RASTER-BASED APPROACHES TO MANIPULATION OF CARTOGRAPHIC DATA

DR. MARBLE: The second session today in the software area is one entitled Raster-Based Approaches to Manipulation of Cartographic Data. In the previous session we talked about data structures, and which refers to the way data is logically organized. Typically, in cartographic work we have worked with a relatively simple and traditional set of data structures. The presentations of this panel are designed to introduce us to an alternate approach to cartographic data handling.

The first presentation is by Dr. Donna Peuquet, from the State University of New York at Buffalo. Donna is the technical chief of the SUNYAB Geographic Information Systems Laboratory, and has been conducting NSF sponsored research on interactive editing of raster-mode line data. Donna?

THE NEED FOR RASTER PROCESSING OF CARTOGRAPHIC DATA

Dr. Donna J. Pequet
Geographic Information Systems Laboratory
State University of New York at Buffalo

The Development of Methodological "Traditions" in Automated Cartography:

The earliest computer programs written to manipulate cartographic data most commonly performed repetitive computations which are tedious and prone to human error when performed by hand. These programs represented direct translations of manual techniques for performing these tasks into computer-executable instructions. The data itself were also stored in a format which most faithfully reproduced man's standard analog model of geographic space, i.e., the map. This usually meant that the map was copied, point for point and line for line, into strings of digital coordinates. The necessity of recording these points by hand kept early volumes of digital spatial data very small. With the advent of efficient graphic input and output devices for the computer, an increasing number of operations with an increasing level of sophistication are being performed on larger and larger volumes of cartographic data. These operations, as well as the storage format of the data on which they are performed generally continue to replicate manual methods. This can be attributed to two factors: first, the software was usually written, or at least designed, by individuals trained in traditional (i.e., manual) cartography and not in computer science. Second, the first available devices designed for graphical i/o lent themselves most readily to replication of existing manual methods.

As reliance on the computer for cartographic data processing continued to increase, large collections of data were shared for a variety of uses. Economy of data storage space, efficiency of computer time needed for processing, and the ease and flexibility of using the data became major concerns. A direct result of this was the development of integrated comprehensive software systems to perform all phases of data management as well as a wide variety of descriptive and analytical processes for cartographic data. Thus, it often became necessary, particularly in the case of such a system, to tailor both the software and data storage methods to the characteristics of the computer hardware for the sake of cost-effectiveness. However, in practice this has meant that traditional cartographic techniques were only modified to the extent necessary and not abandoned.

The Representation of Cartographic Data:

Graphic input devices transform area, line and point structures into numeric, computer-readable form by recording spatial coordinates of mapped entities. The basic problem underlying this transformation is that cartographic data are two- or three-dimensional in nature. The coordinates must therefore be structured so as to preserve the two- or three-dimensional relationships, such as "above" or "left of" which are inherent in these data and yet be capable of being recorded in linear or list fashion so that they can be stored in the normally one-dimensional medium of computer memory.

The many formats which have been developed for storing cartographic data in digital form can be classified into two basic types; vector organization, where the basic logical unit corresponds to a line on a map such as a contour line, or raster organization where the basic logical unit of data is a strip or scan line across a data surface with data values recorded along each scan line (cf., fig. 1). The standard television image is the most common example of a raster display. Data organization on the basis of regular or irregular grids can be viewed as a special case of raster organization, since the data can still be referred and processed as rasters. The basic distinction of gridded or cell formats is that the data can be divided just as easily into either vertical or horizontal strips, and the data can be referenced in both the X and Y directions. This contrasts with a raster structure where there is not necessarily a one-to-one correspondence between locations of equal X-values which occur on different rasters (i.e., have different Y-values.)

The most common format used for computer storage and processing of cartographic data has been vector format. If a raster or grid format is used for storage, the data is almost always vectorized before cartographic processing. The reason for this is that vector format is best adapted to retaining the logical map entities familiar to humans. All rivers, roads, areas, and so on are recorded as distinct lines or groups of lines (i.e., vectors). However, it is a very difficult task to retain spatial and topological interrelationships given that all of these vectors must be stored as lists of coordinates. Often, the relationships which are of particular interest have to be explicitly recorded in some way such as separate data items, links or pointers between items. Not only does this inflate the volume of data to be stored and processed, but other relationships not explicitly recorded either have to be calculated using additional computing time, or are forever lost without redigitization.

In contrast to this, individual map features are not retained as discrete entities in raster format. It is therefore extremely difficult, if not impossible, for people to conceptualize cartographic data in terms of arbitrary slices across the data surface. However, this format easily lends itself to representing cartographic, or spatial, data in list form. Individual rasters can simply be listed in sequence. The X - Y locations are also necessarily pre-sorted and location specific within the data list. This means that all spatial and topological relationships are retained as an important part of the data format and do not need to be explicitly recorded.

In addition, individual data items can be accessed directly on the basis of location. Raster formatted cartographic data therefore has neither the time nor space efficiency handicaps that are unavoidable when processing vector data by computer.

Processing of Cartographic Data:

The internal processing of cartographic data is also most commonly performed in terms of vectors. This is again primarily due to the fact that manual cartographic manipulations are predominantly vector-oriented as a result of the relative ease of human conceptualization. This means that vector-oriented techniques or algorithms are the most commonly known and "traditional". In addition, the repertoire of vector algorithms is more developed than raster-oriented algorithms for performing cartographic manipulations. This does not mean that raster-oriented algorithms do not exist to perform all forms of cartographic manipulations. Based upon a preliminary literature and software search, it has been found that at least one raster-oriented technique presently actually does exist to perform each of a wide range of tasks (Pequet,1977a; Pequet, 1977b). However, the majority of these algorithms were developed within other disciplines, most notably image processing (cf., fig. 2).

Even with the existence of these raster techniques, the growing number of installations which utilize data captured in raster format, such as landsat imagery or drum scanner outputs generally have chosen to undergo the time and expense of reformatting these data into vector format for processing. Then, the vectorized data are often converted back to raster format for output.

These raster-to-vector and vector-to-raster conversions represent not only extra processing steps that one may want to avoid simply for general efficiency considerations, but it also turns out that many raster-oriented algorithms are much simpler and thus more efficient than their vector-oriented equivalents. For example, perimeter and area calculation, simple sums and averages and sums of points within areas are reduced to mere counting operations. Calculating the area of a polygon in raster mode requires counting the total number of pixels or their equivalent inside the polygon and then, given the size of each pixel, converting into the desired unit of measurement. Map overlaying consists of performing a logical "and", or summation, of the contents of each of the corresponding pixels or positions of the separate overlays. Windowing and clipping also become simple, again because of the presorted coordinate structure of raster-formatted data. For both tasks, all data before and after the file locations of the desired minimum and maximum spatial locations are simply disregarded. Each location in the file does not have to be tested for being within the desired limits, unlike vector mode clipping and windowing. Many other procedures are just as simple to perform in raster mode since they are non-contextual in nature. In other words, once certain global parameters have veen calculated, determination of the new value for any point is independent of the values for any other point. Rosenfeld calls these parallel procedures (Rosenfeld and Kak, 1976). By his definition, parallel precedures are those which can be performed "in parallel" or simultaneously, on each pixel or raster element. Included here are scale change and projection conversions.

The conclusion drawn from this can only be that machine efficiency, and therefore time and money, are being sacrificed so that computers can imitate traditional paper-and-pen methods.

Present Hardware and Trends for the Future:

As the volumes of automated cartographic data and the sophistication of cartographic systems increase, users are becoming more sophisticated and numerous. Developments in automated cartography have been greatly facilitated by advancements in hardware technology which have increased the speed, capacity and reliability of computers and related hardware.

I/O and data capture devices for spatial data have experienced significant technological advances in the past two decades. Remote sensing devices for spatial data have advanced from aerial photography to complex aircraft and satellite scanners. Satellite imagery, such as that produced by LANDSAT, generates vast amounts of raster-formatted digital data. For example, take the case of a simple orbital satellite with a single scanning system that uses a six bit (64 gray level) code for each pixel. This can scan the entire earth's surface, approximately 5.1 * 10**8 KM**2, say, once every 17 days. With a ground resolution on the close order of 80 meters per pixel, one global data set (17 day's world of data capture time) would require a storage capacity equivalent to 3,060 nine track, 800 BPI, 2400 foot magnetic computer tapes. As the ground resolution increases, the amount of data increases geometrically, as shown in Table 1. Contemplating the amount of computer time needed to process these data further helps in appreciating the magnitude of these data volumes. The processing time shown in Table 1 is based on a nominal 10 milliseconds per pixel of CPU time.

TABLE 1

MAGNITUDE OF SATELLITE IMAGERY DATA

GROUND RESOLUTION	NUMBER OF DATA BITS	TOTAL PROCESSING TIME
100 KM	0.003×10^8	0.5 SECONDS
10 KM	0.306×10^8	51 SECONDS
1 KM	30.6×10^8	1.4 HOURS
100 METERS	$3,060 \times 10^8$	142 HOURS
10 METERS	$30,600 \times 10^8$	591 DAYS
1 METER	$3,060,000 \times 10^8$	162 YEARS

A major drawback of most X-Y digitizers is that they require an operator. This limits the speed of recording and introduces human error. With drum and flying-spot scanners, in contrast, the data recording process is automatic once the map is mounted on the device. They can record greater volume of data than manual digitizers and are not as subject to error.

A standard device for graphic output is the digital vector plotter. Since these devices are necessarily mechanical, there has always been a trade-off between speed and accuracy The matrix plotter, however, outputs the image in raster format. This means that the complexity of the map has no effect on the plot time, in direct contrast to a vector plotter where plot time is in proportion to the total line, or vector, length drawn. Matrix plotters use an electrostatic process which is more compatible with solid state technology, thus allowing them to be designed with fewer moving parts and avoiding the trade-off between speed and accuracy while increasing reliability.

Another graphic output device which has come to the foreground in recent years is the cathode ray tube. The refresh CRT allows dynamic and interactive graphics and is useful in the display of space-time dependent data. Raster refresh CRT's are not only tailor-made for the output of raster data, but are also more suitable for drawing complex maps since, similar to the matrix plotter, drawing time is not necessarily adversely affected by drawing complexity although complex maps can cause screen flicker. Of course, performance always deteriorates if vector data are to be displayed because of the then needed scan conversion.

Thus, the devices being developed at present which show the greatest capabilities in terms of reliability, flexibility, and data handling capability operate in raster mode (IGU, 1976a). This is particularly true in the areas of spatial data capture. As a result, there has been and will continue to be a distinct trend toward the use of raster devices for graphical input and output (Teicholz, 1975). There is also an increasing need for efficiency in automated cartography systems due both to the size of the data volumes to be handled and the users' needs for fast and economical response. This means that the space in which the data are stored must be kept as small as possible in order to minimize the amount and consequent cost of the hardware required. Cartographic systems also retrieve and process data efficiently in order to keep response time down to a level which is reasonable to the user.

When technological advances are compared with performance requirements of cartographic systems, a problem becomes evident: cartographic data handling techniques have not kept pace with technological advances in computer related hardware. In many systems, the one installed at the Engineering Topographic Laboratory (ETL) at Ft. Belvoir being a good example, data are entered into the computer in raster format, converted to an internal vector format for storage and processing, and then converted back to raster format for output. These raster-to-vector and vector-to-raster processes are very expensive in terms of computer time (Kothe, 1973). In systems which utilize raster input and/or raster output, a raster-type organization for internal data storage and processing could be used to eliminate these expensive conversions. To decrease the load of vectorization put on their system ETL, a part of the Defense Mapping Agency, has found it necessary to purchase a STARAN associative array processor at a cost of $1.3 million dollars. The irony of this situation is that, while it will do vectorization faster, an associative array processor is specifically designed for matrix and raster oriented operations (Goodyear, 1974).

Conclusions:

Given that the total volume and proportions of cartographic data which are initially captured in raster format instead of vector format will continue to grow at a very rapid rate in the foreseeable future and that the trend of hardware technology toward raster-oriented graphic i/o devices is also a continuing fact-of-life, the inherent inefficiencies of vector format storing and processing cartographic data will soon render this approach economically unviable for a number of production applications. The most direct alternative is for authors and users of cartographic software to make an about-face and start conceptualizing in terms of rasters. Not only may this be beyond the capabilities of the human mind, but traditions do die slowly. Human efficiency is a factor which cannot be disregarded. The ideal solution would be to physically store cartographic data and manipulate it in raster format within the computer, while at the same time allowing the user to think in terms of vectors.

This best-of-both-worlds approach of allowing both the machine and the user to operate internally in terms which each can most efficiently use is not new. This is the basic philosophy behind operating systems, language compilers and database management systems. A substantial amount of literature already exists within the field of computer science on the problems of man-machine interfacing encountered in implementing these types of software.

The fact that almost all current cartographic software utilizes some operating system and higher-level language means that we have already been using this approach, either knowingly or otherwise. The file structure, or format in which data is physically recorded, on a disk, is organized and referenced in terms of sectors, tracks and blocks. This format is logically interpreted by the operating system as another format organized in terms of files and records which are usually ordered differently than their physical location on disk. This format or "schema" is again logically interpreted by a given application program as still another format in which data is organized in terms of map lines or pixels. The final translation is made be the user of the program who usually has yet another logical conceptualization into which he interprets the given data. Map lines may be thought of as topographic contours or census tract outlines. All of these do not require any physical reorganization of the data.

In order to utilize raster techniques in software for cartographic manipulations while allowing the user to conceptualize the data and the processes involved in terms of a vector format, a "front end" needs to be built onto the software which acts as an additional level of buffering between the computer and the user. As with the translation between other schema, no physical shuffling of data is necessarily performed. This does not mean that the user should be fooled into thinking that the computer is manipulating vector-formatted data. It merely offers the user a means of communicating with a raster system via the much more convenient vector-oriented terminology.

We do not have, however, a comprehensive body of knowledge with a complete collection of algorithms and design principles ready to be plugged in. A cartographic software system could be built immediately. However, much basic research on raster algorithm development and the comparative merits of different algorithms to perform equivalent tasks needs to be done. Relative efficiency of various raster and vector algorithms and how each is affected by varying data volumes and combinations of algorithms needs to be rigorously quantified. It therefore needs to be drawn together and interpreted in a cartographic context.

This research is necessary not just for the development of raster-oriented automated cartographic manipulation and production systems, but for the advancement of automated cartography in general. Very little basic research has been systematically carried out on development and analysis of either raster or vector algorithms. Only after this has been remedied can we knowledgeably select the best algorithm for any type of automated cartographic application.

DR. MARBLE: Thank you, Donna. The general ground that you laid will now be explored in some depth by discussions of two specific systems, one of which was to be presented by Dr. Nevin Bryant of the Jet Propulsion Laboratory. I was sorry to learn that Nevin has suffered an accident, but the discussion of the IBIS System from the Jet Propulsion Laboratory will be given jointly by David Wherry and Steve Friedman, who have managed to take Nevin's notes and graphics and put together a presentation for us. David, I believe you are going to start.

CARTOGRAPHIC APPLICATIONS OF AN IMAGE BASED INFORMATION SYSTEM

David B. Wherry
Informatics Inc.

Steven Z. Friedman
Jet Propulsion Laboratory, Pasadena

INTRODUCTION

Machine assisted cartography is often thought of synonymously with vector based computing systems. Recently, raster based processing has established itself as a competing technology in this field. Far beyond its beginnings as an offshoot of unmanned space mission imaging systems, cartographic applications based on raster technology blend the ingredients necessary to obtain rapid, flexible, and accurate processing of highly complex data.

The Image Based Information System (IBIS) is a raster based approach for mainpulating spatial data. Characteristic of IBIS operations are modeling applications integrating a variety of data types, and spatial display methods enabling rapid transformations of data into useful cartographic products. General data management considerations, as well as two IBIS case studies undertaken at the Jet Propulsion Laboratory (JPL) will be presented.

IBIS Data Management Considerations

IBIS is a fully automated raster (image) based information system. The IBIS design is based on a sequence of general purpose programs. The logical grouping of these routines into processing steps enables the handling of very complex problems. Easy operation of the routines by the system user has always been a consideration of the system's designers.

The user of IBIS can integrate raster, tabular, and graphical data types for the analysis of spatial phenomena. Image data sets can be acquired from Landsat imagery or other multispectral scanner sources. Still other image data are encoded or scanned from aerial photographic products. Graphical data, such as maps, are electronically digitized into Cartesian coordinate space and are subsequently transformed into image format. Tabular forms of data are entered into IBIS via a table-structured input. In order to establish a link between these different types of data, an interface between image-based data files and all varieties of graphical and tabular data sets has been provided.

IBIS utilizes digital image processing technology to perform most data base storage, retrieval, and analysis operations. A major advantage of this approach is that the locational aspects of data (x,y reference) are implicitly recognized by position in the raster scan. Representation of data in this manner simplifies the algorithms used in data set editing, the construction of multiple overlays, and the reformatting involved in changes of scale or adaptation of maps drawn in different cartographic projections.

Registering data images and the removal of distortions caused by different map projections are important features of IBIS. These geometric corrections are performed by the implementation of an automated "rubber sheet" alignment procedure. The operation is based on feature location, or some common reference grid such as longitude-latitude.

Special purpose algorithms have been developed for the overlay, aggregation, and cross-tabulation of data from one image with data from other images. The analysis capabilities of the system are extended by the implementation of multi-purpose algorithms designed to perform mathematical and logical operations on these data.

IBIS Case Studies

An examination of two case studies: Illinois coal reserves description, and Orlando, Florida, urban growth mapping, will aid in clarifying the flexibility and desirability of the image processing approach to the manipulation of cartographic data.

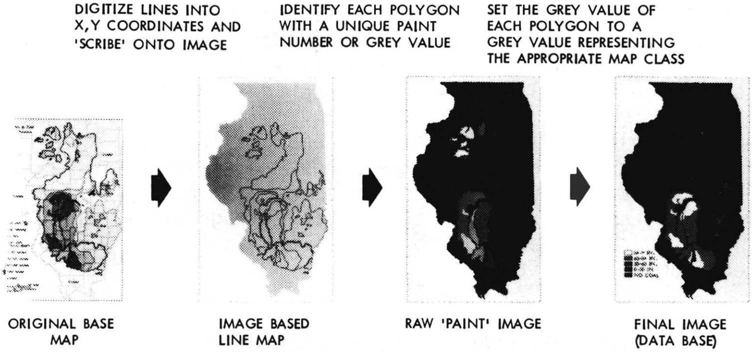
The approach selected in these studies employs the Image Based Information System developed at the Image Processing Laboratory of JPL. IBIS was conceptualized and implemented by N. Bryant and A. Zobrist (1977) as an extension of the VICAR image processing system which was originally designed to reconstruct and enhance image data obtained from spacecraft.

ILLINOIS COAL RESERVES DESCRIPTION

Approach

JPL is currently involved in an effort to define and develop advanced systems for mining deep coal seams. In support of this effort, Image Processing Laboratory (IPL) has assisted in developing an image processing scheme to describe regional coal reserves. The Herrin No. 6 coal seam, the most important seam in Illinois in terms of reserves and production, has been selected as the initial target for description and analysis.

Essentially, the image processing scheme utilized for the description of regional coal deposits is quite simple in concept (an overview of the map to image processing steps are provided in Figures 1 and 2). First, the requisite geological information is assembled in the form of maps which are typically isopleth plots of salient geological variables such as surface and seam structure, thickness, local slope, roof quality, etc. Next, the isopleths are transformed into a computer compatible format as strings of x,y coordinates via digitization. After editing and transformation into a special format required for IBIS processing, each isopleth is converted to an image-base. Each image-based map is then assigned gray values to portray a specific data aggregation chosen for analysis. For instance, a seam thickness map might be assigned brightness values to exhibit thickness classes of 0-12, 13-30, 31-48, 49-60, and more than 60 inches. In subsequent analysis steps, the maps are processed to obtain answers to questions or to test hypothesis regarding the Herrin No. 6 coal seam. At the discretion of the analyst, output is produced in the form of maps or tables.



ORIGINAL BASE MAP

IMAGE BASED LINE MAP

RAW 'PAINT' IMAGE

FINAL IMAGE (DATA BASE)

Figure 1. Processing steps in converting an input map to image format. The Herrin No. 6 seam thickness map.

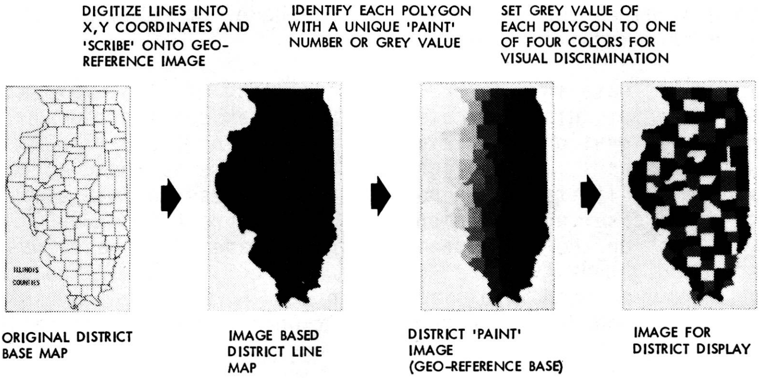


Figure 2. Processing steps in converting an input map to image format. The Illinois county map.

Procedures

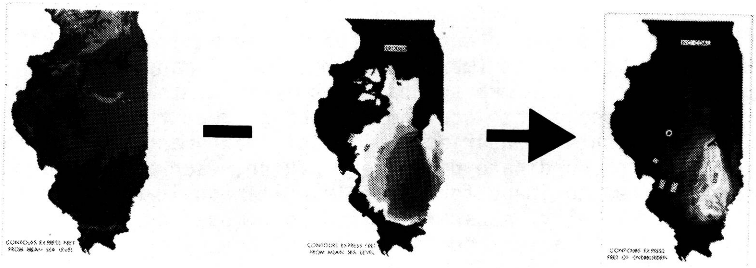
Map Digitization and Transformation to Image Format. Four processing steps are required to convert mapped information into IBIS format. First, tiepoints (markers common to all maps) are identified on a selected base map so that each subsequent map can be registered to those exact locations. Second, features of importance such as county boundaries or isopleths are manually digitized on an electronic coordinate digitizer. Third, each digitized map file is converted to image format and simultaneously registered to the base map. Finally, maps are edited to remove any errors introduced during digitization and map to image conversion.

In the Herrin No. 6 study, all image formatted maps are line-representations of geologic variables or political boundaries. These lines which are "scribed" onto an image-base are constructed of linked paths of grid cells (picture elements or "pixels") within a common grid scheme used for all maps. A grid of 1000 x 586 was selected for the Herrin No. 6 application, resulting in an elemental grid cell area of 0.154 square miles.

Editing. Editing procedures are used to correct either the image formatted map (image) or the original file containing the digitized data. Editing is simplified by "painting" regions in the image with unique values (brightness values or shades of gray) and visually inspecting that image for errors.

Painting. For ease in interpretation, painting is also employed to group data into distinct categories, such as zones of equal thickness or elevation. For example, in creating the Herrin No. 6 coal seam thickness image, brightness levels were assigned to equal the upper class limits of any particular thickness map class. In this specific case, every region or polygon representing 30-60 inch thickness on the final seam thickness image (data base) was assigned a brightness value of 60 (see Figure 2). The Illinois county map was painted in a similar manner. However, brightness values were assigned in a manner so that each county is identifiable in extent and location by a unique gray value (see Figure 1).

With one exception, all images were created via a similar processing stream as described above. The Herrin No. 6 map of overburden depth was obtained by differencing images of seam structure and surface relief (Figures 3 and 4).



TOPOGRAPHY STRUCTURE OVERBURDEN

Figure 3. Construction of the overburden map by a process of image subtraction. Data and results for the Herrin No. 6 seam.

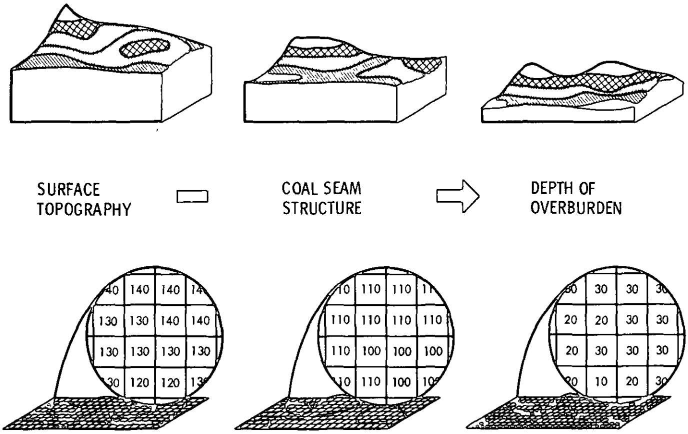


Figure 4. Construction of the depth of overburden map by the process of image subtraction.

Output Format Options. As will become apparent in both case studies, two types of output are available with IBIS -- maps and tables. Tabular output is similar in style and organization to the lists common to many kinds of computer analysis. However, map output is a unique feature of IBIS. Image formatted maps are easily converted into photographic products for display and may be stored on tape or disc as input for future analysis.

Results

Multi-Attribute Analysis. Identification of coal reserves or resources which satisfy a specified list of conditions can be a valuable tool in the evaluation of advanced mining systems. A multi-attribute analysis of maps containing information pertinent to mining systems can produce an inventory of resources containing coal deposits of a specified character. Maps and tables can be output as answers to questions regarding physical, chemical, and locational characteristics of the coal addressed by a new mining technology design. Questions regarding construction access, environmental impact, market transport, and land lease information can also be answered quickly.

A multi-attribute analysis was performed on Herrin No. 6 image maps.

The analysis is focused on identifying the location and tonnage of coal satisfying the following conditions:

Seam thickness of 30 to 60 inches;

Depth or overburden 1000 feet or more; and

Energy content of at least 12,000 Btu/pound[1].

The resource is located by interpreting the above specifications as a logical intersection of the form:

$$(30 \leq t \leq 60) \text{ and } (d \geq 1{,}000) \text{ and } (e \geq 12{,}000);$$

where $\underline{t}$ represents thickness in inches, $\underline{d}$ represents depth of overburden in feet, and $\underline{e}$ represents energy content in BTU/pound. In deriving results, first an attribute window is created for each component in the expression. This is done by masking out areas that do not satisfy the conditions. For instance, coal that does not satisfy the thickness criterion is assigned a value of zero (black) on the final seam thickness image (Figure 5). In like

SEAM THICKNESS MAP

SEAM THICKNESS ATTRIBUTES NOT DESIRED ARE BLOCKED OUT WITH 'BINARY MASK'

IMAGE WITH SEAM THICKNESS ATTRIBUTE WINDOW

Figure 5. Construction of an attribute window by applying a binary mask to the final seam thickness image.

---

[1] This list of specifications describes a portion of Illinois coal resources which may be attractive to developers near the end of this century.

manner, masks are applied to the energy content and depth of overburden images to create windows for those attributes. The three attribute windows are logically intersected to form a multi-attribute window (Figure 6). The multi-attribute window is then re-

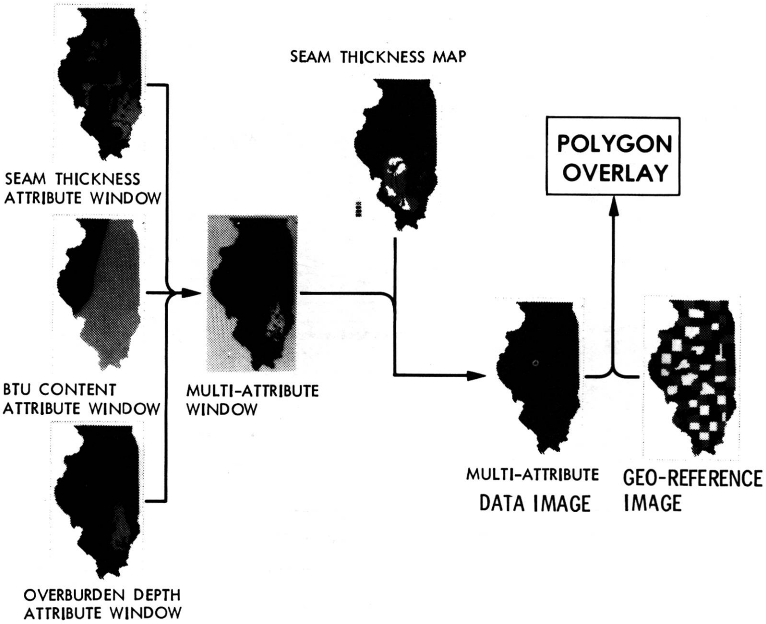
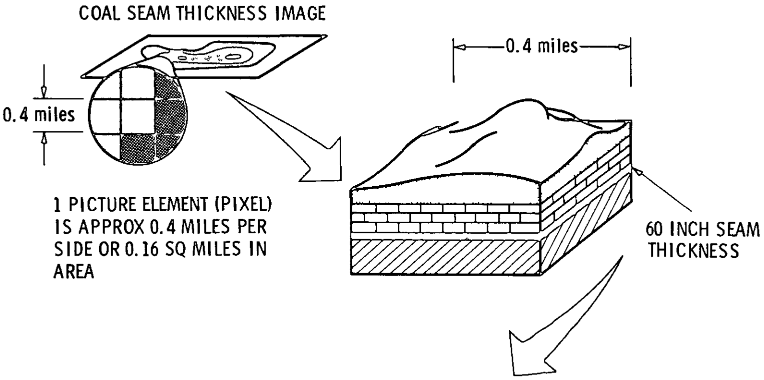


Figure 6. Formation of a multi-attribute window and integration of the Illinois county image (geo-reference base) for IBIS process of polygon overlay.

assigned gray values representing seam thickness which are lost in logical intersection processing. These thickness values are used in subsequent calculations of coal tonnage.

The process of polygon overlay facilitates the aggregation of multi-attribute pixels (image grid cells) per each Illinois county. Subsequent processing transforms these data to an expression of tons of Herrin No. 6 coal satisfying the prescribed conditions in each county (Figure 7). A final output map clearly shows the location of this multi-attribute coal (Figure 8). Associated tonnage calculations are displayed in table form (Table 1)[2]



PIXEL TONNAGE = PIXEL AREA X COAL SEAM THICKNESS X COAL DENSITY

Figure 7. The calculation of Herrin No. 6 coal tonnage.

---

[2] Accuracy of the IBIS Herrin No. 6 output tabulations was assessed by Farrell and Wherry (1978). Areal accuracies were observed to be of a high order -- inaccuracies of less than -0.18 percent difference were observed upon comparison to other published materials. Certain tonnage figures produced via IBIS processing exhibited inaccuracies due to differences in coal reserve definitions between agencies supplying input maps to the IBIS process and other verification sources.

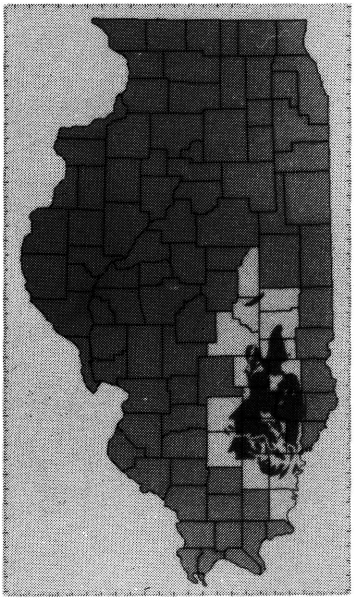


Figure 8. Final Illinois county map displaying the location of Herrin No. 6 multi-attribute coal.

STATE OF ILLINOIS HERRIN NO. 6 COAL BED
COAL RESOURCES BY COUNTY

(TONNAGE EXPRESSED IN THOUSANDS OF TONS)

Table 1. Coal tonnage satisfying restrictions on thickness, Btu content, and overburden depth.

COUNTY	TONNAGE
CLAY	1734339
COLES	283047
CUMBERLAND	783821
DOUGLAS	44226
EDWARDS	621885
EFFINGHAM	1081155
GALLATIN	2041
HAMILTON	410281
JASPER	1204988
JEFFERSON	491929
MARION	467435
MOULTRIE	124513
PIATT	9525
RICHLAND	1469664
SHELBY	217728
WAYNE	2139857
WHITE	558609
	11645043

ORLANDO, FLORIDA, URBAN GROWTH MAPPING

Approach

The Image Processing Laboratory at JPL is developing a data processing package for the Geography Division of the U. S. Bureau of the Census. The package, consisting of IBIS programs and other image processing programs contained in the VICAR system, will be utilized by the Census Bureau to update urbanized area boundary files of the Metropolitan Map Series. It is essential that these maps are kept up-to-date since Federal aid is often appropriated as a function of urban area size. Furthermore, the demarcation of census enumeration districts used in subsequent field applications is partially based on the urbanized area boundary.

Data Processing Requirements. Beginning with the 1980 decennial census, the Census Bureau must provide five year updates of urbanized area boundaries for all Standard Metropolitan Statistical Areas (SMSA) in the United States. Current field based map updating practices cannot be carried into the next decade while still maintaining efficient operating levels. Data must be obtained from new sources, and the time-lag between data collection and map revision must be greatly reduced. Remotely sensed data, such as Landsat imagery, can be processed in a digital image processing system to obtain much of the needed information for map revision. Elapsed time between data collection and map compilation can be greatly reduced with the advanced technology employed.

To complete the map revision process, Landsat data must be integrated with socio-economic data currently available to the Bureau of the Census. Revised maps as well as tabular reports used to summarize urban perimeter changes must be produced. In view of the complex data handling requirements of the task, IBIS has been employed for data management, map revision, and statistical reporting purposes. The Orlando SMSA, located in central Florida, has been selected as the first study area for testing the new package (Figure 9).

Procedures

Since the population of the Orlando SMSA has risen sharply since 1970, it was expected that a significant amount of urban area expansion would be detected. In order to determine if any significant expansion has occurred, three types of data must be integrated: (1) census tract boundary files, (2) census population statistics, and (3) thematic data from Landsat. These data types are integrated through the use of IBIS software.

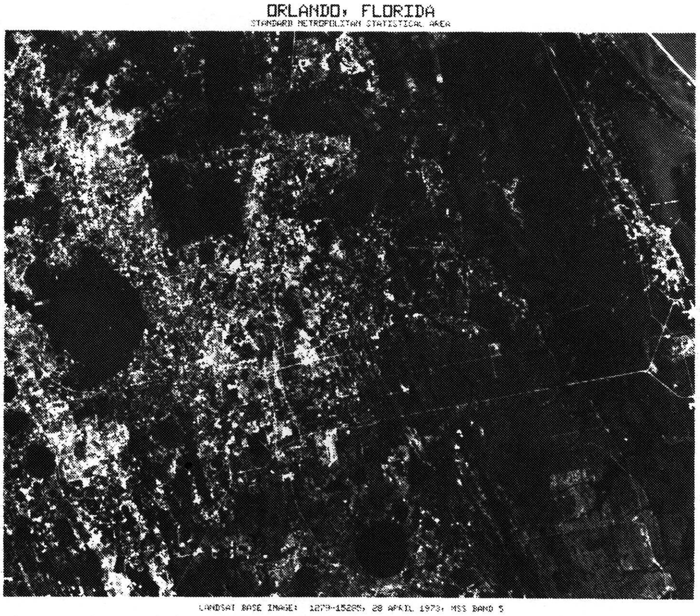
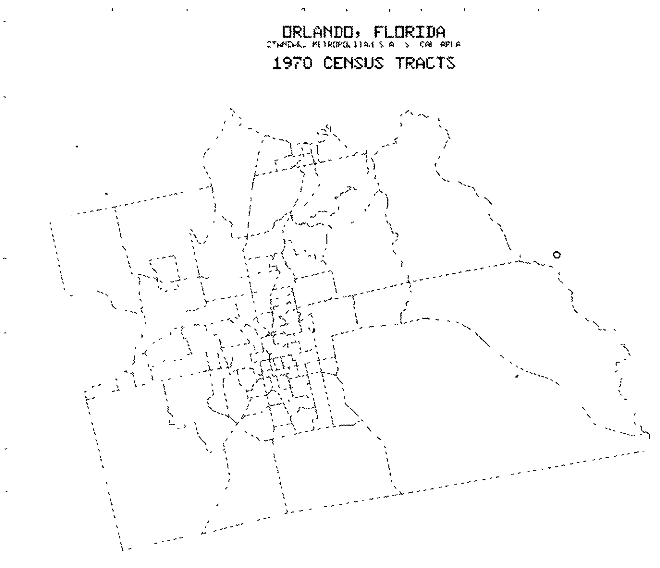


Figure 9. A Landsat image covering the Orlando SMSA.

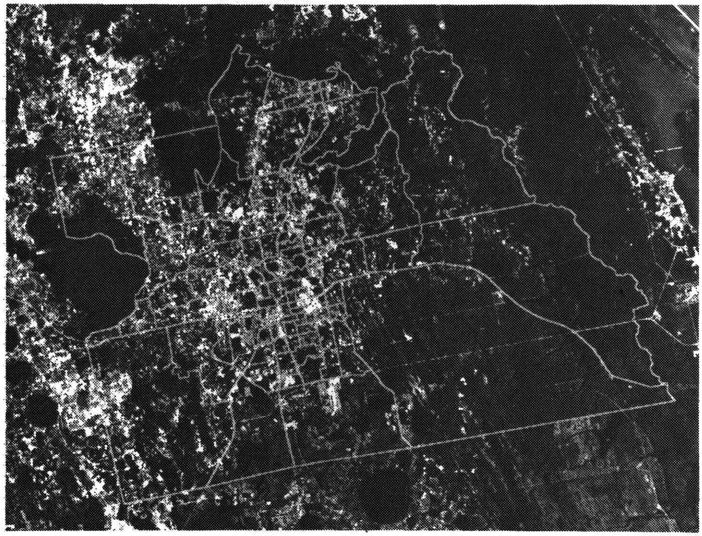
Forming the Geo-Reference Base. The initial task undertaken is the construction of a geo-reference base consisting of census tract boundary information (Figure 10). The geo-reference base is the primary data plane of the information system. Census tract boundary data, obtained in Cartesian reference form, are transformed into a raster type file to facilitate IBIS processing. Since data will also be obtained from Landsat imagery, the geo-reference base must be in registry with the satellite data. A surface fitting algorithm is utilized to obtain the desired geometric correspondence. The data sets may be superimposed for verification purposes (Figure 11). Once the census tract image has been geometrically corrected, the construction of the geo-reference base is completed by assigning a unique gray value to each region (census tract) within the image (Figure 12).

In order to establish a link between the gray tone representing a specific census tract with the geographical name of that census tract, a special file, termed an interface file, is produced. The interface file can also contain data from other image planes in registry with the geo-reference base, and non-image data as well. Additionally, data within the interface file may be utilized in mathematical functions to derive higher-order information.



CENSUS TRACT OUTLINES OBTAINED FROM THE US BUREAU OF THE CENSUS URBAN ATLAS FILE

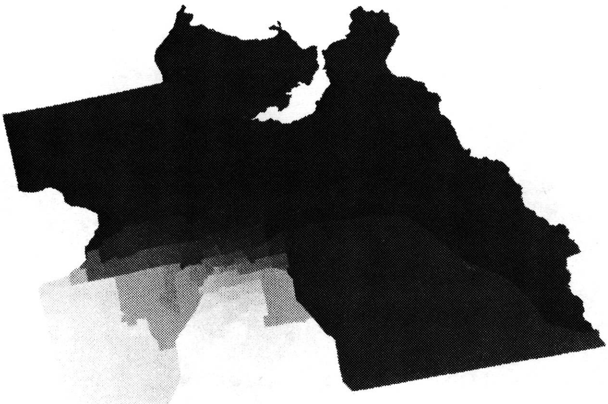
Figure 10. The geo-reference base for Orlando is based on census tract boundaries.



ORLANDO, FLORIDA
STANDARD METROPOLITAN STATISTICAL AREA
1970 CENSUS TRACTS

CENSUS TRACT OUTLINES OBTAINED FROM THE US BUREAU OF THE CENSUS URBAN ATLAS FILE
LANDSAT BASE IMAGE: 1279-15285, 28 APRIL 1973, MSS BAND 5

Figure 11. A census tract outline map has been registered to a Landsat image. The two files are merged for visual analysis.



ORLANDO, FLORIDA
STANDARD METROPOLITAN STATISTICAL AREA
1970 CENSUS TRACTS

CENSUS TRACT OUTLINES OBTAINED FROM THE
US BUREAU OF THE CENSUS URBAN ATLAS FILE

EACH CENSUS TRACT IS ENCODED WITH A UNIQUE GRAY VALUE. HOWEVER,
THE GRADATION IS TOO FINE TO DISCRIMINATE EACH INDIVIDUAL TONE.

Figure 12. The completed geo-reference base for the Orlando area is composed of several regions, each region having a unique gray tone.

Map Generation. In order to produce a map depicting urban land cover in 1970, areal measurements of each census tract and population statistics by census tract are entered into the interface file. Consequently, population density values can be derived for each census tract. The Census Bureau has determined that a census tract is urban if the population density in 1970 is at least 1,000 people per square mile. This criterion can be used to separate census tracts into two classes, urban and non-urban. A map generating routine may be implemented to obtain an urbanized area map for the Orlando SMSA based on these statistics (Figure 13).

The Bureau of the Census has requested information pertaining to urban expansion since 1970. However, no enumerative data has been gathered since the 1970 census. To obtain an updated map of urban land cover in 1975, a Landsat image has been analyzed. Through the implementation of various digital image processing procedures (Friedman and Angelici, 1978), a map of urban and non-urban land has been produced (Figure 14).

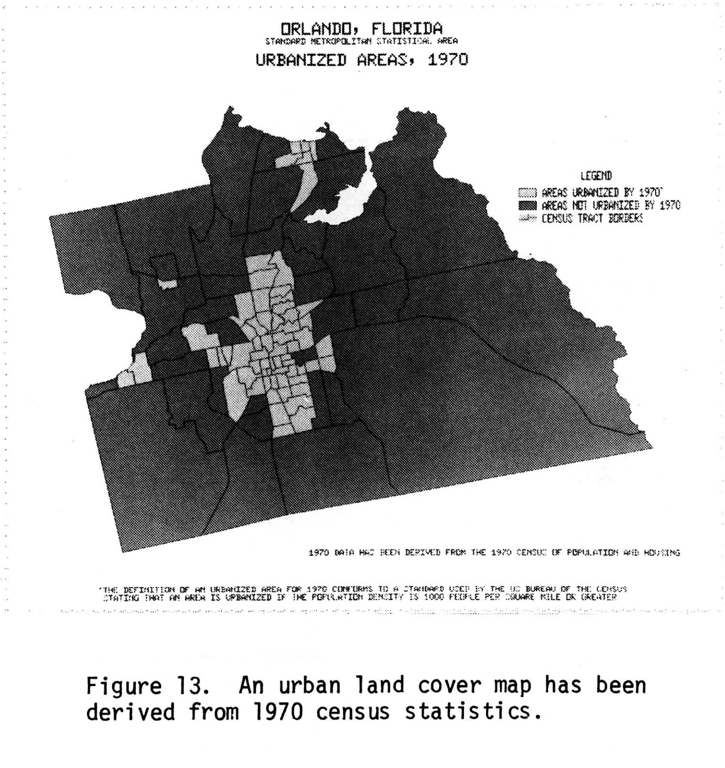


Figure 13. An urban land cover map has been derived from 1970 census statistics.

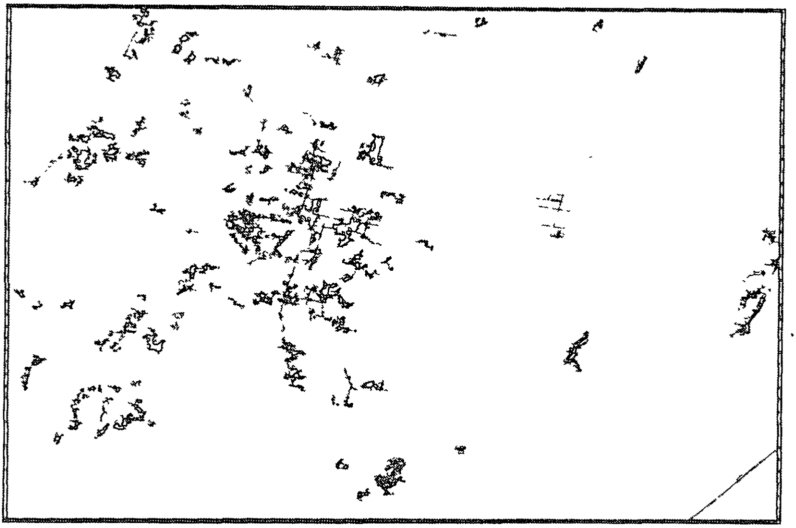


Figure 14. An urban land cover map derived from digital image processing of a Landsat image, 1975.

Map Integration. Before the areas of urban expansion between 1970 and 1975 can be delineated, the two urban area maps must be integrated. The two images are combined in an additive process, reducing the number of data planes to one. A final map product depicting urban land in 1970 and urban expansion between 1970 and 1975 is obtained (Figure 15).

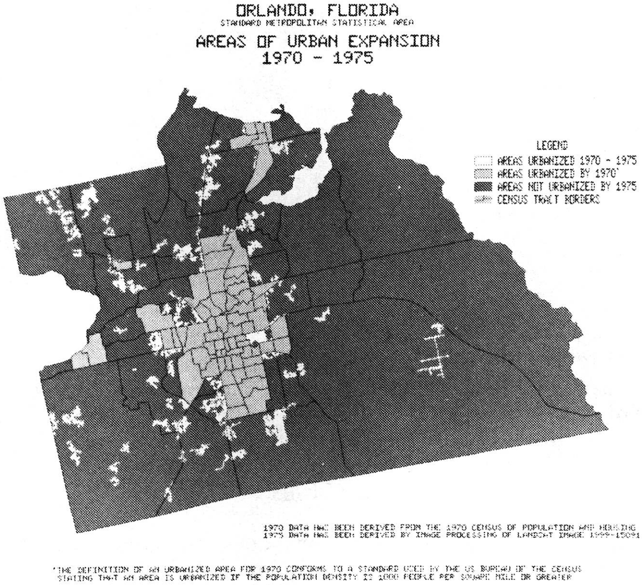
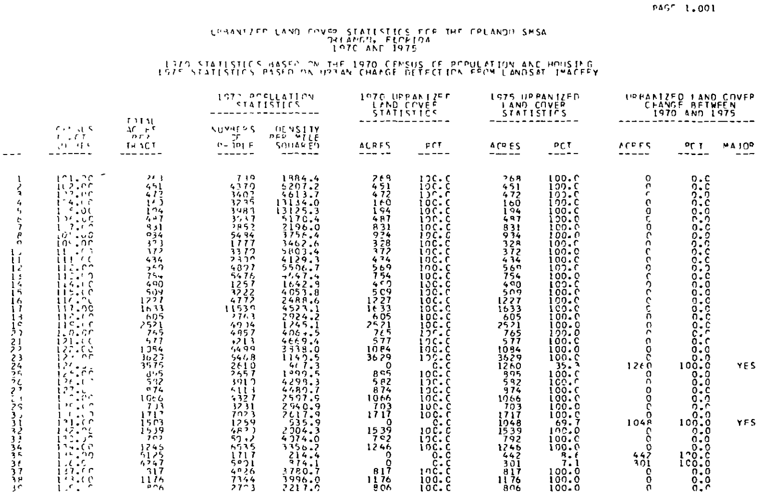


Figure 15. With the integration of remote sensing data sources and conventional data sources, a map depicting urban expansion between 1970 and 1975 is derived.

Results

The Geography Division of the U. S. Bureau of the Census has obtained a depiction of the extent of urbanization within the Orlando SMSA in 1975. Major areas of urban expansion as evident in the final map have indicated where census tract boundaries and enumeration districts may require revision prior to conducting the next census. A tabular report summarizing the land cover changes between 1970 and 1975 has also been generated (Table 2).



PAGE 1.001

URBANIZED LAND COVER STATISTICS FOR THE ORLANDO SMSA
ORLANDO, FLORIDA
1970 AND 1975

1970 STATISTICS BASED ON THE 1970 CENSUS OF POPULATION AND HOUSING
1975 STATISTICS BASED ON URBAN CHANGE DETECTION FROM LANDSAT IMAGERY

			1970 POPULATION STATISTICS		1970 URBANIZED LAND COVER STATISTICS		1975 URBANIZED LAND COVER STATISTICS		URBANIZED LAND COVER CHANGE BETWEEN 1970 AND 1975		
	CENSUS TRACT NUMBERS	TOTAL ACRES PER TRACT	NUMBERS OF PEOPLE	DENSITY PER MILE SQUARED	ACRES	PCT	ACRES	PCT	ACRES	PCT	MAJOR
1	101.00	268	789	1884.4	268	100.0	268	100.0	0	0.0	
2	102.00	451	4370	6207.2	451	100.0	451	100.0	0	0.0	
3	103.00	472	3402	4613.7	472	100.0	472	100.0	0	0.0	
4	104.00	160	3285	13134.0	160	100.0	160	100.0	0	0.0	
5	105.00	194	3983	13125.3	194	100.0	194	100.0	0	0.0	
6	106.00	487	3937	5170.4	487	100.0	487	100.0	0	0.0	
7	107.00	831	2852	2196.0	831	100.0	831	100.0	0	0.0	
8	108.00	934	5494	3756.4	934	100.0	934	100.0	0	0.0	
9	109.00	328	1777	3462.6	328	100.0	328	100.0	0	0.0	
10	110.00	372	3372	5803.4	372	100.0	372	100.0	0	0.0	
11	111.00	434	2800	4129.3	434	100.0	434	100.0	0	0.0	
12	112.00	569	4897	5506.7	569	100.0	569	100.0	0	0.0	
13	113.00	754	5476	4647.4	754	100.0	754	100.0	0	0.0	
14	114.00	490	1257	1642.9	490	100.0	490	100.0	0	0.0	
15	115.00	509	3222	4053.8	509	100.0	509	100.0	0	0.0	
16	116.00	1227	4772	2488.6	1227	100.0	1227	100.0	0	0.0	
17	117.00	1633	11539	4523.1	1633	100.0	1633	100.0	0	0.0	
18	118.00	605	2763	2924.2	605	100.0	605	100.0	0	0.0	
19	119.00	2521	4904	1245.1	2521	100.0	2521	100.0	0	0.0	
20	120.00	765	4857	4064.5	765	100.0	765	100.0	0	0.0	
21	121.00	577	4213	4669.4	577	100.0	577	100.0	0	0.0	
22	122.00	1084	5499	3338.0	1084	100.0	1084	100.0	0	0.0	
23	123.00	3629	5468	1149.5	3629	100.0	3629	100.0	0	0.0	
24	124.00	3575	2610	467.3	0	0.0	1260	35.3	1260	100.0	YES
25	125.00	895	2657	1899.5	895	100.0	895	100.0	0	0.0	
26	126.00	592	3917	4299.3	582	100.0	592	100.0	0	0.0	
27	127.00	974	6114	4480.7	874	100.0	974	100.0	0	0.0	
28	128.00	1066	4327	2597.9	1066	100.0	1066	100.0	0	0.0	
29	129.00	703	3231	2940.9	703	100.0	703	100.0	0	0.0	
30	130.00	1717	7023	2617.9	1717	100.0	1717	100.0	0	0.0	
31	131.00	1503	1259	535.9	0	0.0	1048	69.7	1048	100.0	YES
32	132.00	1539	4820	2004.3	1539	100.0	1539	100.0	0	0.0	
33	133.00	792	5042	4074.0	792	100.0	792	100.0	0	0.0	
34	134.00	1246	6535	3356.2	1246	100.0	1246	100.0	0	0.0	
35	135.00	5125	1717	214.4	0	0.0	442	8.6	442	100.0	
36	136.00	4247	5921	974.1	0	0.0	301	7.1	301	100.0	
37	137.00	817	4826	3780.7	817	100.0	817	100.0	0	0.0	
38	138.00	1176	7344	3996.0	1176	100.0	1176	100.0	0	0.0	
39	139.00	806	2793	2217.0	806	100.0	806	100.0	0	0.0	

Table 2. An excerpt from a statistical report. Urbanized land cover statistics have been derived for 1970 and 1975.

CONCLUSIONS

It is hoped that the two applications covered have facilitated an understanding of the utility of IBIS for cartographic applications. Modeling of data with either of the two data bases described in this paper is not limited to the specific problems addressed. Once an IBIS data base has been constructed, numerous questions may be posed. In the Herrin No. 6 coal seam study, the parameters of depth of overburden, seam thickness, and Btu content, may be modified to derive more extensive information about coal reserves. With the addition of other data planes, questions of a more diverse nature can be answered. Other types of information may be derived from the data base constructed. For Orlando, population density levels may be subdivided into several classes to portray a more complex model of the urban environment (Figure 15).

Several other applications have been implemented with the Image Based Information System. For the Columbia Regional Association of Governments, a pollution potential model was constructed of the Portland, Oregon area (Logan, 1978). A model has been used to predict the potential benefits of a conversion to solar power energy sources for the city of Los Angeles (Angelici, 1978). Other applications requiring an information storage and retrieval system with cartographic display capabilities are coming into view.

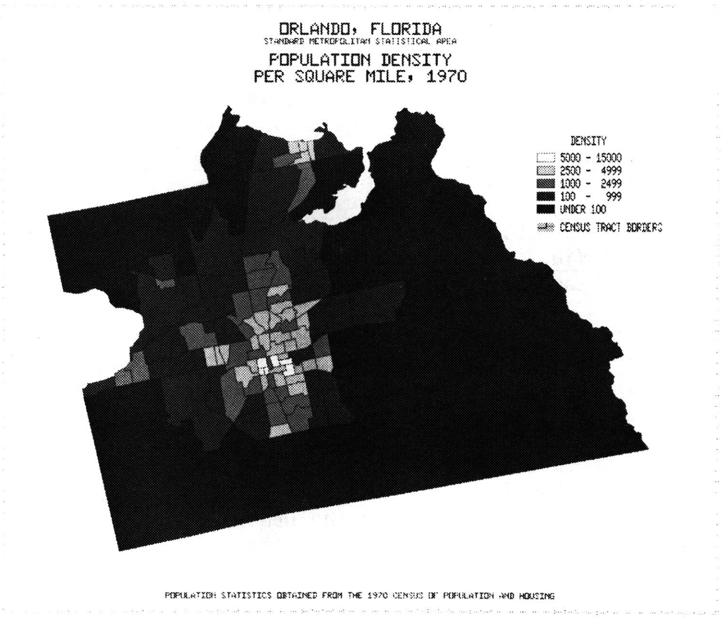


Figure 16. A choroplethic map depicting several levels of population density has been constructed from the Orlando data base.

References

Angelici, G.L. and N.A. Bryant, 1978, "Solar Potential Inventory and Modeling," Data Resources and Requirements: Federal and Local Perspectives, Sixteenth Annual Conference of the Urban and Regional Information Systems Association, 1, Washington, D.C., pp. 442-453.

Bryant, N.A. and A.L. Zobrist, 1977, "IBIS: A Geographic Information System Based on Digital Image Processing and Image Raster Data Type," IEEE Transactions on Geoscience Electronics, Vol. GE-15, No. 3, pp. 152-159.

Farrell, K.W. and D.B. Wherry, 1978, A Synoptic Description of Coal Basins Via Image Processing, unpublished manuscript, Jet Propulsion Laboratory, Pasadena, pp. B-1-B-6.

Friedman, S.Z. and G.L. Angelici, 1978, "The Detection of Urban Expansion from Landsat Imagery," presented at the spring meeting of the Association of American Geographers, 1978.

Logan, T.L., 1978, "Monitoring Airborne Health Impacts Through Spatial Crosstabulation Techniques: The IBIS Approach," Data Resources and Requirements: Federal and Local Perspectives, Sixteenth Annual Conference of the Urban and Regional Information Systems Association, 1, Washington, D.C., pp. 339-350.

DR. MARBLE: One point that I would like to make which is implicit in the previous presentation which I would like to make explicit, is that all this work was done using raster-based data structures. The same type of work could have been undertaken using vector organization of the data, but, in many cases, some of the operations shown, specifically the polygon overlay operations, would have been considerably more difficult to implement.

The third panel presentation is by Dr. Robert DeZur of ESL, Incorporated. Dr. DeZur received his Ph.D in mathematics, and is currently manager of the analysis group in the Earth Resources Laboratory of ESL. He will talk about the IDIMS system.

DR. ROBERT S. DEZUR: I had no really prepared talk. What I thought I would do is talk about the data flow through our system, and, if we have enough time, I will go through some examples of some of our recent projects. First slide. I should say first of all that our GIS type system is quite a bit like the IBIS system, as the previous speakers have discussed. It consists primarily of three basic modules, a GES or geographical entry system; an IDIMS, or interactive digital image manipulation system; and, an ERIS, or earth resources inventory system. Most of our application projects have to do with resource inventories and are reasonably complex. The reason that we do raster type processing is that the bulk of our data is in that type format.

I am speaking now of an operational and production oriented system as evidenced by the number and sizes of the various projects that we are completing or have completed. For example, we have just completed a timber inventory of Western Washington that consisted of an area of about 20 million acres. We did a water resource study for the State of Idaho of about 18 million acres; a Tansy Ragworth inventory in the State of Oregon, of about 15 million acres; and we have done various demonstration projects varying from three to five million acres. So we are talking about a lot of data going through our systems in a rather short time.

The GES system is set up to digitize polygonal boundaries, control points, flight lines or sample points and the like, and provides the framework for generating the transformations by which to register the map base to LANDSAT imagery or whatever the data base might be. IDIMS itself is a collection of image processing routines that are used in digital data processing. ERIS, the earth resources inventory system, basically provides the framework for handling the tabular data, generating and manipulating such files, and providing for the statistical analysis and regression routines that are used in our various studies.

One word about the tabular data, for example a project for the State of Oregon. In the course of this study we had something like 3,000 sample photo points that were photo-interpreted, where at each point information was provided for about 75 different variables. In addition to this PI data, we had some 300 sample ground plots where 75 to 100 variables were measured or reported on for each of these particular data points. Each time an operation is performed in the ERIS system a file is generated. So there are hundreds and hundreds of files being generated and deleted during the course of an inventory analysis.

May we have the next slide, please. I took this slide from one of Dr. Bryant's papers that he wrote in September of last year in which he compared the IBIS system to the ESL system. One unfortunate thing about this particular slide is that it gives one the idea that we are discussing sequential processing when indeed parallel processing is being carried out. A lot of the tasks noted here are going on simultaneously. The Vu-graph is to be read top to bottom, left to right, and continues across the page that way.

Several of the routines mentioned here have been updated, so it is not entirely correct as presented. But the left-hand column essentially has to do with the image processing phase of our analysis, and handles LANDSAT or digital terrain data, or any type of multispectral data. The middle column has to do with our GES system in which we define the study areas. In our particular GES system we begin by defining a geoblock. May I have the next one, please. So our basic GES structure data base looks something like that. We first of all identify geoblocks on either the map base or on LANDSAT imagery. Then, associated with each geoblock, we have a number of overlays to control the data. So, following down through the chart, for example, Geoblock 2, Overlay 2, we have a number of areas, a number of junction points and a number of line segments. These are reassembled in our system then as shown in the lower right-hand corner. Over on the right-hand column, basically we do our data summaries and enter the tabular data into the ERIS portion of our system and process it to generate either graphic products or a formal written report.

I will go through each portion of the data flow more carefully. Next slide, please. Again, this is an overview of the GES system, taking the map data, generating the strata, utilizing the classified results from LANDSAT processing, and providing strata summaries using the various routines that are available in the IDIMS system, and outputting tabular data to the ERIS system. Next slide. This is going to overflow the board. But starting up in the top center with the design of the geoblock, this generally pictures a flow of the data showing some of the parallel processing going on within our system.

First of all, design or choose geoblocks, digitize administrative strata, flight lines, control points and so forth, coming down the center portion, and then start generating various transformations that are used to register the raster type data with other data types. While this is happening over on the left-hand column, we are beginning the pre-processing of LANDSAT data, if that is our data source, and going through more or less standard routines to turn out a classified image. At the same time control points are being selected from the imagery, generating more transformations and storing all of these so that they are available for the strata summaries that will be utilized by the ERIS subsystem. Going now to the extreme right, we are able in our GES or IDIMS routines to summarize data, taking combinations of up to seven strata overlays. So, thinking of these are seven images, looking at the intersections of the various polygons that result from these, we can summarize our data in this particular manner, feed the results into our ERIS system for use in the statistical analysis portion of an inventory study, if that is what we are doing. Also, we generate graphics files, register and geometrically correct our imagery, so that certain customers can have their pretty pictures, I guess.

Next. This is a slide showing some of the transformations occurring during the GES phase of our processing, showing how we register uncorrected LANDSAT data, change to a 50 meter grid, and generate transformations between the LANDSAT and the ground coordinate system to locate various sampling units and that sort of thing. This is a schematic showing how a very simple stratsum algorithm might work. On the upper left-hand corner we have three strata that have been digitized. I am thinking of these as images now. In the upper right-hand corner we schematically represent these three strata, and just counting across each line we note the number of pixels that occur in that particular line segment, and creating a file for this purpose we record that information. Then we manipulate such files in an appropriate manner.

The ERIS portion of our system, is essentially a collection of software routines, again for data and statistical manipulation. We have the general routines noted -- under file manipulation -- filter, append, merge, sort, and sample pairing. Sample paring is, of course, important, because we usually utilize three kinds of data, and we pair them and analyze the paired data quite frequently. Under arithmetic and statistical routines--sample amalgamation, predict, and minitab. Out minitab statistical routine is a standard version that comes from Penn State University. Next, please. This is a description of the minitab capabilities. I will not go through all of them. One thing I can say about the ERIS portion of the analysis is that it is the most time consuming part of the entire study. Image processing probably occupies about ten percent of an effort. The geographic entry system, digitization and so forth, probably about 20 percent. The remainder is taken up by our ERIS processing--a very large volume of data is being processed through the system.

Next. This is an example of some tabular data collection sheets that photo-interpreters use--these generate a large number of variables for each of the data points in question. Next, Again, just a sample of the kind of information that might be collected by photo-interpreters while interpreting certain kinds of imagery. Next, please. Here are several examples from an ongoing project of the type of data flow through the ERIS system, showing how we use various routines, and the data manipulation capabilities of the system. I will not bother going through all of the general routines or the descriptions. But you can see that it is fairly complicated, and, as I say, it presents a horrendous bookkeeping problem keeping track of all the files, what was done where, which can be deleted, what must be kept for posterity, so to speak, and so forth. Here is another example with large scale photographic data. The next one has to do with low altitude photography.

Next, please. Here is an example of some output products, perhaps in a forest inventory project. As a result of our stratsum runs, we would provide the total number of acres by ownership class (in this particular project), and by forest type. In general, there would be a large set of tables of this sort. Or, something more complicated, as in the next. We perhaps would fill in a table like this for a forest inventory. These are basically for in-place mapping type studies. In general, however, as a result of all the ERIS processing, we generate an estimate of a particular attribute for a forest inventory. For example, it may be basal area or volume by ownership class or perhaps just a confinement to a given area of a particular state. We would give standard error estimates, confidence limits and so forth. The inventories are pre-designed for an allowable error and an acceptable confidence limit for the estimates.

Next. To demonstrate the previous material, I chose an Oregon project, Douglas County, an area of about 3.2 million acres. The project required a timber volume inventory, and an in-place thematic mapping study. Both studies were supported by the multistage sampling technology, and we used supporting photographic data for the raster type LANDSAT data. This shows Douglas County approximately in the center of the screen. We can go through these rather rapidly. Next slide, please. Let me summarize what the project was. We were to estimate the net total volume for

Douglas County and provide an analysis of treatment opportunity groups for managerial decisions.

Next, please. These were some flight lines that were going to be flown to help in the second stage sampling procedure. Next. This slide depicts the fact that we are using a multistage sampling approach, which will use ground data, aircraft data and auxiliary LANDSAT information. Next. Again, just a sample of a flight line being flown with a different type of projection. Next. Some low altitude photography. I think it was at 4,000, or something like that. Next. Just a photo of a ground plot, which turned out usually to be an acre in size. Next. That appears to be backwards, but it is a first step output from our classifier, showing a masked LANDSAT classification of a part of Douglas County.

Next. This is an ownership map of a part of Douglas County. I have forgotten what the colors represent, but there are three ownerships for the acreage represented. Next. We have here the same region, showing eight classes that are color coded forest types. Notice that ownership boundaries and township boundaries are also shown. Next. This is a blown-up version of the prior slide showing 18 treatment opportunity groups along with the digitized township boundaries and the ownership areas. Next. That is a similar subsection. Next. This is a similar area in Douglas County again, showing these opportunity groups. I think that is all. (applause.)

DR. MARBLE: A number of years ago I was on the faculty at a major Midwestern university which had a computerized student record system. This had been developed internally by the university, and after some use they decided to improve it. An outside system analyst was brought in and after several days of looking at the system, he said, "This is absolutely remarkable. It is one of the most beautiful examples I have ever seen and should be documented in the professional literature." We said, "Why is it so remarkable?" He replied, "It is the most beautiful example I have ever seen of the one-to-one implementation of quill pin techniques on an IBM 360."

Yesterday a number of people remarked that we should attempt to automate more conventional cartographic operations I would like to take the liberty of suggesting that perhaps we may not want to do just that. One of the points in the discussion in the panel so far is that there are alternate ways of looking at things. As our tools change, so, in many cases, must the way we look at the world that we deal with. This question of the organization of data, whether it is handled in vector (line format) or in raster format, which is a somewhat more difficult one for us to envision, may prove to be quite important to us. Attempts to follow traditional manual techniques and implement them on the computer, including traditional ways of thinking about cartographic data, may cause us to fall into some very serious inefficiencies. The suggestion being made here is not that raster data structures and raster processing is going to save the world. Rather, it is an alternate way of doing things which in some circumstances may be considerably more efficient than traditional approaches.

We must think about new ways of doing things. Yesterday in Dr. Tobler's panel we were exposed to a number of developments in the display area of image processing. The two example systems that were discussed here both evolved out of LANDSAT processing operations where a tremendous volume of data was originally captured in raster or scanned format. The incorporation of digital line data into these systems in raster form was, of course, more efficient than attempting to convert the large volume of raster data back into line format for vector processing. We have a number of devices, some of which are on display in the exhibit area, which are raster-oriented devices. Last night I saw a very interesting little movie put on by one of the exhibitors about a raster device, a scanner, and a raster color processing system. There are other types of devices that handle data in this fashion, and I think that we are going to be confronted with the capture of more and more data in raster format, and we are either going to have to decide to process in this mode or to start working very intensively on very efficient algorithms for the conversion of raster to vector data.

At this point I would like again to open the meeting to general discussion. I know there are other people that have been working on raster systems and are interested in these things. Perhaps one or more of them might like to make a comment.

MR. JON LEVERENZ: I waited, I think, a long enough time to let the people that had something to say about rasters get up. This does not deal exactly with raster but it appeared to me when I heard the discussion on the Herrin No. 6 coal seam that it seemed to be a rather simple process of putting together the three attributes and narrowing it down to where they all occurred together. It probably is not as simple as it looks, but I wondered what the reason was that we, first of all, or why the State of Illinois -- and I have somewhat of a vested interest there, because I pay taxes -- what reason the State of Illinois had for developing an automated system to do something that it appears could be done manually.

Perhaps I am back a few years, but I would like to know what the advantage is, I would like to have them point it out a little more clearly to me the advantages of the raster scan or computerization of the data and manipulation of it this way rather than just a semiautomated type of thing where the data would be input by manual methods and so on.

DR. MARBLE: I think you actually have two questions there. One is the question of why go to an automated analysis scheme for this type of operation. The second is, why do it using raster-mode processing? Do our panelists from JPL want to comment on that?

MR. DAVID WHERRY: Let me describe the processes which were involved in producing this case study. I was going to include in the talk initially the conceptualization of the project and discuss how it does appear to be an automation of a very old process, basically, of that of looking through a series of overlays to find an area exhibiting several attributes. I believe that what we have done in automating this process, although it has taken a great deal of money and a great deal of time to initially encode the maps into digital form, is to provide more flexibility in data analysis. The power of the system is that once the maps are encoded, one might ask any one of a number of questions about BTU content, seam thickness, etc., for whatever kind of maps were encoded -- distance from railroads, topography, land ownership, whatever, and obtain multiattribute results very quickly. The results that you saw presented of the multiattribute analysis after the maps were encoded were pro duced in about five seconds of computer time. Other questions can be asked after the maps are encoded, one after another, with equally rapid results. That was the theory behind the automation of the system.

Traditionally, I believe, that looking through a series of overlays will get you some results, it will define the window, but then the window has to be perimetered, the window then has to be manually manipulated as far as achieving tonnage or whatever. As far as the raster approach, we only work in rasters, number one. Secondly, to use a vector approach, an approach of, let us say, vertices to describe polygons, to look through a series of overlaying polygons to look at an intersection of data sets, is extremely time consuming and takes a great deal more computer time to operate. Also, the algorithms are far more complicated. In the raster approach it is a trivial matter of merely blocking out those sections of the image which are excluded from the window, one at a time, and what you have left is the multiattribute window. It is just a matter of masking the image. I am not sure if that answers the question. Do you need any more clarification?

DR. MARBLE: One point about manual operations on maps. They have been with us for a long time. One of the things that we have learned about them is that they are very costly. If you have a fairly simple question that is going to be asked once, you can extract the information from the map. If you have a lot of questions that are going to be asked in a variety of different forms, then you very rapidly run into rather substantial costs for handling the operation in manual form. Dr. Roger Tomlinson, the Chairman of the IGU Commission, has developed some useful data on this, and it is amazing to me how early the crossover point is and how the low cost effectiveness of manual versus automated methods.

It it were just the case of the single question presented here as an example of the operation of the system, I suspect that we might well worry about the comparative costs of automated versus manual methods. But, as the panelists pointed out, the ability exists now to answer a large number of questions, and this underlies the development of a number of digital, spatial, natural resource data bases, for example, the one that is being developed by the Geology Division of the U.S. Geological Survey dealing with coal resources, where a number of questions are to be asked about the coal resources of the United States and where not only such things as the coal content and overburden computations are available but many, many other things as well. The operation that was illustrated here, and quite well, was that of polygon overlay. And the polygon overlay operation is one which is difficult and complex to carry out in a vector mode data organization. It is a relatively trivial one in raster format.

MR. DAVE DAY: My name is Dave Day. I am with Canadian National Parks. I am not all familiar with the raster system or raster processing, but while you were giving your demonstration of IBIS and the three case studies, you put on the screen quite a few photographic products with different levels of density for your classifications and so on. Are those necessary for processing, or are they just done for our benefit? Is most of the work internally done, or do you have to actually go through those steps and redigitize as you go along?

MR. FRIEDMAN: In order to distinguish each of the polygons, it is implicit that they have to have a unique gray tone value. We can have up to 32,000 polygons in this way. Since you cannot see the gradations, we have enhanced them using contrast stretching so they are digitally discriminable on the screen. Yes, they do have to have a discreet value.

DR. MARBLE: I think the question was whether the photographic products were necessary in every stage of the process, and the answer is that they are not.

MR. JOHN RIDDLESEE: I have a technical question. My name is John Riddlesee, Petty-Ray Geophysical, Houston, Texas. In terms of the storage of raster data, particularly when it is very voluminous -- I am thinking in terms of a thousand dot-per-inch raster on 40-inch square sheets. Are there any techniques developed that relate to the storage of such data and sparse matrices? In other words, where you have particularly the thematic representations in raster form, are there any techniques where, say, only one percent of the cover or a few percent of the area is actually covered by the information you want, while not having to store the regions with blank information. Is there any sparse matrix type storage for raster data techniques that have been developed? That is for anybody on the panel.

DR. PEUQUET: Yes, there have been quite a few techniques developed to compact raster formated data, just as there have been for vector data. One of the most common ones is run length encoding. You simply record the significant data points and the distance between them, instead of recording all the blank space explicitly.

There has been some mathematical research on sparse matrices that I have read which was put out years ago as IBM technical notes, but they have not been applied until very recently when large memories permitted handling really large matrices. But, yes, there are quite a few things you can do.

DR. MARBLE: I might point out that there is some similarity here between the technique used for data compression in the raster area and the techniques that have been successfully used in developing the large scale mathematical optimization programs using sparse matrix storage techniques. Are there any other questions or comments?

MR. RAY DILLAHUNTY: My name is Ray Dillahunty, and I am with Petty-Ray Geophysical in Houston. I have run across the concept of collection of raster data using stereo digitizers or stereo plotters and the consequential compaction of data using that kind of method. Is there some work being done in this area with the people in the cartography industry, or very much work, should I say?

DR. MARBLE: Could you perhaps clarify what you are talking about. Are you talking about digital terrain models?

MR. DILLAHUNTY: In the past I think most contours have been digitized using stereo photographs in a vector form where you digitize a contour of a hundred foot in a polygon type of thing, closing the entire contour. I have run across a little bit of literature where you set up a rather automatic digitizing concept so that your stereo plotter is going in an "X" direction across the stereo photography, and you are recording only the differences in elevation, and the distance between your digitized points in this manner. If you are digitizing flat surfaces, you greatly reduce the number of XY coordinates that you are using. I think that this is sort of the raster concept that she was addressing in her talk. Is that enough information?

DR. MARBLE: Yes, I think it is clear. A number of the automatic devices for deriving terrain information from stereo photographs do work in that manner. For example, the Gestalt photomapper system that in installed at the Geological Survey. There is a fairly complete discussion of this hardware in a recent issue of Photogrammetric Engineering, which reviews a number of the different systems and talks about this approach to it. I also understand from talking to people in the Topographic Division that the Survey plans on releasing a certain amount of terrain data in this form within the next few months. They will be producing DTM's, using the Gestalt photomapper, and this will be made available through NCIC. I am not sure about the timing of this, but I suspect it will not be very long.

As far as data compression in this format is concerned, I do not know. I am not sure I am too aware of that. Is Bob McEwen or someone else from Topo here that could talk about that? Well, I guess not. You might talk to them directly if you would like a more specific answer to that question.

MR. HARRY HEARD: Harry Heard, Institute for Advanced Computation. I have two cost questions I would like to have the panel address. One is, can they define the cost allocation in terms of the percent of total cost related to data capture. The second is, can you give us some idea relative to raster data systems of the cost on a unit basis to process information -- for example, something like the cents per pixel per attribute?

DR. MARBLE: I would be delighted to have the information myself.

MR. FRIEDMAN: So would I.

MR. WHERRY: I am really not prepared to reply to that. I basically do not know anything about the cost effectiveness of the raster-based image processing, so I think it is going to have to be referred to Dr. Bryant at some future date or perhaps somebody else can address the question.

DR. DEZUR: I can say in our studies on a per acre basis for inventories we are about two to four cents per acre, in that range.

DR. MARBLE: There is a very real difficulty in deriving cost figures on systems of any type. This has been something that the IGU Commission has been very concerned with, and we have been very frustrated over it in the past. Many of these systems are developed in a fashion that sort of parallels the way things used to work when I was dealing with NASA some ten, twelve years ago as an earth resources investigator, we were being asked what was the costs of the system? We said, "Well, what is the cost to put up the satellite?" They kept telling us that that was not important!

You have major uncontrolled cost elements in these systems. In many of the areas we really need much more precise cost information, not just about how many CPU seconds were consumed, we need the actual cost functions for it. Digitizing, both table digitizing and mass digitizing is an area in which this information does not exist today. This is an example; some of the mass digitizers like the drum scanners are raster oriented devices. But we are hard-pressed to produce actual cost figures in this area. Even in cases where we have examined existing systems, where they are running these devices, many governmental agencies and scientific institutions that do this work are not set up within a proper cost accounting framework to produce the information. This information is something else we need very badly in order to make intelligent decisions in this area.

MR. BILL JOHNSON: Bill Johnson, Lawrence Berkeley Laboratory. I would like to address the gentleman from JPL and ask what is the availability of this software involved in IBIS, and, in particular, the software involved in registering polygonal files with your LANDSAT raster data? And under what circumstances was the IBIS system developed?

MR. FRIEDMAN: First of all, IBIS is a subset of VICAR, which is our image processing system. That is currently available via COSMIC, which is a government clearing house for data. IBIS as a subset of VICAR has currently been under development for the last couple of years, and the software has been funded via various NASA grants, therefore, it is also in the public domain and can be made available at cost.

DR. MARBLE: The VICAR system is indeed in the public domain and is available and being used in a number of places. I will note that it is somewhat IBM dependent, which you might read "very" IBM dependent.