

# OBIA – Tutorial

Introduction to Object-based Image Analysis

**Stefan Lang, Florian Albrecht, Thomas Blaschke**

Centre for Geoinformatics (Z\_GIS)  
Paris-Lodron University Salzburg

**Z\_GIS**

Centre for Geoinformatics, Salzburg University

## Disclaimer

This tutorial was built upon teaching material for courses on advanced remote sensing, delivered by Dr. S. Lang and Prof. T. Blaschke between 2002 and 2006. Explanations are partly quotes from literature.

Its purpose is to give an introduction to the emerging field of object-based image analysis and to provide a comprehensive overview of methods involved and the respective background.

The tutorial is available as 'slide set only' or 'slide set with additional text notes'. Both are provided in PDF (Acrobat® Reader® required for display).

Dissemination of this **Version 1.0** of the OBIA Tutorial is free. Please obey 'fair-use' conditions as being commonly understood.

We gratefully acknowledge the financial support from Definiens AG ([www.definiens.com](http://www.definiens.com)), which was granted for the compilation of this tutorial.

### Contact

Dr. Stefan Lang  
Centre for Geoinformatics  
Schillerstr. 30, Bauteil XV, 3. Stock  
5020 Salzburg  
+43.662.8044.5262  
[stefan.lang@sbg.ac.at](mailto:stefan.lang@sbg.ac.at)

*„Please cite the tutorial as follows*

Lang S, F Albrecht & T Blaschke (2006)  
OBIA-Tutorial – Introduction to Object-  
based Image Analysis, V 1.0 – Salzburg.

## Table of Contents

- **Chapter 1** – Image interpretation and perception
- **Chapter 2** – Basic concepts of hierarchy theory
- **Chapter 3** – Knowledge representation
- **Chapter 4** – Image segmentation
- **Chapter 5** – Object-based classification
- **Chapter 6** – Accuracy assessment

**Chapter 1** – Image interpretation and perception

*Visual perception; Image context and experience; Pixel- vs. objects; Visual delineation vs. machine-based segmentation*

**Chapter 2** – Basic concepts of hierarchy theory

*Relevance of multi-scale representation; Scales and scaled representation; Decomposition and holarchy; Multi-scale approach with remote sensing; Hierarchical patch dynamics paradigm*

**Chapter 3** – Knowledge representation

*What is knowledge? Cognition networks; Image understanding; Production systems vs. adaptive learning*

**Chapter 4** – Image segmentation

*History; Image segmentation in remote sensing; Segmentation algorithms; Scale space analysis ; Multi-resolution segmentation*

**Chapter 5** – Object-based classification

*Introduction; Available features for classification; Sample- vs. rule-based classification; Fuzzy classification; Class hierarchy; Class-related features; Strengths of object-based classification*

**Chapter 6** – Accuracy assessment

*Definitions; Non-site specific vs. site-specific accuracy assessment; Error matrix; Object-based accuracy assessment*

# OBIA – Tutorial

Introduction to object-based image analysis

## Chapter 1

### Image interpretation and perception

**Z**\_GIS

Centre for Geoinformatics, Salzburg University



## Outline

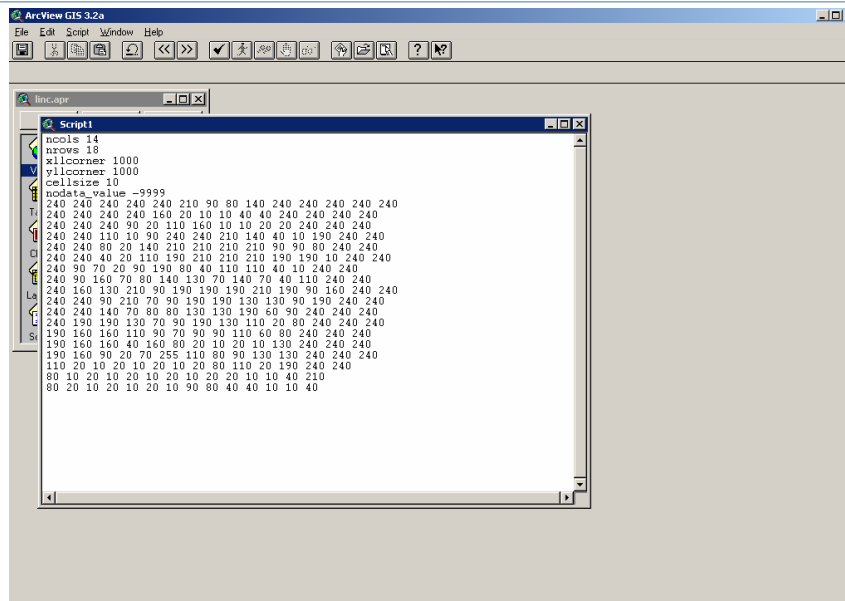
- **Visual perception**
- **Image context**
- **Role of experience**
- **Pixel- vs. object-scope**
- **Using objects**
- **Visual delineation vs. machine-based segmentation**

The first chapter introduces basic ideas behind object-based image analysis (OBIA). The way how human beings perceive imaged information will be covered and we will talk about the role of experience in image interpretation. In this context, the advantages of object-based image interpretation compared to a pixel-based approach will be pointed out.

When using a computer for classifying an image we transfer our knowledge to the machine and make it trying to imitate certain characteristics of the human way of image interpretation. In this respect, and in particular when going beyond spectral cues of certain geographical features, a pixel-based classification approach (i.e. treating single pixels individually) is limited. An object-based approach instead does support considering spatial characteristics of geographical features explicitly. Form-related as well as hierarchical (i.e. scale-related) characteristics can be addressed. We therefore are at the interface between remote sensing and GIScience. But, as we will see later, also the object-based approach has its limitations. Still, our human perception is an ultimate benchmark, still undefeated in analyzing complex scene contents with ease.

Thus, this chapter starts with a short discussion how we perceive the contents of images. An example is given that shows that the content of a picture file can only be recognized given data are coded in the right way. Only then we perceive patterns of color and form, structured in various levels throughout the image. The example shows that the content of an image appears in several scales. The difference between the human way of image interpretation and the way how image data are represented as pixels becomes obvious. The human eye sees much more than only different colors, it also perceives shapes, texture and the spatial arrangement of certain elements. Our own experience influences image interpretation and each of us sees the image in a different context. The context we apply depends on the degree familiarity with nadir-looking satellite imagery or aerial photographs, their different band combinations, and our knowledge about the depicted image content. Gaining experience is a basic prerequisite for skilful image interpretation.

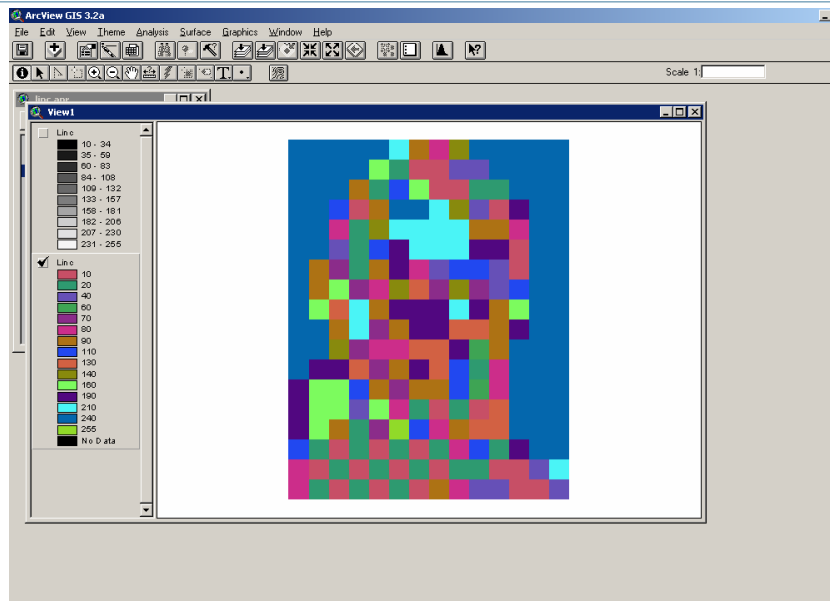
## Visual perception



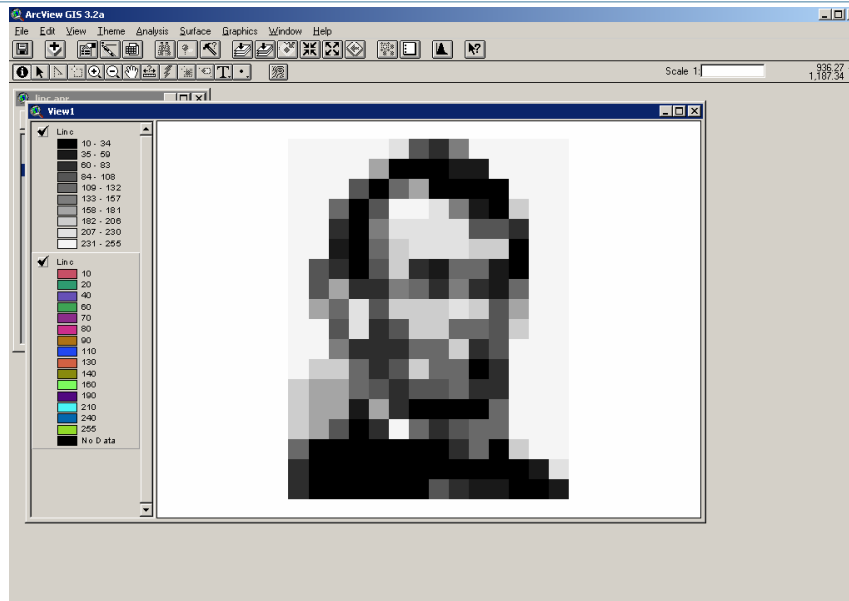
The following example illustrates that data is only valuable for us if coded in the right way. Numbers alone can hardly be interpreted beyond their immediate meaning and transforming rows of number to a 'picture' is nearly impossible for the human brain. Even if coded into (arbitrary) colors, as shown in the following slide, it does not make much difference. Only if there is proper link between coding and color scheme, information is conveyed successfully (see slide after).

❶ » Compared to a scene being represented in pixels the human way of image interpretation is quite different (Pinz, 1994) even if we see single pixels in an image due to a bad resolution we refuse to perceive them. Usually – when looking at an image – we perceive a complex pattern of colour and form, structured in various levels throughout the image. Fine structured features appear but we simultaneously aggregate them into larger ones. That means that the content of a scene appears in several scales at the same time. To illustrate this, an example is given showing an array of grey-scaled pixels. In spite of the bad resolution and the limited number of grey scales, it is possible to convey enough information to recognize a face! The specific arrangement of perceived parts of the face and clothing makes it easy to assign the face to a well-known person. A blurred sight imitates smoothing (suppression of high frequencies). All in all, (exactly) two scale domains are visible, namely the one of the face, and the one of its constituting parts (like eyes, nose, beard, hair, forehead). « (Lang, 2005, p.39)

## Visual perception (2)



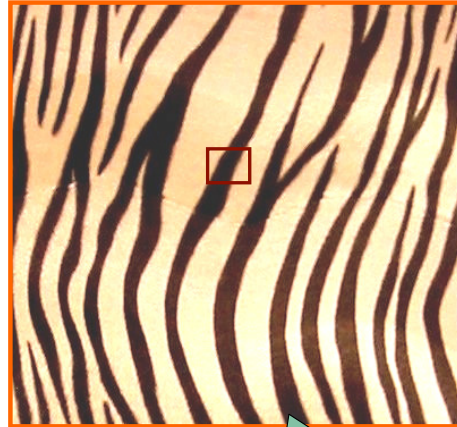
## Visual perception (3)



## Image context



contrast b/w and shape:  
"black line in white area"



mainly shape (pattern  
suppressed): "chair with  
zebra pattern"

contrast b/w and shape  
(elongated but acute):  
"certain stripes pattern  
→ zebra"

Similarly, we can illustrate the role of image context and, in addition, the power of a given spatial arrangement. The picture above, far left, shows a feature which can be interpreted as a piece of a black line on a white background.

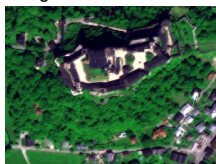
As soon as there is more context provided, we can see that the black area is indeed a small section of a line which belongs to a striped pattern that we recognize as the fur pattern of a zebra.

Hmm, sorry, the last picture makes us rejecting all our previous ideas and there is no way in ignoring that there is a chair with a zebra-like cover!

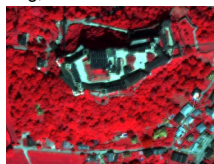
## Role of Experience

- **Human vision is well adapted for complex image interpretation tasks**
  - **Experience built up since early childhood**
  - **But human vision is challenged when dealing with remote sensing imagery:**
    - Applying an overhead view
    - Dealing with spectral characteristics beyond the visual spectrum
    - Working with unfamiliar scales and resolutions
- > Experience is an important prerequisite for skillful and successful interpretation**

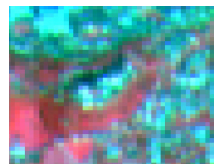
Images of the Fortress in Salzburg, Austria



Quickbird; bands 3,2,1



Quickbird; bands 4,3,2



Aster; green, red, infrared



Color Photo

ZGIS

Centre for Geoinformatics | © 2006 Lang/Albrecht/Blaschke | stefan.lang@sbg.ac.at

❶ » We can hardly describe exactly what really happens if we look at an image and suddenly 'see' something. But indeed we notice that we do any kind of pattern recognition without major effort (Eysenck and Keane, 1995; Tarr and Cheng, 2003). Human perception is a complex matter of filtering relevant signals from noise, a selective processing of detailed information and, of course, experience. « (Lang, 2005)

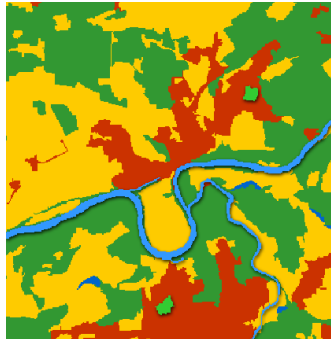
❷ » Three issues distinguish interpretation of remotely sensed imagery from interpretation conducted in everyday experience. First, remotely sensed images usually portray an overhead view – an unfamiliar perspective [...]. Second, many remote sensing images use radiation outside the visible portion of the spectrum [...]. Finally, remote sensing images often portray the earth's surface at unfamiliar scales and resolutions [...]. Students cannot expect to become proficient in image analysis simply by reading about image interpretation. Experience forms the only sure preparation for skillful interpretation. « (Campbell, 2002)

❸ » When starting with manual air-photo or satellite image interpretation we notice that a lot of experience is required. It mostly needs training to match the imaged false colour schemes with natural phenomena and to understand certain texture or structures and the imaged features. Unfortunately, even long time learning cannot prevent us from facing ambiguity when features are very like in structure or colour. According to recent findings in brain research (Spitzer, 2000; Churchland, 2001) signal processing by any of our senses is based on vector coding of signals in a high-dimensional feature space. It remains a challenge to find out more about the vector axes being used when we interpret signals through our senses. 'Experience' means a tighter allocation in our multidimensional feature space. « (Lang, 2005)

❹ » Note that the domain of interest of a skilled interpreter may differ from that of a simple user; the experience of the former makes him specifically look for certain features, whereas the latter is mainly interested in the information he wants to obtain. « (Lang, 2005)

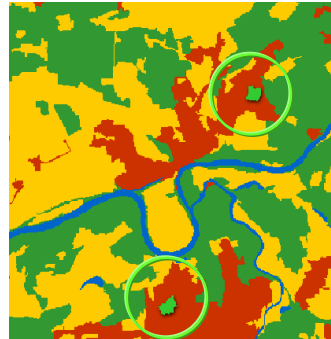
❺ » Nevertheless, human photo-interpreters also implicitly use structural knowledge in the manual classification process. They do not only consider contextual information but also information about the shape of and the spatial relations between the image regions. « (Blaschke, 2003)

## Pixel- vs. object-scape



river

- spectral properties
- specific form/shape



municipal park

- spectral properties
- specific spatial context

*From Definiens, 2004*

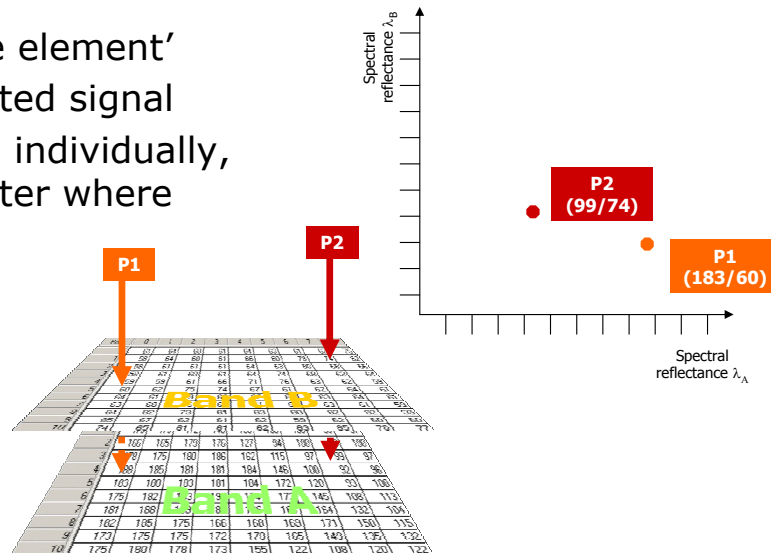
When looking at an image we see several different features that we can describe by looking at their properties.

For example the river in the figure above, left, has specific spectral values and we by its form can distinguish it from other features (e.g. lakes) that have similar spectral values. Another example is a municipal park. Although spectrally similar to grassland we can identify them as parks because they are placed inside urban areas. How can we tell the computer to classify these features properly? By utilizing spatial context and form.

## Pixel- vs. object-scape (2)

- **Pixel**

- 'picture element'
- integrated signal
- treated individually, no matter where located



ZGIS

Centre for Geoinformatics | © 2006 Lang/Albrecht/Biaschke | [stefan.lang@sbg.ac.at](mailto:stefan.lang@sbg.ac.at)

An average Landsat scene with a size of 141 km<sup>2</sup> and a spatial resolution of 30 m is composed by some 22 million pixels per band. The number of panchromatic pixels (15 m resolution) increases to 88 million. An Ikonos scene covering 11 km<sup>2</sup> is built up by approx. 7.5 million pixels (four bands, 4 m resolution) and 121 million panchromatic pixels. No matter, which resolution, a pixel is always an integrated signal of reflection emitted by the observed underlying features (mixed pixel effect). Since usually neighboring pixels over a certain range have similar signals (spatial autocorrelation), it is kind of inappropriate and also 'uneconomically' to treat each of them individually.

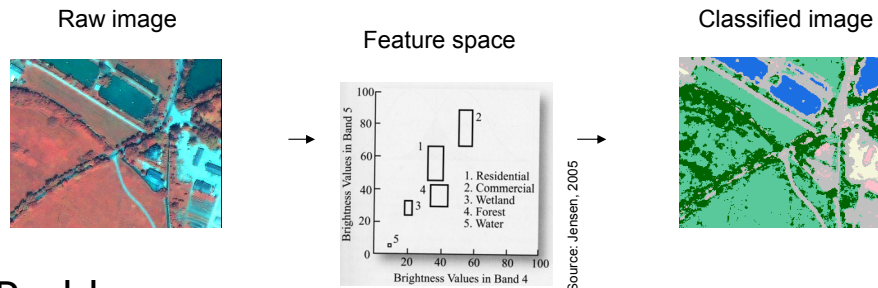
❶ » A digital image is composed of many thousands of pixels ("picture elements"), usually each too small to be individually resolved by the human eye. Each pixel represents the brightness of a small region on the earth's surface recorded digitally as a numeric value usually with separate values for each of the several regions of the electromagnetic spectrum. « (Campbell, 2002)

❷ » For the remotely sensed data classification, many classifiers based on the spectral analysis of individual pixels have been proposed and significant progress has been achieved. However, these approaches have their limitations since most remote sensing image classification techniques are based on per-pixel procedures (Blaschke & Strobl, 2001). They analyze pixels mainly using multivariate statistics. So the current results still cannot compare with human photo-interpreters. « (Blaschke, 2003)



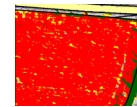
## Pixel- vs. object-scape (3)

## Pixel-based classification process



## Problems

- Spectral values belong to more than one information class
- No spatial relationships used in classification
- Pixel artificial spatial unit
- 'Artifacts' (salt-and-pepper effect)



❶ » By comparing pixels to one another, and to pixels of known identity, it is possible to assemble groups of similar pixels into classes that are associated with the informational categories of interest to users of remotely sensed data. These classes form regions on a map or an image, so that after classification the digital image is presented as a mosaic of uniform parcels, each identified by a colour or symbol. These classes are, in theory, homogeneous: pixels within classes are spectrally more similar to one another than they are to pixels in other classes. In practice, of course, each class will display some diversity, as each scene will exhibit some variability within classes. « (Campbell, 2002)

❷ » The traditional method for analysis of EO data in landscape research is the classification of pixels based on pixels in the same land cover class being close in spectral feature space. This does not hold true for complex environments and their respective classifications. « (Burnett & Blaschke, 2003)

❸ » Sometimes such [pixel-based] classifiers are referred to as spectral or point classifiers because they consider each pixel as a “point” observation (i.e., as values isolated from their neighbours). Although point classifiers offer the benefits of simplicity and economy, they are not capable of exploiting the information contained in relationships between each pixel and those that neighbour it. « (Campbell, 2002; page 319f.)

❹ » As Townshend et al. (2002) point out, a significant, but usually ignored problem with per-pixel characterization of land cover is that a substantial proportion of the signal apparently coming from the land area represented by a pixel comes from the surrounding pixels. This is the consequence of many factors including the optics of the instrument, the detector and the electronics, as well as the atmospheric effects. An alternative is to use contextual procedures in which observations from surrounding pixels are used to assist the characterisation. Although it might be desirable to integrate neighbourhood information continuously or in a fuzzy way, one operational method to work with relatively homogeneous areas is image segmentation, i. e. the use of image objects. « (Blaschke & Strobl 2001)

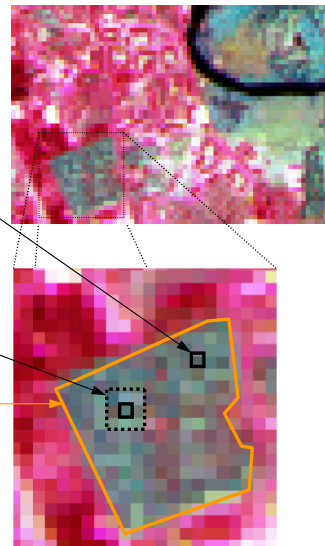
## Pixel- vs. object-scape (4)

Limitations of **pixel-based analysis***considering*

- **Colour** (spectral reflectance in n Bands)
- **Texture** (certain environment, e.g. 3\*3 pixels)

*but not*

- **Form & shape**
- **Neighbourhood**
- **Context**
- **Levels**



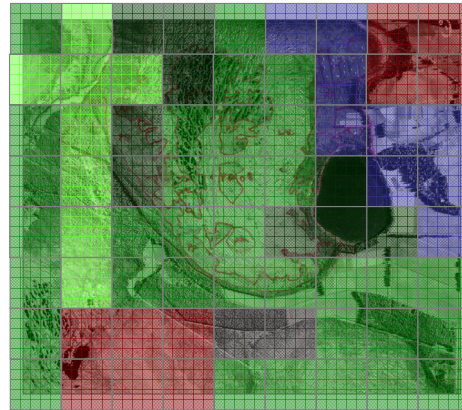
❶ » Human interpreters could derive little information using the point-by-point approach, because humans derive less information from the brightness of individual pixels than they do from the context and the patterns of brightnesses [i.e. texture], of groups of pixels, and from the sizes shapes and arrangements of parcels of adjacent pixels. « (Campbell, 2002)

❷ » Human photo-interpreters also implicitly use structural knowledge in the manual classification process. They do not only consider contextual information but also information about the shape of and the spatial relations between the image regions. [...] One way to make use of this additional information is to organize the image into objects that represent regions of similar pixels prior to the classification. [...] In most cases, information important for the understanding of an image is not represented in single pixels but in meaningful image objects and their mutual relations. « (Blaschke, 2003)

## Using objects

- **Relation between target objects and spatial resolution**

- Increasing importance of VHR EO data
- High level of detail provides extended set of target classes
- Addressing these target classes in a Landsat-imagery would fail

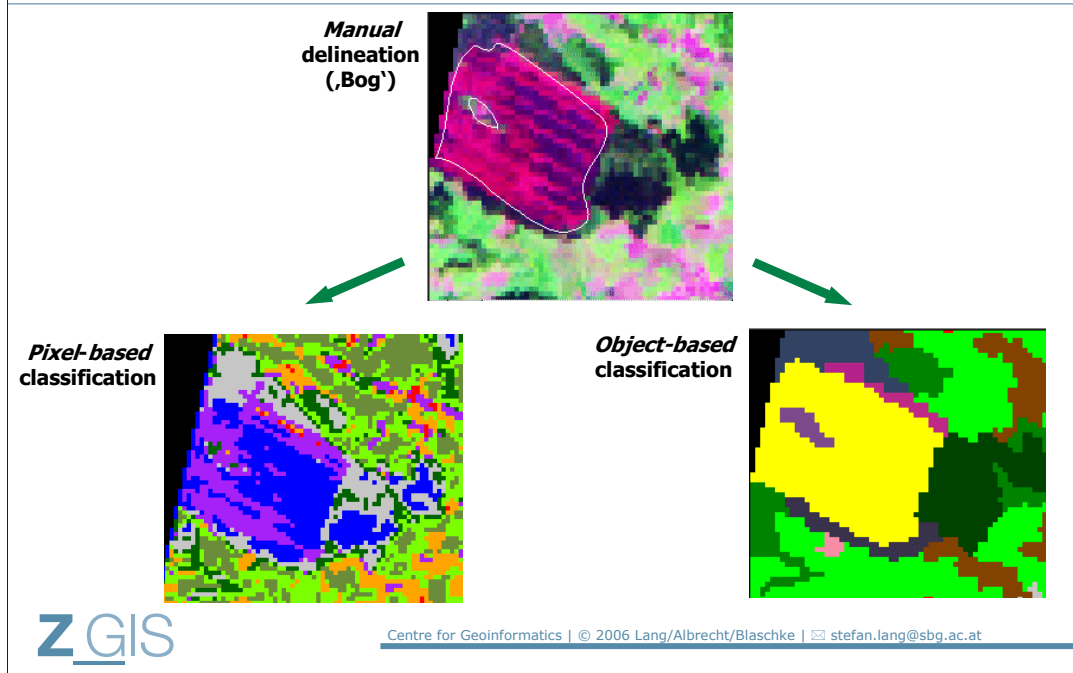


❶ » To extract objects of interest, the statistical analysis of pixels exclusively based on their spectral statistic is not sufficient. As laid out in several recent publications (Blaschke, Strobl 2001; Ehlers et al., 2002; Flanders et al. 2003), the advent of higher resolution image data increased the need for more efficient methods more than ever. Generally, for high resolution data, segmentation as a pre-classification step is preferred over pixel based classification because the resulting division of space tends to involve fewer and more compact subregions. « (Blaschke, 2003)

❷ » Segmentation approaches are generally more suitable for high resolution data, where pixels tend to be spatially clumped. « (Blaschke, 2003)

❸ » For example, in a 1.0-m-resolution image of a forest canopy, where each tree crown exhibits a 10-m diameter, each crown image-object will be composed of many pixels. In this situation each 1.0 m pixel will be part of an individual crown. [...] As a result, an image-object tends to be composed of spatially clustered pixels that exhibit high spectral autocorrelation because they are all part of the same object. Consequently they have similar gray values. These characteristics correspond to Tobler's first law of Geography where 'objects are related to all other objects, but proximal objects are more likely to be related to each other' (Tobler 1970). In an image-object, this relationship is both spectral and spatial. « (Hay et al., 2003)

## Using objects (2)



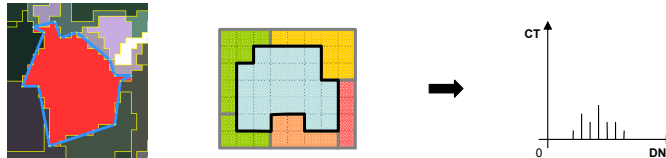
Landsat ETM image of a small bog system in Southern Germany (top), pixel based classification (lower left), segmentation-based classification (lower right).

❶ » The Figure illustrates the main difference when building objects prior to the classification process. Some degree of generalization is applied in this phase. For many applications, e.g. land cover mapping, generalization is intrinsically required to produce tangible target objects of the same class which are relatively homogenous according to the class definition. « (Blaschke, 2003)

The 'salt-and-pepper-effect' occurs, if there are many pixels classified differently but actually belonging to the same land cover type (here: bog). This may result in a unnecessarily detailed classification of the land surface. It can be overcome by segmentation (i.e. grouping the similar pixels first), followed by classification.

## Using objects (3)

- **Meaningful objects**
- **Improved reliability of statistics**
  - Several measurements (pixels) per object
  - Clear boundaries

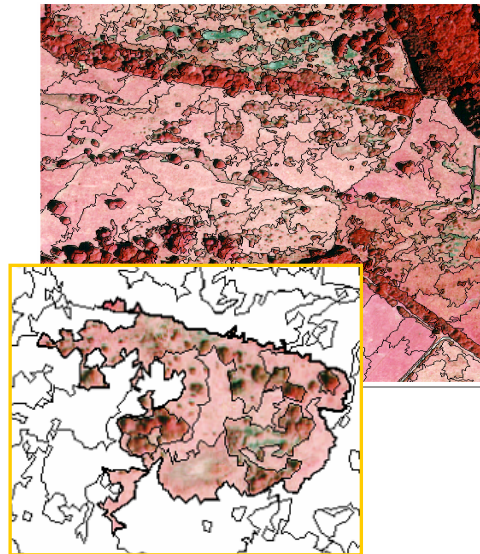


- **Augmented, uncorrelated feature space**
  - Texture within objects, shape, neighbours, hierarchy

❶ » The basic elements of an object-oriented approach are image objects. Image objects are contiguous regions in an image. We distinguish between image object primitives and objects of interest. Only objects of interest match real-world objects, e.g. the building footprints or whole agricultural parcels. Object primitives are usually the necessary intermediate step before objects of interest can be found by segmentation and classification process. The smallest image object is one pixel. Image objects can be linked to a hierarchical network, where they are attributed with a high-dimensional feature space. [...] Within an image object all kind of statistics based on single input layers or combinations within the input image layer stack can be computed, e.g. the ratio of the mean values of two input channels A and B. [...] Using image objects to calculate this statistic instead of boxes of pixels improves the reliability of statistic without smearing edges, since objects do not exceed edges. Of course homogeneous areas of mixed pixels can't be resolved. In ideal cases, this mixture would be detected since it is not matching the signatures of pure classes and therefore result in a reduced reliability of object classification. [...] Advantages of object-oriented analysis are meaningful statistic and texture calculation, an increased uncorrelated feature space using shape (e.g. length, number of edges, etc.) and topological features (neighbour, super-object, etc.), and the close relation between real-world objects and image objects. This relation improves the value of the final classification and cannot be fulfilled by common, pixel-based approaches. « (Benz et al., 2004)

## Using objects (4)

- ➔ **Integration of Remote Sensing and GIS**
- GIS users are 'used to' **polygon** environment
- **Aggregation** of information (highly textured images like VHR data or radar)
- Modelling of scale-specific ecological processes through **multi-scale representation**



❶ » Remotely sensed data are a crucial source for GIS users. From the GIS user's perspective a polygonal structured image representation is like mother tongue. [...] Making use of the objects' attributes and spatial relations brings geographical concepts that are basic to GIS into the field of Remote Sensing. [...] We argue for a somewhat different handling of our entities introducing the concepts of neighborhood, distance and location. All these concepts are not new. In fact, entire disciplines like Geography are based on these concepts. The question therefore is: Why are remote sensing and digital image processing still so much focused on the statistical analysis of single pixels rather than on the spatial patterns they build up? « (Blaschke, Strobl, 2001)

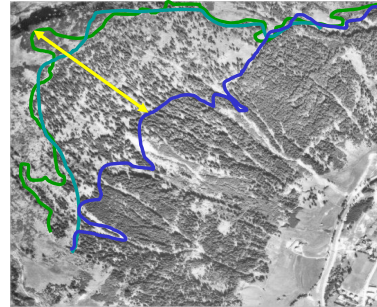
❷ » GIS analysis is usually dominated by an investigation of the horizontal relationships of (map) objects. [...] High-resolution image data carry information in a fine scale resolution and in aggregated super-levels at the same time. [...] Nested relationships in a systemic hierarchy are needed. « (Lang, 2005)



## Visual delineation vs. machine-based segmentation

- **Problems occurring with visual delineation visually (may be solved via segmentation):**

- selection of appropriate levels of generalization
- Individual delineations
- placement of boundaries when there is a graded transition



*Several possibilities for the delineation of 'Forest'*

- **Problems that challenge machine-based segmentation:**

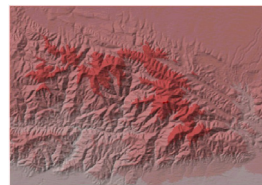
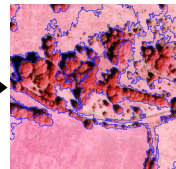
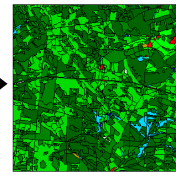
- Delineating conceptual boundaries (e.g. the 'outline' of an orchard, see below)

❶ » The interpreter must often delineate, or outline, regions as they are observed on remotely sensed imagery. The interpreter must be able to separate distinct areal units that are characterized by specific tones and textures, and to identify edges or boundaries between separate areas. Typical examples include delineation of separate classes of forest or of land use – both occur only as areal entities (rather than as discrete objects). Typical problems include: (1) selection of appropriate levels of generalization (e.g., when boundaries are intricate, or when many tiny but distinct parcels are present); and (2) placement of boundaries when there is a gradation (rather than a sharp edge) between two units. « (Campbell, 2002; page 123ff.)

❷ » We can classify the generated segments by modelling their structural properties (ORM). Hierarchical properties of image segments can be expressed by various measures characterizing the averaged properties of (sub)-objects and/or their spatial arrangement (proximity measures). However, in some cases ORM cannot provide a solution for the intended modelling of a target class. This applies when the required geometry of the target class polygons is not provided by segmentation due to restrictions of (region- or edge-based) segmentation algorithms. In many cases the human brain can easily manage to detect and delineate features that otherwise in a machine-based way are hardly to extract. This we can prominently observe for features whose boundaries are mentally constructed and not directly seen in an image. « (Blaschke et al., 2005)

## Visual delineation vs. machine-based segmentation (2)

- **Image Objects = Landscape Units?**
  - **Usually landscape analysis or -planning is based on landscape units**  
→ manual interpretation
  - **Leads image segmentation to similar results?**
  - **Image objects not *per se* 'real world objects'**
    - not **object recognition**
    - Any raster layer can be used for segmentation (e.g. DEM)
  - **object merge can improve appropriateness of objects**

CIR  
interpretationImage  
segmentation

Obviously, any kind of planning is based on spatial units, patches, etc. Landscape analysis and landscape planning utilize units, often derived and updated through images.

When deriving Image objects, we do not have *per se* real world objects in hand. They may be used as representatives for real-world objects, if being (1) appropriate to the respective geographical feature they represent and (2) satisfying our eye. But of course, our eye is not unambiguous either.



## Visual delineation vs. machine-based segmentation (3)

- **Some more aspects**
  - Results not necessarily more correct, but more intuitive, more convincing, more practical
  - object generation suitable for textured or low-contrast image data
    - VHR-satellite imagery
    - Airborne optical scanner data
    - Airborne laser scanning (ALS) data
    - Synthetic aperture radar (SAR) data
  - Semi-automated image interpretation
  - Supports image understanding by solving complex semantic problems
  - Monitoring of known structures
    - e.g. existing land use classification can be used as pre-defined boundaries for segmentation

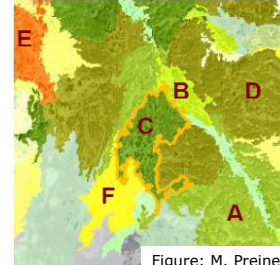


Figure: M. Preiner

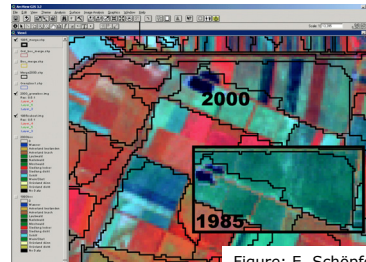


Figure: E. Schöpfer

# OBIA – Tutorial

Introduction to object-based image analysis

## Chapter 2

### Some basic concepts of hierarchy theory



Centre for Geoinformatics, Salzburg University

## Outline

- **Relevance of multi-scale representation**
- **Scales and scaled representation**
- **Decomposition and holarchy**
- **Multi-scale approach with remote sensing**
- **Hierarchical patch dynamics paradigm**

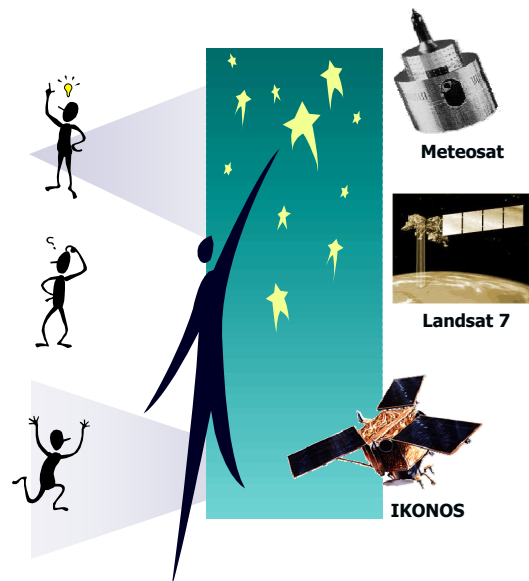
In this chapter we will take a closer look on the concept of scale, and start with a short review on hierarchy theory. Note that applying a coarser scale does not necessarily mean reaching a hierarchical level. In fact hierarchy theory is talking about the level of organization, which is not restricted to average size or extent alone. But drawing on this theory, the Hierarchical Patch Dynamics Paradigm (HPDP) has been introduced in landscape ecology (see below). The term holarchy from hierarchy theory has been applied to the hierarchical organization of landscapes (hierarchically structured patches of increasing average size).

Objects have an inherent scale. An image can be represented in different scales simultaneously (multi-scale representation), depending on the respective scale domain of target objects. In hierarchy theory fundamental parts interacting in a complex system are called holons. Separating and ordering the system components according to their scale can be done by the means of multi-scale analysis. From a remote sensing perspective, image objects are at the same time aggregates of pixels (or smaller objects), as well as parts of larger objects.

The strategy provided by HPDP combines the idea of hierarchically organized patches, their interactions within the ecological system, and the relation between the observed patterns and underlying processes that change with scale.

## Relevance of multi-scale representation

- **General remarks**
  - **Sensor resolution nowadays enters the scale of human activity**
  - **Field of potential applications is increasing**
  - **More and more detailed object levels can be represented**
  - **Complexity of analysis tasks increases as well**



Increasing resolution of image data lets us climb farther down within the 'G-scale'. Because image resolution is now entering the domain of scale of human interaction we need to apply a different view.

In the 1 m resolution domain provided by Ikonos, Quickbird and other operational EO sensors, the variety of application areas is broadly increasing (ecology, urban planning, security issues, etc). Whereas a classical land use / land cover classification may aim at a list of maybe 15-20 classes, the level of detail of recent data allows for a much higher number of classes to differentiate. The degrees of freedom in what we can find our increases heavily.

High spatial resolution supports multi-scale observation, because the human activity resides right in the centre resolution domain. By this, the former 'macroscope' of remote sensing data evolves to a 'homoscope'.

## Scales and scaled representation

- **Scale**

Refers to the size of objects that appear in the landscape or in a representation of it (e.g. a map or a satellite image)

- **Different objects have different scales**

Every object has its inherent scale

It only appears in a certain range of scale

- **Depending on the elevation of our viewpoint we see certain objects**

Different Views – Different Scales – Different Objects?

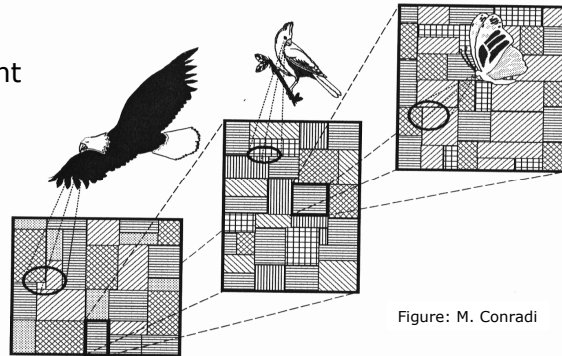


Figure: M. Conradi

❶ » In landscape ecology, there is a growing awareness about continuity of phenomena and discontinuities of scales. Forman (1995) described this ambiguity through the metaphor of a person gradually descending with a spaceship or balloon. Human perception abruptly starts to discover patterns and mosaics. Many mosaics are quasi-stable or persistent for a while, separated by rapid changes that represent the “domains of scale”. Each domain exhibits certain spatial patterns, which in turn are produced by a certain causal mechanism or group of processes. « (Blaschke & Strobl, 2001)

❷ » 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”. « (Lang & Blaschke, 2003)

❸ » 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. « (Lang & Blaschke, 2003)

❹ » Although in the domain of remote sensing a certain scale is always presumed by pixel resolution, the desired objects of interest often have their own inherent scale. Scale determines the occurrence or non-occurrence of a certain object class. The same type of objects appears differently at different scales. Vice versa, the classification task and the respective objects of interest directly determine a particular scale of interest. « (Definiens, 2004, p 62)

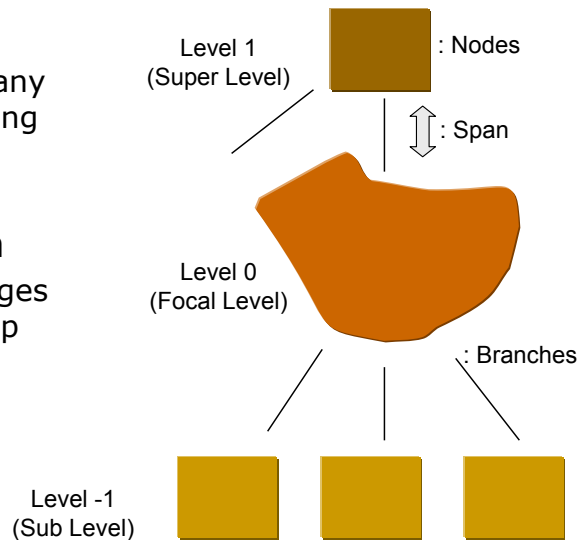
## Scales and scaled representation (2)

- **Scaling Ladder**

Every portion of land contains objects of many different scales resulting in a series of scales

- **Boundaries within the scale spectrum**

Thresholds between ranges of scale are never crisp



❶ » The ladder of hierarchical levels (Wu, 1999) seems to be infinite in either direction – the ultimate constituents of reality are not found yet. « (Lang, 2005)

❷ » The metaphor of a ‘scaling ladder’ (Wu, 1999) marks adjacent scale domains in a continuous scale spectrum. According to hierarchy theory it is assumed that a series of scales is inherent in any portion of land (landscape) no matter what the actual size is (Lang, 2001). O’Neill et al. (1986) have proposed to consider at least three nested scales in any study: the level of interest (‘focal level’ or ‘reporting level’) is constrained by controlling conditions of the level above, which provides significance; lower level components supply explanations (Turner, et al., 2001). Burnett & Blaschke (2003) use ‘candidate discretisations’ to illustrate the formation of scaled representations when working on image data with several inherent scale levels. « (Lang & Langanke, 2006; Lang, 2005)

❸ » Scale thresholds are never crisp, since they mark the boundaries between scale continuums, but are made hard by fiat. « (Burnett & Blaschke, 2003)

## Hierarchical patch dynamics paradigm

- **HPDP – combining hierarchy theory and patch dynamics**

A hierarchical scaling strategy dealing with spatial heterogeneity

- **Holon = patch = ecological unit at a particular scale**

- **Interaction of components**

Loose vertical and horizontal coupling in structure and function

❶ » Patch dynamics provides a powerful way of dealing explicitly with spatial heterogeneity. Wu and Loucks (1995) suggest the integration between hierarchy theory and patch dynamics via the HPD paradigm and lay a theoretical framework for a theory-driven breaking down of ecological complexity through a hierarchical scaling strategy. Wu (1999), drawing on the Koestler's concepts of flux rates in hierarchy, suggests that ecological systems are nearly completely decomposable systems because of their loose vertical and horizontal coupling in structure and function. The term "loose" suggests "decomposable" and the word "coupling" implies resistance to decomposition. When translating hierarchy theory to landscape ecology, holons are synonymous with patches: the ecological unit at a particular scale. Patches interact with other patches at the same and at higher and lower levels of organization through loose horizontal and vertical coupling. [...] Wu & Loucks (1995) and Wu (1999) suggest that the HPD theoretical framework can be used to perceive and model landscape as a hierarchical mosaic of patches although it is difficult in empirical studies to distinguish clearly between nested and non-nested hierarchies (Allen & Starr, 1982), at least prior to investigation. « (Burnett & Blaschke, 2003)

❷ » Ecological systems are hierarchical patch mosaics. On different scales, a patch may be defined as a continent surrounded by oceans, a forest stand surrounded by agricultural lands and urban areas, a fire-burned area or a tree gap within a forest, or a stomata on a leaf. Patches can be characterized by their size, shape, content, duration, structural complexity, and boundary characteristics. The theory of patch dynamics indicates that the structure, function, and dynamics of such patches are important to understanding the systems they comprise, be they populations, communities, ecosystems, or landscapes. « (Wu, 1999)

## Hierarchical patch dynamics paradigm (2)

### ▪ Wu (1999)

- Ecological systems as spatially nested patch hierarchies
- Dynamics of an ecological system derived from the dynamics of interacting patches
- Pattern and process are related and change with scale
- Non-equilibrium and stochastic processes do not necessarily work against stability

❶ » The list below is a digest of the HPD framework found in Wu (1999).

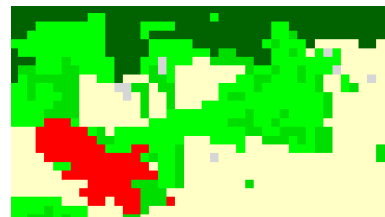
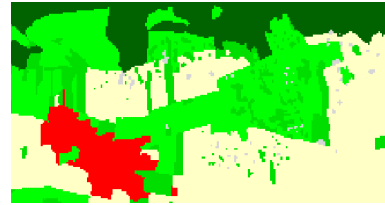
1. Ecological systems can be perceived as spatially nested patch hierarchies, in which larger patches are made up of smaller, functioning patches.
2. The dynamics of a given ecological system can be derived from the dynamics of interacting patches at adjacent hierarchical levels. Patches at higher levels impose top-down constraints to those lower levels by having slower or less frequent processes, while lower levels provide initiating conditions and mechanistic explanations for, and give apparent identity to, higher levels through interactions among patches. Distinctive characteristic time scales of patches at lower versus higher levels are the fundamental reason for the near-decomposability of ecological systems.
3. Pattern and process have components that are reciprocally related, both pattern and process, as well as their relationship, change with scale.
4. Non-equilibrium and stochastic processes are common in ecological systems. In general, small scale processes tend to be more stochastic and less predictable. However, non-equilibrium and stochastic processes do not necessarily work against stability. They usually constitute mechanisms that underlie the apparent stability of systems. (Wu; 1999)

We believe that a better landscape analysis methodology can be built upon a combination of HPD theoretical base, an object-orientated modeling environment and advanced GIS and RS methods.  
« (Burnett, Blaschke; 2003)



## Scales and scaled representation (3)

- Relevant range of the scale spectrum for landscape analysis
  - Grain  
**minimum area at which an organism responds**  
**comparable to resolution (spatial, spectral, temporal) in an image**
  - Extent  
**coarsest scale of spatial heterogeneity**  
**extent of the whole scene (total area, bandwidths, covered temporal duration)**



❶ » Landscape ecology: Grain is the minimum area at which an organism perceives and responds to the patch structure of landscape (Kotliar and Wiens, 1990). Extent is the coarsest scale of spatial heterogeneity at which organisms react (Farina, 1998). « (Burnett & Blaschke; 2003)

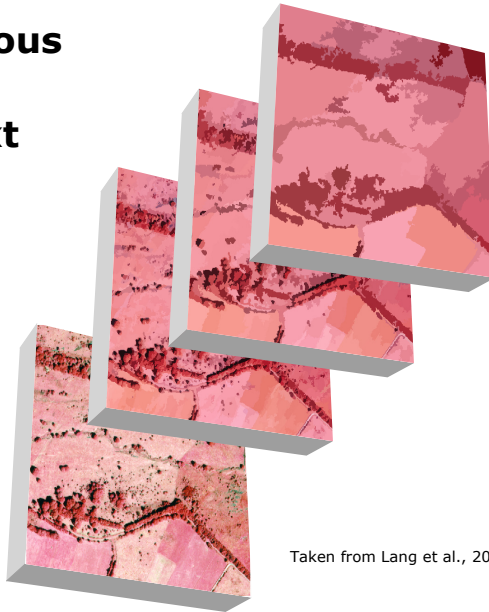
❷ » Remote sensing point of view: Grain refers to the smallest intervals in an observation set, while extent refers to the range over which observations at a particular grain are made (O'Neill and King, 1997). From a remote sensing perspective, grain is equivalent to the spatial, spectral and temporal resolution of the pixels composing an image, while extent represents the total area, combined bandwidths and temporal duration covered within the scene (Hay et al., 2001). « (Hay et al., 2003)

## Scales and scaled representation (4)

- **Representation in various levels simultaneously**
- **Each knows it's context and hierarchical neighbourhood**



**Object hierarchy**



Taken from Lang et al., 2004

❶ » The problem is not to choose the correct scale of description, but rather to recognize that change is taking place on many scales at the same time, and that it is the interaction amongst phenomena on different scales that must accompany our attention (Levin (1992), p. 1947). « (Burnett, Blaschke; 2003)

## Decomposability and Holarchy

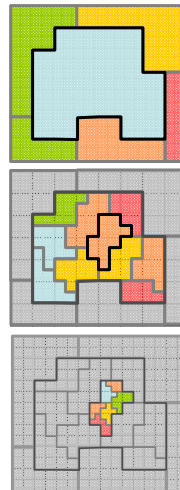
- **Landscape as a system**

Consisting of interacting subsystems

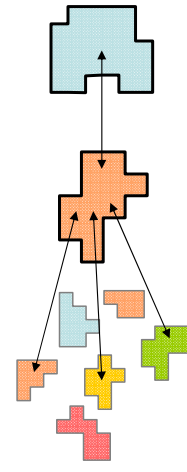
- **Decomposition**

Separating a system into its components according to their scale and ordering them hierarchically

## Separation



## Hierarchical organization



❶ » Systems structure can also be characterized in terms of hierarchical interaction. A system is nearly-decomposable into constituting sub-systems and hierarchical organization is an overarching principle in all living systems (Simon, 1973). « (Lang, Burnett, Blaschke; 2004)

❷ » Decomposition is the process of separating and ordering system components according to their temporal or spatial scales or both. This is done by the means of multi-scale analysis. « (Burnett, Blaschke; 2003)

❸ » In hierarchy theory, objects are apparent as separable entities because of differences in flux rates, by gradients (Simon, 1962; Koestler, 1967). Relatively strong gradients will evoke more apparent boundaries, or local heterogeneity. Boundaries manifest both between objects at the spatial (and temporal) scale and between objects at different scales. « (Burnett, Blaschke; 2003)

❹ » Interactions tend to be stronger and more frequent *within* a level of hierarchy than *among* levels (Allen and Starr, 1982). This important fact enables the perception and description of complex systems by decomposing them into their fundamental parts and interpreting their interactions (Simons, 1962). « (Hay et al.; 2003)

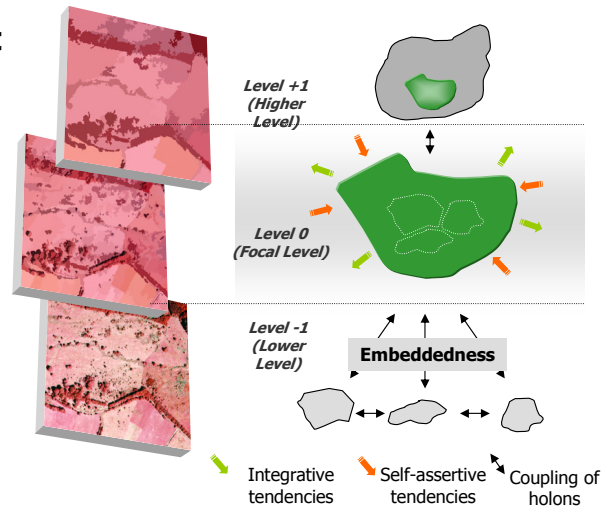
## Decomposition and Holarchy (2)

- **Subsystems are rather independent from each other**

Horizontal and vertical coupling

- **But still have integrative tendencies**

Part-being of constituting elements

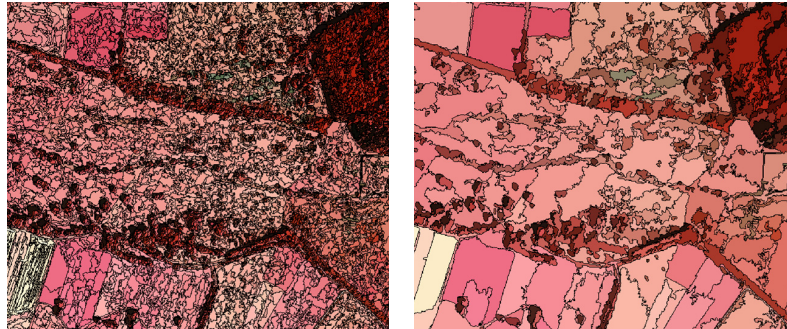


❶ » The existence of vertical and horizontal loose couplings is exactly the basis of the decomposability of complex systems [...]. While the word “loose” suggests “decomposable”, the word “coupling” implies resistance to decomposition. Strictly speaking, complete decomposability only occurs when coupling between components becomes zero, which seems a trivial case because, by definition, a system is composed of interacting parts. Thus, hierarchical complex systems are only nearly completely decomposable or nearly decomposable (Simon, 1962, 1973). « (Wu, 1999)

❷ » This means that constituent elements of a larger system operate in a rather independent way [...]. Koestler (1967) has elaborated on the idea that a system on any level can be considered as both self-assertive and integrative, and suggested the term ‘holon’ (from Greek *holos* and *-on*) to emphasize this dialectic property. Whereas the first underlines the independent and singular character of a system, the latter emphasizes the part-being of a constituting element. In order to highlight the dynamic character of a hierarchy of nested holons, Koestler proposed the term ‘holarchy’. Landscape ecologists (Naveh, 1995; Naveh, Lieberman, 1994) have tried to apply general systems theory to derive organizational forms above and beyond the organism/community dialectic. According to these an ecotope can be defined as a concrete above-organism holon and at the same time as a constituting element of larger landscape mosaics. « (Lang, Burnett, Blaschke; 2004)

## Multi-scale approach with remote sensing

- **Definition of fundamental objects in remote sensing images**
  - Integrated objects vs. aggregate objects
- **Interaction of objects within and across scale domains**
  - What scales should be chosen for the different objects?
  - At what scale should hierarchies be established?



❶ » However, to achieve this [description of complex systems by decomposition/ multi-scale approach], objects, i.e., fundamental parts, have to be clearly defined. Rowe (1961) distinguishes between two fundamental object types: integrated objects and aggregate objects. Integrated objects contain structurally organized parts, while aggregate objects occupy a common area, but have no structural organization. Furthermore, integrated objects have intrinsic scale, whereas aggregates do not. From a remote sensing perspective, image objects are integrated objects that exhibit an intrinsic scale and are composed of structurally connected parts, i.e., H-res pixels. « (Hay et al., 2003)

❷ » To understand how image objects interact within and across scale domains, we need techniques to automatically define them in remote sensing data and the ability to link them within hierarchical structures. The primary unknowns to achieve this are: What are the 'optimal' scales to evaluate the varying sized, shaped, and spectrally distributed image-objects within a scene? At what scale should hierarchies be established? We suggest that there is no single 'optimal' scale for analysis. Rather there are many optimal scales that are specific to the image-objects that exist/emerge within a scene (Hay and Niemann, 1994; Hay et al., 1997, 2001). « (Hay et al., 2003)

❸ » A single scale of image objects is probably insufficient to characterize the hierarchical nature of Nature. O'Neill et al. (1986) has suggested that at least three levels should be used. « (Lang, Burnett, Blaschke, 2004)

❹ » 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. « (Lang, Blaschke; 2003)

# OBIA – Tutorial

Introduction to object-based image analysis

## Chapter 3 Knowledge representation

**Z**\_GIS

Centre for Geoinformatics, Salzburg University

## Outline

- **What is knowledge?**
- **Cognition networks**
- **Image understanding**
- **Production systems vs. adaptive learning**

## What is knowledge?

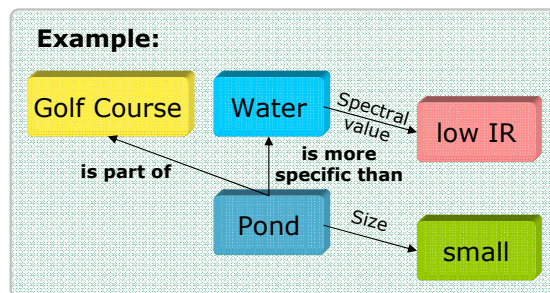
- **Knowledge plays key role in image interpretation part of remote sensing**
  - Implicit human knowledge is supplemented with explicit knowledge by training
- **Artificial intelligence distinguishes knowledge into:**
  - Procedural knowledge (specific computational functions)
    - Represented by a set of rules
  - Structural knowledge (how concepts of a domain are interrelated)
    - For image understanding in remote sensing: Are there links established between image objects and 'real world' geographical features?
    - Rich semantic content
    - Represented by a semantic network

❶ » Knowledge plays a key role in the interpretation-oriented parts of the remote sensing process chain (Campbell, 2001). We have a huge store of implicit knowledge at our disposal and a significant part of it is used in image interpretation [...]. From an artificial intelligence (AI) perspective knowledge can be distinguished in procedural and structural knowledge. Procedural knowledge is concerned with the specific computational functions and can be represented by a set of rules. Structural knowledge implies how concepts of a domain are interrelated: in our case that means, which links between image objects and 'real world' geographical features are established. It is characterized by high semantic contents and therefore by far more difficult to tackle with. « (Lang, 2005)



## What is knowledge? (2)

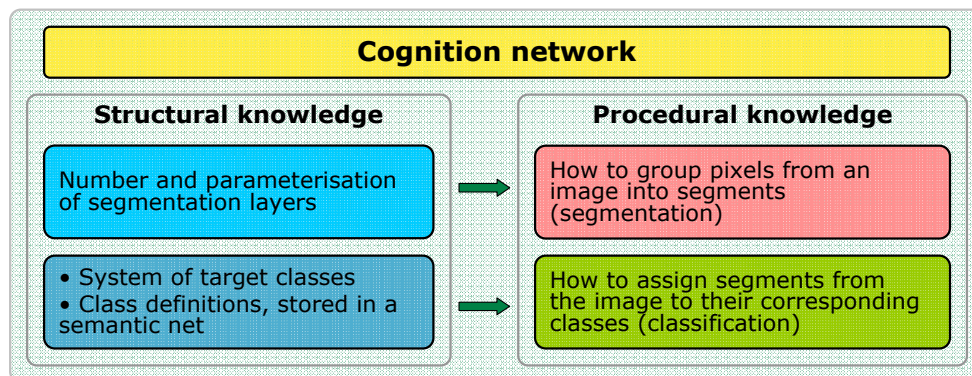
- **Structural knowledge can be organized in knowledge organizing systems**
  - Realised as graphic notations such as semantic nets or frames
  - Semantic knowledge representation (using inheritance concept, e.g. 'is part of', 'is more specific than', 'is an instance of') as formal framework for image analysis
- **Semantic net**
  - To be created
  - Control over existing connections, once established
  - Transparency and operability



compare to Pinz, 1994; p. 94

❶ » Structural knowledge can be organized in knowledge organizing systems (KOSs), realised by graphic notations such as semantic networks (Ibrahim, 2000; Pinz, 1994; Liedtke et al., 1997; Sowa, 1999), and more mathematical theories like the formal concept analysis (FCA, Ganter & Wille, 1996). Within image analysis semantic nets and frames (Pinz, 1994) offer a formal framework for semantic knowledge representation using an inheritance concept (is part of, is more specific than, is instance of). In section 3.3.5 a semantic net is used as part of a cognition network. Though a semantic net needs to be created manually, it allows for controlling each and every existing connection once being established. With increasing complexity the transparency and operability will reach a limit. Bayesian networks are manually built, but the weighting of the connections can be trained, though it has to be trained for every connection. « (Lang, 2005)

## Cognition network



- **Purpose of the cognition network**
  - Documentation of every step and setting in the process of image understanding
- **Usefulness of the approach**
  - Transparent, a suitable means for communication
  - Reproducible
  - To a high degree comprehensible
  - Technically transferable to other scenes

❶ » Within object-based mapping a cognition network (Binnig, et al., 2001) is established which serves as a conceptual framework for the number and parameterization of segmentation layers and the definition of classes. Especially when multi-scale segmentation and object-relationship modelling (MSS/ORM, see Burnett & Blaschke, 2003) is being applied, such a conceptual outline seems to be indispensable. Any step and setting during the entire classification process is documented, and can be assessed and adopted if needed. Although the result is not necessarily more accurate, it can be reproduced and the process is to a high degree comprehensible. The formalized approach of analysis (i.e. the class definitions and composition and the documentation of the workflow and settings in the semi-automated process) technically allows for a transfer of the classification to other scenes (Lang & Langanke, 2004; Benz, et al., 2004). « (Lang & Langanke, 2006)

❷ » The establishment of a cognition network encapsulates the required knowledge for building up a rule set. Though not empirically proved as yet in this case, the transferability seems to be rather a matter of adapting the parameterization (Schöpfer et al., 2005). « (Lang & Langanke, 2006)

❸ » A triple-SN [*self-organizing, semantic, self-similar network*] is essentially a kind of hierarchical world knowledge network containing knowledge about objects, their properties, and their relations, as well as processing knowledge about what to do when certain kinds of objects are present in the real world. By "real world" we mean the varying input that interacts with the triple-SN. This input could be an image, a text, or any complex structure. « (Binnig et al., 2002)

❹ » The networking and structuring of the input transforms information into knowledge and, to a certain extent, constitutes an automatic "understanding". « (Binnig et al., 2002)

❺ » By an optimally hierarchically structured network the classifications and segmentation in an alternating manner, the unstructured input (the single pixels) evolves to a logical structural arrangement of objects according to the cognition network: *"the creation of objects and their relations on and across different hierarchical levels is equivalent to transform information into knowledge"*. « (Binnig et al., 2002)

## Image understanding

## 1. Definition

Image understanding (IU) is a process leading to the description of the image content  
(= reconstruction of an imaged scene)

## 2. Extent of IU

Reaching from signals (image data) to a Symbolic representation of the scene content

## 6. Involved disciplines

- Image processing
- Pattern recognition
- Artificial Intelligence

## Image understanding

## 4. Output description

- Description of real-world objects and their relationships in the specific scene
- Resulting in thoroughly described features (not mere listing and labelling of features)

## 3. Conditions for IU

Outcome depends on the domain of interest of the interpreter, defined by:

- Underlying research question
- Specific field of application
- Pre-existing knowledge and experience of the interpreter

## 5. Knowledge input

Process is driven by

- utilisation of procedural knowledge
- transformation of structural knowledge

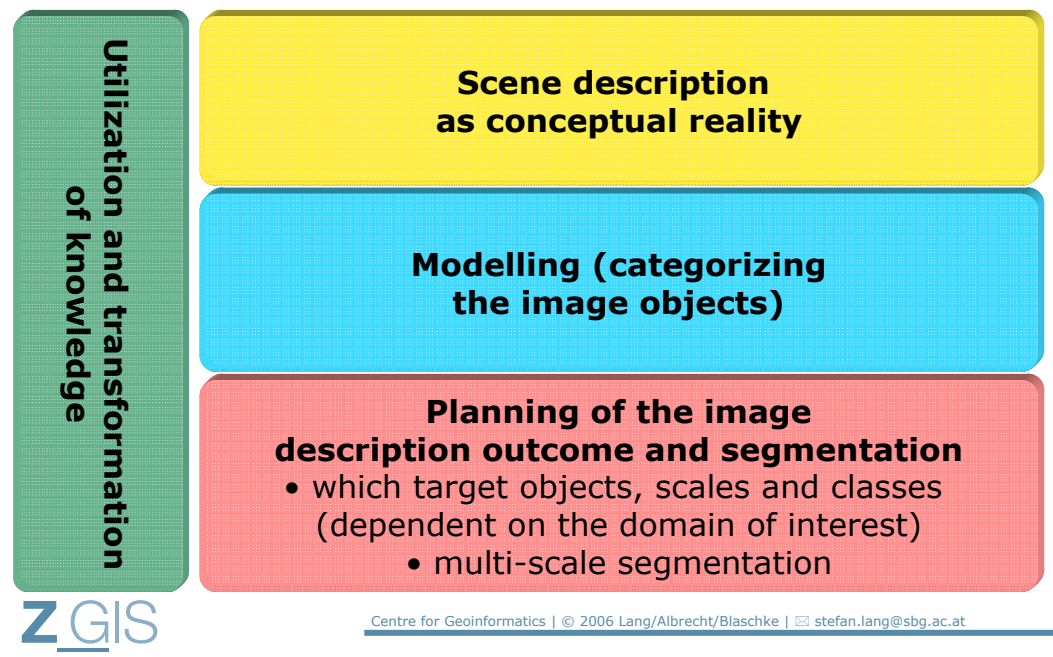
❶ » Major advances and efforts have been made in AI including feature detection algorithms. Yet image understanding is more than just feature extraction on a specified scene (Ibrahim, 2000). In interpreting an ecological scene we are dealing with a high variety of instances of different target classes (& see paper E2). Image understanding (IU) is commonly regarded as a process, through which in the end we arrive at a description of the image content, i.e. the reconstruction of an imaged scene (Pinz, 1994).

Document image understanding employing optical character recognition is often taken as a striking example of successful AI implementation. Whereas in this arena the target symbology is rather clear and well defined, it remains a challenge to correctly identify poor handwriting.

Usually today's image understanding process does not end up with a mere listing and labelling of image primitives (cp. Pinz, 1994). Within image interpretation of EO data the target scheme is usually much less defined and rather ambiguous, especially when dealing with scenes representing natural features. IU more and more aims at providing highly aggregated and application-related information in a defined domain of interest; therefore aiming at a 'full' scene description and is equipped with crucial elements from formal knowledge to reach from signals to a symbolic representation of the scene content. One focus is on the description of (real-world) objects and their relationships in the specific scene (Winston, 1984).

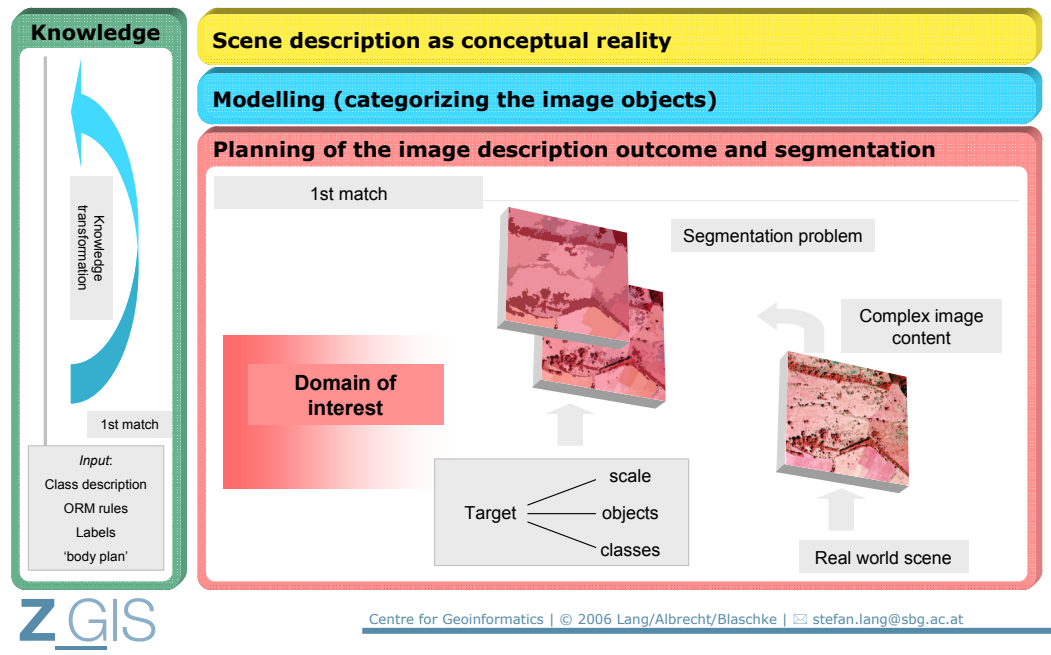
Describing a scene always depends on a conceptual framework constructed by components like (a) the underlying research question within (b) a specific field of application and (c) pre-existing knowledge and experience of the operator. Gaining insight into the content of a scene requires familiarity with the potential content as being realized by personal acquaintance with the imaged area or general experience. This implies recognition of the imaged features and their systemic structure. That means the inference from specific arrangement of units in 2-dimensional space to a certain application context. The field of image understanding is interwoven with disciplines, such as image processing, pattern recognition, and AI. Image processing provides the sources in a pre-processed way. Pattern recognition has made enormous advances in the last decade has incorporated methods of knowledge representation and expert systems to a wide degree. AI covers a major field of computer-based image understanding, yet a certain portion is left uncovered that is related of unsolved challenges of knowledge transfer to an automated system (Ibrahim, 2000). « (Lang, 2005, p 49f.)

## Image understanding (2)



❶ » By forming the conceptual link to human perception image segmentation is considered an essential prerequisite for image understanding (Gorte, 1998). Object-oriented image analysis (OBIA) offers valuable methodological assets in breaking down scene complexity into meaningful image primitives. By providing “candidate discretizations of space” (Burnett & Blaschke, 2003) a scene can be modelled in adaptive scales according to the domain of interest. The relevant steps of using an object-based image analysis approach are depicted in [the following slides]. « (Lang, 2005)

## Image understanding (3)

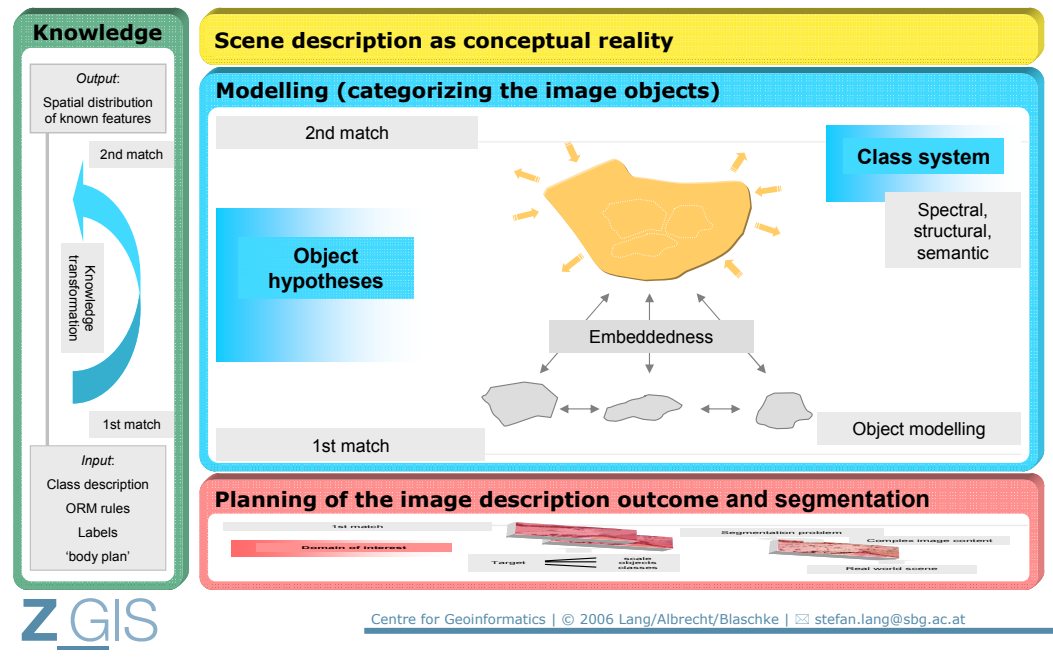


❶ » A profound prerequisite of image object modelling is the provision of a clear underlying concept regarding the domain of interest. This comprises an understanding of the target scale, the target object set, and the target class scheme. Note that the domain of interest of a skilled interpreter may differ from that of a simple user; the experience of the former makes him specifically look for certain features, whereas the latter is mainly interested in the information he wants to obtain. « (Lang, 2005)

❷ » One of the most important aspects of understanding imagery is information about image context. There are two types of contextual information: global context, which describes the situation of the image – basically, time, sensor and location – and local context, which describes the mutual relationships of image regions. [...] Image objects have to be linked to allow low and high-level semantic and spatial context. The image object network becomes a hierarchical image object network, when image objects of different scale at the same location are linked. Together with classification and mutual dependencies between objects and classes, such a network can be seen as a spatial semantic network. « (Benz et al., 2004)

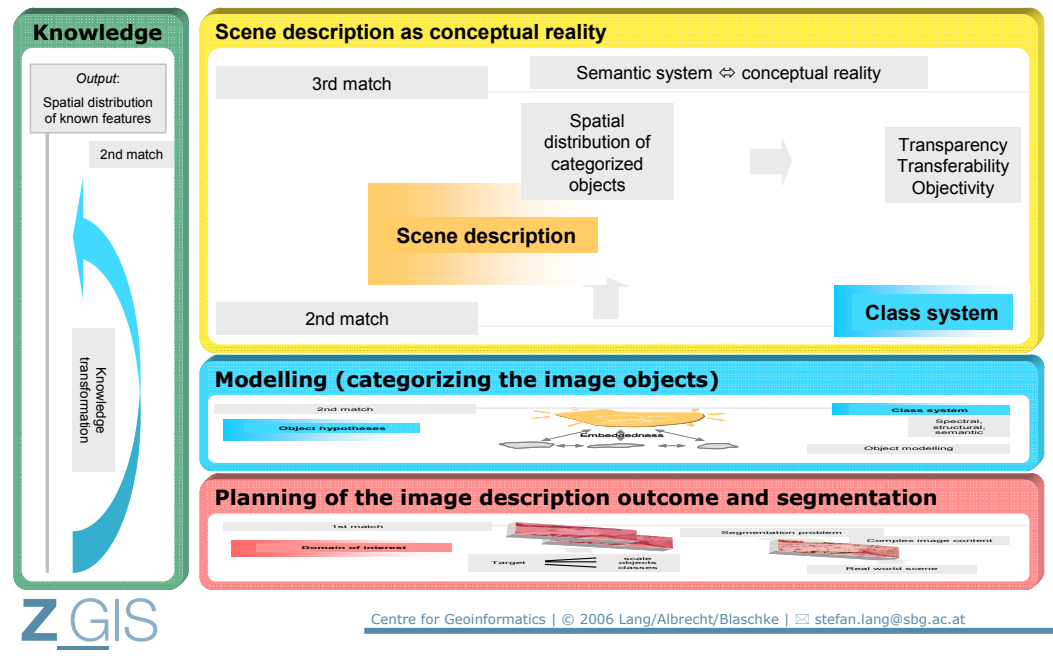
❸ » From image domain to application domain: Starting from a real-world scene subset captured by an image of high complex content the first step comprises the provision of scaled representations by aggregation information and reducing complexity. The multi-scale segmentation should be steered by a clear picture in mind of how target objects are structured by sub-level primitives (or super-level aggregates). The critical choice of appropriate segmentation levels makes up the 1st match of a scene-centred view with conceptual reality. « (Lang, 2005)

## Image understanding (4)



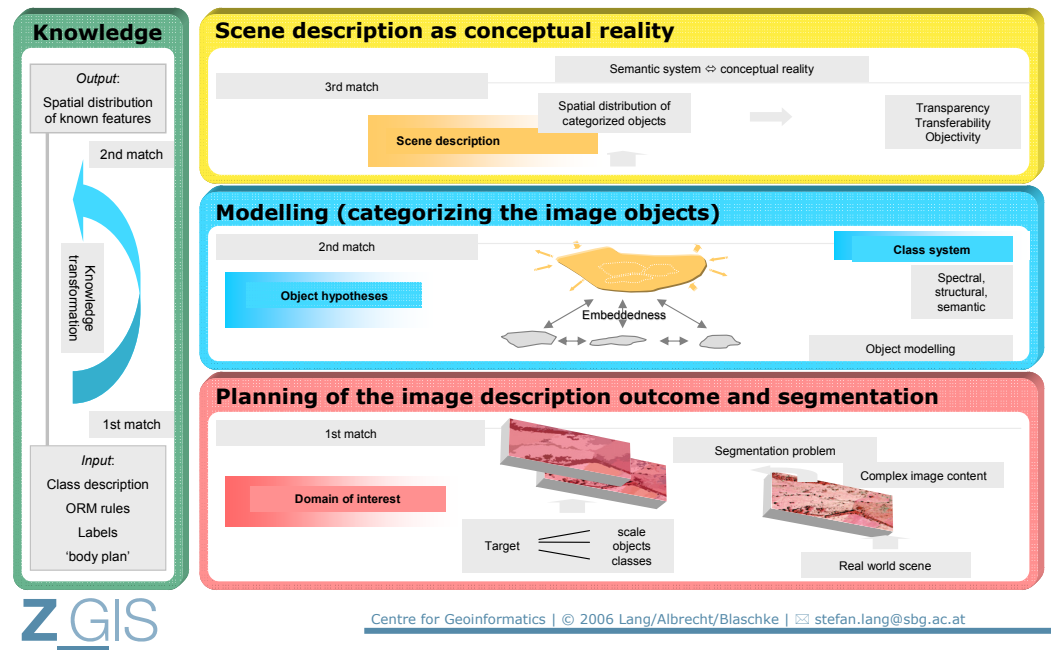
❶ » Having object hypotheses in mind (Bruce & Green, 1990) the modelling is realized by encoding expert knowledge into a rule system. Modelling aims at categorizing the image objects by their spectral and spatial properties and their mutual relationships (Lang & Langanke, 2006). In accordance to various possibilities of grouping this kind of knowledge one can differentiate between spectrally, structurally, and semantically defined classes. This is the 2<sup>nd</sup> match and the shift to an object-centred view is accomplished. « (Lang, 2005)

## Image understanding (5)



❶ » In the final stage of image understanding when arriving at a scene description an exhaustive categorization of any object is achieved. This representation of the image content meets the conceptual reality of the interpreter or user. « (Lang, 2005)

## Image understanding (6)



❶ » The entire process is characterized by the utilization and transformation of knowledge. Knowledge input is realized by establishing a body plan for the relevant features through class descriptions, modelling, rules and labelling. The procedure makes expert knowledge explicit. Knowledge is stepwise adapted and improved through progressive interpretation and object modelling. Knowledge will be enriched from analyzing unknown scenes and the transfer of knowledge may incorporate or stimulate new rules. « (Lang, 2005)



# OBIA – Tutorial

Introduction to object-based image analysis

## Chapter 4 Image segmentation

**Z**GIS

Centre for Geoinformatics, Salzburg University

## Outline

- **Short history**
- **Image segmentation in remote sensing**
- **Categories of segmentation algorithms**
  - Histogram-based/pixel-based segmentation
  - Region-based segmentation
  - Edge-based segmentation
- **Scale space analysis - Image Objects and Blobs**
- **Multiresolution Segmentation algorithm**
  - Region merging and scale
  - Local mutual best fitting approach
  - Colour and shape homogeneity

Chapter 4 discusses segmentation algorithms that can be used to derive objects from an image. We also introduce multi-resolution segmentation. We start with a short history on image segmentation, with a focus on applications in remote sensing. We will explain the concept of image segmentation highlighting the importance of homogeneity as a key criterion of the extracted regions. The practical aim of image segmentation is to find an optimum match between image objects and geographical features of the real world objects.

We may broadly distinguish between three categories of segmentation algorithms, i.e. histogram-based, edge-based, and region-based segmentation algorithms. Histogram-based approaches perform segmentation within the feature space. These clustering approaches differ from other approaches mainly by ignoring the spatial dimension in real space. It is a form of unsupervised classification, leading to classes but not to spatially contiguous regions. Region-based segmentation algorithms, as the name indicates, deliver regions. They can be differentiated into region growth, region merging and splitting techniques and various derivations or combinations. Region growth starts with a set of seed pixels from which regions grow by adding neighboring pixels as long as a homogeneity criterion applies. Region merging starts with initial regions, e.g. single pixels, and merges them together until a scale-dependent threshold in size is reached. The splitting algorithms divide an image into regular sub-regions (e.g. squares), which again will be divided until a certain level of homogeneity is reached. The combination of split and merge is realized in the split-and-merge algorithm. Here, the image is subdivided into squares of a fixed size. Heterogeneous squares are subdivided again and homogeneous squares can be merged together. Edge-based algorithms are searching for edges that occur between homogeneous areas. This process usually includes filtering and enhancement of the image prior to the detection of the edges. The detected edges (groups of pixels) need to be combined in order to form a boundary.

The multi-resolution segmentation algorithm implemented in Definiens software is a region-based, local mutual best fitting approach. The algorithm is utilizing a combined homogeneity and shape concept and allows several segmentation levels to be established within one image.

## Short history

- **1980ies: first developments of image segmentation**  
Major driver: industrial vision
- **Rarely made use of in remote sensing until the late 1990ies**  
Reason: little progress in segmentation of multi-band images, algorithms not made for EO data
- **Since then a high number of segmentation algorithms has been developed**
  - Availability in commercial products made use of segmentation of EO data more common

❶ » Although image segmentation techniques are well known in some areas of machine vision (see Narendra & Goldberg, 1980, Fu & Mui, 1981, Haralick & Shiparo, 1985, Cross et al., 1988) they are rarely used for the classification of earth observation (EO) data. One of the main reasons for this is that most of these algorithms were developed for the analysis of patterns, the delineation of discontinuities on materials or artificial surfaces, and quality control products, in essence. These goals differ from our goals: the discretisation of EO remote sensing imagery aims at the generation of spectrally homogeneous segments, which show the inherent dimensions/objects of the images. « (Blaschke et al., 2004)

❷ » Kartikeyan et al. (1998: 1695) state: “Although there has been a lot of development in segmentation of grey tone images in this field and other fields, like robotic vision, there has been little progress in segmentation of colour or multi-band imagery.” Especially within the last two years many new segmentation algorithms as well as applications were developed, but not all of them lead to qualitatively convincing results while being robust and operational. « (Blaschke & Strobl, 2001)

❸ » The number of methods for segmenting an image is legion (for an overview, see Haralick and Shapiro, 1985; Ryherd and Woodcock, 1996; Kartikeyan et al., 1998; Baatz and Schäpe, 2000; Schiewe et al., 2001). Common approaches use region growing or thresholding algorithms, but many derivatives for specific applications such as grey scale, hyper spectral images or data fusion of different sensors exist. « (Burnett & Blaschke, 2003)

❹ » As stated above, the idea of segmentation is not new but it is becoming more widespread within the EO/RS community recently. While the foundations of the basic principles were laid out in the 80ies (see Haralick & Shiparo, 1985) and various applications demonstrated the potential in the following years for environmental applications (e.g. Véhel & Mignot, 1994, Panjwani & Healey, 1995, Lobo et al., 1996), mainly the availability in commercial software packages catalysed a boost of applications more recently. « (Blaschke et al., 2004)

## Image segmentation in remote sensing

- **Division of an image into regions so that**
  - the **whole scene is covered** by regions (spatially continuous, exhaustive segmentation)
  - the **regions do not overlap**
  - the **regions are homogeneous** within themselves
  - the homogeneity criteria of **neighbouring regions differ**
- **Region (token):**
  - aggregate of pixels grouped together (*directly or indirectly*)
- **Homogeneity as overarching principle**
  - 'relatively' homogeneous regions reflect better the 'Near-decomposability' of natural systems
  - High heterogeneity creates boundary to neighbouring patches, low remaining heterogeneity within patches
  - Homogeneity criterion: grey value, colour, texture, form, altitude, etc.

Segmentation rests upon four mathematical principles of (cp. Horowitz and Pavlidis, 1974): (1) Union set of regions makes up the image, (2) Regions are not allowed to overlap, (3) Homogeneity criterion H applies, (4) Homogeneity criteria of neighbouring regions differ.

❶ » Image analysis implies to deal with image semantics. In most cases important semantic information to understand an image is not represented in single pixels but in meaningful image objects and their mutual relations. Furthermore many types of image data are more or less textured. Airborne data, radar or VHR-satellite data are playing an increasing role in remote sensing. In most cases, analysis of such textured data can only be successful when they are segmented in meaningful 'homogenous' areas. « (Baatz, Schäpe; 2000)

❷ » Segmentation is the division of an image into spatially continuous, disjoint, and homogeneous regions. Segmentation is powerful and it has been suggested that image analysis leads to meaningful objects only when the image is segmented in 'homogeneous' areas (Gorte, 1998, Molenaar, 1998, Baatz & Schäpe, 2000) or into 'relatively homogeneous areas'. The latter term reflects better the 'near-decomposability' of natural systems as laid out by Koestler (1967) and we explicitly address a certain remaining internal heterogeneity. The key is that the internal heterogeneity of a parameter under consideration is lower than the heterogeneity compared with its neighbouring areas. « (Blaschke et al.; 2004)

❸ » Because we believe that 'natural' hard boundaries are antithesis to a view of landscapes as continuum mosaics, we turn to HPD theory for guidance. With multi-scale segmentation, we are searching for the gradient of flux zones between and within holons (patches): areas where the varying strengths of interactions between holons produce surfaces. [...] Methodologically, this equates to searching for changes in image object heterogeneity/homogeneity. « (Burnett, Blaschke; 2003)

## Categories of segmentation algorithms

- **Pixel-based or histogram-based**

Thresholding techniques  
Segmentation in the feature space

- **Region-based**

Region growing, merging and splitting

- **Edge-based**

Laplace filter, Sobel-operator, representativeness, ...

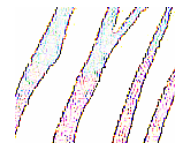
- **Non-image-related/  
non content expecting**

Tiling image with a honeycomb or chessboard structure

*Finding homogenous objects*



*Detecting edges between objects [and background (matrix)]*



*Regions defined without information from the image*

Traditional image segmentation methods have been commonly divided into three approaches: pixel-, edge and region based segmentation methods.

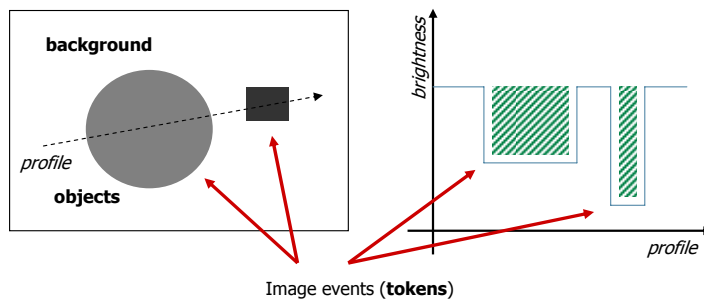
❶ » Pixel based methods include image thresholding and segmentation in the feature space [...]. In edge based segmentation methods, the aim is to find edges between the regions and determine the segments as regions between the edges [...]. Region-based segmentation algorithms can be divided into region growing, merging and splitting techniques and their combinations. « (Blaschke et al., 2004)

❷ » Technically, there are a number of possibilities how to segment an image. Most approaches can be grouped into two classes, namely edge-based algorithms and area-based algorithms. This classification also includes fractal-based approaches aiming at detecting discontinuities as well as fractal-based or texture-based algorithms (Salari & Ling 1995, Ryherd & Woodcock 1996) aiming at finding homogeneous areas. A recent survey of some competing approaches lists advantages but also some potential pitfalls for extracting geoinformation and useful landscape elements on real surfaces (Blaschke et al. 2000). Locally extracted histograms provide a good representation of the local feature distribution, which captures substantially more information than the frequently used mean feature values. The 'representativeness approach' (Hofmann & Böhner 1999) and other boundary- forming techniques (Schneider et al. 1997, Banko et al. 2000) and segmentation approaches (Gorte 1998, Molenaar 1998, Cheng 1999) provide good results in test areas but are not necessarily using all contextual information beyond the spectral information of neighbouring pixels such as texture, shape, directionality, spatial distribution within the study area, connectivity etc. However, from preliminary studies it is concluded, that the most promising recent developments are fractal approaches spearheaded by the developments of INRIA in Paris (Véhel and Mignot 1994) and Definiens AG in Munich (Baatz and Schäpe 2000). « (Blaschke & Strobl, 2001)

## Objects as Tokens

- **Objects** vs. background (**matrix**) in a grey value image
- → **tokens** (token set)

**Region (token):** aggregate of pixel grouped according to homogeneity criterion (*directly or indirectly*)



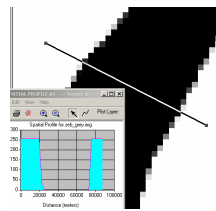
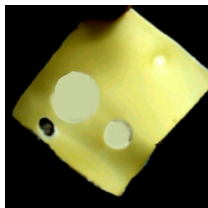
ZGIS

Centre for Geoinformatics | © 2006 Lang/Albrecht/Biaschke | [stefan.lang@sbg.ac.at](mailto:stefan.lang@sbg.ac.at)

Note that after Pinz (1994) tokens can overlap each other.

## Histogram-based

- **Histogram Thresholding:** simplest way to accomplish exhaustive regional segmentation
- → 'Swiss cheese' segmentation for punched parts
- One- or bimodal distribution of grey values, threshold has to be determined



Quickbird: band 1, thresholded



Quickbird: 6 classes Isodata classification



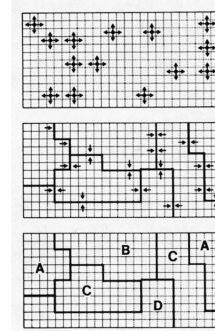
❶ » Some of the simplest approaches are all types of global thresholding. The spectral feature space is separated into subdivisions, and pixels of the same subdivision are merged when locally adjacent in the image data. Typically, this method leads to results of relatively limited quality. Oversegmentation and undersegmentation – i.e., separating into units which are too small or merging regions that do not belong to each other – take place easily without good control of meaningful thresholds. Local contrasts are not considered or not represented in a consistent way and the resulting regions can widely differ in size. « (Definiens, 2004)

❷ » Common alternatives are knowledge-based approaches. They try to incorporate knowledge derived from training areas or other sources into the segmentation process [Gorte, 1998]. These approaches mostly perform a pixel-based classification, based on clustering in a global feature space. Segments are produced implicitly after classification, simply by merging all adjacent pixels of the same class. In doing so, these approaches are typically not able to separate different units or objects of interest of the same classification. Furthermore, the information on which classification can act typically is limited to spectral and filter derivatives. « (Definiens, 2004)

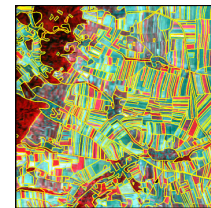
## Region-based segmentation

- **Region growing**

- Seed cells are distributed over image
  - Bottom up (randomly)
  - Top-down (content expected)
- Neighbours (4- or 8-neighbourhood) are included into region, if
  - they do not belong to another region yet
  - the homogeneity criterion H applies
- Two neighbouring regions are unified, if H applies



Campbell, p. 346



❶ » ECHO [extraction and classification of homogeneous objects; after Kettig and Landgrebe, 1975] searches for neighbouring pixels that are spectrally similar, then enlarges these groups to include adjacent pixels that have spectral values that resemble those of the core group. For example, the algorithm can first search for neighbourhoods of four contiguous pixels. For each group, it then tests members for homogeneity [...] Pixels that are not similar to their neighbours are rejected from the group [...].

Each of the homogeneous patches is then compared to each of its neighbours. If similar patches border each other, they are merged to form a larger patch. Patches are allowed to grow until they meet the edges of contrasting patches; when all patches reach their maximum extent within the constraints defined by the operator, the growing process stops. [...]

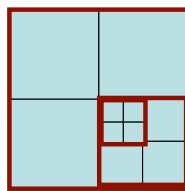
ECHO is a good example of a classifier that operates on fields of pixels rather than on each pixel in isolation. However, it performs the classification upon the average brightness of each patch, so it does not attempt to use image texture. « (Campbell, 2002, p. 346)

❷ » These algorithms depend on a set of given seed points, but sometimes suffering from lacking control over the break-off criterion for the growth of a region. Common to operational applications are different types of texture segmentation algorithms. They typically obey a two-stage scheme (Jain & Farrokhnia, 1991, Mao & Jain, 1992, Gorte, 1998, Molenaar, 1998, Hoffman et al., 1998). « (Blaschke et al., 2004)



## Region-based segmentation (2)

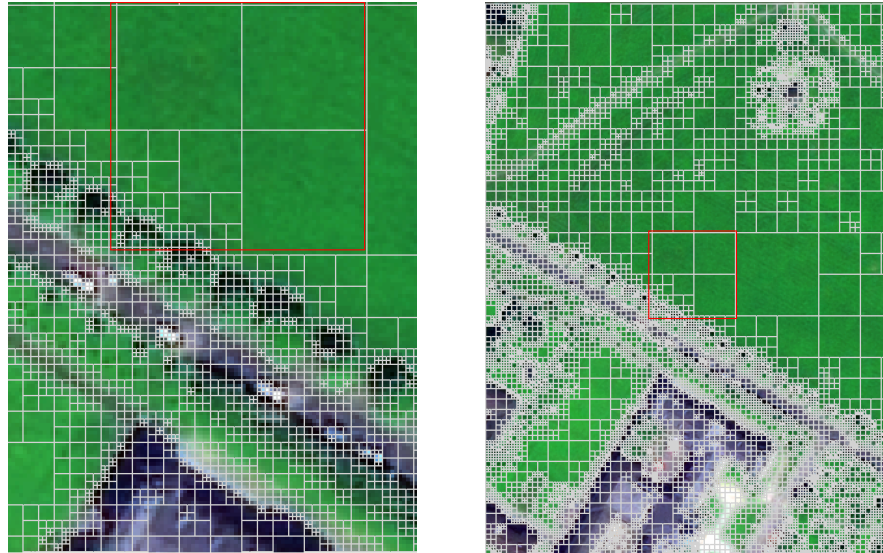
- **'Split and Merge'**
  - Combination **of coarse segmentation and merge**
  - Example: Quadtree
    - **Initially:** image as one object → division into 4 parts, if **H** does not apply
    - Resulting **quadtree** structure
    - Merge of homogenous quadtree areas



❶ » In region merging and splitting techniques the image is divided into sub regions and these regions are merged or split based on their properties. In region merging the basic idea is to merge segments starting with initial regions. These initial regions may be single pixels or objects determined with help of any segmentation technique. In region splitting methods the input usually consists of large segments and these segments are divided into smaller units if the segments are not homogeneous enough. In an extreme case region splitting starts with the original image and proceeds by splitting it into  $n$  rectangular sub-images. The homogeneity of these rectangles is studied and each rectangle is recursively divided into smaller regions until the homogeneity requirement is fulfilled. In both, region merging and splitting techniques, the process is based on a high number of pair wise merges or splits. The segmentation process can be seen as a crystallisation process with a big number of crystallisation seeds. The requirement for the maintenance of a similar size/scale of all segments in a scene is to let segments grow in a simultaneous or simultaneous-like way. « (Blaschke et al., 2004)

## Region-based segmentation (3)

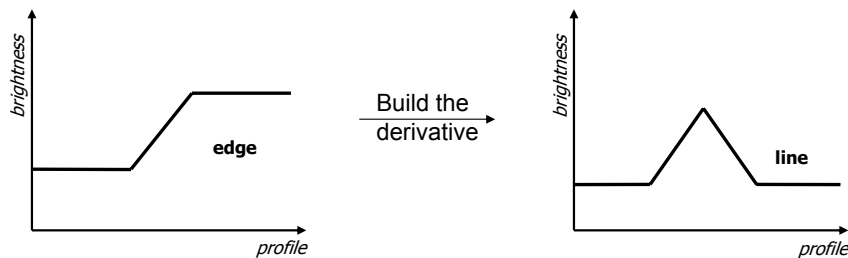
Quadtree structure in a QuickBird Image of Salzburg, generated with eCognition



Application of quadtree algorithm to a QuickBird scene (0.6 m spatial resolution, pan-sharpened).

## Edge-based segmentation

- Region-based segmentation makes sense when **large, compact** and **coherent** objects occur ('blobs')
- ➔ edge-based segmentation for **elongated structures**
- **Edge**: boundary between homogenous areas



Looking at image brightness as a function, the first derivative generates lines where the edge between homogeneous regions lies.

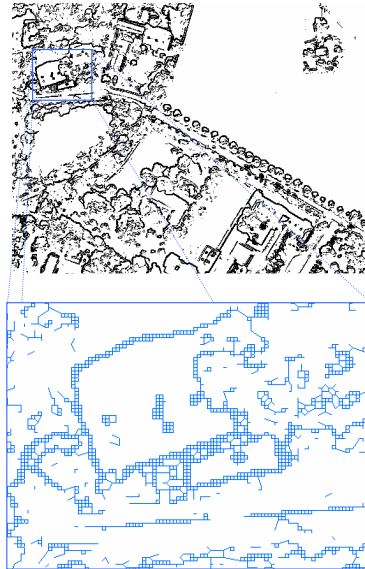
## Edge-based segmentation (2)

▪ **Workflow**

## 1. Edge detection

- Filtering – smoothing to decrease noise in the image
- Enhancement – revealing local changes in intensities
- Detection – select edge pixels, e.g. by thresholding
  - Closing of gaps / deleting artefacts
  - Combining, extending of lines

## 2. Linking the edge pixels to form the region boundaries



## Edge-based segmentation (3)

▪ **Enhancement filters (examples)**

- Sobel operator
- Laplace filter
- Compass edge
- ...

1	2	1
0	0	0
-1	-2	-1

horizontal

-1	0	1
-2	0	2
-1	0	1

vertical

0	1	0
1	-4	1
0	1	0

1	4	1
4	-20	4
1	4	1

-1	-2	-1
0	0	0
1	2	1

0°

-2	-1	0
-1	0	1
0	1	2

45°

-1	0	1
-2	0	2
-1	0	1

90°

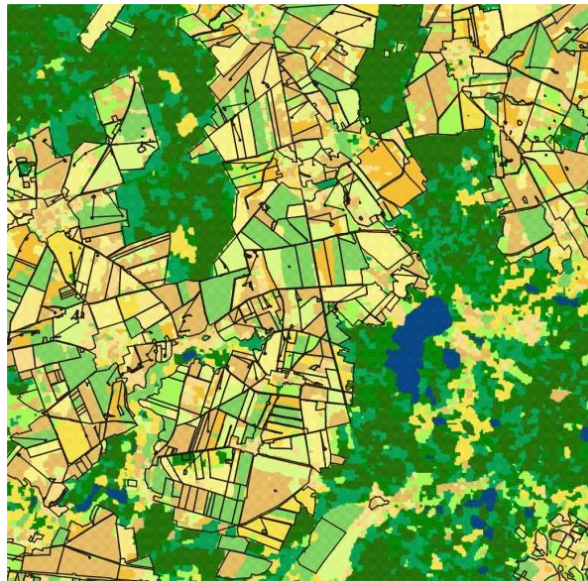
0	1	2
-1	0	1
-2	-1	0

135°

The resulting filter  
is the combination  
of all eight directions.

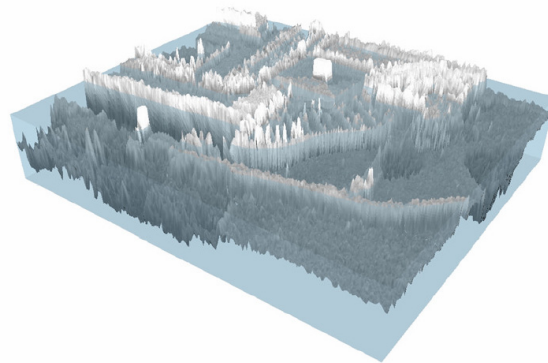
## Edge-based segmentation (4)

- **Segmentation by representativeness measures**
- Calculate a representativeness of each pixel for its neighbours
- The minima represent unsteadiness and indicate edges
- Vectorised minima delimit areas with similar spectral properties



❶ » Hoffman & Böhmer (1999) proposed an edge based method in which they calculate a representativeness of each pixel for its neighbours. The image segmentation is based on the representativeness values of each pixel. At first the values are calculated by a harmonic analysis of the values for each spectral channel. The minima in the matrix of representativeness – typically arranged in pixel lineaments – represents spatial unsteadiness in the digital numbers. For the image segmentation, the vectorised minima of the representativeness delimit areas consisting of pixels with similar spectra properties (spatial segments). A convergence index is combined with a single-flow algorithm for the vectorisation of the representativeness minima. A standardisation is performed through the calculation of a convergence index for every pixel in a 3 by 3 window. « (Blaschke et al., 2004)

## Watershed segmentation

■ **Watershed segmentation**

© by SLang 2005

ZGIS

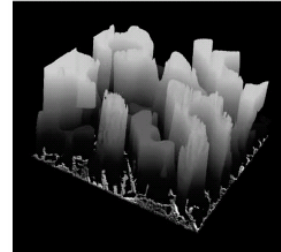
Centre for Geoinformatics | © 2006 Lang/Albrecht/Biaschke | [stefan.lang@sbg.ac.at](mailto:stefan.lang@sbg.ac.at)

❶ » A further relatively common procedure is watershed segmentation [Wegner & al 1997]. It got its name from the manner in which the algorithm segments regions into catchment basins. Typically, the procedure first transforms the original data into a gradient image. The resulting grey tone image can be considered as a topographic surface. If we flood this surface from its minima and if we prevent the merging of the waters coming from different sources, we partition the image into two different sets: the catchment basins and the watershed lines. The catchment basins should theoretically correspond to the homogeneous grey level regions of this image. This method works for separating essentially convex and relatively smooth objects of interest that even may touch slightly in relatively homogeneous image data. When it works, it is convenient, fast and powerful. However, for remote sensing data, which typically contain a certain noise and not always strong contrasts, this method is typically not able to achieve appropriate results. « (Definiens, 2004)

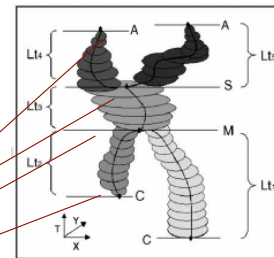
❷ » Watershed segmentation (Haris, 1998) is a segmentation method with a high intuitive character and transparency. Spectral reflectance is modelled as height values and segments are built at gradient magnitudes along similar altitude levels, just in analogy of water flowing into valleys between watersheds. Region growing stops when neighbouring flooding regions meet each other. Higher scale segmentation is achieved by decreasing the number of local minima. One problem of watershed segmentation is that in an initial stage the algorithm leads to over-segmentation (Ibrahim, 2000), and in many cases it has to be actively controlled by markers. It only depends on spectral likeness, so objects may vary significantly in size. « (Lang, 2005)

## Scale space analysis - Image Objects and Blobs

- Blobs
  - Compact image object
  - usually with a clear centroid
  - Temporal aspect → image events
- Scale space blobs
  - Significant within scale domains → defined by spatial extension, brightness, scale (4D)
- Detection
  - Repetitive Gaussian filtering
  - Local maxima are determined
  - Shapes are drawn around centroids (circle, polygon)

Hyper blob with  
image events

- Annihilation
- Split
- Merge
- Creation

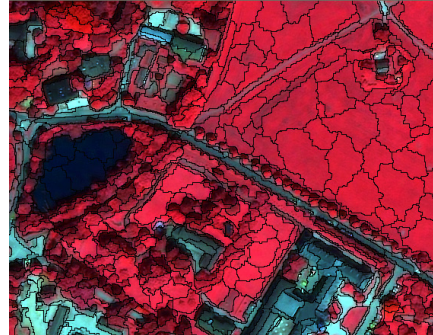


❶ » The following overview represents a multiscale approach as described by Lindberg (1994) that is composed of two principal components: Linear Scale-Space (SS) and Blob-Feature Detection. For a more detailed non-mathematical description of both, see Hay et al. (2002). [...] A SS multiscale representation 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, i.e. smoothing, of corresponding structures at finer scales. [...] This results in a scale-space cube or 'stack' of progressively 'smoothed' image layers, where each new layer represents convolution at an increased scale. [...] The second SS component we use is referred to as Blob-Feature Detection (Lindberg, 1994). The primary objective of this non-linear approach is to link structures at different scales in the scale-space, to higher order objects called 'scale-space blobs' and extract significant features based on their appearance and persistence over scales. [...] An important premise of SS is that structures which persist in scale-space are likely candidates to correspond to significant structures in the image and thus in the landscape. [...] Within a single hyper-blob four primary types of 'bifurcation events' may exist: annihilations (A), merges (M), splits (S) and creations (C). These SS-events represent critical components of SS analysis, as scales between bifurcations are linked together forming the lifetime ( $Lt_n$ ) and topological structure of individual SS-blobs. « (Hay et al.; 2003)



## Multiresolution Segmentation algorithm

- **Design goals**
  - „Convince the human eye“
  - Multi-resolution (strictly hierarchic)
  - **Similar resolution**
  - Reproducibility
  - **Universality**
  - Performance (i.e. speed 😊)
- **Overview**
  - Region merging technique
  - Decision heuristics
  - Homogeneity criterion
    - Colour homogeneity
    - Shape homogeneity
      - compactness and smoothness



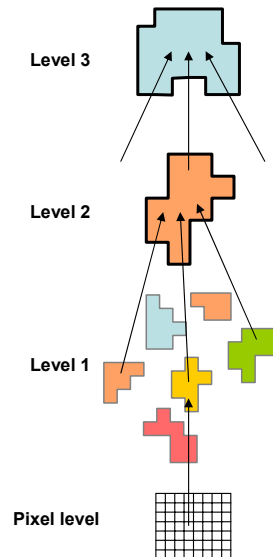
## Region merging and scale

- **Bottom up region merging technique**

- Starting with each pixel being a region
- A pair of regions is merged into one region, each merge having a merging cost (degree of fitting)
- Objects are merged into bigger objects as long as the cost is below a 'least degree of fitting' (scale parameter)  
= the merge fulfils the homogeneity criterion
- Starting points for merging distributed with maximum distance
- Pair wise clustering process considering smallest growth of heterogeneity

- **Establishing segmentation levels on several scales using different scale parameters**

(e.g. 2nd level based on 1st level: larger scale parameter results in larger objects consisting of the objects of the 1st level)

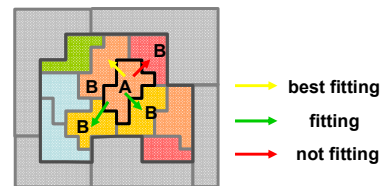
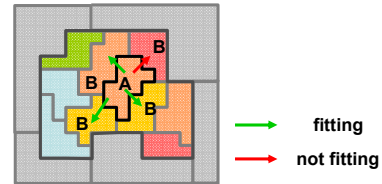


❶ » Segmentation needs to address a certain scale: does the application require information about single bushes or trees or about land cover units such as orchards or mires? Most segmentation approaches don't allow the user to specify a certain scale of consideration and a level of detail or generalization, accordingly. « (Blaschke, 2003)

❷ » The flexibility in performing scale-specific segmentation has led to a growing interest from landscape ecological applications of this approach. Within landscape ecology the hierarchical representation of process-relevant spatial units in various scale domains is one of the fundamental pillars (Wu, 1999). Segmentation can be used to provide a consistent set of image primitives to be used as landscape objects (Lang et al., in press; Burnett & Blaschke, 2003). « (Lang & Langanke, 2005)

## Decision heuristics

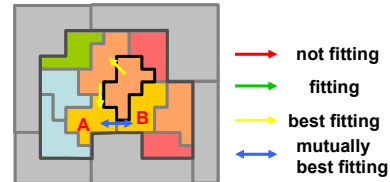
- Finding an adjacent object B for an arbitrary object A for merging them
- 1. Fitting: when the homogeneity criterion is fulfilled
- 2. Best fitting: when the homogeneity criterion is fulfilled, and the merge between B and A produces the best degree of fitting compared to the merge between A and any other adjacent object of A



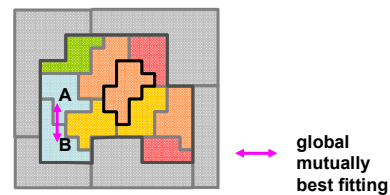
## Decision heuristics (2)

1. Fitting
2. Best fitting

3. **Local mutually best fitting:**  
find the best fitting object B for the object A, then find the best fitting object C for the object B. Confirm that object C is the object A, otherwise take B for A and C for B and repeat the procedure.  
=> find best fitting pair of objects in the local vicinity of A following the gradient of homogeneity



4. **Global mutually best fitting:**  
merge the pair of objects for which the homogeneity criterion is fulfilled best in the whole image



- **Distributed treatment order**

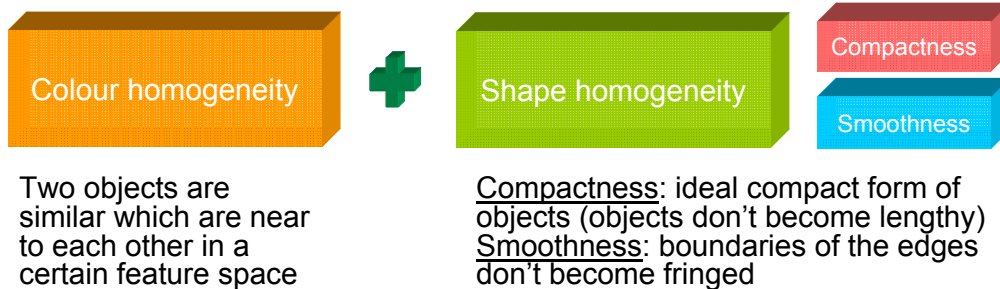
Use starting points with maximum distance to all other points treated before (treatment order defined over pixels or segments)

Find the best fitting object B for the object A and confirm that the best fitting object C for the object B is indeed the object A (homogeneity criterion is fulfilled mutually). If not, take B for A and C for B and try again.

## Homogeneity criterion

**Definition of the degree of fitting**

- Colour and shape homogeneity are weighted against each other
- Compactness and smoothness make up the shape homogeneity and are weighted against each other



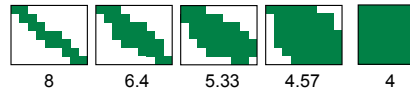
## Homogeneity criterion (2)



$$h_{\text{compact}} = \frac{l}{\sqrt{n}}$$

Relation between boundary length  $l$  of the object and the square root of the number  $n$  of the pixels of the object (square root of  $n$  equals the side of a square with  $n$  pixels)

$h_{\text{compact}} =$

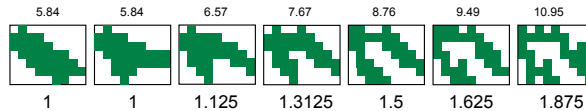


$$h_{\text{smooth}} = \frac{l}{b}$$

Relation between boundary length  $l$  of the object and the perimeter of the bounding box of the object (bounding box: shortest possible boundary length)

$h_{\text{compact}} =$

$h_{\text{smooth}} =$



The smooth shape homogeneity criterion only allows the objects to grow into shapes that have no fringed boundaries. The compact shape homogeneity criterion suppresses fringed boundaries, too, but additionally favours compact object shapes.

# OBIA – Tutorial

Introduction to object-based image analysis

## Chapter 5 Object-based classification

**Z**GIS

Centre for Geoinformatics, Salzburg University

## Outline

- **Introduction**
- **Available features for classification**
- **Sample- vs. rule-based classification**
- **Fuzzy classification**
- **Class hierarchy**
- **Class-related features**
- **Strengths of object-based classification**

In this chapter we explore object-based classification, its commonalities and specifics as compared to other pixel-based classification techniques in remote sensing. Looking at the peculiarities of object-based classification, there are at least two major aspects to be considered.

(1) We are dealing with a (largely) augmented, high-dimensional feature space. The number of features is virtually unlimited, and several hundreds can easily be constructed. Two reasons account for that: first, since objects are aggregates of  $n$  pixels we can use for any feature statistical derivatives like *mean*, *standard deviation* etc. Second there are a variety of geometrical features to be used. In addition, when working with hierarchical representations, there are features which characterize the relationships among objects on different hierarchical levels. All of these features can be used for classification. The 'art' is to find out which ones of these are really distinctive for a certain class. Consider taking samples for distinguishing between two pairs of classes, coniferous forest vs. deciduous forest and grassland vs. football ground. Whereas in either case you may establish shape-related features (area, rectangularity, etc.) for your samples, only in the latter case these are distinctive.

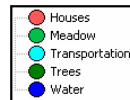
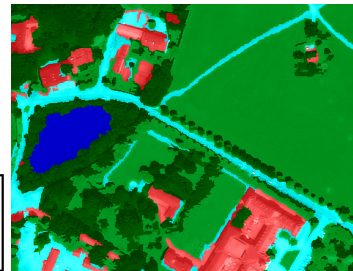
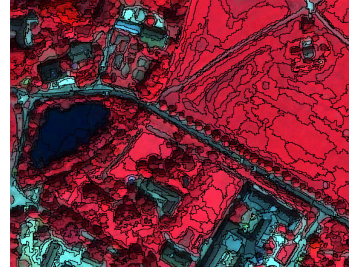
(2) Depending on the features we use for our classification, we are facing a high degree of freedom. Using shape information we can for example differentiate between grassland and a football ground. Using size information, we can tell small from big football grounds. Using spatial context information we can distinguish a big football ground within a city from a big football ground in the countryside. And so on. As this process becomes has a strong modeling component, working with samples alone may be limited. We therefore put a strong focus on formulating rule sets and perform rule-based classification in the sense of a productions system.



## Introduction

■ **Classifying**

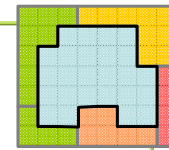
- Assigning objects to a certain class according to the class's description
- Feature space segmented into distinct regions which leads to a many-to-one relationship between the objects and the classes
- Definition of the class descriptions
  - Available object features
  - Sample-based or rule-based classification strategy



❶ » Usually classifying means assigning a number of objects to a certain class according to the class's description. Thereby, a class description is a description of the typical properties or conditions the desired classes have. The objects then become assigned (classified) according to whether they have or have not met these properties/conditions. In terms of database language one can say the feature space is segmented into distinct regions which leads to a many-to-one relationship between the objects and the classes. « (Definiens, 2004)

## Available features for classification

- **layer values**
  - mean
  - std-dev
- **geometrical properties**
  - size, shape, ...
- **textural properties**
  - layer value texture (e.g. mean of sub objects: std-dev)
  - shape texture (e.g. directions of sub objects)
- **hierarchical properties**
  - number of higher levels
  - number of super or sub objects

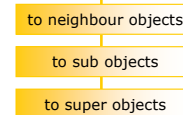


## object features



- **relations to classes of ...**
  - neighbour objects
  - sub objects (relative area of ...)
  - super objects
- **membership to ...**

## class related features



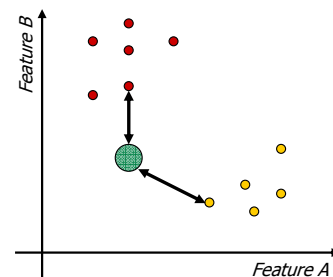
❶ » Pixels can be characterised by their spectral behaviour, as being expressed by their specific allocation in an  $n$ -dimensional feature space (where  $n$  equals the number of spectral bands of a sensor). With multispectral data (such as colour and false-colour orthophotos, QuickBird data, Landsat data)  $n$  ranges from 3 to 6 or 7. When considering image objects instead of pixels, a wealth of additional features can be used for characterisation. Besides statistically aggregated spectral features also geometrical and neighbourhood-related features can be used, by which the dimensionality of the resulting feature space is getting significantly higher and raises to virtually limitless extent. If still a network of image objects is created (full multi-scale decomposition, Burnett and Blaschke, 2002), even vertical relationships between super-objects and sub-objects in the object hierarchy can be used to expand the feature space even further. « (Lang, 2005)

## Sample- vs. rule-based classification

- **Sample-based classification**

- Define class membership by similarity to selected samples
  - Sample has to be representative for its class
  - Use features clearly distinguishing the sampled class from other classes
- Nearest neighbour classifier
  - Object will be assigned to the class whose samples are closest to it in the feature space

Useful approach, if knowledge about the scene's content is limited



❶ » Multispectral classification, in a pixel-oriented view, can be defined in a rather straightforward definition as the act of “*assigning pixels to classes*” (Campbell, 2001). This implies transforming reflectance values and ordinal scaled pixel values into nominal classes. In a statistical sense this means pixels are grouped according to their proximity in the spectral feature space, i.e. their spectral homogeneity. [...] By their respective spectral behaviour pixels can be allocated in an  $n$ -dimensional feature space, created by the  $n$  bands of a sensor. Classifiers are used to assign pixels to certain classes, which correspond to specific regions in the feature space. These regions are usually defined by training areas. Minimum distance (or nearest neighbour) calculates the Euclidean distance for each pixel to the centroid of each class and then assigns a given pixel to the nearest class. When working with objects instead of pixels, the feature space is significantly augmented; in this case a minimum distance classifier is more suitable, since normal distribution of features can hardly be presumed in all cases. [...]

Image objects can be classified by (1) classification based on samples and (2) using prior external knowledge stored in rule bases (Schöpfer et al., in press). The first is a supervised classification process, which unifies advantages of manual and statistical procedures. Rather than delineating training areas sample objects are iteratively selected. These sample objects should have the most representative and clearly distinguished features. « (Lang, 2005)

❷ » In comparison to pixel-based training, the object-based approach of the nearest neighbour requires less training samples: One sample object already covers many typical pixel samples and their variations. « (Definiens, 2004)

## Sample- vs. rule-based classification (2)

- **Rule-based classification**
  - Define a class by a rule on one feature or by rules on several features
  - Fuzzy or crisp rule definition
  - Hierarchical relations of classes
  - Rules can address different kinds of features
    - Object features
    - Class-related features
- **Advantages compared to sample-based classification**
  - Incorporation of expert knowledge in the classification
  - Formulation of complex class descriptions
  - Transparency (especially compared to neural networks)
  - Transferability

❶ » The second, more transparent and better transferable approach of encapsulating prior knowledge in rule-bases has been theoretically discussed in section 2.7.3. In Lang, Blaschke, 2005, Lang, Schöpfer, Langanke, 2005 and Lang, Langanke, 2006 we demonstrate how rule-based classification has been successfully applied. Integrating knowledge is a way to overcome the spectral similarity of different geographical features (Schöpfer et al., in press). Rules can address any of the spectral, spatial or hierarchical features the objects are equipped with. Because of the enormous range of potential features, a rule-based approach needs formalised representation of the target class system and the way how to interpret them beforehand. By fuzzification the degree of class membership can be modelled for a certain range of values. « (Lang, 2005)

❷ » Fuzzy logic classification is a simple technique, which basically translates feature values of arbitrary range into fuzzy values between 0 and 1, indicating the degree of membership to a specific class. It was chosen for the analysis of image objects in eCognition because

- by translating feature values into fuzzy values it standardizes features and allows the combination of features, even of very different range and dimension.
- it provides a transparent and adaptable feature description especially compared to neural networks.
- it enables the formulation of complex feature descriptions by means of logical operations and hierarchical class descriptions. « (Definiens, 2004)

## Fuzzy classification

▪ **Fuzzy vs. crisp classification**

- Uncertainties about class descriptions
  - Approximate real-world knowledge in its complexity
  - Uncertain human knowledge about the world, imprecise human thinking
  - Vague (linguistic) class descriptions
  - Other:
    - Uncertainty in sensor measurements
    - Class mixtures due to limited (spatial and spectral) resolution
- Possibilities of fuzzy classification
  - Express each object's membership to more than one class
  - Probability of an object to belong to other classes

⇒ No sharply defined boundaries between classes as in crisp classification

❶ » Supervised and unsupervised classification algorithms typically use *hard classification* logic to produce a classification map that consists of hard, discrete categories (e.g., forest, agriculture). Conversely, it is also possible to use fuzzy set classification logic, which takes into account the heterogeneous and imprecise nature of the real world. « (Jensen, 2005; p. 389)

❷ » Fuzzy classification is beside neural networks (Gopal and Woodcock, 1996) and probabilistic approaches (Curlander and Kober, 1992) a very powerful soft classifier. As an expert system for classification (Tsatsoulis, 1993) it takes into account:

- uncertainty in sensor measurements,
- parameter variations due to limited sensor calibration,
- vague (linguistic) class descriptions,
- class mixtures due to limited resolution. [...]

Avoiding arbitrary sharp thresholds, fuzzy logic is able to approximate real world in its complexity much better than the simplifying boolean systems do. Fuzzy logic can model imprecise human thinking and can represent linguistic rules. « (Benz et al., 2004)

❸ » It makes it also possible to express each object's membership in more than just one class or the probability of belonging to other classes, but with different degrees of membership or probabilities. With respect to image understanding these soft classification results are more capable of expressing uncertain human knowledge about the world and thus lead to classification results which are closer to human language, thinking and mind. « (Definiens, 2004)

❹ » Thus, fuzzy set does not have sharply defined boundaries, and a set element (a pixel in our case) may have partial membership in several classes. [...] All that has been learned before about traditional hard classification is pertinent for fuzzy classification because training still takes place, feature space is partitioned, and it is possible to assign a pixel to a single class, if desired. However, the major difference is that it is possible to obtain information on the various constituent classes found in a mixed pixel. If desired (Foody, 2000). It is instructive to review how this is done. « (Jensen, 2005)

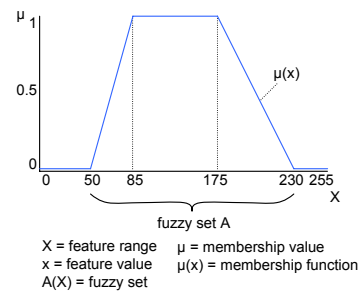
## Fuzzy classification (2)

- **Transition from a crisp to a fuzzy system**

(To decide if a feature value belongs to a fuzzy set)

- The fuzzy set (A):
    - Is a certain subset of values of the whole range of an object feature X (e.g. NIR-band)
    - represents an object feature class (e.g. forest) within one object feature
  - Replace boolean logic ("false" and "true") of the membership value  $\mu$  by a continuous range of  $[0, \dots, 1]$
  - Define membership function  $\mu(x)$ 
    - Assigning to every object feature value  $x$  a membership value  $\mu$
    - If  $\mu > 0$ , then  $x$  belongs to the fuzzy set A
    - Relation between object feature and classification
- ⇒ Choice and parameterisation of the membership function influence the quality of the classification
- ⇒ Introducing expert knowledge

Example:



❶ » Fuzzy logic is a multi-valued logic quantifying uncertain statements. The basic idea is to replace the two boolean logical statements "true" and "false" by the continuous range of  $[0, \dots, 1]$ , where 0 means "false" and 1 means "true" and all values between 0 and 1 represent a transition between true and false. « (Benz et al., 2004)

❷ » Fuzzification describes the transition from a crisp system to a fuzzy system. It defines on an object feature certain fuzzy sets. These fuzzy sets represent object feature classes, e.g. "low", "medium" or "high". [...] These fuzzy object feature classes are defined by so-called membership functions. These functions assign a membership degree between 0 and 1 to each object feature value with respect to the considered object feature class. Depending on the shape of the function, the transition between "full member" and "no member" can be crisp (for a rectangular function) or fuzzy. All feature values, which have a membership value higher than 0 belong to a fuzzy set. « (Benz et al., 2004)

❸ » Fuzzy set theory provides some useful tools for working with imprecise data (Zadeh, 1965; Wang, 1990a, b). Fuzzy set theory is better suited for dealing with real-world problems than traditional logic because most human reasoning is imprecise and is based on the following logic. First, let  $X$  be a universe whose elements are denoted  $x$ . That is,  $X = \{x\}$ . As previously mentioned, membership in a classical set  $A$  of  $X$  is often viewed as a binary characteristic function  $f_A$  from  $X$  to  $\{0 \text{ or } 1\}$  such that  $f_A(x) = 1$  if and only if  $x \in A$ . Conversely, a fuzzy set  $B$  in  $X$  is characterised by the membership function  $f_B$  that associates with each  $x$  a real number from 0 to 1. The closer the value of  $f_B(x)$  is to 1, the more  $x$  belongs to  $B$ . « (Jensen, 2005)

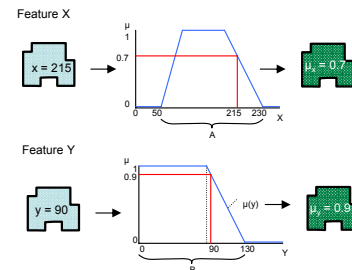
❹ » For successful classification a deliberate choice and parameterization of the membership function is crucial. The function has to model the underlying relation between object features and classification as good as possible. The design is one of the most important steps to introduce expert knowledge into the system. Therefore, the better the knowledge about the real system is modelled by the membership functions, the better the final classification result (Civanlar and Trussel, 1986). It is possible to define more than one fuzzy set on one feature. « (Benz et al., 2004)

## Fuzzy classification (3)

- **Fuzzy rule-base**

- Fuzzy rule "if – then" for assigning an object to a class
  - If feature value  $x$  (of the object) is member of the fuzzy set (e.g. associated with the class forest), the image object is a member of the land-cover forest
- Combination of fuzzy sets to create advanced fuzzy rules
  - Operator "AND" – Minimum operation
  - Operator "OR" – Maximum operation
  - Operator "NOT" – inversion of a fuzzy value: returns  $1 - \text{fuzzy value}$
- Fuzzy rule-base (combination of the fuzzy rules of all classes) delivers a fuzzy classification
  - Every object has a tuple of return values assigned to it with the degrees of membership to each class/degrees of class assignment
  - Since these values are possibilities to belong to a class, they don't have to add up to 1 (unlike probabilities)

Class "forest" defined by  
Feature X AND Feature Y



$$\mu_{\text{forest}} = \mu_x \text{ AND } \mu_y = \min(\mu_x, \mu_y) = \min(0.7, 0.9) = 0.7$$

$\mu_{\text{forest}}$	= 0.7
$\mu_{\text{pasture}}$	= 0.4
$\mu_{\text{water}}$	= 0.05
...	

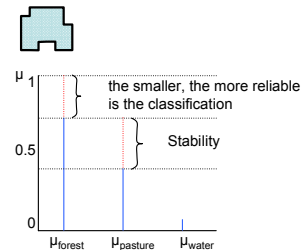
## Fuzzy classification (4)

■ **Comparison of membership degrees**

- Reliability of class assignment  
The higher the degree of the most possible class, the more reliable is the assignment
- Stability of classification  
Stable classification for differences between highest membership value and other values
- Equal membership degrees
  - high values – reliability for both classes: classes cannot be distinguished with the provided classification
  - Low values – unreliable classification (use threshold of a least required membership degree to ensure quality of classification)

■ **Defuzzification**

- Maximum membership degree of fuzzy classification used as crisp classification

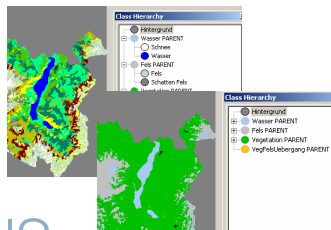


❶ » To produce results like maps for standard land-cover and land-use applications, the fuzzy results have to be translated back to a crisp value. To this end, the maximum membership degree of the fuzzy classification is used as crisp class assignment. This process is a typical approach for defuzzification of fuzzy classification results. If the maximum membership degree of a class is below a threshold, no classification is performed to ensure minimum reliability. As this output removes the rich measures of uncertainty of the fuzzy classification, this step should be only performed if necessary and as late as possible in the whole information extraction process. Further information on fuzzy systems in image analysis and remote sensing can be found in Bezdek and Pal (1992), Maselli et al. (1996), Benz (1999) and Jaeger & Benz (2000). « (Benz et al., 2004)



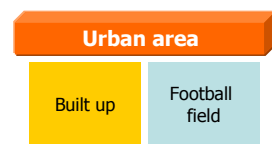
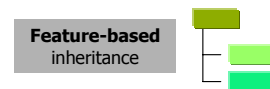
## Class hierarchy

- Classes are embedded in a **heritage system** → i.e. child classes inherit all spectral props from their parents
- Classes can be grouped **semantically**
- Classification can be shown on different **semantic levels**



ZGIS

Centre for Geoinformatics | © 2006 Lang/Albrecht/Biaschke | stefan.lang@sbg.ac.at



❶ » The class hierarchy is the frame of eCognition's "language" for formulating the knowledge base for classifying image objects. It contains all classes of a classification scheme in a hierarchically structured form. The relations defined by the class hierarchy are twofold: the inheritance of class descriptions of child classes on the one hand, and semantic grouping of classes on the other. Each class is represented by a semantic group. The semantic objects can have different relationships to each other. [...]

**Inheritance:** Class descriptions defined in parent classes are passed down to their child classes. A class can inherit descriptions from more than one parent class. Based on the same inherited feature descriptions, the inheritance hierarchy is a hierarchy of similarities.

*Purpose:* reduction of redundancy and complexity in the class descriptions.

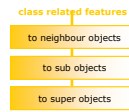
**Groups:** Combination of classes to a class of superior semantic meaning. Beyond that, the groups hierarchy has a direct functional implication: each feature which addresses a class is automatically directed to this class and all its child classes in the groups hierarchy. A class can be part of more than one group. The group register displays the hierarchy of semantic meaning.

*Purpose:* combination of classes previously separated by the classification in a common semantic meaning. [...]

Inheritance is a common technique in object orientated modelling. More general parent objects pass on their properties to child objects. Inheritance is not only used for the sake of simplicity, but also ensures the synchronization of all child objects. Changes in a parent class do not need to be redone in each of the child classes since these inherit the changes automatically. [...] Circular inheritance, however, is impossible (*A* passes on to *B*, *B* passes on to *C*, *C* passes on to *A*). « (Definiens, 2004)

## Class-related features

- **Classifying an object based on the classification of other objects**
  - Class of other objects works as a context feature
    - E.g. a *green* area (class contains parks, meadows,...) is classified as *urban park* because it is embedded in *urban* area
  - Newly classified object may be a context feature itself for other objects
- **Iterative Process**
  - Possibly indeterministic or even unstable
  - Mutual and circular dependencies (should be avoided if possible)



❶ » The use of class-related features is more complex. When an object changes its classification because of the classification of networked objects, the problem arises that the object itself might be a context feature for the evaluation of other objects. Therefore, classification must be an iterative process in cycles in which each object is classified over and over taking into account the changes in the classification of networked objects. The number of cycles can be specified for this purpose.

With context classification a new complexity arises: while classification without context is a deterministic process, context classification can become indeterministic and even unstable due to the possibility of circular dependencies between different classes. Classification becomes an optimizing problem in which convergence to a global best classification must be ensured.

This problem of unstable classification can basically be avoided by the sensible generation of class descriptions. Mutual or circular dependencies between classes should be avoided whenever possible. Class A should not be described by means of class-related features which refer to Class B if Class B itself depends on Class A because of its class description. If this can be ensured classification might need more than one classification cycle, but it is not an optimizing problem. « (Definiens, 2004)

## Strengths of object-based classification

- **Additional information** can be utilized (**shape, texture, relationship to neighbours**)
- 'Overcoming' texture through **segmentation**
  - Especially useful with VHSR-data and radar data
- **Objects**: increased signal/noise ratio
- Decreased number of units to be classified
- Adaptable **image resolution**
- 'Click and classify' – since based on objects
- Classification: **data base query**

❶ » Any step and setting during the entire classification process is documented, and can be assessed and adopted if needed. Although the result is not necessarily more accurate, it can be reproduced and the process is to a high degree comprehensible. The formalized approach of analysis (i.e. the class definitions and composition and the documentation of the workflow and settings in the semi-automated process) technically allows for a transfer of the classification to other scenes (Lang & Langanke, 2004; Benz et al., 2004). « (Lang & Langanke, 2006)

# OBIA – Tutorial

Introduction to object-based image analysis

## Chapter 6 Accuracy assessment

**Z**\_GIS

Centre for Geoinformatics, Salzburg University

## Outline

- **Definitions**
- **Non-site specific accuracy assessment**
- **Site-specific accuracy assessment**
- **Error matrix**
- **Limitations of the error matrix**
- **Specifics of object-based accuracy assessment**

Accuracy assessment is an important part in image analysis, in order to have at hand a n evaluation about the level of potential confusion, the reliability of the class assignments and the overall quality of the results. In this chapter we discuss different approaches to accuracy assessment. Starting with non-site-specific map comparison the chapter moves on to more advanced site-specific approaches with the error matrix as a core element. There are several different accuracy measures that can be retrieved from the error matrix, including overall accuracy, consumer's and producer's accuracy as well as the kappa statistic. The chapter concludes with elaborating on some problems that explicitly apply to object-based accuracy assessment. Especially the spatial geometrical characteristics of the extracted objects need to be evaluated, a challenging task which is still an open field of research.

## Definitions

▪ **Accuracy**

- Degree of correctness of a map or classification (degree to which the derived image classification agrees with reality or conforms to the 'truth') (Foody, 2002)

▪ **Error**

- Discrepancy between the situation depicted on the thematic map and reality (Foody, 2002)

▪ **Significance of accuracy for**

- Usefulness of thematic maps for land management
- The validity of maps for scientific investigations

❶ » The quality of spatial data sets is a very broad issue that may relate to a variety of properties (Worboys, 1998) but frequently, and here, the property of interest is map or classification accuracy. [...] It is important that the quality of thematic maps derived from remotely sensed data be assessed and expressed in a meaningful way. This is important not only in providing a guide to the quality of a map and its fitness for a particular purpose, but also in understanding error and its likely implications, especially if allowed to propagate through analyses linking the map to other data sets (Arbia, Griffith, & Haining, 1998; Janssen & van der Wel, 1994; Veregin, 1994). « (Foody, 2002)

❷ » The scientists who create remote sensing-derived thematic information should recognize the sources of the error, minimize it as much as possible, and inform the user how much confidence he or she should have in the thematic information. Remote sensing-derived thematic maps should normally be subjected to a thorough accuracy assessment before being used in scientific investigations and policy decisions (Stehman and Czaplewski, 1998; Paine and Kiser, 2003). « (Jensen, 2005)

❸ » Accuracy has many practical implications: for example, it affects the legal standing of maps and reports derived from remotely sensed data, the operational usefulness of such data for land management, and their validity as a basis for scientific research. « (Campbell, 2002)

## Definitions (2)

▪ **Accuracy assessment**

- Meaningfully quantify the accuracy of digital land cover classifications; “A classification is not complete until its accuracy is assessed” (Lillesand, Kiefer; 2000)
- Comparison of
  1. Pixels or polygons in a remote sensing-derived classification (**the map to be evaluated**)
  2. Ground reference test information (**reference map**) (Jensen, 2005)

❶ » The accuracy assessment task can be defined as one of comparing two maps, one based upon analysis of remotely sensed data (the map to be evaluated), and another based upon a different source of information (reference map, assumed to be accurate, standard for the comparison). « (Campbell, 2002)

❷ » To correctly perform a classification accuracy (or error) assessment, it is necessary to systematically compare two sources of information: 1. pixels or polygons in a remote sensing-derived classification map, and 2. ground reference test information (which may in fact contain error). « (Jensen, 2005)

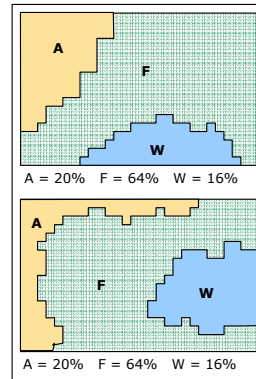
## Non-site-specific assessment

- Comparing area percentages of categories in classified and reference map (inventory error can be assessed)
- Disadvantage:
  - Inaccurate method (e.g. classes have similar proportions but may appear in different locations of the mapped area)

Example of non-site-specific accuracy assessment

	Classified Image	Reference Map	Difference
Forest	42,00%	40,00%	2,00%
Meadow	13,00%	17,00%	-4,00%
Sealed	25,00%	22,00%	3,00%
Water	18,00%	19,00%	-1,00%
Bare Rock	2,00%	2,00%	0,00%
	<b>100,00%</b>	<b>100,00%</b>	

Two maps that are similar according to non-site-specific accuracy assessment



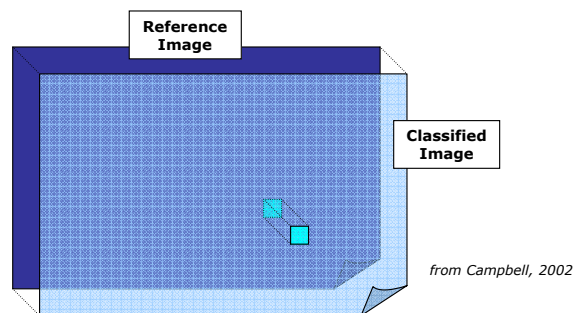
❶ » In this, accuracy assessment was based on comparisons of the areal extent of the classes in the derived thematic map (e.g., km<sup>2</sup> or % cover of the region mapped) relative to their extent in some ground or other reference data set. The non-site-specific nature of this approach is, however, a major limitation as a map could easily display the classes in the correct proportions but in the incorrect locations, greatly limiting the value of the map for some users. Thus, a major problem with this approach to accuracy assessment is that the apparent and quantified accuracy of the map would hide its real quality. « (Foody, 2002)

❷ » This method is inaccurate in itself due to the possibility of compensating errors that don't show up and only overall figures of the two images are compared. Nevertheless, the "inventory error" of the classification map can be estimated. « (compare to Campbell, 2002)



## Site-specific accuracy assessment

- Agreement between categories in classified and reference map **at specific locations**
  - Based on site-by-site comparison (using pixels, clusters of pixels or polygons)
  - Every site can only be occupied by one class (for a clear statement of "error" or "correctly classified")
- Calculation of error matrices



## Error Matrix

- **Identification of overall errors and misclassifications (by category)**
- **$n \times n$  array; where  $n$  is the number of categories**
- **Every cell summarizes the number of sample units assigned to a particular category in the classification relative to the actual category**
  - Diagonal cells (upper left to lower right): correctly classified units
  - Off-diagonal cells: error in the remote sensing classification relative to the ground reference information

Error matrix (schematic representation)

	IMAGE TO BE EVALUATED						TOTAL
	URBAN	CROP	RANGE	WATER	FOREST	BARREN	
REFERENCE IMAGE	URBAN						
	CROP						
	RANGE						
	WATER						
	FOREST						
	BARREN						
TOTAL							
COLUMN MARGINALS							
DIAGONAL ENTRIES GIVE CORRECTLY CLASSIFIED PIXELS ("RANGE CLASSIFIED AS RANGE")							
SUM OF DIAGONAL ENTRIES GIVE TOTAL NUMBER OF CORRECTLY CLASSIFIED PIXELS							

from Campbell, 2002

❶ » The standard form for reporting site-specific error is the error matrix, sometimes referred to as the confusion matrix because it identifies not only overall errors for each category, but also misclassifications (due to confusion between categories) by category. Compilation of an error matrix is required for any serious study of accuracy. It consists of an  $n \times n$  array, where  $n$  represents the number of categories. The left-hand side (y-axis) is labelled with the categories on the reference ("correct") classifications; the upper edge (x-axis) is labelled with the same  $n$  categories; these refer to those on the map to be evaluated. (Note that the meanings of the two axes can be reversed in some applications, as the convention is not universal). « (Campbell, 2002)

❷ » The intersection of the rows and columns summarize the number of sample units (e.g. pixels, clusters of pixels, or polygons) assigned to a particular category (class) relative to the actual category as verified in the field. [...] The diagonal of the matrix summarizes those pixels or polygons that were assigned to the correct class. Every error in the remote sensing classification relative to the ground reference information is summarized in the off-diagonal cells of the matrix. Each error is both an omission from the correct category and a commission to the wrong category. The column and row totals around the margin of the matrix (referred to as marginals) are used to compute errors of inclusion (commission errors) and errors of exclusion (omission errors). The outer row and column totals are used to compute the producer's and user's accuracy. Some recommend that the error matrix contain proportions rather than individual counts (Stehman and Czaplewski, 1998). « (Jensen, 2005)

❸ » Presently, the confusion or error matrix is at the core of accuracy assessment but there is much scope to extend the analysis beyond it (Congalton, 1994; Congalton & Green, 1999). « (Foody, 2002)

## Error Matrix (2)

- Percentage correct (overall accuracy)

- Sum of diagonal entries divided by total observations

- Errors of omission  
⇔ Errors of commission

- Regarding an error from two different viewpoints
- Error of omission:** correct class hasn't been recognised by the classification process (exclusion from category)
- Error of commission:** by mistake the unit has been assigned to the wrong class (error of inclusion)

- Producer's accuracy

- ⇔ Consumer's accuracy

- Accuracies of individual categories
- Producer's accuracy** (a measure of omission error): indicates probability of a reference unit being correctly classified
- Consumer's accuracy** (user's accuracy, a measure of commission error): probability of a classified unit on the map actually representing that category on the ground

Example of an error matrix

		IMAGE TO BE EVALUATED							
		URBAN	CROP	RANGE	WATER	FOREST	BARREN	TOTAL	PA%
REFERENCE IMAGE	URBAN	150	21	0	7	17	30	225	66.7
	CROP	0	730	93	1	115	21	973	75.0
	RANGE	33	121	320	23	54	43	594	53.9
	WATER	3	18	11	83	8	3	126	65.9
	FOREST	23	81	12	4	350	13	483	72.5
TOTAL		248	979	451	134	555	225	1748	
CA%		60.5	74.6	71.0	61.9	63.1	51.1		

Note: Percentage correct = sum of diagonal entries/total observations = 1748/2592 = 67.4%; CA, consumer's accuracy; PA, producer's accuracy

Error of omission → Error of commission ←  
 Producer's accuracy → Consumer's accuracy

from Campbell, 2002; modified

❶ » The overall accuracy of the classification map is determined by dividing the total correct pixels (sum of the major diagonal) by the total number of pixels in the error matrix. Computing the accuracy of individual categories, however, is more complex because the analyst has the choice of the dividing the number of correct pixels in the category by the total number of pixels in the corresponding row or column. Traditionally, the total number of correct pixels in a category is divided by the total number of pixels of that category as derived from the reference data. This statistic indicates the probability of a reference pixel being correctly classified and is a measure of omission error. This statistic is called the producers accuracy because the producer (the analyst) of a classification is interested in how well a certain area can be classified. If the total number of correct pixels in a category is divided by the total number of pixels that were actually classified in that category, the result is a measure of commission error. This measure, called the user's accuracy or reliability, is the probability that a pixel classified on the map actually represents that category on the ground (Story and Congalton, 1986). « (Jensen, 2005)

❷ » The figure shows an example of an error matrix. Each of the 2592 pixels in this scene was assigned to one of six land-cover classes. The resulting classification was then compared, pixel by pixel, to a previously existing land-use map of the same area, and the differences were tabulated, category by category, to form the data for this table. The total number of pixels reported by the matrix (in this instance, 2592) may constitute to the entire image, or may simply be a sample selected from the image. Also the land-use classes here simply form examples; the matrix could be based on other kinds of classes (including forest types, geology, etc.) and could be smaller or larger, depending upon the number of classes examined. « (Campbell, 2002)

## Error Matrix (3)

▪ **Kappa coefficient ( $\hat{\kappa}$ )**

- Need for a more objective accuracy assessment
- Compensation of the effect of chance agreement
- Example:
  - **Random assignment** of pixels into 3 classes
    - ➔ Results in 33% correctly classified pixels (= overall accuracy)
  - 4 classes ➔ 25% correct
  - 5 classes ➔ 20% correct

⇒ **Kappa coefficient**

Measure of difference between **observed agreement** (between a classification and reference information) and **agreement by chance**

❶ » After an initial inspection of the error matrix reveals the overall nature of the errors present, there is often a need for a more objective assessment of the classification. [...] A shortcoming of usual interpretations of the error matrix is that even chance assignments of pixels to classes can result in surprisingly good results, as measured by percentage correct. Hord and Brooner (1976) and others have noted that the use of such measures is highly dependent upon the samples, and therefore upon the sampling strategy used to derive the observations used in the analysis.

K (kappa) is a measure of the difference between the observed agreement between two maps (as reported by the diagonal entries in the error matrix) and the agreement that might be attained solely by chance matching of the two maps. Not all agreement can be attributed to the success of the classification. K attempts to provide a measure of agreement that is adjusted for chance agreement. « (Campbell, 2002)

❷ » The kappa coefficient has many attractive features as an index of classification accuracy. In particular, it makes some compensation for chance agreement and a variance term may be calculated for it enabling the statistical testing of the significance of the difference between two coefficients (Rosenfield & Fitzpatrick-Lins, 1986). This is often important, as frequently, there is a desire to compare different classifications and so matrices. To further aid this comparison, some have called for the normalization of the confusion matrix such that each row and column sums to unity (Congalton, 1991; Smits et al., 1999). « (Foody, 2002)

## Error Matrix (4)

- **Kappa coefficient ( $K_{\text{hat}}$ )**

- $\hat{K} = \frac{\text{Observed agreement} - \text{expected agreement}}{1 - \text{expected agreement}}$

- **If:** Observed agreement  $\uparrow 1$   
Expected agreement  $\downarrow 0$  } **Then:  $\hat{K} \uparrow 1$**

$\kappa = 1 \rightarrow$  perfect agreement between classification and reference data

$\hat{\kappa} = 0 \rightarrow$  agreement is not better than a random classification

^

- Explanation of the formula:

- Observed agreement = overall accuracy

- Expected agreement = sum of the products of the consumer's accuracy (CA) and the producer's accuracy (PA) of each class

$$\text{Kappa coefficient} = \frac{n \sum_{k=1}^q n_{kk} - \sum_{k=1}^q n_{k+} n_{+k}}{n^2 - \sum_{k=1}^q n_{k+} n_{+k}}$$

Quelle: Foody, 2002; S. 188

❶ » The statistic serves as an indicator of the extent to which the percentage correct values of an error matrix are due to “true” agreement versus “chance” agreement. As true agreement (observed) approaches 1 and chance agreement [expected agreement] approaches 0,  $\kappa$  approaches 1. This is the ideal case. In reality,  $\kappa$  usually ranges between 0 and 1. For example, a  $\kappa$  value of 0.67 can be thought of as an indication that an observed classification is 67% better than one resulting from chance. A  $\kappa$  of 0 suggests that a given classification is no better than a random assignment of pixels. In cases where chance agreement is large enough,  $\kappa$  can take on negative values – an indication of very poor classification performance. « (Lillesand and Kiefer, 2000)

❷ »  $\kappa$  values  $>0.80$  (i.e.,  $>80\%$ ) represent strong agreement or accuracy between the classification map and the ground reference information.  $\kappa$  values between 0.40 and 0.80 (i.e., 40 to 80%) represent moderate agreement.  $\kappa$  values  $<0.40$  (i.e.,  $<40\%$ ) represent poor agreement (Landis and Koch, 1977). « (Jensen, 2005)

## Limitations of the error matrix

- **Error matrix not a standardized measure**  
Many different indices, none of them ideal for every single problem
- **Used samples**  
Sampling design (stratified, random...) and sample size can be limiting factors
- **Type of error**  
thematic error vs. error of location
- **Accuracy of the reference data accuracy**
  - Ground "truth" data is a classification itself and may contain error
  - Often remotely sensed data is used as a surrogate for ground data
- **No assessment of the spatial distribution of error**
- **Severity of an error is equalized**
  - Minor errors between relatively similar classes vs. major errors between very different classes
  - Attempts to represent continua by a set of discrete classes
- **Problem with mixed pixels containing more than one class**  
Classes are not mutually exclusive within this pixel (which is a problem for site-specific accuracy assessment)

❶ » There is no single universally acceptable measure of accuracy but instead a variety of indices, each sensitive to different features (Stehman, 1997a).

The design of an accuracy assessment programme has several elements including the definition of an appropriate sample size and sampling design [...]. The sample size, for example, must be selected with care and be sufficient to provide a representative and meaningful basis for accuracy assessment (Hay, 1979). [...] If, for example, a probability-based measure of classification accuracy is to be used (Stehman, 1997a), it is essential that the cases were acquired according to an appropriate sampling design (Hay, 1979; Stehman, 1999b).

A variety of errors are encountered in an image classification. Typically, interest focuses on thematic accuracy. [...] Unfortunately, however, other sources of error contribute to the pattern of misclassification depicted in the confusion matrix [...]. Non-thematic errors can be large and particular concern focuses on errors due to misregistration of the image classification with the ground data (Canters, 1997; Czaplewski, 1992; Muller et al., 1998; Stehman, 1997a; Todd et al., 1980). In fact, the ground data are just another classification which may contain error (Congalton & Green, 1999; Khorram, 1999; Lunetta, Iiames, Knight, Congalton, & Mace, 2001; Zhou, Robson, & Pilesjo, 1998). Problems with ground data accuracy may be particularly severe if a remotely sensed data set is used as the reference data. The erroneous allocations made by a classification are typically not randomly distributed over the region (Congalton, 1988; Steele, Winne, & Redmond, 1998). [...] Unfortunately, however, the confusion matrix and the accuracy metrics derived from it provide no information on the spatial distribution of error.

In classical accuracy assessments all misallocations are equally weighted. Often, however, some errors are more important or damaging than others (Forbes, 1995; Naesset, 1996; Stehman, 1999a). In many instances, the errors observed in a classification are between relatively similar classes and sometimes these may be unimportant while other errors may be highly significant. (Felix & Binney, 1989; Foody, 2000a; Steele et al., 1998; Townsend, 2000).

A further source of error associated with the use of a standard (hard) classifier that allocates each pixel to a single class is the implicit assumption that the image is composed of pure pixels (Foody, 1996; Gong & Howarth, 1990). [...] As many remotely sensed data sets are dominated by **mixed pixels**, the standard accuracy assessment measures such as the kappa coefficient, which assume implicitly that each of the testing samples is pure, are, therefore, often inappropriate for accuracy assessment in remote sensing (Foody, 1996; Karaska et al., 1995). « (Foody, 2002)

## Specifics of object-based accuracy assessment

Need for a different handling of object-based accuracy assessment

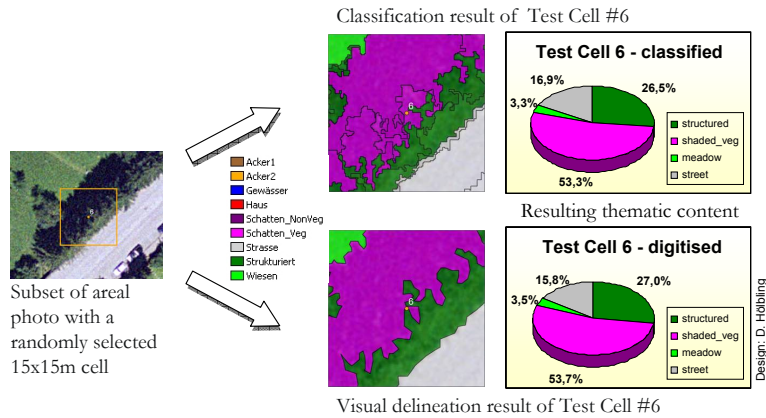
- **Thematic assessment of object-based classification**
    - Site-specific approach (making use of **random points** within objects)
    - Selecting **objects** prior to training process to be used as reference information
    - Selecting **small test areas** where every single object is assessed in terms of its label
  - **Geometry assessment of image objects**
    - Visually checking classified images against manual delineation
    - Quantitative assessment with GIS overlay techniques
  - **Difficulties in object-based accuracy assessment**
    - No 100% geometrical fit between reference objects and objects to be evaluated (due to different ways of segmentation and delineation)
    - When using a fuzzified rule base thematic classes are not mutually exclusive
- ⇒ Accuracy is a matter of geometrical and semantic agreement

❶ » Using an object-based classification approach requires to adapt existing methods of accuracy assessment and develop new techniques that explicitly assess the accuracy of object-specific features.

Within object-based image analysis thematic assessment can be performed by generating random points within objects and checking the labels against a ground truth layer (Lang, Langanke, 2006). Alternatively a set of objects can be selected prior to the training process to be used as reference information. In smaller test areas with a limited number of larger objects, every single object could be assessed in terms of its label (Lang, Langanke, 2006). However, geometrical accuracy is by far harder to evaluate. Classified image objects can be visually checked against manual delineation (Koch et al., 2003), but a quantitative assessment requires GIS overlay techniques (Lang, Schöpfer, Langanke, in press). As outlined in Lang, Schöpfer, Langanke, in press and Lang, Langanke, 2006 we encounter difficulties in performing object-based accuracy assessment, which will satisfy the needs as being discussed by Congalton and Green (1999). Two reasons account for this at least: a) a 100% geometrical fit between reference and evaluation data is usually not given due to different segmentation algorithms and other ways of delineation; b) the thematic classes are not mutually exclusive when using fuzzified rule bases. In other words, the accuracy is also a matter of geometrical and semantic agreement (see also section 3.4.2; p.82). « (Lang, 2005)

## Specifics of object-based accuracy assessment (2)

- Visually checking classified images against manual delineation





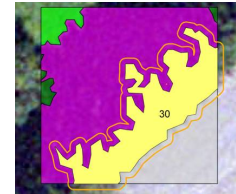
## Specifics of object-based accuracy assessment (3)

- Quantitative assessment of object geometry accuracy

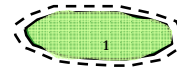
- Defining a tolerance (spatial overlay strictness) to allow minor differences in the object outlines
- Testing of objects
  - Classification object has the same extent as the reference polygon ("stable object")
  - Reference object doesn't match and is split into several objects of the classification
  - Complex geometry with objects not having the same boundaries
- Characterisation of object relationship ("object fate", similar to object-based change analysis)
  - Good objects
  - Expanding objects
  - Invading objects

⇒ Approach needs further development

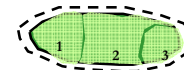
**Tolerance buffer:**  
Selected digitised polygon with 0.5m buffer



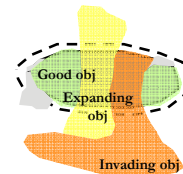
**Case 1:**  
Object stays the same



**Case 2:**  
Object splits into n sub-objects



**Case 3:**  
Complex geometry with expanding and intruding objects



❶ » LIST [Landscape Interpretation Support Tool] offers GIS routines for combining polygon themes from different sources and calculating object characteristics and object fate. [...] However, for several reasons we are mostly dealing with spatially inconsistent features. This requires a more flexible concept of spatial coincidence. [...] LIST offers a way to prove spatial coincidence of two data sets. Using a threshold an accepted overlap can be determined. In order to consider uncertainties in the geometrical properties of the boundaries of related objects a spatial overlay sensitivity (SOS) factor has been introduced. [...] Image objects are subject to change. At the same time they may change topological relationships to neighbouring objects over time. In principle changes may occur with regard to type, intensity and shape of the object. However, when an object is shrinking or expanding, then instantly the surrounding is affected. Raza and Kainz, 2001, for example, include subdivision and amalgamation of objects in their list of spatiotemporal characteristics of parcels. Here it is assumed that one object at time  $t_0$  is split into  $n$  parts, or an object in  $t_1$  is a product of  $m$  parts. [...] Straub [2004] presents eight topological relations being reducible to the principal relationships disjoint, equal, overlap and containment. [...] All fate categories as depicted are derived by combining basic spatial relationship types. In general with all  $t_1$  objects that intersect (operation *inters*), the respective  $t_0$  objects are considered to be related to it. The number of objects that are completely within (operation *compwi*) the buffered shape of the  $t_0$  object is considered to equal the number of 'good' objects (*ngood*). If *compwi* equals 1 and the area of the  $t_1$  object is equal to the area of the  $t_0$  object (considering the SOS factor), the  $t_0$  object has remained the same. If this is not the case or if *compwi* is greater than 1, the object  $t_0$  has been split over time. The product of the splitting is further distinguished by additional spatial relationship types, such as 'have their centre in' (*centin*) and 'intersect' (*inters*). The operation *centin* reveals the number of  $t_1$  objects originating from the  $t_0$  object. Consequently, the difference of *centin* and *compwi* is the number of objects that are originating from  $t_0$ , but expanding. And finally the difference of *inters* and *centin* reveals the number of invading objects. « (Lang et al. in press)

## Outlook

- A next version of this tutorial is under preparation. We will include the following topics:
  - **Scale-specific** vs. **scale-adaptive** segmentation
  - **Class modelling** ⇒ segmentation and classification as a cyclic process
  - Basics of **CNL** (cognition network language)
  - More on **object fate** ⇒ object-based change detection and object-based accuracy assessment
  - Application **showcase** (forestry, urban studies, habitats and wetlands, etc.)

Version 2.0 of the OBIA Tutorial is under preparation and will be issued in autumn 2007. Please watch out for the announcement under [www.uni-salzburg.at/zgis/research](http://www.uni-salzburg.at/zgis/research).

## References

▪ **A – J**

- Note that – with a few exceptions – only literature is listed from which text was taken
- Figures in brackets ([ ]) indicate the respective chapters

Baatz M & A Schäpe (2000) Multiresolution Segmentation: an optimization approach for high quality multi-scale image analysis. In: J Strobl & T Blaschke (eds.) *Angewandte geographische Informationsverarbeitung*, vol. XII., pp 12 -23 – Heidelberg **[4]**

Benz U, P Hofmann, G Willhauck, I Lingenfelder & M Heynen (2004) Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. In: *International Journal of Photogrammetry and Remote Sensing* (58), pp 239– 258 **[1, 3, 4, 5]**

Blaschke T (2003) Object-based contextual image classification built on image segmentation – IEEE proceedings, Washington DC, CD-ROM **[1, 4]**

Blaschke T, S Lang, M Möller (2005) Object-Based Analysis of Remote Sensing Data for Landscape Monitoring: Recent Developments. In: *Anais XII Simpósio Brasileiro de Sensoriamento Remoto*, Goiânia, Brasil, 16-21 abril 2005, INPE, pp 2879-2885 **[1]**

Blaschke T, C. Burnett, & A. Pekkarinen (2004): New contextual approaches using image segmentation for object-based classification. In: F de Meer & S de Jong (eds.) *Remote Sensing Image Analysis: Including the spatial domain*. Kluwer Academic Publishers, Dordrecht, pp 211-236 **[4]**

Burnett C & T Blaschke (2003) A multi-scale segmentation / object relationship modelling methodology for landscape analysis. In: *Ecological Modelling* 168(3), pp 233-249 **[1, 2, 3, 4, 5]**

Blaschke T & J Strobl (2001) What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. *GIS – Zeitschrift für Geoinformationssysteme* 14(6), pp 12-17 **[1, 2]**

Campbell JB (2002) *Introduction to remote sensing*, 3rd ed. – New York, London **[1, 6]**

Definiens Imaging (2004) User Guide 4. [www.definiens\\_imaging.com](http://www.definiens_imaging.com) (accessed Feb 01, 2006). **[2, 4, 5]**

Foody GM (2002) Status of land cover classification accuracy assessment, *Remote Sensing of Environment* 80, 185-201 **[6]**

Hay G, T Blaschke, D Marceau & A Bouchard (2003) A comparison of three image-object methods for the multiscale analysis of landscape structure. In: *International Journal of Photogrammetry and Remote Sensing* (57), pp 327-345 **[1, 2, 4]**

Jensen JR (2005) *Introductory Digital Image Processing: A Remote Sensing Perspective*, 3rd edition, Prentice Hall – Upper Saddle River, New Jersey, pp. 393 – 401 **[4, 5, 6]**

## References

- **L – Z**
  - Note that – with a few exceptions – only literature is listed from which text was taken
  - Figures in brackets ([ ]) indicate the respective chapters

Lang S (2005) Image Objects and Landscape Objects – Interpretation, Hierarchical Representation and Significance – unpublished PhD thesis, Salzburg **[1, 3, 5, 6]**

Lang S (2002) Zur Anwendung des Holarchiekonzeptes bei der Generierung regionalisierter Segmentierungsebenen in höchstauflösenden Bilddaten. In: T Blaschke (ed.) Fernerkundung und GIS: Neue Sensoren – innovative Methoden. Heidelberg, pp 24-33 **[2]**

Lang S & T Blaschke (2003) Hierarchical object representation. Comparative multi-scale mapping of anthropogenic and natural features. In: International Archives of Photogrammetry and Remote Sensing, vol. 24 (3/W8), pp 181-186 **[2]**

Lang S, E Schöpfer & T Langanke (in press) Combined object-based classification and manual interpretation - Synergies for a quantitative assessment of parcels and biotopes. Geocarto International **[6]**

Lang S & T Langanke (2006) Object-based mapping and object-relationship modelling for land use classes and habitats. PFG – Photogrammetrie, Fernerkundung, Geoinformatik, vol. 1, pp 1-18 **[3, 4, 5]**

Lang S, C Burnett & T Blaschke (2004) Multi-scale object-based image analysis – a key to the hierarchical organisation of landscapes. In: Ekologia , Supplement 23, pp 1-9 **[2, 5]**

Lillesand TM & RW Kiefer (2000) Remote Sensing and Image Interpretation, 4th ed., Wiley **[6]**

Marceau D & G Hay (1999) Contributions of remote sensing to the scale issues. In: Canadian Journal of Remote Sensing, Vol. 25, pp 357-366 **[2]**

Pinz A (1994) Bildverstehen. Springer, Wien **[3]**

Wu J (1999) Hierarchy and scaling: extrapolating information along a scaling ladder. In: Canadian Journal of Remote Sensing, Vol. 25, pp 367-380 **[2]**