FUNCTIONAL COMPONENTS OF A SPATIAL DATA PROCESSOR

Stephen C. Guptill
National Mapping Division
U.S. Geological Survey
Reston, Virginia 22092

ABSTRACT

A study to determine the feasibility of developing a spatial data processor (a standalone, office environment, image processing and geographic information system) capable of manipulating imagery and digital cartographic data has been completed. A key phase of this study involved determining the set of operational functions required for spatial data processing. A set of 16 functional components was identified. This set provides useful criteria for not only evaluating existing systems but also determining the capabilities of new systems. The extent to which each functional component is used is dependent on the user application. A review of several existing systems determined that no one system provides all functional capabilities working with either raster or vector data. Although the majority of current image processing and geographic information system applications can be handled by using raster data, the functions of a spatial data processor require both raster and vector data types. The most expeditious way to build a spatial data processor would be to utilize existing raster-processing software functions; identify functions that require vector capabilities and develop and incorporate software modules to perform those tasks; and integrate the raster and vector functions.

BACKGROUND AND OBJECTIVES

Digital cartographic data bases and digital imagery holdings are increasing rapidly in size and becoming more widespread in use. Image processing and geographic information system (GIS) technology is used to merge, integrate, and analyze data from these data bases, and the demand for such systems is increasing. Although a number of systems exist, they are implemented on a wide variety of hardware configurations and do not have an integrated approach to performing both image processing and GIS functions with both raster and vector data. Two additional desirable criteria, that the software be in the public domain and that the hardware be relatively low-cost and able to operate in an office environment, are met by few systems. The lack of a system that meets these general requirements has led towards multiple efforts to develop standalone systems

---

Publication authorized by the Director, U.S. Geological Survey, on January 4, 1985. Any use of trade names and trademarks in this publication is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

to do either GIS or image processing functions. In response to this situation the National Mapping Division (NMD), U.S. Geological Survey (USGS), conducted a study of these topics.

The study determined the feasibility of developing a spatial data processor (SDP)(a standalone, office environment, integrated, image processing and geographic information system) capable of manipulating imagery and digital cartographic data. The portions of the study concerning assumptions, functional requirements, hardware characteristics, and findings (after studying a number of existing systems) are presented here.

ASSUMPTIONS

In conducting the SDP feasibility study, certain initial assumptions about the system were made. These assumptions concerned issues relating to users, functions, data, hardware, operating environment, and system architecture.

Users

The users of an SDP are assumed to represent a wide range of interests and expertise. They could range from scientists in a research environment to technicians in a production center. Although an SDP design could be optimized for specific applications, a single design cannot optimally perform tasks in research, production, demonstration, and application with USGS and other data sources. These four tasks conflict in optimization requirements. Production operations must be focused to attain maximum throughput and allow adequate data security. Research operations must be diverse and varied to allow the interplay of ideas and operations to stimulate new developments. A system to use USGS digital cartographic data must be flexible and accessible. Where design conflicts were discovered, preference was given to performing research operations rather than production operations.

Functions

The varied users of an SDP are assumed to require practically the entire range of GIS and image processing functions. These functions have been limited, however, to those available today, proven possible by a demonstrated implementation. The first implementation of an SDP is not expected to extend the state-of-the-art functionally in a significant way, except possibly to integrate functions previously available only in separate existing systems. However, it is assumed that the sum of all functions required by all users is possible in one system. This one system may actually be a family of systems, with a given function modularized in perhaps more than one implementation, each implementation optimized for a given set of user/performance specifications.

Data

The SDP will handle any data considered image, geographic, or cartographic and will readily accept USGS digital cartographic and image products. The SDP is expected to accommodate all or most existing models of these spatial data to be able to utilize and integrate data from the widest possible range of sources. A full range of non-spatial-attribute data handling facilities is also assumed.

The internal data representations employed by the SDP might, by necessity, imply innovative design approaches. The integration of multi-user requirements and various data models will place an enormous demand on flexibility and interchangability of data. The design of the internal data models and data management approaches is a key to creating an SDP for a wide range of user environments.

Hardware

The hardware necessary to perform all functions for all users at satisfactory performance levels is assumed to be available at a cost of $50,000 to $100,000. A basic system configuration can be outlined, with various hardware options added or interchanged to optimize the configuration for a given class of users. Once again, modularity is important. It may be that the only hardware component which remains constant through all SDP configurations is the central processing unit.

Operating Environment

The set of SDP software is expected to operate as an application within an existing vendor-supplied operating system and using standard data processing support facilities. The operating system or utilities will not be modified except perhaps with respect to device drivers. One operating system will be chosen within which all SDP configurations will run.

System Architecture

An SDP is assumed to be a standalone system having the capability of performing all functions locally. With the increasing availability of various local-area-networking and longer-range distributed processing options, the system design incorporates the ability to participate in a distributed processing environment. For some requirements, such as those associated with production tasks, distributed processing may be preferred.

Modularity is a possible key to solving the problem of multiple user scenarios. Modularity is assumed to be required for both hardware and software components. Conflicting requirements may be resolved by the same or similar function implemented in different, optimized, hardware and (or) software configurations.

FUNCTIONAL REQUIREMENTS

A functional analysis forms the basis of the SDP feasibility study. Once the objectives and assumptions are specified, the functional components of an SDP need to be determined. The functional components list (and companion set of data models needed by the functional components) is a generic description of the capabilities that an SDP should have.

The investigation into the operational functions required for spatial data processing began by detailing the components of five major GIS processes: data input, data editing, data storage and retrieval, data manipulation, and map, image, and report generation. Each component has counterparts that work on vector or raster data types.

The resulting lists of components lacked a unifying theme and varied in level of detail. Therefore, another approach was taken in which the underlying spatial data models were defined, and then the processes that operate on the data models were developed.

Data Models

Two schemas are suggested for defining spatial data model types--one for vector data, the other for raster or grid data. With the vector schema, two factors are considered in examining the data model: the model's suitability for handling point data, line data, area data, and mixed cases of these data types; and the degree of topological structuring (the ability to address points of intersection among lines, common edges between areas, etc.) The vector schema may also employ alternate representations of two-dimensional (cartesian) space, such as Generalized Balanced Ternary (Gibson and Lucas, 1982) addresses, or Peano keys (Marvin White, 1984, written commun.).

Two factors are also considered with the raster schema: the cell shape (rectangular, triangular, hexagonal, or some other n-sided shape); and whether the data are explicitly coded cell by cell or whether some run-length encoding is employed.

With either schema, three additional factors need to be examined: the model's accommodation of features; the model's accommodation of attribute information; and the model's accommodation of temporal changes.

There are three ways each schema may accommodate features: data oriented to explicit features (data may be directly accessed by individual feature); explicit feature tags (individual features may be extracted by exhaustive search of the set of vector topological elements, raster pixels, or grid cells); and feature encoding included within attribute classes (in which case individual features are not explicitly encoded).

Among the factors to consider in handling attribute information are: the number of codes or amount of attribute data which may be associated with a given feature, topological element, or raster pixel/grid cell; and the types of attribute information handled--either numeric (with measurement level specified as nominal, ordinal, interval, or ratio) or alphanumeric (text).

The manner in which temporal data are handled may be specified by the types of other data that carry temporal information: entire map or overlay set; spatial or other subsets of a map/overlay; individual features or topological elements; and individual attributes.

Functional Components

The earlier lists of spatial data processes and subprocesses were regrouped into generic functional components which are applicable to either a raster or vector data model. This list of functional components is given in Table 1. These functional components and seven system-level evaluation criteria (functionality, operating environment, performance, long-term support, modularity, software transportability, and expandability) were used to evaluate various existing spatial data processing systems.

Hardware Considerations

A general set of hardware requirements was developed during the study. These requirements include: 32-bit CPU with virtual memory; color raster and vector displays (512 x 512 resolution for imagery, 1024 x 1024 resolution for line graphics); and bus design to support mass storage devices (Winchester disk, tape drives), local area networks, and various graphic and alphanumeric input and output.

FINDINGS

The systems examined during this study covered the gamut of functions desired in an SDP. No single system provides all functions; however, some generalizations may be drawn from an examination of several existing systems. It is evident that raster-based systems are more standardized in functionality and more consistent in data base design and structure. For these reasons the raster processing components of a SDP can probably be extracted directly from existing systems. A system that ranks high in most of these functional components is the USGS Mini Image Processing System (MIPS) (Chavez, 1984).

The vector-based systems are extremely diverse both in functionality and data base structure. The systems have various methods for encoding and storing the attribute, coordinate, and topological information. It will be difficult to draw from these existing systems to support an SDP. Certainly, no existing vector system can supply all the requirements of the SDP vector schema in a modular and easily transportable form.

The development of the SDP functional components, the existing system studies, and the list of hardware considerations allowed a consensus to be reached in a number of areas. These findings are summarized as follows:

- Functionality and software base are more important than the performance specifications of any given hardware/operating system configuration.
- Acceptable hardware configurations are available for \$50,000 to \$75,000.
- Final choice of hardware and operating environment will be driven by the existing software modules that are chosen to form the basic components from which a system will be constructed.

Table 1.--Functional components of an SDP

Data Capture: assembling analog source data in digital form; for example, line digitizing, raster scanning.

Structuring: processing data from intial digital form into a resident model; for example, skeletonizing scanned line data, deriving topology, polygon chaining.

Editing: inserting, deleting, and changing attribute and geometric elements to correct and (or) update model; for example, node snapping, sensor noise removal.

Representation/Structure Conversion: moving between representations and the structures associated with them; for example, raster-to/from-vector, polygon-to/from-grid, digital elevation model-to/from-contour, polygon-to/from-arc-node.

Geometric Correction: fixing model to ground or image space in some referencing system; for example, adjustment of map or image to control points.

Projection Conversion: transforming coordinates between alternative referencing systems; for example, geographic-to/from-UTM.

Spatial Definition: paneling and clipping to achieve the spatial limits for data in a model; for example, limiting data to within a county boundary.

Generalization: reducing detail in the model; for example, resampling to larger spacing, reduction of points in a line.

Enhancement: modification of detail in the model; for example, edge definition.

Classification: analysis and interpolation of the model to form classes; for example, classification of spectral response data, choropleth mapping.

Statistical Generation: deriving descriptive statistics and (or) measurements from model; for example, histograms.

Retrieval: selective extraction of data from the model by attribute and (or) spatial searches or neighborhood analysis; for example, categories within a circle of given radius from a point.

Overlaying: relating two models in a Boolean and (or) arithmetic manner; for example, creating composite maps, image ratios.

Display: generating a graphic image from the model; for example, color CRT displays, symbolized line maps.

Analytical Technique Support: using analytical manipulations and computations on data model; for example, Markov chaining, network analysis, location/allocation.

Data Management: managing access and archiving of data models; for example, storage, retrieval, update, security protection, data base sub-schemas, transaction records.

- Digital Equipment Corporation's VAX computers and VMS operating system probably offer the best environment in which to implement existing software.
- Both raster and vector data models must be supported.
- No one existing software system provides all functional components for both data types.
- Automated cartography applications require a full set of functional components that use a vector data model.
- Existing public-domain vector software systems offer very little software from which to build a system.
- Raster data model functions are supported by a number of image processing software systems.
- Vast majority of image processing/GIS applications can be met by functional components that use a raster data model.
- By using a raster data model to satisfy most of the functional capabilities, an operational SDP could be created in a timely fashion and with limited resources by taking the "best" components from existing raster systems and adding necessary vector data handling functions.

RELATIONSHIP TO OTHER SPATIAL ANALYSIS RESEARCH ACTIVITIES

The development of an SDP serves a dual role in USGS's spatial analysis research activities. First, it would provide a system with some basic functionality in image processing, GIS, and automated cartography. These capabilities can be used to conduct specific applications projects. Secondly, it would provide a software base and hardware configuration conducive for future enhancements to the system.

SDP represents an important component in the effective design of a geographic information system. It would provide the general-purpose geoprocessing capabilities about which user specific applications may be integrated. The future development and enhancement of an SDP will draw from developments in software, hardware, firmware, and artificial intelligence.

Research is directed along several fronts. Under study are existing systems for graphic, geographic, and image processing to define a fundamental set of capabilities that respond to a range of user queries. Offering these basic capabilities as spatial operators, much like arithmetic and relational operators in programming languages, provides a sparse syntax for expressing desired types of manipulations. Additionally, by defining generic forms of spatial representations, spatial data stored in a particular structure may be reduced to its generic form, rather than the more extensive process of restructuring. An enhanced SDP with capabilities developed about generic representations and their combinations will result in less restructuring.

Developments in logical structures and expert systems provide a likely shell for an enhanced spatial data processor. With the hypergraph-based data structure (Bouille, 1978), for example, the concept of topological structuring, which had advanced digitial cartography in earlier years, is extended to other features of the map model. Rule-based systems may utilize these relationships, together with patterns of previous queries, to direct processing in the most effective manner.

In short, SDP development does not replace more long-term research activities. It provides a near-term capability as well as a hardware and software base for implementing enhancements resulting from ongoing studies.

CONCLUSIONS

A study to determine the feasibility of developing a spatial data processor (SDP) (a standalone, office environment, image processing and geographic information) system capable of manipulating imagery and digital cartographic data has been conducted. In the first phase of this study a set of the functional components required for spatial data processing, as well as the characteristics of the underlying data models, were determined. The functional components and seven other system evaluation criteria (functionality, operating environment, performance, long term support, modularity, software transportability, and expandability) were used to evaluate various existing spatial data processing systems. The general hardware characteristics of the system also were developed. After these studies, it has been concluded that an SDP using Digital Equipment's VAX hardware, the VMS operating system, utilizing the raster processing capabilities of the MIPS software, and adding necessary vector data handling functions offers the best opportunity, considering the resources and time available, for creating an SDP.

REFERENCES

Bouillé, François, 1978, Structuring Cartographic Data and Spatial Process with the Hypergraph-Based Data Structure, in Dutton, Geoffry, ed., First Advanced Symposium on Topological Data Structures for Geographic Information Systems: Cambridge, Mass., Harvard University

Chavez, Pat, 1984, U.S. Geological Survey Mini-Image Processing System: USGS Open-File Report 84-880, Reston, Va.

Gibson, L., and Lucas, D., 1982, Vectorization of Raster Images using Hierarchial Methods: Computer Graphics and Image Processing, v. 20, p. 82-83