

HIERARCHICAL OBJECT REPRESENTATION – COMPARATIVE MULTI-SCALE MAPPING OF ANTHROPOGENIC AND NATURAL FEATURES

Stefan Lang ^{a,*}, Thomas Blaschke ^a

^aLandscape Analysis and Resource Management Research Group (LARG),
Department of Geography and Geoinformatics, University of Salzburg, 5020 Salzburg, Austria–
(stefan.lang, thomas.blaschke)@sbg.ac.at

KEY WORDS: Object-related modelling, hierarchical representation, anthropogenic vs. natural features, target dimension, multi-scale analysis.

ABSTRACT:

Hierarchical feature representation through multi-scale segmentation offers new possibilities in mapping complex systems. We lay out that the recognition of natural features is more difficult than the recognition of anthropogenic features such as houses or roads. For the latter group spectral and spatial characteristics can be anticipated and rules can be defined. Consequently, the automated extraction of roads or the 3D extraction of buildings is very advanced. In contrast, the automatic extraction of natural features like habitats is still far from being operational. We discuss reasons for that and highlight an object-based image segmentation methodology which incorporates spectral, spatial, topological and hierarchical characteristics of the objects within a semantic network.

1. INTRODUCTION

When dealing with the issue of an appropriate scale or the optimum resolution of images we must keep in mind what categories of target objects we investigate. This includes considerations regarding the scale domain in which those features are represented, as well as specific spectral or spatial properties. Before the advent of object-based classification techniques (Blaschke and Strobl, 2001), there was a sharper methodological distinction between classification of entire scenes (wall-to-wall, land cover/land use type) and pattern or object recognition. Whereas the typical sensors for large areas include multispectral scanners of medium resolution (e.g. Landsat), objects such as roads or houses were typically extracted from aerial photographs. Land cover mapping as the key application has been governed by the resolution and therefore more detailed habitat mapping remained in the domain of manual interpretation of aerial photographs. Recent developments in digital camera technology (e.g. HRSC-A) and the operation of high resolution satellite sensors (Ikonos, Quickbird), and more importantly the development of advanced analysis tools have brought the two sides closer together. As in other application fields, ecologists and nature conservationists become more aware of the potential of the spatial and temporal data resolutions and begin to believe that those might be useful.

When trying to heuristically describe natural target features it becomes obvious that this is more difficult as to extract anthropogenic features. As we will see the reason is partly that we must go beyond the human scale to accurately make representation of nature. Working outside the typical time and spatial scale and perhaps try to see the environment from another

species' view is a new and challenging task. Now - what exactly are we trying to delineate in a natural scene? Are we really addressing specific target objects or rather aggregates of something?

This article highlights some considerations and consequences why specific heuristics to differentiate between categories of anthropogenic features are easier to be set up than for categories of natural features. Although it can be argued that this simple fact is relatively obvious it is rarely addressed in image analysis and photogrammetry literature and object recognition and extraction work mainly concentrates on anthropogenic features. The set of heuristics to be used seems to be relatively clear in an anthropogenic environment: this is largely explained by the argument of homogeneity. Another aspect is the understanding of anthropogenic features. It is much clearer than the understanding of natural ones, because the features emerged through human action. Moreover we are dealing with a typical human scale, the scale where human action takes place and changes the environment. Data are optimized for this scale and so is typically our understanding and methodology. Nevertheless, mapping of anthropogenic settings is still challenging. It is a complex task but the underlying conceptual constructs which image interpretation exploits are more intrinsic.

2. HIERARCHICAL OBJECT REPRESENTATION

2.1 Object Representation vs. Object Recognition

Object representation in our understanding addresses a whole scene. Contrarily to object detection it aims at a

complete representation of the entire image information in several levels of aggregation. One solution to dissecting reality or a subset of a landscape is image segmentation. Image segmentation is not new (Haralick and Shapiro, 1985), but only recent software developments push a wider use in practice. When performing image segmentation a complete set of image objects is created. This is usually done for one specific object scale. We will demonstrate an application of multi-scale image segmentation whereby we aim for a hierarchical and interlinked set of objects. We use a segmentation approach, which segments images in several scales at the same time (Batz and Schaepe, 2000). In contrast to other segmentation approaches the user can somewhat control the process by defining a target scale parameter.

Image segmentation is a first step in our methodology based on the model of Burnett and Blaschke (2003). The segments at various levels can or cannot correspond to the scale of the objects of interest. In the second part of the methodology we build heuristics based on the image segments through categorization and classification. Object recognition then focuses on the detection of specific objects or a certain object type. It aims at the delineation of features and is based on clear conception. The potential to be correctly detected is a function of proper description and suitable resolution.

2.2 Heuristics for the Description of Anthropogenic, Semi-natural and Natural Features

We use 'anthropogenic' to refer to all features, which are constructed or planned by humans. 'Semi-natural' addresses all features which are mainly managed or maintained by human activity, such as fields, loggings, intensively used meadows and exhibit a certain variation within the objects and between the objects e.g. through seasonal conditions. 'Natural' finally is used for relative undisturbed areas of highly changeable complexity with smooth transitions and gradients. Because they are less dominated by human activities their spatial characteristics are not necessarily attached to human scales.

A study was performed to illustrate these ideas. We focused on a small study area represented on a colour orthophoto with a resolution of 0.2m. It shows the discussed feature types in a rather distinct manner: a typical rural setting with a village in the North surrounded by agricultural fields and intensively used grassland. In the south of the image part there is a small portion of a bog area, which shows several stages of degradation. The example illustrates that rules addressing the spatial properties of the target features can be more clearly established in the case of anthropogenic objects. The width of a road can be determined and different types of roads can be

differentiated. The houses can be defined by a specific arrangement of rectangular parts of the roof.

The area representing natural features exhibits a rather complex system. The entire bog consists of a complex pattern of vegetation patches which create sub-units of ecological relevance. The sub-units form habitats that represent a specific degradation stage of the bog through drainage activity over the last decades. In a classical manual interpretation such habitats are delineated by an experienced interpreter, who mentally aggregates the elements to a whole. Usually this process is accompanied by a certain generalization. Habitats themselves are typically not homogeneous, but bear a certain structure. The habitat '*Pinus mugo bog*' is characterized by a high within-patch-heterogeneity. Figure 1 illustrates that the arrangement of the *Pinus mugo* (mountain pine) patches can be categorized or typified, but not standardized as in the road example. The spatial arrangement of the constituting parts can be characterized using structural signatures (Lang et al, 2003). There is no strict rule set constructed, the spatial arrangement is not fixed in terms of e.g. the distance between the patches or by their absolute lengths or widths. Thus, structural signatures, as long as they represent an ideal status, must be built upon a set of fuzzy rules.

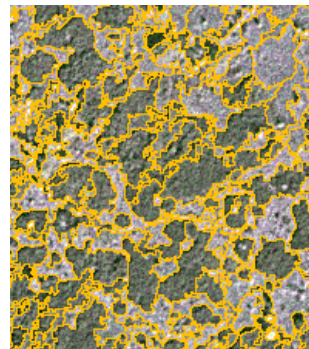


Figure 1: A complex arrangement of *Pinus mugo* patches in a bog area. Based on the specific spatial configuration experts would call it *Pinus mugo bog*.

When dealing with image objects and using a multi-scale segmentation approach, the spectral feature space is augmented significantly by various additional feature axes that describe spatial features as well as hierarchical ones. Thus an extended set of heuristics can be encoded to specify the features of interest. The categorization of features then relies on more information than the spectral reflectance, e.g. object size, object complexity, number of sub-objects, number of neighbouring objects, spectral distance to neighbouring objects etc. We hypothesize that the building of rule bases is less ambiguous for anthropogenic or semi-natural features compared to natural ones. For the sake of a clear distinction these two categories are contrasted against each other

although some environments may exhibit characteristics of both types.

Since anthropogenic objects are planned and constructed by humans, construction plans usually exist that will show, how these objects are supposed to look like. For these objects being technically planned and built, it is obvious that one can retrieve them on images. In general, a sound concept exists of how these well understood objects (roads, houses etc.) are being represented on aerial photographs or other high resolution data. Although technically challenging, many good examples and operational applications exist. If the image resolution and the set of individually resolvable target objects are known, a rule base can be established to classify them (figure 2).

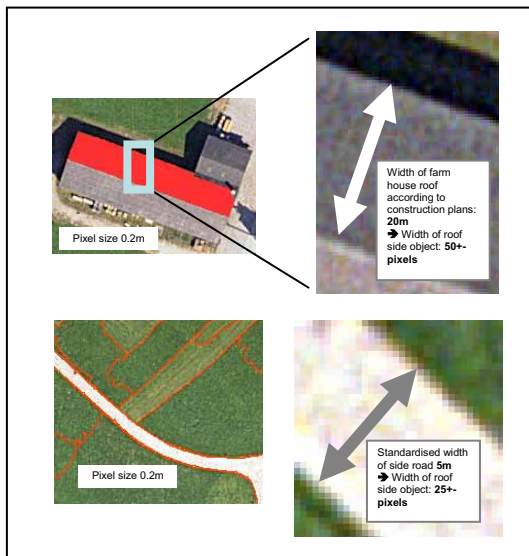


Figure 2: Anthropogenic features like roofs and roads can be described by the constant parameter 'width'.

2.3 Hierarchical Representation

We started this paper from the hypothesis that anthropogenic features are somehow related to a certain scale, natural features are not necessarily related to one specific scale. Conceptually, scale corresponds to a 'window of perception' (Hay et al., 2001). As we try to model anthropogenic features using hierarchical relationships among constituting elements (i.e. sub-objects), we end up with a rather shallow hierarchy with relatively distinct levels. In this respect the vertical structure of planned objects can be modelled relatively straightforward. Since anthropogenic features (like roads, houses, gardens) have low variance in size and shape, standardized sets of rules can be used to describe their vertical structure. A medium highway consists of four lanes, each of it with certain markers and separated by a narrow strip of lawn. The highway as well as its constituting parts (lanes, lawn strips, markers, etc.) has a standardized

width and specific ordered arrangement. In other words, in object recognition we know what we are searching for.

Contrarily, natural objects are less suited to be detected by object recognition methods. There is no such clear imagination of how they are organized in terms of spectral behaviour and spatial properties (size, form, etc.). The rules for objects we are searching for are difficult to determine. Many applications are based on both types of features. For example, in non-planned settlements the problems of interpretation come closer to that of natural phenomena. Hofmann (2001) has shown how difficult it is to detect scattered settlements being erected in a spontaneous manner.

A multi-levelled structural representation of natural habitats reflects the idea that natural living systems or medium-number systems (Weinberg, 1975) are complex systems, but hierarchically structured and therefore near-decomposable. As Simon (1973) illustrates, hierarchic systems can be decomposed to the levels of subsystems without breaking completely apart. A loose vertical coupling of subassemblies accounts for the stability of the whole. Koestler (1967) has elaborated the idea that any system or subsystem should be considered as a 'holon' (from Greek *holos* and *-on*) with self-assertive as well as integrative tendencies. Ecosystems are supposed to work as assemblies above the organism (Naveh and Lieberman, 1994). In a vertical dimension the hierarchical structure implies the evidence of distinct levels. Concrete systems emerge in specific levels, within which respective dynamic behaviour occurs. Thereby higher hierarchical levels correspond with low-frequency dynamics and lower levels with high-frequency ones. This way of organization results in a nested hierarchy of emergent entities in certain scale domains and at the same time it reveals, through horizontal coupling, specific spatial patterns within levels.

2.4 Applying Hierarchy Theory: Object-of-Interest Level

The conceptual framework of hierarchy theory offers a solution to the objects-of-interest problem. There are a number of levels, one of which is the focal one, the one under consideration (level of reference, Miller, 1975). The level above level 0 is referred to as level +1, the one below as level -1. Assuming that the remote sensing data exhibit a fine enough grain and broad enough extent to model landscape structure and pattern, we utilize hierarchy theory to identifying pattern components, i.e. real-world objects, at their respective scales of expression.

An object-related modelling approach (ORM, Burnett and Blaschke, 2003) provides the flexibility to model a landscape at various scales, i.e. to perform target-oriented landscape partitioning. These scales are to be

defined by experts or according to conventional findings. Studies (Hay, et al. 2003, Pinz, 1994) have tried to detect inherent scales by investigating the spectral content of a scene through different scales and aim to identify ‘significant scale levels’ within an image and to utilize information about the corresponding objects and their patterns. As Hay et al. (2001) have shown the ramifications for inappropriately using remote sensing data to understand landscape patterns are profound. Hay et al. (2003) demonstrated that the representation at different scales corresponds more to the objects of interests rather than only referring to statistical measures in an image.

Ecosystems are of overwhelming complexity and an object-of-interest level cannot be easily defined. If we step beyond the level of land cover classification, we are not automatically reaching another meaningful level. Seemingly any homogenous patch shows heterogeneity at a closer look (within-patch-heterogeneity). Whereas the scale of the target dimension is clear at anthropogenic features, it is not for natural ones (see figure 3).

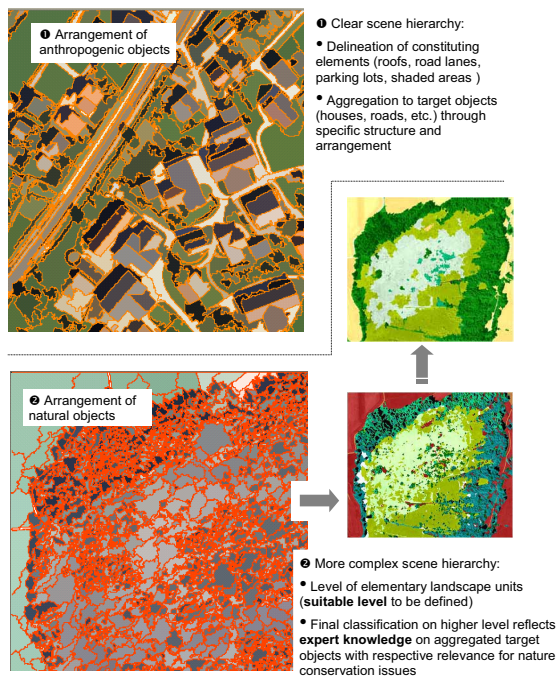


Figure 3: Target objects of segmentation.

According to Allen and Starr (1982) the concept of scale is illustrated by the analogy of a window, through which the constant flows of signals is filtered or weighted. So scale is defined by “the period of time or space over which signals are integrated [...] to give message”. Anthropogenic features usually cover a narrow scale window. They demand a shallow hierarchy. Natural objects demand a multi-scale representation, reflecting several scales which are corresponding to functional hierarchies. A ‘shallow’

hierarchical representation is faced with a ‘deep’ or flexible one, the bottom of which is not clearly defined.

When objects are to be modelled by heuristics beyond their spectral properties, we have to determine appropriate geometric properties. Based on segmentation, the objects intrinsically reveal numerous spatial characteristics known from GIS systems, such as area, width, area-perimeter ratio, or more complex compactness measures. Similarly to that we can describe semi-natural features (under an anthropocentric view) through geometric properties. But to find proper heuristics for natural features is more difficult. As they show a high degree of complexity only a vague concept about a typical structure exists. The structural representation remains an approximation. One way to deal with that high degree of variability is to establish structural signatures, which work as structural ‘samples’ for the spatial arrangement of natural objects of interest is one possible methodology. But they have to be thoroughly defined and the typical relevant configurations have to be found out. Additionally, a sound concept is needed how species respond to different scales.

In the multiscale segmentation approach results can be reported at different levels. While the calculation of the underlying metrics is mainly mechanistic, reporting is intrinsically connected to a selection process and subjective decisions choosing appropriate aggregation levels. Aggregation is a mental (perceptual) process which is based on knowledge and experience. It includes manual interpretation, the delineation of features depending on the experience and personal study context of the interpreter. If more complex systems are to be classified, an interpretation tends to be more biased by the underlying concepts; at least it becomes less transferable.

With increasing resolution, we can establish a relationship between the cell size and (anthropogenic) features being resolvable at that very resolution. After carefully checking average object size against image resolution a clear understanding should exist of which features are being represented and which not. But the same is difficult for natural features. We need a thorough expert knowledge about which ecological units can be interpreted based upon which spatial structure at what scale. It is not trivial to model specific relationships between the critical constituents in order to derive a mosaic of habitats. Even if the target dimension is clear (e.g. by being defined in a classification key) it is relatively ambiguous how these habitats are delineated. The success strongly depends on the expertise of people who are very familiar with the specifications of the area under investigation.

2.5 Homogeneity/Heterogeneity

For the sake of simplicity we refer to heterogeneity as a variation in space and leave the temporal aspect

aside in this paper. We assume a relative low amount of heterogeneity within the features of interest and a certain gradient to their surrounding at least for the underlying mapping criterion. To be 'relatively homogeneous' is the main concept of any dissection of reality, manual interpretation or image analysis. Homogeneity, of course, is a relative measure depending on the observer and the scale. One of the most striking criteria of homogeneity for the human visual perception system is likeliness of spectral characteristics. In multi-spectral classification techniques we are using similar spectral behaviour to classify certain land cover types, but only few pixel-based approaches incorporate homogeneity criteria. Usually the similarity is evaluated through the proximity in a spectral feature space. Whereas our visual system is only sensitive for the 'visual' spectrum of light, and therefore multi- or hyperspectral sensors seem to be superior, we use in fact more dimensions to represent features. Through a high-dimensional feature space we are locating features very precisely, and the property of homogeneity becomes a very complex one, because it is hard to define, what exactly makes them homogenous (beyond colour, shape). In an object-based environment image objects will bear a lot of features based on spectral and spatial information. They comprise colour, size, shape and relationships to neighbours or sub/super-objects.

Anthropogenic features are either homogenous or of high contrast in a more or less binary manner. For example considering roads, the pavement shows only low variance and so does the linearity of the shape. Whenever the homogenous characteristics is interrupted, this is caused by high-contrast (and again internally homogenous) features like painted white lines. Natural features show an intrinsic within-patch-heterogeneity, i.e. with increasing zoom we face a next level of organization. Therefore homogeneity is always a matter of disposition depending on the target scale dimension. Finally, empirical relationships have to be derived between measures of heterogeneity and ecological features of interest.

We characterised homogeneity of both complementary object types (anthropogenic vs. natural) by the parameters listed in figure 4. The scale parameters (SP) of the different segmentation levels refer the nomenclature of Baatz and Schaepke (2000) used in the software eCognition. SP expresses a mean summed heterogeneity across all information bands used for the resulting objects. It is correlated with the resulting mean object size. Mean values of parameters like the *length/width ratio* or the distribution of the parameter *spectral difference to neighbouring objects* clearly show that both categories of features bear specific properties regarding homogeneity. Although the segmentation level was nearly the same (100 vs. 120), the number of objects being created and thus the average area in this level differs a lot. Mean brightness

values show nearly a tri-modal distribution in the anthropogenic area.

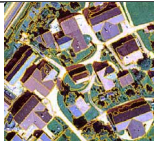
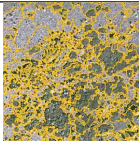
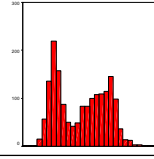
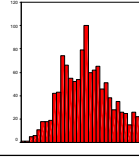
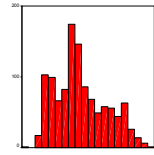
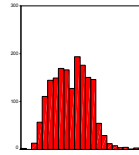
Parameter	 <i>anthropogenic</i>	 <i>natural</i>
Segmentation level	SP 120	SP 100
# of objects	1159	1727
Mean value of brightness	110.3	92.7
Length/width	11.4	2.5
Spectral difference to neighbour		
Area	307,1	137,4
Distribution of brightness		

Figure 4: Parameters to compare homogeneity in scenes dominated by anthropogenic and natural target objects.

3. CONCLUSION

We started this paper from the hypothesis that natural and semi-natural features are more difficult to be mapped from images than anthropogenic features. The remote sensing literature offers surprisingly little theory beyond of radiation and reflectance, especially for complex environments. Complex natural environments were opposed to the sophisticated arena of extracting and classifying roads and buildings. To elucidate the differences and reasons for them one core concept discussed herein is heterogeneity. While some ecologists were focusing on heterogeneity, others were building spatial variation into theory in the form of patches. To retain analytical tractability, the patches were usually assumed to be internally homogeneous and equivalent in size and quality. We discussed a multi-level image segmentation approach and conclude that it can in principle support the information extraction from images. In contrast to many photogrammetry approaches which concentrate on single classes it aims for a complete coverage of an image scene or a landscape. We demonstrated the potential of the use of spatial information beyond pixels. This is realized through object modelling using the different level of segmentation. This way, image analysis bridges 'classic' remote sensing techniques and GIS methods.

Clearly, we need a more thorough understanding of how to model complex natural systems and we need a sound concept how to actually delineate instances of these. The greatest methodological problem remaining is to scientifically determine the appropriate levels of segmentation. While this is determined by relatively well-known parameters for anthropogenic features our level of understanding of objects in a natural environment is still very low. Vegetation types exhibit certain statistical parameters, but sizes, shapes, distinctness from the surroundings or embeddedness in certain settings can vary enormously. This makes information extraction out of image data more difficult. Why are we not able to automatically delineate a certain habitat type from a satellite image automatically when at the same time surveillance cameras and attached software can relatively accurately identify persons on a street? The answer is partially that in the latter case we precisely know what we are searching for. In analogy, for natural environments, we would need hundreds of different examples of bogs to have a good chance to identify one in an image.

More interdisciplinary research will be needed between remote sensing experts, biologists and ecologists about the heuristics, which can be used to address characteristic structural properties. Of course, it is difficult to find the proper set of rules to delineate features that are for instance partially disrupted by shadow. But again, when aiming to extract a house, we have clear assumptions how to bridge the missing pieces of information. For natural systems, we do not. Still, we have discussed ways to overcome this difficulty. Some approaches investigate image characteristics in order to learn about typical arrangements or discontinuities within relatively homogeneous transitions. Finally, we need data for the appropriate scale. But we also need the right analysis scale for our data. Although a 0.5m resolution laser scanner data set seems to offer all demanded domains, the primary constituents of a natural complex system may not be resolvable, at least not automatically and in a repeatable and transferable manner.

ACKNOWLEDGEMENTS

We gratefully thank Dr. Charles Burnett for critical statements and a stimulating discussion.

REFERENCES

Allen, T.F.H and T.B. Starr, 1982. *Hierarchy*. University of Chicago Press, Chicago and London.

Baatz, M. and A. Schaepke, 2000. Multiresolution segmentation – an optimization approach for high quality multi-scale segmentation. *Angewandte*

Geographische Informationsverarbeitung, Salzburg, pp. 12 - 23.

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

Burnett, C. and T. Blaschke, 2003. A multi-scale segmentation / object relationship modelling methodology for landscape analysis. In: *Ecological Modelling* (pages pending).

Haralick, R. and L. Shapiro, 1985. Survey: Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 29(1), 100-132.

Hay, G.J., D.J. Marceau, P. Dubé and A. Bouchard, 2001. A Multiscale Framework for Landscape Analysis: Object-Specific Analysis and Upscaling. *Landscape Ecology*, 16(6), pp. 471-490.

Hay, G.J., T. Blaschke, D.J. Marceau and A. Bouchard, 2003. A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry & Remote Sensing*, 1253, pp. 1-19.

Hofmann, P., 2001. Detecting informal settlements from Ikonos image data using methods of object oriented image analysis. *Proc. of the Symposium on Remote Sensing of Urban Areas, Regensburg, Germany, June 2001*.

Koestler, A., 1967. *The ghost in the machine*. Hutchinson, London.

Lang, S., T. Langanke, K. Aaviksoo, C. Burnett and T. Blaschke, 2003. An Object-based Monitoring Methodology for Mires: Examples from Austria and Estonia. *Proceedings of the US-IALE Conference*, Banff, Canada, March 2-6, 2003.

Miller, J.G., 1975. The nature of living systems. *Behavioural Science*, vol. 20, pp. 343-365.

Naveh, Z. and A. Lieberman, 1994. *Landscape ecology – theory and applications*. Springer Verlag, New York.

Pinz, A., 1994. *Bildverstehen*. Springer, Heidelberg.

Simon, H. A., 1973. The organization of complex systems. In H. H. Pattee (ed.), *Hierarchy Theory: The Challenge of Complex Systems*, George Braziller, New York. Cambridge, pp. 1-27.

Weinberg, G., 1975. *An introduction to General Systems Thinking*. Dorset House Publishing, New York