Analysis of Remote Sensing Data in Geographical Information Systems

Dieter Fritsch

Munich

ABSTRACT

The progress in imagery sensors of remote sensing demands for an efficient integration of remote sensing data in geographical information systems (GIS). This integration leads to hybrid GIS which should take care for vector data, image data and semantic data (attributes) simultaneously. Amongst different problems to be solved for an efficient treatment of image data the following ones are very important: data management, data analysis, and data visualizations and output.

The paper deals with analysis aspect of remote sensing data. In particular, intersections and generalizations will be investigated. Data being used are classified images in form of chorophlet maps - the main advantage of these maps is its homogeneity to be coded very efficiently by means of quadtrees.

With this in mind the paper deepens functional aspects of intersections in a quadtree environment. The regular geometrical raster data should to be intersected with topologically ordered vector data and attributes. One problem is the maintenance of topology which is necessary for back transforms from raster to vector data.

In the last part of the paper some ideas on object shrinking - also called generalization - are introduced. An underlaying data structure should be of object type, which means that topological primitives must be classified and ordered with regard to different semantic contents. In this sense, a city of two-dimensional extent shrinks to a point when a medium scale map is generalized in a cartographic presentation of small scale. Some examples demonstrate this shrinking procedure.

INTRODUCTION

The information contents of remote sensing data demands for efficient preprocessing and analyses. Meanwhile the methods of remote sensing are gaining an operational mode (G. Konecny, 1989, J. Albertz, 1991), however, they still lack use afterwards. With the integration of these data in geographical information systems (GIS) an environment is given in which both sides should make profit of: on the one hand GIS can use the geometrical and non-geometrical information (attributes) of classified image data to update the database, to intersect vector data with image data or attributes with image data. On the other hand, interpretation of preprocessed image data becomes easier if vector data are integrated in the processing steps (M. Ehlers, et al., 1989, D. Fritsch, 1991).

With regard to the object definition in GIS (see Fig. 1) the remotely sensed data should be defined in a consistent reference frame. Consistency means here not having the same origin for each raster theme but it has also to correspond with the geometry of vector data. In raster the cell position x, y serves as a vehicle for the semantics, e.g., every raster theme is linked through the position which is also called spatial reference.

Fig. 1 - Object definition in GIS - a) overall model; b) feature links.
The spatial reference presupposes a rectification of the image. In many cases original or raw image data has to be preprocessed firstly by image restoration algorithms for instance by look-up tables and filtering respectively. Typical histograms of LANDSAT TM for channels 1, 2 and 3 can be seen in figure 2 - the lower part is obtained after histogram equalization. Very often contrast is improved nonlinear having the advantage that contrast is improved much in the range which contains most of the grey values. The important step in preprocessing of image data is feature extraction. This can be done by different methods as explained later on. Once the areas of the same semantic content are found they should be compressed. The data compression shrinks the physical data extent to a fraction of it. Within the postanalysis process one data set is combined with other ones, which need not be pure geometrical data, also gridded attributes of a GIS can be intersected with image data. The last box visualizes the results. Figure 3 gives an overview on the necessary steps to demonstrate the data flow and also to indicate the different branches of processing which are not fixed a priori, but the data can run through to obtain reasonable results.

In the following aspects of feature extraction will be discussed in general; the main part of the paper is quadtree coding and quadtree based intersection. Quite another point of interest is object generalization: a concept for raster geometry is shown in the last part of the paper.
1. FEATURE EXTRACTION

The extraction of features is the most important step in processing of remotely sensed images. The results are areas of the same semantic contents: vegetation in general, special vegetation characteristic, sealing up surface, areas of different heat emission and others. In the recent past many contributions were given for feature extraction (J. Albers, 1991) which resulted in classification of different reliability. In general it can be differentiated in mainly three categories:

- visual classification, which will be carried out interactively
- heuristic classification as semi-automated or fully automated process
- supervised and non-supervised classification as a feedback loop

Visual classification: Here the image scene is interpreted by the human observer. Area based features are identified and its polygons are stored. The corresponding data structure consists of points, nodes (n) and edges (e)

\[ f = f(\,n,\,e) \]

Because of the vector representation this data structure has no further relations with the original raster.

Heuristic classification: Heuristic classifications are computer-controlled and lead to semiautomated or fully automated interpretations. One example which is often used especially in Landsat TM and Kosmos KFA 1000 images is vegetation detection. The combination of both cameras uses the high spectral resolution of the Thematic Mapper and the high geometric resolution of the KFA 1000 camera. Vegetation detection is carried out by the computation of vegetation indices which have to be grouped afterwards. Different ratios for the index determination
can be applied; their results differ only slightly. A ratio commonly applied is the vegetation index proposed by C.J. Tucker (1979).

\[ v_i = \frac{IR - R}{IR + R} \quad (2) \]

in which IR is the infrared and R is the red channel. Once the features are extracted they represent altogether a more or less homogeneous raster map. The most disturbing point features might be eliminated by filtering techniques, e.g. by median filtering, so that the final output should look like figure 4.

![Homogenized raster map](image)

**Fig. 4 - Homogenized raster map.**

2. FEATURE CODING BY QUADTREES

Feature coding by quadtrees has been proposed by H. Samet (1984) who made also comprehensive investigations on data compression. The underlying idea is simple: let be given a choropleth map which should be compressed to a fraction of the original raster data. Divide the map into four subimages of equal size and store them as four line segments one after the other so that it can replace the original map. Every subimage or part of the map represented by a line segment is once again subdivided and replaced by a chain of smaller segments. This process can successively be continued down to the pixel level. Figure 5 shows the successive subdivision of a quadrangle down to its smallest homogeneous elements. The resulting tree of three levels has branches only in those nodes which are not homogeneous, therefore the end nodes of the tree are in closed correspondance with the semantic of the choropleth map.

![Quadtree coding - a simple example](image)

**Fig. 5 - Quadtree coding - a simple example.**

3. INTERSECTION WITH QUADTREES

The quadtree coding provides the necessary geometric structure for quadtree based intersection. Intersections are necessary when different quadtree coded layers are combined for a GIS data analysis. In general the orientation of quadtree coding of one layer does not know the orientation of another layer. Therefore if quadtree based intersection will be applied the orientation for every layer must be fixed a priori as shown by figure 6. Every layer (theme) contributes to the extension of the semantical dimension, which is given by the variable \( z \). The quadtree decomposition can be seen in the context of a fixed zero datum - in consequence the result is not homogeneous and not isotropic which can be neglected in particular for medium and small scaled images.

![Prerequisites for quadtree based intersections](image)

**Fig. 6 - Prerequisites for quadtree based intersections.**
The intersection process itself can be described by set theory. Let $n$ be a defined plane square also to be called overall quadtree element and $N$ the set composed of all quadtree elements with $n \in N$. The overall quadtree element is represented by

\[
\begin{align*}
    n &= n_1 \cup n_2 \cup n_3 \cup n_4 \\
    &\quad \quad \quad \quad (n_{11} \cup n_{12} \cup n_{13} \cup n_{14}) \cup n_2 \cup n_3 \cup \\
    &\quad \quad \quad \quad (n_{41} \cup n_{42} \cup n_{43} \cup n_{44}) \\
\end{align*}
\]

(3)

with

\[
\begin{align*}
    n_1, n_2, n_3, n_4 &\subset n \\
    n_{11}, n_{12}, n_{13}, n_{14} &\subset n_1 \subset n \\
    n_{41}, n_{42}, n_{43}, n_{44} &\subset n_4 \subset n \\
\end{align*}
\]

(transitive) (4)

and

\[
\begin{align*}
    n_1 \cup n &= n \\
    n_1 \cap n &= n_1
\end{align*}
\]

(5)

In general, for two intersecting basic elements $m$ and $n$ it holds

\[
m \cup n = n
\]

(6)

and

\[
m \cap n = m
\]

(7)

The intersection operation $\cap$ can be expanded to pairs of no intersecting basic elements $m$ and $n$ by setting

\[
m \cap n = 0 \in N \quad \text{(empty)}
\]

(8)

The intersection process itself is solved pairwise, what means that only two choropleth maps are intersected with each other at one moment in time. This will be illustrated by the following example.

EXAMPLE: Intersect the boundaries of municipalities with the highway layer to show which municipalities are disturbed by highways and which not (see Fig. 7).

\[
\begin{align*}
    \text{Intersection of two layers.}
\end{align*}
\]
CONCLUSION

The classification of satellite imagery leads to homogeneous raster data which are called choropleth maps. These maps can be quadtree coded to compress the data set and to allow for a basic data structure for intersections. The advantage of this data structure is the efficient algorithmization of intersections although it is not homogeneous and not isotropic. Another advantage is its tree structure which is strictly hierarchical. This hierarchy can be used in generalizations.

The procedure to generalize choropleth maps makes use of this hierarchy: it starts with a rough approximation and improves the geometry by the residuals of different grid size. Running through the tree structure the finest level contains the original quadtree structure which is dependent on the resolution a priori chosen. The applicability of this proposal must be proven in future investigations.

REFERENCES


