Integration of satellite imagery using a GIS 'toolbox'

A Stevens M J Jackson W J James

Laser-Scan Laboratories Ltd.
Science Park
Cambridge
England

Commission Number I/II

Abstract

In order to satisfy current and future user needs it is argued that a toolbox of GIS components is required allowing flexible configuration and extensibility. This paper reports on the requirements of such a 'toolbox' and illustrates its functionality.

Central to the GIS is the input and maintenance of verified structured data from a variety of sources with additional information to describe its accuracy and source.

Increasingly satellite imagery is to be incorporated into GIS systems. Suitable datastructures for storage and display of the images are discussed and methods for registration to existing cartographic data described.

Practical examples of GIS use are given and methods of visualisation of pictorial data in 2D and 3D are described.

Introduction

The development of Geographic Information Systems (GIS) is not new but there has been a remarkable recent surge of academic, political and commercial interest [Jackson *et al* 88]. Many of these initiatives have been stimulated by remote sensing developments including:

- the availability of large quantities of remotely sensed data with good radiometric, temporal and geometric resolution; especially from the Landsat Thematic Mapper and French SPOT sensors.
- advances in computer processing and display technology now making practical the manipulation of large quantities of geographically referenced data for commercial applications.
- investments in space programmes, and political pressure to demonstrate commercial returns.

GIS design encompasses many fields of endeavour including spatial data handling, image processing, database technology, statistical analysis, computer science and intelligent knowledge based (IKBS) systems. Development programmes, especially of national agencies are correspondingly large (e.g. [Estes 84], [Smith *et al* 87]). However, it has been our experience that the requirements of single organisations are generally much more specific and that facilities offered by large systems are either too general or too expensive.

The development of GIS, and especially those integrating remotely sensed and other data types (so-called Integrated GIS or IGIS), is now recognised as a pre-requisite for the effective exploitation of remotely sensed data. This paper illustrates Laser-Scan's approach which is to provide an integrated framework and modules or tools which can be flexibly configured to tailor a system to requirements and budget. The toolkit concentrates on the fundamental requirements of providing simultaneous access to, and manipulation of, independent raster and vector datastructures. Facilities are provided to interface with general purpose commercial database systems at a higher level, and some examples of these are described.

IGIS data types

Geographic data exists in a wide variety of forms, but may be grouped according to ***type***. Each data type requires distinct structures and methods of handling within the IGIS. The data types are:

- Image data, commonly raster data, especially satellite imagery.
- Object data, commonly vector data, especially digital cartographic data.
- Terrain height data, commonly digital terrain models, contour/spot-height maps or surface models
- Tabular data, especially qualifying data, attributes and indexes
- Rule/knowledge data, for support of decision making processes.

The importance of qualifying data

Qualifying data should accompany any geographical dataset and may generally be divided into status and historical information. It is particularly important when bringing together geographic data from different sources.

Status information includes geographical coverage, description of content, scale, accuracy, currency etc. Laser-Scan's existing tools to input satellite imagery in a variety of formats also fully decode the 'header' records into an accessible tabular form.

Historical qualifying data references the data and procedures from which the dataset was created and modified. It allows the user to trace the sources of the data. In the case of object data, the history information is contained within the object file (see below). For operations involving data of various types, it is more appropriate to store the information in project oriented database systems. Each record will specify a procedure, the data inputs, and any other data used by the procedure. For example, a classified satellite image may record:

- creation by a segmentation process using a specified source image, reference object, terrain height and rule data, followed by
- modification through a classification process using further reference object data and a different set of rules.

Description of Image storage

The simplest form of image storage is probably a single 2D array whose indices represent the pixel coordinates and whose values represent the image pixel data. Satellite images usually consist of multi-bit data and may also be recorded in a number of frequency bands. Although specialised image processing computers are available with suitably large addressable memories, Laser-Scan aims to provide efficient image handling capability on relatively low-cost, and general purpose, workstations. This allows satellite imagery to be more closely integrated with other data processing and IGIS functions.

For modest sized images and processing tasks, 'virtual' memory techniques, which transfer portions of the image array from disk to real memory, provide acceptable performance. For more intensive operations Laser-Scan have developed a 'structured image' storage mechanism (LSI) which is a a form of grid file optimised for rapid access. A hierarchical index is superimposed on the structured image to provide summary information which may be exploited for IGIS operations.

Description of Object storage

As one component of a GIS, Laser-Scan have developed a flexible feature based storage mechanism called IFF. Features are assigned describing codes and contain strings of floating-point coordinates representing spatial position. Each coordinate can contain any number of dimensions (usually 2 or 3) and any number of per-point attributes. A whole IFF file may be separated into layers (typically to separate different groups of features) and contain ***history records*** which log changes and uses of the file. A ***feature representation table*** determines how feature codes are represented on graphic devices. Higher level objects, such as polygons and link-node structures, can be built by pointers into IFF. Applications of Laser-Scan's structured vector (LSV) format are described later in the paper.

Description of Terrain storage

Describe triangles and grid files

Laser-Scan's powerful terrain modelling software is based on Delaunay unique triangulation [McCullagh 83]. Every measured data point is honoured directly, since they form the vertices of the triangles used to model the terrain. Particular constraints allow fault or break lines to be incorporated. The triangular representation allows visualisation, editing, and storage of surface form in a compact representation. In addition the triangles may be processed for subsequent generation of regular grids or contours.

IGIS toolkit components

Data input

Modular facilities for input include:

- Format conversion — to take 'popular' digital datasets, and published interfaces and convert to IFF object data.
- Remotely sensed data — by image and header decoding; images to raster files, header information to tabular files (with a 'keyword = value' format).
- Aerial photographic data — via stereoplotters into 3D object format.

- Map digitising — Laser-Scan support manual, interactive (capturing linework and junctions) and fully automatic data capture from a wide variety of source documents. Output is into IFF object data.
- Voice recognition — for rapid coding of specific features.

Data selection and orientation

Before multiple data types can be used to satisfy IGIS enquiries, the individual data types have to be retrieved, seamlessly joined, orientated and presented.

Image data can undergo a variety of image processing operations (filtering, edge enhancement, histogram equalisation etc.). It can be geolocated by interactive matching and can be transformed within a number of standard coordinate systems. 'Panelling' operations allow a number of images to be amalgamated and window operations allow sub-images to be selected. Manipulation of colour lookup tables provide a method to change display representation.

Object data can be extensively processed by Laser-Scan's IFF manipulation utilities (including geometry checking and structuring, smoothing, statistics generation, sorting and selection). It can be transformed to a wide variety of coordinate systems. Edge matching between map sheets allow amalgamation and window operations allow data selection. Feature display on a variety of output devices can be modified by manipulation of feature representation tables.

Terrain data in triangular form can be viewed checked and edited. Constrained triangulation can be selected to honour particular geographic features such as cliff and break lines. Interpolation to user defined grids allows output as a raster image, where pixel value represents height. Edge matching between adjacent models, sub-model selection and perspective displays allow flexible output.

Of particular relevance to this paper is the registration of satellite imagery to map data. For vertical or near-vertical imagery, simple interactive registration using corresponding map and image points may be used to calculate affine transformation parameters to be applied to the imagery. For off-nadir images sensed with a linear array detector, such as SPOT, a 'dynamic space resection' approach is required [Gugan 87].

Software has been written to allow pixel images and map data to be displayed and manipulated simultaneously on a colour graphics workstation. The data may be displayed in separate windows or superimposed in a single window. An operator may then point and record corresponding locations (to sub-pixel accuracy) using the workstation mouse in map and image. Such points constitute a *registration control point* file which serves as the basis for absolute orientation calculations.

Integrated data editing

A basic tool for data manipulation within an IGIS is the editing of object(vector), image(raster) and terrain data, either separately, or in combination. To this end Laser-Scan have developed integrated editing capability allowing the simultaneous viewing of multiple data types on a single workstation screen. Using the facilities, vector data may be digitised or edited against a backdrop of image data (from a scanned map or satellite image).

Alternatively, raster operations can be achieved with reference to overlayed or adjacently displayed vector files or terrain models. Digital terrain matrix values can similarly be modified with reference to image or vector data. Tabular or textual data can easily be modified in the presence of pictorial data using the multi-window facilities of modern multi-tasking workstations.

Image/object and Object/image conversion

In some circumstances it is desirable to convert between data types:

Image to object conversion may be achieved by a number of methods, depending on the complexity of the data and the quality of the desired result. Manual digitising from a table or screen allows an operator to include 'intelligence' information such as the type of feature being digitised. It is most appropriate with poor quality documents but is time consuming. Laser-Scan's unique vector digitising equipment (LASERTRAK) uses a semi-automatic line following approach and scans a laser beam through a photographic negative of the source document. More recent developments (Laser-Scan's VTRAK equipment) apply image understanding techniques to structured images and perform feature extraction with partial feature coding. Usually some form of post-processing to structure the data is also required.

Object to image conversion is generally easier and is achieved with a 'gridding' program (I2GRID) which allows individual features to be assigned pixel widths. The program allows the final pixel size to be chosen and generates binary data.

Data output

Modular facilities for output include:

- Conversion to popular formats and published interfaces, allowing users to generate their own output datasets.
- Display and plotting of combined data (e.g. 3-D views comprising image, object and terrain data) using raster and vector colour plotters and photoplotters.
- Generation of cartographic products.
- Generation of statistical or other tabular data, against specific criteria. This is achieved via libraries of standard facilities which can be enhanced by user programs. These can be used, for example, to derive statistics about average length of features within particular datasets or average greylevel of an image.

Data structuring and polygon formation

Figure 1 shows a typical example of unstructured data to be included in an IGIS. It consists of linear road features and polygons, such as might be obtained from manual digitising of geolocated satellite imagery. As can be seen, in its unchecked form it contains a number of digitising errors. The locations of the polygons text codes may later be used later as 'seed' points. The codes themselves may be used as a pointer to classification or descriptive data stored in a relational database.

To integrate such data into an IGIS, Laser-Scan have developed a powerful set of editing and structuring facilities which may be linked together to process the ('spaghetti') data into true polygons with minimal operator intervention. The steps are as follows:

- Removal of double digitising — common boundaries are aligned and multiply digitised features are removed. Specification of immovable features may be made.
- Extending and clipping of lines — junctions of specified features are tidied, within given tolerances.
- Polygon closing — again within tolerances and for specified features line ends may be joined.
- Feature breaking at junctions — This creates a link and node structure where each link between junctions becomes a separate entity. Each of the links is recorded once only.

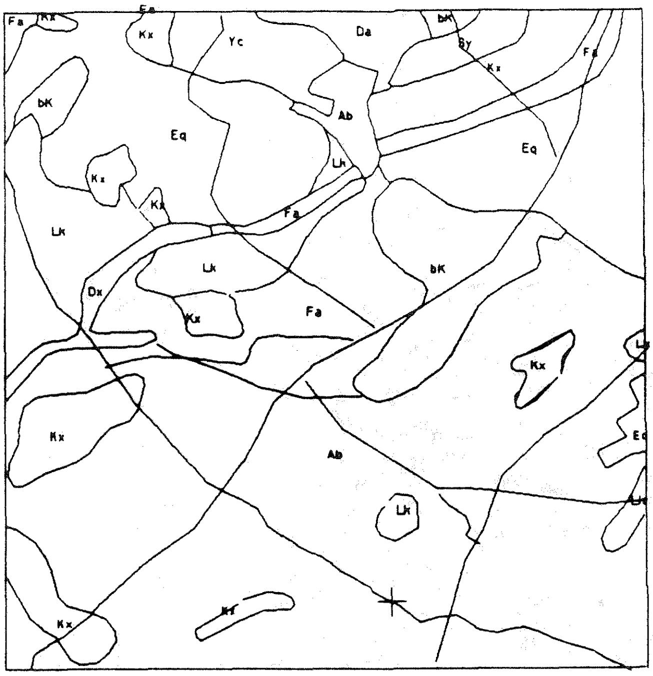

Figure 1: Unstructured polygon data with roads

- Polygon formation — The above links and nodes are processed into polygons. Where necessary the above links are duplicated and each polygon becomes a separate feature.

Figure 2 shows the unstructured data processed into polygons and filled with different patterns using internal seed points. Information about areas, bounding boxes, or perimeter length is easily derived by standard routines and may be interfaced to a wide variety of statistical analysis packages.

Digital elevation modelling and visualisation

A Digital Elevation Model (DEM) refers here to a set of elevation data recorded on a regular rectangular grid. Terrain models may be derived directly from stereo satellite imagery [Alvey] or from existing map data.

Software modules exist to calculate and manipulate absolute height, slope and aspect, and to generate shaded overlay perspective views and visibility profiles. These parameters and others such as terrain roughness are necessary in considering many environmental issues. Examples include the distribution of selected flora and fauna, modelling snow melt or run off characteristics, geological interpretation and cross-country mobility studies. The surface form is also important to image classification, particularly with the greater look angles which are obtained with TM, off-nadir SPOT data, SAR images and airborne scanners.

Despite the increased complexity of 3-D systems compared with 'flat' storage, a future trend is anticipated towards representations and display of the third dimension within IGIS systems.

Figure 2: Structured polygon data from figure 1

Flowline and Database Capabilities

Laser-Scan have identified a number of levels of data organisation which are appropriate for different users and applications. These can be grouped as follows:

- Use of individual packages and system facilities — Starting from a core system, individual tools can be added without changing the basic computer hardware. Use of operating system facilities, directories and protection mechanisms, along with specifically written software can produce a well tailored and cost-effective environment.
- Flowlines constructed from command scripts — To integrate individual GIS components for specific tasks, such as map production, command script facilities are used to link primitive operations. This level of information management is appropriate for many national and military mapping production agencies.
- Data management with relational databases — Production flowlines often benefit from the use of relational databases to keep track of data flow. Laser-Scan has recently implemented a map production flowline for a major customer integrating data from a variety of sources. This incorporates a data catalogue and allows browse, search and select mechanisms to find relevant local or remote datasets. In a separate development, a relational database has been used to organise descriptive data relevant to geographic areas with specific report generation facilities.
- Output to map information management systems — Laser-Scan market Informap III from Synercom Technology. This is a fully supported product on the DEC range of workstations. It includes engineering facility and work management, geographic information processing and analysis, database management with a continuous map environment and report generation facilities.

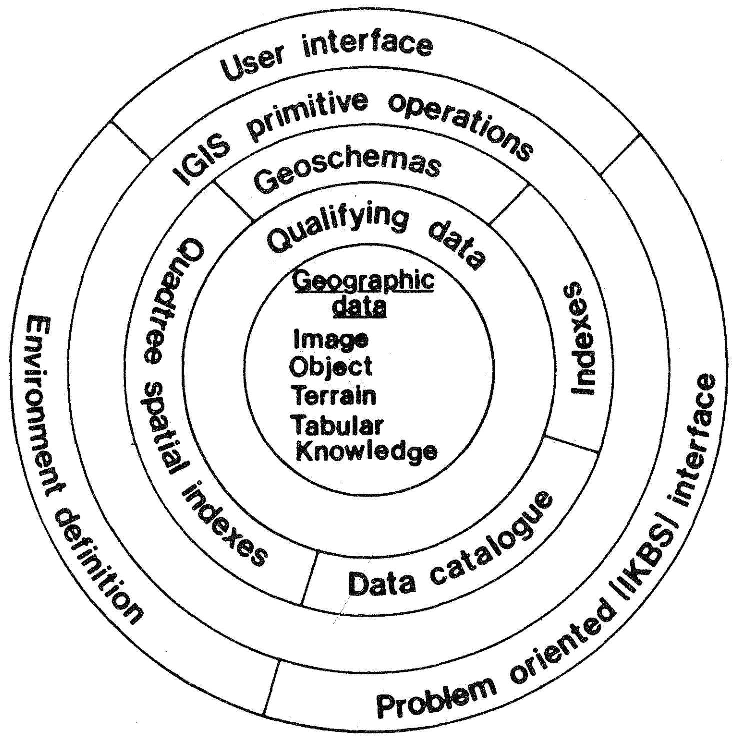

Figure 3: Schematic of integrated GIS system

Current development directions

IGIS Data Organisation

In order to use most effectively the tools described above Laser-Scan is continuing the development of structures for overall data organisation. These permit the maintenance of data in its primitive input form to preserve precision and information content, whilst offering the user an integrated view of the whole data environment. In addition, the structures must provide fast and efficient spatial processing. A schematic of the organisation proposed is presented in Figure 3 and has been described in [Jackson *et al* 86].

The term *geoschema* is analogous to the use of 'schema' within a traditional database. It describes the organisation of the geographical datasets. In particular it defines the spatial referencing system, indexes, classification and attribute coding schemes.

Quadtree spatial indexing

Much of the research effort to date on integrated GISs has concentrated on low level data structures (such as the quad-tree [Samet 84]) which can hold both vector and raster data in a uniform manner. While this use of quad-trees may be justified for certain types of operation, especially in the context of pre-defined projects, they do not offer a sensible general purpose solution to IGIS integration. However, they do have benefits in providing an integrated *spatial index* [Ibbs and Stevens 87].

At the highest level of area storage, the operator is aware of one common spatial index into the data. At 'lower' levels the index contains references from the structure into the particular (type specific) datasets. Obviously, the design and implementation of the spatial index and the underlying datastructures have a significant effect on the speed and flexibility of the IGIS system.

Hardware and User Interface

The design of a consistent and appropriate user interface is rightly acknowledged as a vital link in any IGIS. All of the facilities for image integration described above are available on the same hardware — the DEC range of colour workstations. This common hardware environment needs to be complimented with a common software interface across the wide spectrum of available facilities. Much work has been done in this area with WYSIWYG (what -you-see-is-what-you-get) menu and icon interfaces and on-line help facilities. Emerging 3D graphics and window standards and object oriented programming techniques allow further sophistication, which if sensitively applied offer enhanced capability with portability.

In addition specific new hardware and software developments for 2.5D, 3D and stereo vision are being researched. These include an Alvey project [Muller *et al* 88] involving the transputer [Inmos 86] as a flexible general purpose vision engine, and more specific graphic displays [BCS 87]. Such devices offer the computational power necessary for real-time display and animation.

Summary

As functional capabilities increase, making use of 3-D data sets, IKBS techniques and special purpose hardware, an integrated approach to GIS design will become necessary to control the large volumes of data involved. In particular IGIS qualifying data is important to understand the pedigree and reliability of any output.

Laser-Scan have well proven utilities and tools for storage, retrieval and manipulation of spatial data. The approach presented here allows a flexible and upgradeable route towards a general purpose IGIS.

Bibliography

References

[Alvey] *Real time 2.5D vision System.* Alvey MMI project 137. Participants: Thorn EMI Central Research Laboratories, Laser-scan Laboratories Ltd., University College London, RSRE Malvern.

[BCS 87] *3D Graphics Hardware.* BCS display group documentation 1987. British Computer Society. London.

[Estes 84] J Estes. *The pilot land data system: Report of the program planning workshop.* NASA Technical memorandum 86250.

[Gugan 87] D. Gugan *Practical aspects of topographic mapping from SPOT imagery* Photogrammetric Record vol 12(69), April 1987, pages 349–355

[Ibbs and Stevens 87] T. J. Ibbs and A. Stevens *Quadtree storage of Vector data for Geographic Information Systems* Presented at: GIS-LP '87—International Conference on Geographic Information Systems for Land Planning, Lisbon, 14–18 September 1987

[Inmos 86] *The transputer Family.* Product Information. March 1986. Inmos.

[Jackson *et al* 86] Jackson M J, James W J, Stevens A. *British National Space Centre GIS Demonstrator Feasibility Study* BNSC Report 1987.

[Jackson *et al* 88] M. J. Jackson, W. J. James, A. Stevens *The design of environmental Geographic information systems* Royal Society —in press.

[Muller *et al* 88] J-P Muller, K A Collins, G P Otto, J B G Roberts. *Structure of stereo matching using transputer arrays.* Presented at ISPRS July 1988.

[McCullagh 83] M J McCullagh. *Transformation of contour strings to a regular grid based digital elevation model.* Euro-Carto II. 18pps.

[Samet 1984] H. Samet. *The quadtree and related hierarchical data structures.* A C M Computing Surveys. **16**, 67pp, 2 June 1984. (also University of Maryland Computer Science TR-1329).

[Smith *et al* 87] T R Smith, S Menon, J Star and J E Estes. *Requirements and principles for the implementation and construction of large-scale geographic information systems.* Int. J. Geographical Information Systems, 1987, Vol. 1, No. 1, 13-31.