Expert Systems and Spatial Data Models for Efficient Geographic Data Handling

William J. Ripple and *Veit S. Ulshoefer*
Environmental Remote Sensing Applications Laboratory (ERSAL) and Department of Geography, Oregon State University, Corvallis, OR 97331

ABSTRACT: Existing geographic information systems need to be more flexible and efficient to effectively handle large quantities of spatial data acquired by remote sensing systems and from conventional map sources. Partial solutions to this problem may come from the implementation of improved spatial data models in expert systems. It is suggested that expert systems offer possibilities for making geographic information systems more efficient and user-friendly. Improved spatial data models for expert geographic information systems include hierarchical tesselation models. Research is suggested to determine the most appropriate spatial data model.

INTRODUCTION

THE QUANTITY AND VARIETY of spatially referenced data to be stored, manipulated, and displayed using geographic information systems (GIS) is rapidly increasing. Data processing times are often slow due to inefficiencies in data handling procedures. Current GIS are also limited in their ability to effectively process spatial data from remote sensing sources as well as from a wide variety of traditional map sources. This results from the fact that data from remote sensing sources are encoded in raster format, whereas many traditional map sources are digitally represented in vector format.

Despite many inefficiencies, GIS are becoming widely used in natural resource analysis and management. A major problem of current GIS is that they have to be operated by "experts" with experience in the study of complex spatial relationships using computerized systems. User-friendly systems that could assist a user in solving a problem would greatly improve the overall application and performance of GIS.

Partial solutions to some of these problems seem readily apparent. The development of expert systems for GIS should allow for more efficient data processing and analysis, as well as help users with little experience in computerized spatial data processing and modeling to become proficient users of GIS. The problem of accommodating spatial data from a wide variety of sources may be solved through the development of new spatial data models. These models should be compatible with artificial intelligence methods for the reduction of search space and, therefore, complement the development of expert systems for GIS.

The following sections give an introduction to expert systems and their applicability in a GIS context, followed by a discussion of the role of improved spatial data models for expert GIS.

EXPERT SYSTEMS FOR GIS

The term "expert system" is difficult to define. Feigenbaum (1984) asserts that an expert system is a computer program that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise for their solution. Expert systems are based on knowledge which is acquired from experts in a specific domain. The domain knowledge is stored in the knowledge-base of the system. The knowledge is applied and processed by an inference engine which controls the reasoning of the expert system. The separation of knowledge and procedures of applying the knowledge is one of the main characteristics of expert systems (Duda and Gaschnig, 1981). This makes it easy to change or update the system as new knowledge is acquired. Other qualifications of expert systems are the use of a natural language interface and the ability to explain their line of reasoning. Areas currently under active research are the problem of reasoning under uncertainty and the development of learning capabilities of expert systems (Buchanan and Shortliffe, 1984).

The complexity of geographically referenced data and the large data bases to be manipulated make GIS a good candidate for the application of artificial intelligence techniques. Expert systems have been built and successfully used for a variety of GIS related tasks. Robinson and Frank (1987) provide a review in this volume of how expert systems have recently been applied to problems in GIS, including map design, terrain/feature extraction, geographic database management, and geographic decision support. Estes *et al.* (1986) discuss applications of expert systems to remote sensing. An expert system has recently been built for land-cover change detection (Wang and Newkirk, 1987). These authors demonstrate how expert system concepts can be applied to GIS and related problems. There are some subsystems of GIS that seem particularly amenable to expert systems:

- *Intelligent User Interface* - to guide an inexperienced user through the most efficient use of the system (Jackson and Mason, 1986);
- *Image Classification* - to increase the consistency of image classifications, and to improve the information extraction potential from remotely sensed images by automated merging of spatial ancillary data with spectral data;
- *Database Search* - to make the search of large geographic databases more efficient by using heuristic search methods, i.e., search methods based on judgemental rules, which eliminate major portions of the database from consideration as early as possible (Peuquet, 1984a);
- *Learning Capability* - to allow results of computationally expensive queries to be added to the knowledge-base to process frequent queries faster (Smith and Pazner, 1984); and
- *Cartographic Output* - to produce high quality maps and graphs and avoid misinterpretation. Rules about cartographic design can be included in the system that guide a user through the production of cartographic products.

A framework for an expert GIS is suggested in Figure 1. The intelligent user interface interacts with the user in a problem-oriented language, such as a restricted version of English (Hayes-Roth *et al.*, 1983). It guides the inexperienced user through an application session. The user interface also acts as a query parser (i.e., it translates the high-level user input into a form that can be understood by the system, identifies a task, and defines a goal that is to be reached by the system).

The inference engine is the problem solving mechanism of the system. The inference engine spits the overall goal of a query into lower-level subgoals and controls the reasoning steps taken by the system. The working memory contains relevant data and intermediate hypotheses for the current problem. The justifier explains the line of reasoning of the system upon user request. This is commonly performed by collecting the rules that have been used to solve the problem and translating them into English for presentation to the user (Hayes-Roth *et al.*, 1983). The knowledge-base contains the facts and rules of the

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 53, No. 10, October 1987, pp. 1431–1433.
0099-1112/87/5310–1431$02.25/0

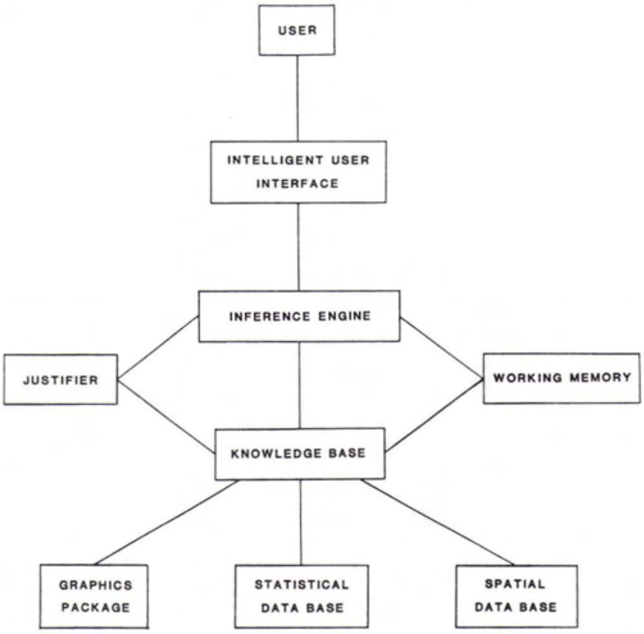

FIG. 1. Conceptualization of an expert geographic information system.

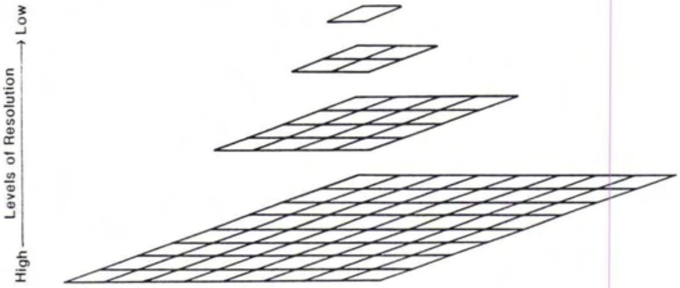

FIG. 2. Example of a pyramid as multiple resolution model.

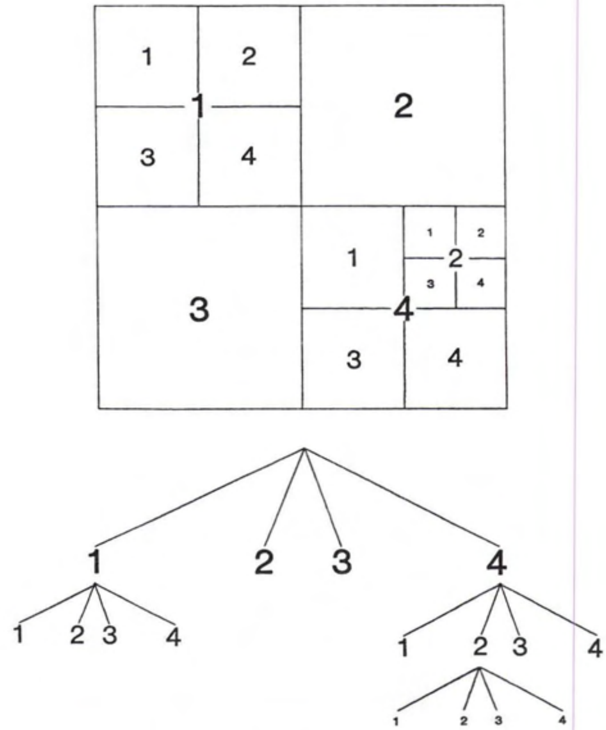

FIG. 3. Example of the decomposition of a square grid data plane (top) and a corresponding quadtree showing its hierarchical structure (bottom).

system. These represent knowledge from the areas of image processing and classification, database search, cartography, and learning as well as knowledge related to a problem domain, such as forestry or planning. Rules of the knowledge-base may invoke algorithms that manipulate data of the spatial and statistical databases. Standardized interface procedures are required for communication among the knowledge-base, the other databases, and the graphics package (Robinson and Jackson, 1985).

SPATIAL DATA MODELS FOR EXPERT GIS

The development of an appropriate spatial data model is one of the foremost problems to be solved in the construction of expert GIS. Spatial data models also need to consider the integration of remote sensing and conventional map data in a GIS. Major problems continue to occur within existing raster and vector models in terms of efficiency, storage volume, manipulation, and flexibility of application (Peuquet 1984a). Recent developments in the use of hierarchical tesselation models provide hope for more efficient systems in the future. Tesselations are the decomposition of the data plane into polygons. In regular tesselations, the data plane is subdivided into equal sized and equal shaped polygons, such as squares, triangles, or hexagons. A regular hierarchical tesselation is based on the recursive decomposition of the data plane: larger cells are subdivided into smaller cells of the same shape. Regular hierarchical tesselations typically involve the use of a square mesh, because triangles and hexagons cannot be recursively subdivided with the maintenance of both shape and orientation (Peuquet, 1984b). Irregular tesselations are also possible as variable resolution data planes, with the most common application in terrain modeling using the triangulated irregular network (TIN). Irregular tesselations, however, can be computationally expensive and, therefore, may not be the most appropriate data model for solving efficiency problems discussed here.

Two regular hierarchical tesselation models are the pyramid and the quadtree. The pyramid model uses an exponential stack of discrete arrays, each one-fourth the size of the previous array (Peuquet, 1984b). Data values higher in the pyramid (i.e., at a lower cell resolution) are based on averages from aggregated cells lower in the pyramid (Figure 2). The pyramid model has been used for image processing to speed up operations involving edge detection and the isolation of objects (Tanimoto and Pavlidis, 1975).

The quadtree model is based on the recursive decomposition of a square grid with the resolution dictated by the spatial heterogeneity of the data. The grid is recursively subdivided in selected areas of the data plane until every grid cell represents a homogeneous area at the lowest possible resolution (Figure 3). The pyramid is a multiple resolution model while the quadtree is a variable resolution model (Samet *et al.*, 1986).

The quadtree model has become very popular as a subject of research relating to GIS development (Mark and Lauzon, 1984; Palimaka *et al.*, 1986). A major advantage of using quadtree models in expert GIS is that their hierarchical nature is compatible with problem solving techniques in artificial intelligence based on traversing trees using Boolean set operations (Peuquet, 1986).

Other advantages of spatial quadtree models for use in GIS as discussed by Smith *et al.* (1987), include

- Spatial relationships are implicitly coded in the model;
- They allow for faster search and retrieval of data using higher levels in the quadtree;
- Data compactness increases as the homogeneity of the data plane increases; for example, if four sibling cells have the same value, they are replaced by a parent cell at the next higher level in the tree;
- The resolution is variable in that the data plane may be viewed

at a low resolution for browsing using higher levels in the quadtree; and

- The structure of decomposition of a quadtree allows for efficient storage of, and access to, the data by area.

Less attention has been paid to the disadvantages of quadtrees. A primary question involves the efficiency of representing vector data within a quadtree. Different types of quadtrees may be necessary for holding point, line, and raster data (Jackson and Mason, 1986). Waugh (1986) presents other disadvantages of quadtree models, including

- Quadtrees are currently expensive to create and unsuitable for certain types of manipulations, such as set operations on heterogeneous data;
- Changes in the data plane require recalculating the quadtree; and
- They are not efficient for the storage of heterogeneous data such as unclassified satellite or grid digital terrain data.

Recursive hierarchical tesselation models hold promise for the future, but machine independent theoretical research is needed on the spatial data models that were conceptually described by Peuquet (1984b). A hybrid model that offers both quadtree and pyramid characteristics was suggested by Jackson and Mason (1986). Peuquet (1984b) proposed a "vaster" (combined from the words *vector* and *raster*) model that incorporates the virtues of both raster and vector data. Antony and Emmerman (1986) suggested a quadtree/frame hybrid model for GIS based on artificial intelligence techniques. However, the most appropriate spatial data model for expert GIS is still to be determined, and may take the form of a hybrid model or may evolve from an existing model.

CONCLUSIONS

There is a need to focus concurrently on efficient and flexible methods for both storing and manipulating digital geographic data. Fundamental research in the development of expert systems using improved spatial data models promises to be of great value for the development of better GIS. Expert systems offer possibilities for making GIS computationally efficient and user-friendly using expert knowledge and high-level reasoning procedures. The storage and analysis of spatial data in expert GIS can be made more efficient using a common spatial data model, such as a hierarchical tesselation model, for incorporating spatial data from a variety of sources.

ACKNOWLEDGMENTS

The authors would like to thank A. Jon Kimerling, Thomas L. Logan, Vincent B. Robinson, and Barry J. Schrumpf for their helpful suggestions for improvement of this paper.

REFERENCES

Antony, R., and P.J. Emmerman, 1986. Spatial Reasoning and Knowledge Representation. In B.K. Opitz (ed.), *Geographic Information Systems in Government*, Vol. 2, A. Deepak Publishing, Hampton, Virginia, pp. 795–813.

Buchanan, B.G., and E.H. Shortliffe (eds.), 1984. *Rule-Based Expert Systems. The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, Massachusetts, 748 p.

Duda, R.O., and J.G. Gaschnig, 1981. Knowledge-Based Expert Systems Come of Age. *Byte*, Vol. 6, No. 9, pp. 238–281.

Estes, J.E., C. Sailer, and L.R. Tinney, 1986. Applications of Artificial Intelligence Techniques to Remote Sensing. *The Professional Geographer*, Vol. 38, No. 2, pp. 133–141.

Feigenbaum, E.A., 1984. Knowledge Engineering: The Applied Side of Artificial Intelligence. *Annals of the New York Academy of Science*, Vol. 426, pp. 91–107.

Hayes-Roth, F., D.A. Waterman, and D.B. Lenat, 1983. An Overview of Expert Systems. In F. Hayes-Roth, D.A. Waterman, and D.B. Lenat (eds.), *Building Expert Systems*, Addison-Wesley, Reading, Massachusetts, pp. 3–29.

Jackson, M.J., and D.C. Mason, 1986. The Development of Integrated Geo-Information Systems. *International Journal of Remote Sensing*, Vol. 7, No. 6, pp. 723–740.

Mark, D.M., and J.P. Lauzon, 1984. Linear Quadtrees for Geographic Information Systems. *Proceedings of the International Symposium on Spatial Data Handling*, Zurich, Switzerland, pp. 412–430.

Palimaka, J., O. Halustchak, and W. Walker. 1986. Integration of a Spatial and Relational Database within a Geographic Information System. *Proceedings of the American Society of Photogrammetry and Remote Sensing - American Congress on Surveying and Mapping Annual Convention*, Washington, D.C., pp. 131–140.

Peuquet, D.J., 1984a. Data Structures for a Knowledge-Based Geographic Information System. *Proceedings of the International Symposium on Spatial Data Handling*, Zurich, Switzerland, pp. 372–391.

———, 1984b. A conceptual Framework and Comparison of Spatial Data Models. *Cartographica*, Vol. 21, No. 4, pp. 66–113.

———, 1986. The use of Spatial Relationships to Aid Spatial Database Retrieval. *Proceedings of the Second International Symposium on Spatial Data Handling*, Seattle, Washington, pp. 459–471.

Robinson, G., and M. Jackson, 1985. Expert Systems in Map Design. *Proceedings of the AUTO-CARTO 8 Symposium*, Washington, D.C., pp. 430–439.

Robinson, V.B., and A.U. Frank, 1987. Expert Systems for Geographic Information Systems. *Photogrammetric Engineering and Remote Sensing*, Vol. 53, No. 10, pp. 1435–1441.

Samet, H., C.A. Shaffer, R.C. Nelson, Y. Huang, K. Fujimura, and A. Rosenfeld, 1986. Recent Developments in Quadtree-Based Geographic Information Systems. *Proceedings of the Second International Symposium on Spatial Data Handling*, Seattle, Washington, pp. 15–32.

Smith, T.R., and M. Pazner, 1984. Knowledge-Based Control of Search and Learning in a Large-Scale GIS. *Proceedings of the International Symposium on Spatial Data Handling*, Zurich, Switzerland, pp. 498–519.

Smith, T.R., S. Menon, J.L. Star, and J.E. Estes, 1987. Requirements and Principles for the Implementation and Construction of Large-Scale Geographic Information Systems. *International Journal of Geographical Information Systems*, Vol. 1, No. 1, pp. 13–31.

Tanimoto, S., and T. Pavlidis, 1975. A Hierarchical Data Structure for Picture Processing. *Computer Graphics and Image Processing*, Vol. 4, pp. 104–119.

Wang, F., and R. Newkirk, 1987. A GIS-Supported Digital Remote Sensing Land Cover Change Detection System. *Proceedings of the American Society of Photogrammetry and Remote Sensing - American Congress on Surveying and Mapping Annual Convention*, Vol. 6, Baltimore, Maryland, pp. 53–62.

Waugh, T.C., 1986. A Response to Recent Papers and Articles on the use of Quadtrees for Geographic Information Systems. *Proceedings of the Second International Symposium on Spatial Data Handling*, Seattle, Washington, pp. 33–37.