Chapter 12
Image Segmentation Methods for Object-based Analysis and Classification

Thomas Blaschke, Charles Burnett & Anssi Pekkarinen

12.1 Introduction

The continuously improving spatial resolution of remote sensing (RS) sensors sets new demand for applications utilising this information. The need for the more efficient extraction of information from high resolution RS imagery and the seamless integration of this information into Geographic Information System (GIS) databases is driving geo-information theory and methodology into new territory. As the dimension of the ground instantaneous field of view (GIFOV), or pixel (picture element) size, decreases many more fine landscape features can be readily delineated, at least visually. The challenge has been to produce proven man-machine methods that externalize and improve on human interpretation skills. Some of the most promising results in this research programme have come from the adoption of image segmentation algorithms and the development of so-called object-based classification methodologies. In this chapter we describe different approaches to image segmentation and explore how segmentation and object-based methods improve on traditional pixel-based image analysis/classification methods.

According to Schowengerdt (1997) the traditional image processing/image classification methodology is referred to as an *image-centred approach*. Here, the primary goal is to produce a map describing the spatial relationships between phenomena of interest. A second type, the *data-centred approach,* is pursued when the user is primarily interested in estimating parameters for individual phenomena based on the data values. Due to recent developments in image processing the two approaches appear to be converging: from image and data centred views to an *information-centred* approach. For instance, for change detection and environmental monitoring tasks we must not only extract information from the spectral and temporal data dimensions. We must also integrate these estimates into a spatial framework and make *a priori* and *a posteriori* utilization of GIS databases. A decision support system must encapsulate manager knowledge, context/ecological knowledge and planning knowledge. Technically, this necessitates a closer integration of remote sensing and GIS methods. Ontologically, it demands a new methodology that can provide a flexible, demand-driven generation of information and, consequently, hierarchically structured semantic rules describing the relationships between the different levels of spatial entities.

Several of the aspects of geo-information involved cannot be obtained by pixel information as such but can only be achieved with an exploitation of neighbourhood information and context of the objects of interest. The relationship between ground objects and *image objects*

S.M. de Jong and F.D. van der Meer (eds.),
Remote Sensing Image Analysis: Including the Spatial Domain, 211–236.

examined in remote sensing representations must be made explicit by means of spatial analysis and the construction of a semantic network. In the following sections we contrast conventional image classification/analysis methods to the new segmentation-based methods; review some current progress in image segmentation and RS/GIS integration, which adds topological and hierarchical rules to build databases of context information; and present two examples to demonstrate the utility of the segmentation-based methodology.

12.1.1 The RS/GIS image analysis continuum

Thematic mapping from digital remotely sensed images is conventionally performed by pixelwise statistical classification (Schneider & Steinwender, 1999, Blaschke & Strobl, 2001). Pixelwise analysis utilizes three of the attributes of a pixel; *Position*, *Size* and *Value* (de Kok et al., 2000). The size attribute can be usually considered as constant, except for imagery acquired in high relief areas (Burnett et al., 1999). The main drawback of pixelwise classification is that it largely neglects shape and context aspects of the image information, which are among the main clues for a human interpreter. A limited form of contextual information can be stored in the Value parameter. For example, texture or other relevant information can be analysed from the immediate neighbourhood of the pixel and result can be assigned to the central pixel. Examples of this are moving window filters which can be implemented with help of convolution masks (Jain et al., 1995). In object oriented analysis shape and context are

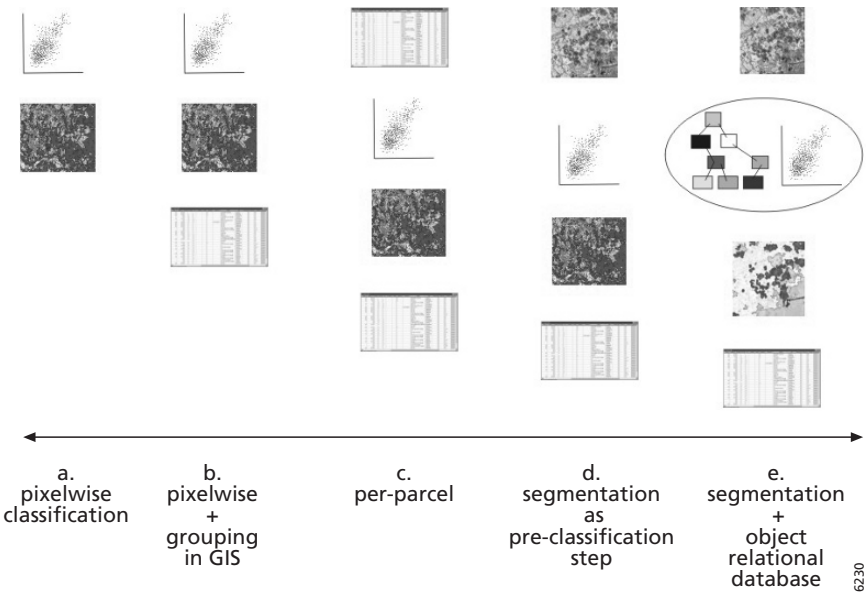

Figure 12.1 – A continuum of classification methods: (a) pixelwise classification, utilizing spectral information; (b) pixelwise with grouping of results into groups in a GIS; (c) per-parcel classification, where GIS data is used to partition the scene before a pixelwise classification; (d) single-scale segmentation prior to pixelwise classification, and; multi-scale segmentation as part of object relationship database building, with classification done by querying spectral and spatial object parameters. Please consult the enclosed CDROM for a full colour version.

clumped into a fourth attribute, that defining *Fellowship*; 'to which pixel population does this pixel belong' (de Kok et al., 2000). It has been suggested that classifying remotely sensed images in pixelwise fashion (using only the first three pixel attributes) is a special case of the super-set of object-based classification methodologies which utilise all four (Schneider & Steinwender, 1999; de Kok et al., 2000). This is not illogical and we propose the following continuum of classification methods incorporating more or less fellowship information (figure 12.1).

12.1.2 New sensors, new image/ground object relationships

Until recently, the 20m spatial resolution of SPOT was regarded as 'high spatial resolution'. Since the launch of IKONOS 2 in 1999 a new generation of very high spatial resolution (VHR) satellites was born, followed by Quick Bird late 2001. The widely used Landsat and Spot sensors are now called 'medium-resolution'. Especially the new satellite sensor generation meet the strong market demands from end-users, who are interested in image resolution that will help them observe and monitor their specific objects of interest. The increasing variety of satellites and sensors and spatial resolutions lead to a broader spectrum of applications but not automatically to better results. The enormous amounts of data created a strong need for new methods to exploit these data efficiently. In addition, the complexity of the relationship of pixel and object make it necessary to develop additional methods of classification.

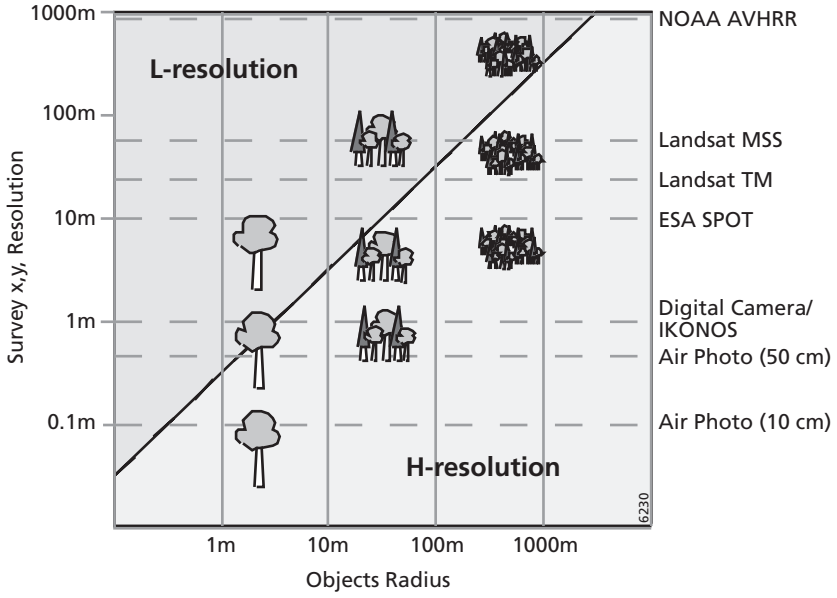

Figure 12.2 – Relationship between the remotely sensed image (survey) resolution (GIFOV) and real-world object diameter. Close to the diagonal, the diameter of the objects and the GIFOV will be nearly equal and image objects (groups of H-res pixels representing a single real-world object) may reduce to 1-pixel image-objects. For situations where the sensor GIFOV is larger than the object of interest, we are in L-res or 'mixed-pixel' territory. (Image modified from Burnett 2002.)

The number of pixels containing more than one land cover type is a function of the complexity of the scene, the spatial resolution of the sensor and the classification scheme. Therefore spatial resolution is among other factors important to the definition of classes. The relationship is based on the simple fact that higher resolution images contain a smaller percentage of pixels regarded as 'boundary pixels', falling into two or more different land cover classes (Cao & Lam, 1997, Mather, 1999). The term 'mixed pixel' does not imply that the scale of observation is inappropriate or does not match the scale of variation of the phenomenon under investigation, although a large proportion of mixed pixels are often associated with a too detailed classification system and/or an inappropriate scale for the respective application. For example, imagine a class 'buildings' consisting of sparsely distributed single houses of 10 by 10m. A Landsat 30m image will never result in 'pure pixels' but an IKONOS image would, although of course there will be many mixed pixels remaining but their percentage decreases. In natural environments we have another problem that there will always be small gaps e.g. in forest's canopy which influence the result, or their might be classes which are specifically mosaics of single entities of specific spatial arrangements of classes. It may be necessary to stress the fact that even though the relative number of mixed pixels decreases their absolute number usually increases. The mixed pixel problem can be examined with the aid of figure 12.2, which describes the relationship between pixel size and object-of-interest size. The nomenclature L-res and H-res refers to the relationship between the pixel GIFOV and the median radius of the real-world objects being imaged (figure 12.1): if the GIFOV is significantly smaller than this diameter, than we are dealing with an H-res mode – at least for that object type! The corollary is the L-res scenario (Woodcock & Strahler, 1987).

As Mather (1999) points out, where this is clearly not the case then a small proportion of pixels will be located in a 'boundary region', e.g. between adjacent fields, and may, therefore, described as 'mixed'. But Cihlar (2000) emphasises, that even in high-resolution and radiometrically well corrected data, some variation will remain which can be regarded as noise or can lead to mixed pixels falling partially into two or more classes. Sensor systems have a specific GIFOV – simply put: a certain spatial resolution. Several targets/classes of interest may be found within one unit of GIFOV. Usually, only a single category is assigned to each pixel. But in fact one pixel could represent more than one target/class.

12.1.3 From pixels to image-objects

It will be useful to clarify the terms pixel and image-object. A pixel is normally the smallest unit of analysis of RS imagery. A pixel's dimensions are determined by the sensor and scene geometry models, giving the GIFOV. The phrase 'normally the smallest' is applied because there have been attempts to decompose the spectral signature of pixels and thus do sub-pixel analysis (Aplin & Atkinson, 2001; Asner & Heidebrecht, 2002; Lucas et al., 2002; Verhoeye & de Wulf, 2002). *Image-object*s are defined by Hay et al. (2001) as... basic entities, located within an image that are perceptually generated from H-res pixel groups, where each pixel group is composed of similar digital values, and possesses an intrinsic size, shape, and geographic relationship with the real-world scene component it models.

Schneider & Steinwender (1999) suggest a simpler definition for image-objects, 'groups of pixels with a meaning in the real world'. As image GIFOV decreases we are faced with new challenges: we can resolve more and more types of real world objects. The internal or 'within-object heterogeneity' increases and the spectral separability between image objects drops. The

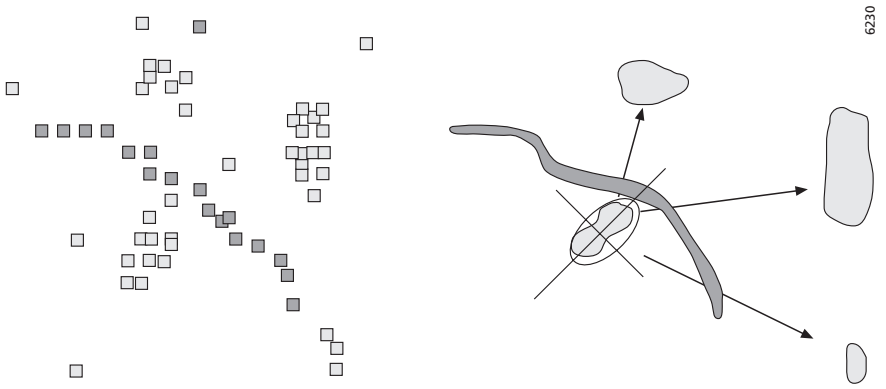

Figure 12.3 – Pixels with same spectral entities vs. objects with averaged spectral properties and evolving spatial properties.

benefit of going to a segmentation approach is that you permit the incorporation of more expert knowledge – more specifically, you can incorporate more context information. This may seem at first to be counter-productive: are we not searching for methods which limit the amount of a priori information necessary to do the classification? The answer is that 'simple' methods to monitor earth processes automatically is a myth propagated in the 1980s. The fact is that our world is complex and our 'windows to the world' (RS images) are limited (Burnett & Blaschke, in press). Segmentation begets image objects, permitting the incorporation of spatial information as mutual relationships to these objects.

The incorporation of segment-level spatial information to the analysis of RS imagery introduces new dimensions to the actual analysis of the data. Instead of relying on only the Value, Location and Size attribute of single pixel, we can incorporate Fellowship (topology-related) attributes, for example image object size, shape or number of sub-object. This allows us to better utilise sophisticated GIS functions in analysing the data, e.g. to describe the spatial complexity of the objects, their spatial and spectral embeddedness in relation to neighbouring objects etc. We speak of objects if we can attach a meaning or a function to the raw information. Generally, the object is regarded to be an aggregation of the geometric, thematic and topologic properties. The topologic relations between the cells the object consists of can be examined once the user has defined his or her objectives, classification scheme and scale of analysis.

12.2 Image segmentation review

12.2.1 What is image segmentation?

One possible strategy to model the spatial relationships and dependencies present in RS imagery is image segmentation. Image segmentation is the partitioning of an array of measurements on the basis of homogeneity. To be more exact, 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 'homogenous' 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.

Although image segmentation techniques are well known in some areas of machine vision (see Narendra & Goldberg, 1980, Fu & Mui, 1981, Haralick & Shapiro, 1985, Cross et al., 1988), they were 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 of 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.

Before delving more deeply into the different algorithms that have been developed for image segmentation, and more specifically, for remotely sensed EO image segmentation, we would like to demonstrate the complexity of the process. For a 1 by 1 array, we can only partition the array only into 1 segment (figure 12.4). For a 2 by 2 array, we get a theoretical maximum of 16 combinations of segments or partitions. Note that partitions can be of any size greater than or equal to 1, and are of course, limited to the size of the array. Partitions must not overlap (they are disjoint) and the whole array must be partitioned. This means that for a typical 1:20,000 scale aerial photograph scanned to produce a 60 cm GIFOV, and which thus has 3000 rows by 3000 columns the theoretical maximum number of partitions is as large as 2 to the power 17994000! In real life, however, we would never be interested in exploring all these different

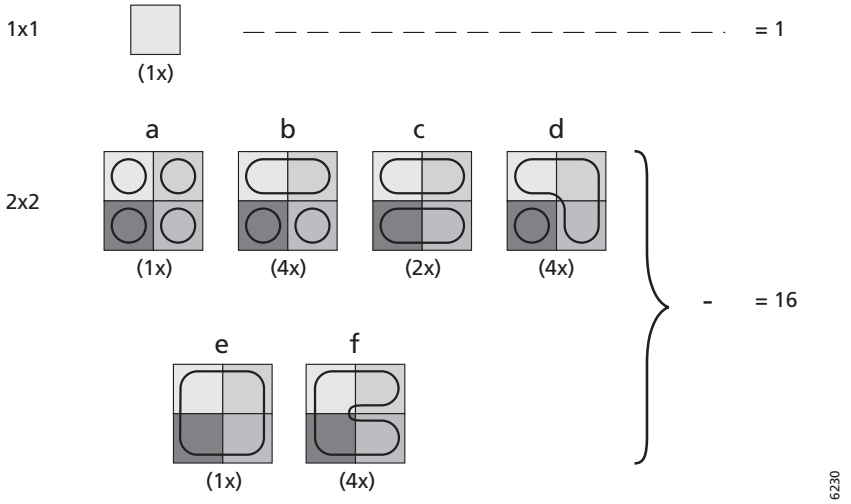

Figure 12.4 – Theoretical partitioning of 2D arrays. The grey boxes are meant to be the array, with a grey level value (GLV) shown by the 'greyness'. The coloured lines (circles, ovals, etc.) are vectors that 'partition' or 'discretize' these 2D arrays. Please consult the enclosed CDROM for a full colour version.

discretization instances, rather we want to intelligently chose an optimized partitioning. In fact, we probably want to chose 3 or more useful partitions of a single scene – but we will reserve a discussion of the use of multiple partitions (or multi-scale segmentation) until later. Still, it should be noted that we still have to use our expert knowledge to calibrate the segmentation algorithm so that the image objects that the segments delineate and the real world objects-of-interest match as closely as possible. However, let's first take a look at some of the tools that have been developed for image segmentation.

12.2.2 Types of segmentation

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

Pixel based methods include image thresholding and segmentation in the feature space. These methods do not necessarily produce a result which fulfils the requirement and definition of segmentation, and therefore the resulting output needs to be clumped. In other words, each spatially continuous unit (often referred as connected component in machine vision literature) needs to be assigned a unique label.

In edge based segmentation methods, the aim is to find edges between the regions and determine the segments as regions within these edges. From this point of view, edges are regarded as boundaries between image objects and they are located where changes in values occur. There are various ways to delineate boundaries but in general the first step of any edge-based segmentation methods is edge detection which consists of three steps (Jain et al., 1995): filtering, enhancement and detection. Filtering step is usually necessary in decreasing the noise present in the imagery. The enhancement aims to the revealing of the local changes in intensities. One possibility to implement the enhancement step is to us high-pass filtering. Finally, the actual edges are detected from the enhanced data using, for example, thresholding technique. Finally, the detected edge points have to be linked to form the region boundaries and the regions have to be labelled.

Region-based segmentation algorithms can be divided into region growing, merging and splitting techniques and their combinations. Many region growing algorithms aggregate pixels starting with a set of seed points. The neighbouring pixels are then joined to these initial 'regions' and the process in continued until a certain threshold is reached. This threshold is normally a homogeneity criterion or a combination of size and homogeneity. A region grows until no more pixels can be attributed to any of the segments and new seeds are placed and the process is repeated. This continues until the whole image is segmented. 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).

In region merging and splitting techniques the image is divided into subregions 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 of objects determined with help of any segmentation technique. In region splitting methods the input usually consists of large segments and these segments are divided to smaller units if the segments are not homogeneous tough. In an extreme case region splitting starts with the

original image and proceeds by slitting 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 (figure 12.5). In both, region merging and splitting techniques, the process is based on a high number of pairwise merges or splits. The segmentation process can be seen as a crystallisation process with a big number of crystallization 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.
Sometimes seen separately, is the group of 'split-and-merge' algorithms (Cross et al., 1988). They start by subdividing the image into squares of a fixed size, usually corresponding to the resolution of a certain level in a quad tree. These leaves are then tested for homogeneity and heterogeneous leaves are subdivided into four levels while homogeneous leaves may be combined with three neighbours into one leaf on a higher level etc.

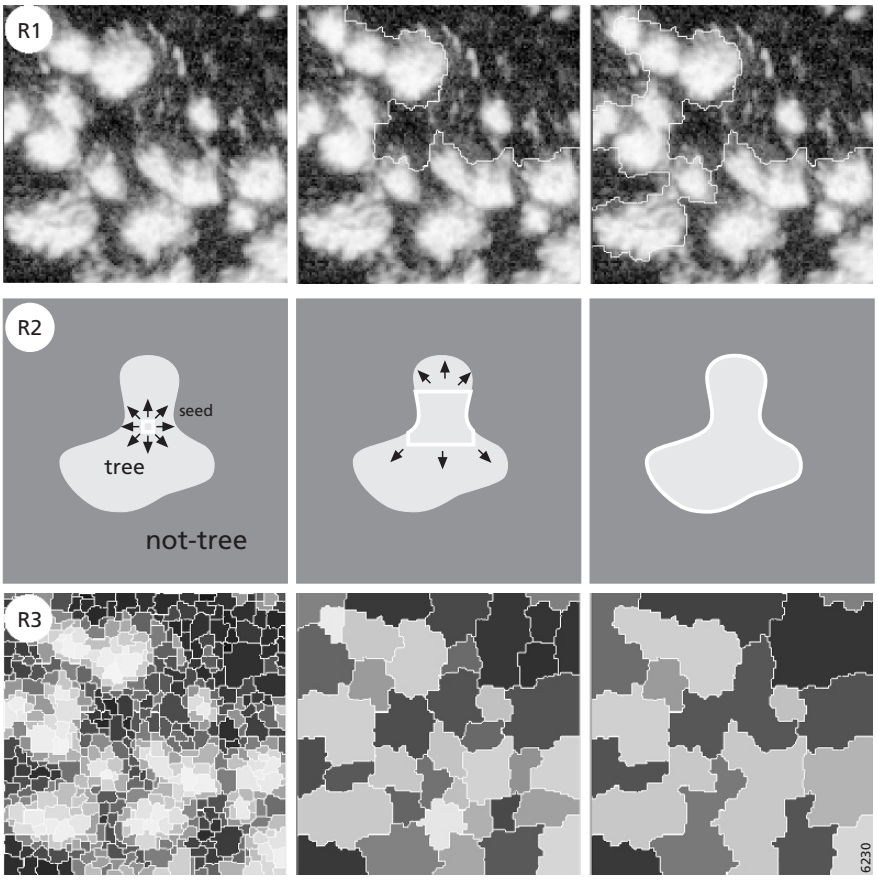

Figure 12.5 – Examples of region splitting and merging techniques.

12.2.3 Segmentation of Remotely Sensed data: state of the art

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 & Shapiro, 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 (Baatz & Schäpe, 2000, Blaschke & Strobl, 2001). Most approaches create segments which are in any sense regarded as being homogeneous by utilising geostatistical analysis (Hofmann & Böhner, 1999) and similarity approaches such as: unsupervised texture recognition by extracting local histograms and a Gabor wavelet scale-space representation with frequency (Hofmann et al., 1998); image segmentation by Markov random fields and simulated annealing; or Markov Random Fields (MRF) using a Maximum *a posteriori* (MAP) probability approach. The MRF method generally classifies a particular image into a number of regions or classes. The image is modelled as a MRF and the MAP probability is used to classify it. The problem is posed as an objective function optimisation, which in this case is the *a posteriori* probability of the classified image given the raw data which constitutes the likelihood term, and the prior probability term, which due to the MRF assumption is given by the Gibb's distribution. MRF was already exploited for an unsupervised classification by Manjunath & Chellappa (1991).

Hoffman & Böhner (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 these 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 – represent spatial unsteadiness in the digital numbers. For the image segmentation, the vectorised minima of the representativeness delimit areas consisting of pixels with similar spectral 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.

Dubuisson-Jolly & Gupta (2000) developed an algorithm for combining colour and texture information for the segmentation of colour images. The algorithm uses maximum likelihood classification combined with a certainty based fusion criterion. One of the most promising approaches is developed by Hofmann et al. (1998) and based on a Gabor wavelet scale-space representation with frequency-tuned filters as a natural image representation. Locally extracted histograms provide a good representation of the local feature distribution, which captures substantially more information than the more commonly used mean feature values. Homogeneity between pairs of texture patches or similarity between textured images in general can be measured by a non-parametric statistical test applied to the empirical feature distribution functions of locally sampled Gabor coefficients. Due to the nature of the pairwise proximity data, this algorithm systematically derives a family of pairwise clustering objective functions based on sparse data to formalize the segmentation problem. The objective functions are designed to possess important invariance properties. A clustering algorithm has been developed, that is directly applicable to the locally extracted histograms. It applies an optimisation technique known as multi-scale annealing to derive heuristic algorithms to

efficiently minimize the clustering objective functions. This algorithm has not been tested comprehensively and never been implemented within an operational/commercial software environment.

From most research following a segmentation approach it is argued that image segmentation is intuitively appealing. Human vision generally tends to divide images into homogeneous areas first, and characterises those areas more carefully later (Gorte, 1998). Following this hypothesis, it can be argued that by successfully dividing an image into meaningful objects of the land surface, more intuitive features will result. The problem is to define the term 'meaningful objects'. Nature hardly consists of hard boundaries but it is also not a true continuum. There are clear, but sometimes soft, transitions in land cover. These transitions are also subject to specific definitions and subsequently dependant on scale. Therefore, segments in an image will never represent meaningful objects at all scales, for any phenomena.

In the modelling stage characteristic features are extracted from the textured input image which include spatial frequencies (Jain & Farrokhnia, 1991, Hoffman et al., 1998), Markov Random Field models (Mao & Jain, 1992, Panjwani & Healey, 1995), co-occurrence matrices (Haralick et al., 1973), wavelet coefficients (Salari & Zing, 1995), wave packets (Laine & Fan, 1996) and fractal indices (Chaudhuri & Sarkar, 1995). In the optimisation stage features are grouped into homogeneous segments by minimising an appropriate quality measure. This is most often achieved by a few types of clustering cost functions (Jain & Farrokhnia, 1991, Mao & Jain, 1992, Hoffman et al., 1998). A further possibility is the watershed transformation. Bendjebbour et al. (2001) defined a general evidential Markovian model and demonstrated that it is usable in practice [to do what?]. Different simulation results show the interest of evidential Markovian field model-based segmentation algorithms. Furthermore, they described a variant of generalized mixture estimation, making possible the unsupervised evidential fusion in a Markovian context. It has been applied to the unsupervised segmentation of real radar and SPOT images showing the relevance of these models and corresponding segmentation methods. These approaches are just examples of what's available in scientific computing but most of these approaches are far from being operational.

12.2.4 Operational image segmentation frameworks

Per-field classification approaches have shown good results in studies (e.g. Lobo et al., 1996). Their results are often easier to interpret than those of a per-pixel classification. The results of the latter often appear speckled even if post-classification smoothing is applied. 'Field' or 'parcel' refers to homogenous patches of land (agricultural fields, gardens, urban structures or roads) which already exist and are superimposed on the image. Some studies (e.g. Janssen, 1993, Aplin et al., 1999) indicate that the methodology is positively contributing to the classification of remote sensing imagery of high to medium geometric resolution. This classification technique is especially applicable for agricultural fields (Janssen, 1993, Abkar & Mulder, 1998). Distinct boundaries between adjacent agricultural fields help to improve the classification due to the fact that boundaries in an agricultural landscape are relatively stable while the cropping pattern (also within the lots) changes often.

An alternative to approaches based on the idea of finding homogeneous areas in an image is the multi-fractal image analysis. The only operational approach widely available is implemented in a software called FracLab. FracLab is a Matlab toolbox for the multi-

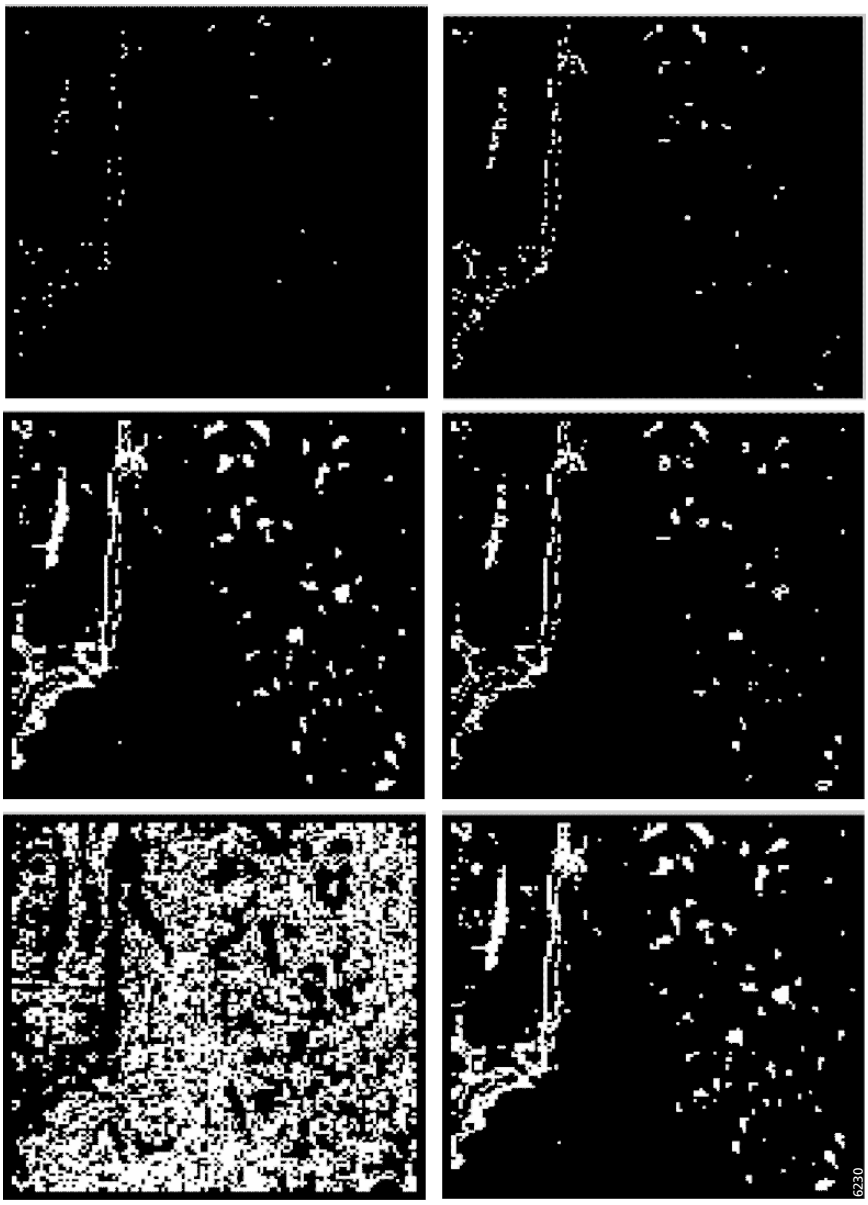

Figure 12.6 – Images a – f represent MAX Hoelder exponent of multifractal segmentation images of different parameters set-up for Band 4 (segment I).

fractal analysis of sets. It is produced by the *Groupe Fractales* of INRIA, Paris, France. The descriptions here are based closely on the works of Jacques Vehel. In the Multi-fractal approach the image is modelled not by a function but by a measure μ. This allows the role of

resolution in image interpretation to be emphasised. The basic assumptions of the approach are the:

- Relevant information for image analysis can be extracted from the Hölder regularity of the measure μ
- Analysis is at three levels
- The pointwise Hölder regularity of μ at each point
- Variation of the Hölder regularity of μ in local neighbourhoods
- The global distribution of regularity of a whole scene
- The analysis is independent of translation and scale

Compared to other approaches to image segmentation or filtering, information about whole images is used to analyse each point instead of local comparison. The main difference between classic and multi-fractal methods is in the way they deal with regularity. The former try to obtain smoother versions of images, possibly at different scales, but multi-fractal analysis tries to obtain information directly from singular measures. Edges, for instance, are not considered as points where large variations of a signal still exist after smoothing, but as regions whose regularity is different from the background regularity in the raw data. This approach has merit for analysis of complex images.

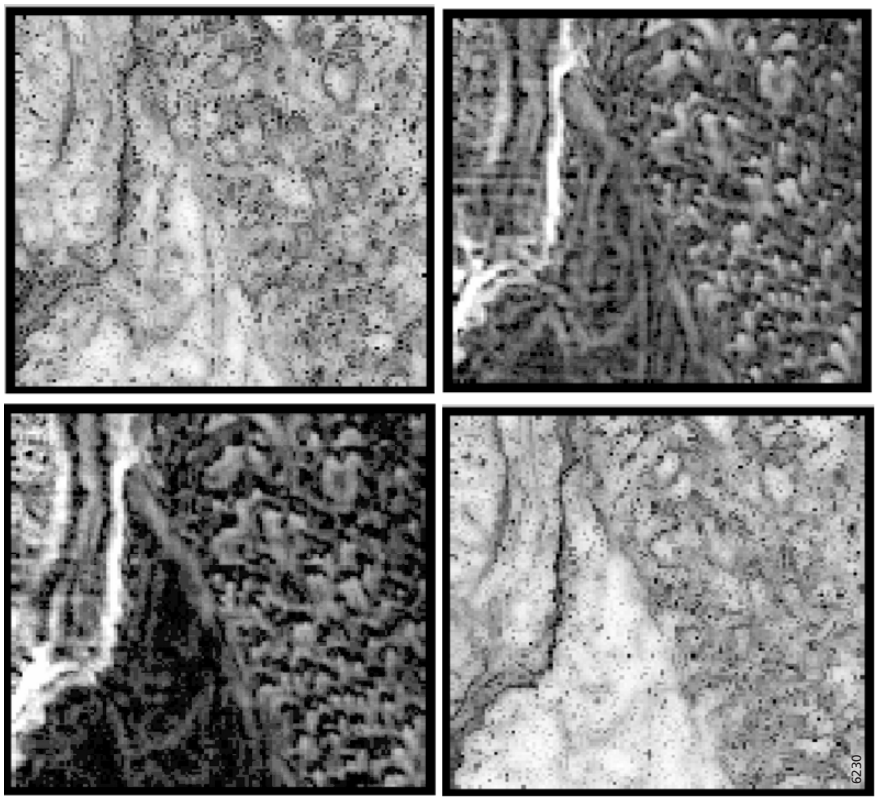

Figure 12.7 – Images a–d represent different Hoelder exponent images for Band 4 and 7 of Landsat 5.

Where grey levels of images are used in classic analysis, Hölder regularity is used for multi-fractal analysis. This is justified in situations where, as is common in complex images, important information is contained in the singular structure of image elements. Multi-fractal analysis will, for instance, find boundaries between texture regions as opposed to boundaries within textures, which is normal in classic methods.

The fractal net evolution approach (FNEA) was documented by Baatz & Schäpe (2000) and successful applications already exist (de Kok et al., 1999, Blaschke et al., 2000, Blaschke et al., 2001, Schiewe & Tufte, 2002, Neubert & Meinel, 2002, Hay et al., 2003). The FNEA is a region merging technique and starts with 1-pixel image objects. Image objects are pairwise merged one by one to form bigger objects. In this conceptualisation the procedure becomes a special instance of an assignment problem, known as pairwise data clustering. In contrast to global criteria, such as threshold procedures, decisions are based on local criteria, especially on the relations of adjacent regions concerning a given homogeneity criterion. In such an optimisation procedure each decision concerning a merge is based on all previous decisions or merges at least in its local vicinity. Therefore such a procedure includes to a certain degree historicity which can cause problems for reproducibility. The solution for this problem is the optimisation procedures and the homogeneity criteria which are maximizing the constraints in the optimisation process (Baatz & Schäpe, 2000).

12.3 Extending segmentation to an object-based analysis and classification

For many applications, a segmentation procedure is only a first, mechanistic step. Exceptions are approaches where the segmentation is performed on classification data and their respective class uncertainty values (e.g. Abkar & Mulder 1998, Klein et al., 1998). But, generally, the research interest is much deeper and image processing goes much further. Most research projects aim to map 1 to 1 the delineated (segmented) image objects to real-world entities within the geographic extent of the scene being assessed. The term image objects refers to the individually resolvable entities located within a digital image which are perceptually generated from images (Hay et al., 2001). In high resolution images a single real-world object is modelled by many individual pixels whereas low resolution implies that a single pixel represents the integrated signal of many (smaller) real world objects (Hay et al., 2003). In a remote sensing image, both situations occur simultaneously. For example, a 1 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. The 1m pixel is high resolution in relation to the crown object it models. However, each 1 m pixel will also be composed of the integrated reflectance from many needles/leaves and branches. Thus, it will be low resolution in relation these individual crown components. As a result, image and objects tend to be composed of spatially clustered pixels that exhibit high spatial autocorrelation because they are all part of the same object. Because an 'ideal' object scale does not exist (Marceau, 1999), objects from different levels of segmentation have to be utilised for many applications.

In remote sensing, a single sensor correlates with range of scales rather than a single scale. The detectability of an object can be treated relative to the sensor's resolution. A coarse rule

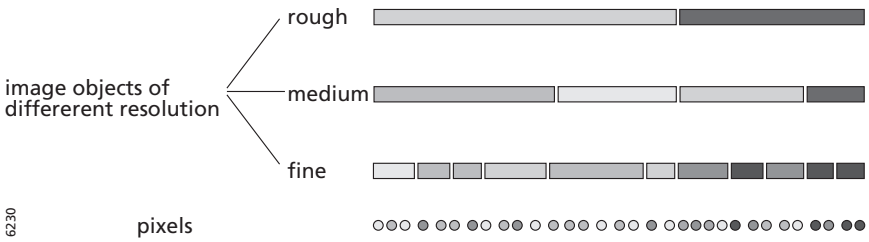

Figure 12.8 – A multi-resolution image segmentation approach.

of thumb is that the scale of image objects to be interpreted must be significantly bigger than the scale of image noise relative to texture (Haralick & Shapiro, 1985). This ensures that subsequent object oriented (OO) image processing is based on meaningful image objects. Among the most important characteristics of any segmentation procedure is the homogeneity of the objects. Only if contrasts are treated consistently, good results are expected (Baatz & Schäpe, 2000). Furthermore, the resulting segmentation should be reproducible and universal which allows the application to a large variety of data. Baatz & Schäpe argue that multi-resolution image processing based on texture and utilising fractal algorithms can fulfil all the main requirements at once.

Existing approaches show advantages but also some potential pitfalls of segmentation approaches for extracting geoinformation and useful landscape elements on 'real' (earth) surfaces (Blaschke et al., 2000, Blaschke & Strobl, 2001). The 'representativeness approach' (Hofmann & Böhner, 1999) mentioned earlier provides a good representation of the local feature distribution, which captures substantially more information than the usually used mean feature values. This one and other boundary-forming techniques (Schneider et al., 1997) and segmentation approaches (Gorte, 1998, Molenaar, 1998, Cheng, 1999, Dubuisson-Jolly & Gupta, 2000) provide good results for test areas but are not necessarily using all context information beyond the spectral information of neighbouring pixels such as texture, shape, directionality, spatial distribution within the study area, connectivity etc. But we strongly believe that this contextual information is the key to advanced classifications.
Although practically all segmentation procedures result in crisp objects, there are several ways to treat continuous transitions and fuzziness of objects. One strategy is to partition an image at several levels as discussed before and to utilize information at the finest, in most cases mechanistic level to express graininess at the higher level. The smallest spatial unit is still crisp except in the approach of Gorte (1998) but the entities at the superior level can be built up in a fuzzy decision process.

An important family of methods that strive to improve accuracy of classification are those using fuzzy sets. With this concept each pixel may have fuzzy membership with more than one class expressed as degree of its membership to each class (values range between 0 and 1). Training data for fuzzy classification need not be homogeneous as is desirable for conventional hard classifiers. Throughout the classification procedure one needs to assign known portions of mixing categories. Popular fuzzy set based approaches are the fuzzy c-means clustering (FCM) or the possibilistic c-means clustering (PCM). The fuzzy classifiers

produce images showing the degree of membership of pixels to stated categories. One caveat of the fuzzy set based methods is that he accuracy of fuzzy classification depends to a high degree on the complete definition of training data sets. Foody (2000) remarks that untrained classes will only display membership to trained classes, which can introduce a significant bias to classification accuracy.

An advanced method is the use of (artificial) neural network classifiers (ANN) borrowed from artificial intelligence research. Training data together with a known land-cover class (the input layer) are fed into the neural network system (the hidden layer). The algorithms inside the network try to match training data with the known class spectra patterns and produce an output layer together with errors of non-matching neural nodes. The procedure restarts trying to minimize errors. The process can be repeated several times. For the classification of specific objects neural networks have proven to be more accurate than conventional methods (Civco, 1993; Foschi & Smith, 1997; Skidmore et al., 1997). Main points of critique include:

- Accurate meaningful results require good training data sets; otherwise outputs will not be very reliable.
- The classification procedure needs the adjustment of various parameters which highly increases complexity of the whole system and seems to limit its usefulness.

In the following chapter, we demonstrate two applications and go into the classification stage whereby we use explicit rules and a semantic network for classification aiming to overcome these shortcomings of ANN by making the rules transparent.

12.4 Examples of applications

12.4.1 Segmentation in multi-source forest inventory

One, and increasingly popular, application field of remote sensing is multi-source forest inventory (MSFI). Since the introduction of first MSFI applications in the late 1960's (Kuusela & Poso, 1970) the rapid development of sensors and image analysis methods have resulted in many, although few operative, MSFI applications. The first operative multi-source national forest inventory began in Finland in 1990. The method employs field plot data, satellite (mostly Landsat TM) images, digital map data and *k*-nearest neighbour (*k*-NN) estimator and produces georeferenced thematic maps and detailed forest statistics for any given area (Tomppo, 1996). Similar MSFI methods have been tested in many different conditions (e.g. Franco-Lopez et al., 2000; Nilsson, 2002; Tomppo et al., 1999; Tomppo et al., 2001). In a typical MSFI application, the field data consists of sparse field sample and the interesting forest attributes are estimated for the rest of the area with help of measured field data and image information. In addition, the field data is usually gathered using relatively small field sample plots and training data is often built in a straightforward way: each field sample plot is assigned spectral information from the pixel it is located on. Methods based on this kind of approach and use of medium resolution satellite imagery have proven to produce reliable forest resource information for large and medium sized areas (Tomppo, 1996; Katila et al., 2000). However, estimation errors have been high at the plot- and stand-levels (Tokola et al., 1996, Mäkelä & Pekkarinen, 2001, Katila & Tomppo, 2001). This is due to many reasons, among which are the poor spatial resolution of the image material and the difficulty to assure

Figure 12.9 – An example of generalised AISA data. NIR channel.

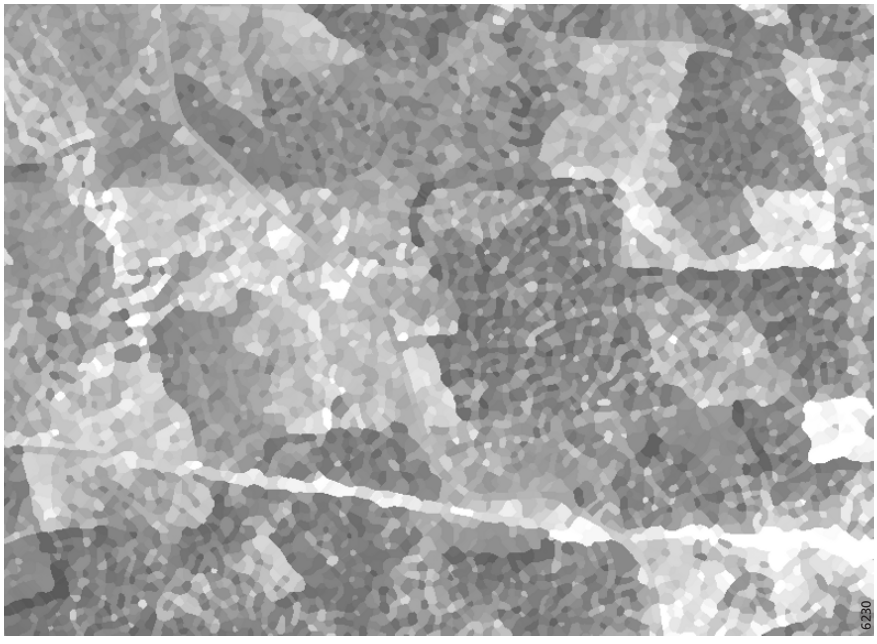

Figure 12.10 – An example of the resulting segmentation. Segment level averages of NIR channel.

exact geo-referencing in both image and field data. These problems can be avoided, at least to some extent, if the MSFI application is based on VHR data.

The increasing spatial resolution in image data results in increasing number of pixels per area unit. This is obvious if one considers the number of pixels falling into the area of a single field plot. Therefore, when comparing VHR and medium resolution data, locational errors of similar magnitude should result in smaller discrepancies between spectral and informational classes of the field plots. The problem is that the pixel-based approach to feature extraction and image analysis does not apply to VHR imagery. From a forest inventory point of view a sub-meter resolution VHR pixel is too small a unit for image analysis. It represents spectral characteristics of only a small portion of the target: a stand, a plot or a tree. In order to be able to fully utilise the improving spatial resolution, need a way to aggregate pixels into appropriate spatial units for the image analysis. One way to accomplish this is image segmentation.

Our MSFI example demonstrates the effect of segment-level feature extraction and image analysis to the resulting MSFI estimates at plot and region level. The material employed in the example has been gathered from a study area that was originally established for MSFI research employing imaging spectrometer data (Mäkisara et al., 1997). The area has been imagined with Airborne Imaging Spectrometer for Applications (AISA). The instrument has been developed in Finland and it has bee employed in several studies since the development of the first prototype AISA in early 1990's (e.g., Mäkisara et al., 1993). Currently the AISA family consists of three different systems: AISA+, AISA Eagle and AISA Birdie (http://www.specim.fi/). The image data were pre-processed in Finnish Forest Research Institute and a mosaic of the seven original flight lines was composed. The radiometric differences between adjacent flight lines were normalised using the overlap area and histogram matching technique. The GIFOV of the resulting mosaic was 1.6 metres. Finally, the number of spectral channels of the original AISA images was reduced from 30 to 4 by means of spectral averaging. The averaging was accomplished in such a way that the resulting output would correspond as well as possible to the spectral characteristics of new generation VHR satellite images. An example of the generalised image data is presented in figure 12.9.

The averaged AISA image was segmented using a two-phase approach. In the first phase, a large number of initial segments were derived using a modified implementation of the 'Image segmentation with directed trees'-algorithm (Narendra & Goldberg, 1980; Pekkarinen, 2002). In the second phase, the minimum size of the segments was set to 10 pixels and all segment smaller than that were merged to their spectrally nearest adjacent segment. An example of the resulting segmentation is presented in figure 12.10.

The field data of our example consists of systematic grid of 262 Bitterlich (relascope) field sample plots (Mäkisara et al., 1997; Pekkarinen, 2002). Each field sample plot was assigned spectral information from A) the pixel the plot centre was located on and B) from the segment the plot was located on.
The performance of pixel- and segment-level features was compared in the estimation of volume of growing stock at plot. The plot level estimation tests were carried out using cross-validation (leave one out) technique. In other words, the total volume was estimated for each

plot with help of the rest of the plots. The actual estimate was derived using a weighted *k*-nearest neighbour estimator (*k*-NN). In weighted *k*-NN, the estimate is computed as a weighted average of *k* spectrally nearest neighbouring observations in the feature space (e.g. Tomppo, 1991, Pekkarinen, 2002). Three different values for *k* were tested, namely 1, 3 and 5. The accuracy of the estimates was judged using root mean square error (RMSE, equation 12.1) of the estimates. In addition, the within volume class distributions of the estimates and observed values were examined.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n - 1}} \qquad (12.1)$$

Where
y_i = measured value of variable y on plot i
$\hat{y}$ = estimated value of variable y on plot i
n = number of plots

The results of our experiment show that the estimates derived using segment level features have significantly lower RMSEs that estimated derived with pixel-level features. The decrease in relative RMSE was from about 11 to 13% depending on the number of *k* (figure 12.11). The benefits of the segment-level approach are also obvious at the region-level. The distribution of the resulting wall-to-wall estimated is much closer to the distribution of the field data in the segment-level estimation approach (figure 12.12).

The example shows that the segmentation-based approach gives significantly better estimation results than the pixel-level approach. However, there is still a lot of room for improvement. One fundamental prerequisite for any image analysis is that the phenomena under investigation and the unit of analysis are of similar scale. This is not the case in our example. Even though the segment level analysis gives better results it is not an optimal solution to the estimation problem: the spectral characteristics of a segment do not

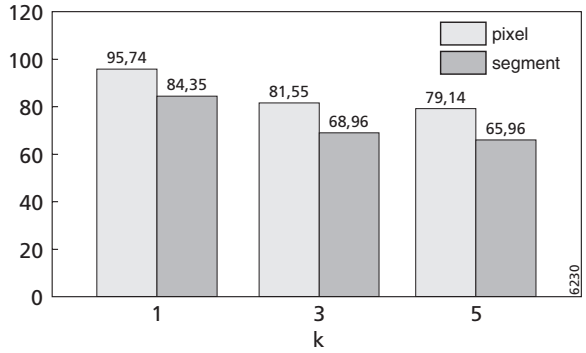

Figure 12.11 – Relative RMS errors of total volume estimates with different numbers of k. Pixel- and segment-level approaches.

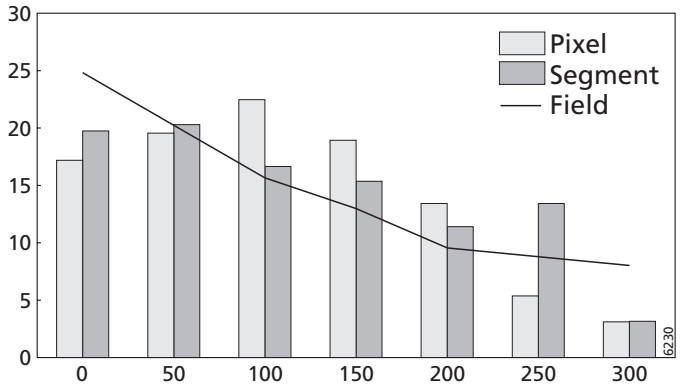

Figure 12.12 – Volume class distribution of the pixel and segment based estimates and the field data.

necessarily represent the spectral characteristics of the area from which the field information was gathered.
In an optimal case one would know the exact location of the plot, and its dimensions and could use this information in construction of the training data. In such a case the extracted spectral information would represent the plot but the problem of determining the spatial units for image analysis in unsampled areas would remain the same. Image segmentation provides a visually appealing solution to this problem but does not necessarily give significantly better plot-level estimation results than a straightforward spectral averaging in local neighbourhood (Pekkarinen, 2002).

12.4.2 Object-based forest stands mapping in an urban forest

In this second example, we demonstrate the use of segmentation as part of a two step multiscale segmentation/object relationship modelling MSS/ORM approach (Burnett & Blaschke, in press) to in a study delineating habitat patches in a mixed hardwood/deciduous urban forest. Just as in the above example, we utilise VHR remotely sensed imagery. And also in this example, 'meaningful' image objects are also deleniated, although using a different segmentation algorithm. At this point the two example diverge. Instead of using a spectral (plus textural) feature space (i.e. the *kNN*) estimator, the image is further segmented so that we have a minimum of 2 levels (scales) of image objects. Following this multiscale segmentation, the ORM step then begins, wherein expert knowledge is integrated into the classification and heuristics using both spectral as well as 'fellowship' (object inter-relationships) information. The software eCognition by Definiens AG of Munich was used for both the segmentation step and the object-relationship modelling.

The study site is located on the 11 km long island of Ruissalo (figure 12.13), west of the city of Turku in SW Finland. The forest patches in the site differ in tree species, stem density, age and stand species homogeneity; ranging from sparsely treed rocky meadows with medium sized Scots pine (*Pinus sylvestris*) to mature mixed stands of lime (*Tilia cordata*), Scots pine, Norway Spruce (*Picea abies*) and oak (*Quercus robar*) exhibiting early patch-phase dynamics. Topographic variation is slight but because of the recent emergence of the island during the

Figure 12.13 – Ruissalo Island study site marked with white box, centred on 60 25 42N & 22 08 53E. Please consult the enclosed CDROM for a full colour version.

Quaternary, organic soil layers are often very thin on upland areas. In addition to the varied microclimatic and soil types, long term human management (eg. introduced species), the island is home to one of the richest species communities in Finland (Vuorela, 2000). The island is now managed as a recreation area with a large proportion of the land area in nature reserves.

The goal of the MSS/ORM analysis was to differentiate between and map upland dry pine sites and mature mixed forests. In addition, the compositional break-down of deciduous and coniferous species in the mixed stands would be useful. The data used was 1m GIFOV digital still camera mosaic acquired with a professional Minolta digital camera, and rectified and mosaiced using software developed by the Enso Forest Company and Technical Research Centre of Finland (VTT). We also had access to the City of Turku's cadastre and road mapping data in vector format.

Multiscale segmentation

The goal of the MSS step is to find three levels (scales) of image objects: level -1 are the sub-units which are mostly characterised by their spectral features; level 0 is the main mapping units which are classified using inter-object relationships; and level +1 which is an aggregate or reporting level. The segmentation of the image was carried out so that in the forest, the smallest segmentation level comprised image objects such as tree crowns, small groups of crowns, shadow and sunlit bare rock (figure 12.14). It was found that for this image resolution/image object combination, suitable segmentation levels could be generated in eCognition using scale parameters of 10, 50 and 100 (always with colour and shape parameters of 0.8 and

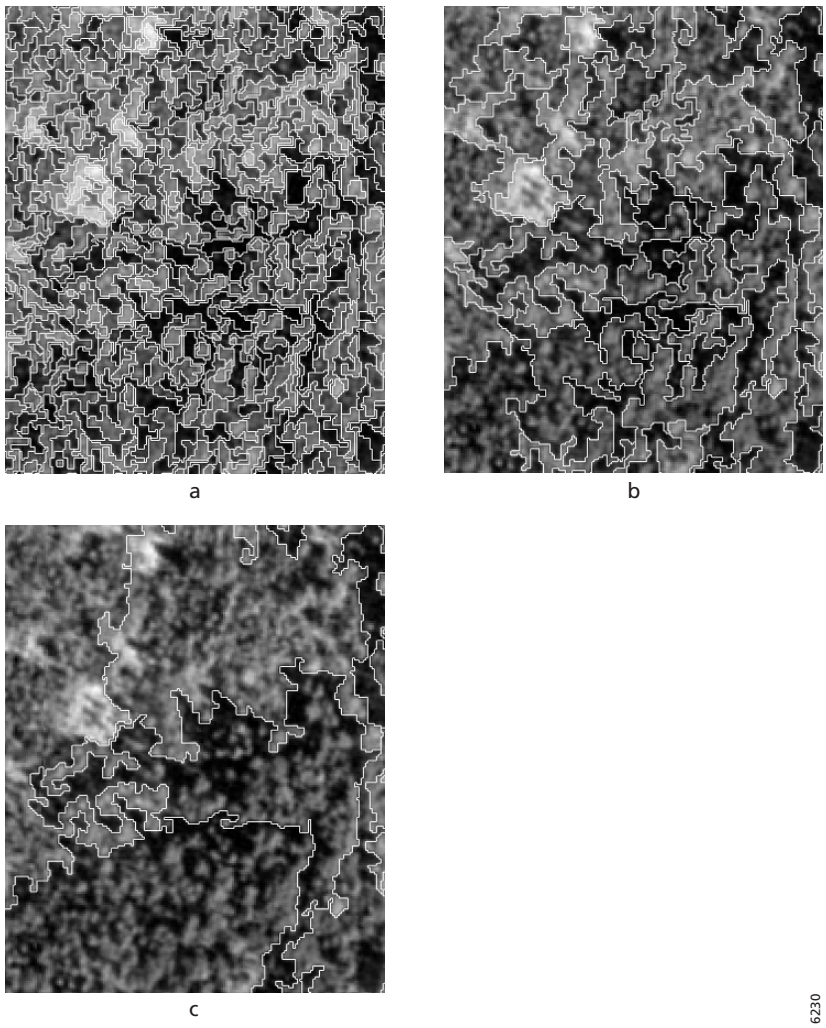

Figure 12.14 – Examples of segmented image objects created in the forested area in the SW of the study area. This is an area that transitions from the mature mixed habitat to the dry upland pine. The images show the forest canopy at the three segmentation levels -1, 0 and +1 (eCognition segmentation scale parameters of 10, 50 & 100).

0.2). We then took samples of the different tree crowns. We had field survey plots to guide us in this step. We also took samples from agricultural fields, pasture, houses and roads. Finally we classified the level -1 objects using a nearest neighbour classification routine.

Object relationship modelling

In the second step, we began to incorporate rules in order to guide the software to properly classify each of the level 0 objects. We use the term modelling, because we are in effect building rules that model (abstract) the functional relationship that exists in the real world

Figure 12.15 – Sub-object relationship mapping guide (see table 12.1).

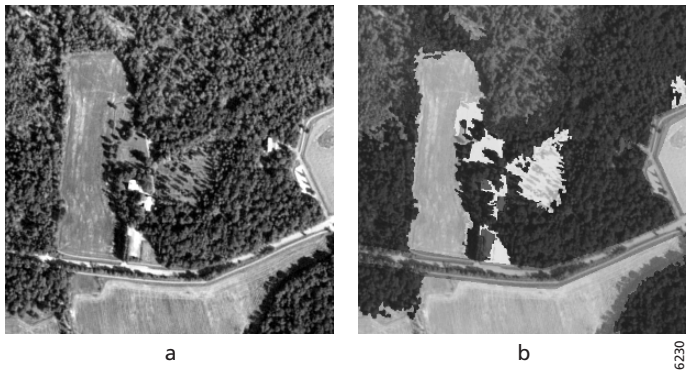

Figure 12.16 – Original image and completed classification at the Level +1 (super-object) level.

between the objects, sub-objects and super-objects. Please see figure 12.15 and table 12.1 for a concise description of these rules. We started by ruling that any object under the main road mask that came from the City of Turku cadastre should be classed 'road' (rule 1). The rules that govern agricultural areas (rule 2) involved both spectral means of the level -1 objects and their standard deviation. These objects are generally larger and smoother then those found in the forest areas. Houses (rule 3) were perhaps the simplest level 0 class: the spectral differences made only a mean of level -1 (sub-object) rule necessary.

There were large areas of shadowed fields that required special consideration (rule 4). We didn't want to confuse shadow in the forest. Here we had to examine level 0 objects to the northwest. If there were level 0 agriculture objects in this direction, we classified them as shadowed field. We could have also used a textural value such as the number of sub-objects or the standard deviation of the spectral values in the sub-objects, since these are smoother then their forest shadow cousins, but this was not needed. Pasture (rule 5) perhaps the most difficult to model. Here, there are relatively small expanses of fields but with large single trees with elongated shadows in the centre. The field sub-objects (level -1) had different spectral signatures than the bare rock found in the upland pine sites, so there was no confusion there. In the end, a heuristic that said any groups of agriculture level 0 objects with greater than 30% are covered by tree and shadow level -1 objects was arrived at after several iterations of testing. Finally, we were able to concentrate on separating the to focal forest units. We had fairly good separability of the coniferous and deciduous level -1 objects. We also used the amount of bare rock/soil level -1 objects. And we knew that the upland sites rarely touched agricultural areas. Using a combination of these rules, we came up with heuristics that modelled this 'behaviour'.
In the final step, we aggregated all neighbouring level 0 objects of the same class into single polygons (figure 12.16). These were the polygons upon which the reporting statistics and spatial indices were calculated. Values for average amount of level -1 sub-objects (i.e. tree crowns) were also reported.

The result of the multiscale analysis using the MSS/ORM methodology is a flexible mapping of these semi-natural forests, but with the added benefit of having a record of sub-object characteristics. Thus we could go further in our analysis and further classify the upland pine into areas with more or less bare rock. Or we may want to identify zones in the mature mixed forest that are within a fixed radius of an oak crown of a certain dimension. With

Table 12.1 – Object-relationship modelling rules

Diagram label	Class	Heuristic
1	Roads	Road vector layer used as mask
2	Agriculture	Classified from spectral characteristics at Level 0
3	Houses	Classified from spectral characteristics at Level 0
4	Shadowed fields	Level -1 segments classified relative to segments to NW.
5	Pasture	Agricultural level 0 segments having >30% 'single tree + shadow' sub-objects
6a	Mature mixed	Classified by spectral, textural and sub-object rules
6b	Upland dry pine	Classified by spectral, textural and sub-object rules

the increasing availability of scanning LIDAR data, this ORM modelling will be even more powerful.

12.5 Discussion and conclusions

We began the chapter with a comparison of different image processing strategies. The image analysis presented here provides methodology and examples of dealing with image semantics rather than pixel statistics. 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. Prerequisite for the successful automation of image interpretation are therefore procedures for image object extraction which are able to dissect images into sets of useful image objects. As stated above, segmentation is not new, but only a few of the existing approaches lead to qualitatively convincing results while being robust and operational. One reason is that the segmentation of an image into a given number of regions is a problem with a huge number of possible solutions. The high degrees of freedom must be reduced to a few which are satisfying the given requirements.

The multiscale consideration of landscape pattern gains much attraction recently but the realisation in practise becomes difficult and data-intensive. Only for some small areas, field surveys and mapping at different scale is possible. Not a solution but one step forward to support this approach is a nested multi-scale image processing of the same data sources. The resulting different object scales have to be logically connected. This is achieved through an OO approach where each object 'knows' its intrinsic relation to its superobject (*is within*) and its subobjects as well as the relations to the neighbouring objects at the same scale.
So far, it is concluded, that context based, object-oriented image classification is a promising development within integrated GIS/RS image analysis. Comprehensive studies using multi-sensor data sets which explore the 'behaviour' (stability and consistency of the image objects and their respective classification results due to different data situations) are still urgently required. However, several studies indicate that current description schemata for landscape objects are dependent on scale, resolution and class definition (Hargis et al., 1998, Herzog & Lausch, 2001). Very few studies already illustrated the potential of context-based approaches to improve classification results in real-world studies, e.g. Lobo et al. (1996). Although the literature mentions the possibilities of object-based image analysis since two decades (Kettig & Landgrebe, 1976, Haralick & Shapiro, 1985), only latest-technology hardware, intelligent software and high resolution images can advance this concept.

While using image segmentation, a somewhat implicit hypothesis is that results from objects or regions based on segmentation are often easier to interpret and more meaningful than those of per-pixel classification. Only recently, some studies compare the accuracy of both approaches. A main finding is, that the results of the latter often appear speckled even if post-classification smoothing is applied (Blaschke & Strobl, 2001, Ivits et al., 2002). The second strategy originates in conceptual ideas of landscape ecology and information structuring and puts forward a conceptionalisation of landscape in a hierarchical way utilising remote sensing and GIS data at different scales resulting in an object-oriented modelling approach and the construction of semantic networks.

This chapter took up the challenge to the technology which lies in an ontology inherent to modern landscape consideration: landscapes are composed of a mosaic of patches (Forman, 1995). But patches comprising the landscape are not self-evident; patches must be defined relative to the given situation. From an ecological perspective, patches represent relatively discrete areas (spatial domain) or periods (temporal domain) of relatively homogeneous environmental conditions, where the patch boundaries are distinguished by discontinuities in environmental character states from their surroundings of magnitudes that are perceived by or relevant to the organism or ecological phenomenon under consideration. The technical potential of oo-based image processing will be applied to the delineation of landscape objects. The involves issues of up- and downscaling and the need for an ontology and methodology of a flexible 'on-demand' delineation of landscape objects hypothesising that there are no 'right' and 'wrong' solutions but only 'useful' and 'meaningful' heuristic approximations of partition of space. The GIS development shall be the medium for the transfer of the indicators and techniques developed into operational use.

The object-oriented approach is also a philosophy of improved image understanding. Human vision is very capable of detecting objects and object classes within an image. To pursue this type of analysis, it is important to stay close to the 'intuitive' image understanding. Object-based classification starts with the crucial initial step of grouping neighbouring pixels into meaningful areas, especially for the end-user. The segmentation and object (topology) generation must be set according to the resolution and the scale of the expected objects. The spatial context plays a modest role in pixel based analysis. Filter operations, which are an important part of the pixel based spatial analysis, have the limitation of their window size. In object analysis, this limitation does not exist. The spatial context can be described in terms of topologic relations of neighbouring objects. But how to define the image objects? What should be the rule of thumb for a segmentation or a pre-segmentation of image primitives which can build up the corresponding objects? Only very recently, several approaches in image analysis and pattern recognition are exploited to generate hypothesis for the segmentation rules as an alternative to knowledge-based segmentation (Blaschke & Hay, 2001, Lang, 2002, Hay et al., 2003).

The new data available necessitate improved techniques to fully utilise the potential resulting from the combination of high-resolution imagery and the variety of medium-resolution multi-spectral imagery widely available. High-resolution panchromatic images now show variety within a so far 'homogeneous' area (a pixel in the medium-resolution image). The understanding of local heterogeneity in a panchromatic image has a strong effect on the standard deviation value in the same 'window' area of the image. Nevertheless, they can be simultaneously used to classify areas, where spectral values in multispectral bands are less important compared to local texture.

The Earth Observation data are not sufficient to characterise completely the natural environment. They need so to be associated to other data. We have to further investigate synergy effects between satellite images and GIS-derived vector data, such as a digital topographic database. The use of such topographic databases, which are built up in most countries can support the satellite image analysis. This digital database offers a geometric as well as a semantic prediction for objects in the satellite image.

Current average geometric resolution of common satellite sensors allows to develop main area-based land use classes like 'settlement', 'forest', 'water' or 'agriculture'. In combination with high-resolution imagery described above, expected results of this feature extraction allow a symbolic description of more complex image content such as 'urban forest' or 'traditional, small-sized, complex land use'. Both, the symbolic description and the digital database, are transferred in a semantic network, a compact formalism for structuring the entire knowledge. None of the various pixel-based classification methods seems to really satisfy all the needs for the production of reliable, robust and accurate information similar to objects identified by a human interpreter.

Important next research steps if image segmentation is used as a core methodology in an image analysis include:

- To develop a methodological framework of object-based analysis which offers an option to improve (semi)automatic information derivation from a rapidly increasing amount of data from different sensors;
- To achieve meaningful representations of spatial physical parameters derived from original reflectance values
- To develop new methods for accuracy assessment based on objects rather than pixels;
- To develop data analysis tools for environmental management based on multi-scale object representation in order to support decision making.