GENERALIZATION OF IMAGE DATA TO GIS POLYGONS FOR CHANGE DETECTION AND DATA BASE REVISION

Katarina Johnsson, PhD
Dept. of Geodesy and Photogrammetry
Royal Institute of Technology, KTH
S - 100 44 Stockholm
Sweden

Commission IV, Working Group 3

KEY WORDS: map revision, forest maps,.environmental monitoring, remote sensing

ABSTRACT: Semi-automatic revision of geographic data bases from satellite imagery requires methods to perform change detection between map and image data. Most geographic databases are stored in vector-based geographic information systems, while analysis of satellite data requires raster GIS. It is argued that image data must be generalized into one or several polygon attributes to enable change detection between map and image data, and that change detection should take place in an attribute data base environment, rather than in a raster image environment. A framework for image generalization functions is presented, based on descriptive statistics. Statistical measures to describe central tendency, dispersion and spectral heterogeneity are discussed for quantitative and qualitative image data respectively. A case study is presented, which describes how forest change detection can be performed by regression between polygon attributes derived from map and image data of different dates. High detection accuracy is achieved for logged forest stands, confusion with objects of similar spectral characteristics is avoided, and the influence from natural spectral variation is minimized.

1. INTRODUCTION

There is a need for efficient methods to revise digital geographic data bases. This paper is concerned with revision of small-scale geographic data bases (>1:20,000) for natural resources management, and specifically with revision of map polygon features. Satellite imagery provides an adequate data source for revision for geographic databases at this scale.

A geographic data base will normally contain one or several layers of digitized maps. Most geographic data bases have been created and are being maintained in vector-based Geographic Information Systems (GIS), where map features are represented either as points, lines or polygons. Polygon features are especially common in geographic databases for natural resources management, for example forest stands, animal habitats, soil compartments, geomorphological terrain components, and watersheds.

Satellite imagery is stored and processed in raster based GIS or image analysis systems (IAS). Traditionally, change detection has been focussed on pixelwise analysis between images of different dates (Singh, 1989).

Revision of geographical databases can take place either with respect to map attributes or with respect to geometric properties. For natural resources databases, geometry and attributes are often interconnected, since the boundaries of the map polygons are drawn based their attributes (Veregin, 1989). This is for example true for soil and forest maps. In these cases, the revision process has two distinct steps: 1) detection of changes with respect to database attribute(s), and 2) adjustment of geometric properties, e.g. by drawing new boundaries and deleting old. This paper discusses the first step.

Map revision from satellite data requires that change detection takes place from map to image data. This can be done manually, by on-screen image interpretation and map overlay. However, operational revision of large databases require the revision process to be, at least, semi-automated. Several issues need to be addressed before semi-automatic database revision becomes reality, for example differences in data formats and in levels of abstraction between database and image data.

2. CHANGE DETECTION FROM MAP TO IMAGE

Since integrated vector/raster analysis is currently not possible, one of the data sources must be processed to fit the other, before map-to-image change detection can take place. This is not only a question of format conversion, but also a question of data abstraction. Maps are highly abstracted representations of the real world, while image data is a lower form of information (Ehlers et al., 1991; Goodchild, 1989).

There are several reasons for why image data should be generalized to map polygons, rather than converting the vector map to raster data and performing pixelwise change detection (the traditional image analysis approach):

- The image data becomes immediately comparable to other polygon attributes in the attribute database.

Current affiliation: T-Kartor Sweden AB, Box 5097,
S-291 05 Kristianstad, Sweden
email: katarina.johnsson@t-kartor.se

- The map polygons represent meaningful units in terms of the purpose of the database. For example, the forester is primarily interested in which forest stands have been logged, not which pixels. In a paper explaining how GIS can be applied in animal ecology, Haslett (1990) recommends using vector- rather than raster-systems, since analysis based on vector polygons, defined by parameters of relevance to the animals, are of much greater ecologial significance than analysis based on individual pixel values.

- When two data sets with different spatial resolution are combined, the data set with the coarser resolution must determine the resolution of the analysis.

- Map attributes are generalizations which are valid on the level of the map polygons, but which are not necessarily valid in every single pixel.

- The influence from natural spectral heterogeneity present in the image data is reduced.

- Standard methods for data analysis, such as one-dimensional statistics, can be applied, e.g. to change detection.

These arguments support the approach of concentrating the image analysis efforts to extraction of one or several relevant parameters, which are valid at a map polygon level. New polygon attribute(s) are created from the image data, and change detection is performed in the attribute database environment, rather than in the spatial (vector/raster) environment. This approach has previously been adopted to extract model parameters from remotely sensed data for ecological modelling (Band et al., 1991).

3. IMAGE GENERALIZATION FUNCTIONS

As of yet, only a handfull GIS functions exist to generalize image data to GIS-polygons. These are found in raster-based GIS. The need for development in this area has previously been pointed out e.g. by Trotter (1991).

Image generalization functions in raster GIS operate on regions of connected pixels, rather than on individual pixels. Pixel values are generalized from one data layer to regions defined in another data layer (figure 1).

Conceptually, two principally different categories of image generalization functions can be distingushed, based on image data type. Raw image data is *quantitative* in nature, i.e. the pixel values correspond to measurements on an interval scale. Image data which has been categorized is *qualitative* in nature, i.e. the pixel values correspond to classes on a nominal scale. Image categorization can take place e.g. by statistical classification, clustering or image segmentation.

3.1 Quantitative image generalization functions

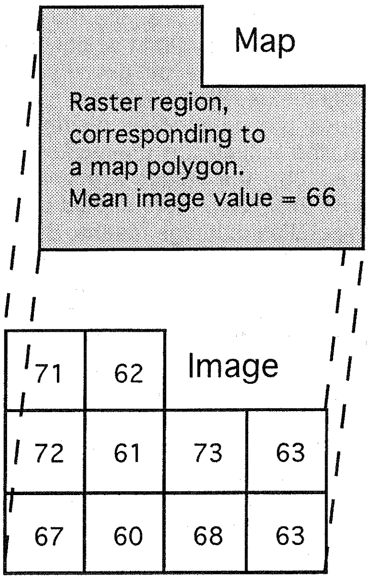


Figure 1. The shaded pixels represent a region in the map data layer, defined by 4-connectivity. Image pixel values are generalized into one mean value (66), valid for the map region.

Generalization of raw image data to GIS-polygons may include as well condensation of multiple image bands into a single band, as extraction of representative values from a set of pixels in this single band. None of these problems are new to the remote sensing community, but they need to be expanded and adopted to work on sets of image pixels, defined by regions or polygons in other GIS layers.

The normalized vegetation index (NDVI), Prinicipal Component Analysis, Tasseled cap transformation, as well as other image transformations are commonly accepted methods to reduce the dimensionality of the data (Lillesand and Kiefer, 1987). In some applications, image transformations are required, in others it may be sufficient to use individual image bands. By using existing methods for reduction of data dimensionality, it is possible to concentrate development efforts to image generalization functions which operate on singel image bands.

Primarily, a full set of statistical functions to compute measures of central tendency and dispersion for image pixels within GIS-polygons is required. These statistical functions should include descriptive measures such as mean, median, trimmed mean, variance, and standard deviation (table 1).

Table1. Statistical measures and methods to generalize quantitative image data to GIS polygons

	Measure
Central tendency	mean, trimmed mean, median
Dispersion	variance, standard deviation
Shape of distribution	skewness, kurtosis
Uni-modal distribution?	histogram peak detection, clustering

Additionally, it is important to describe the shape of the distribution of pixel values within a polygon. This is crucial to indicate whether or not the polygon is spectrally homogeneous. Basically, three situations can occur. The values are uni-modally distributed without tails (figure 2a), the values are uni-modally distributed with a tail (figure 2b), the values are not uni-modally distributed (figure 2c). Such tests should be performed prior to the computation of image means. Skewness and kurtosis can be applied to find tails in the distribution, e.g. to detect local within-unit variation, such as that introduced by the road in the forest stand (figure 2b). There is no standard statistical method to determine if a set of values are uni-modally distributed. Several alternatives exist, such as histogram peak detection algorithms, or clustering algorithms. One concern here is processing time, another is that the method should be applicable to small data sets, since a map polygon may include few pixels. Research and tests are ongoing.

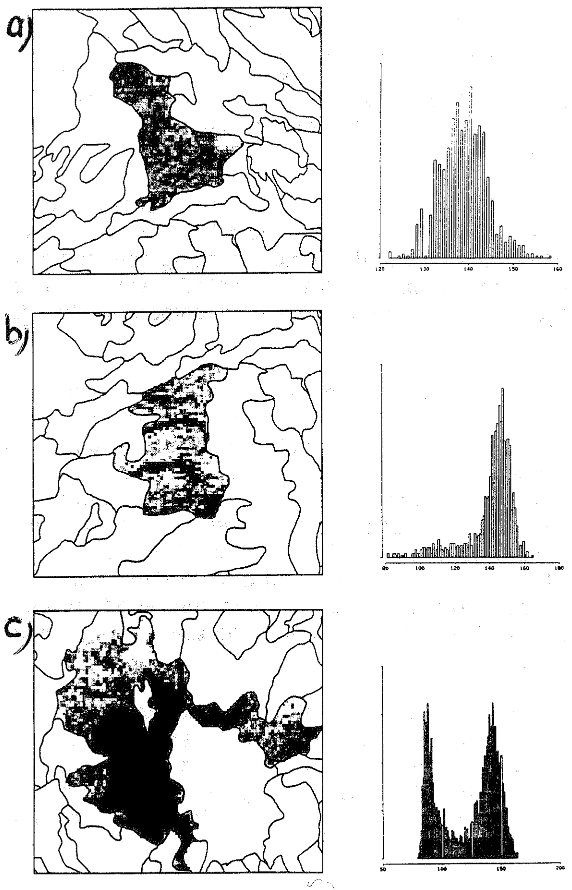


Figure 2. a) A spectrally homogeneous forest stand, b) a forest stand with uni-modally distributed pixel values, with a tail caused by logging roads, c) a forest stand that has been partially logged, resulting in a bi-modal histogram.

3.2 Qualitative image generalization functions

Multi-spectral image classification reduces the dimensionality of the data, at the same time converting raw measurements into information. Mode (highest frequency of occurence) is a measure of central tendency in the data (table 2). Percentage coverage of a certain class, and composition (i.e. the relative proportion) of all classes are two measures which describe the "dispersion" or heterogeneity (figure 3).

Table 2. Measures to generalize qualitative image data to GIS polygons

	Measure
Central tendency	mode
"Dispersion" or homogeneity	percentage coverage of each class relative proportion of classes
Spatial distribution of classes	fragmentation index, other spatial pattern indeces.

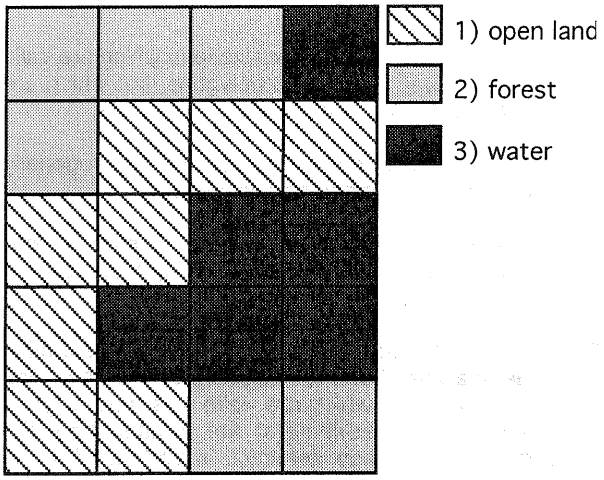
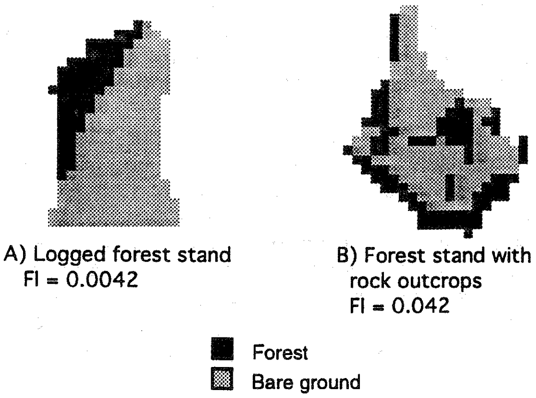


Figure 3. Measures for qualitative image data: mode = 1; percent coverage of water is 30, and composition of categories is {40,30,30}

The spatial distribution of classes can be just as important as the proportional coverage classes, when extracting image information for GIS polygons. Consider for example two forest stands composed of approximately the same proportion of the image categories forest and bare ground (figure 4). One has been partially logged, while the other contains numerous rock outcrops. The major difference between the two stands is the spatial distribution of categories.


A) Logged forest stand
FI = 0.0042

B) Forest stand with rock outcrops
FI = 0.042

Figure 4. The partially logged forest stand has a lower pattern complexity than the forest stand with rock outcrops. Pattern complexity has been measured using a fragmentation index, FI (Monmonier, 1974)

For this reason the fragmentation index, originally intended as a measure of pattern complexity for choroplethic maps (Monmonier, 1974) was implemented and applied as a region-based GIS operator (Johnsson, 1995). The fragementation index is computed as:

$$FI = (M-1)/(N-1),$$

where M = number of image regions in the categorized image, and N = number of pixels in the categorized image. The FI is computed for each map polygon in the base map (figure 5). The FI was implemented in a raster-based GIS.

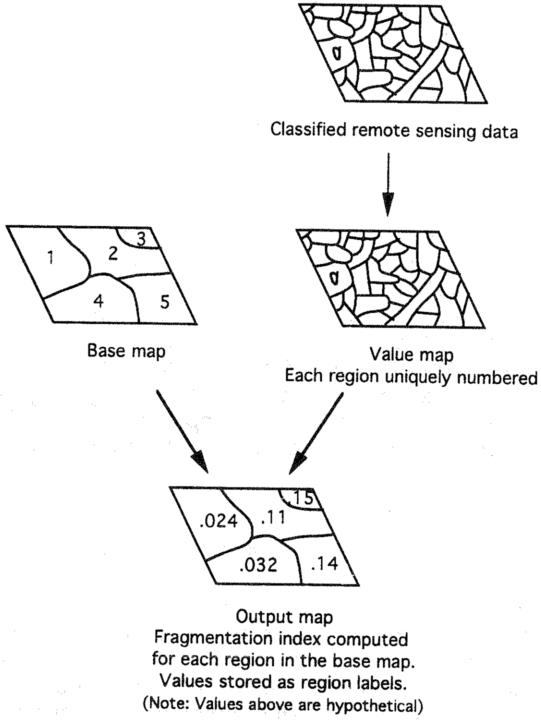


Figure 5. Region-based fragmentation index (FI) computations in a raster GIS.

The spatial pattern of landuse categories is being recognized as increasingly important within landscape ecology and several indices have been developed to capture significant spatial patterns (e.g. McGarigal and Marks, 1994). These are designed to work on categorized image data, and would provide another source for qualitative image generalization functions.

4. POLYGON-BASED CHANGE DETECTION - A CASE STUDY

4.1 Background

The case study relates to forest management. The study was carried out at the Pacific Forestry Centre in British Columbia, Canada, within the framework of the SEIDAM (System of Experts for Intelligent Data Management) project.* The issue was to develop and test a method to automatically extract clear-cuts from a Landsat TM image for revision of forest inventory maps in a vector-based GIS.

Clear-cut detection by multi-spectral image classification yielded unsatisfactory result due to the natural heterogeneity of the mountaineous test area, and due to confusion with other features of similar spectral characteristics, such as rock outcrops and roads.

Instead an approach was adopted that relied on comparison of existing forest density data in the GIS data base with meausures of forest density computed from the image data.

The study has been described in detail in Johnsson (1994).

4.2 Material and methods

The method was developed and tested on a digital forest inventory map in scale 1:20,000 (BC Ministry of Forests, 1991), covering an area of approximately 11x14 km (figure 6). The digital forest map consists of map polygons, which correspond to forest stands (forest management units). Each forest stand has a number of attributes, stored in a separate attribute database.

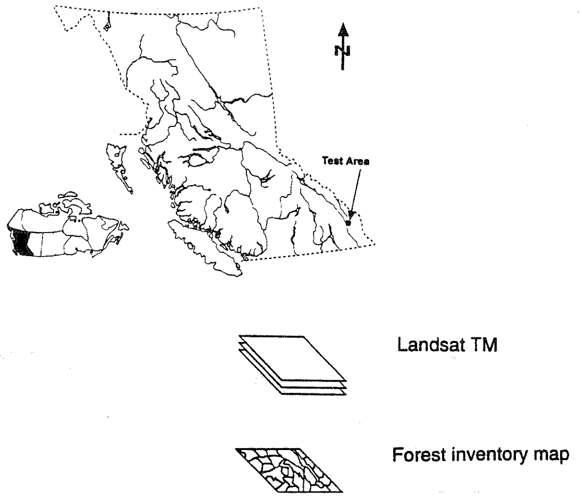


Figure 6. Study area and data

A subsection of a Landsat TM scene was used as image material. The temporal difference between the map and the image was approximately a year, during which logging was known to have occurred.

The forest database attribute crown closure provided an estimate of expected forest density for each forest stand. Crown closure is defined as the percentage of ground area covered by the vertically projected crowns of all living, commercial tree species in the main canopy, rounded off to the nearest 10 % (BCMoF, 1992). The map contains forest stands with crown closure ranging from 0% to 65%.

A normalized difference image of Landsat TM4 and TM5 (ND45) was computed according to:

---

* SEIDAM is a project under NASA's Applied Information Systems Research Program. The SEIDAM Project is also supported by Natural Resources Canada, Industry and Science Canada, the BC Ministry of Forests, the BC Ministry of Environment, Lands and Parks, and the BC Forest Resources Development Agreement.

```
ND45 = (TM4 - TM5)/(TM4 + TM5)*128 + 128,
```

The image values were generalized for the forest stands by computing the mean image value for each stand. These computations were performed in a raster-based GIS environment. Boundary pixels and known roads were omitted from the computations. The computed values were stored for each forest stand, as an additional attribute in the forest data base.

The correlation between crown closure and ND45 was tested for a set of undisturbed, SE-facing (illuminated) forest stands (N=71), and a linear regression function was computed and tested for preliminary evaluation.

Forest stands with a potential for logging were extracted from the database and considered as candidates for change (N=228). Change detection was carried out by robust regression between map attribute crown closure and image attribute ND45. The algorithm has been described in Baarda (1968). Outliers were automatically detected in the regression. Outliers with unexpectedly low ND45 values were labeled as logged forest stands.

4.3 Results and discussion

It was demonstrated that crown closure and ND45 were correlated on a polygon level for undisturbed forest stands (figure 7). The regression coefficient for the linear regression function was 0.97. The two data sets had low correlation on a per-pixel basis. This can be explained by the natural heterogeneity of the landscape, and by the influence of factors which could not be kept under control, such as understory vegetation or secondary species in the canopy.

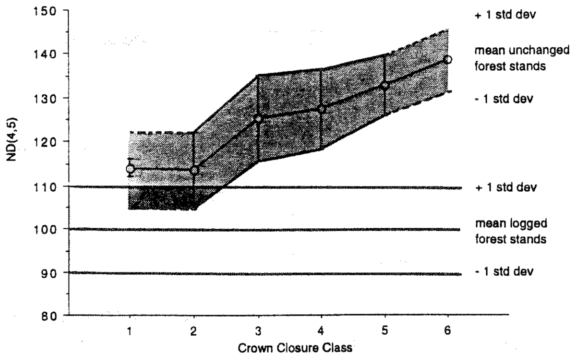


Figure 7. The relationship between crown closure, as derived from the map, and ND45, a standwise image measure of forest density. The means and the standard deviations for undisturbed and logged forest stands indicate the separability between the two groups of data, for the various crown closure classes.

10 out of 10 completely (> 60%) logged forest stands in the test area were detected, with one error of commission. Confusion with rock outcrops and forest of low density was avoided. The influence from natural spectral variation was minimized.

Partially logged (< 60 %) were not detected. As discussed previously (section 3), standwise image means should not be computed for these stands since they contain at least two statistically separate populations of pixel values. Some clustering algorithms, combined with fragmentation index computations were tested to initially detect and identify partially logged forest stands, with promising preliminary results.

4.4 Conclusions from the case study

The polygon-based approach radically reduced the problems related to landscape heterogeneity, in comparison to multi-spectral classification. In combination with the GIS-based sub-selection of forest stands, it also reduced the confusion with features of similar spectral characteristics.

The case study underlines the need for improved image generalization functions, especially to deal with spectrally heterogeneous map polygons.

A revision system is envisioned where polygon-based change detection is the first step. The result, map polygons which exhibit strong spectral anomalies, is signalled to the operator, who will make the final decision. The second step will be semi-automatic geometric adjustment of polygon boundaries. Another advantage of the polygon-based method is that existing polygon boundaries can be used as first approximations in this process. Preliminary investigations are being undertaken in this area.

5. OPERATIONAL MAP REVISION

Currently, semi-automated revision of geographical databases from satellite image data requires that raster and vector GIS are operated in parallell. The map with its attribute database is stored in a vector GIS. It is converted from vector to raster format and imported to a raster GIS, where it is combined with the image to computer generalized image values. These region-based values must be exported back to the attribute database of the vector-GIS, for integrated analysis with the other polygon attributes.

This situation would hardly be acceptable in an operational situation. Too much time is spent converting data back and forth, and it is required that parallell databases (raster and vector) are maintained. There is an apparent risk of introducing errors in the data.

Several authors (e.g. Star et al., 1991; Ehlers et al., 1991; Goodenough, 1988) have suggested development of high level data management systems or expert systems to perform low level tasks such as format conversion in a manner transparent to the user. This level of integrated analysis, where two systems are working in parallell under a common user interface is referred to as "seamless integration" by Ehlers et al. (1989).

Integrated systems should certainly be a development goal, but perhaps they lie far in the future. Most geographic databases have been created and are being maintained in vector systems. In addition, a large number of data bases with environmentally related point measurements exist (e.g. runoff, precipitation, ground water quality), which can be linked to vector GIS based on the X,Y-coordinate pairs. Operational map revision procedures and monitoring programs will be based mainly on these existing systems. If image data is intented to provide input to data bases and models, vector GIS must incorporate raster based image analysis functions, especially functions that operate on regions of connected pixels, as defined by polygons in GIS layers.

Future geographic information systems should provide the user with a complete set of tools to generalize image values for meaningful analysis units. For raster systems this means expanding on the concept of region-based GIS analysis. For vector systems it means integrating image analysis functions. Statistical measures, as well as spatial pattern indices can provide the base.

6. ACKNOWLEDGEMENTS

The author wishes to thank the SEIDAM project leader, Dr. David G. Goodenough at the Pacific Forestry Centre, for his valuable support.

The financial support from the Swedish Institute and the Sweden-America Foundation is gratefully acknowledged.

The data for the case study was provied by the British Columbia Ministry of Forests, the British Columbia Minsitry of Environments Lands and Parks, and by the Canada Centre for Remote Sensing.

Thank you also to my former collegues at Dept. of Geodesy and Photogrammetry, KTH, for support in practical as well as in scientific matters.

7. REFERENCES

BAND. L., PETERSON, D., RUNNING, S., COUGHLAN, J., LAMMERS, R., DUNGAN, J, AND R. NEMANI: 1991. Forest ecosystem processes at the watershed scale: basis for distributed simulation. *Ecological Modelling*, Vol 56., pp. 171-196.

BAARDA, W.: 1968. A Testing Procedure for Use in Geodetic Networks, *Neth. Geod. Comm.*, Vol. 2, no. 5.

BCMoF, BRITISH COLUMBIA MINISTRY OF FORESTRY: 1991. *Forest Inventory Manual*, chapter 11, Forest Resource Geographic Information System, pp. 44-45.

BCMoF, BRITISH COLUMBIA MINISTRY OF FORESTRY: 1992. *Forest Inventory Manual*, Photographic Interpretation, pp. 1-25.

EHLERS, M., EDWARDS, G., AND Y. BÉDARD: 1989. Integration of remote sensing with geographic information systems: a necessary evolution. *Photogrammetric Engineering and Remote Sensing*, Vol. 55, No.11, pp. 1619-1627.

EHLERS, M., GREENLEE, D., SMITH, T. AND J. STAR: 1991. Integration of remote sensing and GIS: data and data access. *Photogrammetric Engineering and Remote Sensing*, Vol. 57, No. 6, pp. 669-675.

GOODCHILD, M.: 1989. Modeling error in objects and fields. *In: The accuracy of Spatial Databases* (editors: Goodchild, M. and S. Gopal), Taylor and Francis, London, pp. 107-113.

GOODENOUGH, D. G.: 1988. Thematic Mapper and SPOT integration with a geographic information system. *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No. 2, pp. 167-176.

JOHNSSON, K.: 1994. *Integrated digital analysis of regions in remotely sensed imagery and thematic data layers*. PhD thesis. Dept. of Geodesy and Photogrammetry, Royal Institute of Technology, Stockholm, Sweden.

JOHNSSON, K.: 1995. Fragmentation index as a region based GIS operator. *Int. J. of Geographical Information Systems*, vol. 9, No. 2, pp. 211-220.

LILLESAND,T. and R. Kiefer: 1987. *Remote Sensing and Image Interpretation, second ed.* John Wiley & sons. 721 pages.

MCGARIGAL, K. and B. MARKS: 1994. FRAGSTATS - spatial pattern analysis program for quantifying landscape structure. Forest Science Department, Oregon State University, Corvallis.

MONMONIER, M.: 1974. Measures of pattern complexity for choroplethic maps. *The American Cartographer*, Vol. 1., pp. 159-169.

SINGH, A.: 1989. Review article. Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, Vol. 10, No. 6, pp. 989-1003.

STAR, J., ESTES, J., AND F. DAVIS: 1991. Improved integration of remote sensing and geographic information systems: a background to NCGIA Initiative 12. *Photogrammetric Engineering and Remote Sensing*, Vol. 57, No. 6, pp. 643-645.

TROTTER, C. M.: 1991. Remotely-sensed data as an information source for geographical information systems in natural resources management: a review. *International Journal of Geographical Information Systems*, Vol. 5, No. 2, pp. 225-239.

US CERL: 1993. *GRASS4.1 Users Reference Manual.* United States Corps of Engineers, Construction Engineering Research Laboratories, Champaign, Illinois, 556 pages.

VEREGIN, H.:1989. A taxonomy of error in spatial databases. *NCGIA Technical paper 89-12,* University of California, Santa Barbara. 115 pages.