PHOTOGRAMMETRIC APPLICATIONS FOR GIS/LIS
UNSTRUCTURED DATA CAPTURE AND TOPOLOGY BUILDER

B. Studemann
R&D LIS/Photogrammetry
A. Chapuis
Product Manager Photogrammetry
P. Sonnenfeld
Product Manager LIS
Leica Heerbrugg PMU
CH-5035 Unterentfelden
Switzerland

ISPRS Commission IV

PURPOSE:

Spatial data are important for geographical information systems (LIS/GIS) to describe the form and dimension of the earth surface. Since geometry (topology and metrical information) as well as thematic information have to be managed over a long period of time, they have to be kept up-to-date, otherwise they loose their inestimable value.

Working with the standard photogrammetrical data capture on a feature-oriented workstation, a lot of geometrical information unfortunately is lost in the translation process between the analytical plotter and the GIS database.

Photogrammetry is in the process of changing from an instrument for data acquisition and map production to a method for collecting and updating spatial information for land information system (LIS). It is, therefore, of great importance that photogrammetric application software provides the means to create and maintain LIS data bases.

Leica is in the position to fullfill the requirements of modern photogrammetry by offering an integrated product line from data acquisition systems to sophisticated LIS software. Photogrammetry will thus give the GIS/LIS community the means to update large-area data bases accurately and within a short space of time.

KEY WORDS: Data Capture, Topology, Data Consistency, GIS/LIS, Map Revision

1. SPATIAL DATA

1.1. The Need of Spatial Data

Mankind has been collecting various data about his environment for a variety of reasons ever since time began. The main interest lies primarily in geographical and topographical data, because it describes the form and dimensions of the earth surface. The need for geographical and topographical data is growing in the 20th century.

Spatial information is also today of major interest to LIS/GIS, because it builts the background information or the skeleton of such systems.

In many countries GIS/LIS databases have been built and a great number of applications for data analysis, management and planing are now available in combination with these systems. Their main goal is to generate results and forecasts. But relevant and useful results can only be calculated on the basis of up-to-date data in order to respond correctly to the demand of the users. Otherwise the GIS/LIS database loose their inestimable value. In the near future many of these databases will have to be revised to meet these demands.

1.2. Features of the LIS/GIS Data

At first sight one gains the impression that there is no difference between spatially related systems and conventional databases. The analysis of LIS/GIS data structure shows that only specific and complex solutions are possible and must take into account the following features:

- **Spatial relativity:** (geometrical aspect of data
  LIS/GIS deals with geometric objects that are spatially fixed with respect to place and extent. **Geometric modelling** enables the metrical and topological features of geographic data to be processed and recorded.

  The topology describes the associations and relationships between the objects using topological basic figures (nodes, edges and surfaces), whilst the metrics describes the shape and extent of the object with angle, distances and parameter shapes.

- **Thematic aspect of data:**
  Thematic modelling separates objects into subject groups and processes the corresponding descriptive data and attributes.

- Consistency of complex objects:
  The checking of the consistency of geometrical complex objects is more complex than in commercial or mapping applications. It is no longer sufficient just to check the standard data types or the value areas. Topology must be unambiguous: e.g. a network of surfaces can neither show overlapping nor holes. Thematics themselves also makes demands, but demands which are dependent on the uses: e.g. roads and houses cannot intersect.

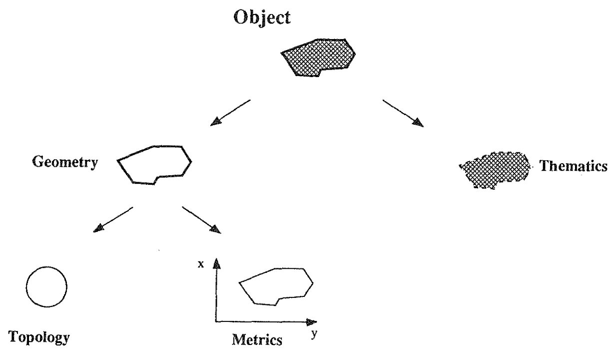

Fig. 1: Geometry and Thematics

2. DATA CAPTURE

Since geometry (topology and metrical information) as well as thematic information have to be captured, their complexity bears an influence on the method of data acquisition.

2.1. Current Photogrammetrical Data Capture

Unfortunately, most current applications of photogrammetry are connected with graphic oriented mapping systems. These standard photogrammetrical software only records line and point data without classifying topological relationships and thematic groups, i.e. it only provides a copy of the map in digital form. As a consequence of the weaknesses of these data models used hitherto, much useful information were not recorded.

Working with a photogrammetrical data capture on a feature-oriented but off-line connected workstation, a lot of geometrical information are unfortunately lost in the translation process between the analytical plotter and the GIS database.

2.2. Request to Data Acquisition Procedures

Methods of data acquisition must:

- support geometric and thematic modelling
- process complex spatial, geometrical objects
- apply complex rules of consistency
- permit relationships between objects
- be simple and quick to use and have a good cost/benefit ratio
- facilitate data acquisition over large areas

Two different procedures have to be considered with respect to data acquisition:

1. **acquisition of basic data**
   During basic data capture the operator has to record a great number of geometric and thematic data within a short time. He wants to see immediately the status of his work.

2. **up-dating spatial data bases**
   During up-dating, the operator records new data selectively. Therefore he needs to compare the old data status with the new one.

With both methods the user wants to make on-line consistency checks to establish a consistent database in order to reduce subsequent corrections. However the verification procedures are time consuming and analytical photogrammetric restitution systems are expensive.

3. DATA CAPTURE STRATEGIES

Two data capture strategies have to be differentiated:

- the unstructured data capture
- the structured data capture

Ultimately, the data at the outputs of all the data acquisition systems conforms to the data base model. The final product does not therefore represent the difference between them, but rather the way in which it is produced, and this depends on the volume and complexity of the data.

3.1. Structured Data Capture

Every element or object is entered in a single operation together with all its topological and thematic characteristics, while at the same time observing all the relevant conditions, consistencies and parameters.
Even at the digitization stage, it is of advantage to distinguish between the geometrical symbol, line and region elements, i.e. data acquisition should be structured:

- when "only few new" data have to be added to a (much larger) block of old data, because data is accessed selectively during up-dating.
- when the data is of a complex nature (e.g. each element belongs to a different thematic group and has different attributes).
- when a number of regulations and special conditions require continuous interactive intervention on the part of the user (e.g. general surveying).

3.2. Unstructured Data Capture

During the unstructured data capture the operator can digitize a large amount of data, called "spaghetti, quickly and efficiently. He only has to assign one (or several) thematic codes to each spaghetti and he is thus free from having to do structuring or description work.

He can digitize a large amount of data, quickly and efficiently, which is topologically structured and prepared for entry into the information system in a further work stage.

Functions such as "snap to line" or "snap to point" are available when digitizing spaghetti. Extensive elements do not have to be closed.

Unstructured data acquisition is to be preferred:

- when a "large volume of new" data has to be recorded, e.g. when recording data for the first time. It is of advantage not to have to pay attention to data structure or topology when recording basic data. Complicated consistency checks can be carried out later.
- when the structure of the data is uncomplicated, i.e. when there are only few (or perhaps only one) thematic groups and all the attributes are the same.
- when data acquisition can be performed automatically and structuring for the most part requires no intervention by the user.

As a solution we propose an on-line data capture, which takes place in three steps, and the extension of the photogrammetic restitution system with an image superimposition system.

4. PROCEDURE FOR UNSTRUCTURED DATA ACQUISITION

The data go through at least four states between data acquisition and becoming a fully structured, topological correct block of data:

- **State 0:** The data exist outside the system in analogue (photographs) or in raster form (digital image).

- **State 1:** The data have been acquired and are available in unstructured vector-oriented form. The unstructured data of lines are called **SPAGHETTI.** At this stage, not all the spaghetti intersections are known, some pieces of spaghetti are too short and others too long and the thematic group is only indicated by a code.

```
                 |                          /
      ----------|                         /
     /           |              ---------------
    /            |                      /
   ----------------                    /
```

  After data capture, spaghettis have to be automatically structured by a topology builder. Geometric and thematic consistencies are then checked.

- **State 2:** The data have been topologically pre-sorted and are in the form of an edge/node structure, which does not yet suit the user's purpose. The diagram shows 8 edges, but it is neither possible to recognise the area enclosed by edges 1 to 4, nor that edges 8 and 6 and edges 5 and 7 merge to lines from the users point of view.

```
                                               o
          1                                   / 5
      o---------o                     8     /
  4 /           | 2               o-------o-------o
   /            |                        /      6
  o------------o                    7 /
          3                            o
```

  Areas and lines are influenced by the choice of thematic groups, which is not taken into account at this stage. Attributes, for example, have also not been assigned yet.

- **State 3:** Structuring has been completed and areas (regions), lines and symbols are all clearly visible and the attributes have been assigned.

```
                                               o
                                              /
      o---------o                   Main / Street
     /  Plot    |                 o=======o=======o
    /   171/2   |                        /
  o------------o                        /
                                       o
```

  Objects such as regions or lines are made up by comparing the thematic codes. Any irregularities are recognised by the system and can be cleaned up with support from the program. Then if the geometry has been correctly filed, the technical data (attribute) is entered in order to complete the information for the data base.

Structuring thus takes place in three steps, which are the transitions between states 0-->1, 1-->2 and 2-->3.

5. DATA CAPTURE WITH INFOCAM/MAPIT

Both structured and unstructured acquisition have their place and must be available in an fully operative system.

INFOCAM applications concentrate on the capture, management and evaluation of geometrical - thematical information. The system is particularly suited for the following tasks:

- On-line 3D-data capture with the LEICA analytical plotters SD2000, DSR 15.
- Off line connection with the WILD BC2/BC3 analytical plotters
- 2D on-line digitizing
- topography
- real estate cadastre
- multipurpose cadastre (including networks documentation)
- digital terrain models
- automated maps and plan drawing.

The system has:

- no limitations on the size of the database in terms of information except for the physical limits on the amount storage space available.
- no limitation on project size.
- no limitations on map sheet size.

INFOCAM is available in two versions:

- INFOCAM - ORACLE/SQL for clients requiring complex interrogations and multi users systems
- INFOCAM excluding ORACLE for applications that do not require a relational database.

In both versions the database management system takes maximum account of the requirements for long term storage and rapid access for interactive graphic processing.

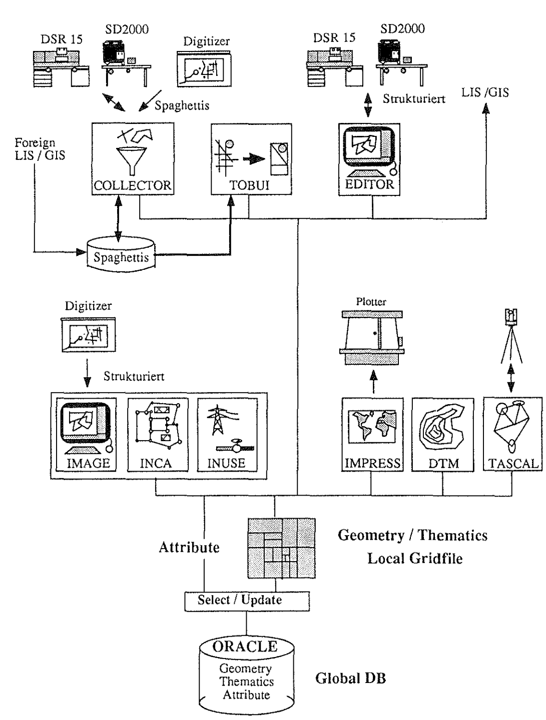

Fig 2: INFOCAM Design

INCOME (INFOCAM COMPILATION OF MAP ELEMENTS), as the photogrammetric data capture module of INFOCAM/MAPIT, allows with an analytical stereo plotter, such as the Leica SD2000, the on-line capture of large volumes of unstructured **three--dimensional** vector data (mainly POINTS and LINES) and the automatic structuring of the topology. INCOME supports fully the image superimposition of the SD2000 (COLORISS) and of the DSR 15 (KRISS). The INCOME module comprises the following three sub-modules:

- COLLECTOR
- TOBUI (TOPOLOGY BUILDER)
- EDITOR

5.1. STEP A / COLLECTOR

The most important tasks can be described as follows:

- recording of zero- and one-dimensional geometrical elements (points and sections of line called "spaghetti")
- assignment of thematic codes
- placing of region seeds for the subsequent formation of areas
- setting thematic labels for the subsequent attachment of attributes.

At the conclusion of Step A, each zero- and one-dimensional element is available on its own. The relationship between the elements has not yet been established and the connection between attributes and their subsequent owners is only implied (by their relative positions).

5.2. STEP B / BUILDER - Part 1

The most important tasks can be defined as follows:

- determination of topological relationships
- intersections of lines and the determination of nodes
- elimination of sections of lines, which occur more than once,
- marginal lengthening or shortening of lines in those cases where a node can be created within the tolerance band.

This step is performed automatically without intervention by the user. The corresponding BUILDER RULES to be observed can be defined in advance. At the conclusion, a valid edge/node structure has been completed and the data are topologically consistent. Thematic lines and areas (regions), however, have not yet been created and neither has the explicit assignment of thematic groups and attributes taken place.

5.3. STEP C / BUILDER - Part 2

The most important tasks can be defined as follows:

- automatic determination of symbol, line and region elements on the basis of thematic criteria and rules defined at the beginning of the procedure (BUILDER RULES).
- Firstly, symbol points are transformed into symbol elements. However, since a symbol point can have several thematic codes, some points will result in several symbol elements having the same geometry.
- The mosaic of edges, which was created in the preceding steps, is then reduced to line and region elements on the basis of the thematic codes attached during data acquisition.
  Connecting edges with a common thematic (line) code become a line.
  Edges, to which a region code was attached, are included in a region construction process. This process starts at a node and searches for the shortest possible route back to the starting point.

The process is repeated as many times as is necessary.

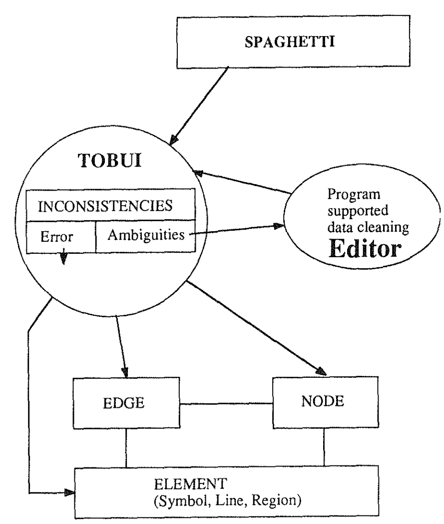

Fig 3: Data flow in the Topology Builder

Problems are registered by the system and can be processed interactively (INCOME EDITOR) after the automatic procedure has finished. Finally, the attributes are assigned to the symbol, line and region elements, which have been created during this step. At the conclusion of Step C, the data are available in a **thematically and topologically correct form.**

6. POSSIBILITIES AND LIMITATIONS OF AUTOMATIC STRUCTURING

A characteristic of INCOME is that it tolerates errors which can be both to the user's advantage or disadvantage as the case may be.

One **advantage** is most certainly that a number of "trivial" errors such as the merging of closely adjacent points and cases of slight overshooting and undershooting are corrected automatically by the system itself, and that the user does not have to explicitly define thematic significance and certain of the attributes, but simply implies them by stating the location (REGION SEEDS and LABELS).

Most of the predictable problems can be combated by establishing BUILDER's RULES, BUILDER's, which is of substantial benefit for the user.

Problems, which BUILDER is unable to resolve on its own, are correspondingly marked and dealt with in a subsequent "correction phase" (INCOME EDITOR).

The **disadvantage** of a system which tolerates errors is obvious. If data acquisition is too "careless", the number of errors, which have to be corrected interactively can become overwhelming. Thus the time gained when working with COLLECTOR is more than outweighed by the additional (generally considerably greater) time needed with EDITOR.

By far the better solution is to avoid "predictable" conflicts in advance by fully exploiting the facilities provided in COLLECTOR for increasing data acquisition accuracy (e.g. SNAP). In this respect, it would be wrong to look upon BUILDER as a remedy for all ills, which offers just the right strategy in every conceivably confused situation.

+ Bartelme N. und Spaeni, B.: "Spatial Access to Thematic and Thopological Structures in GIS", presented paper at the 16th ISPRS Congress, Kyoto, 1988.
+ Frank A.: "Requirements for Database Management System for GIS", PERS, Vol LIV, Nr 11, November 1988
+ INFOCAM Product-Information, Version 5, Leica Heerbrugg PMU, 1991.