click operations. We note that SCRM results are only meant to be an intermediate imperfect aid for the work of the interpreter, since (a) SCRM only considers radiometric features for the segmentation, and (b) the correspondence between radiometric similarity and semantic similarity is not straightforward. We also note that although the SCRM sequence includes procedures other than region merging, we named the algorithm after this condition because it is the most influential step.

The objective of this paper is threefold. First, we provide conceptual foundations that underlie the SCRM approach to image segmentation. Such foundations constitute an unprecedented contribution underpinning GEOBIA, and therefore are treated in depth. Second, we briefly describe the SCRM workflow and illustrate with examples how it may be used as an automated delineation aid for computer-assisted photointerpretation. We then discuss its strengths and limitations, and summarize relevant points and future work.

Background SCRM Algorithm Foundations

The conceptual foundation, software development, and implementation of SCRM were initially conducted by the first author (Castilla, 2003). Since then, it has been tested and refined with the aid of others (Castilla *et al.*, 2004; Hay *et al.*, 2005). SCRM has been inspired and or influenced by a number of important contributions, which are described in the following sections. In the Limitations of then Algorithms section, we highlight their major limitations, and in the following sections, we expound upon the conceptual foundations of our approach.

Beaulieu and Goldberg's Algorithm

The *stepwise optimization algorithm* of Beaulieu and Goldberg (1989) begins by considering single pixels as the initial regions of analysis. At each iteration, the two adjacent regions that show the highest similarity or degree of fitting are merged. That is, the pair of adjacent regions merged at each iteration is the one showing the least distance in the feature space from which similarity is quantitatively estimated. Merging proceeds sequentially in this way until there is no pair with a dissimilarity distance below a user-defined threshold (i.e., merging tolerance). The final partition is optimal regarding the minimization of the internal heterogeneity of regions, but the procedure is too slow, since it allows only one merge per iteration. In addition, it leads to an uneven growth of regions between areas of smooth and coarse texture. As a result, areas marked by coarse texture

will consist of many small regions (often individual pixels), whereas smooth uniform areas will be segmented into large regions.

Woodcock and Harward's Algorithm

Based on Beaulieu and Goldberg's (1989) work, Woodcock and Harward (1992) introduced a faster algorithm that allowed multiple merges per iteration and included size constraints. At each iteration, a list of candidate pairs is prepared. Each candidate pair consists of two adjacent regions where both are the nearest neighbor (in feature space) of the other. Then, all candidate pairs whose dissimilarity distance is less than the current *merging tolerance* (T_{pass}) are allowed to merge, providing neither region has previously merged during this iteration. After each iteration, the new value of T_{pass} is computed automatically from (a) the current histogram of dissimilarity distances to nearest neighbors, and (b) a user-defined *merge coefficient* ($0 < C_m < 1$), where the latter is roughly the proportion of regions that merge at each iteration. In this way T_{pass} increases monotonically until it reaches a user-defined *maximum tolerance* (T_{glob}). The idea behind this procedure is to keep the order of the merging sequence as similar as possible to the *single merge per iteration* strategy, since segmentation quality is highly sensitive to that order. Therefore, the smaller the C_m, the closer the results are to the optimal ones, but also the slower the algorithm.

As these authors noted, the global threshold alone leads to a great disparity in size of output regions. The cause of that difference is the variety of textures usually found in remote sensing images. Size disparity is a problem mainly for two reasons. First, output regions (if used as basic units in an automated classification procedure) should represent landscape objects of the same hierarchic level, therefore their size should be in the same order of magnitude. Second, for cartographic purposes, regions smaller than the minimum mapping unit are a hindrance. Therefore, Woodcock and Harward (1992) supplemented their algorithm with size constraints that prevented excessive growing in smooth areas during normal merging; and once the global threshold was reached, they forced the development of regions exceeding the minimum required size in areas with coarse texture.

Baatz and Schape's Algorithm

An alternative strategy for tackling size disparity is that followed by the segmentation algorithm of Baatz and Schape, (2000), which is embedded in eCognition®. Similar to Woodcock and Harward's (1992) method, it starts with individual pixels as the initial regions, but instead of using a stepwise increase of the dissimilarity threshold, at each iteration it distributes the candidate pairs as far as possible from each other over the image. The actual way this distribution takes place is proprietary, although according to the authors it is devised to allow for repeatability. In any case, it achieves a uniform growth of regions throughout the image, so that the final regions all have a similar scale (size). Since a conservative (small) threshold permits fewer merges than a greater one, the mean size of segments will increase with the value of the threshold. For this reason it is called the *scale parameter*. A particularity of this algorithm is the optional inclusion of a *form heterogeneity factor* in the overall dissimilarity between two adjacent segments. In this case, dissimilarity is measured as a linear combination of radiometric heterogeneity (expressed by the mean of the variance in each band of pixels within the segment) and form heterogeneity (expressed by the ratio between the actual edge length of a segment, and the edge of a square with the same number of pixels as the segment). In this way, the segmentation favors the construction of regions with smooth edges and a more or less compact form. Although such an approach to tackling the fractal nature of landscape yields visually appealing results, it is conceptually moot, as it uses a dissimilarity metric that combines two incommensurable features, color and shape. Another disadvantage is that the scale parameter, being a unitless threshold that is image dependent, has no functional relationship with the number of output regions and therefore with their size. Hence, users need to find useful segmentation levels in a trial and error basis, even if they can optionally impose some size constraints at the outset as in the previous method (Blaschke and Hay, 2003).

Limitations of these Algorithms

We suggest that the previous three segmentation methods bear a number of limitations. First, they start the merging with individual pixels. Apart from being computationally expensive, the result is very sensitive to the merging order at early stages of the segmentation, when a single pixel still has a considerable weight in the features of a small region. Second, the control of region growth, when exerted, is made through somewhat arbitrary means, such as the merging coefficient in Woodcock and Harward's (1992) method, and the heuristic spatial distribution of candidate pairs in Baatz and Schape's (2000) algorithm. Third, they require the user to set parameters that are non-intuitive, such as the *scale parameter* in eCognition®. And fourth, they are conceptually inconsistent with the object-oriented approach (OOA): an underlying hypothesis of any segmentation method is that there is a correspondence between radiometric similarity in the image and semantic similarity in the imaged landscape. Thus, it is expected that image objects (segments) coincide with landscape objects (patches). However, if this is so, object boundaries should correspond to some discontinuity on the ground. Pixels can be thought of as arbitrarily delimited square plots; therefore, single pixels can hardly correspond to landscape objects. So starting the segmentation with single pixels is inconsistent with the OOA.

Blobs: The Basic Perceptual Units

SCRM lacks the incongruity of the other approaches starting with individual pixels, because it begins the region merging process with a fine partition made of primal regions or *blobs* derived from image morphology (see the Image Smoothing section for details). A *blob* is a perceptual concept that refers to a tiny homogeneous region in the image, darker, brighter or of different hue than its surroundings. Being internally homogeneous and different from its exterior, a blob may be also viewed as the basin of attraction of a perceptual attractor, where all pixels within that basin may be perceived as forming a distinct single entity. Under this view, blobs are the basic perceptual units, just as pixels are the physical basic units. Perceptual attractors may be associated to local minima in a transformed image where the value (DN) of pixels is proportional to the difference with adjacent pixels in the original image. Thus, blobs can be formally defined as the basins of attraction of local minima in a gradient magnitude image derived from the original one through some dissimilarity metric. This definition enables us to establish a powerful innovative link between catastrophe theory, graph theory, and morphological image analysis (for additional detail, see Castilla, 2003).

Linking Catastrophe and Graph Theories

Thom's (1975) topological theory of attractors, also known as *catastrophe theory*, may be interpreted, when applied to image perception, as a dividing of the image into *regular*

pixels (those belonging to some basin of attraction) and *singular* pixels (those separating the basins). According to this theory, every object in an image can be represented as an attractor of a dynamical system on a space of internal variables. Thus, an object may be recognized only when the corresponding attractor is *stable*. The stability is attained by a process, called *morphogenesis*. This consists in the disappearance of the attractors representing the initial unstable forms, and their replacement (i.e., capture) by the attractors representing the final forms, which is the observable state of the object(s). An attractor can be thought of as the centroid of an object, so that pixels belonging to the object are more attracted to it than to the centroid of neighboring objects. However, in the boundaries between objects, this attractive force becomes unstable and an infinitesimal move in one or another direction may produce a change of attractor: these are the *singular* pixels (i.e., edges or boundaries) that define the spatial structure of the image.

We have translated this account into graph theory by considering the image as the initial state of a planar dynamic network consisting of triangular meshes made up of nodes (pixels) connected through links through which the nodes interact. This interaction consists of quantitative luminance exchanges between the nodes. The intensity of the interaction is regulated by proximity in feature space, decreasing rapidly with dissimilarity distance, and it is formalized through a weight allocated to each link. A link may be active, if there is some noticeable interaction through it, or inactive if its weight is nearly zero. To simulate morphogenesis, we allow the network (i.e., image) to evolve through several cycles in which the state of a node is dependent on the state of adjacent nodes in the previous cycle, according to the above interaction mechanism. During this process, some nodes interact more strongly between themselves, while some others stop interaction. This induces a coherent behavior of nodes within local groups. Eventually, some inactive links may become active, opening paths between nearby groups, so that small groups are captured by larger ones. After relatively few cycles, the network reaches a steady state far from equilibrium (where equilibrium would represent a uniform distribution of luminance across the network, i.e., a flat image). Thus, the network evolves towards a piecewise constant image in which the pixels within each local group have roughly the same value. Each local group has a node that acts as an attractor in feature space, i.e., a node whose basin of attraction is the local group itself. Then, the remaining attractors within the network are the stable attractors, and their respective basins of attraction coincide with blobs. As a result, this evolution can be viewed as a self-organization of the image into perceptual units (Castilla, 2003). The practical implementation of this process is a non-linear diffusion filtering, as explained in the Image Smoothing section.

Linking Gradient Watersheds and Morphogenesis

The set of singular pixels demarcating the stable basins of attraction is obtained using the watershed algorithm, a morphological segmentation method commonly used in computer vision and biomedical imaging (Meyer, 2001). This algorithm extracts the network of ridges (watersheds, or drainage divides) that exist in the input image (usually a gradient magnitude image) when it is considered as a Digital Elevation Model (DEM). In our method, since the filtered image (i.e., the final state of the above network; see the previous section) is a quasi-piecewise constant, its corresponding gradient magnitude image resembles a DEM from a lunar plain full of craters of different shape and size. Here, each crater is the basin of attraction of a local gradient minimum, which is located at the bottom of the crater. Now, the concept of *attractor* becomes more evident by introducing an analogy with gravity. For example, if a basketball is dropped at any location within a crater, it will be *attracted* towards the bottom and eventually come to rest there. The event of the ball getting trapped in some pit (i.e., an unstable attractor) near the bottom is precluded by filtering, which smoothes out the crater surface. However, at the very top of the crater, the attractive force exerted by the base of the crater becomes unstable, as any small shift may result in the ball falling in or out. Therefore, ridges may be defined as the set of singular pixels bounding the basins of attraction, which conveniently also happen to be the pixels that the watershed algorithm foregrounds as boundaries of catchment basins.

When blobs are viewed in this way, it becomes apparent that the watershed algorithm is more than just another segmentation method: it is *the* tool for applying Thom's (1988) *semiophysics* to image analysis (Castilla, 2003). The merit of Thom's work is that it provides a link between the quantitative physical theory of the world and our qualitative daily experience (Smith, 1995). Our claim, which we believe is an important contribution to this field, is that this link can be realized using gradient watersheds when we observe the world through images. We further suggest that within Marr's (1982) definition of *vision*, this algorithm provides a first step in the transformation of a numerical representation (a digital image) of a territory into a shape-oriented representation (a thematic vector layer).

Simplifying the Watershed Partition

Watersheds alone are not sufficient to derive a meaningful map. Users are typically interested in structures larger than blobs. Indeed, the profuse number of watershed regions is normally regarded as a disadvantage known as *oversegmentation*. Therefore, the watershed partition has to be further simplified. There are three common ways of achieving this: (a) embedding the watersheds in a linear scale-space framework, (b) thresholding the flood dynamics of watershed arcs (see below), and (c) merging of catchment basins according to a similarity criterion. The first two methods are discarded for reasons explained below, and the third is treated in the Methodology: SCRM section.

The first method (Jackway 1996; Olsen and Nielsen 1997; Gauch 1999) consists in creating a family of increasingly blurred images from the original one. This family forms a scale-space representation (Lindeberg, 1994; Hay *et al.*, 2002), in which the scale of each image is given by the width of the Gaussian filter used to smooth it. As a result of this blurring, successive images are simpler, meaning that there are less gradient minima. Next, the watershed partition is computed at each scale, and the resulting catchment basins are linked across scales, yielding a hierarchical partition. However, there are limitations with this method: first, it is computationally intensive; and second, the linkage is not as straightforward as one might think. Gaussian blurring deforms the structure of the original image, not only annihilating but also displacing both edges and gradient minima. Therefore, there may be situations where the assignment of previous regions to new ones cannot be clearly established. These matching ambiguities between successive images, which also occur for image-pyramids (a stack of images of increasingly larger pixel size), are known as the *correspondence problem* in computer vision (Cox, 1993).

The second method uses the concept of *geodesic saliency* of watershed contours (Najman and Schmitt, 1996). This procedure registers the height of the water table at which overflow between adjacent catchment basins takes place. Since ridges usually do not follow contour lines, water will begin to overflow from saddle points (passes across the ridges). Therefore, a hierarchical partition can be obtained

using the altitude of these saddle points as an index to flood dynamics of watershed arcs. This approach is convenient if the goal is to foreground the most contrasted objects within an image, but not to obtain a partition where the size distribution of regions is bounded and to a degree controlled by the user. As a point of fact, successive partitions corresponding to increasing thresholds of watershed dynamics will show an increasing disparity in size, with many small high-contrast regions littered throughout larger regions, and this is not convenient for a photointerpretation template. Then, the only remaining method to reduce the number of watershed segments is *region merging*, a common segmentation technique with a long history in image processing.

Methodology: SCRM

Proprietary (to the author G. Castilla) SCRM source code is written in IDL, and can be run either within the commercial remote sensing software ENVI® (RSINC, 2005b) as a user extension or stand-alone in conjunction with the IDL Virtual Machine (RSINC, 2005a). A freeware version may be downloaded from the ITTVIS Code Contribution Library (**www.ittvis.com/codebank**). In order to use SCRM, four parameters must be specified: (a) the *desired mean size of output polygons* (DMS, in hectares), (b) the minimum size required for polygons, or *minimum mapping unit* (MMU, in hectares), (c) the *maximum allowed size* (MAS, in hectares), and (d) the minimum distance between vertices in the vector layer, or *minimum vertex interval* (MVI, in meters). MVI is an indication of the positional accuracy of boundaries and is internally used to define the working pixel size (i.e., spatial resolution). SCRM workflow is as follows (Figure 1). The input image (previously ortho-rectified to some cartographic projection) is (if necessary) resampled to a suitable pixel size, and then filtered with *Gradient Inverse Weighed Edge Preserving Smoothing* (GIWEPS; Castilla, 2003). The output of this step is an almost piecewise constant image, from which the gradient magnitude is computed. This gradient magnitude image is then searched for local minima, and the area of influence of each minimum is contoured and labeled with the watershed algorithm. The resulting regions are aggregated iteratively by increasing dissimilarity until they all exceed the size of the minimum mapping unit (MMU). Then, the labeled image containing the final partition is converted into a vector layer.

We note that a segmentation sequence consisting of image smoothing and/or gradient magnitude simplification, followed by a watershed algorithm and region merging is not new in image analysis (e.g., Haris *et al.*, 1998; Weickert 1998; Fjørtoft *et al.*, 1998; Bleau and Leon, 2000; Chen *et al.*, 2004). However, to the best of our knowledge, no other has explicitly linked gradient watersheds to the theory of attractors. Furthermore, except for the watershed transform, the techniques proposed here are original, and our method is grounded on a solid conceptual basis (Castilla, 2003), an asset that many segmentation algorithms lack.

Image Resampling

The default minimum vertex interval (MVI) is double the pixel size of the input image, since this is the spline interval that is applied to interpolate the centers of watershed pixels (see the Vectorization section). If the final product is intended to depict the theme of interest at a coarser scale, then segmentation should be performed at a resolution that matches this cartographic requirement in the same way interpreters should not zoom beyond a certain visualization scale while digitizing arcs. On the one hand, too fine a scale often overwhelms the operator (be it man or machine) with too much detail and limited context. Conversely, as a consequence of the fractal nature

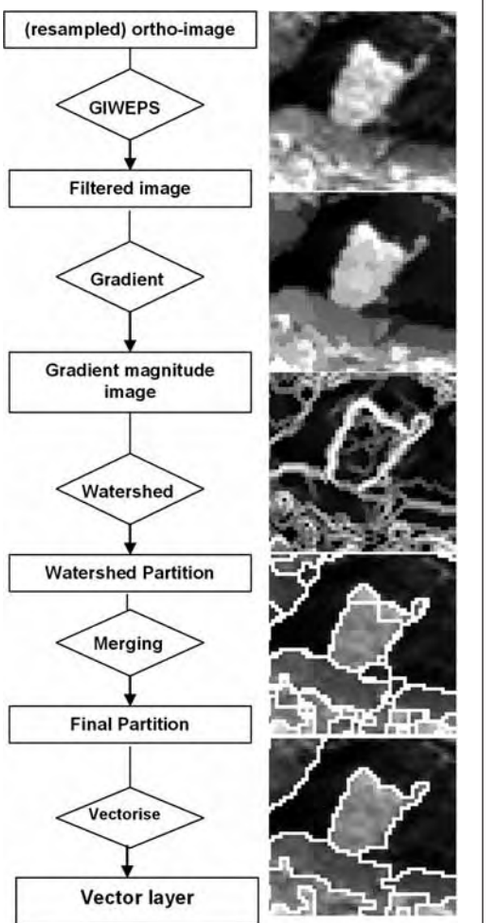

Figure 1. Overview of the SCRM workflow.

of landscapes, the length of arcs increases indefinitely as the resolution increases. Thus, the finer the pixel size, the more intricate the arcs (Figure 2). Therefore, in this case, the input image should be up-scaled to a resolution balancing edge simplicity and accuracy. This is done by resampling the image to half MVI by simple pixel averaging. When the user is unsure about what minimum vertex interval fits their application, they may set MVI equal to the boundary positional accuracy required for the final product, as it will on average be better than this value.

Image Smoothing

A consequence of hierarchical patchiness in landscapes is the presence of areas of coarse texture in all types of remote sensing imagery, irrespectively of spatial resolution. Coarse texture, or high local variance, is mainly due to the existence of recurrent elements that are large enough to produce some variation in the image, but too small to be resolved by the sensor, such as individual trees in a Thematic Mapper image. Since no information can be retrieved about their individual shape, they cannot be included individually in a shape-oriented representation of the scene derived from that image. Therefore, they have to be captured within the larger element they are a part of (e.g., a forest stand). On the other hand, coarse texture areas are characterized by a high density of local luminance extrema that produce a considerable amount

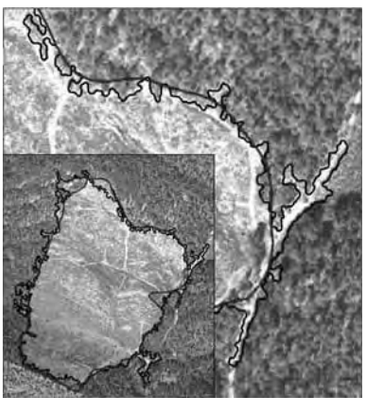

Figure 2. Two automated delineations of a burn scar at two different resolutions, 1 m and 10 m; the latter is smoother. The larger image shows a zoom on the northeast corner of the scar. A color version of this figure is available at the ASPRS website: **www.asprs.org.**

of spurious minima in the corresponding gradient magnitude image. Following the discussion in the background section, these minima may be regarded as perceptually unstable attractors, which we remove by processing the image with a non-linear diffusion filter that simulates the aforementioned morphogenesis. Non-linear diffusion is an iterative process where diffusivity (the rate of luminance exchange between adjacent pixels) changes according to the evolution of the local gradient (*difference* in brightness for single channel images, or *dissimilarity* for multichannel images between each pixel and their immediate neighbors). Unlike linear diffusion schemes, such as the iterative Gaussian filter used in Scale-Space, non-linear filters produce minimal blurring and displacement of edges. Rather, they only act only upon the unresolved elements that lead to coarse texture while preserving larger structures and their edges (Weickert, 1997).

Our implementation, called *Gradient Inverse Weighted Edge Preserving Smoothing* (GIWEPS) has as precursors the filters by Wang *et al.* (1981) and by Perona and Malik (1990). In GIWEPS, the new digital number (DN) of a given pixel is the weighted mean of the current DNs of its eight neighbors. The weight of each neighbor is proportional to its Euclidean distance (in the feature space) to the pixel under examination, and the proportion is governed by an arbitrarily fixed diffusivity parameter. The filter is applied iteratively, until the difference between consecutive output images is negligible, which typically occurs in a few iterations. Further details can be found in Castilla (2003).

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. In the case of multi-band images, the slope describes the variations in similarity of adjacent pixels across the image. The dissimilarity measure used here is the Euclidean distance between points (pixel signatures) in the feature space. At each pixel of the image, the gradient magnitude is computed as the square root of the sum of squared dissimilarity distances between the East and West neighbors, and the North and South neighbors, respectively. Note that the *gradient minima* are those pixels whose value is lower than that of their eight neighbors in the gradient magnitude image.

Watershed Partition

The application of the topographic concept of watershed to the field of image analysis was introduced by Beucher and Lantejoul (1979) and later implemented into an efficient algorithm by Vincent and Soille (1991). In the SCRM workflow, the idea is to consider the gradient magnitude image as a DEM. The goal is to find the drainage divides, or watersheds, of that virtual territory, which is achieved by simulating a gradual immersion of the DEM. In the output partition, watershed (boundary) pixels are set to zero, whereas non-zero pixels have as DN the numeric label of the region (blob) to which they belong.

Region Merging

In this step, the regions of the watershed partition (i.e., blobs) are aggregated until all regions in the partition are larger than the specified minimum size (MMU). The merging sequence is such that the homogeneity of the resulting regions is maximal given the size constraints. The dissimilarity measure used as merging criterion is the Euclidean distance in a feature space, where each dimension corresponds to a channel of the image. Thus, the signature of a region, from which dissimilarity to its adjoining neighbors is computed, corresponds to the coordinates of the region centroid in the feature space. That is, a signature is an n-component vector where each component is the mean value in each of n channels of pixels belonging to the region. After a merge, the signature of a new region is the weighted (by size) mean of the signatures of the two merged regions. In this way, region signatures are computed from the original image only once, at the beginning of the procedure. The same can be said about the adjacency table (an array returning the list of neighbors of any given region), 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 neighbor (MSN) to each region is trivial.

The table-based updating of both signatures and adjacency, which was first implemented in a recent version (Castilla, 2004), is much faster than our former image-based updating. This fact enables us to adopt the optimal *single merge per iteration* strategy, which otherwise would be too slow. Consequently, in each iteration the two adjacent regions that merge are those best fitting, (i.e., those having the least dissimilarity distance from the set of candidate pairs). Next, the AT, SL, and MSN arrays are updated, the new best fitting candidate pair is identified, and a new iteration proceeds. During the first several hundred iterations, the list of candidate pairs consists of every pair of adjacent regions in the image. As the merging proceeds, smooth low contrast areas become occupied by increasingly larger regions, and the maximum size constraint (MAS) comes into play. At a given iteration, if the best fitting pair consists of two regions both exceeding MAS, then it is not allowed to merge. Furthermore, this pair is permanently withdrawn from the candidate list, and the second best fitting pair is selected for merging. The merging continues this way until the sum of (a) the number of regions currently larger than the minimum allowed size MMU, plus (b) the expected number of final regions that may result from the area currently occupied by regions smaller than MMU, is less than the expected number of final regions (i.e., the image area divided by DMS). This is

a partial stop criterion that guarantees that the final mean size of regions will be close to DMS. Thereafter, the candidate list is restricted only to those pairs where at least one of both regions is smaller than MMU. In this way, homogeneous regions are formed first, and then dissimilar regions smaller than MMU are progressively incorporated to the former until all regions are larger than MMU.

The actual merging of two regions involves replacing the numeric label of one of them (arbitrarily the lowest one) with the label of the other in the *final label list* (FLL). FLL is an array of length equal to the number of regions (blobs) in the initial (watershed) partition. At any point during the merging process, there is a link that keeps track of blobs composing each new region, so that FLL can be easily updated. Once the merging is completed, a new image (the SCRM partition) 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 regions are filled with the numeric label of the corresponding region.

Vectorization

The last step in the workflow is to convert the SCRM partition into a polygon vector layer and save it as an ESRI shapefile. In order to proceed, the centers of boundary (zero-valued) pixels are considered the initial vertices forming the vector layer. This is analogous to considering boundary pixels as a transition zone between patches that can be represented by its medial axis. Then, nodes (junctions connecting arcs) are the centers of those boundary pixels having more than two non-diagonal neighbors that are boundary pixels. Arcs on the other hand correspond to chains of boundary pixels that start and end by a node. Finally, vector units (polygons) are delimited by the set of arcs bounding the corresponding region. In order to give a *natural* appearance to arcs, a spline interpolation is applied to the centroid of each three consecutive vertices within the arc. The smoothed arc is further simplified with a proprietary implementation of the Douglas-Peucker (1973) algorithm, which deletes redundant vertices using a tolerance of half the pixel size (Figure 3). These results are then saved as a shapefile, and the associate database file (.dbf) is filled with radiometric statistics about each polygon (i.e., min, max, mean DN, and standard deviation).

Results

In this section, we present results from three different types of remote sensing images, and note that the third image (Figure 6) illustrates how a SCRM output vector layer may be used as an initial template for computer assisted photointerpretation. SCRM may be applied to any kind of RS imagery (in Geotiff, jpeg, or ENVI® format), of any size (images larger than two Megapixels at working resolution are subject to tiled processing) and any number of channels. Typical processing time (using an Intel Pentium IV, 2.3 GHz with 1 GB of RAM) is less than three minutes per Megapixel. Before proceeding, the user should have a sense for how broken up the scene needs to be. For example, if we are interested in delineating forest stands with a 10 ha average size, a suitable Desired Mean Size (DMS) would be some 2.5 ha. In this way, we would generate sufficient units to avoid excessive manual digitization in a latter stage. The Maximum Allowed Size (MAS) could be set to 10 ha, so that units larger than 20 ha would be rare, or left blank, if this is not a concern. The minimum mapping unit (MMU) of the final product must be known in advance, or the user must recognize that any region below the default size will not be retained in the output partition no matter how distinct the region is. The last SCRM input parameter, the Minimum Vertex interval (MVI), can also be set intuitively. If there is no formal requirement, a good rule-of-thumb is the recommended digitizing visualization scale. For example, for a visualization scale of 1:10 000, a reasonable MVI would be 5 m, and 50 m for 1:100 000.

Figure 4 represents a 5 km × 3.3 km area of agricultural land (near Barrax, Spain) that shows the results of applying SCRM to a Normalized Difference Vegetation Index (NDVI) multitemporal image derived from three Landsat ETM+ images (resampled to 25 m) acquired respectively on 20 May (displayed as the red channel), 28 June (green) and 16 July 2000 (blue). The input parameters were DMS = 25 ha, MMU = 2 ha, and MVI = 50 m (no MAS restriction was enforced). Figure 5 shows SCRM results (DMS = 2.5 ha, MMU = 0.5 ha, MAS = 5 ha, MVI = 5 m) on a 1.25 km × 0.82 km sub-scene from a pan-sharpened 2.5 m pixel SPOT5 multispectral image (displayed as RGB bands 3, 2, 1) acquired on May 2005 near Sant Boi de Llobregat, Catalonia, Spain. Figure 6 represents a 1.4 km × 0.92 km sub-scene from a semi-natural landscape centered on Muskilda Hill, a 50 ha forest near Estella (Navarra, Spain)

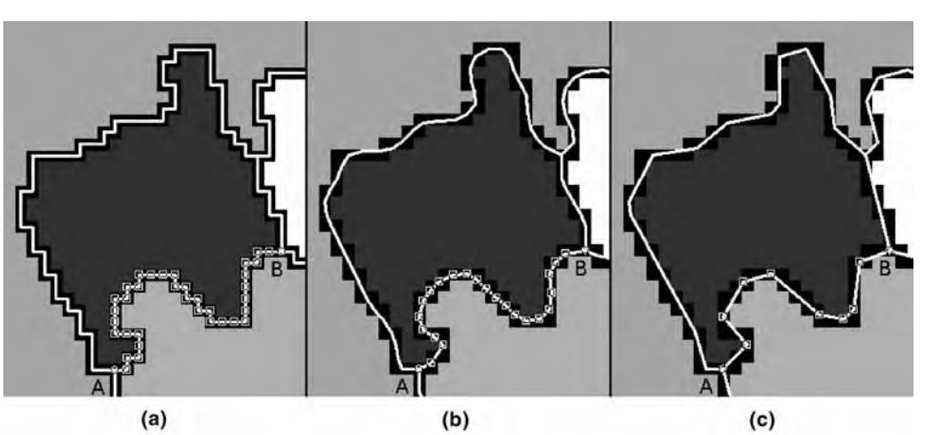

Figure 3. Vectorization sequence in a sample partition: (a) raw vector, (b) vector after spline interpolation, and (c) final vector after Douglas-Peucker simplification. The vertices of arc AB are highlighted.

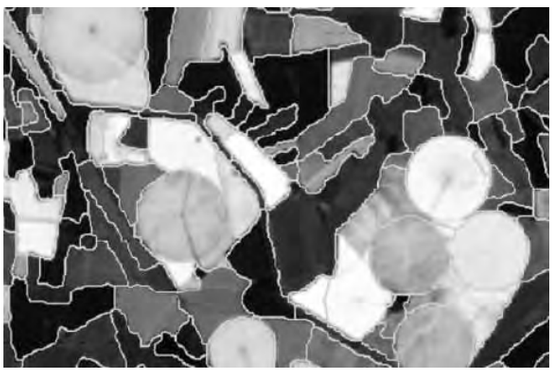

Figure 4. SCRM sample result in a NDVI multitemporal image derived from Landsat ETM (dimensions: 5 km × 3.3 km). A color version of this figure is available at the ASPRS website: **www.asprs.org.**

Figure 5. SCRM sample result in a SPOT5 multispectral (RGB 321) image (dimensions: 1.25 km × 0.82 km). A color version of this figure is available at the ASPRS website: **www.asprs.org.**

Figure 6. A 1.4 km × 0.92 km QuickBird (02) RGB 432 sub-scene showing Muskilda Hill, a 50 ha oak forest that requires digitizing. A color version of this figure is available at the ASPRS website: **www.asprs.org.**

covered by *Quercus faginea* trees and shrubs, and surrounded by vineyards, pastures, and ploughland. The color composite (RGB bands 4, 3, 2) is a 2.8 m pixel QuickBird-2 multispectral image acquired in September 2004. This is the base image we will use in Figures 7 through 9 to illustrate how SCRM may be used for photointerpretation.

Figure 7 shows SCRM results for DMS = 1.5 ha, MMU = 0.5 ha, and MVI = 10m (no MAS restriction). Imagine we want to compile a forest map with a minimum mapping unit of 2 ha. The partition produced from Figure 7 SCRM-inputs would be too profuse for this purpose. In addition, it has units, like the 1.4 ha rectangular patch of scrub to the left of the forest, that even if they were densely populated by trees, would not qualify as forest in this map. Figure 8 illustrates SCRM results for DMS = 6 ha, MMU = 2 ha, and MVI = 10m (no MAS restriction). Here, the aforementioned scrub no longer constitutes a separate unit, as it is smaller than the MMU. Also, some regions now form part of a different aggregate than in the previous partition, like the small area marked with an "X" in the right side of the forest in Figures 7 and 8. In Figure 7, this region had as its most similar neighbor, the scrub located below it. However, in Figure 8, the bright barren area surrounding it has previously merged with some slightly darker agricultural fields, lowering the average brightness of the aggregate, while at the same time the scrub has merged with the (also darker) forest, so that now the most similar neighbor is the barren area instead of the scrub.

The partition of Figure 8 is better suited than that of Figure 7 as an initial template to delineate our forest map, as it has no unit smaller than the specified 2 ha MMU. Therefore, the basic operation is to manually merge connected regions that in our opinion are forests. In this example, this would be done in a few seconds (in any GIS with editing capabilities) by hold-clicking and dragging with the left mouse button over the polygons we want to merge as to select them, and then right clicking to confirm the merge. The rectangle shown in Figure 8 represents an instance of such movement that would produce the merging of the eight polygons within Muskilda Hill. Next, we would need to correct several small areas along the perimeter of the newly formed polygon that look similar to the forest but that are actually scrub with less trees than would qualify as forest (Figure 9). This is the case of the two patches lying at

Figure 7. SCRM results applied to Muskilda Hill (Figure 6) using DMS = 1.5 ha, MMU = 0.5 ha, and MVI = 10 m. A color version of this figure is available at the ASPRS website: **www.asprs.org.**

Figure 8. SCRM result used as template for the example in Figure 9 (DMS = 6 ha, MMU = 2 ha, and MVI = 5 m). A color version of this figure is available at the ASPRS website: **www.asprs.org.**

Figure 9. Final delineation of Muskilda forest. Only arc segments AB, CD, EF, and GH have been digitized manually. A color version of this figure is available at the ASPRS website: **www.asprs.org.**

both extremes of the lower half of Muskilda, and also of its northeast corner. Finally, there is a gravel pit surrounded by a thin corridor of trees in the upper right half of Muskilda that we have decided not to retain within the forest because of the narrowness of the corridor. All these operations can be done with a stream digitizing tool that manually splits the undesired parts into separate polygons, where the latter would be subsequently merged to the surroundings using the above standard procedure. Figure 9 illustrates how Muskilda forest would look in the final map. In this example, less than ten percent of the polygon (those arc segments that appear labeled) has been delineated manually.

Discussion

SCRM Strengths

After having applied SCRM to many different images, none of the output partitions produce a visual impression of a *bad segmented* image. This is to say that for most polygons, there seems to be some sense of cohesiveness throughout the area enclosed within the polygon, and contrariwise, some sense of discontinuity across the boundary between the polygon and its neighbors. Also, since our vectorization scheme leaves no trace of jagged pixel boundaries, the general impression the vector arcs provide is that of a human-made delineation. Well-defined regions in the image are invariably cleanly delineated, providing they are larger than the specified MMU. If one considers only radiometric aspects and neglects semantic ones, then there are very few segmentation errors; and, those that apparently are errors can be explained after changing the image enhancement settings used for display. This is due to the fact that perceived chromatic differences are not isometric to the (Euclidean) distance employed within SCRM, which in turn is provoked by the differential sensitivity of the human eye to different wavelengths and luminance (Kaiser and Boynton, 1996). With regards to the level of cartographic generalization applied to the image (be it expressed by the size distribution of polygons or the edge complexity) which can be easily and explicitly controlled through SCRM input parameters.

SCRM Limitations

Notwithstanding the above strengths, SCRM results are not perfect neither are they intended to provide a final solution for manual digitization. SCRM uses only one criterion, i.e., radiometric distance between region centroids (mean value of inner pixels), whereas the interpreter uses a wide-ranging suite of criteria, and resorts to external information (experience and ancillary data). An underlying premise of any segmentation method is that radiometric similarity and semantic similarity go hand in hand. Wherever this assumption is invalid, segmentation results will go awry. Whenever the interpreter unconsciously uses Gestalt principles (e.g., proximity, good continuation, etc.) to complete a polygon, the corresponding automated delineation will, in all likelihood have to be redrawn. The same can be said of *fiat* (imposed) boundaries separating continua in transitional zones. In short, the rationale of using SCRM segmentation as a guiding template for photointerpretation is that it can save time. The automated polygons must require little adjustment for time savings to occur. Depending on both the complexity of the imaged landscape, and the degree of fit between radiometry and semantics, there will be parts of the scene where the interpreter would have digitized polygons very akin to those generated by SCRM; whereas in other parts the automated polygons may even be a hindrance. If both situations are spatially segregated, then the automated vector layer is still worthwhile, as one can easily discard automatic results in complex areas, by adopting the described manual merging procedure. If this is not the case and *bad* and *good* polygons are jumbled, the point at which vetting of the automated polygons requires more time than the standard manual delineation is an important consideration that will certainly be addressed in future research.

We also have some concerns regarding robustness and consistency. There are three user-independent choices that are somewhat arbitrary and that impact these two issues: namely (a) the filter diffusivity parameter and convergence criterion, (b) the dissimilarity measure, and (c) the intermediate stop criterion. The first affects the initial watershed partition, and the other two influences the merging order. In general, the shape of high contrasted regions does not suffer significant changes by varying these internal criteria, and the same can be said for strong edges. But weak edges are combined in very different ways, producing disparate regions. A similar situation would occur also for human interpreters, whose individual interpretations of low contrasted areas are likely to differ too. Although this problem is common to any segmentation method

(e.g., Baatz and Schape, 2000), it deserves further research. In the end, such inconsistencies actually arise because a given scene may be represented in many different ways, and often, none of them can be said to be strictly preferred to the others. Perhaps for this reason, visual verification remains the basic evaluation procedure for newly developed algorithms, although there are some empirical methods (Zhang, 1996) that try to mitigate the inevitable subjectiveness of the evaluation. Since each individual and institution has different interests, conceptions and methods, they may hold different views of the same scene. Therefore, we concur with Openshaw (1978) in rejecting the premise of objectivity in the design of zoning systems of the like of thematic maps.

Conclusions and Future Work

In this paper, we have introduced Size Constrained Region Merging (SCRM), a novel segmentation method that may be used for computer-assisted photointerpretation. SCRM transforms a single or multi-channel ortho-image into a polygon vector layer (shapefile) that may be used by the interpreter as an initial template. After having applied SCRM to many images, none of the output partitions produce a visual impression of a *bad segmented* image. Furthermore, compared to other segmentation algorithms embedded in already available commercial software, SCRM is less demanding computationally, only requires intuitive user parameters explicitly related to the size of output regions and the accuracy of their boundaries, and tackles the fractal nature of landscape and its hierarchical structure in a conceptually coherent manner. Moreover, SCRM is grounded on a solid conceptual basis (Castilla, 2003), an asset that many segmentation algorithms lack. Notwithstanding, a thorough comparison with other segmentation methods is desirable in order to fully evaluate the pros and cons of our procedure. In particular, the usability of SCRM vectors (with regard to time savings during the interpretation process) needs to be confirmed by additional empirical studies, which we will address in future work. In addition, we intend to investigate new filtering options where diffusivity is self-controlled by the evolving spatial structure of the image. Also, we will test new dissimilarity distances that include (a) some boundary saliency metric, so that the merging of similar regions separated by a strong edge (like a narrow road separating two fields with the same crop) is precluded, and (b) some measure of the internal texture of regions, possibly at different scales. Finally, we intend to embed SCRM in a multi-scale framework, and note that Multi-scale Object Specific Segmentation (MOSS; Hay *et al.*, 2005) is a first attempt that has produced encouraging results. In MOSS, SCRM parameters are automatically derived from the original image and their subsequent upscale versions using a proprietary procedure of the second author (Hay and Marceau, 2004). The result is a family of increasingly generalized representations of the imaged scene that may have utility in multi-scale habitat mapping studies.

Acknowledgments

The authors wish to thank Geosys (**www.geosys.es**), IDR (**www.idr-ab.uclm.es**), and ICC (**www.icc.es**) for providing the satellite images appearing in the paper. This research has been generously supported in part by a University of Calgary Start-up Grant, and an Alberta Ingenuity New Faculty Grant to Dr. Hay. The opinions expressed here are those of the Authors, and do not necessarily reflect the views of their funding agencies.

References

Bleau, A., and L.J. Leon, 2000. Watershed-based segmentation and region merging, *Computer Vision and Image Understanding*, 77:317–370.

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

Beaulieu, J.M., and M. Goldberg, 1989. Hierarchy in picture segmentation: A stepwise optimisation approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:150–163.

Beucher, S., and C. Lantuejoul, 1979. Use of watersheds in contour detection, *Proceedings of the International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation*, September, Rennes, France.

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

Blaschke, T., and G.J. Hay, 2001. Object oriented image analysis and scale-space: Theory and methods for modeling and evaluating multiscale landscape structure, *International Archives of Photogrammetry and Remote Sensing*, 34(4):22–29.

Blaschke, T., S. Lang, E. Lorup, J. Strobl, and P. Zeil, 2000. Object-oriented image processing in an integrated GIS/remote sensing environment and perspectives for environmental applications, *Environmental Information for Planning, Politics and the Public* (A. Cremers and K. Greve, editors), Vol. 2, Verlag, Marburg, pp. 555–570.

Booch, G., 1991. *Object-Oriented Design with Applications*, The Benjamin/Cummings Publishing Company, Inc., Redwood City, California.

Castilla, G., 2004. Size-constrained region merging: A new tool to derive basic land-cover units from remote sensing imagery, *Theory and Applications of Knowledge Driven Image Information Mining with Focus on Earth Observation*, ESA SP- 553, pp. 54–62.

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

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, Madrid, Spain, URL: http://oa.upm.es/133/01/07200302_castilla_castellano.pdf (last date accessed: 15 January 2008).

Chen, Q., C. Zhou, J. Luo, and D. Ming, 2004. Fast segmentation of high-resolution satellite images using watershed transform combined with an efficient region merging approach, *Lecture Notes in Computer Science*, Vol. 3322, pp. 621–630.

Cox, I.J., 1993. A review of statistical data association techniques for motion correspondence, *International Journal of Computer Vision*, 10(1):53–66.

Definiens, 2005. Definiens AG, Geospatial Solutions München, Germany. URL: http://www.definiens-imaging.com/ecognition/ (last date accessed: 28 December 2007).

Douglas, D., and T. Peucker, 1973. Algorithms for the reduction of the number of points required to represent a line or its caricature, *The Canadian Cartographer*, 10(2):112–122.

Fjørtoft, R., A. Lopès, P. Marthon, and E. Cubero, 1998. An optimal multiedge detector for SAR image segmentation, *IEEE Transactions on Geoscience and Remote Sensing* 36(3):793–802.

Gauch, J.M., 1999. Image segmentation and analysis via multiscale gradient watershed hierarchies, *IEEE Transactions on Image Processing*, 8(1):69–79.

Goodchild, M.F., 1994. Integrating GIS and remote sensing for vegetation analysis and modelling: Methodological issues, *Journal of Vegetation Science*, 5:615–626.

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

Haris, K., S.N. Efstratiadis, and A.K. Katsaggelos, 1998. Hybrid image segmentation using watersheds and fast region merging, *IEEE Transactions on Image Processing*, 7(12):1684–1699.

Hay, G.J., and G. Castilla, 2008. Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline, *Object-Based Image Analysis–Spatial Concepts for Knowledge-driven Remote Sensing Applications* (T. Blaschke, S. Lang, and G.J. Hay, editors), Springer-Verlag, in press.

Hay, G.J., G. Castilla, M.A. Wulder, and J.R. Ruiz, 2005. An automated object-based approach for the multiscale image segmentation of forest scenes, *International Journal of Applied Earth Observation and Geoinformation*, 7:339–359.

Hay, G.J., and D.J. Marceau, 2004. Multiscale object-specific analysis (MOSA): An integrative approach for multiscale landscape analysis, *Remote Sensing and Digital Image Analysis, Including the Spatial Domain* (S.M. de Jong and F.D. van der Meer, editors), Kluwer Academic Publishers, Dordrecht, pp 71–92.

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

Hay, G.J., D.J. Marceau, A. Bouchard, and P. Dubé, 2001. A multi-scale framework for landscape analysis: Object-specific upscaling, *Landscape Ecology*, 16:471–490.

Jackway, P.T., 1996. Gradient watersheds in morphological scale-space, *IEEE Transactions on Image Processing*, 5(6):913–921.

Kaiser, P.K., and R.M. Boynton, 1996. *Human Color Vision*, Optical Society of America, Washington D.C.

Leckie D.G., M.D. Gillis, F. Gougeon, M. Lodin, J. Wakelin and X. Yuan, 1998. Computer-assisted photointerpretation aids to forest inventory mapping: Some possible approaches, *Proceedings of the International Forum on Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry*, 10–12 February, Victoria, British Columbia, Canadian Forest Service, Pacific Forestry Center, Victoria, B.C., pp. 335–343.

Lindeberg, T., 1994. Scale-space theory: A basic tool for analysing structures at different scales, *Journal of Applied Statisitics*, 21(2):225–270.

Marr, D., 1982. *Vision*, Freeman Publishers, San Francisco.

Meyer, F., 2001. An overview of morphological segmentation, *International Journal of Pattern Recognition and Artificial Intelligence*, 15(7):1089–1118.

Najman, L. and M. Schmitt, 1996. Geodesic saliency of watershed contours and hierarchical segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1163–1173.

Olsen, O.F., and M. Nielsen, 1997. Multi-scale gradient magnitude watershed segmentation, *Lectures Notes in Computer Science*, Vol. 1310, pp. 6–13.

Openshaw, S., 1978. An empirical study of some zone-design criteria, *Environment and Planning*, Vol. 10, pp. 781–794.

Perona, P., and J. Malik, 1990. Scale space and edge detection using anisotropic diffusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(629):639.

RSINC, 2005a. RSI, Boulder, Colorado, URL: http://www.rsinc.com/idl, (last date accessed: 28 December 2007).

RSINC, 2005b. RSI, Boulder, Colorado, URL: http://www.rsinc.com/envi, (last date accessed: 28 December 2007).

Smith, B., 1995. The structures of common-sense world, *Acta Philosophica Fennica*, 58:290–317.

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

Thom, R., 1975. *Structural Stability and Morphogenesis: An Outline of a General Theory of Models*, Benjamin-Cummings Publishing.

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

Vincent, L., and P. Soille, 1991. Watersheds in digital spaces: An efficient algorithm based on immersion simulations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598.

Wang, D.C., A.H. Vagnucci, and C.C. Li, 1981. Gradient inverse weighted smoothing scheme and the evaluation of its performance, *Computer Graphics and Image Processing*, 15:167–181.

Weickert, J., 1998. Fast segmentation methods based on partial differential equations and the watershed transformation, *Mustererkennung 1998* (P. Levi, R.J. Ahlers, F. May, and M. Schanz, editors), Springer, Berlin.

Woodcock, C.E., and V.J. Harward,1992. Nested-hierarchical scene models and image segmentation, *International Journal of Remote Sensing*, 13(16):3167–3187.

Woodcock, C.E., and A.H. Strahler, 1987. The factor of scale in remote sensing, *Remote Sensing of Environment*, 21:311–332.

Wu, J., and O.L. Loucks, 1995. From balance-of-nature to hierarchical patch dynamics: A paradigm shift in ecology, *Quarterly Review of Biology*, 70:439–466.

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

(Received 28 March 2006; accepted 20 June 2006; revised 18 August 2006)