OBJECT-ORIENTED IMAGE ANALYSIS AND SCALE-SPACE: THEORY AND METHODS FOR MODELING AND EVALUATING MULTISCALE LANDSCAPE STRUCTURE

Thomas Blaschke[1] and Geoffrey J. Hay[2]

[1] *Department of Geography and Geoinformation, University of Salzburg, Austria, Hellbrunner Str. 34, A-5020 Salzburg, Austria. Email: thomas.blaschke@sbg.ac.at*

[2] *Geocomputing Laboratory, Département de géographie, Université de Montréal, C.P. 6128, succursale Centre-Ville, Montréal, Québec, Canada, H3C 3J7. Email: ghay@sympatico.ca*

KEY WORDS: Object-oriented, image analysis, multiscale, landscape structure

ABSTRACT

Landscapes are *Complex Systems*, which by their very nature necessitate a multiscale approach in their monitoring, modeling and management. To assess such broad extents, remote sensing technology is the primary provider of landscape sized data sets, and while tremendous progress has been made over the last thirty years in terms of improved resolution, data availability, and public awareness, the vast majority of remote sensing analytical applications still rely on basic image processing concepts: in particular, per-pixel classification in multi-dimensional feature space. In this paper we describe and compare two technically and theoretically different image processing approaches, both of which facilitate multiscale pattern analysis, exploration, and the linking of landscape components based on methods that derive spatially-explicit multiscale contextual information from a single resolution of remote sensing imagery. Furthermore, we suggest how both methods may be integrated for improved results.

1 BACKGROUND

To fully exploit the information content of high-resolution (H-res) images, methodologies are required which go beyond traditional statistical analysis and the classification of individual pixels. In particular, Blaschke and Strobl (2001) argue that classic per-pixel classification approaches do not explicitly make use of the spatial characteristics inherent within the image. For example, in H-res images it is extremely likely that neighboring pixels belong to the same land-cover class as the pixel under consideration. Consequently, Blaschke and Strobl and others (Hay et al. 1994, 2001a, b, Schiewe et al. 2001) argue for the classification of homogeneous groups of pixels that reflect real-world objects of interest, and the use of algorithms to delineate such 'image-objects' based on contextual image information i.e., image-texture (Hay et al 1996, 1997) or fractal dimensions.

Like the landscape itself, defining patches that comprise the landscape is not a self-evident task. Instead, patches must be defined relative to a given situation. For example, from a timber management perspective, a patch may correspond to a forest stand; however, the stand may not function as a patch from a particular organism's perspective. From an ecological perspective, patches represent relatively discrete areas (spatial domain) or periods (temporal domain) of relatively homogeneous environmental conditions. Their boundaries are distinguished from their surroundings by discontinuities in environmental character states of magnitudes that are perceived by, or relevant to, the organism or ecological phenomenon under consideration. From an organism-centered view, patches may be defined as environmental units where habitat quality differs more between them than within them. As a result, a necessary, yet non-trivial task in landscape analysis involves delineating appropriate and meaningful patches. Difficulty in this task exists because there are an infinite number of solutions for defining lines/polygons (i.e., patches) on a relatively continuous surface. While, manual interpretation of patches is common and well established, as monitoring needs increase, results tends to be expensive and require more time to produce. Furthermore, maps generated by different interpreters often provide limited – or poorly comparable - results.

Based on these concepts, the underlying goal of this short paper is to report our findings regarding our efforts to combine and integrate components of the *fractal net evolution approach (FNEA)* with the *Scale-Space (SS)* approach for analyzing and classifying a H-res IKONOS satellite image. More specifically, we intend to evaluate whether *Scale-Space* can be used to automatically detect the most relevant/appropriate levels of information (i.e., patches, scales) required by the multi-scale segmentation approach used in FNEA.

2 MATERIAL AND METHODS

2.1. Study site

The data used in the FNEA study is a 1000x1000 pixel sub-image (4m resolution) of an 11-km^2 IKONOS image that was acquired in August 2001. Geographically, it represents a portion of the highly fragmented agro-forested landscape typical of the Haut Saint-Laurent region of southwest Quebec, Canada. Due to the memory and computational demands required by SS, the image used in the SS study portion, represents a 500x500 pixel sub-section from the same near-infra red image used for FNEA analysis. In addition, one hundred individual SS levels we generated and evaluated.

2.2. Fractal Net Evolution Approach (FNEA)

The fractal net evolution approach incorporates an object-oriented framework and image segmentation techniques. In particular, it utilizes fuzzy set theory to extract the objects of interest, at the scale of interest, by segmenting images simultaneously at both fine and coarse scales, then building image semantics between levels and their elements. The principal challenge and flexibility of this multiscale approach lies in defining the aggregation rules for the lower level entities, which result in improved image classifications, and a new framework for integrating semantic rules in image processing. Contrary to neural net classifiers, a semantic network based on fuzzy set theory is not a black box, instead it allows for a transparency of all classification steps. Although, FNEA is already embedded in a commercial software environment, its usability is not operational as long as a theoretical framework is not provided and users have to find useful segmentation levels in a 'trial and error' style.

Since its recent introduced by Baatz and Schäpe (2000), it has been applied in various research projects in Europe (Blaschke et al 2000, Blaschke and Strobl 2001, Schiewe et al 2001). For example, Blaschke et al. (2001) reported on the applicability of FNEA to distinguish different phases of bush and shrub encroachment in an old pasture-dominated landscape in Germany. This involved a multi-scale delineation of images and image semantics that incorporated image-structure and -texture characteristics. It was found that they could distinguish three levels of delineation appropriate for three different key species, which resulted in the construction of a hierarchical network of image-objects, and semantic rules between these levels. Additional studies have demonstrated the potential of this multi-scale segmentation approach, where the results of unsupervised delineations of landscape objects are surprisingly appealing. The 'realistic' look of the resulting patches of forests, pastures, fields and built-up areas has motivated several agencies in Europe to evaluate the applicability of the approach and the commercial software environment. Experienced image interpreters have already expressed their concern of becoming obsolete one day.

In a remote sensing scene, image information may be considered as fractal in nature. That is, structures typically appear on different scales in an image simultaneously. However, to extract meaningful image regions the user has to take into account the scale of the problem that is to be solved and the type of image data. With the FNEA, users are required to aim for different scale levels by hypothesizing that almost all attributes of image structure – color, texture or form – are essentially scale-dependent. This is different from many other approaches, which do not require any user-defined parameters (i.e., region growing algorithms, watershed algorithms, multi-fractal based segmentation, Markov random fields, etc). In FNEA, defining a specific level of analysis leads to objects at a certain scale. This means, an image-object can easily be 10

times the size of a small object but it is very unlikely that (at the same semantic level of consideration) an image object is 1000 times the size of another object.

FNEA starts with single a single pixel and a pairwise comparison of its neighbors with the aim to minimize the resulting summed heterogeneity. The common solution for this pairwise cluster problem is described as *global mutual best fitting*. In fact, global mutual best fitting is the strongest constraint for the optimization problem and it reduces heterogeneity primarily over the scene following a pure quantitative criterion. However, there is a significant disadvantage to global mutual best fitting. It does not use the distributed treatment order and – in connection with a heterogeneity definition for color - builds initial segments in regions with a low spectral variance. This leads to an uneven growth of image-objects over a scene and to an unbalance between regions of high and low spectral variance. Local mutual best fitting always performs the most homogeneous merge in the local vicinity following the gradient of best fitting. An iterative heuristic optimization procedure aims to get the lowest possible overall heterogeneity across an image. The basis for this is the degree of difference between two regions. As this difference decreases, the fit of the two regions is said to be closer. These differences are optimized in a heuristic process by comparing the attributes of the regions (Baatz and Schäpe 2000). That is, given a certain feature space, two image-objects are considered similar when they are near to each other in this feature space. For a *d*-dimensional feature space the heterogeneity *h* is described as :

$$h = \sqrt{\sum_d (f_{1d} - f_{2d})^2}$$

Equation 1

Examples for appropriate object features are, for instance, mean spectral values or texture features, such as the variance of spectral values. These distances can be further standardized by the standard deviation of the feature in each dimension using Equation 2.

$$h = \sqrt{\sum_d \left(\frac{f_{1d} - f_{2d}}{\sigma_{fd}} \right)^2}$$

Equation 2

Equation 3 defines the homogeneity of two adjacent regions by describing the difference of heterogeneity *h* of the two regions before (*h1* and *h2*) and after a virtual merge (*hm*). Given an appropriate definition of heterogeneity for a single region, the growth of heterogeneity in a merge should be minimized. There are different possibilities for describing the change of heterogeneity *hdiff* before and after a virtual merge – but they are beyond the scope of this paper.

Equation 3 $hdiff = \text{hm} - (\text{h1} + \text{h2})/2$

We note that this attribute allows us to distinguish between two types of objects with similar mean reflectance values but different 'within-patch heterogeneity'. One application based on this type of heterogeneity was described by Blaschke et al (2001) utilizing '*mean spectral difference between all sub-objects*" as one manifestation of heterogeneity applied to pastures and conservation aspects of changes in a cultural heritage landscape in central Germany.

2.3. Scale space

The second multiscale approach is composed of two principle components. The first is referred to as *Linear Scale-Space* (Lindeberg 1994). *Linear Scale-space* (SS) is an uncommitted framework for early visual operations that was developed by the computer vision community to automatically analyze real-world structures at multiple scales – specifically, when there is no *a priori* information about these structures, or the appropriate scale(s) for their analysis (see Hay et al, 2001a for a detailed description). Its basic premise is that a multi-scale representation of a signal (such as a remote sensing image of a landscape) is an ordered set of derived signals showing structures at coarser scales that constitute simplifications of corresponding structures at finer scales.

In practice, Gaussian filters are applied to an image at a range of kernel sizes resulting in a scale-space cube or 'stack', where each layer in the stack represents convolution at a specific scale (Figure 1). The use of Gaussian filters is essential to SS theory as they satisfy necessary conditions for an uncommitted framework. These include linearity (i.e., no knowledge, no model, no memory), spatial shift invariance (i.e., no preferred location), isotropy (i.e., no preferred orientation) and scale invariance (i.e., no preferred size or scale). In addition, a Gaussian kernel satisfies the linear diffusion equation, thus Gaussian smoothing can be considered as the diffusion of the grey-level intensity over scale (*t*), instead of time.

Figure 1. Colorized Grey-level stack

The second SS component is referred to as *Scale-Space Blob Feature Detection* (Hay et al 2001a). The primary objective of this non-linear approach is to link structures at different scales in scale-space, to higher-order objects called 'scale-space blobs', and to extract significant features based on their appearance and persistence over scales. The main features that arise at each scale within a stack are smooth regions, that are brighter or darker than the background and which stand out from their surrounding. These regions are referred to as 'grey-level blobs' (Figure 2a). When blobs are evaluated as a volumetric structure within a stack, it becomes apparent that some structures visually persist through scale, while others disappear (Figure 1). Thus, an important premise of SS is that blobs-like structures which persist in scale-space, are likely candidates to correspond to significant structures in the image, and thus in the landscape.
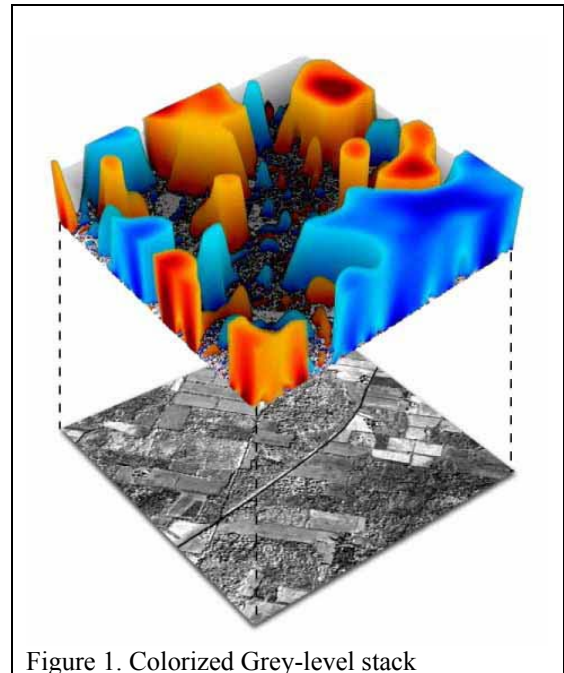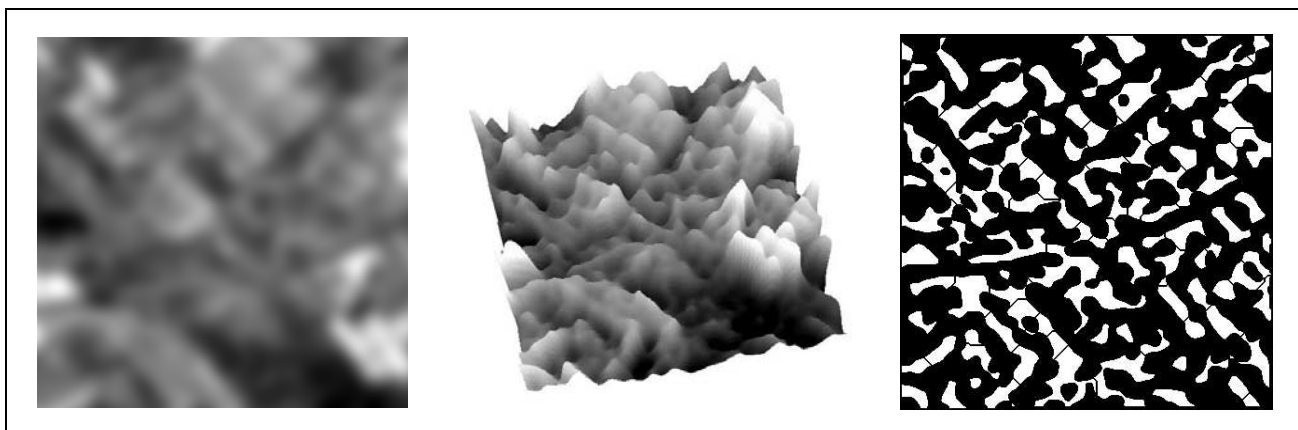

Figure 2a. 2D Grey-level blob (t 20)   2b. 3D Grey-level blobs (t 20)   2c. Binary blob (t 20)

In simple terms, grey-level blobs at each scale in the stack are treated as objects with extent both in 2D space (x, y) and in grey-level (z-axis) – thus 3D. Grey-level blob delineation may best be defined with a watershed analogy. At each scale in the stack, the image function of all blobs may be considered as a flooded 3D landscape (i.e., a watershed, see Figure 2b). As the water level gradually sinks, peaks will appear. At some instance, two different peaks become connected. The corresponding 'connected' elevation levels are called the 'base level' of the blob. They are used for delimiting the 2D spatial extent or 'region of support' of each blob - defined as a binary blob (Figure 2c). 2D binary blobs at all scales are then combined within a new stack to create 3D hyper-blobs (Figure 3a). Within a single hyper-blob there are four primary types of visible structures or 'bifurcation events': *annihilations* (A), *merges* (M), *splits* (S), and *creations* (C) (Figure 3b).

The ability to define these SS-events is a critical component of SS, as scales between bifurcations are linked together forming the lifetime (Lt_n) and structure of individual SS-blobs. Next, the integrated (4D) volume (x, y, z, t) of each individual SS-blobs is defined. These resulting 4D volumes are then ranked, and an arbitrary number of significant SS-blobs are defined, from which the scale (t) representing the maximum 3D grey-level blob volume (x, y, z) of each hyper-blob is extracted. From these layers the 2D spatial support (i.e., binary blob) is identified and related back to the corresponding structures in the image for further examination (Figure 3c). Thus, based on the underlying initial premise, 4D scale-space blobs are simplified to 3D grey-level blobs, which are further simplified to their 2D support region (x, y), and then to their corresponding 'significant' real-world objects in the original image.
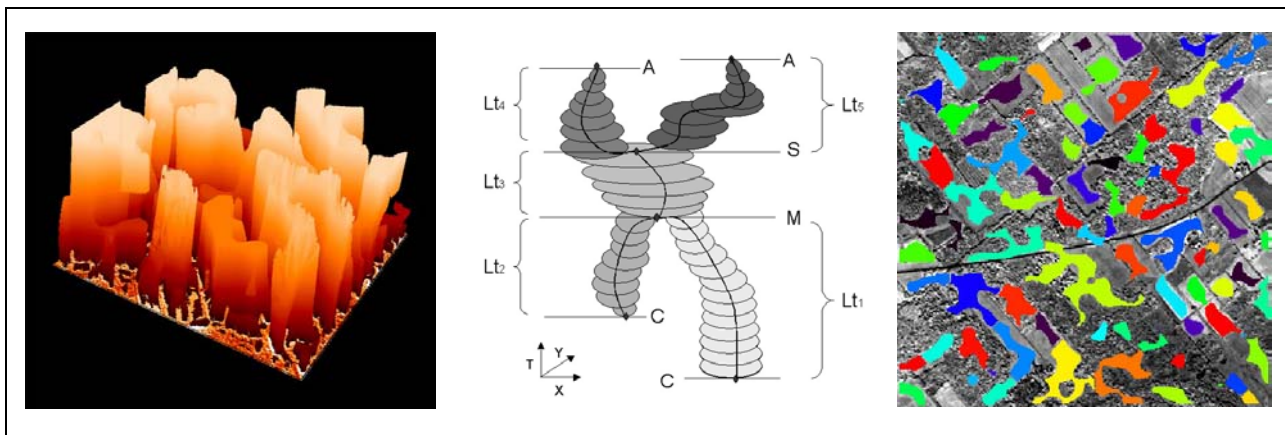

Figure 3a. Hyper-blob stack   3b. Idealized hyper-blob   3c. Ranked blobs overlaid study site

At present, our SS analysis is performed exclusively within a raster environment. We are currently working on converting raster blob layers to vector datasets with their accompanying attribute lists that will include, the identification number and rank of all 2D blobs, their scale of expression, their spatial boundary (i.e., polygon dimensions), and the area of each ranked blob-polygon. As a result, we hypothesize that the scale of expression of significant scale-space blobs may be used as early visual operators to automatically define and or refine the aggregation semantics of the FNEA. In generic terms this refinement process is referred to as '*feature localization*' (Lindeberg 1999).

3 RESULTS AND DISCUSSION

Applying FNEA to the IKONOS data resulted in several levels, which are expressed by their 'object scales'. These object scales represent the mean heterogeneity the user allows for the resulting objects at a certain level, and corresponds to the mean object size. From the many levels produced, three levels were selected for further analysis and classification of the image. Visual interpretation (Figure 4 a-c) proves all three levels to be meaningful at their respective levels. Fig. 4b shows an image scale of 60 and corresponds very well to the

target level of forest stands, agricultural parcels and sealed surfaces. The latter are problematic because roads can be very long and can reach an extreme perimeter-area ratio. Due to the applied heuristics and fractal theory, these segments cannot be too long or they are sometimes cut off. This is one of the reasons for a level 'above' including the related super-objects. In Figure 4, the landscape is primarily segmented into objects that already include some generalization. As forest patches get larger, they include small forest gaps and different textures related to tree heights, etc. Although this level (i.e., 60) is relatively coarse, clearly different land cover classes are still separated; however, within an 'agriculture' super-class, fields start to grow together to form larger patches. Although the object scale 60 was chosen to best fit the target level, a much finer segmentation (Figure 4a) is very important for the classification. While, this level can be regarded as being 'over segmented', the information of the 'within' patch heterogeneity allows the expression of texture based on sub-objects. By incorporating single entities within patches such as groups of shrubs, hedges or forest gaps, and their corresponding information about size, shape, color or relative location within their super-objects, the texture of an image-object is expressed more explicitly than using moving window style texture filters.

The most innovative aspect beyond a simple improvement of image classification is the potential to differentiate different 'object-classes' within the same image 'on-demand' for different applications. Contrary to the static view of a map, all forest areas could be treated as relatively homogeneous areas (although in reality they aren't) and open grasslands can be explored in further detail (see Blaschke et al 2001) or vice versa. Figure 5 illustrates three ways to express texture. In 5a, the standard deviation per object is calculated for

Figure 5. Analysis examples, object parameter 60, sub objects stem from object parameter 20.
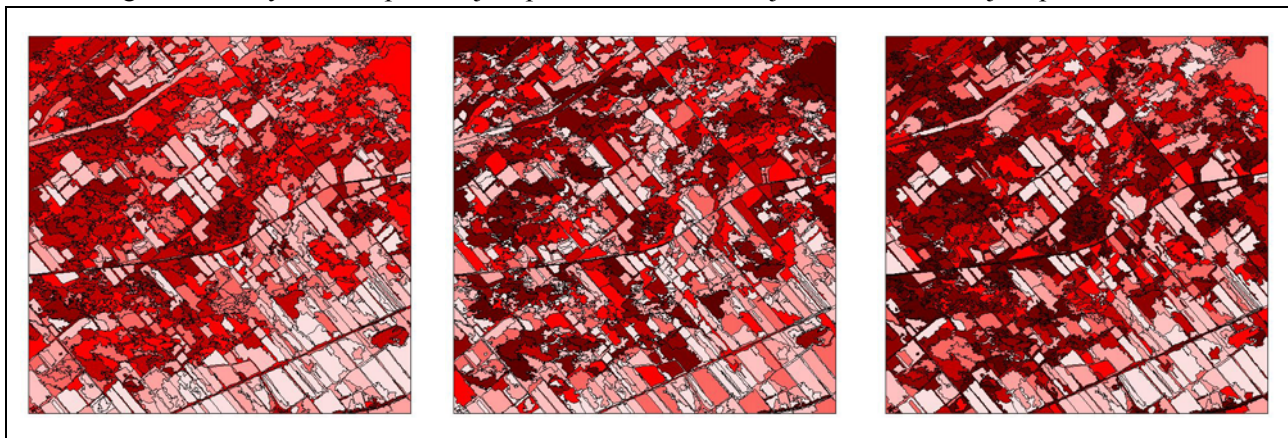
5a. Avg-std of all bands per object   5b. No# of sub object at level below   5c. Relative area of sub object

each band and the average of all 4 bands is shown. The darker areas represent high values and include most forest stands; while in 5b and 5c, the results are based on sub-objects. The general patterns of 5a, b and c coincide with minor exceptions. In ongoing studies, we are specifically interested in these resulting differences i.e., between 'traditional' ways to express texture, based primarily on the spectral variation of pixels and the spectral differences of sub-objects within super-objects (i.e., 5b and c). Initial results suggest further research on expressing the 'graininess' of sub-objects.

Figure 6 shows a certain level of generalization similar to what human interpreters intuitively incorporate in their work. In 6a, single trees within fields are subsumed within the object 'agricultural area'. If the single tree is of interest, it still exists at the level below and is logically a sub-object with information relating to its size, shape, relative position within its super-object, and spectral differences to its neighbors. Figure 6b shows that a large object can border to small objects like hedgerows. This does not contradict the hypothesis that all resulting objects have to be a similar dimension. In reality, a large and relatively uniform forest patch or a large field may exist next to a small yet distinct patch like a road. As long as the resulting heterogeneity per object is within a certain range, an object can even have 500 times the size of the smallest objects. Nor-

mal region growing segmentation algorithms were compared in Blaschke et al (2000) and did not result in similar semantic objects.
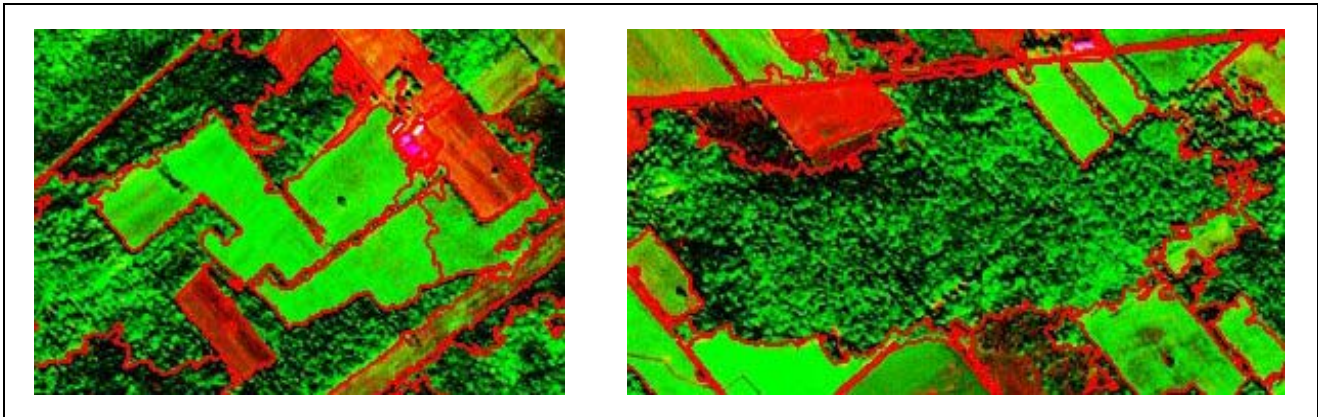

Figure 6a. Close up of fields at an object scale 140   6b. Close up of forest stands at an object scale 140

4 CONCLUSIONS

In this paper, we describe, apply and compare two technically and theoretically different multiscale approaches with a common theme: The focus on intrinsic pattern and their exploration at different scales simultaneously. The key characteristic of these multiscale approaches is their evolution beyond individual pixel analysis to analyzing the explicit contextual structure of landscape components. In this work, we have begun to examine the sensitivity of both methodologies to landscape structure, and provide a case study were both approaches are compared. While the methodological comparison is technically interesting, the overall goal of this study is to contribute to a more coherent understanding of landscape structures, their representation in images, and their linking through scales. Both authors independently started from paradigms that most remote sensing and GIS methodologies do not readily support; that is, the representation of geographic entities at a variety of scales and levels of abstraction, simultaneously within a single image. Both approaches incorporate a *bottom-up* approach designed to convert lower level observational data into higher-level geographic entities, which - ideally - can then be related to the morphological processes that formed them. Our current results suggest that we can directly link both approaches by exporting SS-blob vectors (and their associated attributes) and using them as *apriori* information in the FNEA.

REFERENCES

Baatz, M. and Schäpe, A., 2000. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. In: Strobl, J. and Blaschke, T. (Eds.): Angewandte Geogr. Informationsverarbeitung XII, Wichmann, Heidelberg, pp. 12-23.

Blaschke, T. and Strobl, J., 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. GIS – Zeitschrift für Geoinformationssysteme 6/2001, pp. 12-17.

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

Blaschke, T., Conradi, M., Lang, S., 2001. Multi-scale image analysis for ecological monitoring of heterogeneous, small structured landscapes. Proceedings of SPIE, Toulouse.

Hay, G. J., P. Dubé, A. Bouchard, and D. J. Marceau. 2001a. A Scale-Space Primer for Exploring and Quantifying Complex Landscapes. Ecological Modelling. Accepted. April, 2001

Hay, G. J., D. J. Marceau, P. Dubé, and A. Bouchard, 2001b. A Multiscale Framework for Landscape Analysis: Object-Specific Analysis and Upscaling. Landscape Ecology (In Press, October, 2001).

Hay, G. J., K. O. Niemann, and D. G. Goodenough, 1997. Spatial Thresholds, Image-Objects and Upscaling: A Multi-Scale Evaluation. Remote Sensing of Environment, 62: 1-19.

Hay, G. J., K. O. Niemann, and G. McLean, 1996. An Object-Specific Image-Texture Analysis of H-Resolution Forest Imagery. Remote Sensing of Environment, 55: 108-122.

Hay, G. J., and K. O. Niemann, 1994. Visualizing 3-D Texture: A Three Dimensional Structural Approach to Model Forest Texture. (Cover Article) Canadian Journal of Remote Sens. Vol. 20, No.2, pp. 90-101.

Lindeberg, T. 1994. Scale-Space Theory in Computer Vision. Kluwer Academic Publishers, Dordrecht, Netherlands. 423 pp.

Lindeberg, T. 1999. Principles for Automatic Scale Selection, Technical report ISRN KTH NA/P--98/14--SE. In: B. Jähne (et al. eds.), Handbook on Computer Vision and Applications, volume 2, pp 239--274, Academic Press, Boston, USA.

Schiewe, J., Tufte, L., Ehlers, M., 2001. Potential and problems of multi-scale segmentation methods in remote sensing. GIS – Zeitschrift für Geoinformationssysteme 6/2001, pp. 34-39.