METSÄNTUTKIMUSLAITOKSEN TIEDONANTOJA 926, 2004
FINNISH FOREST RESEARCH INSTITUTE, RESEARCH PAPERS 926, 2004

Image segmentation in multi-source forest inventory

Anssi Pekkarinen

VANTAAN TUTKIMUSKESKUS – VANTAA RESEARCH CENTRE

METLA

METSÄNTUTKIMUSLAITOKSEN TIEDONANTOJA 926, 2004
FINNISH FOREST RESEARCH INSTITUTE, RESEARCH PAPERS 926, 2004

Image segmentation in multi-source forest inventory

Anssi Pekkarinen

Academic Dissertation

VANTAAN TUTKIMUSKESKUS – VANTAA RESEARCH CENTRE

Pekkarinen, Anssi. 2004. Image segmentation in multi-source forest inventory. Finnish Forest Research Institute. Research papers 926. ISBN-951-40-1929-6. 35 p. + 4 original papers.

Supervisor: Professor Erkki Tomppo
Finnish Forest Research Institute
National Forest Inventory

Pre-examiners: Professor Alan Ek
Department of Forest Resources
University of Minnesota, USA

Professor Håkan Olsson
Department of Forest Resource Managements
and Geomatics
SLU, Umeå, Sweden

Opponent: Ph.D. Tuomas Häme
Chief Research Scientist
VTT Information technology
Remote sensing group
Espoo, Finnland

Publisher: Finnish Forest Research Institute, Vantaa Research Centre, P.O. Box 18, FI-01301 Vantaa, Finland. Accepted by Kari Mielikäinen, Research Director, June, 2004.

Author : Anssi Pekkarinen
Finnish Forest Research Institute
National Forest Inventory
Unioninkatu 40 A, FI-00170 Helsinki

Front cover: AISA image: ©METLA/NFI

Printed in Dark Oy, Vantaa 2004.

Contents

Abstract

This thesis examines whether the applicability of remote sensing aided forest inventory methods can be improved by using the image segment -based approach to feature extraction and image analysis. Several image segmentation techniques are developed, implemented, and tested in forest inventory applications employing both, satellite and aerial remote sensing material. The inventory applications in which the segment-based approach is tested include the estimation of plot-level timber volumes (I -III) and the stratification of forested areas (IV).

All the tested and developed segmentation algorithms are applicable to the determination of feature extraction and image analysis units. The incorporation of image segment and sub-segment features into the estimation procedure improved the plot level volume estimates in most of the cases (I-III). The achieved reductions in the RMSEs are smaller than expected and it is questionable whether these small improvements justify the use of the segment-level estimation approach. Better results are achieved in the stratification of forested areas, even though some of the results are controversial (IV). The segment-aided approach can be recommended for stratification purposes.

The selection of the appropriate segmentation algorithm depends on the application, but in general a two-phase approach starting with initial segmentation and proceeding with merging of the initial segments is advisable. The initial segmentation method presented in III, and a region merging algorithm that can be guided with minimum segment size and similarity parameters are recommended methods especially for forest inventory applications that employ VHR imagery.

List of papers

I Mäkelä, H. & Pekkarinen, A. 2001. Estimation of timber volume at the sample plot level by means of image segmentation and Landsat TM imagery. Remote Sensing of Environment 77(1): 66-75.

II Pekkarinen, A. 2002. A method for the segmentation of very high spatial resolution images of forested landscapes. International Journal of Remote Sensing 23(14): 2817-2836.

III Pekkarinen, A. 2002. Image segment-based spectral features in the estimation of timber volume. Remote Sensing of Environment 82(2-3): 349-359.

IV Pekkarinen, A. & Tuominen, S. 2003. Stratification of a forest area for multi-source forest inventory by means of aerial photographs and image segmentation. In: Advances in Forest Inventory for Sustainable Forest Management and Biodiversity Monitoring. Forestry Sciences. Vol. 76: 111-124. Kluwer Academic Publishers, Dordrecht, Netherlands, ISBN 1-4020-1715-4, 460 pp.

In I, Pekkarinen was responsible for developing and implementing image segmentation and estimation algorithms. The report was written together with Mäkelä. In IV, Pekkarinen was responsible for developing and implementing of image segmentation and clustering algorithms. The report was written together with Tuominen.

Acknowledgements

The majority of this research was conducted from April 1998 to June 2002, during which time I was a part-time post-graduate student within the "Forests in Geographical Information Systems" Graduate School, University of Helsinki and part-time research scientist at the Finnish Forest Research Institute (Metla). I greatly appreciate the financial support provided by the Foundation of Foresters (Metsämiesten Säätiö) through the Graduate School, without that this work could not have been realised. In addition, I would also like to thank both, Metla's National Forest Inventory research program and the University of Helsinki, Department of Forest Resource Management for providing the research material and all necessary work facilities for this thesis.

Numerous people encouraged and helped me throughout the course of this study.

My supervisor Professor Erkki Tomppo supported my idea to study image segmentation and helped me on several occasions. Both, he and Dr. Jari Varjo also initially encouraged me to apply to the Graduate School. I greatly appreciate their support. I would also like to thank Prof. Jouko Laasasenaho and Dr. Markus Holopainen for guiding us doctorants through the Graduate School and for arranging inspiring post-graduate seminars in Helsinki and Hyytiälä. Thanks also go to Emeritus Professor Simo Poso, with whom I have had many interesting and thought-provoking debates during both, my MSc and doctoral studies. It may actually be, that my interest in remote sensing can be traced back to an enlightening discussion that we had in a mature spruce stand in Toivala, May 1992.

I have been very fortunate to work with the Finnish National Forest Inventory team. I would particularly like to thank my excellent co-authors and colleagues Lic. Sc. Helena Mäkelä and Mr. Sakari Tuominen In addition to my co-authors, I am especially grateful to Lic. Tech. Kai Mäkisara who provided me with an outstanding programming environment for implementing the concepts that I had. Kai also helped me on numerous occasions to debug my awkward 'C' code and patiently answered all of my questions related to the AISA data. I would also like to thank Dr. Juha Heikkinen and Dr. Helena Henttonen for helping me to understand what the Ohkola field data was about.

Beyond the scope of this thesis, I would like to express my gratitude to my former and present colleagues that I have had the pleasure to work with. In particular I would like to acknowledge Ms. Reija Haapanen with whom I wrote my first scientific article, Ms. Tarja Tuomainen with whom I have shared many of my work-related headaches over the past couple of years and my dear colleagues and 'co-lunchers': Ms. Sirpa Rajaniemi, Mr. Markus Haakana and Mr. Mikael Strandström. I have shared many enjoyable and cheerful moments with all of them.

I would like to express my sincere gratitude to Prof. Annika Kangas, her encouragement and comments on the synthesis part of this thesis were invaluable. I would also like to thank my pre-examiners Prof. Alan Ek and Prof. Håkan Olsson for their constructive and valuable preview comments. Thanks also go to Dr. Ashley Shelby for editing the language of sub-studies I-III and the synthesis part of this thesis and to Ms. Anna-Kaisu Korhonen and Ms. Anne Siika for their friendly assistance in the publishing phase of this work.

Finally, I would like to thank my wife Heli for initially encouraging me to start the post-graduate studies and for her continued support during this process and our children Pihla and Pietu for constantly and concretely reminding me that every day really is an adventure.

Helsinki, June 2004

Anssi Pekkarinen

1 Introduction

Remote sensing (RS) provides invaluable information concerning spatial and temporal distribution of land use and forests. The oldest earth observation (EO) program provides spaceborne data dating back to 1972 with the launch of Landsat 1 (originally ERTS 1). In addition, aerial RS archives contain data that originate from the first decades of the 20th Century. Combining the temporal and spatial dimensions of these data enables the assessment of not only the present state of the environment but also the changes that have occurred over various periods of time. The value of these data has been recognised among the foresters and the number of RS based forest inventory applications is constantly increasing.

Although RS aided forest inventories have been shown to produce valuable information for large and medium sized areas, their applicability at the forest holding and stand level has been limited. Stand- and plot level estimation errors have remained high with different RS information sources and estimation techniques. These results imply that the high estimation errors for small areas do not straightforwardly result from the spatial or spectral resolution of the employed RS material, but the fundamental reasons for partial failures are elsewhere.

This thesis examines whether the applicability of remote sensing aided forest inventory methods can be improved by using the image segment -based approach to feature extraction and image analysis. Several image segmentation techniques are developed, implemented and tested in forest inventory applications employing both, satellite and aerial RS material. In order to present the issue in a broader perspective, let us first review the common history of remote sensing and forest inventory.

2 Remote sensing in forest inventory

In Finland, the first tests of airborne RS aiming at cartographic mapping were conducted by the Topographic Service of Finland in 1926-1927 (Löfström 1946). The results were promising and the tests stimulated significant technical and methodological developments. The developed techniques and equipment were soon given wider applications. Before the end of the 1930's, both the Topographic Survey of Finland and the Finnish National Land Survey provided maps based on aerial photographs in order to *"satisfy the constantly increased map-demand of economic and social life"* (Löfström 1946).

The potential of aerial photographs was soon recognised among the foresters. According to Sarvas (1938) and Nyyssönen (1955), the first forestry related applications were reported in Central Europe at the beginning of 1920's, but the initiative was later *"transferred to U.S.A and Canada"* (Nyyssönen 1955). In these first applications, aerial photos were used as maps in fieldwork and in timber surveys (Nyyssönen 1962, Lillesand et al. 2004).

After World War II, the increasing use of the infrared films stimulated the use of aerial material, and the number of applications increased. Probably the first significant Finnish effort to estimate the amount of growing stock from aerial photographs was presented in 1955 (Nyyssönen 1957). Nyyssönen concluded that the sensible way to accomplish aerial photography aided forest inventory would be the combination of fieldwork and aerial photo interpretation. The interpretation of aerial photographs could not satisfy all information needs of the forest management but, on the other hand, they provided possibilities that should not be ignored (Nyyssönen 1957). However, aerial photos were still used mainly as *"maps facilitating the work"* at the beginning of 1960's (Nyyssönen 1962).

In the late 1960's, foresters' expanded their attention from aerial photos to satellite imagery (e.g. Kuusela and Poso 1970), but a significant increase in forest inventory related satellite RS application was seen only after the launch of first satellite of Landsat program in 1972. Landsat 1 carried a multispectral scanner (MSS) instrument and provided multispectral EO data in digital format (Lillesand et al. 2004). This new data source and the rapid development of computer technology resulted in major developments in image analysis techniques and in an increasing number of RS applications.

So far, most of the environmental RS research has concentrated on examining the present and monitoring recent changes. The focus has been in land use and forest type classification (Bauer et al. 1994, Holopainen 1998, Haapanen et al. 2004), forest change (Goldberg et al. 1982, Saukkola 1982, Varjo 1996, Häme et al. 1998, Woodcock et al. 2001) and prediction of values of forest variables, such as timber volume and woody biomass, for areas of various sizes (Hagner 1987, Tomppo 1992a, Tokola et al. 1996, Mäkisara et al. 1997, Trotter et al. 1997, Kilpeläinen and Tokola 1999, Tomppo et al. 1999, Hyyppä et al. 2000, Halme and

Tomppo 2001, Holmström et al. 2001, Tomppo et al. 2001, Tuominen and Poso 2001, Anttila 2002, Dong et al. 2002, Tomppo et al. 2002).

Since Landsat 1, the spatial resolution of the RS material available in digital format has continuously improved. Currently, several satellite imaging systems provide data that possess spatial resolutions higher than 5 meters (Nieke et al. 1997). Furthermore, the availability of aerial imagery has significantly increased. New image compression algorithms and the development of web-based services provide ready access to digitized aerial photographs (METRIA 2003, ILMARI 2003) and the introduction of digital aerial cameras such as Z/I DMC (Z/I Imaging 2004) and Leica ADS40 (Leica Geosystems 2004) to operative aerial imaging services is approaching. In addition to aerial photography, aerial very high spatial resolution (VHR) data is available from airborne imaging spectrometers (e.g. AISA, CASI, DAIS 7915, ROSIS and HyMap) and active sensors such as airborne laser scanners (e.g. TopoSys, Optech ALTM and Leica ALS50) and airborne radars (e.g. CARABAS and GEOSAR). Many of these new data sources have shown to provide interesting data for forest inventory applications (e.g. Fransson et al. 2000, Holopainen 1998, Hyyppä et al. 1999, Mäkisara et al. 1997, Næsset 1997).

The challenge that the very high spatial resolution image material introduces to image analysis can be explained using the H- and L-resolution concepts (Strahler et al. 1986). In the case of L-resolution, the resolution of RS material is coarse in relation to the size of object of interest, i.e. a single L-resolution pixel contains information on several objects of analysis. In such a case, the image analysis can be approached using fuzzy or sub-pixel level image processing techniques. With these methods, the user can model the informational content of the pixel despite the fact that the objects of interest cannot be directly observed in the spatial domain. In the case of H-resolution, the problem is reversed and reasonable analysis of the object properties requires that the contextual information present in the image is taken into account. One, and probably the most simple, way to accomplish this is to use the square or rectangular neighbourhood of each pixel and to utilize spectral information from that area in the analysis instead of the information from a single pixel. The problem with this kind of approach is: how to determine the appropriate size and shape of the neighbourhood to be employed in the analysis of the phenomena of interest? Even though an adaptive solution to the definition of the size of the geographic window may be possible (Franklin et al. 1996), the determination of the shape remains a problem. It can be solved with methods that are able to compose spatial entities i.e. regions that can be used as basic units in image analysis instead of single pixels (Blaschke and Strobl 2001). These regions can be determined with help of image segmentation.

3 Image segmentation

3.1 General

Image segmentation is the division of an image into spatially continuous, disjointed and homogeneous regions. More formally, following the notation presented by Pal and Pal (1993), if a digital image is presented as

$$F_{PxQ} = [f(x,y)]_{PxQ}$$

where PxQ is the size (columns x rows) of the image and $f(x,y) \in G_L = \{0,1,...,L-1\}$ is the set of possible grey level values, image segmentation is partitioning of the set F into a set of homogeneous regions S_i in such a manner that

$$\bigcup_{i=1}^{n} S_i = F \text{ with } S_i \bigcap S_j = \varnothing, \quad i \neq j$$

The homogeneity of the regions is controlled with a homogeneity criterion, denoted by $P(S_i)$. The criterion has to be true for each region and false for adjacent regions. This ensures that every region is distinct from every other region. More formally: $P(S_i \bigcup S_j)$ is false when S_i is adjacent to S_j. $P(S_i)$ can be determined as convenient. It can, for example, be set in such a way that a segment may include only pixels that carry the same grey level value. In real-world applications, however, the criteria are usually much more complicated and may consist of, for example, a set of spectral and geometrical rules.

As pointed out by Haralick and Shapiro (1985), there is no theory of image segmentation. In addition, even though image segmentation is precisely defined, the word "segment" may be sometimes confusing. The online Merriam Webster (2004) dictionary gives, among many others, following meanings to the word "segment":

"One of the constituent parts into which a body, entity, or quantity is divided or marked off by or as if by natural boundaries" and *"Portion cut off from a geometric figure by one or more points, lines, or planes".*

Thus, the word "segment" does not explicitly involve the requirement of spatial continuity that is characteristics to image segments. Therefore RS students and even specialists often misunderstand the meaning of an image segment by confusing it with a cluster. This is understandable when one considers the fact that the segmentation or clustering of an image may sometimes yield an identical result. This is an exception, however, and is found only in cases where the image consists of a background and a single, spatially continuous and separable, object.

3.2 Classification of image segmentation techniques

Image segmentation techniques can be classified in many different ways depending on the level of details included. Fu and Mui (1981) use three relatively coarse classes: 1) characteristic feature thresholding or clustering, 2) edge detection and 3) region extraction. Haralick and Shapiro (1985) use a more detailed classification and divide the methods into 1) measurement space guided clustering, 2) region growing schemes that include single, hybrid and centroid linkage region growing methods, 3) hybrid linkage combination techniques, 4) spatial clustering schemes and 5) split and merge schemes. Also Pal and Pal (1993) use a quite detailed classification and even separate colour image segmentation to its own class: 1) grey level thresholding, 2) iterative pixel classification, 3) surface based segmentation, 4) segmentation of colour images and 5) edge detection and 6) methods based on fuzzy set theory.

The latter two classifications are unduly complicated and technically oriented for the brief introduction to segmentation of remotely sensed images. The division into pixel-, edge and region based methods is sufficient for that.

Pixel based image segmentation methods include image thresholding, clustering in the feature space and other methods that rely on pixel-level information and employ it in the global feature space. The pixel-based methods may also include spatial components. For example, the input image may be a smoothed version of the original image. In such a case, the original pixel value has been replaced with a weighted average of its neighbouring pixels. Another example is a case in which the pixel has been assigned texture information describing its neighbourhood. These examples can, however, be classified as pixel-based segmentation approaches if the actual segmentation is conducted in the global feature space.

Image thresholding is a pixel-based technique in which an image is turned into a binary image in such a way that the objects of interest are separated from the background. The selection of an appropriate threshold value is usually based on *a priori* known properties of the object and background. Even though image thresholding may, in many cases, seem trivial that is not usually the case. The contrast between the objects may be poor and the illumination conditions may vary and cause artefacts (such as shadows) that make the determination of appropriate threshold difficult. In addition, many applications require that the appropriate thresholds can be determined automatically. However, from a forester's point of view, image thresholding, as such, is usually applicable only to relatively simple segmentation problems such as binarization an image into bright and dark areas for local maxima detection (e.g. Pitkänen 2001), and extraction of water bodies or clouds from remotely sensed images. Therefore, the discussion about different thresholding techniques is beyond the scope of this thesis. Examples of these techniques can be found in, for example, Weszka (1978) and Jain et al. (1995).

Image clustering can be seen as a multi-dimensional extension of thresholding (Fu and Mui 1981). A typical image clustering algorithm, such as ISODATA, is

an iterative process that seeks to find natural classes within the feature space with help of user provided parameters. Depending on the implementation, the required parameters may include the number of clusters, the maximum number of iterations, the convergence threshold etc. A common solution is to start the clustering with a set of initial cluster centres that have been located in the multi-dimensional feature space in such a way that the distance between the centres is maximized. During the first iteration, each image pixel is assigned to that cluster centre that is closest to it in the given feature space. The locations of the cluster centres are subsequently re-determined with help of the pixels that fell to each cluster. The process is iterated until all the pixels remain in same clusters during two sequential iterations or until the proportion of the pixels changing clusters is smaller than the given convergence threshold (e.g. ERDAS 1994).

Because image thresholding and clustering methods produce results that may have several spatially discontinuous units that carry the same label, the result does not fulfil the definition of segmentation until the spatially continuous regions have been identified and re-labelled. This can for example be done using connected component labelling (CCL) -algorithm (Jain et al. 1995).

The nature of edge-based image segmentation methods differs significantly from that of pixel-based methods. The first phase in all edge-based segmentation algorithms is, of course, the detection of edges. An edge point (pixel) in an image can be defined as:

"... a point in an image with coordinates [i,j] at the location of a significant intensity change in the image." (Jain et al. 1995).

Given this definition, to decide whether a pixel is an edge pixel or not one needs to analyse it and its neighbourhood. In general, edge detection consists of the following steps: a) filtering, b) enhancement and c) detection (Jain et al. 1995). The filtering step is required because most of the edge enhancement methods are relatively sensitive to image noise and therefore they perform better when using a smoothed input. The edge enhancement phase is usually carried out using specific edge operators that emphasize pixels having significantly different values than their neighbours. Most of these operators, such as Roberts, Sobel and Prewitt operators, are based on discrete approximation of the gradient that in the case of images is a two-dimensional equivalent of the first derivative (Jain et al. 1995). They usually produce sufficient results for most applications, even though they typically result in relatively thick edges. In case a more precise location of the edges is required, second derivative operators, such as Laplacian and Second Directional Derivative (Jain et al. 1995), can be used.

After image filtering and edge enhancement, the remaining step in edge detection is the recognition of edge points (pixels) among the edge candidates. This is usually carried out with help of thresholding. In the simplest case, all pixels having an edge magnitude above a threshold T are considered as edge pixels. In many real-world cases that deal with noisy images, it may be very difficult to find a threshold that keeps the probability of detecting false edges low while finding all the relevant edges. It may therefore often be necessary to use

several thresholds. For example, Canny (1986) suggests the use of two thresholds T_1 and T_2. The idea is to detect an edge contour using the higher threshold T_2 and to also mark as edges all connected edge pixels of that contour that have edge magnitude higher than T_1. The suggested relation for the thresholds is: $2T_1<T_2<3T_1$.

Despite the method with which the edge pixels are detected, the final phase of the edge-based methods is to link the detected edges and to compose meaningful boundaries. The simplest way to represent a boundary is to use an ordered list of its points, but more compact representations are usually preferred because they provide more efficient basis for subsequent operations. Examples of different means to represent boundaries can be found in, for example, Jain et al. (1995).

Region-based image segmentation techniques differ from pixel- and edge-based methods in the way they deal with spatial relationships. Region-based techniques can be all seen as region growing techniques (e.g. Zucker 1976) or further divided into region growing, merging and splitting techniques, and their combinations. Here, the latter classification is used.

There are several approaches to region growing. The algorithm may require a set of seed pixels or regions with which the process is started, or it may simply start with the initial image and process it pixel-by-pixel. If seeding is required, the seed pixels or areas may be shown interactively on screen or selected automatically. Where seeding is not required, the processing usually begins from the top left corner of the image and proceeds from left to right and top to bottom. Despite the processing details, the region growing techniques usually join neighbouring pixels to a same region if their spectral properties are similar enough. The similarity can be determined in terms of a homogeneity criterion or a combination of homogeneity, size or some other characteristics criteria (Zucker 1976). Following the definition of segmentation, the region growing process terminates after every pixel has been assigned to a segment.

In region merging and splitting techniques, the image is divided into sub-regions and these regions are merged or divided according to their properties. The basic idea is to start with initial regions and merge similar adjacent regions. These initial regions may be single pixels, or areas determined with help of any low-level segmentation technique. Region splitting methods operate in the opposite fashion; the input usually consists of large segments that are divided into smaller sub-segments with help of a simple geometric rules. If the sub-segments are not homogeneous enough, they are further divided and the process is continued. A common way to implement the region-splitting technique is to use a quad tree structure (Figure 1). The basis on which the splitting or merging is done may be, for example, the spectral similarity of the segments or the magnitude and length of their common edge (Zucker 1976).

The above presented classification of image segmentation techniques is not comprehensive. Many segmentation methods fuse properties of algorithms of several classes. An example of a segmentation method that combines features of edge- and region-based approaches is the "Image segmentation with directed trees" -algorithm (Narendra and Goldberg 1980).

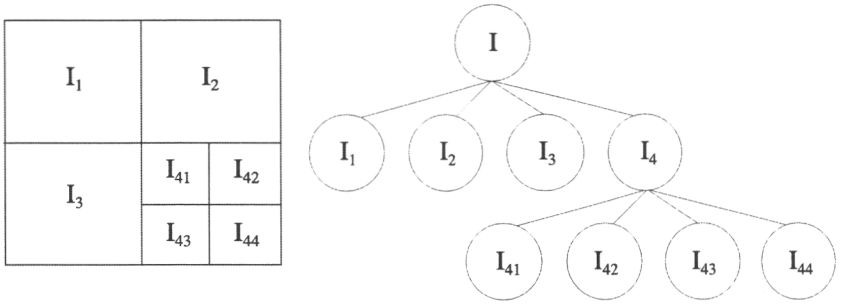


Figure 1. Segmented image and corresponding quadtree (modified from Gonzales and Woods, 1993).

The selection of the appropriate image segmentation approach and algorithm for a specific task on the basis of the algorithm description may be difficult. The segmentation approaches may have advantages and disadvantages that cannot be recognised prior to testing the algorithms with actual imagery. In addition, several different algorithms may result in similar or not dissimilar segmentation output. In practice, the decision is usually made between the algorithms that are commercially available. Unfortunately, there are few such algorithms.

Practically, the first commercially available segmentation software packages that were designed for the analysis of remote sensing data were released in 2000 (Schieve et al. 2001). Since then the interest in the development of such segmentation packages has increased and currently segmentation tools are available for many leading image processing software packages. One software package that deserves explicit mentioning in forestry context is *eCognition*. It is currently the leading commercial image segmentation and object-oriented image analysis software designed for analysis of RS data and has recently been strengthened with a new tool designed for automated tree crown delineation (Definiens 2003).

eCognition's multivariate segmentation is based on a region merging technique that starts with regions of one pixel in size. The region-merging algorithm is iterative and it merges adjacent regions based on their spectral and spatial properties. The main parameters controlling the algorithm are "scale" and "homogeneity criteria". The "scale" restricts the allowed heterogeneity of the resulting segments with help of the homogeneity criteria that can be controlled by weighting the "colour" and "shape" parameters. "Colour" refers to spectral and "shape" to geometric properties of the segments. Furthermore, the shape parameter is a combination of segments "smoothness" and "compactness" that can be weighed by the user (Baatz et al. 2002).

4 Segmentation in forest inventory

The use of image segmentation as a tool for analysing earth observation imagery is, of course, not new. In their segmentation review, Haralick and Shapiro (1985) cite EO image segmentation studies that have been conducted as early as the mid-1970's. The actual need for image segmentation tools was, however, recognised later, soon after the launch of the Landsat 4 (1982) and SPOT 1 (1986) satellites. These satellites introduced new sensors that provided images with considerably improved spatial resolution. In addition to early satellite image segmentation pioneers, many other scientist among the RS community were convinced that these sensors would motivate an increased use of contextual methods in the analysis of satellite remote sensing images.

Despite the fact that the need for contextual image analysis methods was recognised, only few forestry applications for segmentation of EO images were presented. In Scandinavia, many of the forestry related approaches were developed on the basis of directed-trees -algorithm presented by Narendra and Goldberg (1980). Tomppo (1987) tested the algorithm in stand delineation for the estimation of several stand attributes, namely total volume, mean diameter at breast height, mean age and volume proportion by species for pine, spruce and deciduous trees. Later, a similar or not dissimilar method has been applied to the segmentation of Landsat and SPOT imagery by Parmes (1992), and in the spatial generalization of pixel-level change-detection (Häme 1991) and forest site fertility classification results (Tomppo 1992b).

Algorithms designed for more specific purposes were also developed. Hagner (1990) presented a method he calls *"t-ratio segmentation"* that is used for the automatic delineation of stands. Hagner describes the method as *"a type of region growing algorithm"*, but it can also be classified as a region merging method. The same segmentation method has later been used for change detection (Olsson 1994). Another example of an algorithm that has been designed for a specific purpose is a method that was aimed at the delineation of stands for the construction of forest canopy reflectance models. The method has been presented by Woodcock and Harward (1992) and was later employed in the generalization of change detection results (Woodcock and Macomber 2001).

In general, the segmentation methods produced promising results in stand delineation, in the estimation of forest parameters and in post-processing of the results of pixel-based analysis. For example, it has been concluded that stand delineation *"seems to work quite well"* and that *"it is possible to develop a stand-wise forest inventory method based on the satellite images"* (Tomppo 1987), and that segmentation based stand delineation with SPOT imagery followed by manual editing is comparable to results achieved with visual interpretation of aerial images (Hagner 1990). Further, the precision of stand-level estimates of stand volume and mean diameter was found to be comparable to the results of subjective field inventory (Hagner 1990).

In spite of the promising early results, the number of segment-based forestry applications remained low. There are two probable reasons for this. First, it was soon observed that the improved spatial resolution of the satellite imagery did not necessarily require contextual image analysis. Despite the preliminary doubts, the pixel-level analysis worked reasonably well with these imageries. Another reason for few reported applications was the lack of commercial image segmentation software.

The need for segmentation was perceived again in the 1990's. Increasing availability of digitized aerial photographs and the approach of a new generation of satellites providing VHR data re-stimulated the discussion concerning contextual image analysis. As a consequence of that the interest in object oriented image analysis has steadily increased, the pixel-by-pixel approach to image analysis has been increasingly criticized (e.g. Blaschke and Strobl 2001) and image segmentation has been tested in numerous VHR applications. Examples of these applications are delineation of habitats for biodiversity assessment (Holopainen 1998), delineation of individual tree crowns from aerial and other high spatial resolution imagery (Pitkänen 2001, Gougeon, 1995, Burnett 2003) and change detection (Pekkarinen and Sarvi 2002, Saksa et al. 2003). Even commercial services that are based on segmentation technology and aim at stand-level inventories are already available (FACT 2004). In addition to the optical imagery, image segmentation has been increasingly used to analyse airborne laser scanning (ALS) data. Examples of segment-aided ALS applications include the delineation of trees for change detection and growth estimation (Yu et al. 2003) and for the extraction of forest inventory parameters (Diedershagen et al. 2003). It seems, that image segmentation of RS imagery is experiencing its second renaissance.

5 Objectives

The main objective of this thesis has been to study whether the applicability of multi-source forest inventory (MSFI) methods can be improved using the image segment based approach. More specifically, the objective has been to study the applicability of segment-level analysis to the estimation of timber volume and stratification of forested areas. The particular objectives of sub-studies I-IV have been:

I To study the effect of segment-level feature extraction and image analysis on the accuracy of plot-level multi-source forest inventory estimates with help of Finnish National Forest Inventory field data and Landsat TM imagery. In addition, the selection of appropriate segmentation method has been addressed.

II To develop a method for the segmentation of very high spatial resolution imagery of forested landscapes and to evaluate its applicability to MSFI, specifically with respect to the estimation of timber volume.

III To introduce and test multi-scale segment based features in MSFI, specifically in the estimation of plot-level timber volume with help of field data and VHR imagery.

IV To evaluate an image-segment based approach to the stratification of forested area with help of aerial photographs.

6 Material

6.1 Field data

The sub-studies were carried out at four different test sites. The field data included four data sets two of which were gathered by the personnel of the National Forest Inventory (NFI) of Finland. The first of those sets included a subset of plots from 9th NFI (I) and the second set consisted of a dense grid of systematically sampled NFI-like field plots (II and III). The third and fourth field data sets (IV) consisted of circular and relascope field sample plots located in two study areas in Southern-Finland and measured by the Department of Forest Resource Management of the University of Helsinki.

6.2 Image material

The sub-studies of this thesis employ images from several different RS data sources. Sub-study I was based on an analysis of spaceborne imagery, namely two Landsat TM images. The thematic mapper (TM) is mounted on a satellite platform that orbits the earth at a nominal altitude of 705 kilometres. It sweeps the earth from west-to-east and east-to-west and collects data during both sweeps. It has seven bands, a quantization range of 8 bits and a spatial resolution of 30 (bands 1-5 and 7) and 120 meters (band 6) (Lillesand et al. 2004).

In sub-studies II and III, the imagery employed was acquired with a pre-series version of Airborne Imaging Spectrometer for Applications (AISA). AISA is a pushbroom type scanner recording radiation in the range 450 to 900 nm. The pre-series version of AISA has 286 spectral channels and the number of pixels per line is 384. The instrument is programmable and has four operating modes. The selectable parameters of AISA include the number of channels, wavelength and bandwidth of each channel, operating mode and integration time. The instantaneous field of view (IFOV) of the instrument is 1 milliradian and its dynamic range 2500 digital numbers. Across track pixel size of the instrument depends on the IFOV and flight height and along track pixel size on the velocity of the aeroplane and the integration time. For example, one meter pixel size is achieved with flight height of 1000 m, speed of 50 m/s and integration time of 20 ms (Mäkisara et al. 1993). The details of the AISA data employed here can be found in II and III. The first prototype of AISA was developed in the early 1990's, and currently the AISA family consists of three different systems: AISA+, AISA Eagle and AISA Hawk (SPECIM 2004).

In sub-study IV, the analysis was carried out using CIR aerial imagery. The images were obtained with a Wild RC30 camera, UAGA-F 13158 optics and

Kodak Aerochrome II Infrared Film 2243. The film characteristics curve is presented in figure 2. The antivignetting AV520 nm, and IR80% filters were used. The images were scanned using Zeiss Scai -scanner and 14 μm resolution and resampled to the pixel size of 0.5 metres.

All these data sources have different resolution characteristics that affect to their applicability in forest inventory applications. Note, that the term "resolution" refers to spatial, spectral or radiometric resolution (Lillesand et al. 2004). Spatial resolution describes the sensors capability to record spatial details, whereas spectral resolution determines the wavelength area to which the sensor is sensitive. The sensor's radiometric resolution determines the magnitude of the differences in the radiation that can be observed. In the case of aerial films, the radiometric resolution is usually described with help of the film characteristics curves (Lillesand et al. 2004). In real imaging systems, there is always a trade-off between these different types of resolution, and the choice of the appropriate sensor depends on the task to be conducted. In the following, only the differences in spectral and spatial resolution characteristics of the employed imagery are discussed.

The wavelength areas of the imagery employed in the sub-studies are presented in Table 1. From the standpoint of multi-source forest inventory the best performing sensors are TM and AISA. TM covers the widest range of spectrum and AISA is capable of dividing the spectrum into very narrow bands. This may be useful in the analysis of a phenomenon that can be observed only in a narrow range of the spectrum. Note, that in II, the estimation was carried out using the original 30 spectral AISA channels whereas in III these channels were generalised to four channels imitating the spectral characteristics of new generation VHR satellites (e.g., IKONOS). Detailed spectral characteristics of the employed AISA imagery are presented in II and III.

The drawback in both aerial AISA and CIR imagery is that the spectral sensitivity of the sensor (or film) is limited to the range of about 400 nm to about 900 nm. However, both of these data sources provide superior spatial resolution when compared to that of the TM sensor. The spatial resolution of channels of Landsat TM imagery employed is 30 meters. The corresponding figures with AISA and aerial imagery were 1.6 and 0.5 meters, respectively. In IV, however, the aerial imagery was resampled to a pixel size of 1.5 meters prior to the analysis

In addition to resolution characteristics, there are other factors that affect the applicability of remote sensing imagery to multi-source forest inventory. The radiance that a given remote sensing sensor observes is affected by the sun-object-sensor geometry, atmospheric attenuation, bidirectional reflectance and for non-lambertian surfaces also factors such as land cover or vegetation type (Leckie 1987). The magnitude with which each of these factors affects the observed radiance depends mainly on the view angle and imaging altitude. In satellite imagery, the main disturbing factors are usually scattering and absorption in the atmosphere (Song et al. 2001) and in low altitude imaging systems the sun-object-sensor geometry and bidirectional reflectance of the land surface (Pellikka et al. 2000).

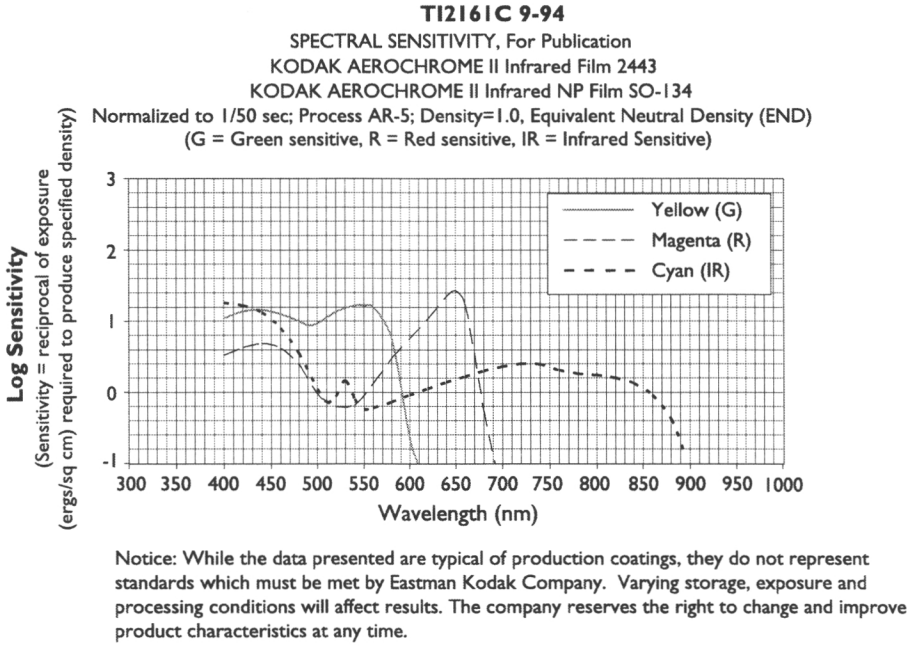


Figure 2. Spectral sensitivity curve of KODAK AEROCHROME II CIR film.

In spite of the fact, that the correction of atmospheric attenuation is often necessary prior to the classification and analysis of satellite EO imagery, that is not always the case. If the training data and the imagery are on the same relative scale, as in I, atmospheric correction has only minor effect on the image analysis results and is therefore unnecessary (Song et al. 2001).

The radiometric distortions in aerial imagery are often larger than in spaceborne material. The sun-object-sensor geometry and bidirectional reflectance effects cause radiometric distortions that may complicate the analysis of the imagery. The most important factors affecting the bidirectional reflectance of the forests include the hotspot effect and effects caused by mutual shadowing between trees, branches and leaves. The hotspot effect is observed when the viewing and illumination positions coincide, because the shadows are hidden behind illuminated objects and only bright object are registered on the scene (Li and Strahler 1992). In addition, the objects in the sun-side of the image appear darker, because only the shadowed part of the crowns and trunks are visible to the sensor. The phenomenon is more obvious at large viewing angles.

The AISA imagery employed in II and III consisted of data collected from seven flight lines. The analysis of the radiometric differences between the lines revealed, that adjustment of the pixel values was necessary. The adjustment was carried out using overlap areas of adjacent flight lines and cumulative histogram matching. In IV, the radiometric quality of the image mosaic was generally good. Similar objects had similar or close-to-similar properties in different parts of the

images and therefore the radiometric correction was considered unnecessary after the exclusion of one sub-area that had significantly different spectral properties than the rest of the area.

Table 1. Characteristics of the image material employed.

Image type	Channel	Sensitivity, nm
Satellite Landsat TM)	1	0.450 - 0.520
	2	0.520 - 0.600
	3	0.630 - 0.690
	4	0.760 - 0.900
	5	1.550 - 1.750
	6	10.400 - 12.500
	7	2.080 - 2.350
Imaging spectrometer (AISA)	1	0.470 - 0.477
	2	0.488 - 0.495
	3	0.505 - 0.512
	4	0.523 - 0.530
	5	0.542 - 5.488
	6	0.550 - 0.556
	7	0.565 - 0.572
	8	0.579 - 0.587
	9	0.600 - 0.607
	10	0.623 - 0.630
	11	0.647 - 0.654
	12	0.668 - 0.676
	13	0.676 - 0.683
	14	0.686 - 0.691
	15	0.697 - 0.701
	16	0.701 - 0.706
	17	0.711 - 0.715
	18	0.726 - 0.733
	19	0.733 - 0.738
	20	0.742 - 0.747
	21	0.750 - 0.755
	22	0.776 - 0.780
	23	0.785 - 0.790
	24	0.794 - 0.799
	25	0.803 - 0.808
	26	0.818 - 0.823
	27	0.844 - 0.849
	28	0.853 - 0.858
	29	0.861 - 0.866
	30	0.865 - 0.870
Aerial CIR photographs	See figure 2	

7 Methods

7.1 Image segmentation

In I, the Landsat TM imagery was segmented using two different approaches: a) measurement space guided clustering followed by connected component labelling (ISOCCL) and b) modified implementation (NG) of the "Image segmentation with directed trees" -algorithm (Narendra and Goldberg 1980). The resulting initial segmentations were fine-tuned using two different region merging (RM) algorithms that differ in the way they compute the similarity of adjacent segments. The similarity was determined using either the Euclidean distance between the segments or their t-ratio (see e.g. III, page 353). The RM algorithms were guided by a minimum segment size parameter that was set to 0.5 hectares, and in the case of t-ratio RM by a spectral similarity threshold.

In II, a new two-phase segmentation algorithm was developed. The method is based on the assumption that a VHR image of a forested area is a composition of certain spectral classes that have spatial relationships that can be modelled with their co-occurrence statistics. The developed method forms initial segments via clustering and CCL (as in I) and merges the initial regions to applicable segments with the help of a novel RM algorithm, namely co-occurrence region merging (CRM). The CRM is an iterative algorithm that combines each segment smaller than a given minimum size to a neighbouring segment with which it has the largest content co-occurrence. During the first iteration, the co-occurrence is determined using the original cluster labels of the segments. In subsequent iterations, the mode of the cluster labels of possibly merged segments is employed; see II for a more detailed description of the algorithm.

The applicability of the developed algorithm was tested in the segmentation of AISA imagery. Two different initial segmentations were derived, one based on clustering of the first three principal components (PCs) and row and column coordinates (S1), and one based on the use of the three first PCs only (S2). In the subsequential region merging phase the CRM -algorithm was guided by a minimum segment size parameter that was set to 0.03 hectares.

Paper III describes the implementation of a segmentation algorithm (INISEG) that is based on the ideas of Narendra and Goldberg (1980). The algorithm is used as an initial segmentation method in a two-phase segmentation process that aims to delineation of feature extraction and image analysis units for VHR imagery. The initial segments produced by the algorithm were further processed using a region-merging algorithm that compares the spectral similarity of adjacent segments with the aid of their mutual t-ratio. The two-phase segmentation algorithm was tested in the segmentation of AISA imagery that was spectrally generalized to correspond to the spectral properties of new generation VHR satellite images. The minimum segment size was set to 10 pixels and three

Table 2. Summary of applied image segmentation and region-merging algorithms.

Initial segmentation algorithms	
Algorithm	Employed in sub-studies
Modified implementations of "Image segmentation with directed trees": NG and INISEG	I, III, IV
Measurement space guided clustering followed by connected component labelling (CCL)	I, II
Region merging (RM) algorithms	
Algorithm	Employed in sub-studies
t-ratio RM	I, III
Co-occurrence RM	II
Euclidean distance RM	I, IV

different segmentations (SA, SB and SC) were derived using t-ratio threshold values 0, 24 and 40.

In IV, the initial segmentation of the aerial imagery employed was derived using the INISEG algorithm described in III. The initial segments were merged to larger entities using RM algorithm that was guided by a minimum segment size parameter. All segments smaller than the given minimum size were merged to the most similar neighbouring segment. The similarity of the segments was determined by means of their Euclidean distance in the spectral feature space. Two different minimum sizes, 380.25 and 675 m^2, were tested.

A summary of the developed, implemented and tested segmentation algorithms is presented in Table 2.

7.2 Feature extraction and selection

In I, the Landsat TM spectral average features were extracted in two different ways: from square shaped windows surrounding the plots (reference features) and from those pixels within the windows that belonged to the same segments as the plot pixel (segment-restricted features). Window sizes from 1 x 1 (0.06 ha) to 11 x 11 pixels (7.56 ha) were tested. A basically similar approach was applied in II, but due to the better spatial resolution of the AISA imagery, a larger number of different window sizes was tested. Both, reference and segment-restricted features were extracted from windows of sizes from 3 x 3 pixels (about 0.002 hectares) to 121 x 121 pixels (about 3.75 hectares). In II, the extracted spectral feature set consisted of sub-optimal features found in an earlier AISA study (Mäkisara et al. 1997).

The objective in III was to test the performance of multi-scale segment based features. The analysis was therefore extended to include tests of segment-level features that were derived using all pixels within the segments in addition to reference and segment-restricted features. Segment-restricted and reference features were extracted from square shaped windows of 31 x 31 pixels (about 0.25 hectares). The features extracted included spectral averages and standard deviations computed from the spectrally generalized AISA imagery. Similar segment-level and reference features were extracted and employed in IV, but in the case of reference features, the size of extraction window was set to about 20 x 20 m^2, which corresponds to the size recommended in Holopainen and Wang (1998).

In III, the different number of reference and segment-based spectral features complicated the comparison of their performance. A subset of five best-performing features from both datasets was therefore selected for estimation tests. The feature selection was carried out using a sequential forward selection algorithm. It started by selecting the feature giving the lowest RMSE and proceeded by adding the feature that gave the best performance with the already selected features. The five best features were chosen from both reference and segment-based datasets for the evaluation tests.

7.3 Evaluation of segment-based approaches to estimation and stratification

7.3.1 General

The applicability of the segment-aided approach to feature extraction and image analysis was evaluated by employing segment-based spectral features, namely spectral averages (I-IV) and standard deviations (II-IV) and their sub-optimal combination (II), in the estimation of plot-level timber volume (I-III) and stratification of an inventory area (IV). The performance of segment-based spectral features was compared to that of reference features that were extracted in a more straightforward manner. These methods were selected for evaluation of segment-based approaches because they provide a more objective basis for comparisons than, for example, visual analysis of segmentation result. The evaluation of different segmentation results (I and II) could have been based on GIS analysis of the location of the segments borders, shape of segments and other segment properties, but because there is no analytical way to determine the "correct" segmentation, indirect analysis was needed. An alternative way to compare the segmentation results would have been the analysis of the within and between segment variances of stand characteristics with help of dense grid of field plots (Hagner 1990).

7.3.2 Estimation of timber volume

The estimation tests (I-III) were carried out using field sample plot data and extracted spectral features. The estimates for the plot level timber volumes, i.e. total volume and volumes by tree species, were derived using an inverse distance weighted non-parametric k -nearest neighbour estimator (k-NN, e.g. III equation 4) (e.g. Tokola et al. 1996, Tomppo 1996, Franco-Lopez et al. 2001) and a leave-one-out cross-validation technique. In cross-validation, every field plot was in turn omitted from the dataset and its characteristics were predicted with the aid of the other plots. The number of employed nearest neighbours was chosen separately for each study and was 5 in III and 10 in I and II. The performance of reference and segment-based features in the estimation was compared on the basis of root mean square error (RMSE) and relative RMSE (I-III), and empirical bias (III). In sub-study III, the analysis was extended to volume classes with the help of confusion matrices and their user's, producer's and overall accuracy (Stehman 1997).

7.3.3 Stratification of forest area

In IV, the evaluation of the segment-based approach to stratification was based on an analysis of within-strata variation in the spectral information and forest attributes. The stratification was conducted for segment-based and reference data using extracted spectral average and standard deviation features and a k-means algorithm (MacQueen 1967). Different strata numbers (20 - 50) were tested and the homogeneity of each stratum was characterized with the aid of area weighted mean standard deviations of forest attributes.

8 Results

8.1 Estimation tests

8.1.1 Landsat TM imagery (I)

The sensitivity of the traditional pixel-by-pixel image analysis methods to locational errors has been one reason for high plot- and stand-level estimation errors of MSFI applications employing satellite imagery. One possibility to diminish the effect of locational errors in the plot-based training data is to extract the spectral information for each plot not only from the plot pixel, but also from its spatial neighbourhood. This task can be conducted with the aid of image segments. The approach was tested with Finnish NFI data and Landsat TM imagery. Four different segmentations were derived using two different initial segmentation (NG and ISOCCL) and region merging (NN and TR) algorithms.

In general, segment based features gave lower estimation errors than features extracted from square-shaped windows. In the case of the estimation of total volume, the features derived using the combination of ISOCCL and TR gave the best estimation result. Best estimates for pine and spruce volumes were achieved with features derived using the combination of ISOCCL and NN, and best estimates for the volume of broadleaved species with features extracted from the image segmented using a combination of NG and TR. The differences in the performances of segment-based features and features extracted from square-shaped windows were, however, insignificant.

The clustering phase of ISOCCL is computationally demanding. In addition, it produced about 2.5 times higher number of initial segments than the NG. The NG segments also corresponded better to forest stand structure. Consequently, NG was judged to be the better choice as the initial segmentation method for forest inventory applications.

8.1.2 AISA imagery (II and III)

In II, a new two-phase algorithm that was designed for the segmentation of VHR images was developed and tested in the segmentation of AISA imagery. Two different segmentations (S1 and S2) were derived. The input of S1 consisted of the first three principal components (PCs) of the AISA imagery, and row and column coordinates. S2 was derived on the basis of the three first PC's only.

The visual analysis of the segmentation results showed clear differences between S1 and S2. The use of row and column coordinates as additional input channels resulted in improved recognition of spectrally homogeneous areas, such as recent clear cuttings, but at the same time some relevant segment borders were lost. The exclusion of the coordinate channels from the analysis resulted in a more

distinct segment structure. The drawback of S2 was that even spectrally very homogeneous areas were in many cases divided into several segments.

The estimation results showed only minor differences in the performances of the segment-based and reference feature sets. Segment-based features extracted using S1 gave slightly smaller RMSEs for all estimated variables than features extracted using S2. In the case of estimates of total volume and the volume of deciduous species, S1 features gave the best results, whereas the lowest RMSEs for the estimates of pine and spruce volumes were achieved with the reference feature set. S1 features were less sensitive to the change in the size of the extraction window. The differences in the RMSEs were insignificant in most of the cases.

One means of ensuring that the extracted spectral information is representative at the field sample plot level, is to extract multi-scale features and select the best performing features for the actual analysis. That approach was tested in III. The spectrally generalized AISA imagery was segmented with a two-phase segmentation method. Initial segmentation was carried out with the INISEG algorithm and the resulting output was further processed with a t-ratio-based region merging algorithm with three different similarity thresholds.

The segmentation algorithm performed well and produced segments that were visually appealing. In addition, the results of the estimation tests showed that segment based features (SF) performed better than reference features (RF) in the estimation of all tested variables except in the estimation of pine volume. The differences in the performances of the feature sets were, however, small. Even though SF performed better than RF, the evaluation of confusion matrices and user's (UA) and producer's (PA) accuracies revealed problems especially in the estimation of high timber volumes.

8.2 Stratification test (IV)

In Finland, the data for forest management planning has been traditionally gathered by stand-level visual field inventories. The inventory method has been criticized because of its subjectivity, and the introduction of more objective methods have been suggested. One of the alternatives suggested is a two-phase sampling scheme (e.g. Poso and Waite 1996). The problem with the suggested approach is that it is difficult to determine the appropriate density for the first phase sample. Paper IV suggests that this problem could be solved by replacing the two-phase scheme with a segment-based approach. Stratification based on clustering of spectrally homogeneous segments is assumed to result in strata that are more homogeneous in their spectral and forest characteristics than strata derived from the two-phase sampling approach.

The results were controversial. In both study areas (S1 and S2) the distributions of the extracted spectral values proved that segment-level features retain better the original spectral variation of the images than features extracted from square

shaped windows surrounding the first phase sample plots. This observation was evident in both study areas and holds true for both the minimum segment sizes employed. In spite of these results, the within-strata variation of forest attributes did not reveal clear difference between the approaches. In general, the results for S1 imply that when forest characteristics are considered the stratification based on a two-phase sampling scheme results in more homogeneous clusters than the segment-based approach. In S2, the results were partly reversed. The spectral features employed in the clustering phase have, however, a relatively large effect on the results. In S2, for example, the stratification based on merely spectral averages showed that segment based stratifications produce more homogeneous strata than the two-phase sampling approach. The inclusion of standard deviation features into the clustering process, however, implied that the two-phase sampling strategy would result in better strata.

9 Discussion and conclusions

The sub-studies of this thesis developed and tested several different approaches to image segmentation and their MSFI applications. The developed and implemented algorithms were tested in the estimation of plot-level timber volumes (I-III) and in the stratification of forested areas (IV). The incorporation of image segment- and sub-segment features into the estimation procedure improved the plot level volume estimates in most of the cases (I-III). The achieved reductions in the RMSEs were, however, smaller than expected and it is questionable if these small improvements can be used as basis for the recommendation of segment-level estimation approach. Better results were achieved in the stratification of forested areas, even though some of the results were controversial. In spite of that, the segment-aided approach can be recommended for stratification purposes.

The finding that the segment-based approach did not significantly improve the estimation results may be due to many reasons. First, the type of field data employed in the estimation studies (I-III) is sensitive to locational accuracy and may be unrepresentative for their neighbourhood (Koivuniemi 2003). The suitability of the employed data to segment-level analysis is therefore questionable. Furthermore, the field data employed in II and III had been pre-processed in such a way that the studies may give an over-optimistic impression concerning the performance of window-based feature extraction approaches that were used as benchmarks in the evaluation of segment-based results.

In general, the results of remote sensing based estimation and stratification applications are largely dependent on the spectral and radiometric properties of the imagery employed. Atmospheric attenuation, sun-object-sensor geometry and bidirectional reflectance effects may cause radiometric distortions that hinder the remote sensing -based analysis of phenomena that manifest themselves in slight spectral changes. The effect of these factors is generally larger in data that has been acquired from low-altitudes and using wide-angle lenses (Pellikka et al. 2000). From the viewpoint of this thesis, however, these factors are of minor importance. Even though these factors may have affected the absolute estimation errors (I-III) and within strata variations (IV), it is unlikely that they affect the mutual relationship of results of segment and non-segment based analysis on which the judgement of the segment-based approach was based.

All the tested and developed segmentation algorithms were applicable to the determination of feature extraction and image analysis units. However, bearing in mind that the selection of appropriate segmentation parameters is often based on trial-and-error, segmentation methods in which the creation of initial segments is based on computationally intensive iterative clustering algorithms may be too time consuming, at least from an operational point of view. In addition, initial segmentations created with these algorithms and VHR imagery usually produce

a large number of initial segments that are complex in shape and do not necessarily correspond well with the stand structure. The complexity of the initial segments can, at least in some cases, be diminished by introducing spatially varying components (e.g. image coordinates) to the clustering process. However, that is not advisable in general, because some of the relevant borders may be lost and the use of other than spectral information complicates the interpretation of the segmentation result. Most of these problems can be avoided if initial segments are created using algorithms that are based solely on local image properties (see III).

Region merging algorithms that can be guided with minimum segment size and similarity parameters provide a meaningful way to aggregate small, potentially irrelevant regions to meaningful spatial entities that are, in most cases, more applicable in MSFI analysis than the initial areas. Where there exists areas that are small, but spectrally separable and relevant, the similarity parameter can be used in such a way that it prohibits the merging of two segments that are spectrally very dissimilar. In a forestry context, this kind of functionality might be needed in, for example, the separation of small underproductive rocky areas from the surrounding productive forest land.

Even though image segmentation is often referred as an objective way to isolate and determine spatial units for image analysis, segmentation result may be very sensitive to the user-defined parameters. The selection of various spectral or other characteristics parameters, such as input channels and minimum segments size, is often based on subjective decisions that may have a drastic influence on the final segmentation result. In addition, image segmentation is highly sensitive to the quality of image material. Special attention has to be paid to segmentation of imagery acquired from low altitudes and using wide-angle lenses because they typically include a large proportion of shadows. In some forestry applications, such as automated stand delineation, these shadows may cause serious locational errors in the final product. Even with its drawbacks, however, image segmentation is a tool that should not be bypassed when considering alternatives for VHR image analysis.

The algorithms developed here were implemented in such a way that they aim at a high level of automation. They are therefore guided by few parameters and the user has practically no other control over the segmentation procedure. Even though the implemented software has been successfully applied in this thesis and for many purposes not reported here (e.g. Pekkarinen and Sarvi 2002, Sell 2002, Saksa et al. 2003, Tuominen and Pekkarinen 2004), the highly automated approach is probably not the best alternative for all forestry applications. If segmentation is used in tasks that require a knowledge-based interpretation, such as semi-automatic stand delineation, a different approach is recommendable. For such purposes the algorithms should be implemented in such a way that they allow better interaction and thus the incorporation of superior image analysis capabilities of human beings into computer-aided image processing systems.

References

Anttila, P. 2002. Nonparametric estimation of stand volume using spectral and spatial features of aerial photographs and old inventory data. Canadian Journal of Forest Research 32: 1849-1857.

Baatz, M., Benz, U., Dehghani, S., Heynen, M., Höltje, A., Hofmann, P., Lingenfelder, I., Mimler, M., Sohlbach, M., Weber, M. & Willhauck, G. 2002. eCognition - User Guide 3.

Bauer, M.E., Burk, T.E., Ek, A.R., Coppin, P.R., Lime, S.D., Walsh, T.A. & Walters, D.K. 1994. Satellite inventory of Minnesota forest resources. Photogrammetric Engineering and Remote Sensing 60: 287-298.

Blaschke, T. & Strobl, J. 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. GeoBIT/GIS 6: 12-17.

Burnett, C. and Blaschke, T. 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. Ecological modelling 168: 233-249.

Canny, J. 1986. Computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence PAMI-8(6): 679-898.

Gougeon, F.A. 1995. A system for individual tree crown classification of conifer stands at high spatial resolution. pp. 635-642 In: Proceedings of 17th Canadian Symposium of Remote Sensing, Saskatoon, Saskatchewan, Canada, June 13-15, 1995.

Definiens 2003. http://www.definiens-imaging.com/ecognition/forester/.

Diedershagen, O., Koch, B., Weinacker, H. & Schutt, C. 2003. Combining Lidar- and GIS data for the extraction of forest inventory parameters. In: ScandLaser Scientific Workshop on Airborne Laser Scanning of Forests. Umeå, Sweden. Swedish University of Agricultural Sciences, Department of Forest Resource Management and Geomatics.Working paper 112. 273 p.

Dong, J., Kaufmann, R.K., Myneni, R.B., Tucker, C.J., Kauppi, P.E., Liski, J., Buermann, W., Alexeyev, V. & Hughes, M.K. 2002. Remote sensing estimates of boreal and temperate forest woody biomass: carbon pools, sources and sinks. Remote Sensing of Environment 84(3): 393-410.

ERDAS Inc. 1994. ERDAS FIELD GUIDE, Third Edition. ERDAS, Inc. Atlanta, GA.

FACT 2004. http://www.falconinformatics.com/

Franco-Lopez, H., Ek, A.R. & Bauer, M.E. 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbour method. Remote Sensing of Environment 77: 251-274.

Franklin, S.E., Wulder, M.A. & Lavigne, M.B. 1996. Automated derivation of geographic window sizes for use in remote sensing digital image texture analysis. Computers & Geosciences 22: 665-673.

Fransson J.E.S., Walter F., and Ulander L.M.H., 2000. Estimation of Forest Parameters Using CARABAS-II VHF SAR Data, IEEE Transactions on Geoscience and Remote Sensing, 38: 720-727

Fu, K.S. & Mui, J.K. 1981. A survey of image segmentation. Pattern Recognition 13(1): 3-16.

Gonzales, R. C. and Woods, R. E. 1993. Digital Image Processing. Addison-Wesley Publishing Company. USA.

Goldberg, M., Schlaps, D., Alvo, M. & Karam, G. 1982. Monitoring and change detection with Landsat imagery. Proceedings of 6th International Conference on Pattern Recognition, Munich, Germany, Oct. 19-22, 1982 1: 523-526.

Haapanen, R., Ek, A.R., Bauer, M.E. & Finley, A.O. 2004. Delineation of forest/nonforest land use classes using nearest neighbor methods. Remote Sensing of Environment 89: 265-271.

Hagner, O. 1987. Remote sensing-aided forest inventory. Helsingin yliopiston metsänarvioimistieteen laitoksen tiedonantoja 19: 130-139.

—, 1990. Computer aided forest stand delineation and inventory based on satellite remote sensing. The usability of remote sensing for forest inventory and planning. Proceedings from SNS/IUFRO workshop in Umeå 26-28 February 1990. Swedish University of Agricultural Sciences, Remote Sensing Laboratory, Umeå. pp. 94-105 .

Halme, M. & Tomppo, E. 2001. Improving the accuracy of multisource forest inventory estimates by reducing plot location error - a multicriteria approach. Remote Sensing of Environment 78: 321-327.

Häme, T. 1991. Spectral interpretation of changes in forest using satellite scanner images. Acta Forestalia Fennica 222: 1-11.

—, Heiler, I. & Miguel-Ayanz, J.S. 1998. An unsupervised change detection and recognition system for forestry. International Journal of Remote Sensing 19(6): 1079-1099.

Haralick, R.M. & Shapiro, L.G. 1985. Survey: Image Segmentation Techniques. Computer Vision, Graphics, and Image Processing 29(1): 100-132.

Holmström, H., Nilsson, M. & Ståhl, G. 2001. Simultaneous estimations of forest parameters using aerial photograph interpreted data and the *k* nearest neighbour method. Scandinavian Journal of Forest Research 16: 67-78.

Holopainen, M. 1998. Forest habitat mapping by means of digitized aerial photographs and multispectral airbone measurements. University of Helsinki, Department of Forest Resource Management. Publications 18: 1-49.

— & Wang, G. 1998. The calibration of digitized aerial photographs for forest stratification. International Journal of Remote Sensing 19(4): 677-696.

Hyppänen, H. 1996. Spatial autocorrelation and optimal spatial resolution of optical remote sensing data in boreal forest environment. International Journal of Remote Sensing 17: 3441-3452.

Hyyppä, J., Hyyppä, H., Inkinen, M., Engdahl, M., Linko, S. & Zhu, Y.-H. 2000. Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. Forest Ecology and Management 128: 109-120.

—, Hyyppä, H., Samberg, A., 1999, Assessing Forest Stand Attributes by Laser Scanner, Laser Radar Technology and Applications IV, 3707: 57-69.

ILMARI 2004. www.ilmari.fi

Jain, R., Kasturi, R. & Schunck, B.G. 1995. Machine vision. McGraw-Hill International Edition.

Katila, M. 2004. Error variations at the pixel level in the k-nearest neighbour estimates of the Finnish multisource National Forest Inventory. In: Controlling the estimation errors in the Finnish multisource National Forest Inventory. The Finnish Forest Research Institute, Research Papers, 910.

Kilpeläinen, P. & Tokola, T. 1999. Gain to be achieved from stand delineation in Landsat TM image-based estimates of stand volume. Forest Ecology and Management 124: 105-111.

Koivuniemi, J. 2003. The accuracy of compartmentwise forest inventory based on stands and located sample plots. Faculty of Agriculture and Forestry. Department of forest resource management. University of Helsinki, Helsinki. 143 p.

Kuusela, K. & Poso, S. 1970. Satellite pictures in the estimation of the growing stock over extensive areas. Photogrammetric Journal of Finland 4(1).

Leckie, D. G. 1987. Factors affecting defoliation assessment using airborne multispectral scanner data. Photogrammetric Engineering and Remote Sensing (53)12: 1665-1674.

Lillesand, T.M., Kiefer, R.W. & Chipman, J. W. 2004. Remote Sensing and Image Interpretation. John Wiley and Sons Inc.

Li, X. and Strahler, A. H. 1992. Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: effect of crown shape and mutual shadowing. IEEE Transactions on Geoscience and Remote Sensing 30(2): 276-292.

Leica Geosystems 2004. http://gis.leica-geosystems.com/products/

Löfström, K. 1946. Ilmakuvakartoitus Suomessa. The Photogrammetric Journal of Finland 1(1): 78-109.

MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. Volume 1 of Proceedings of the Fifth Berkeley Symposium on Mathematical statistics and probability, pp 281-297. Berkeley, 1967. University of California Press.

Merriam -Webster Online Dictionary 2004. "segment". www.merriamwebster.com

METRIA 2004. www.lantmateriet.se.

Mäkisara, K., Heikkinen, J., Henttonen, H., Tuomainen, T. & Tomppo, E. 1997. Experiment with imaging spectrometer data in large-area forest inventory context. Proceedings of the Third International Airbone Remote Sensing Conference and Exhibition. Development, Integration, Applications & Operations 7-10 July 1997, Copenhagen, Denmark II: 420-427.

—, Meinander, M., Rantasuo, M., Okkonen, J., Aikio, M., Sipola, K., Pylkkö, P. & Braam, B. (eds.). 1993. Airborne Imaging Spectrometer for Applications (AISA). 1993 International Geoscience and Remote Sensing Symposium (IGARSS'93). Tokyo, Japan. pp 479-481.

Narendra, P.M. & Goldberg, M. 1980. Image segmentation with directed trees. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-2: 185-191.

Næsset, E. 1997. Estimating timber volume of forest stands using airborne laser scanner data. Remote Sensing of Environment, 61(2): 246-253.

Nieke, J., H., S., Neumann, A. & Zimmermann, G. 1997. Imaging Spaceborne and Airborne Sensor Systems in the Beginning of the Next Century. The European Symposium on Aerospace Remote Sensing (IEE); Conference on Sensors, Systems and Next Generation Satellites III. London, UK. SPIE3221-71p.

Nyyssönen, A. 1955. On the estimation of the growing stock from aerial photographs. Communicationes Instituti Forestalis Fenniae 46: 1-57.

—, 1962. Aerial photographs of tropical forests. Unasylva 16(1).

Olsson, H. 1994. Monitoring of Local Reflectance Changes in Boreal Forests using Satellite Data. PhD thesis. Research report 8. Department of Forest Resource Management and Geomatics. Swedish University of Agricultural Sciences, Umeå.

Pal, N.R. & Pal, S.K. 1993. A review on image segmentation techniques. Pattern recognition 26(9): 1277-1294.

Parmes, E. 1992. Segmentation of SPOT and Landsat satellite imagery. The Photogrammetric Journal of Finland 13(1): 52-58.

Pellikka, P., King, D. J. & Leblanc, S. G. 2000. Quantification and removal of bidirectional effects in aerial CIR imagery of deciduous forest using two reference land surface types. Remote Sensing Reviews, Special issue on "Multi-angle Measurements and Models", 19: 259-291.

Pekkarinen, A. & Sarvi, V. 2002. Detection of clearcuttings with help of high-altitude panchromatic aerial photographs and image segmentation. In: Operational Tools in Forestry using Remote Sensing Techniques. ForestSAT Symposium. Edinburgh, Scotland, August 5th-9th 2002.CD-ROM.

Pitkänen, J. 2001. Individual tree detection in digital aerial images by combining locally adaptive binarization and local maxima methods. Canadian Journal of Forest Research 31: 832-844.

Poso, S. & Waite, M.-L. 1996. Sample based forest inventory and monitoring using remote sensing. Remote sensing and computer technology for natural resource assessment. Joensuu. Research notes of University of Joensuu 48.

Saksa, T., Uuttera, J., Kolström, T., Lehikoinen, M., Pekkarinen, A. & Sarvi, V. 2003. Clear cut detection in boreal forest aided by remote sensing. Scandinavian Journal of Forest Research 18(6): 537-546.
Sarvas, R. 1938. Ilmavalokuvauksen merkityksestä metsätaloudessamme. Silva Fennica 48: 2-45.
Saukkola, P. 1982. Monitoring regeneration fellings by satellite imagery. Technical Research Centre of Finland - Reports 89: 1-108.
Schieve, J., Tufte, L. & Ehlers, M. 2001. Potential and problems of multi-scale segmentation methods in remote sensing. GeoBIT/GIS 6: 34-39.
Sell, R. 2002. Segmentointimentelmien käyttökelpoisuus ennakkokuvioinnissa. Metsätieteen aikakauskirja 3: 499-507.
Song, C., Woodcock, C.E., Seto, K.C., Pax Lenney, M., and Macomber, S.A. 2001. Classification and change detection using Landsat TM data: when and how to correct atmospheric effects? Remote Sensing of Environment 75:230-244.
SPECIM 2004. www.specim.fi
Stehman, S. V. 1997. Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment 62: 77-89.
Strahler, A.H., Woodcock, C.E. & Smith, J.A. 1986. On the nature of models in remote sensing. Remote Sensing of Environment 20: 121-139.
Tokola, T., Pitkänen, J., Partinen, S. & Muinonen, E. 1996. Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. International Journal of Remote Sensing 17: 2333-2351.
Tomppo, E. 1987. Stand delineation and estimation of stand variates by means of satellite images. Remote Sensing-Aided Forest Inventory. University of Helsinki, Department of Forest Mensuration and Management Research Notes 19. 60-76 p.
—, 1992a. Multi-source national forest inventory of Finland. Metsäntutkimuslaitoksen tiedonantoja 444: 52-60.
—, 1992b. Satellite image aided forest site fertility estimation for forest income taxation. Acta Forestalia Fennica 229: 1-70.
—, 1996. Multi-source national forest inventory of Finland. New Thrusts in forest inventory, EFI proceedings 7: 27-41.
—, Goulding, C. & Katila, M. 1999. Adapting Finnish multi-source forest inventory techniques to the New Zealand preharvest inventory. Scandinavian Journal of Forest Research 14: 182-192.
—, Korhonen, K.T., Heikkinen, J. & Yli-Kojola, H. 2001. Multi-source inventory of the forests of the Hebei Forestry Bureau, Heilongjiang, China. Silva Fennica 35: 309-328.
—, Nilsson, M., Rosengren, M., Aalto, P. & Kennedy, P. 2002. Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass. Remote Sensing of Environment 82: 156-171.
Trotter, C.M., Dymond, J.R. & Goulding, C.J. 1997. Estimation of timber volume in a coniferous plantation forest using Landsat TM. International Journal of Remote Sensing 18: 2209-2223.
Tuominen, S. & Pekkarinen, A. 2004. Local radiometric correction of digital aerial photographs for multi source forest inventory. Remote Sensing of Environment 89: 72-82.
— & Poso, T. 2001. Improving multi-source forest inventory by weighting auxiliary data sources. Silva Fennica 35: 203-214.
Varjo, J. 1996. Detecting manmade forest activities and natural disasters using Landsat TM satellite data - a method presented for controlling continuously updated forest information in Finland. Presented in joint meeting of the Council on Forest Engineering and International Union of Forest Research Organizations Subject Group S3.04.00, Marquette, MI, July 29-August 1 1996: 1-9.
Weszka, J.S. 1978. A survey of threshold selection techniques. Computer Graphics and Image Processing 5: 382-399.

Woodcock, C.E. & Macomber, S. 2001. Monitoring large areas for forest change using Landsat: Generalization across space, time and Landsat sensors. Remote Sensing of Environment 78: 194-203.

—, Macomber, S.A., Pax-Lenney, M. & Cohen, W.B. 2001. Monitoring large areas for forest change using Landsat: generalization across space, time and Landsat sensors. Remote Sensing of Environment 78: 194-203.

— & Strahler, A.H. 1987. The factor of scale in remote sensing. Remote Sensing of Environment 21: 311-332.

Yu, X.W., Hyyppä, J., Rönnholm, P., Kaartinen, H., Maltamo, M. & Hyyppä, H. 2003. Detection of harvested trees and estimation of forest growth using laser scanning. In: ScandLaser Scientific Workshop on Airborne Laser Scanning of Forests. Umeå, Sweden. Swedish University of Agricultural Sciences, Department of Forest Resource Management and Geomatics. Working paper 112. 273 p.

Z/I Imaging 2004. http://www.ziimaging.com/

Zucker, S., W. 1976. Region Growing: Childhood and Adolescence. Computer Graphics and Image Processing 5: 382-399.

Paper I

I

Reprinted from Remote Sensing of Environment, vol 77, Mäkelä, H. & Pekkarinen, A., Estimation of timber volume at the sample plot level by means of image segmentation and Landsat TM imagery, pp 66-75, copyright (2001), with permission from Elsevier.

ELSEVIER

Remote Sensing of Environment 77 (2001) 66–75

Remote Sensing of Environment

www.elsevier.com/locate/rse

Estimation of timber volume at the sample plot level by means of image segmentation and Landsat TM imagery

Helena Mäkelä, Anssi Pekkarinen*

Finnish Forest Research Institute, Unionikatu 40 A, FIN-00170 Helsinki, Finland

Received 16 July 2000; received in revised form 25 January 2001; accepted 27 January 2001

Abstract

The use of image segments in the feature extraction for the estimation of timber volumes using a Landsat TM image was investigated by applying the *k* nearest neighbour estimation method (*k*nn) and Finnish National Forest Inventory (NFI) sample plots. The estimates of the volumes by tree species at the plot level were derived by means of the cross-validation technique. Ten nearest neighbours (NNs) were applied in the estimation. Image segments were derived by two different methods: (1) a measurement space-guided clustering followed by the connected component labeling (ISOCCL) and (2) a directed trees algorithm (NG). The segmentations were fine-tuned by means of two different region-merging algorithms. The spectral features were extracted in two ways: from a fixed window (FW) around the field sample plot, and from those pixels within the FW that belonged to the same segment as the sample plot pixel. Window sizes from 1 to 11 × 11 pixels were tested, and the average of the extracted pixel values was used in the estimation. Features from the ISOCCL-based segments gave the best estimates for the volumes of pine and spruce, as well as for the total volume. Best estimates for the volume of broad-leaved trees were obtained from NG-based segments. Compared to the estimates of the FW approach, the improvements were, however, quite small and relative root mean square errors (RMSEs) remained high. The minimum and maximum improvements of relative RMSEs were 1% and 11.3%, respectively. The NG was considered a more applicable segmentation method for forest inventory purposes at the stand level, even though the ISOCCL gave slightly better estimation results in this study. The use of image segmentation in the stratification of the image material into stand margin and within stand areas could be more suitable for the estimation of forest variables. This is the case especially if only the plot-level field information is available.

1. Introduction

Remote sensing is a cost efficient source of information for large-area forest inventories, and satellite image data has been widely applied in various inventory tasks. One of the first operative multisource inventory methods was introduced by the National Forest Inventory (NFI) of Finland (Tomppo, 1993). This multisource NFI combines field measurements of sample plots, digital image information (mostly Landsat TM), and numerical map and elevation data by means of nonparametric image analysis. The information of the field sample plots located in the area of an image are generalised over the whole image area by applying the spectral properties of the field sample plots and image pixels (Kilkki & Päivinen, 1987). In the multisource NFI, this has been implemented by means of the *k* nearest neighbour estimation method (*k*nn) (Keller, Gray, & Givens, 1985; Tomppo, 1990).

Although the multisource forest inventory methods have proven to produce reliable forest resource information for large- and medium-size areas, their accuracy at the forest stand-level is not sufficient for forest management purposes (Holmgren & Thuresson, 1998; Kilpeläinen & Tokola, 1998; Tokola & Heikkilä, 1997; Trotter & Dymond, 1997). This is due to the errors in image registration and the location of sample plots, as well as the limited spatial and spectral resolution of the satellite image data. The combination of these errors may easily lead to large estimation errors at the pixel level because erroneous spectral features, e.g., features of adjacent pixels, may have been assigned to some field plots of the training data.

One commonly applied solution to this kind of problem is to apply some low-pass filtering and use the average spectral properties of the immediate spatial neighbourhood of a pixel (Trotter & Dymond, 1997). This reduces the effect of the errors mentioned above, but at the same time

* Corresponding author. Fax: +358-9-625-308.
E-mail address: anssi.pekkarinen@metla.fi (A. Pekkarinen).

0034-4257/01/$ – see front matter
PII: S0034-4257(01)00194-8

some details of the image information are lost and the spectral features of the informational classes are shifted towards the mean. Thus, small objects with extreme spectral properties may become impossible to recognise. This can be avoided if the extraction of spectral features is carried out by units, which are homogeneous both in the sense of their spectral and forest characteristics. In the context of forestry, these units could result, for example, from forest management planning.

In Finland, the inventory data for forest management planning purposes is collected by means of an ocular stand-level field inventory. Forest stands are usually delineated by means of visual interpretation of paper copies of colour-infrared aerial photographs in the scales of 1:10,000–1:30,000. The use of these areas as feature extraction units is, however, somewhat doubtful. Because the purpose of the stand delineation is to define operational units for silviculture treatments and harvesting, a part of a stand or combination of several small and even heterogeneous stands may be combined to a same unit. In addition to this, the delineation of these units is subjective, and therefore the result may be different between foresters (Poso, 1983). Furthermore, forest management planning is typically carried out every 10 years, and thus the delineation may be out of date and its positional accuracy poor (Mäkinen, 1999). Finally, forest management planning does not cover all forested areas. For these reasons, a more objective and flexible way to define the units of feature extraction is needed. This study suggests the use of image segments for that purpose.

Image segmentation is the division of an image into spatially continuous and homogeneous regions. Segmentation methods can be classified into three groups: pixel-, edge-, and region-based methods. Pixel-based methods include thresholding and clustering in the measurement space. The result of these methods is turned into segmentation by identifying spatially continuous regions by means of connected component labeling (CCL). In edge-based methods, the edges between the segments are detected and linked into contour chains. Segments are defined as regions inside these contours. Region-based methods include region growing, merging, splitting, and their combinations. In region growing approaches, adjacent pixels that do not differ more than a given criterion are assigned to the same segment. In the region merging and splitting approaches, areas are combined or divided based on their spectral similarities and homogeneity. Reviews of segmentation methods have been presented by Fu and Mui (1981), Haralick and Shapiro (1985), and Pal and Pal (1993).

Even though image segmentation is a widely used image processing technique, its reported applications to forestry are few. In Scandinavia, the method "Image segmentation with directed trees" presented by Narendra and Goldberg (1980) has been tested in forest inventory applications (Fransson, Walter, & Olsson, 1999; Hagner, 1990; Häme, 1991; Kilpeläinen & Tokola, 1998; Parmes, 1992; Tokola, 1990; Tomppo, 1987, 1992). Tomppo (1987) concluded that it is possible to develop a stand-level inventory application based on this method and satellite images. A modified implementation of the method for large images was presented later (Parmes, 1992). Tokola (1990) applied segments instead of forest stands in order to predict forest parameters with the help of regression models in a large area inventory application. The method has also been applied in postprocessing of change detection and site fertility classification results (Häme, 1991; Tomppo, 1992). Kilpeläinen and Tokola (1998) applied the method in a stratification of a forest area into stand margin and within stand areas.

Hagner (1990) went a step further, and presented a method called "*t*-ratio segmentation," (TR), which is a type of region-merging algorithm. The method can use the original image or any low-level segmentation as input, and proceeds by merging adjacent segments if they are similar enough. Similarity is defined by means of the *t*-ratio. The process is iterative, and the most similar segments are merged first. The author stated that a result, comparable to visual interpretation of aerial photos and field checking, could be achieved with this method and SPOT XS and PAN data (Hagner, 1990). Fransson et al. (1999) applied the same method in the identification of clear felled areas with help of SPOT P and Almaz-1 SAR data.

Woodcock and Harward (1992) presented an interesting centroid linkage region growing algorithm for forest stand delineation from satellite images. The method is a multiple pass approach, which uses a global distance threshold as a merging criterion. The merging process is controlled by this threshold, a merging coefficient, and size parameters. The approach yielded segments corresponding to fairly large forest stands, which were applicable for further analysis (Woodcock & Harward, 1992).

The study in hand presents results from testing two segmentation methods in the feature extraction for *k*nn-based multisource estimation of timber volumes using Landsat TM imagery. A measurement space-based clustering (Haralick & Shapiro, 1985) and a modified implementation of the method of Narendra and Goldberg (1980) were applied to produce low-level segmentations, which were further fine-tuned with two alternative region growing algorithms. The aim of the study was to investigate whether the accuracy of volume estimates can be improved when the features for volume estimation are extracted from a homogeneous neighbourhood around the sample plot pixels instead of the plot pixels only.

2. Materials

2.1. Image data

The image material consisted of two Landsat TM scenes (WRS 188/16 and 188/17) both acquired on August

22, 1996. Images were rectified to the Finnish uniform coordinate system and resampled to a pixel size of 25 × 25 m. Second-order polynomial models and a nearest-neighbour resampling method were applied in the rectification. The root mean square errors (RMSEs) of the rectification were 0.62 (188/16) and 0.65 pixels (188/17). Only the parts of the images, which covered the area of the Keski-Suomi and Pohjois-Savo Forestry Centres, were applied because the field data measured in 1996 was only available for these areas. The size of the final, merged image was 60 × 52 km (Fig. 1). The thermal channel (TM 6) was excluded. The digital map data from the Land Survey of Finland was utilised to exclude land use classes other than forest land.

The correspondence of the field information of the sample plots and the intensities of the image was examined by determining the correlation coefficients between the volume of growing stock and the intensities of the TM channels 3 and 4 in a 3 × 3 window around the sample plots (Table 1). The highest correlation coefficients were found in the bottom-left corner of the 3 × 3 window, which indicated that there were some errors in the image rectification or in the locations of the sample plots.

Table 1
The correlation coefficients between the total volume of the growing stock and the intensities of the Landsat TM channels 3 and 4 in a 3 × 3 window centred on the sample plots (n = 466)

	TM 3			TM 4		
	−25 m	0 m	+25 m	−25 m	0 m	+25 m
+25 m	−.365	−.420	−.350	−.263	−.344	−.307
0 m	−.497	−.491	−.401	−.447	−.470	−.395
−25 m	−.540	−.497	−.403	−.584	−.546	−.400

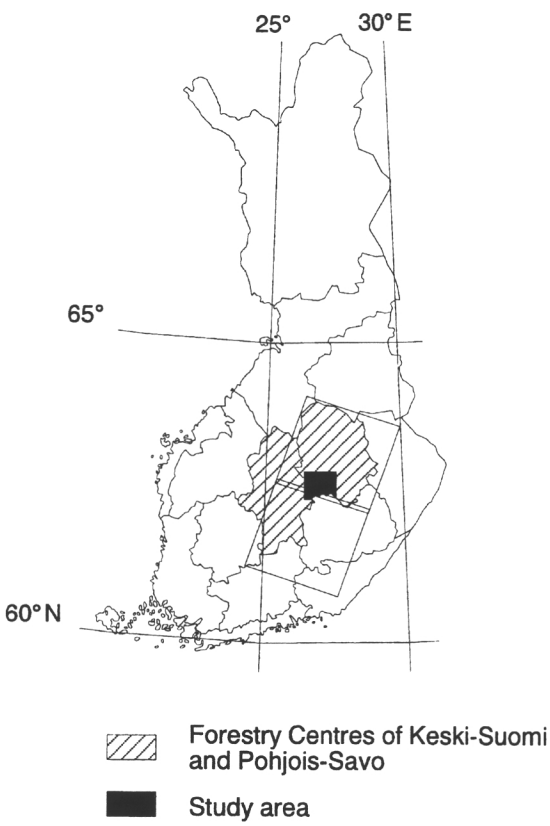


Fig. 1. The location of the study area inside the Landsat TM scenes. The Keski-Suomi and Pohjois-Savo Forestry Centres are shaded.

2.2. Field data

The sample plot data of the 9th NFI of Finland were used in the study. One sampling unit (cluster) of the NFI in the study area consisted of 18 relascope sample plots (basal area factor 2 m^2/ha), which were located at 300-m intervals along the sides of a rectangle. The distance between two clusters was 7 km both in the north–south and in the east–west directions (Fig. 2). The field data of the study area were measured in 1996.

In the field survey, the positions of sample plots were located using bearings and distances measured from points that were identifiable in the field and on the 1:20,000 base map. The total number of sample plots in the study area was 1065, of which 466 were located on forestry land and were completely inside their respective forest stands. Sample plots intersected by a stand boundary were excluded from the data set. Of the sample plots, 81% were on mineral soil land and the remaining 19% on peat land.

The variables of interest were the total volume of growing stock and the volumes of pine (*Pinus sylvestris*), spruce (*Picea abies*) and broad-leaved species, e.g., birch (*Betula pendula* and *Betula pubescens*) and aspen (*Populus tremula*). The volumes per area unit were calculated

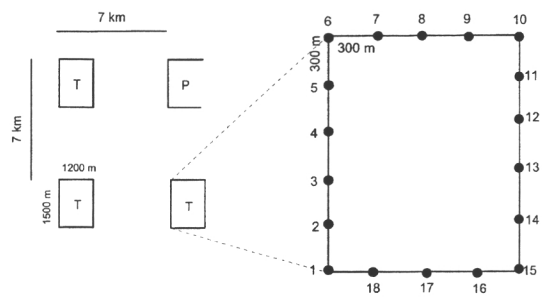


Fig. 2. The layout of the NFI sample plots in the study area.

Table 2
Statistics of timber volumes in the applied NFI sample plot data

Volume (m^3/ha)	Mean	S.D.	Min	Max
Total	110.8	104.0	0.0	493.0
Pine	40.2	56.6	0.0	332.0
Spruce	50.1	81.5	0.0	493.0
Broad-leaved	20.5	36.8	0.0	202.0

The total number of plots = 466.

from the volumes of tallied trees on a sample plot. The main statistics of the variables of interest are shown in Table 2.

3. Methods

3.1. Image segmentation

Two different low-level segmentation methods were tested. The first one (ISOCCL) was a measurement space-guided clustering (see Haralick & Shapiro, 1985) carried out by means of ISODATA clustering followed by the CCL (Jain, Kasturi, & Schunck, 1995). The second method (NG) was based on the "Segmentation with directed trees" (Narendra & Goldberg, 1980). The results of both segmentations were fine-tuned using two region-merging algorithms in order to remove segments smaller than a predefined minimum size.

In the first segmentation method, the clustering of the input image was carried out by means of the ISODATA clustering algorithm (ERDAS, 1994) and turned into segmentation by means of the CCL. The ISODATA clustering was controlled by two parameters: number of clusters and the convergence threshold. The values of these parameters were set heuristically to 50 and 0.95, respectively. The initial cluster centres were arbitrarily determined, and each pixel was assigned to the nearest cluster in the feature space. The Euclidean distance measure and simple linkage were applied. The algorithm was run until the convergence threshold was reached. The convergence threshold is the proportion of those pixels whose cluster value has not been changed between two sequential iterations. Prior to the clustering, two additional channels including the x- and y-pixel locations were merged to the original six-channel image in order to increase the spatiality of the clustering. To restrict the weight of these coordinate channels in clustering, they were linearly scaled to the range of 1–30.

The second segmentation method tested, the NG algorithm, combines the features of both edge- and region-based segmentations and, at least to some extent, avoids their drawbacks. The NG starts by dividing an image to edge and plateau pixels. This task is carried out by means of an edge (e) and a gradient (G) operator and a gradient threshold value (T). First, the edge value e is defined for each image pixel (i,j) with the help of its eight-connected neighbourhood (Fig. 3) and the original grey value $g(i,j)$ (Eq. (1)). The result is used as input to the gradient operator (Eq. (2)). If the value of $G(i,j) \geq T$, the pixel is defined as an edge pixel and, if not, as a plateau pixel. Several different values for T were tested and the value applied ($=6$) was based on visual validation and the mean size of the segments.

$$e(i,j) = \sum_{n=1}^{\text{nbr}} \left(\sum |g_n(i,j) - g_n(k,l)| \right) \quad (k,l) \in N_8(i,j) \tag{1}$$

where nbr = the number of channels.

$$G(i,j) = \max[e(i,j) - e(k,l)] \quad (k,l) \in N_8(i,j) \tag{2}$$

The next step of the NG is the actual image segmentation. It is carried out by two sequential passes over the image. During the first pass, all edge pixels are considered. If an edge pixel has $G(i,j) < 0$, it is marked as a root pixel, otherwise, it is linked to the direction of the smallest edge gradient. During the second pass, all the plateau points are arbitrarily linked to the pixels of the same plateau in their N_8. The links are tracked and the formation of directed cycles is prohibited. If an allowed neighbour is not found, the plateau pixel is labeled as a root pixel. Finally, all the root pixels are labeled and the pixels of each directed tree are given the label of the root pixel of that tree. More detailed description of the algorithm can be found in Narendra and Goldberg (1980) or Tomppo (1992).

In order to remove segments of single or few pixels caused by image noise, the ISOCCL and NG results were fine-tuned with two region growing algorithms. The first one (nearest neighbour [NN]) was based on a minimum size parameter and the Euclidean distance to the spectrally nearest-neighbouring segment. The Euclidean distance between the segment means and standard deviations of all channels was applied. The minimum segment size of 8 pixels (0.5 ha) was applied, which is the recommended minimum size of an operational unit in forest management planning in Finland. The second region-merging

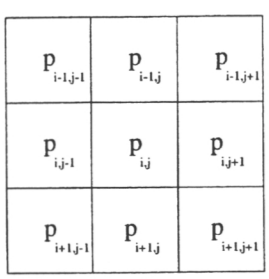


Fig. 3. The eight-connected neighbourhood N_8 of the pixel $p_{i,j}$.

method (TR) was based on the t-ratio (Eq. (3)) (Hagner, 1990). The three first principal components of the original TM image were used as an input to the algorithm, and the square root of the summed squares of the t-ratios of each channel was applied (Hagner, 1990). The method was controlled by the minimum size and the t-ratio threshold (T_t) parameters. If a segment was smaller than the minimum size, it was merged to its nearest-neighbouring segment. The distance was measured by means of the t-ratio. If the segment was larger than the minimum size, but the t-ratio between it and its NN was smaller than T_t, the segments were merged. If the segment size was only 1 pixel, the NN rule was applied. The minimum size applied was same as in the NN method and the value of T_t (=2) was determined heuristically.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{3}$$

where $\bar{x}_i$ = the mean intensity of the segment i; s_i = the variance of the segment i; n_i = the number of pixels in the segment i.

3.2. Feature extraction and the estimation of timber volumes

The spectral features for each sample plot were extracted from the immediate neighbourhood of the plot pixel. The plot pixel was defined as the pixel whose coordinates were closest to the plot centre.

Two methods were tested for the feature extraction. In the first one, the average of spectral values was extracted from a square-shaped window around the plot pixel. This method is referred from now on as a fixed window (FW) approach. In the second method, the average value of those pixels within the window that belonged to the same segment as the plot pixel was extracted. Window sizes from 1 (0.06 ha) to 11×11 pixels (7.56 ha) were applied.

A stepwise algorithm based on forward selection was tested in order to find out the best spectral features for the estimation. The algorithm tested feature combinations of the original TM channels, their products, ratios, and logarithms, and found the suboptimal feature set that minimised the RMSE of the volume estimates. Only the features of the plot pixel were used as input to the algorithm. However, the resulting suboptimal feature set was found to be unstable, and it did not perform well with the segment-based features. Therefore, the extracted averages of the pixel values of the original TM channels were applied in the estimation.

A nonparametric k nearest neighbour method ($k$nn) was applied for the estimation of timber volumes. The $k$nn-estimator has been widely used in multisource forest inventory applications (e.g., Nilsson, 1997; Tokola et al., 1996; Tokola & Heikkilä, 1997; Tomppo, 1993; Tomppo, Gouldding, & Katila, 1999; Trotter & Dymond, 1997). Advantages of the method are that several forest variables can be estimated simultaneously for each pixel and that the natural dependence structure between variables is preserved (Moeur, 1987). The estimates were determined as a weighted mean of k spectrally nearest neighbours. These neighbours were determined with the help of the Euclidean distance between the extracted spectral features of the sample plots. Each of the k nearest neighbours were proportionally weighted by inverse squared Euclidean distance (Eq. (4)). The weight was applied in order to decrease the bias of the $k$nn-estimator (Altman, 1992).

$$\hat{y} = \left(\sum_{i=1}^{k} \frac{1}{d_i^2} y_i \right) \Big/ \sum_{i=1}^{k} \frac{1}{d_i^2} \tag{4}$$

where $\hat{y}$ = the estimate; y_i = the value of y of the ith nearest neighbour; d_i = the Euclidean distance to the ith nearest neighbour; k = the number of neighbours applied.

The actual estimation tests were carried out by means of the cross-validation technique (leave one out). Every plot was in turn left out of the data set and its characteristics were estimated with the aid of the other plots. The accuracy of the estimates was measured by means of the RMSE and the relative RMSE ($\mathrm{RMSE_r}$) (Eqs. (5) and (6)).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{5}$$

$$\mathrm{RMSE_r} = \frac{\mathrm{RMSE}}{\bar{\hat{y}}} \tag{6}$$

where $\hat{y}_i$ = the estimate; $\bar{\hat{y}}$ = the mean of the estimates; y_i = the observed value of y; n = the number of observations.

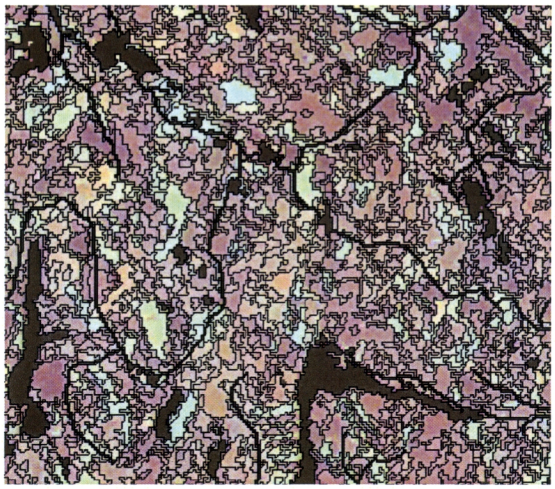
4. Results

4.1. Segmentation

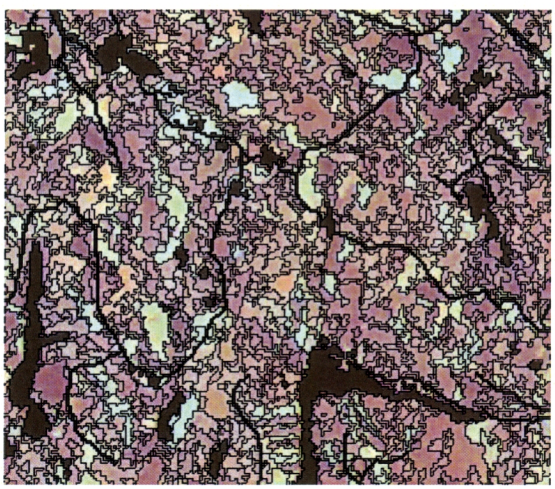
Four different segmentations were derived. Two low-level segmentations were derived by means of two methods: (1) measurement space-guided clustering (ISOCCL) and (2) the NG algorithm. The low-level segmentations were fine-tuned by means of both the NN and TR region-merging algorithms. The resulting number of segments from different segmentations is shown in Table 3 and examples of segmentations in Figs. 4 and 5.

Table 3
The total number of segments after different segmentations with and without region merging

Method		Region merging	
		NN	TR
ISOCCL	469,527	95,123	99,358
NG	183,689	132,728	98,017



(a)



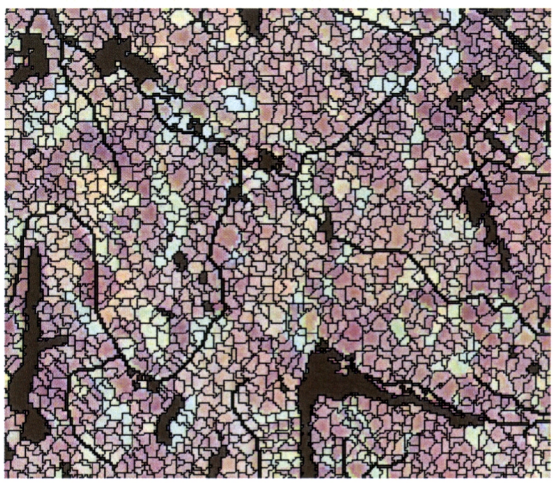
(b)

Fig. 4. The ISOCCL-based segment borders on the original Landsat TM image (©ESA 1996, Eurimage, Novosat Oy): (a) the ISOCCL and NN; (b) the ISOCCL and TR. The displayed channels in RGB order are TM 4, TM 5, and TM 3.

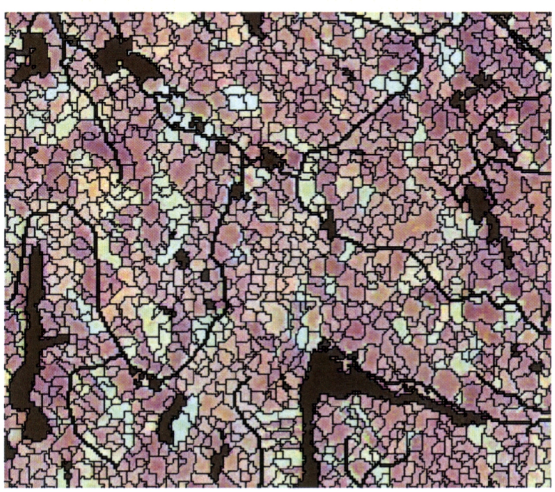
The number of the ISOCCL segments before region merging was about 2.5-times higher than that of the NG method. The mean area of the ISOCCL segments was 0.39 ha and the mean area of the NG segments was 0.97 ha. In the fine-tuning of the ISOCCL segmentation, the NN and TR methods resulted in fairly similar segmentations, the mean sizes of segments being 1.91 and 1.83 ha, respectively. With the NG method, the TR region merging led to a larger segment size than the NN method. This was due to the fact that the result of the NG did not have as many smaller segments than the given minimum size. The mean sizes of the NN and TR fine-tuned NG segments were 1.34 and 1.82 ha, respectively.

4.2. Volume estimates

The estimates of timber volume were derived for each sample plot by means of the *k*nn estimator and the cross-validation technique. Values from 1 to 30 for parameter *k*



(a)



(b)

Fig. 5. The NG-based segment borders on the original Landsat TM image (©ESA 1996, Eurimage, Novosat Oy): (a) the NG and NN; (b) the NG and TR. The displayed channels in RGB order are TM 4, TM 5, and TM 3.

were tested. The results of these tests are shown in Figs. 6 and 7. The RMSEs of the volume estimates decreased rapidly when the number of neighbours was increased from 1 to 10, but only slightly after that. The volume estimates for the comparisons were computed using the 10 nearest sample plots (Table 4). This number of neighbours has been reported in the previous studies to give sufficient accuracy for estimates of timber volume (Nilsson, 1997; Tokola et al., 1996). The window sizes, which gave the lowest RMSEs are shown in Table 5.

The effect of the window size is illustrated in Figs. 8 and 9, which show the FW approach and the segmentation method giving the best estimation result for each tree species. In all the tested window sizes, the segment-based features were more stable and gave lower RMSEs than the FW features.

In the estimation of the total volume of growing stock, the lowest RMSEs were achieved with window sizes of 3×3 pixels. This was the case with all the tested segmentation methods, and also with the FW approach. The use of the ISOCCL and TR in the feature extraction gave the best result with 10 NNs. The RMSE of the estimate of the total volume improved from 89.6 to 86.1 m^3/ha when the ISOCCL and TR segments in 3×3 window were used

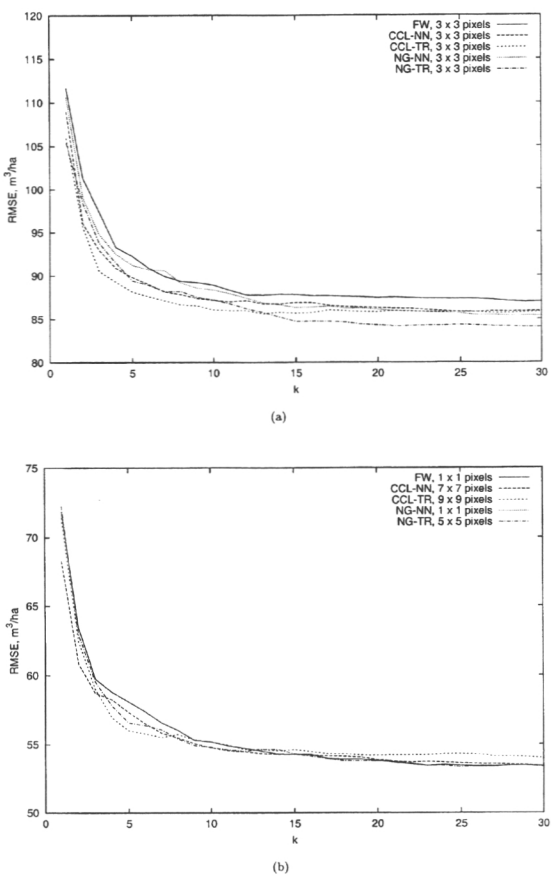


Fig. 6. The RMSEs of volume estimates by tree species using the window size giving the best estimate of each segmentation method for feature extraction: (a) total volume of the growing stock and (b) volume of pine.

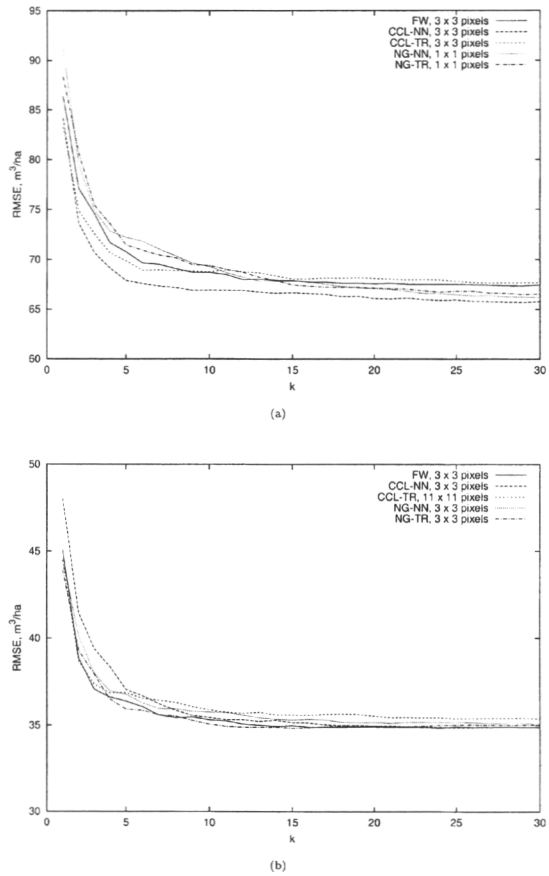


Fig. 7. The RMSEs of volume estimates by tree species using the window size giving the best estimate of each segmentation method for feature extraction: (a) volume of spruce and (b) volume of broad-leaved trees.

Table 4
The RMSEs and relative RMSEs of the volume estimates by tree species and by feature extraction methods

	Pine		Spruce		Broadl.		Total	
Method	m^3/ha	%	m^3/ha	%	m^3/ha	%	m^3/ha	%
FW	55.2	139.9	68.7	139.9	35.3	170.6	88.9	80.4
ISOCCL, NN	**54.8**	**132.7**	**66.9**	137.6	35.5	186.3	87.2	80.4
ISOCCL, TR	55.2	138.1	68.8	141.9	35.9	193.5	**86.1**	**79.3**
NG, NN	55.2	139.9	69.3	138.2	35.8	182.9	88.4	82.7
NG, TR	54.8	138.2	69.3	**136.2**	**35.1**	**181.9**	87.2	82.3

The *k*nn-estimator ($k = 10$) and NFI sample plots ($n = 466$) were applied. The best estimates are typed as bold.

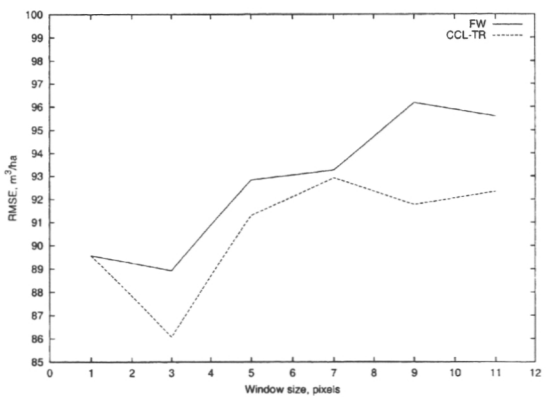
Table 5
The window sizes (pixels) giving the best estimation result with different feature extraction methods

Method	Pine	Spruce	Broad-leaved	Total
FW	1	3	3	3
ISOCCL, NN	7	3	3	3
ISOCCL, TR	9	3	11	3
NG, NN	1	1	3	3
NG, TR	5	1	3	3

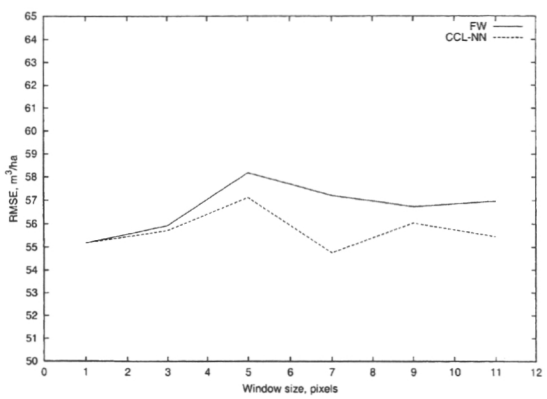
The *k*nn-estimator ($k = 10$) and NFI sample plots ($n = 466$) were applied.

instead of the plot pixel only. Applying any of the segmentation methods in feature extraction improved the estimates compared to the use of the FW.

The ISOCCL segmentation with the NN gave the best estimates for pine and spruce volumes. A window size of 3×3 pixels was the best one for feature extraction for spruce, but 7×7 pixels for pine. In the case of spruce, an improvement of 2.4 m^3/ha in RMSE was achieved when the ISOCCL and NN segments were applied instead of the plot

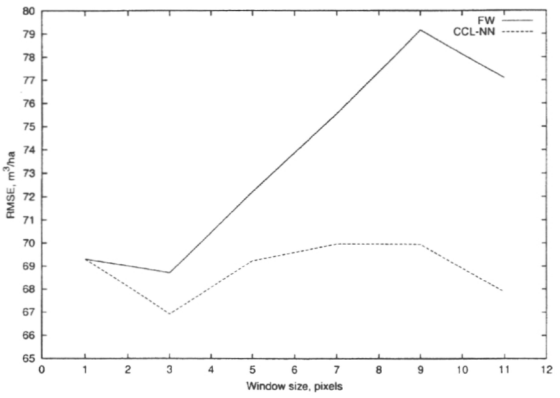


(a)

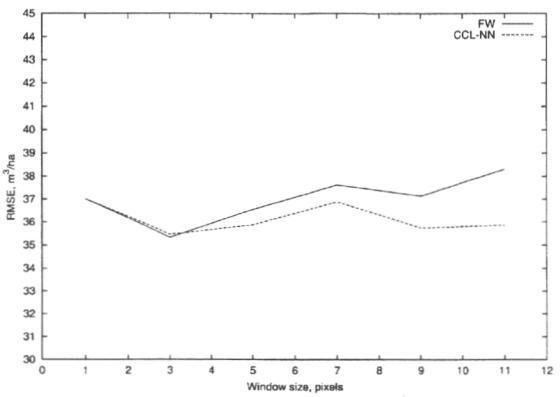


(b)

Fig. 8. The RMSEs of volume estimates by tree species using different window sizes in feature extraction: (a) total volume of the growing stock and (b) volume of pine.



(a)



(b)

Fig. 9. The RMSEs of volume estimates by tree species using different window sizes in feature extraction: (a) volume of spruce and (b) volume of broad-leaved trees.

pixel only. In the FW approach, the use of adjacent (3×3) pixels also gave a better result than the use of the plot pixel. The improvement of the volume estimate of pine was only 0.4 m^3/ha, although the advantage of use of the segments was more evident when a smaller number of neighbours was applied (Figs. 6 and 7). In the FW approach, the features extracted from the plot pixel provided the best estimate for pine.

In the estimation of the volume of broad-leaved trees, the best result was achieved when the window size of feature extraction was 3×3 pixels or larger. The combination of the NG and TR methods and a window size of 3×3 pixels gave a slightly better result than other methods.

5. Discussion

The aim of this study was to test if the plot-level estimates of timber volume can be improved by applying

image segments in feature extraction. Segment-based and traditional feature extraction methods were compared. In the case of traditional methods, the spectral features were extracted from the sample plot pixel and its local neighbourhood. Different sizes of the neighbourhood were tested. In the segment-based approach, this neighbourhood was restricted with the help of image segments.

In general, the best volume estimates were obtained if image segments were used in the feature extraction. When segment restricted feature extraction was applied instead of the FW approach, the improvement of the RMSE was 0.2–2.8 m^3/ha depending on the tree species. The estimates of the total volume and the volume of spruce improved more clearly than those of pine and broad-leaved trees. However, the overall improvement of the estimates was very small.

The relative standard errors of the estimates were high. The relative RMSE of the estimate of the total volume was the only one under 100%. High errors in plot-level estimates have also been reported in previous studies. Tokola et al. (1996) reported errors of about 67% for total volume and more than 100% for volumes by tree species. The errors reported by Tokola and Heikkilä (1997) are of the same order of magnitude. However, the results of these studies are not directly comparable due to the differences in the estimation parameters and the field data.

Features extracted from the ISOCCL segments gave better results compared to the features extracted from NG segments. This difference was most evident in the estimates of the total volume and the volume of spruce. The differences between the segment-based methods were minor with respect to the estimates of the volumes of pine and broad-leaved trees.

Although the segment-based feature extraction performed better than the FW approach in most of the cases, there were some exceptions. For example, the NG-based feature extraction gave worse estimate for the volume of spruce than the FW. In addition, the best result of the NG approach was obtained by using the centre pixel only, while the best result of the FW approach was achieved with a window size of 3×3 pixels. This phenomenon is due to the filtering effect of the two sequential neighbourhood operators applied in the NG, a consequence of which is that the location of the edge of the segment may not be exact. Thus, if a sample plot is located close to the forest stand boundary, it may be assigned to the segment that represents the neighbouring stand. The larger the extraction window is, the more erroneous features are obtained because the proportion of the pixels of the neighbouring stand increases.

The use of segments in the feature extraction provides a way to diminish some of the errors of image analysis caused by errors in the image registration and sample plot positions. The traditional pixel-based analysis methods are more sensitive to these errors. In this study, the differences between the results of pixel and segment-based methods are, however, quite small. The use of segments did not significantly improve the accuracy of the estimates. This is due to two facts. First, the stand parameters recorded from a relascope plot are very sensitive to the location of the plot. Thus, the relascope sample plot may not be an optimal data collection unit for satellite image-based estimation of forest parameters for small areas. Secondly, in areas where the mean size of the forest stands is small, the relative proportion of mixed pixels becomes very large, which confuses image analysis (Tokola & Kilpeläinen, 1999).

When the selection of an image segmentation method is considered, the NG is more feasible for operational purposes, even though the ISOCCL gave slightly better estimation results. The clustering step of the ISOCCL is computationally very time-consuming and controlled by several parameters. Furthermore, the ISOCCL produces large number of small low-level segments, which have to be merged by means of a region-merging algorithm in order to obtain applicable segments for further analysis. The criteria for merging can be difficult to determine and are usually found heuristically. In the NG, less parameter selection is required and low-level segments are applicable, even though the region-merging step improves the result. After that, the size and shape of NG segments seem to correspond better to those of forest stands.

Image segmentation provides a tool, which can be applied in the extraction of features of the units of interest, i.e., forest stands. In the present study, no stand-level information was available. Consequently, because the sample plot data was not considered representative at the stand-level, the feature extraction had to be restricted to the immediate neighbourhood of a plot. If such an analysis would be based on the data recorded by homogeneous forest stands, this restriction may not be needed and segments may directly serve as image interpretation units. An alternative way to apply image segments in the estimation of forest parameters could be the stratification of an image into stand margin and within stand areas. Such an approach would not be as sensitive to the minor errors in the location of the segment edge as the one presented here.

Acknowledgments

We would like to thank D.For. Jari Varjo and MSc (for.) Reija Haapanen for their review comments, and PhD Ashley Selby for revising the English language. This study was partly funded by the "Forests in GIS Graduate School" of the University of Helsinki.

References

Altman, N. (1992). Introduction to kernel and nearest-neighbour nonparametric regression. *The American Statistician*, *46* (3), 175–184.

ERDAS. (1994). *ERDAS field guide* (3rd ed.). Atlanta, GA: ERDAS.

Fransson, J., Walter, F., & Olsson, H. (1999). Identification of clear felled areas using SPOT P and Almaz-1 SAR data. *International Journal of Remote Sensing*, *20*, 3583–3593.

Fu, K., & Mui, J. (1981). A survey of image segmentation. *Pattern Recognition*, *13* (1), 3–16.

Hagner, O. (1990). Computer aided forest stand delineation and inventory based on satellite remote sensing. In: *Proceedings from SNS/IUFRO workshop in Umeå 26–28 February 1990: the usability of remote sensing for forest inventory and planning* (pp. 94–105). Umeå: Swedish University of Agricultural Sciences, Remote Sensing Laboratory.

Häme, T. (1991). Spectral interpretation of changes in forest using satellite scanner images. *Acta Forestalia Fennica*, *222*, 1–111.

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

Holmgren, P., & Thuresson, T. (1998). Satellite remote sensing for forestry planning — a review. *Scandinavian Journal of Forest Research*, *13*, 90–110.

Jain, R., Kasturi, R., & Schunck, B. (1995). *Machine vision*. McGraw-Hill.

Keller, J., Gray, M., & Givens, J. J. (1985). A fuzzy *k*-nearest neighbor algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, *15* (4), 580–585.

Kilkki, P., & Päivinen, R. (1987). Reference sample plots to combine field measurements and satellite data in forest inventory. In: Remote sensing-aided forest inventory. Proceedings from seminars organised by SNS, Hyytiälä, Finland, Dec. 10–12, 1986. University of Helsinki, Department of Forest Mensuration and Management, Research Notes, 19, 209–215.

Kilpeläinen, P., & Tokola, T. (1998). Gain to be achieved from stand delineation in Landsat TM image-based estimates of stand volume. *Forest Ecology and Management*, *124*, 105–111.

Mäkinen, T. (1999). Metsäsuunnittelun kehittämisprojekti 1998–1999. Technical report, Tapio (in Finnish).

Moeur, M. (1987). Nearest neighbor inference for correlated multivariate attributes. In: Proceedings of IUFRO conference on forest growth modeling and prediction, Minneapolis, MN, 23–27 August, 1987. USDA Forest Service, General Technical Report NC-120, pp. 716–723.

Narendra, P., & Goldberg, M. (1980). Image segmentation with directed trees. *IEEE Transactions on Pattern Analysis and Machine Intelligency Pami*, *2* (2), 185–191.

Nilsson, M. (1997). Estimation of forest variables using satellite image data and airborne lidar. Licensiate thesis, Swedish University of Agricultural Sciences, Department of Forest Resource Management and Geomatics, Umeå.

Pal, N. R., & Pal, S. K. (1993). A review on image segmentation techniques. *Pattern Recognition*, *26* (9), 1277–1294.

Parmes, E. (1992). Segmentation of SPOT and Landsat satellite imagery. *The Photogrammetric Journal of Finland*, *13* (1), 52–58.

Poso, S. (1983). Kuvioittaisen arvioimismenetelmän perusteita. Summary: basic features of forest inventory by compartments. *Silva Fennica*, *17* (4), 313–349 (in Finnish).

Tokola, T. (1990). Satelliittikuvan ja VMI-koealatiedon käyttö metsätalousalueen puuston inventoinnissa. Licensiate thesis, University of Joensuu, Faculty of Forestry.

Tokola, T., & Heikkilä, J. (1997). Improving satellite image based forest inventory by using a priori site quality information. *Silva Fennica*, *1* (31), 67–78.

Tokola, T., & Kilpeläinen, P. (1999). The forest stand margin area in the interpretation of growing stock using Landsat TM imagery. *Canadian Journal of Forest Research*, *29*, 303–309.

Tokola, T., Pitkänen, J., Partinen, S., & Muinonen, E. (1996). Point accuracy of a nonparametric method in estimation of forest characteristics with different satellite materials. *International Journal of Remote Sensing*, *17* (12), 2333–2351.

Tomppo, E. (1987). Stand delineation and estimation of stand variates by means of satellite images. In: Remote sensing-aided forest inventory. University of Helsinki, Department of Forest Mensuration and Management, Research Notes, 19, pp. 60–76.

Tomppo, E. (1990). Satellite image-based national forest inventory of Finland. *Photogrammetric Journal of Finland*, *12* (1), 115–120.

Tomppo, E. (1992). Satellite image aided forest site fertility estimation for forest income taxation. *Acta Forestalia Fennica*, *229*, 1–70.

Tomppo, E. (1993). Multisource national forest inventory of Finland. In: Proceedings of the Ilvessalo Symposium on National Forest Inventories, 17–21 Aug., 1992, Finland. The Finnish Forest Research Institute, Research Papers, 444, pp. 52–60.

Tomppo, E., Goulding, C., & Katila, M. (1999). Adapting Finnish multisource forest inventory techniques to the New Zealand preharvest inventory. *Scandinavian Journal of Forest Research*, *14*, 182–192.

Trotter, C., & Dymond, J. (1997). Estimation of timber volume in a coniferous plantation forest using Landsat TM. *International Journal of Remote Sensing*, *18* (10), 2209–2223.

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

Paper II

Reprinted from International Journal of Remote Sensing, vol 23 no 14, Pekkarinen, A., A method for the segmentation of very high spatial resolution images of forested landscapes, pp 2817-2836, copyright (2002), with permission from Taylor & Francis Ltd.
<http://www.tandf.co.uk>

II

INT. J. REMOTE SENSING, 2002, VOL. 23, NO. 14, 2817–2836

Taylor & Francis
Taylor & Francis Group

A method for the segmentation of very high spatial resolution images of forested landscapes

A. PEKKARINEN

Finnish Forest Research Institute, Unioninkatu 40 A, FIN-00170 Helsinki, Finland; e-mail: anssi.pekkarinen@metla.fi

(*Received 22 February 2000; in final form 4 April 2001*)

Abstract. Pixel-by-pixel image analysis methods are not applicable when using VHR image material in multisource forest inventory applications. One possible solution to this problem is to define the units of image analysis by means of image segmentation. The paper presents a two-phase segmentation method (COS) based on segmentation in the feature space and co-occurrence region merging (CRM). The resulting segments were tested in the spectral feature extraction, and the estimation of plot-level total volume and volumes by tree species. The study material consisted of an AISA spectrometer image and 254 relascope field data plots. Two different segmentations were derived and the performance of segment-based feature sets were compared to that of a feature set extracted from the local neighbourhood of the field plots. Cross-validation techniques and a *k-nn* estimator were applied in the estimation tests. The estimation results which had the smallest rms errors were achieved with segment-based features in the cases of total volume and the volume of deciduous species. In the cases of pine and spruce, the features from the local neighbourhood performed best. The study suggests that COS produces segments which can be used as units for further image analysis.

1. Introduction

Remote sensing is a cost-efficient source of information for forest inventory and monitoring purposes. Satellite images, e.g. Landsat TM, have been widely applied in different forest inventory and monitoring tasks. Examples of these tasks include change detection (Häme 1991, Varjo 1997, Häme *et al.* 1998) and the estimation of forest parameters (Tomppo 1996, Tokola and Heikkilä 1997, Trotter *et al.* 1997). Although some of these applications have proven to be useful for large-area inventories (Tomppo 1996), doubts have been raised concerning the applicability of satellite remote sensing for forest management planning purposes (Holmgren and Thuresson 1998). One reason for such doubts is the limited spatial resolution of the applied image material.

The errors arising from coarse spatial resolution may be divided into two classes. Locational errors include errors in the image registration, and errors in the location of the training areas. Both of these may lead to a mismatch between spectral and informational classes. This problem is obviously more serious in the case of small training areas. The other class of errors include errors caused by mixed pixels, mixels. Mixels are pixels which carry spectral information from more than one informational

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160110076162

class. In Finland, the average size of an operative forest management unit is about 1.5 hectares. Thus, if the spatial resolution of the remote sensing material is about 30 metres, a typical stand consists of a small amount of pixels, most of which are mixels. Thus, the increasingly available very high spatial resolution (VHR) remote sensing material provides interesting new possibilities for forest inventory applications.

The use of VHR images also sets new demands for image analysis methods. So far, most of these methods have operated pixel-by-pixel and have not utilized the spatial information present in the image. More spatially oriented methods are required in order to fully exploit the improvements in spatial resolution. One possible solution to this exploitation problem is the use of image analysis units larger than a single pixel. This kind of a unit can be determined, for example, by means of image segmentation.

Image segmentation seeks to divide an image into spatially continuous, disjoint and homogenous regions. Segmentation methods can be divided into pixel-, edge- and region-based methods.

Pixel-based methods include thresholding and clustering in the feature space. Because these methods may produce a result in which two or more spatially unconnected areas carry the same label, their result cannot be generally considered as segmentation. This problem is solved by labelling spatially connected components of the resulting image (Haralick and Shapiro 1985).

Edge-based image segmentation methods are based on the idea that the regions of an image can be detected by finding the edges between them. In general, edge detection algorithms consist of the following phases: filtering, enhancement, detection and localization (Jain *et al.* 1995). The edge enhancement step is usually carried out in the local neighbourhood of a pixel by means of an edge operator. Examples of commonly applied edge operators are those of Sobel, Prewitt and Laplace (Jain *et al.* 1995). The edge detection phase is usually based on thresholding the result of the edge enhancement. Thus, only those pixels which have an edge magnitude larger than a certain threshold are considered as edge candidates. Finally, these pixels are linked to a contour and the region surrounded by it is considered to be a segment.

Region-based segmentation methods include region growing, region merging, region splitting and their combinations. In region growing approaches, the first step is to identify starting points of the segmentation, often referred to as 'seed' pixels. The regions are built around these pixels by joining the similar neighbouring pixels to them. In region merging methods, the adjacent regions are merged if they are similar enough. In region splitting techniques, a region is divided into subregions if the original region is not homogeneous.

Image segmentation is a commonly applied technique in the fields of machine vision and pattern recognition. For reviews of different segmentation methods, see, for example, Fu and Mui (1981), Haralick and Shapiro (1985) and Pal and Pal (1993).

The need for the use of segmentation in remote sensing-based forestry applications was recognized during the 1980s, when the studies began concerning the applicability of Landsat TM and SPOT images to forest inventories. Several segmentation methods were studied. Tomppo (1987) tested the applicability of the method presented by Narendra and Goldberg (1980). The method is based on the construction of directed trees guided by local edge values. Tomppo (1987) concluded that it was possible to develop a method which could be used to assist stand-level forest inventories based on satellite images and their segmentation. The implementation of

the applied segmentation method for large images was presented later (Parmes 1992). Parmes suggested that the result of the segmentation of textured areas could be improved with the help of additional texture channels. A method called 't-ratio segmentation' was also tested in a stand delineation and inventory application by Hagner (1990). The method used any low-level segmentation or the original image as an input and proceeded by iterative merging of similar segments. The similarity was defined with the help of the t-ratio. The minimum and maximum sizes of segments and number of merging iterations were defined by the user. Hagner (1990) argued that a stand delineation comparable to the visual interpretation of aerial photos and field checking could be achieved by applying this method and SPOT data. Another kind of multiple pass algorithm has also been reported to yield applicable results in the segmentation of Landsat TM images of forested landscape (Woodcock and Harward 1992). In addition to region size constraints, this method used a global threshold as a rule in the distance-based merging process.

Segmentation methods have been also tested with filtered aerial photographs for assessing the diversity of forested landscapes (Holopainen 1998). The method applied was basically similar to the one presented by Hagner (1990), but the low-level segmentation was created by means of k-means clustering.

Even though these methods have been successfully used in the segmentation of Landsat TM and SPOT images and filtered aerial orthophotos, they are quite problematic if the segmentation of VHR images is considered. A VHR image of a forested area is a composition of certain spectral classes, such as illuminated and shaded crown, and ground pixels. The spectral properties of the neighbouring pixels within a stand may thus be very different. Therefore methods which apply absolute intensity difference thresholds may not work sufficiently well. Similarly, segmentation in the feature space will lead to unsatisfactory results due to the high local variation. These methods typically produce very small segments which may may consist of, for example, pixels of the shaded or the illuminated part of the crown. One possible solution to this problem is to use some kind of a region merging algorithm in order to merge neighbouring segments into larger units which include the spectral variation of the stand. If the merging criteria is based on the spectral similarity of the neighbouring segments, however, the merging process is more or less arbitrary because all the neighbouring segments may be spectrally very different from the one being processed.

This paper suggests a new two-phase co-occurrence segmentation algorithm (COS). The method derives low-level segments by applying segmentation in the feature space, and final segments from the low-level segments by means of co-occurrence region merging (CRM). The CRM is based on the minimum segment size criteria and the gray level co-occurrence of the spectral classes of the neighbouring segments. The use of co-occurrence as the basis of region merging makes it possible to merge in a meaningful way segments which have spectrally different properties. The segmentation results will be illustrated, and evaluated in the context of multisource forest inventory.

2. Material

2.1. *Field data*

This study was carried out using the previously established test area of Ohkola in the municipality of Mäntsälä, Southern Finland (figure 1). The test area was originally established for imaging spectrometer studies by the National Forest

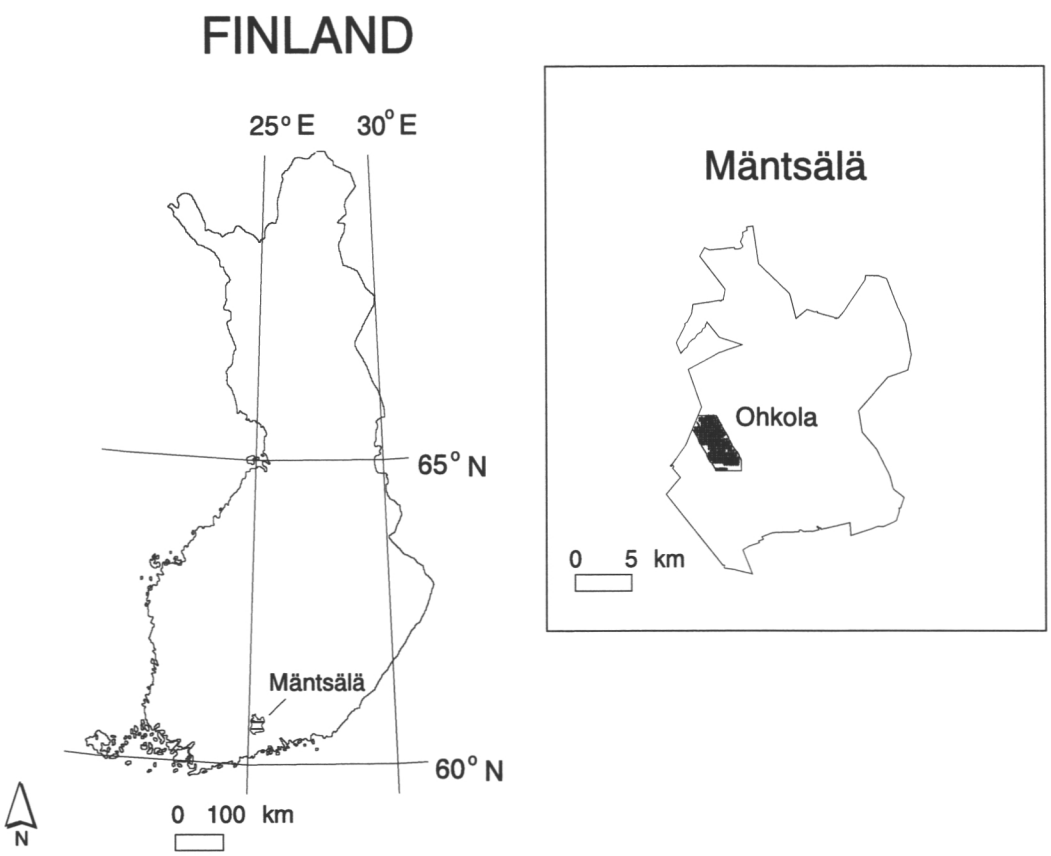


Figure 1. Location of the municipality of Mäntsälä within the borders of Finland and the location of the Ohkola test area within the borders of Mäntsälä.

Inventory (NFI) team of the Finnish Forest Research Institute (FFRI) (Mäkisara *et al.* 1997).

The test area was covered by a systematic square grid of 254 field plots. The distance between the plots was 250 m in both north–south and east–west directions. All the plots were located using a surveying tape and a special compass. The coordinates of the plots were also recorded using a GPS device (Trimble Explorer) which employed the Radio Data System-based differential correction (RDS 3000). Relascope plots (factor 2) with a fixed maximum radius of 12.52 m were applied. Every seventh tree of these relascope plots was measured in detail as a sample tree.

The volumes of all sample trees of the test area were estimated using the volume functions of Laasasenaho (1982) employing three stem parameters measured in the field: diameter at breast height ($d_{1.3}$), diameter at the height of six metres ($d_{6.0}$), and height. The volumes of the trees tallied from each plot were interpolated from the estimated volumes of these sample trees with help of $d_{1.3}$, tree species and some stand characteristics (Tomppo *et al.* 1998). Tree level volumes were converted to volumes per area unit ($m^3\ ha^{-1}$) and summed by tree species. If the plot intersected several stands, it was divided into subplots and the volumes were summed for these subplots. Finally, the subplots having a proportional area smaller than 0.5 were removed from the material because their volume estimates were considered to be unreliable. The characteristics of this data set are given in table 1.

Each plot was overlaid on the AISA image and its location was visually checked

Table 1. Forest characteristics of the input dataset.

Species	mean ($m^3\ ha^{-1}$)	std ($m^3\ ha^{-1}$)	min ($m^3\ ha^{-1}$)	max ($m^3\ ha^{-1}$)	Pure stands (%)*
Pine (*Pinus sylvestris*)	28.5	44.7	0	220.2	28.5
Spruce (*Picea abies*)	70.3	99.5	0	397.6	25.9
Deciduous (*Betula* spp., *Alnus* spp.)	16.6	25.2	0	142.7	9.4
Total	115.3	103.3	0	408.8	–

*At least 90% of the total volume of the plot consists of the particular species.

on the screen. If the plot (or subplot) was located near the stand border, the stand-level information of the plot (or subplot) was assigned to an artificial plot which was established inside the stand. All measurements were carried out between July and September 1996. A more detailed description of the field data, is given by Mäkisara *et al.* (1997).

2.2. *Image data*

An imaging spectrometer (AISA) image of the test area was acquired on 23 June 1996. The AISA is a low-cost spectrometer developed in Finland (Mäkisara *et al.* 1993). The spectral and spatial configurations of the instrument are programmable (Mäkisara 1998). The original image mosaic consists of data from seven flight lines and 30 spectral channels. The size of the subset image used in this study was 3500×3750 pixels and the pixel size was $1.6 \times 1.6\,m^2$. The spectral configuration applied is shown in table 2. The image was radiometrically and geometrically corrected by the NFI team of the FFRI. An example of the AISA image of the study area is shown in figure 2.

3. Methods

3.1. *Overview*

A new co-occurrence segmentation method (COS) was developed and tested with the AISA image. The method consists of two steps: (1) low-level segmentation carried

Table 2. The spectral channel configuration of AISA (Mäkisara *et al.* 1997).

Ch	Centre (nm)	Width (nm)	Ch	Centre (nm)	Width (nm)
1	470.1	7.3	16	701.4	4.6
2	487.6	7.3	17	710.5	4.6
3	505.1	7.3	18	725.7	7.6
4	522.6	7.3	19	733.3	4.6
5	541.5	7.3	20	742.4	4.6
6	550.2	7.3	21	750.0	4.6
7	564.8	7.3	22	775.8	4.6
8	579.4	7.3	23	784.9	4.6
9	599.8	7.3	24	794.1	4.6
10	623.1	7.3	25	803.2	4.6
11	646.7	7.6	26	818.4	4.6
12	668.0	7.6	27	844.2	4.6
13	675.6	7.6	28	853.3	4.6
14	686.2	4.6	29	860.9	4.6
15	696.8	4.6	30	865.4	4.6

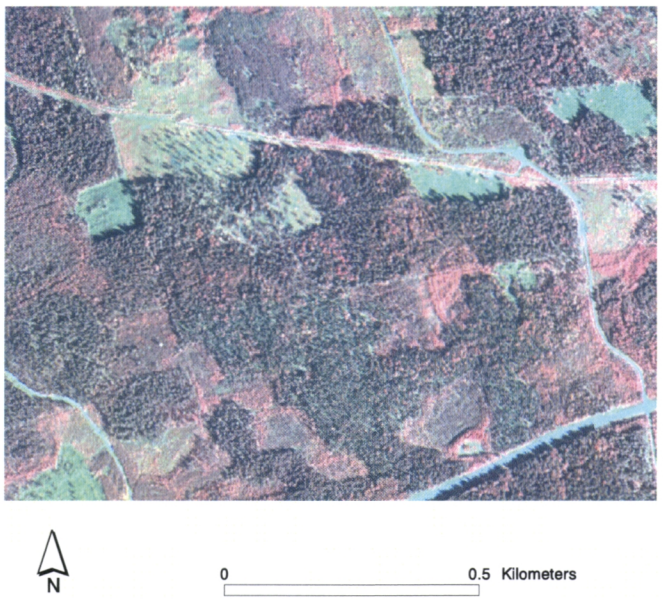


Figure 2. Example of the AISA data. RGB composition of channels 26, 16 and 6 respectively.

out by means of clustering and labelling of connected components (CCL) and (2) co-occurrence region merging (CRM). A principal component analysis step was also applied in this study, but it is not an essential phase of the method itself.

3.2. *Preprocessing*

To reduce the computation time needed in the actual image segmentation steps, principal component analysis (PCA) was applied to the original AISA image. PCA is a commonly applied dimension reduction technique which transforms the original variables into uncorrelated principal components (PCs). It aims to include the maximum amount of variation of the original variables into a few principal components (Afifi and Clark 1984).

Here, the first three PCs based on the covariance matrix of the channels of the original AISA image were applied. These PCs, which accounted for 99.8% of the total variation of the original image, were linearly scaled in a range of 0 to 255 and employed as input to the image segmentation steps.

3.3. *Low-level segmentation*

The first step in the actual image segmentation was to cluster the image and produce a low-level segmentation for further processing. The clustering step was carried out by means of the iterative self-organizing data analysis technique (ISODATA) (Erdas 1999). ISODATA starts by arbitrarily locating the given number of cluster centres in the feature space. It then classifies each pixel of an image into one of these initial clusters. The classification is carried out by assigning each pixel to the nearest cluster in the feature space. A Euclidean distance measure is applied. After the image has been processed, the cluster centres are redetermined as a mean vector of the observations of the cluster. This process is iterated until one of the

user-defined parameters, convergence threshold or the maximum number of iterations, is reached. The convergence threshold is the percentage of those observations which remain in a same cluster during two sequential iterations (Erdas 1999).

To introduce a spatial aspect into the clustering process, two additional variables, row and column image coordinates, were added to the three-channel PC image prior to the clustering. These coordinate channels were scaled from 0 to 255.

Reasonable parameters for the clustering process were found by means of error and trial. It was found that a small number of classes, say less than 50, tended to produce segments which were too large. This problem became more obvious if the coordinate channels were included in the clustering process. Based on these observations, the number of classes was increased to 255 and the clustering process was continued until at least 95% of the pixels remained in the same class during two successive iterations. Areas outside the actual image data were assigned a null value. Two different clusterings were produced: one based on the three-channel PC image and one based on the five-channel image including PC and coordinate channels. Examples of the results of the clustering processes are shown in figure 3.

The clustered images were transformed into segmented images by labelling the connected components. Neighbouring pixels were considered to be connected if they both carried the same cluster label. The sequential CCL implementation was applied (Jain *et al.* 1995).

Sequential CCL was carried out with help of two passes over the image. The first pass began with the upper left pixel of the image and proceeded from left to right, and from top to bottom. The first pixel was labelled '1'. In the case of all other pixels, the following procedure was applied: if the pixel being processed was connected to any already processed pixel in its neighbourhood, it was given the label of the neighbourhood pixel already processed. If more than one of the neighbouring pixels already processed fulfilled the connectivity criteria, the one with the highest label number was selected and the pixel being processed was given the same label. The equivalency of the labels of all the connected pixels was recorded in an equivalency table. If none of the neighbouring pixels already processed was connected to the pixel being processed, a new unique label was created and given to the pixel being processed.

The first pass of such a 'retrospective algorithm' was unable to label some kinds of continuous regions during the first pass (figure 4). This problem was solved by means of the equivalency table and second pass over the image.

At the beginning of the second pass, the recorded equivalences were processed and a new unique label was assigned to each equivalent label set. The segment labels of the output of the first pass were then replaced with the new labels and the final output was composed. For a more general description of the sequential CCL algorithm, see Jain *et al.* (1995).

3.4. *Region merging step*

The core of the COS is the CRM; the region merging step based on the global gray level co-occurrence matrix (GCM) (Haralick *et al.* 1973). The first step of the CRM is the calculation of the GCM G. Let P be the image having C possible different gray values, and p a pixel of that image $p \in P$ having a gray value of $v(p)$. Furthermore, let p_k^n, $k=1,...,8$ be an element in the 8-connected neighbourhood $N_8(p)$ of pixel p (figure 5). Let us form a $C \times C$ matrix M for P in such a way that an element $m(i,j)$ presents the number of times the value $v(p_k^n)=j$ occurs in the N_8

(*a*)

(*b*)

Figure 3. Examples of the results of the ISODATA clustering processes of applied input images using 255 classes. (a) Input: First three PCs and row and column image coordinates. (b) Input: First three PCs.

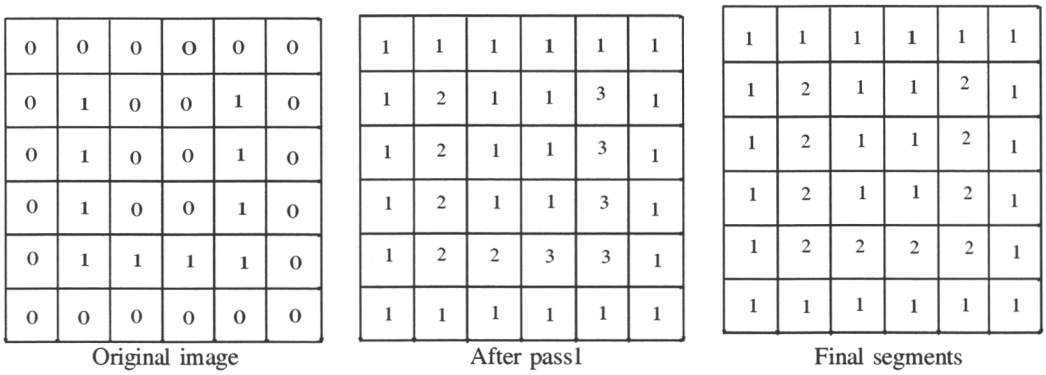


Figure 4. Phases of the sequential CCL. After pass one, the connected region carries two labels: 2 and 3. The equivalency of these labels is recorded during the processing and both of them are replaced with a new label value (2) in the resulting image.

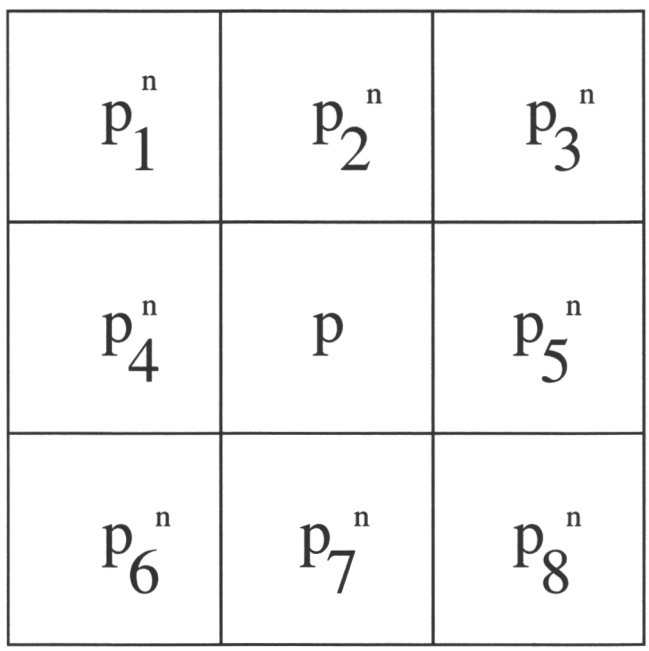


Figure 5. 8-connected neighbourhood of pixel p.

neighbourhood of value $v(p)=i$ (Equation 1). Let $g(i,j)$ be the probability that the value j belongs to the N_8 neighbourhood of value i (Equation 2). Thus, $\hat{g}(i,j)$ (Equation 3) is the estimate of that probability in the GCM G computed from an image P.

This is a simplified presentation of the calculation of the GCM, because the different directions and distances between the elements in the N_8 neighbourhood and the central pixel p are not taken into account. In textural feature extraction, for example, the textural measures are derived separately for each direction (Haralick *et al.* 1973). Here, an inverse distance weighting was used in order to take these directions into account. Thus, elements p_2^n, p_4^n, p_5^n and p_7^n increased the frequency by

one and all the others by $1/\sqrt{2}$.

$$m(i,j) = \sum_{p \in P} \#(v(p_k^n) = j,\ k = 1, ..., 8 \mid v(p) = i) \tag{1}$$

$$g(i,j) = P(v(p_k^n) = j, \text{ for some } k = 1, ..., 8 \mid v(p) = i) \tag{2}$$

$$\hat{g}(i,j) = m(i,j) / \sum_{i=1}^{C} m(i,j) \tag{3}$$

The second step of the CRM was the actual region merging. This process was guided by a minimum region size parameter and the GCM. The merging process was iterated until segments smaller than the minimum size parameter no longer existed. Iterative processing was required because a segment resulting from the merging of two or more segments may still be smaller than the given minimum size. The merging process was controlled as follows: if a segment smaller than the minimum size was found, the co-occurrences between its mode cluster and the mode clusters of neighbouring segments were studied and the segment was merged to the segment with which the co-occurrence value was largest. The applied CRM algorithm was as follows:

```
CRM (ClusterInput, SegmentInput, SegmentOutput) {
 WHILE{ SmallSegmentsExist
  IF { IsFirstIteration
   GCM=ComputeGCM(ClusterInput)
   TmpSegment=MergeSegments(GCM,SegmentInput)
   TmpCluster=ClusterInput
   }
  ELSE{
   TmpCluster=ComputeSegmentModes(TmpCluster, TmpSegment)
   GCM=ComputeGCM(TmpCluster)
   TmpSegment=MergeSegments(GCM,TmpSegment)
   }
  }
 SegmentOutput=TmpSegment
 }
```

At the beginning of the first iteration of the merging process, all pixels of a particular segment belong to the same cluster, while none of the pixels in the adjacent segment belong to that cluster. Both of these properties of the low-level segments result from the definition of segmentation in the feature space (Haralick and Shapiro 1985). Thus, during the first iteration, the GCM can be computed from the original cluster input image. In all other iterations, a segment may consist of pixels of several original clusters. The segment mode cluster value is therefore computed with the help of the template cluster and segment images of the previous iteration. The GCM is then re-computed.

Two different segmentations, one applying low-level segmentation based on clustering of the three first PCs and row and column coordinates (S1), and one applying the three first PCs only (S2), were derived for further evaluation. The applied minimum segment size was 121 pixels (≈ 0.03 hectares). The minimum size of a segment is approximately the same as the optimal pixel size suggested for the stratification of forest stands by Holopainen and Wang (1998).

3.5. *Evaluation of the results*

The results of the segmentations were evaluated in a multisource forest inventory application by utilizing image segments in the spectral feature extraction. The extracted features were applied in the estimation of plot-level forest parameters. The parameters included volumes of Scots pine, Norway spruce and deciduous trees, as well as the total volume of the plot. Spectral features were employed which were as close as possible to the suboptimal features found in an earlier AISA study (Mäkisara *et al.* 1997). They are presented in table 3. The exact suboptimal features could not be used because the spectral configuration of the AISA instrument was different between the studies. The differences between the configurations were, however, quite small.

Feature extraction was carried out for each ground data plot within a square shaped window surrounding it. Window sizes from 3×3 pixels (≈ 0.002 hectares) to 121×121 pixels (≈ 3.75 hectares) were applied. Only odd window side lengths were used.

Three different spectral feature sets (F1, F2 and F3) were derived. The segment-based feature sets (F1 and F2) were based on the derived segmentations (S1 and S2). The spectral features were extracted for each applied window size employing those pixel within a window which carried the same segment label as the centre pixel of the window. All pixels of a segment were not used because the plot-level estimates were not considered to be representative at the stand level. Thus, the employment of all the pixels of a segment may have led to excessive rms errors. The performance of the segment-based features was compared to that of features extracted from all pixels in the local neighbourhood of the plot (F3) (see figure 6).

The estimation tests were carried out applying a squared inverse distance-weighted *k-nn*-estimator (Tomppo 1990, 1996, Altman 1992, Tokola *et al.* 1996) and the described suboptimal features (Mäkisara *et al.* 1997). The cross-validation technique was applied. In this technique, every plot was in turn left out of the dataset and its characteristics were estimated with the aid of the other plots. Spectral k-nearest neighbours, determined by means of the Euclidean distance between plots in the feature space, were applied. Each of the k-nearest neighbours received a weight which was in proportion to its squared inverse Euclidean distance in the suboptimal feature space.

The evaluation was based on the rms error and relative standard error (s_e) of the estimates obtained with the cross-validation (Equations 4 and 5). The applied number k was 10, determined by running the cross-validation with k values from 1 to 30 using the feature set F3 and applying a 3×3 pixel window. These tests showed that the rms error decreases with increasing k (figure 7). The rate of this decrease began

Table 3. Suboptimal features used in the estimation tests. Letters *a* and *s* refer to the average and standard deviation respectively, thus notation *s*21 means standard deviation of channel 21.

Variable	Feature set
Pine (*Pinus sylvestris*)	*s*21 *a*20 *a*30 *s*14 *a*26 *s*2
Spruce (*Picea abies*)	*a*17 *a*1 *a*13 (*s*4 − *s*5) (*s*1 − *s*2) (*a*3 − *a*4)
Deciduous (*Betula* spp., *Alnus* spp.)	*a*19 *a*9 *s*19 *s*11
Total	*a*1 (*a*28 + *a*29 − *a*30)

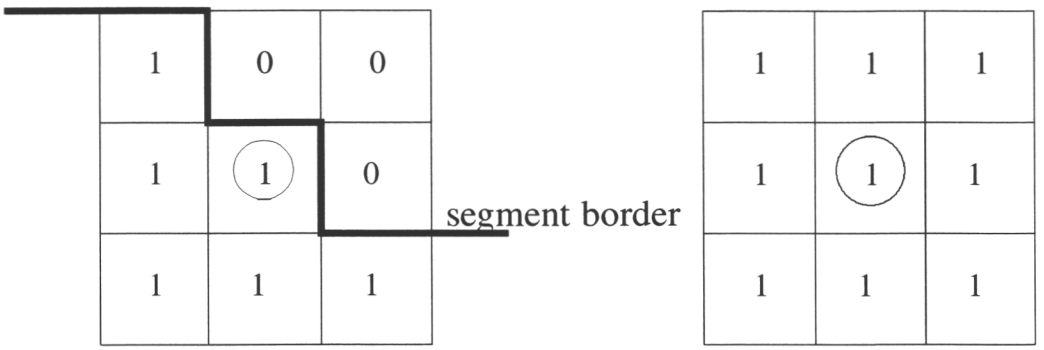


Figure 6. Example of the weighting employed in feature extraction. In the extraction of feature sets F1 and F2, only those pixels which belonged to the same segment as the centre pixel of the window, were used. The location of the field plot has been marked with a circle. In the case of F3, all the pixels within the window were employed. The length of the window side in the example is three pixels.

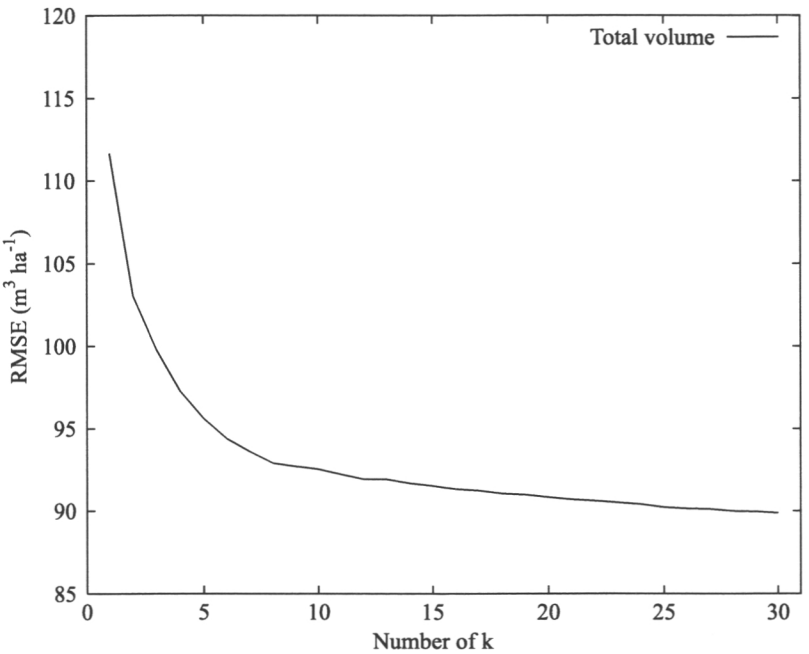


Figure 7. The effect of k on the rms error of the cross-validation estimates using F3. The spectral features were extracted as an average of the pixels within a 3×3 pixel window around the plot.

to slow down with k values >5. This observation is similar to the one in some earlier studies applying Landsat TM material (Tokola *et al.* 1996, Trotter *et al.* 1997).

$$rms\ error = \sqrt{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 / n} \tag{4}$$

where n = number of observations, y_i = ith observed value of y, and $\hat{y}_i$ = estimated value of y_i.

$$s_e = \frac{rms\ error}{\bar{y}} \tag{5}$$

where $\bar{y}$ = mean of observed values y_i.

4. Results

Examples of the results of both of the derived segmentations are shown in figures 8 and 9. Both S1 and S2 consisted of quite small segments: the average segment size of S1 was 498 pixels (0.127 ha) and that of S2 was 296 pixels (0.076 ha).

Some objects which consisted of several small segments in S2, were more clearly detected in S1. Examples of these objects included clear-cut areas and roads. This difference was due to the coordinate input channels applied in the low-level segmentation of S1. The use of coordinates smooths local differences because two out of the five input channels were locally very similar. In areas which were spectrally quite homogeneous, this smoothing was large enough to force neighbouring elements into the same cluster.

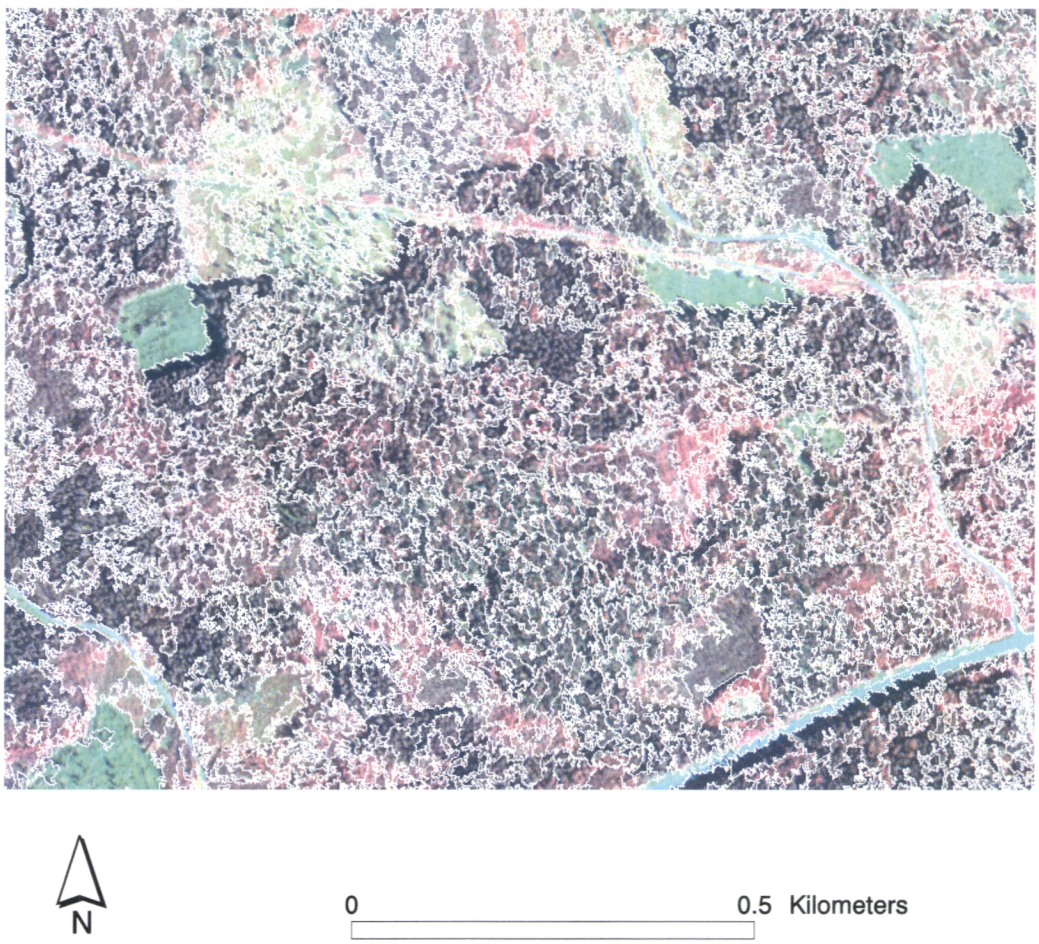


Figure 8. Example of segmentation based on the first three PCs and row and column image coordinates of the original AISA image.

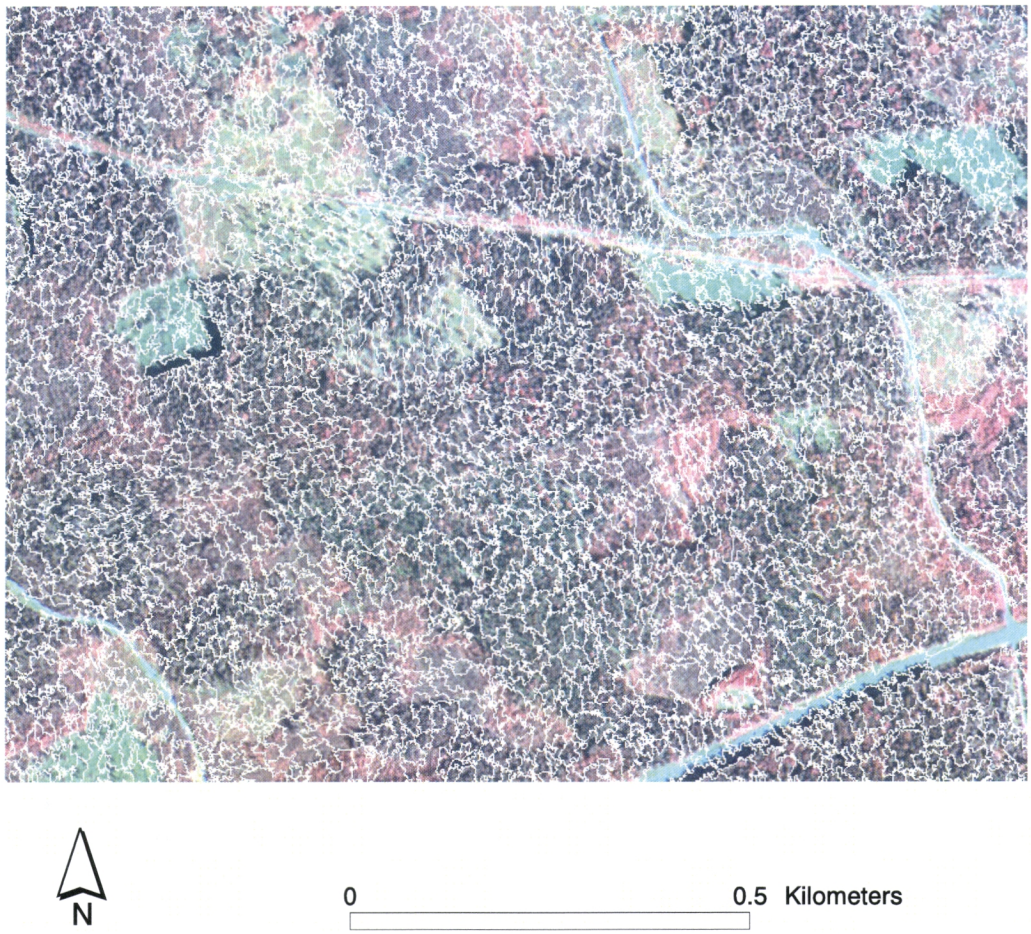


Figure 9. Example of segmentation based on the first three PCs of the original AISA image.

The best results of the cross-validation based *k-nn* estimation test applying a *k* value of 10 to each derived feature set (F1, F2 and F3) are shown in tables 4 and 5. The window sizes which resulted in the best estimate of each feature set are given in table 6. Set F1 yielded the best estimates for the total volume and volume of deciduous species, and set F3 was best for volumes of pine and spruce. The window sizes which resulted in the best estimates varied according to the method. The best results for set F1 were achieved when employing larger window sizes than in the case of F2 and F3. The effect of the window size on the performance of each feature set is illustrated in figures 10–12.

Table 4. Root mean square (rms) errors ($m^3\ ha^{-1}$) of the best estimates by feature sets and species. The best overall estimates are shown in **bold**.

Feature set	Total	Pine (*Pinus sylvestris*)	Spruce (*Picea abies*)	Deciduous (*Betula* spp., *Alnus* spp.)
F1	**70.0**	41.8	68.3	**21.1**
F2	72.7	42.5	70.2	21.9
F3	71.7	**40.9**	**64.5**	21.4

Table 5. Relative stardard errors (s_e) of the best estimates by feature sets and species. The best overall estimates are shown in **bold**.

Feature set	Total	Pine (*Pinus sylvestris*)	Spruce (*Picea abies*)	Deciduous (*Betula* spp., *Alnus* spp.)
F1	**0.61**	1.47	0.97	**1.27**
F2	0.63	1.49	1.00	1.32
F3	0.62	**1.44**	**0.92**	1.29

Table 6. The length of the window side (pixels) giving the smallest rms error in the estimation tests by feature sets and species. The length of a pixel side is 1.6 m. The sizes of the windows applied in the feature extraction giving the best overall estimate have been shown in **bold**.

Feature set	Total	Pine (*Pinus sylvestris*)	Spruce (*Picea abies*)	Deciduous (*Betula* spp., *Alnus* spp.)
F1	**49**	111	21	**23**
F2	11	47	15	17
F3	19	**11**	**13**	13

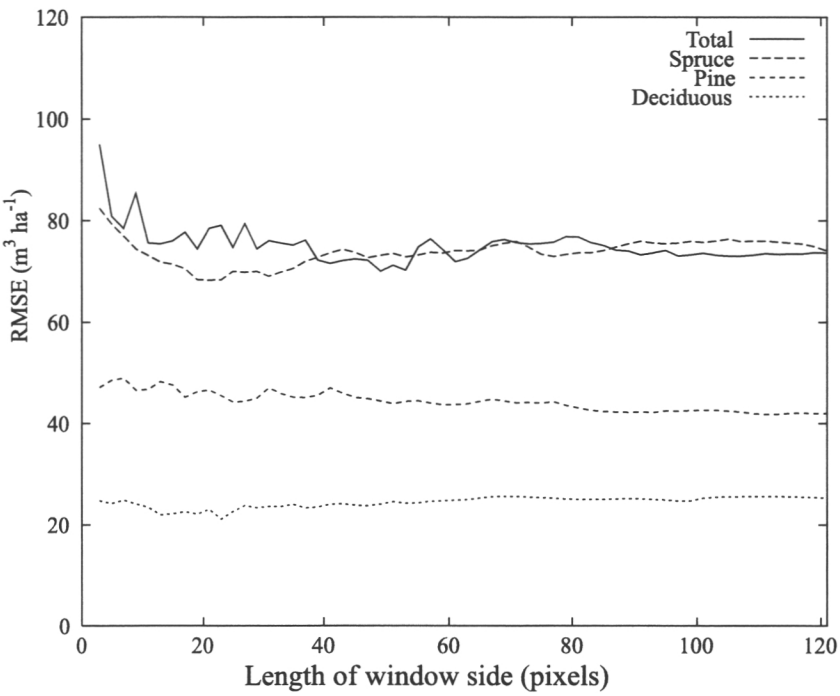


Figure 10. Root mean square (rms) error of cross-validation tests using F1.

Set F1 gave slightly better results than set F2 for all estimated variables. The estimates derived from set F1 were also more stable than those of set F2 with respect to different window sizes. The rms errors of the estimates remained lower even if the window size was increased. This was especially so in the case of the estimates of

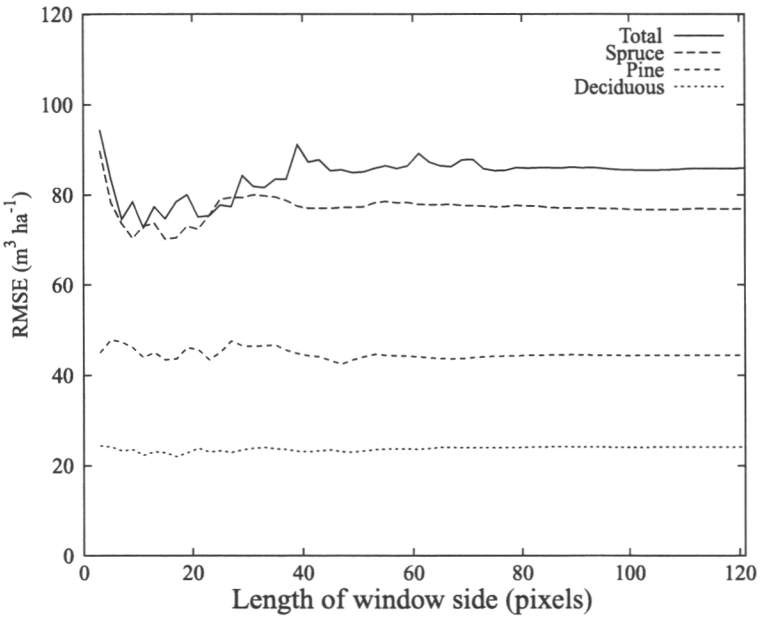


Figure 11. Root mean square (rms) error of cross-validation tests using F2.

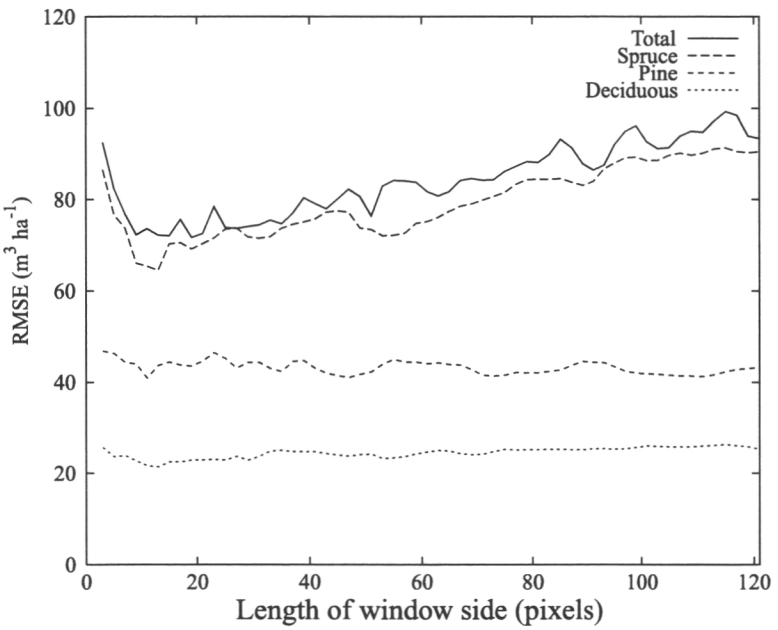


Figure 12. Root mean square (rms) errors of cross-validation tests using F3.

total volume and the volume of spruce. In fact, if the length of the side of the extracting window was increased beyond about 25 pixels, the estimation errors increased if S2 was applied. This suggests that the large segments of S2 were not as applicable as those of S1. Note also, that the rms errors derived from set F2 changed only slightly after the length of the window size had reached about 80 pixels. This was due to the small average segment size; the majority of segments of S2 fit completely within the applied 80×80 pixels window.

The rms errors of the total volume and volume of spruce were of the same magnitude with all the applied feature sets. This was due to the fact that the stands of the study area were dominated by spruce. Thus, most of the total volume consisted of the volume of spruce. In general, the rms errors of the estimation tests were high and the use of image segments in the feature extraction had no significant effect on the performance of the estimation. The relative standard errors were more than 60% for all estimates. Only the total volume and volume of spruce could be estimated at a relative error of less than 100%.

5. Discussion

The co-occurrence segmentation method (COS) produces segments which are applicable in multisource forest inventory and monitoring applications. It can be applied to the derivation of image analysis units for tasks such as the estimation of stand-level forest parameters. In this study the use of derived segments in the spectral feature extraction did not significantly improve the performance of the features in the estimation of plot-level forest characteristics. This was partly due to the segmentation result, as well as to the field data employed in the estimation tests.

The result of the COS depends on the input data, parameters and clustering method employed in the the low-level segmentation, and the minimum size parameter of the co-occurrence region merging (CRM) step. In this study, the ISODATA algorithm was employed in the derivation of low-level segments, but any other clustering algorithm could have been used. The ISODATA was chosen because it is a common algorithm which is available in many commercial remote sensing related software packages.

The use of row and column image coordinates as additional input information in the clustering process improved the segmentation result in some of the spectrally more homogeneous areas. However, the weight given to the coordinate channels was slightly too large. Consequently, some of the obvious segment borders were lost. When the coordinate channels were not employed, the sizes of the resulting segments were more evenly distributed and even some of the homogeneous areas were divided into different segments. This implies that the applied minimum segment size, 121 pixels (≈ 0.03 hectares), was too small. On the other hand, the fundamental assumptions behind the CRM rely on the small-scale variation pattern typically present in the forests. Thus, the increment in the minimum size parameter may lead to a situation where these assumptions are no longer valid.

One reason why the employment of the segment-based feature sets failed to improve the estimation results may be the type of field plots employed. The plot-level estimation result with this kind of field data and Landsat TM images have also resulted in high rms errors (Tokola *et al.* 1996, Tokola and Heikkilä 1997). With Landsat TM material, this has been attributed to the small size of the collection unit of the training data compared to the spatial resolution of TM images (Tokola *et al.* 1996). Even though this may apply to image material with coarser spatial resolution, it does not apply to VHR images. A more probable reason for the poor estimates is the fact that the relascope plots are very local in nature and some of the recorded forest characteristics, e.g. basal area and the volume of growing stock, may be very different from those a few metres from the plot centre. Thus, even very small errors in the image registration or location of the plot may result in an erroneous combina-

tion of spectral and informational classes in the training data. On the other hand, the use of relascope plots is an efficient way to collect field data for traditional forest inventory purposes especially if the area to be inventoried is large. Furthermore, if the errors in the image registration and location of the plots are small, most of the plots of the training data will carry relevant information. Thus, the plots carrying misleading information may be considered as 'noise' in the training data. Therefore, instead of rejecting relascope plots as a collection unit for the training data, methods for removing the noise and stabilizing the training data should be developed.

In addition to the error sources mentioned, another reason for the fact that the differences in the performance of the applied feature sets in this study were small is the preprocessing of the field plot data. The way in which the locations of the plots intersecting several stands were modified improved the performance of set F3 because with small window sizes all the pixels inside the extracting window were within one stand.

Even though the segment-based feature sets did not improve the plot-level estimation results in this study, the use of COS or some other segment-based approach can be recommended if the use of VHR images is considered. However, most of the segmentation methods presented in the machine vision literature are not suitable for this task. Image segmentation is usually employed in tasks where an object needs to be recognized, and isolated from its background. The properties of an object are usually well known and in many cases at least some control over the imaging environment is possible. Unfortunately, this is not the case in most of the multisource forest inventory and monitoring applications. The object of interest may vary from, for example, a forest stand to a single tree. Furthermore, the spectral properties of the objects within an image may differ. Therefore, the role of image segmentation in forestry-related inventory and monitoring applications is rather the determination of meaningful units for further analysis than the actual recognition or isolation of real-world objects.

The COS is a tool for the segmentation of VHR images of forested landscapes. Its main advantages are that it can employ many kinds of input data and that it can combine neighbouring segments having spectrally different properties. The method can be employed, for example, in the determination of units for feature extraction, but is is not suitable for automated stand delineation purposes. For such a problem, the result of COS must be further processed by means of a region merging algorithm based on spectral similarity or some other properties of the segments.

Acknowledgments

This work was funded by the Finnish Forest Research Institute and the 'Forests in GIS' graduate school. Finnish Forest Research Institute provided all the material and equipment applied in this study. I would like to thank the institute's National Forest Inventory team for collecting and preprocessing the field plot and image data. I also thank Professor Erkki Tomppo and Dr Jari Varjo for their review comments, and Tech. Lic. Kai Mäkisara for practical implementation tips.

References

Afifi, A. A., and Clark, V., 1984, *Computer Aided Multivariate Analysis* (Belmont, CA: Lifetime Learning Publications).

Altman, N., 1992, Introduction to kernel and nearest-neighbour nonparametric regression. *The American Statistician*, **46**, 175–184.

ERDAS, 1999, *Erdas FieldGuide™* (Atlanta, Georgia, USA: Erdas Inc.).

FU, K., and MUI, J., 1981, A survey of image segmentation. *Pattern Recognition*, **13**, 3–16.

HAGNER, O., 1990, Computer aided forest stand delineation and inventory based on satellite remote sensing. *Proceedings of SNS/IUFRO Workshop in Umeå , 26–28 February 1990: The usability of remote sensing for forest inventory and planning* (Umeå : Swedish University of Agricultural Sciences. Remote Sensing Laboratory).

HÄME, T., 1991, Spectral interpretation of changes in forest using satellite scanner images. *Acta Forestalia Fennica*, **222**, 111.

HARALICK, R., and SHAPIRO, L., 1985, Survey: image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, **29**, 100–132.

HARALICK, R., SHANMUGAM, K., and DINSTEIN, I., 1973, Textural features for image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligency*, **SMC-3**, 610–621.

HÄME, T., HEILER, I., and SAN MIGUEL-AYANZ, J., 1998, An unsupervised change detection and recognition system for forestry. *International Journal of Remote Sensing*, **19**, 1079–1099.

HOLMGREN, P., and THURESSON, T., 1998, Satellite remote sensing for forestry planning—a review. *Scandinavian Journal of Forestry Research*, **13**, 90–110.

HOLOPAINEN, M., 1998, Forest habitat mapping by means of digitized aerial photographs and multispectral airborne measurements. PhD thesis. Department of Forest Resource. University of Helsinki.

HOLOPAINEN, M., and WANG, G., 1998, The calibration of digitized aerial photographs for forest stratification. *International Journal of Remote Sensing*, **19**, 677–696.

JAIN, R., KASTURI, R., and SCHUNCK, B. G., 1995, *Machine Vision* (New York: McGraw-Hill International Editions).

LAASASENAHO, J., 1982, Taper curve and volume functions for pine, spruce and birch. *Communicationes Instituti Forestalis Fenniae*, **108**, 1–74.

MÄKISARA, K., 1998, AISA data user's guide. Technical Report 1894. Technical Research Centre of Finland (VTT). ISBN 951-38-5281-4 ISSN 1235-0605.

MÄKISARA, K., HEIKKINEN, J., HENTTONEN, H., TUOMAINEN, T., and TOMPPO, E., 1997, Experiments with imaging spectrometer data in large area forest inventory context. *Proceedings of the Third International Airborne Remote Sensing Conference and Exhibition, Copenhagen, Denmark* (Piscataway, New Jersey, USA: IEEE), pp. 420–427.

MÄKISARA, K., MEINANDER, M., RANTASUO, M., OKKONEN, J., AIKIO, M., SIPOLA, K., PYLKKÖ, P., and BRAAM, B., 1993, Airborne imaging spectrometer for applications (AISA). *Proceedings of the 1993 International Geoscience and Remote Sensing Symposium (IGARSS'93), Tokyo, Japan* (Piscataway, New Jersey, USA: IEEE), pp. 479–481.

NARENDRA, P., and GOLDBERG, M., 1980, Image segmentation with directed trees. *IEEE Transactions on Pattern Analysis and Machine Intelligency*, **PAMI-2**, 185–191.

PAL, N. R., and PAL, S. K., 1993, A review on image segmentation techniques. *Pattern Recognition*, **26**, 1277–1294.

PARMES, E., 1992, Segmentation of spot and landsat satellite imagery. *The Photogrammetric Journal of Finland*, **13**, 52–58.

TOKOLA, T., and HEIKKILÄ, J., 1997, Improving satellite image based forest inventory by using a priori site quality information. *Silva Fennica*, **31**, 67–78.

TOKOLA, T., PITKÄNEN, J., PARTINEN, S., and MUINONEN, E., 1996, Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. *International Journal of Remote Sensing*, **17**, 2333–2351.

TOMPPO, E., 1987, Stand delineation and estimation of stand variates by means of satellite images. Remote sensing aided forest inventory. University of Helsinki. Department of forest mensuration and management. Research notes, pp. 60–76.

TOMPPO, E., 1990, Satellite image-based national forest inventory of Finland. *Photogrammetric Journal of Finland*, **12**, 115–120.

TOMPPO, E., 1996, Multi-source national forest inventory of finland. New thrusts in forest inventory. EFI proceedings. number 7, pp. 27–41.

TOMPPO, E., HENTTONEN, H., KORHONEN, K. T., AARNIO, A., AHOLA, A., HEIKKINEN, J., IHALAINEN, A., MIKKELÄ, H., TONTERI, T., and TUOMAINEN, T., 1998, Eteläpohjanmaan metsäkeskuksen alueen metsävarat ja niiden kehitys 1968–97 (in Finnish). *Metsätieteen aikakauskirja—Folia Forestalia*, **2B/1998**, 293–374.

TROTTER, C. M., DYMOND, J. R., and GOULDING, C., 1997, Estimation of timber volume in a coniferous plantation forest using landsat tm. *International Journal of Remote Sensing*, **18**, 2209–2223.

VARJO, J., 1997, Calibration and change detection method for controlling continuously updated forest information by landsat tm material. *Acta Forestalia Fennica*, **258**, 1–64.

WOODCOCK, C., and HARWARD, V. J., 1992, Nested-hierarchical scene models and image segmentation. *International Journal of Remote Sensing*, **13**, 3167–3187.

Paper III

Reprinted from Remote Sensing of Environment, vol 82, Pekkarinen, A., Image segment-based spectral features in the estimation of timber volume, pp 349-359, copyright (2002), with permission from Elsevier.

III

ELSEVIER

Remote Sensing of Environment 82 (2002) 349–359

Remote Sensing of Environment

www.elsevier.com/locate/rse

Image segment-based spectral features in the estimation of timber volume

Anssi Pekkarinen *

Finnish Forest Research Institute, Unioninkatu 40 A, FIN-00170 Helsinki, Finland

Received 22 October 2001; received in revised form 5 April 2002; accepted 13 April 2002

Abstract

Plot- and stand-level errors associated with satellite image-based multisource forest inventory (MSFI) applications have been relatively high. The reasons suggested for that are related to the limited spatial resolution of the image material. The introduction of very high spatial resolution (VHR) images to MSFI applications should, therefore, diminish these errors. The use of VHR images is, however, problematic, because pixel-by-pixel analysis methods are no longer applicable. The paper presents an image segment-based approach to the determination of feature extraction and image analysis units. The study was carried out in Southern Finland and employed a spectrally averaged imaging spectrometer (AISA) image and field data gathered from sample plots. A two-phase segmentation method was applied and a large number of segment-based spectral features was extracted and used as input to a feature selection procedure. Forward selection based on an improvement of RMSE was applied. The performance of segment-based features (SF) was compared to that of reference features (RF) extracted from square-shaped windows. The estimation results revealed that even though the applied segmentation method succeeded well in the determination of units of feature extraction and image analysis, the differences between the performance of SF and RF were small and the plot-level estimation errors remained high. The study suggests that large estimation errors are due to the local nature of the field data and may be diminished using data that is representative at the segment level.

1. Introduction

Satellite remote-sensing-aided multisource forest inventory (MSFI) methods have two major advantages when compared to traditional forest inventories based solely on field data. First, MSFI makes it possible to compute reliable forest statistics for smaller areas than is possible with only sparse field samples. Secondly, MSFI produces thematic maps for any area of interest covered by the imagery employed (Franco-Lopez, Ek, & Bauer, 2001). The fact that MSFI can be easily integrated into an existing inventory scheme, by employing the field plots of traditional inventories as training data in MSFI, makes it even more appealing. This kind of approach has been applied in, e.g. the National Forest Inventory (NFI) of Finland (Katila & Tomppo, 2001; Tomppo, 1996).

Although satellite image-based MSFI methods have been successfully applied to the estimation of forest characteristics of large- and medium-size areas in different conditions (Franco-Lopez et al., 2001; Tomppo, 1996; Tomppo, Goulding, & Katila, 1999; Tomppo, Korhonen, Heikkinen, Yli-Kojola, 2001; Trotter, Dymond & Goulding, 1997), their general applicability to forest management planning has been questioned (Holmgren & Thuresson, 1998). These doubts are supported by the fact that the stand and sample plot-level estimation errors of the multisource inventory approaches have been rather high (Katila & Tomppo, 2001; Mäkelä & Pekkarinen, 2001; Poso, Paananen, & Similä, 1987; Tokola, Pitkänen, Partinen, & Muinonen, 1996).

Several reasons for high stand- and plot-level estimation errors have been suggested. In applications employing relatively small, circular or relascope, sample plots and Landsat TM images, the sampling area is typically smaller than the pixel size, which may adversely affect the image analysis (Katila & Tomppo, 2001; Poso et al., 1987; Tokola et al., 1996). In addition, errors in the location of the sample plots and georeferencing of the satellite images are still potential sources of error (Katila & Tomppo, 2001; Tokola et al., 1996) although a method for relocating the plots relative to the georeferenced satellite imagery and reducing the error has been recently presented (Halme & Tomppo,

* Fax: +358-9-625-308.
E-mail address: anssi.pekkarinen@metla.fi (A. Pekkarinen).

0034-4257/02/$ - see front matter © 2002 Elsevier Science Inc. All rights reserved.
PII: S0034-4257(02)00052-4

2001). Furthermore, plots located near the stand boundaries introduce errors to the analysis (Poso et al., 1987; Tokola et al., 1996).

The problems mentioned above are related to the limited spatial resolution of the image material employed. The introduction of increasingly available very high spatial resolution (VHR) image material, such as digital and digitised aerial photographs, and images of imaging spectrometers and VHR satellites, should, therefore, provide a way to overcome these problems. However, the use of VHR material is problematic because the pixel-based feature extraction methods and traditional pixel-by-pixel image analysis methods are no longer applicable. This is due to the fact that a pixel of a VHR image does not necessarily represent the object of interest (stand, tree), but may cover only a small part of it. Therefore, special attention has to be paid to the determination of the feature extraction and image analysis units.

Two alternatives for feature extraction have been usually considered in MSFI applications employing satellite images and field data gathered from sample plots. Features have been extracted from a single pixel, on which the plot is located, or from its local neighbourhood. In applications employing VHR image material, the choice is obvious; in order to extract features that describe the area of a sample plot and/or its surroundings, a neighbourhood of pixels has to be applied.

Probably the simplest approach to the local feature extraction is to extract the features from square-shaped windows surrounding the field plots. The approach is efficient, but has a drawback that becomes increasingly serious as the size of the extraction window increases: the window may intersect several stands having different spectral and forest characteristics. Therefore, the extracted features may not describe the forest characteristics of the plot on which the centre of the extracting window was located.

One possible solution to this problem is to apply an image segment-based approach. In an ideal case, image segmentation succeeds in the delineation of stands and, thus, the within-segment forest characteristics are homogeneous. Thus, the spectral characteristics of a stand could be extracted by segments and directly employed in the image analysis. However, the within-stand variation of some forest characteristics may be large and information gathered from a sample plot is not necessarily representative at the stand (segment) level. Spectral features that also describe the immediate neighbourhood of the plot centre should, therefore, be employed if plot-based field data is used.

Considering a typical image segmentation method that starts from initial segmentation and proceeds by merging the initial areas into larger units, a natural way to extract the features is to apply initial segments in the extraction of local features and derive larger-scale features from larger segments. An approach lying between these two methods is that of restricted segment-based feature extraction, in which features are extracted from homogeneous parts of local windows (Mäkelä & Pekkarinen, 2001; Mäkisara, Heikkinen, Henttonen, Tuomainen, & Tomppo, 1997; Pekkarinen, 2002).

One drawback in any approach employing multiple-scale features is that the number of features increases and places an additional computational load on the estimation process. In addition, the performance of the different features may be very different and irrelevant features may confuse the image analysis. An approach to this problem is to select a relevant subset of features to be employed in the estimation. In this study, a forward selection procedure starting with an empty feature subset and adding features one by one was applied. The selection of features was based on improvement of the global RMSE.

The aim of this study was to determine whether the errors of MSFI estimates can be reduced using segment-based spectral features instead of features extracted in a more straightforward way. The extracted features included spectral averages and standard deviations. The segment-based feature set consisted of the features derived from three different segmentations and the reference features (RF) were extracted from local square-shaped windows. The actual estimation tests were carried out using a selected subset of the extracted features, and a k nearest neighbour (k-NN) estimator. The global accuracy of the estimates was evaluated using RMSE and the within volume class accuracy was studied with the help of confusion matrices.

2. Material

2.1. Field data

This study was carried out using a previously established test area in Ohkola in the municipality of Mäntsälä, Southern Finland (Fig. 1). The test area was originally established for imaging spectrometer studies, and the field data was

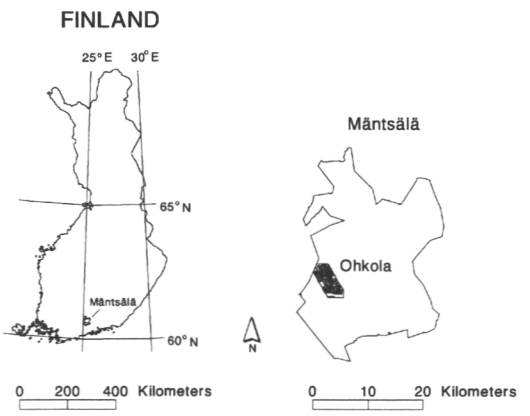


Fig. 1. Location of the municipality of Mäntsälä within the borders of Finland and the location of the Ohkola test area within the borders of Mäntsälä.

gathered and processed by the National Forest Inventory (NFI) Group of the Finnish Forest Research Institute (FFRI). For details about the field data, see Mäkisara et al. (1997).

The main tree species of the study area are Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies*), birch (*Betula* spp.) and aspen (*Alnus* spp.). For the purposes of this study, the tree species were classified into three groups: pine, spruce and deciduous species.

The test area was covered by a systematic grid of 254 field plots. The distance between the plots was 250 m in both north–south and east–west directions and they intersected 329 stands. The locations of the plots were determined with help of a compass and measuring tape. Circular sample plots with a fixed radius of 12.52 ms were employed. Every seventh tree, sampled with help of a relascope, was selected as a sample tree, which was then measured in more detail than the tally trees. All measurements were carried out from July to September 1996 (Mäkisara et al., 1997).

The volumes of sample trees were determined with help of volume functions of Laasasenaho (1982) using the following predicting variables: diameter at breast height ($d_{1.3}$), diameter at height of 6 m ($d_{6.0}$) and height. The volumes of tallied trees were estimated with the help of estimated volumes of sample trees using mixed models determined by tree species groups. The variables employed in the estimation were $d_{1.3}$, basal area of the stand and some other stand characteristics. Plot volume per area unit (V, m^3/ha) was determined using the total volume of the trees of the sample plot and the area of the plot. If the plot intersected several stands, the volumes of trees of each stand were summed and divided by the area of the plot belonging to that particular stand (Mäkisara et al., 1997). Subplots that had a proportional area smaller than 0.5 were removed from the study material because their volume estimates were considered to be unreliable. Plot-level characteristics of the resulting input data set consisting of the remaining 262 field plots are given in Table 1.

The locations of the plots were verified by overlaying each plot on the averaged imaging spectrometer (AISA) image and checking its location on the screen. The ground-truth information of the plot (or subplot) was assigned to an artificial plot that was established clearly inside the stand where the plot (or subplot) was located near the stand border (Mäkisara et al., 1997).

Table 1
Plot-level timber volume characteristics (m^3/ha) of the input data set and the percentage of pure stands by tree species

Species	Mean	Std.	Min.	Max.	Pure stands (%) *
Pine	28.5	44.7	0	220.2	17.4
Spruce	70.3	99.5	0	397.6	20.7
Deciduous	16.6	25.2	0	142.7	7.5
Total	115.3	103.3	0	408.8	–

* Percentage of plots on which at least 90% of the total volume of the plot consists of the particular species of all plots having total volume>0 m^3/ha.

Table 2
The original spectral channel (Ch_o) configuration of AISA and the channels applied in the generation of new (Ch_g) blue (B), green (G), red (R) and near-infrared (NIR) channels

Ch_o	Ch_g	Centre (nm)	Width (nm)
1	B	470.1	7.3
2	B	487.6	7.3
3	B	505.1	7.3
4	G	522.6	7.3
5	G	541.5	7.3
6	G	550.2	7.3
7	G	564.8	7.3
8	G	579.4	7.3
9	G	599.8	7.3
10		623.1	7.3
11	R	646.7	7.6
12	R	668.0	7.6
13	R	675.6	7.6
14	R	686.2	4.6
15		696.8	4.6
16		701.4	4.6
17		710.5	4.6
18		725.7	7.6
19		733.3	4.6
20		742.4	4.6
21		750.0	4.6
22	NIR	775.8	4.6
23	NIR	784.9	4.6
24	NIR	794.1	4.6
25	NIR	803.2	4.6
26	NIR	818.4	4.6
27	NIR	844.2	4.6
28	NIR	853.3	4.6
29	NIR	860.9	4.6
30	NIR	865.4	4.6

2.2. Image data

An imaging spectrometer (AISA) image of the test area was acquired on June 23, 1996. The image mosaic covering the study area consisted of data from seven flight lines and 30 spectral channels. The subset image used in this study was 3500 × 3750 pixels and the pixel size was 1.6 × 1.6 m^2. The image data was radiometrically and geometrically corrected by the NFI team at the Finnish Forest Research Institute.

The number of spectral channels was reduced from 30 to 4 by means of spectral averaging in order to reduce the computation time needed in the image-processing steps. The original channels employed in the averaging were selected so that the wavelength areas of the resulting blue (B), green (G), red (R) and near-infrared (NIR) bands would correspond as well, as possible, to the bands of the new generation VHR satellite images. The averaged pixel values were computed as weighted averages of original pixel values. Each of the original channels was weighted by its width.

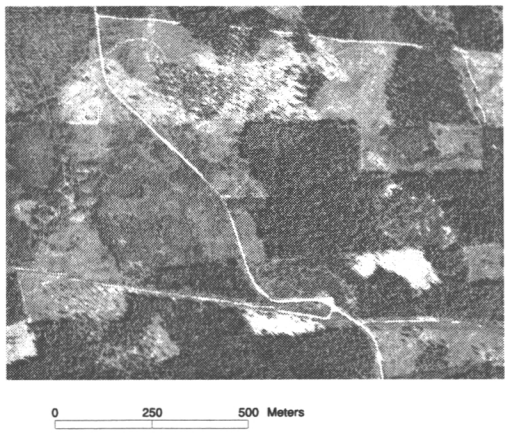


Fig. 2. An example of spectrally averaged AISA data. Channel 3 (Red).

The pixel values of the spectrally averaged image were linearly scaled to a range of 0 to 255 with the use of the global minimum and maximum values of the averaged channels. The original spectral configuration of the AISA channels and the channels employed in the spectral averaging are presented in Table 2 and an example of the averaged image is given in Fig. 2.

3. Methods

3.1. Image segmentation

Prior to the segmentation, the spectrally averaged image was smoothed with help of a Gaussian kernel (see, e.g. Jain, Kasturi, & Schunck, 1995). An iterative approach to the smoothing was applied using a 3×3 pixel window and 10 sequential passes over the image. After that, a two-phase segmentation approach was applied to the image. First, initial segments were created by a method (INISEG) based on the "Image segmentation with directed trees" (Narendra & Goldberg, 1980) and, after that, the final segments were composed with the help of a region merging (REGMERG) algorithm.

The original idea of the "Image segmentation with directed trees" algorithm was to detect regions without using absolute thresholds (Narendra & Goldberg, 1980). The absolute thresholds were avoided with the help of an edge image which was determined using an edge operator and the original image data. The edge image was used in computing the maximum difference in the edge value between a pixel and its neighbours. After that, the pixels of the resulting edge gradient image were divided into plateau and non-plateau elements with help of the edge gradient value and a given threshold. Plateau and non-plateau elements represented areas of small and large local variation, respectively. All non-plateau pixels were linked to the direction giving maximum positive edge gradient. If no such direction existed, the pixel was assigned as a root. After the processing of the non-plateau pixels, plateau elements were arbitrarily linked to the pixels of the same plateau. The links were tracked and the formation of directed cycles was prohibited. If an allowed neighbouring pixel was not found, the plateau pixel was labelled as a root pixel. Finally, all the root pixels were given unique labels and the pixels of each directed tree were given the label of the root pixel of that tree (Narendra & Goldberg, 1980).

$$e(i,j) = \sum_{k=1}^{n} \left(\sum_{(l,m) \in N_8(i,j)} |p_k(i,j) - p_k(l,m)| \right), \quad (i,j) \in I \tag{1}$$

where n = the number of channels.

$$G(i,j) = \max_{(l,m) \in N_8(i,j)} [e(i,j) - e(l,m)], \qquad (i,j) \in I \tag{2}$$

The initial steps of INISEG are somewhat similar to the implementation presented by Narendra and Goldberg (1980). INISEG begins by dividing an image I to edge and plateau pixels using edge (e) and directed maximum edge (G) operators and a threshold value (T). Edge value e is defined for each image pixel $(i,j) \in I$ with help of its eight spatially nearest neighbours $N_8(i,j)$ and the original pixel values $p(i,j)$ of the utilised n channels(Eq. (1)). The resulting edge image is used as an input to the directed maximum edge operator (Eq. (2)).

The pixels are classified into plateau, root and edge pixels using Algorithm 1. If a pixel (i,j) is an edge pixel, it is linked to the direction giving $G(i,j)$.

Algorithm 1. *Classification of pixels of edge image.*

```
if |G(i,j)| ≤ T then
   (i,j) is a plateau pixel
else
   if G(i,j) < 0 then
      (i,j) is a root pixel
   else
      (i,j) is an edge pixel
   endif

endif
```

The next step of INISEG is the actual image segmentation and it differs from the method presented in Narendra and Goldberg (1980) as the formation of segments is implemented using sequential connected component labelling (CCL) algorithm (see, e.g. Jain et al., 1995). CCL is usually applied in labelling spatially continuous areas having similar pixel values. Thus, it can be applied in turning

the result of clustering into segmentation. Here, however, the CCL is used in a different way and (i,j) and its neighbouring pixel $(l,m) \in N_8(i,j)$ are connected if:

A: (i,j) is an edge pixel AND,

(l,m) is an edge pixel and linked to (i,j) OR,
(i,j) is linked to (l,m) OR,
(l,m) is a plateau pixel OR a root pixel and $| e(i,j) - e(l,m) | \leq T$;

B: (i,j) is a plateau pixel OR a root pixel AND,

(l,m) is a plateau pixel OR,
(l,m) is a root pixel OR,
(l,m) is an edge pixel and linked to (i,j),
(l,m) is an edge pixel and $| e(i,j) - e(l,m) | \leq T$.

The sequential CCL requires two passes over the image. The first pass begins with the upper left pixel of the image and proceeds from left to right and from top to bottom. The first pixel is labelled '1'. In the case of the rest of the pixels, the segment label is determined as follows: if the pixel being processed is connected to any of the already processed pixels, it is given the label of that pixel. If such a pixel does not exist, a new unique label is created for it. If a pixel being processed is connected to more than one already processed pixels having different labels, the largest of the labels is assigned to it and the equivalency of the selected label and labels of other connected pixels is recorded.

After processing each row, the labels of the pixels of that row are written in a temporary file. In the second pass, all the labels recoded as equivalent are given common unique labels. After that, each of the pixels of the temporary file is renumbered with the new labels and the final output is composed. An example of the result of the initial segmentation derived with a threshold value $T = 1.5$ is shown in Fig. 3.

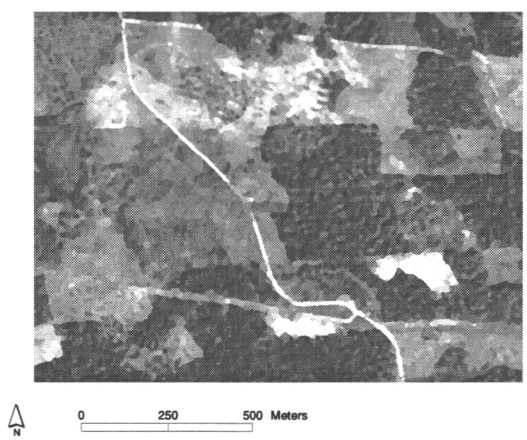


Fig. 3. An example of initial segment averages of the red channel of spectrally averaged AISA data.

Table 3
Parameters employed in the segment derivation employed in the feature extraction

Result	Input image	Minimum segment size	t-Ratio threshold
SA	Initial segmentation	10 pixels	0
SB	SA	10 pixels	24
SC	SB	10 pixels	40

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{3}$$

where $\bar{x}_m$ = the mean intensity of the segment m, $m = 1, 2$; s_m^2 = the variance of the segment m, $m = 1, 2$; n_m = number of pixels in the segment m, $m = 1, 2$.

Segmentation methods based on local edge detection are sensitive to image noise and the high local variation present in the VHR images. The initial segments, therefore, need to be post-processed in order to remove segments resulting from irrelevant small-scale variation and so produce relevant feature extraction and image analysis units. A method close to the 't-ratio segmentation' (Hagner, 1990) was applied in the present investigation. The iterative algorithm sought segments smaller than a specified minimum size and combined them to the most similar neighbouring segment. In addition, if two spectrally similar neighbouring segments were observed, they too were combined. The similarity of the adjacent segments was determined for each channel by means of the t-ratio (Eq. (3)) and summed over the channels employed. Note that if the summed t-ratio would have been used in statistical testing of the significance of the difference of the mean vectors of two segments, it would have required the channels to be uncorrelated. Here, however, the t-ratio was employed simply as a measure of similarity.

3.2. Feature extraction and selection

Three different segmentations (SA, SB, SC) were derived for the feature extraction. SA, SB and SC and their parameters are presented in Table 3. The features extracted from each segment included spectral averages and standard deviations of each of four input channels of the spectrally averaged AISA image. An example of SC is presented in Fig. 4.

In addition to segment-based features, a local neighbourhood approach to the feature extraction was applied. Pixels belonging to the same segment as the plot pixel were employed in the segment-restricted local feature extraction. The size of the extraction window was 31×31 pixels. The

size of the extraction window was chosen in such a way that each window included at least 10 pixels belonging to the same segment as the plot pixel. A similar square-shaped window, without segment restriction, was applied in the extraction of a reference feature set (RF). The purpose of RF was to produce baseline results for the evaluation of the performance of segment-based features. Finally, all segment-based features, including segment averages and standard deviations, as well as segment-restricted spectral features from SA, SB and SC, were combined to a segment feature set (SF).

Because there were 48 features in SF (16 features for each of the three segmentations) compared to the 8 features of the RF, a subset of original features was selected from both of these sets for the estimation tests. The purpose of the selection was partly to reduce the number of segment-based features in order to decrease computational load of the estimation and partly to harmonise the number of features in RF and SF.

The selection was carried out with an algorithm based on sequential forward selection. It started by selecting the feature giving the lowest RMSE and proceeded by adding a feature that gave the best performance with the already selected features. The procedure was repeated until the five best features were found after which only minor improvements in RMSE was observed when additional features were employed (see Fig. 5). The procedure allowed each of the original features to be selected one or more times.

The RMSE was determined during each feature selection step by means of a cross-validation technique and a non-parametric k nearest neighbour estimator (k-NN) (Nilsson, 1997; Tokola et al., 1996; Tomppo, 1996; Tomppo et al., 1999; Trotter et al., 1997). The estimates for each plot were determined as a weighted mean of k spectrally nearest neighbours found among the rest of the plots. These

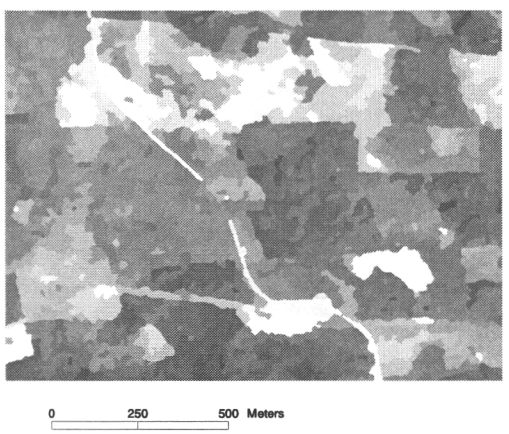


Fig. 4. An example of segment (SC) averages of the red channel of spectrally averaged AISA data.

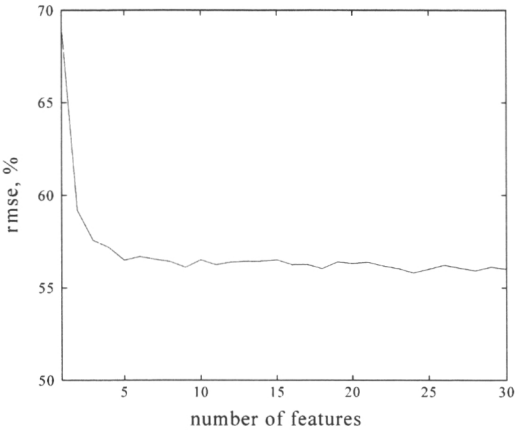


Fig. 5. The effect of number of employed segment-based features to the relative RMSE of the total volume estimate.

neighbours were determined using the Euclidean distance in the feature space. Each of the k nearest neighbours were weighted by the inverse squared Euclidean distance (Eq. (4)). The weight was applied in order to decrease the bias k-NN of the estimator (Altman, 1992).

The number of applied nearest neighbours was set to five. In general, the RMSE of the volume estimates decreases with increasing k, but at the same time, some of the variation present in the original data may be lost and the estimates may become biased (Altman, 1992; Franco-Lopez et al., 2001; Katila & Tomppo, 2001; Mäkelä & Pekkarinen, 2001; Tokola et al., 1996). The chosen number of nearest neighbours was a trade-off between these factors and it was selected based on the analysis of the RMSE and variance curves of the total volume estimates. The curves were determined using RF and k values from 1 to 60. Bearing in mind that the aim of the study was to evaluate the performance of different spectral features in the estimation, the largest k retaining at least 85% of the variation present in the field data was chosen.

$$\hat{y}_s = \left(\sum_{r=1}^{k} \frac{1}{d_{r,s}^2} y_r \right) \Big/ \sum_{r=1}^{k} \frac{1}{d_{r,s}^2} \tag{4}$$

where $\hat{y}_s$ = the estimate of variable y on plot s; y_r = the value of y on the rth nearest neighbour; $d_{r,s}$ = the Euclidean feature space distance between plots s and its rth nearest neighbour; and k = the number of neighbours applied.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{s=1}^{n} (y_s - \hat{y}_s)^2}{n}} \tag{5}$$

$$\mathrm{RMSE},\ \% = \frac{\mathrm{RMSE}}{\bar{y}} \tag{6}$$

where $\hat{y}_s$ = the estimate; $\bar{y}$ = the mean of the observed values; y_s = the observed value of y; and n = the number of field plots.

$$\bar{e} = \frac{\sum_{s=1}^{n}(\hat{y}_s - y_s)}{n} \quad (7)$$

$$s(\bar{e}) = \frac{\delta_e}{\sqrt{n}}, \quad (8)$$

where $s(\bar{e})$ = standard error of empirical bias, $\bar{e}$; δ_e = standard deviation of estimation errors $(\hat{y}_s - y_s)$; and the rest of the notation as in Eq. (5).

3.3. Validation

The performances of the reference (RF) and segment-based features (SF) were compared with the following statistics: RMSE, relative RMSE, empirical bias ($\bar{e}$) and standard error of empirical bias ($s(\bar{e})$) (Eqs. (5)–(8)). The value of $s(\bar{e})$ was employed in such a way that empirical biases larger than $2 \times s(\bar{e})$ were considered to be significant (see Katila and Tomppo, 2001). It has to be noted, however, that the $s(\bar{e})$ is probably an overestimate due to a possible positive spatial autocorrelation of the residuals (Matern, 1960). It should, therefore, be interpreted with care.

In addition to the evaluation of global accuracy by means of RMSE, confusion matrices (CM) were built for volume classes of 50 m^3/ha. The CM statistics employed in describing the accuracy of the estimation were user's (UA) and producer's (PA) accuracies and the overall proportion of area (plots) correctly classified (P) (Stehman, 1997). The quantity of UA is the proportion of correctly classified observations of all observations of a particular class. PA is the proportion of correctly classified observations of those classified to a particular class. For example, if the UA of a class is 1 and the PA of that class is low, it indicates that all the observations of that class were correctly classified, but observations from other classes were also classified to that particular class. Finally, P is the proportion of the observations correctly classified of all observations.

4. Results

4.1. Feature selection

The features selected for the estimation of each variable are presented in Table 4. In reference feature sets (RF), standard deviation features were selected more often than average features. In RF, the average of channel two gave the best estimates for the total volume and volume of spruce. In the case of the volumes of pine and broad-leaved trees, the features giving the lowest RMSEs were the standard deviations of green and red channels, respectively. In the segment-based approach, most of the selected features had been extracted from window-restricted segments, which implies that local features perform better in the estimation than stand-level features. The best segment-based features (SF) for the estimation of total volume and the volume of spruce were averages of the red channel. In the case of spruce, this window-restricted average was extracted from the SC and in the case of total volume from SB. The single feature giving the lowest RMSE for pine was the window-restricted average of the blue channel of SA; for broad-leaved trees, it was the window-restricted standard deviation of NIR channel of SB.

Table 4
Features selected for the estimation of timber volume

Species	Features
Reference features	
Pine	s2,s1,a2,a3,a1
Spruce	a2,s2,a3,s3,s1
Broad-leaved	s3,s1,s3,s3,s3
Total	a2,s1,a3,s3,s1
Segment features	
Pine	RA-a1,RA-a4,RC-s4,RB-s2,RB-s1
Spruce	RC-a3,RB-s3,SB-s3,SB-s1,RC-a3
Broad-leaved	SC-s4,SC-s2,SC-s3,SB-s1,SB-s1
Total	RB-a3,RB-s1,SC-s3,RB-a3,RB-a3

an and sn = Average and standard deviation of channel n, respectively; Sn = segmentation n as described in the Methods section; Rn = restricted feature of segmentation n. Thus, for example, notation SC-s3 stands for segment-level standard deviation of channel 3 of segmentation SC.

4.2. Estimation

The RMSEs of the estimation test showed that SF performed better in the estimation of all the variables employed except for the volume of pine (Table 5). The differences in the relative RMSEs were, however, quite small, varying from about 1% (total volume) to about 9% (volume of broad-leaved trees). An interesting result was that RF seems to perform better in the estimation of the volume of pine, while SF succeeds better in the estimation of the spruce volume. Taking into account that the mean volume of spruce was almost 2.5 times higher than the mean pine volume in the study area, the result may imply that RF performs better in the estimation of small volumes, whereas SF gives more accurate estimates for high volumes. However, considering that k-NN tends to underestimate high volumes and overestimate small ones, the observed empirical biases do not support this conclusion. Biases of estimates employing RF showed slight overestimation, whereas most of the biases of SF were negative. All the biases were, however, considered insignificant given that they were in every case less than $2 \times s(\bar{e})$.

Table 5
RMSE, relative RMSE and empirical bias of the estimates employing reference and segment-based features

Species	RMSE (m^3)	RMSE (%)	$\bar{e}$	$s(\bar{e})$
Reference features				
Pine	40.57	142.53	2.68	2.50
Spruce	69.56	99.01	1.39	4.30
Broad-leaved	23.04	139.11	1.09	1.42
Total	66.67	57.83	3.14	4.11
Segment features				
Pine	42.82	150.43	−0.16	2.65
Spruce	64.34	91.58	−4.88	3.96
Broad-leaved	21.47	129.65	0.94	1.33
Total	65.21	56.56	−0.29	4.03

The evaluation of the confusion matrices using volume classes of 50 m^3/ha supported the RMSE-based observation that the differences in the performance of RF and SF are small. The differences in the proportions of plots correctly classified varied from about 1% to about 6% between the feature sets.

RF gave slightly better P for pine due to its better performance in volume classes from 50 to 150 m^3/ha. Both, RF and SF, gave high UA and PA in the smallest volume class, including most (76%) of the observations, but failed in the estimation of volumes of observations of class 150 m^3/ha or more (Table 6).

The confusion matrix of volume estimates of spruce presented in Table 7 shows that there is no clear difference in the performance of the feature sets. SF gives slightly better P, and its UA is better in every other volume class starting from class 0 to 50 m^3/ha. Both RF and SF gave high PA for the smallest volume class, but RF performed significantly better in the estimation in the highest volume class than SF.

Even though the relative RMSEs of the estimates of volume of broad-leaved trees were high with both feature sets, their confusion matrices show high P, UA and PA values (Table 8). This is due to the fact that most (90.8%) of the field plots fell into the smallest volume class. In addition, observations existed in three volume classes only. Considering that the k-NN produces estimates as weighted averages, it is clear that with this kind of material it is not likely to produce large estimation errors for individual observations. Nevertheless, in such a case, the RMSE of the mean may show a high value due to the small mean volume of the variable.

Table 9 shows the confusion matrix of the estimates of total volume. RF succeeded better in the estimation of volume classes and gave P about 6% higher than SF. Once again, both of the feature sets performed best in the volume class 0 to 50 m^3/ha even though SF could not differentiate the first two classes. In fact, most of the observations of class 50 to 100 m^3/ha were classified to the class 0 to 50 m^3/ha. The difference between RF and SF was also clear in the

Table 6
Confusion matrix of estimation of pine volume

Class	0	50	100	150	200	UA
Reference features						
0	174	24	1	0	0	0.87437
50	24	11	5	0	0	0.27500
100	4	10	2	0	0	0.12500
150	3	1	1	0	0	0.00000
200	1	1	0	0	0	0.00000
PA	0.8447	0.2340	0.2222	null	null	P=0.7137
Segment features						
0	174	25	0	0	0	0.87437
50	30	8	1	1	0	0.20000
100	6	9	1	0	0	0.06250
150	3	1	1	0	0	0.00000
200	1	0	1	0	0	0.00000
PA	0.8131	0.1860	0.2500	0.0000	null	P=0.6985

Table 7
Confusion matrix of estimation of spruce volume

Class	0	50	100	150	200	250	UA
Reference features							
0	130	15	9	6	0	0	0.81250
50	7	10	3	3	3	0	0.38462
100	3	7	4	6	3	1	0.16667
150	2	1	7	6	1	2	0.31579
200	0	3	1	4	3	2	0.23077
250	0	4	3	2	4	7	0.35000
Pa	0.9155	0.2500	0.1481	0.2222	0.2143	0.5833	P=0.6107
Segment features							
0	139	10	8	1	1	1	0.86875
50	13	4	5	4	0	0	0.15385
100	5	4	9	3	2	1	0.37500
150	1	5	2	5	6	0	0.26316
200	0	1	1	4	7	0	0.53846
250	1	1	4	6	6	2	0.10000
PA	0.8742	0.1600	0.3103	0.2174	0.3182	0.5000	P=0.6336

Table 8
Confusion matrix of estimation of volume of broad-leaved trees

Class	0	50	100	UA
Reference features				
0	228	9	1	0.95798
50	14	3	1	0.16667
100	2	4	0	0.00000
PA	0.9344	0.1875	0.0000	P=0.8817
Segment features				
0	234	4	0	0.98319
50	12	6	0	0.33333
100	3	2	1	0.16667
PA	0.9398	0.5000	1.0000	P=0.9198

Table 9
Confusion matrix of estimation of total volume

Class	0	50	100	150	200	250	UA
Reference features							
0	60	22	8	4	1	0	0.63158
50	9	18	11	3	3	0	0.40909
100	2	8	15	4	5	2	0.41667
150	3	2	8	6	7	2	0.21429
200	0	2	1	8	11	5	0.40741
250	0	1	2	7	9	13	0.40625
PA	0.8108	0.3396	0.3333	0.1875	0.3056	0.5909	$P=0.4695$
Segment features							
0	57	25	10	3	0	0	0.60000
50	16	15	10	2	1	0	0.34091
100	4	6	9	9	6	2	0.25000
150	2	3	7	7	6	3	0.25000
200	0	1	2	9	8	7	0.29630
250	0	2	2	8	9	11	0.34375
PA	0.7215	0.2885	0.2250	0.1842	0.2667	0.4783	$P=0.4084$

volume class 200 to 250 m^3/ha, where RF gave about 11% better UA than SF.

5. Discussion

The two-phase image segmentation method applied in this study produced segments which can be efficiently applied as units of feature extraction and image analysis. The visual verification of the segmentation results revealed that after the removal of very small segments present in the initial segmentation due to image noise and irrelevant local variation, the resulting within-stand segments were homogeneous in their forest characteristics. With high values of *t*-ratio threshold, however, some of the relevant segment borders were lost in areas where the forest characteristics changed gradually.

Even though the segmentation succeeded well in the determination of the units for feature extraction and image analysis, the estimation results of this study showed only small differences in the performance of segment-based features (SF) and reference features (RF) extracted from square-shaped windows surrounding the plots. Both feature sets produced high plot-level RMSEs and the confusion matrices of the estimates showed problems, especially in the estimation of high timber volumes.

One reason for the small differences between RMSEs of the estimates produced using RF and SF may be the relatively small size of the extraction window of RF. Taking into account the way in which the locations of the field plots had been post-processed, it is likely that the number of extraction windows intersecting several stands was small. This notion is supported by an earlier study using AISA data and practically the same field data as was employed in this study (Pekkarinen, 2002). That study showed that the RMSE of volume estimates increases with increasing size of the extraction window if segment-based restriction was not employed. The lowest RMSEs were reported with window sizes from 11×11 to 19×19 pixels, but the RMSEs remained at the same level until the window size reached about 40×40 pixels. The results reported here may, therefore, give an overoptimistic impression of the general performance of window-based feature extraction approaches.

In general, the results of this study supported earlier observation that if similar features and estimation technique are employed, the RMSEs of plot-level estimates can be slightly improved using VHR images instead of high resolution image material, e.g. Landsat TM (Hyyppä et al., 2000; Mäkisara et al., 1997). The results presented here are slightly better than those observed with Landsat TM image material and employing somewhat similar estimation methods. Tokola et al. (1996) reported relative standard errors from 66.3% to 68.2% for total volume and estimation errors of larger than 100% for individual species. Errors of approximately the same magnitude were recently reported by Katila and Tomppo (2001). In another recent study employing a Landsat TM image, segment-based local features and *k*-NN estimator errors of about 80% for total volume and errors larger than 100% for individual species were reported (Mäkelä & Pekkarinen, 2001).

The results of this study are also comparable to results obtained using VHR data sources. For example, Tuominen and Poso (2001) recently reported errors from 58.3% to 67% in the estimate of total volume in a study applying different kinds of auxiliary data sources including digitised aerial orthophotos. At the stand level, the accuracy increases and estimation errors of about 45% for total volume have been reported (Hyyppä et al., 2000). In general, the accuracy of the multisource inventory estimates tends to improve rapidly with increasing size of the area for which the estimates are produced (Tokola and Heikkilä, 1997).

An interesting comparison can be made between the results reported here and those of Mäkisara et al. (1997) because both studies employed practically the same study area and field data, and both employed segment-aided feature extraction and applied a somewhat similar estimator. The differences between the studies were that Mäkisara et al. (1997) employed manually derived image segments and original spectral channel configuration of AISA and tested more sophisticated spectral and spatial features in the estimation. In addition to actual results obtained with AISA data, reference results estimated with Landsat TM material and the same field data were also reported.

The comparison showed that better plot-level results were obtained with SF of this study than with Landsat TM data. The improvement in the relative RMSE was most evident, about 20%, in the estimates of spruce and total

volume. In the case of pine volume, however, the TM data gave better results than reported here. The results obtained with averages and standard deviations of the original spectral channels of the AISA image and manually derived image segments were slightly better than the results obtained with SF in this study. The largest differences in relative RMSEs were observed in estimates of volumes of pine and broad-leaved trees, about 18% and 11%, respectively. In the case of spruce volume, the difference was 7.5% and in the case of total volume, only 1%. Taking into account that both of the studies reported relative RMSEs of over 100% for volumes of pine and broad-leaved trees, it can be concluded that the differences in the results of this study and those obtained employing manually derived segments and basic features in Mäkisara et al. (1997) are insignificant.

Even though there may be a larger difference in the performance of segment and window-based features than could be detected with the material used for this study, it can be concluded that segment-based feature extraction barely improves the plot-level estimates of MSFI employing field data gathered from small sample plots. This also seems to apply to images of high and very high spatial resolution (Mäkelä & Pekkarinen, 2001; Pekkarinen, 2002). One reason for this may be the type of field data employed. Considering that reported within-stand variation of total volume is, in Finnish conditions, typically from 26% to 33% (Laasasenaho & Päivinen, 1986; Poso, 1983), it is evident that the small field plots may be very local in nature, and some of the recorded forest characteristics may change significantly if the plot centre is moved even a few meters. From the viewpoint of image analysis, the problem is how to define an analysis unit for which the plot data is representative.

There are at least three possible ways to try to solve the problem presented above. First, in applications employing large numbers of field plots, the segments on which the field plots are located can be sought using a recently presented post-optimisation methodology (Halme & Tomppo, 2001). Another alternative applicable to MSFI applications employing VHR images is to try to find the exact location of a plot by template matching. In such an approach, the tree map of a plot would be matched to the pattern of automatically detected individual crowns in areas near the assumed location of the plot. After that, the training data could be extracted from units of the size and shape of the employed plots and the actual analysis could be carried out by image segments. Thirdly, instead of gathering the field data as point-like observations, it could be stabilised by gathering more data in the field. A relatively cheap, and probably effective, way to collect such data might be the measurement of additional basal area observations near the plot centre and the determination of stand-level volume estimates using these extra measurements. Which, if any, of these solutions will be the most effective way to fully utilise VHR data is a topic for further research.

Acknowledgements

This work was funded by the Finnish Forest Research Institute and the 'Forests in GIS' Graduate School. The field measurements and imaging of the study area were partly funded by the he National Technology Agency(TEKES). I would like to thank the National Forest Inventory Group of the Finnish Forest Research Institute for collecting and preprocessing the field plot and image data. In addition, I would like to thank Prof. Erkki Tomppo, PhD JuhaHeikkinen and LicSc (Tech.) Kai Mäkisara for their valuable comments on an early version of the manuscript.

References

Altman, N. (1992). Introduction to kernel and nearest-neighbour nonparametric regression. *American Statistician*, *46*(3), 175–184.

Franco-Lopez, H., Ek, A. R., & Bauer, M. E. (2001). Estimation and mapping of forest stand density, volume, and cover type using the *k*-nearest neighbour method. *Remote Sensing of Environment*, *77*, 251–274.

Hagner, O. (1990). Computer aided forest stand delineation and inventory based on satellite remote sensing. *Proceedings from SNS/IUFRO workshop in Umeå 26–28 February 1990: The usability of remote sensing for forest inventory and planning*. Umeå: Swedish University of Agricultural Sciences. Remote Sensing Laboratory.

Halme, M., & Tomppo, E. (2001). Improving the accuracy of multisource forest inventory estimates by reducing plot location error—a multicriteria approach. *Remote Sensing of Environment*, *78*(3), 321–327.

Holmgren, P., & Thuresson, T. (1998). Satellite remote sensing for forestry planning—a review. *Scandinavian Journal of Forest Research*, *13*, 90–110.

Hyyppä, J., Hyyppä, H., Inkinen, M., Engdahl, M., Linko, S., & Zhu, Y.-H. (2000). Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. *Forest Ecology and Management*, *128*, 109–120.

Jain, R., Kasturi, R., & Schunck, B. (1995). *Machine Vision*. New York: McGraw-Hill.

Katila, M., & Tomppo, E. (2001). Selecting estimation parameters for the Finnish multisource national forest inventory. *Remote Sensing of Environment*, *76*(1), 16–32.

Laasasenaho, J. (1982). Taper curve and volume functions for pine, spruce and birch. *Communicationes Instituti Forestalis Fenniae*, *108*, 1–74.

Laasasenaho, J., & Päivinen, R. (1986). Kuvioittaisen arvioinnin tarkistamisesta (Summary: On the checking of inventory by compartments). *Folia Forestalia*, *664*, 1–19 (In Finnish).

Mäkelä, H., & Pekkarinen, A. (2001). Estimation of timber volume and the sample plot level by means of image segmentation and Landsat TM imagery. *Remote Sensing of Environment*, *77*(1), 66–75.

Mäkisara, K., Heikkinen, J., Henttonen, H., Tuomainen, T., & Tomppo, E. (1997). Experiments with imaging spectrometer data in large area forest inventory context. *Proceedings of the third international airborne remote sensing conference and exhibition. Copenhagen, Denmark* (pp. 420–427).

Matern, B. (1960). Spatial variation. *Meddelanden fran statens skogsforskningsinstitut*, *49*(5), 144.

Narendra, P., & Goldberg, M. (1980). Image segmentation with directed trees. *IEEE Transactions on Pattern Analysis and Machine Intelligency*, *Pami-2*(2), 185–191.

Nilsson, M., 1997. Estimation of forest variables using satellite image data and airborne lidar. Licensiate thesis. Swedish University of Agricultural Sciences, Department of Forest Resource Management and Geomatics, Umeå.

Pekkarinen, A. (2002). A suggested method for segmentation of very high spatial resolution images of forested landscape. *International Journal of Remote Sensing*, *23*(14), 2817–2836.

Poso, S. (1983). Kuvioittaisen arvioimismenetelmän perusteita (Summary: Basic features of forest inventory by compartments). *Silva Fennica*, *17* (4), 313–349 (In Finnish).

Poso, S., Paananen, R., & Similä, M. (1987). Forest inventory by compartments using satellite imagery. *Silva Fennica*, *21*(1), 69–94.

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, *62*, 77–89.

Tokola, T., & Heikkilä, J. (1997). Improving satellite image based forest inventory by using a priori site quality infrmation. *Silva Fennica*, *1*(31), 67–78.

Tokola, T., Pitkänen, J., Partinen, S., & Muinonen, E. (1996). Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. *International Journal of Remote Sensing*, *17*(12), 2333–2351.

Tomppo, E. (1996). Multi-source national forest inventory of Finland. *New thrusts in forest inventory. EFI Proceedings* 7, 27–41.

Tomppo, E., Goulding, C., & Katila, M. (1999). Adapting Finnish multisource forest inventory techniques to the New Zealand preharvest inventory. *Scandinavian Journal of Forest Research*, *14*, 182–192.

Tomppo, E., Korhonen, K., Heikkinen, J., & Yli-Kojola, H. (2001). Multisource inventory of the forests of the Hebei forestry bureau, Heilongjiang, China. *Silva Fennica*, *35*(3), 309–328.

Trotter, C., Dymond, J., & Goulding, C. (1997). Estimation of timber volume in a coniferous plantation forest using Landsat TM. *International Journal of Remote Sensing*, *18*(10), 2209–2223.

Tuominen, S., & Poso, S. (2001). Improving multi-source forest inventory by weighting auxiliary data sources. *Silva Fennica*, *35*(2), 203–214.

Paper IV

Reprinted from Forestry Sciences, vol 76, Pekkarinen, A. & Tuominen, S., Stratification of a forest area for multi-source forest inventory by means of aerial photographs and image segmentation, pp 111-124. In: Advances in Forest Inventory for Sustainable Forest Management and Biodiversity Monitoring. Kluwer Academic Publishers, Dordrecht, Netherlands, ISBN 1-4020-1715-4, 460 pp.

ERRATA Paper IV

Abstract:

Printed:

"The segments based stratification produces spectrally more homogeneous segments in both study areas, but it did not result in more homogeneous strata in forest characteristics".

Should be:

"Even though the segment-based features retained more of the original spectral variation in both study areas, they did not unabmiguously result in more homogeneous strata in forest characteristics".

CHAPTER 9

STRATIFICATION OF A FOREST AREA FOR MULTISOURCE FOREST INVENTORY BY MEANS OF AERIAL PHOTOGRAPHS AND IMAGE SEGMENTATION

A. Pekkarinen, S. Tuominen
Finnish Forest Research Institute, Unioninkatu 40 AFIN-00170 Finland; Tel. +358-9-85705246, Fax +358-9-625308, e-mail: anssi.pekkarinen@metla.fi; sakari.tuominen@metla.fi

Abstract

In Finland, the data for forest management planning has been gathered by stand-level visual field inventories. From the viewpoint of forest inventory and monitoring, this kind of a method has several drawbacks: the delineation of the stands is subjective and the stand borders tend to change during the planning period. As an alternative, a two-phase plot sampling has been suggested, where first phase sample plots are generated and stratified on the basis of remote sensing images and other auxiliary data sources. The second phase sample is allocated to these strata and measured in the field. The gathered information is generalised to the first phase sample plots using suitable estimators. The method has been applied with digitalized aerial photographs by extracting spectral features for first phase sample plots from square windows surrounding them. This method is unable to fully exploit the spatial resolution of aerial orthophotos. As an alternative we suggest image segment-based feature extraction and stratification. The main hypothesis is that the spectral strata produced by segments are more homogenous in relation to the actual stand characteristics than strata derived by the sample plot-based method. The hypothesis was tested using plot- and segment-based approaches as the basis of the stratification. Spectral average and standard deviation features were extracted from square windows surrounding the plots and from image segments. Two different segmentations were tested. The results from two study areas (S1 and S2) were contradictory. The segment-based stratification produces spectrally more homogeneous segments in both study areas, but it did not result in more homogeneous strata in forest characteristics. In S1 the plot-based approach gave better results than the segment-based approaches. In S2, segment-based approach performed better if only spectral average features were employed in the stratification. The

P. Corona et al. (eds.), Advances in Forest Inventory for Sustainable Forest Management and Biodiversity Monitoring, 111-123.

introduction of spectral standard deviation features to the analysis significantly improved the performance of plot-based approach.

1. INTRODUCTION

In Finland, the data for forest management planning has been traditionally gathered by visual stand-level field inventories. These inventories have been carried out in three phases. First, the pre-delineation of stands has been carried out with help of hardcopies of aerial photographs of the area. In the fieldwork phase, each of the pre-delineated stands has been visited and its delineation has been confirmed. The forest characteristics of each stand have been estimated visually. Finally, the delineation and inventory data have been post processed and the final output composed.

From the viewpoint of forest inventory and monitoring, this kind of a method has several drawbacks. First, the delineation of the stands is subjective and the delineations carried out by different interpreters are seldom similar. Second, the stand borders tend to change during the planning period due to management operations and natural disturbances, which makes the compartments unsuitable for monitoring. Finally, the reliability of the field data may be poor due to the subjective way in which the field data is gathered.

As an alternative to the visual inventory by stands, a two-phase plot sampling method has been suggested (Poso et al. 1987, Holmgren and Thuresson 1995, Poso and Waite 1996). The method is based on the idea of establishing a dense systematic grid of first phase sample plots (1pp) to the area of interest. These plots are assigned remote sensing or other auxiliary information, and stratified into homogeneous strata. The second phase sampling is done within these strata, and based on, for example, the size of the strata. The second phase plots (2pp) are measured in the field. Finally the gathered information is generalised to the 1pp using an appropriate estimator like the mean vector (within strata) or the k-nearest-neighbour (*k*-NN) method. The optimal grid density of 1pp depends on the forest stand-class and applied auxiliary data. As a general recommendation, a grid density of 20 x 20 m^2 has been suggested for studies aiming at stratification of forested areas with help of aerial orthophotos (Holopainen and Wang 1998).

Even though the sample plot-based approach may serve forest-monitoring purposes well, it has some drawbacks when the allocation of the field sample is considered. For example, in applications employing aerial photographs or very high spatial resolution (VHR) images in general, the suggested grid density is sparse in relation to the spatial resolution of the image material. Considering the fact that a single pixel of a VHR image does not represent the spectral properties of a stand, it is obvious that the spectral information for the first phase sample plots has to be generalised from the local neighbourhood of a sample plot. In practice, the average of the pixel values within a square window surrounding each plot has been used (Holopainen and Wang 1998). The problem in that kind of an approach is that the spectral and informational contents of the extracting window may be heterogeneous. This is the case especially near the edges between stands and stands and other land-use classes. Thus, the stratification of 1pp using spectral features extracted this way may assign 1pp observations of different informational classes to the same stratum. This can be avoided at least to some extent if a more sophisticated feature extraction method is applied.

Several studies on multi-source forest inventory have shown that the accuracy of the forest attribute estimates is usually poor at a sample plot level, when optical satellite remote sensing images are used as an auxiliary data in the estimation (e.g., Tokola et al. 1996, Poso et al. 1999). This is mainly due to inadequate correlation between the spectral values of the images and the actual forest attributes. Digitized aerial photographs have performed better in the estimation when compared to satellite images like Landsat TM (Poso et al. 1999). Also Hyyppä et al. (2000) have compared various remote sensing data sources in estimation of forest attributes. Their results show that aerial photographs and imaging spectrometer images give significantly better estimation accuracy at stand level compared to satellite imagery.

This study suggests an image segment-based approach as a solution to the allocation problem. Instead of plot-based feature extraction we determine homogeneous image segments and apply their spectral content in the stratification. Our study hypothesis is that the spectral content and forest characteristics within the strata derived from segment-based features are more homogeneous than in the strata derived from sample plot-based features. The hypothesis was tested by creating three alternative feature sets based on plot-based and segment-based approaches. Based on these

features the image data was stratified into spectrally homogeneous strata and the within-strata variability of timber volume was examined.

2. MATERIAL

The study was carried out in two separate areas in Southern Finland (Figure 1). Study area 1 (S1) is located in the municipality of Leivonmäki. Its forest area covers about 1500 ha and it is a part of a state owned forest area managed by the Finnish Forest and Park Service (Metsähallitus). Study area 2 (S2) is located in the municipality of Kirkkonummi. Its forest area covers about 1000 ha and it consists mainly of small forest holdings in private ownership.

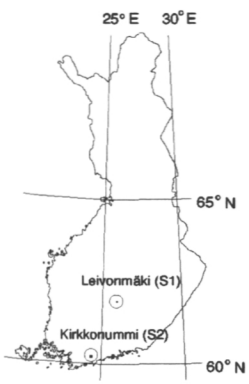


Figure 1. Location of the study areas within the borders of Finland.

The main forest characteristics of the study areas are presented in Table 1. The mean volume of S1 was significantly smaller than the mean volume of S2 due to different site fertility conditions and management practice. In addition, the forests of S1 were dominated by pine whereas the dominating species was spruce in S2.

Both study areas were covered by ortho-rectified colour-infrared aerial photographs. The images from S1 were acquired on July 1999 and the images from S2 on June 1998. The original negatives of the images were scanned using red, green and blue filters and an image mosaic was composed for both areas. The spatial resolution of both mosaics was 0.5

meters. The quality of the image mosaics was good except in south-eastern part of S1. This part of S1 had significantly different spectral properties from the rest of the image mosaic covering S1 and was excluded from the analysis. Finally, both images were re-sampled to the pixel size of 1.5 m using the nearest neighbour method. The re-sampling was applied in order to decrease the computation time needed in the image processing steps.

Table 1. Main forest characteristics of the study areas S1 and S2.

	Volume, m^3ha^{-1}			
Area	Total	Pine	Spruce	Broadl.
S1	95.6	44.0	36.8	14.8
S2	157	50.2	68.1	38.6
Standard deviations				
	Volume, m^3ha^{-1}			
Area	Total	Pine	Spruce	Broadl.
S1	99.4	62.0	74.6	28.7
S2	106.4	62.5	92.5	52.0

Even though the spectral channels of the image mosaics had quite similar mean values the spectral variation in S1 was larger than in S2 (Table 2). Considering that the variation in forest characteristics is larger in S2 it is likely that the differences in the spectral variation are caused by the radiometric properties of the employed images.

Table 2. Spectral characteristics of the study areas.

	S1		S2	
BAND	Mean	Std	Mean	Std
Blue	60.4	31.4	66.2	28.5
Green	93.1	37.7	97.7	26.0
Red	90.3	42.4	86.1	39.9

Field data of earlier studies (Tuominen et al. 2001) were available from both study areas. In S1, 388 field sample plots were measured during the summer of 1999; 194 plots had been measured as circular field sample plots with variable radii and 194 plots as relascope sample plots (factor 1). The plots had been allocated proportionally into strata that were based on

information from aerial orthophotos and old stand inventory data (Tuominen et al. 2001). Of these plots, 59 were located in the excluded part of the image mosaic. Thus the number of field plots employed in the analysis was 329.
In S2, the field data consisted of 233 relascope field sample plots (factor 1), measured during the summer 2000. The plots had been allocated proportionally into strata that were based on information from a Landsat 7 image, aerial orthophotos and old stand inventory data.
The sample plots were located in the field using a compass, distance measurement and a hardcopy of the digitized aerial orthoimage mosaic on which the plot centre points were printed. Plots intersecting several stands were moved 20 metres in the north-south or east-west direction in order to retain the whole plot inside the stand in which the plot centre point was originally located.

3. FEATURE EXTRACTION AND STRATIFICATION

Three different feature sets were derived for the stratification of both study areas. For extraction of feature set F1 systematic grids of first phase sample plots (1pp) were generated for S1 and S2. The distance between centres of 1pp was 20 m in both north-south and east- west directions. An example of the overlay of the 1pp grid on an aerial photograph is presented in Figure 2. The spectral features extracted were averages and standard deviations of pixel values within 20 m x 20 m (13 x 13 pixels) square windows surrounding each 1pp. All three spectral channels (RGB) were employed in the extraction.
Prior to the extraction of F2 and F3, the original image material was pre-processed and segmented. The pre-processing step included Gaussian smoothing of the image. The size of the smoothing window was 3 x 3 pixels and 3 sequential smoothing passes over the image were applied. The initial segmentation of the pre-processed image was carried out with help of a modified implementation of the "Segmentation with directed trees" algorithm (Narendra and Goldberg 1980, Pekkarinen 2002). The algorithm tends to produce a large number of small segments when VHR material is applied, and therefore the result of the initial segmentation was further processed by means of a region-merging algorithm. The applied algorithm searched segments smaller than a user-defined minimum size

and merged them to the spectrally most similar adjacent segment. The spectral similarity of the segments was measured by means of Euclidean distance in the three-dimensional (RGB channels) feature space. For F2 the applied minimum segment size was 169 pixels (380.25 m^2) and for F3 300 pixels (675 m^2). For F2 the minimum size was set to correspond to the size of extraction unit of F1 (20m x 20m) and for F3 a larger segment size was employed in order to study the effect of segment size on the spectral homogeneity of the strata. The spectral averages and standard deviations of pixel values within the segments were extracted for both F2 and F3. An example of the segments is shown in Figure 3.

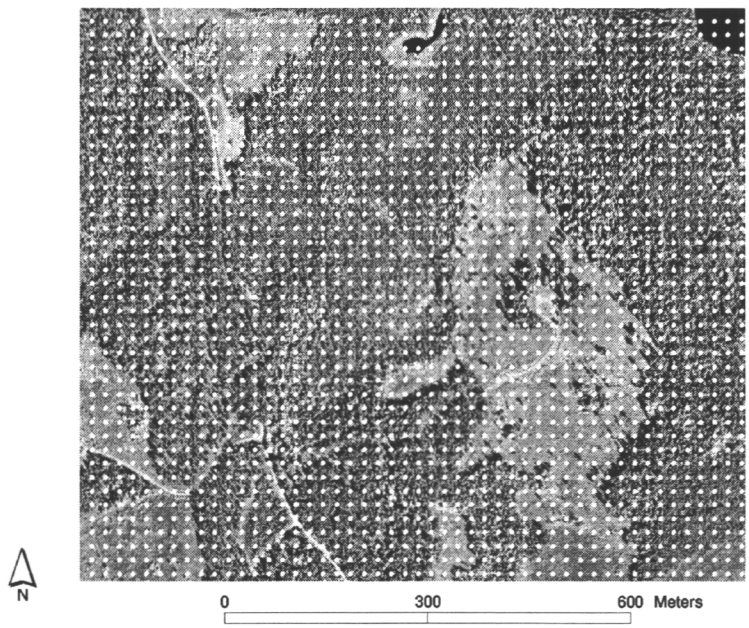


Figure 2. An example of image material and 1st phase sample plots in S1.

In order to study the within strata variation of timber volume the first phase plots and segments were stratified on the basis of the extracted spectral features. The stratification was carried out in two ways: using only spectral averages and using spectral averages and standard deviations. The stratification was carried out using k-means algorithm

(MacQueen 1967). The stratification was carried out as follows: First, the initial stratum centres were selected as the set of observations that maximises the distance between the stratum centres. After that each observation of the feature set was assigned to the spectrally nearest stratum centre. Finally, the centroid vector of each stratum was recalculated as a mean vector of the observations assigned to that stratum. The process was iterated until the stratum centres remained unchanged. Because the result of stratification may be sensitive to the number of strata employed, different numbers of strata (20-50) were tested.

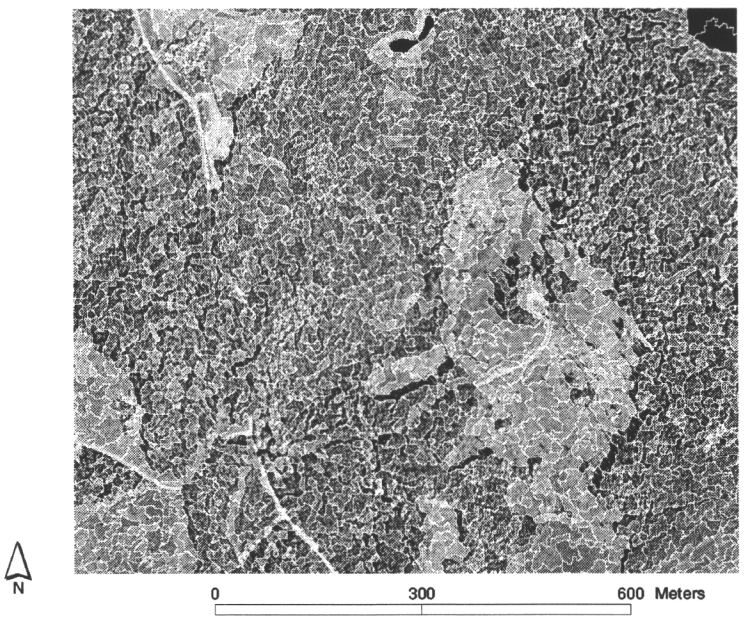


Figure 3. An example of borders of image segments employed in the extraction of F2 in S1.

After the stratification the field sample plots of S1 and S2 were assigned to the strata. In the sample plot-based approach each plot was assigned to the spectrally nearest stratum. Euclidean distance measure was employed. In the segment-based approach, the field plots were assigned to the strata of the segment on which the plot centre was located. Therefore also the plots intersecting several segments were assigned to a single stratum.

Dividing the plots was not practicable because most of the plots were measured as relascope plots. After the allocation of all plots to the strata, the within strata standard deviations of the total volumes were examined. The strata having zero or single field observations (and therefore zero standard deviation) were excluded from further analysis, because their standard deviation had no meaningful content for the analysis. The strata resulting from different feature sets were compared by examining the mean standard deviations within the strata. The mean standard deviations were calculated as area-weighted averages of the standard deviations of the strata.

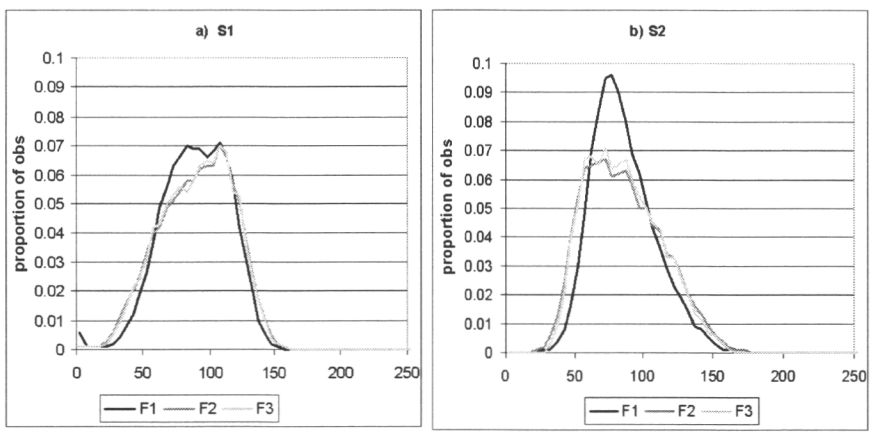


Figure 4. Spectral distributions (red channel averages) of feature sets F1, F2 and F3 in S1 (a) and S2 (b).

4. RESULTS AND CONCLUSIONS

There were small differences in the spectral characteristics of F1, F2 and F3 in both of the areas. Figure 4 shows the distributions of the spectral values in F1, F2 and F3. In both study areas, the distributions of the mean values of segment-based spectral features were wider than those extracted from square neighbourhoods surrounding the plots. This supports our hypothesis that the segment-based approach retains more of the original spectral variation present in the image. The hypothesis is also supported by the fact that the spectral variation in the feature set F3 was larger than in F1, even though the area from which the features were

extracted was larger in F3. In general, the use of larger feature extraction should result in more averaged features. However, using the segments in the feature extraction seems to avoid this drawback at least to some extent.

In S1 the plot-based stratification gave clearly better results than the segment-based approach if the stratification was based on only segment averages of the employed three channels (Figure 5a). The within strata variation of total volume was smaller with all the examined numbers of strata. In the case of segment-based approaches, the smaller segment size produced more homogeneous strata. In addition, the smaller segments produced more stable results than the larger ones.

The introduction of spectral standard deviations of the three channels to the analysis clearly decreased the within stratum variation of total volume in the segment-based approaches. However, the plot-based strata were still more homogeneous when a relatively small number (20-30) of strata was employed (Figure 5b). With increasing number of strata, however, the variation within both segment-based and plot-based strata was comparable.

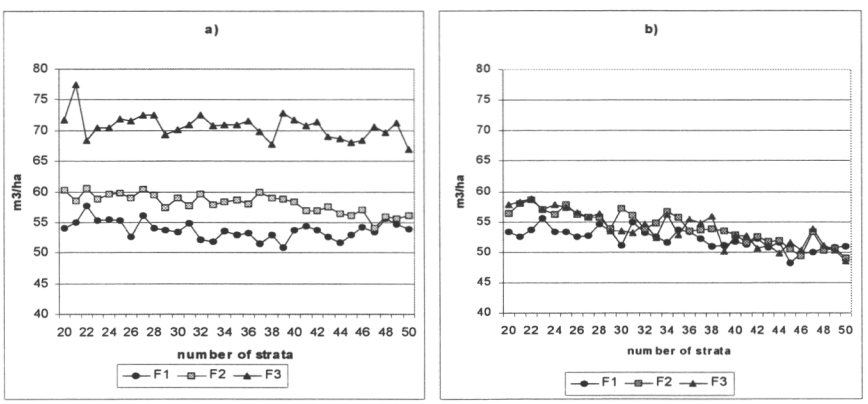


Figure 5. Weighted within strata standard deviation of total volume in S1. Strata resulted using spectral average (a) and average and standard deviation features (b) extracted from RGB channels.

In S2, the segment-based strata derived using spectral averages only produced more homogeneous strata than the plot-based approach (Figure 6a). The difference between the plot-based and segment-based approaches increased with increasing number of strata. In addition, with most (65%)

of the examined numbers of strata, the spectral averages extracted from larger segments produced more homogeneous strata than averages extracted from smaller segments.

The inclusion of the spectral standard deviation to the stratification significantly increased the homogeneity of the plot-based strata in S2 (Figure 6b). In fact, the comparison of the results obtained with different number of strata revealed that the plot-based strata were more homogeneous than any of the segment-based strata in 87% of the cases.

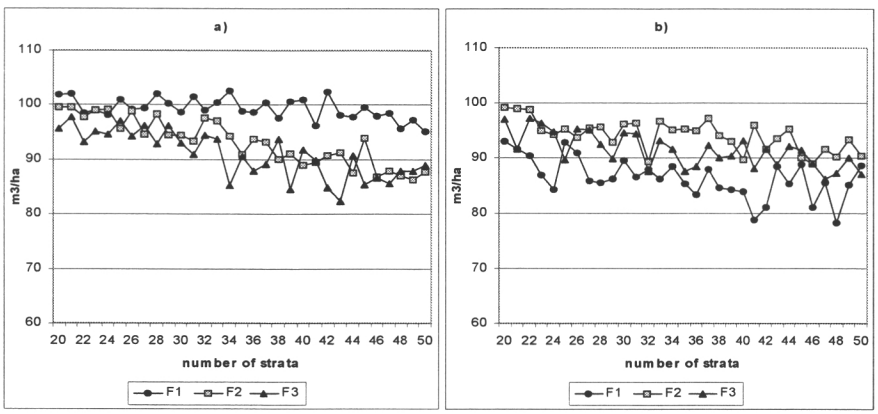


Figure 6. Weighted within strata standard deviation of total volume in S2. Strata resulted using spectral average (a) and average and standard deviation features (b) extracted from RGB channels.

The results were controversial. The use of the segment-based spectral features retained more of the original spectral variation of the images than the features extracted from square neighbourhoods of the plots. This difference in the spectral variation was clearer in S2. The within strata variation of timber volume was generally larger in segment-based strata in both study areas. The introduction of spectral standard deviation features to the stratification diminished the within strata variation of timber volume in plot-based approach only slightly in S1 whereas this improvement was more obvious in S2. In segment-based strata, the effect was most evident in S1 when larger segments were employed. In S2 the inclusion of the standard deviation features did not have a clear, positive effect on the results. In fact, the combination of segment-based average-

and standard deviation features tended to produce more heterogeneous strata in S2 than spectral averages only.

The fact that the strata derived from segment-based features were not more homogeneous in timber volume may be partly due to the too small segment size employed. This conclusion is supported by the fact that if both spectral average and standard deviation features were employed, the larger segment size resulted in more homogeneous strata in most of the examined cases in both study areas. Small segments tend to produce strata having larger mean standard deviation of total volume than strata derived from large segments. This is due to the high within stand spectral variation present in the images. Small segments are more homogeneous in their spectral contents and tend to separate bright and dark areas into different segments. Therefore, the pixels of one stand may fall to several strata having different spectral properties. Thus, the result of the analysis of within strata variation of timber volume may be more or less arbitrary depending on the locations of the employed field plots.

We conclude that the segment-based stratification is a promising tool for stratification of very high resolution imagery. The results presented here indicate that the segment-based approach has some advantages in forest inventory utilising very high resolution images. However, some matters need to be taken into account. The use of segment-based stratification in a multi-source inventory requires field data that is representative at the segment level. In practice, this may cause some problems because the field personnel need to know the exact locations of the borders of the segments to be assessed. Furthermore, the segmentation may not detect all relevant borders if the stand structure is not clear. This, however, should be of minor relevance because the size of the segments is typically smaller than the average size of the stand. Therefore segment level features should result in representative stratification even in gradually changing areas.

References

Hagner, O. 1990. Computer aided forest stand delineation and inventory based on satellite remote sensing. In Proceedings from SNS/IUFRO workshop in Umeå 26–28 February 1990: The usability of remote sensing for forest inventory and planning. Swedish University of Agricultural Sciences. Remote sensing laboratory. Umeå.

Holmgren, P., Thuresson, T. 1995. Avdelningsfritt skogsbruk. Faktaskog Nr 14. Sveriges Lantbruksuniversitet. In Swedish.

Holopainen, M., Wang, G. 1998. The calibration of digitized aerial photographs for forest stratification. International Journal of Remote Sensing 19: 677-696.

Hyyppä, J., Hyyppä, H., Inkinen, M., Engdahl, M., Linko, S., Zhu, Y-H. 2000. Accuracy estimation of various remote sensing data sources in the retrieval of forest stand attributes. Forest Ecology and Management 128: 109-120.

MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. Volume 1 of Proceedings of the Fifth Berkeley Symposium on Mathematical statistics and probability, pp 281-297. Berkeley, 1967. University of California Press.

Narendra, P., Goldberg, M. 1980. Image segmentation with directed trees. IEEE Transactions on Pattern Analysis and Machine Intelligence, Pami-2: 185–191.

Pekkarinen, A. 2002. Image segment-based spectral features in the estimation of timber volume. Remote Sensing of Environment 82: 349-359.

Poso, S., Paananen, R., Similä, M. 1987. Forest inventory by compartments using satellite imagery. Silva Fennica 21: 69 -94.

Poso, S., Waite, M-L. 1996. Sample based forest inventory and monitoring using remote sensing. *In* Remote sensing and computer technology for natural resource assessment. Research notes 48. University of Joensuu, Faculty of Forestry.

Poso, S., Wang, G., Tuominen, S. 1999. Weighting alternative estimates when using multi-source auxiliary data for forest inventory. Silva Fennica 33: 41 –50.

Tokola, T., Pitkänen, J., Partinen, S., Muinonen, E. 1996. Point accuracy of a non-parametric method in estimation of forest characteristics with different satellite materials. International Journal of Remote Sensing 17: 2333-2351.

Tuominen, S., Fish, S., Poso, S. 2001. Combining remote sensing, old inventory data and geostatistical interpolation in multi-source forest inventory. Submitted to Canadian Journal of Forest Research.

ISBN 951-40-1929-6
ISSN 0358-4283