Eighth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Crop Inventory and Monitoring

July 7-9, 1982

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

INFORMATION PROCESSING OF EARTH RESOURCES DATA

A.L. ZOBRIST, N.A. BRYANT

Jet Propulsion Laboratory
Pasadena, California

ABSTRACT

Current trends in the use of remotely sensed data include integration of multiple data sources of various formats and use of complex models. These trends have placed a strain on information processing systems because an enormous number of capabilities are needed to perform a single application. A solution to this problem is to create a general set of capabilities which can perform a wide variety of applications. General capabilities for the Image-Based Information System (IBIS) are outlined in this report. They are then cross-referenced for a set of applications performed at JPL.

I. INTRODUCTION

Extremely difficult problems have been squarely faced and successfully overcome to produce the level of technology in remote sensing that we enjoy today. The Landsat MSS is a marvelous instrument, and the numerous image processing techniques which are used to derive data sets for applications are in an excellent state of development. The use of remotely sensed data sets by discipline scientists has reached a high degree of sophistication with integration of multiple data sources and complex modelling as key trends. One bottleneck that may be found in this stream of data capture and use is in our general capabilities for information processing of earth resources data. This would include both remotely sensed data and other data necessary for modelling in the discipline areas. Poorly developed methods of information processing do allow small tasks to be performed because the analyst can adjust to problems that arise. Large tasks require careful approaches to spatial data integration and geographic analysis if analysts are to be freed from a vast amount of exhausting and stupifying work.

At JPL, a complete system for the integration of Landsat data with geographic data has been developed[1]. It is actually a bit more general purpose, and has been used with remotely sensed data other than Landsat. The system at JPL (Image Based Information System, IBIS) operates on the three basic datatypes used in geographic analysis: Image, graphics, and tabular. This report will attempt to identify some basic elements of the JPL system and cross reference the use of those elements on some actual applications performed at JPL. This should serve to identify the types of information processing needed to further the utilization of remotely sensed data.

II. SYSTEM ELEMENTS

The purpose here is to give a systematic arrangement of the basic capabilities needed for a hybrid raster-vector-tabular geographic information system. The basic capabilities are grouped under the subheadings that follow here.

A. GRAPHICS TO RASTER

The most important case here is the conversion of a region boundary file to a raster file containing a unique identifier for each region in the cells corresponding to each region. An associated tabular file is created to link image identifier with region names from the graphics file. In some cases it is desired to convert only lines or boundaries to raster form. A set of routines to accomplish these tasks consist of the following:

A. vector linear transformation routine
B. vector to raster scribe
C. Raster connectivity routine
D. point overlay routine
E. map color assignment routine

Routine A operates in vector mode only to

This article presents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract No. NAS 7-100, sponsored by the National Aeronautics and Space Administration.

transform a geographic coordinate in a linear fashion to agree with an image coordinate system. Routine B writes a raster file, setting raster cells to a selected value corresponding to the geographic locations of lines in a vector file. Routine C finds connected regions in the file produced by Routine B, marking each with a unique identifier. Since the border lines constitute a significant amount of area, they must be assigned in a systematic or random manner to neighboring regions. Routine D uses a tabular file containing centroids of regions and their corresponding names to perform lookup of raster identifiers to create a link between names and raster identifiers in the tabular file. Routine E is used to produce a visual product from the raster region file. It assigns four colors to a set of regions so that no adjacent regions have the same color. This routine operates with a heuristic instead of an algorithm, but is backed up by a guaranteed five color algorithm.

B. SURFACE APPROXIMATION

Problems here can be categorized according to type and according to mathematical requirements that the surface must meet. The types of problems are handled by the following routines:

F. point to raster surface
G. vector contour to raster surface
H. aggregated region tabular data to raster surface
I. generation of raster surface by mathematical function

Here, the term raster surface refers to an image in which the cell value represents a geographic Z-value. Routine F is commonly used where point measurements of a physical variable are used to extend to a surface. A variety of techniques are known which can give a mathematical surface approximation to the points, and the raster is calculated from the mathematical surface. Routine G is a difficult and highly specialized routine. Its most likely use is to obtain digital elevation from elevation contour data because so much of these data exist. However, the new generation of photogrammetry machines are producing raster elevation directly. Routine H might be used to create a population density surface from a map of regions and a table of population by regions. See Tabler and Lau[2] for methods of obtaining a continuous surface while maintaining correct aggregates by region. Routine I is useful in a variety of modelling situations.

C. RASTER PROCESSING

The system considered here will have to handle images of varying sizes and should handle up to 6000 x 6000. The most useful image processing routines are in the following categories:

J. file handling
K. thematic classification
L. mathematical function
M. image correlation
N. image brightness ruber sheet
O. image stretch

Routines in J include image copying, cutting, and pasting. Routines in K convert raw Landsat into a raster representation of ground cover or land use. See Bauer, et. al.[3], for a view of the status of these techniques. Routine L allows a user specified mathematical function to be applied to one or more images, resulting in an output image. An expression compiler is needed for the user specified function and a hash table is used for storage and lookup of computer results to save computation on large images. Routine M uses two-dimensional FFT phase correlation to find matching points in two images over a common area. Routine N corrects an image according to a brightness correction surface specified on a grid of points. Routine O changes the brightness values on a per pixel basis using a table specifying new values for each old value.

D. SPATIAL REGISTRATION

This is necessary when two maps or data planes are not in positional agreement when overlaid. In simple cases where the maps are accurate, agreement can be obtained by changing scale, aspect, or rotation (affine transformations). More commonly, the maps disagree by a complex distortion surface. The distortion is usually measured by obtaining the location in each data plane of a set of control points. The following routines can register data sets using control points:

P. random to grid control point transformation
Q. raster rubber sheet
R. vector rubber sheet

Routine P is needed because randomly spaced control points do not allow efficient image processing in routine Q. One of the surface fitting techniques used for routine F is used here to fit the random points and the fit is applied to a rectilinear grid of control points. Routine Q distorts an input image according to a grid of control points to obtain an output image. The distortion is precisely followed at the control points and interpolated in between. An interpolation is also carried out to obtain the cell value in the output since it usually corresponds to a between-cell location in the input. Routine Q is very complex because one line of the output is often calculated from more lines of the input than a computer can hold in main memory. Routine R applies the same operation to a vector data set but calculations are much simpler here.

E. RASTER TO TABULAR

Two operations are important here. Measurement of overlapping areas of two data planes and spatial integration of a surface data plane by regions in another data plane.

S. polygon overlay
T. spatial integration

Routine S reads two inputs and creates three columns in a table. The cell values are placed in two columns and a cell count is placed in the third (a simple joint histogram). Hash table storage is usually useful for this routine. Routine T is similar except that the cell values in one input are summed for the regions represented in the other input. A combined version of these routines that integrates a surface over the intersecions of two region rasters is also useful. These routines can be used for point and line representations as well as areas.

F. TABULAR TO RASTER

Routine H falls in this category. A simpler version is:

U. choropleth mapper

This routine reads a table which contains region identifiers and some derived statistic that is to be mapped into colors or textures in that region. The raster that contains the region identifiers is converted into a choropleth map by table lookup.

G. TABULAR PROCESSING

A basic set of tabular operations gives the system enough power to support a variety of applications.

V. sort
W. report
X. aggregate
Y. delete on key
Z. merge correlate
AA. mathematical function
BB. concatenate
CC. map projection
DD. disaggregation by proportions

Routine V operates on numeric or alphabetic keys occupying one or more columns. Routine W allows column selection, specification of headings, column headings and numeric format, and subtotaling of selected columns using a control column. Routine X allows totalling or other operations for groups of records specified by a control column. Routine Z merges columns from two files based on matching keys. Routine AA allows a user specified function to be applied to values in the file placing the results in the file. Routine DD breaks down a total in one column for groups of records specified by a control column according to proportions specified in a third column.

H. INTERFACE

The essential interfaces are as follows:

EE. tabular input merge correlate
FF. tabular output copy
GG. graphics internal format conversion
HH. image internal format conversion
II. remote sensed data cleanup

Routine EE is exactly like routine Z except that one of the merge files can have external labels and format, such as a tape from the U.S. Bureau of the Census.

Routine FF outputs tabular files in arbitrary format. Routine(s) GG are needed because a number of different external formats for graphic data exist, but routines such as A, B, and R need a particular internal format. Since external formats may or may not be chained, the internal format should allow both so that chaining routines will not be necessary. Usually one new routine is needed for each new application. Routine(s) HH serve an analogous function for image data. An example of image format conversion is the disinterleaving of multispectral components of Landsat to produce a single spectral band image. Routines for II fall into a highly complex area of removing sensor aberrations as necessary to perform an application. An example would be de-skewing a Landsat image to compensate for Earth rotation. There is a trend for the data providers (such as NASA) to provide ready to use products but it is likely that data customers will still find it necessary to perform these operations. Some customers prefer to begin with raw data while others find that the ready to use product has had an undesirable step performed.

III. USE OF SYSTEM ELEMENTS

Four prototype applications are taken from a recent paper by Zobrist and Nagy[4]. For each, a listing of major processing steps is given without much explanation or detail. The letter codes after each step indicates which system elements were used in the processing steps.

A. DESERT CONSERVATION

As part of the planning activity for the 25-million-acre California Desert Conservation Area, the Bureau of Land Management needed a survey of vegetaton and animal forage in a form that could be integrated with other elements of the plan. Multispectral classification of Landsat imagery was chosen as the best means of obtaining this survey, considering time and money constraints. The goal was to integrate digital terrain data, such as elevation, slope, and aspect, that have a strong effect on vegetation, with ground information from maps, surveys, and photointerpretation to aid in classification. The output was required in 1° x 1° quadrangles in Lambert Conformal Conic projection and in tabular form, categorized by grazing allotments, land ownership, and other types of boundaries.

The procedure for vegetation and forage classificaton consisted of 24 major steps:

1. Obtain Landsat raw data (10 frames).
2. Obtain map control points from conventional maps.
3. Convert the map coordinates to Lambert Conformal Conic grid (GG, N).
4. Find the coordinates in the Landsat frames (J).
5. Find edge-match control points for all overlapping Landsat frames (M).
6. Adjust edge-match control points according to a distortion model determined by the map control points (V, X, Z, AA).
7. Perform geometric correction and map projection of the Landsat frames according to the control points (II, P, Q).
8. Perform brightness adjustment of the Landsat frames according to differences noted in the edgematching control points (P, N).
9. Trim edges of the Landsat frames to remove bad data (J).
10. Mosaic the 10 frames together to obtain a single, large 7400 x 7500 image (J).
11. Repeat steps 7-10 for three more spectral bands of Landsat.
12. Section Landsat mosaic into 1° quadrangles (J).
13. Perform geometric correction on digital terrain file to obtain 1° quadrangles in the same grid map projection (P, Q, J).
14. Classify land cover unsupervised initial clustering (1993 clusters) with supervised pooling (25 final classes) by a discipline scientist (K, J, V, W, X, Y, AA).
15. Test the statistics on 12 areas of 400,000 acres each. Refine the statistics (K, U, J, S, V, W, X, Y, AA).
16. Extend the classification to the entire California Desert Conservation Area (K, U, J).
17. Coordinate digitize various boundary files such as land ownership.
18. Change the digitizer coordinates to Lambert Conformal Conic grid (GG, CC).
19. Change the boundary file from vector to raster format (A, B).
20. Identify the polygons determined by the boundaries (C, E).
21. Perform polygon overlay of classification files with boundary files. This yields a table of acreages of classes per district or parcel (S).
22. Perform data management operations such as sort, merge, merge correlate, aggregate, and cross-tabulate on the results of polygon overlays (V, X, AA, CC).
23. Report the tables by printer listing or magnetic tape (W, FF).
24. Report the image products by photographic image playback or magnetic tape (J).

The results of the California Desert Task were used as a primary input to the Desert Master Plan for range forage allocation for livestock and wild burros. The work was performed within a two-year deadline, primarily in digital format to facilitate 10-year updates mandated by Congress. Details of the work and utilizaion of the results can be found in McLeod and Johnson[5] and in the Bureau of Land Management's report,[6], respectively.

B. AIR POLLUTION STUDY

Several agencies in the Portland, Oregon, area have been working under the general coordination of the Pacific Northwest Regional Commission to determine the utility of Landsat digital images for urban applications. A map of land use from Landsat was prepared at the USGS EROS Data Center, with assistance from the NASA Ames Research Center. The application described here used the map to allocate population from the 1973 census update to traffic zones and two-kilometer grid cells. The grid-cell population was combined with air pollution data from the Oregon Department of Environment Quality to obtain a health impact map.

The procedure for the health impact interpretation consisted of 14 major steps:

1. Find control points in boundary files.
2. Find the same control points in the Landsat image (J).
3. Calculate a distortion surface from the control points (P).
4. Geometrically correct the boundary files according to the distortion surface (R).
5. Change the boundary file from vector to raster format (A, B).
6. Identify the polygons determined by the boundaries (C, E).
7. Perform polygon overlay of classification files with boundary files. This yields a table of acreage of classes per zone or cell (S).
8. Rearrange the table of acreages and produce a printed report (V, W, X, AA).
9. Perform polygon overlay of the census tracts with the grid cells taking the sum of residential pixels in each intersection (S, T).

10. Merge the population data according to census tract keys (Z).
11. Disaggregate the population data over the tract-cell intersections in proportion to the sum of residential pixels (DD).
12. Reaggregate the population data by grid cell (X).
13. Multiply population per grid cell by pollution per grid cell to obtain a tabulation of health effect and produce a printed report (AA, W).
14. Inject the health-effect tabular values into the grid-cell image to give a choroplethic map of health effect (U).

The Portland agencies can now combine these results with new data sets produced from the grid pollution model to calculate health effect on 1973 population. It is also easy to use census updates or projections to obtain health effects at a desired date. Parenthetically, census updates and projections are rather hard to do, but are undertaken by the Bureau of the Census and other agencies. Details are given in Bryant et al.[7].

C. URBAN EXPANSION DELINEATION

The Geography Division of the US Bureau of the Census was interested in the urban growth of the Orlando, Florida, region during 1970-1975. Since Landsat 1 was not placed into orbit until 1972, this study had to compare a 1975 land cover map derived from Landsat to a 1970 base constructed from the decennial census.

The procedure for assessing urban expansion consisted of 16 major steps:

1-6. Same as Portland case (above).
7. Count the pixels in each polygon to produce a table of areas according to census tract keys (Z).
9. Divide population by area and use 1000 people per square mile as a threshold for urban/non-urban (AA).
10. Inject the urban/non-urban values into the census track image to give a choroplethic map (U).
11. Use classification to extract basic land cover from the 1975 Landsat (K, U, J).
12. Use an image cell-value translation routine to convert the classes to urban/non-urban (O).
13. Use polygon overlay to produce tables of urban/non-urban land use by census tract (S).
14. Merge the tables according to census-tract key to produce a single table comparing urban growth by census tract for 1970-1975 (Z, AA, W).
15. Digitally overlay the census tract boundaries on the two urban area images to produce maps.
16. Use an image arithmetic combination routine to produce a change map (L).

It is interesting to note that Landsat data forms an historical archive over large areas of the earth dating back to 1972. Here, the status of an area in 1975 was retrieved. For a report, see Friedman and Angelici.[8]

D. SOLAR POWER STUDY

Our final example, a rooftop inventory for photovoltaic potential in California's West San Fernando Valley, consisted of 14 major steps:

1-6. Same as steps 1-6 of the Portland case using power station district boundaries.
7. Use classification to extract land cover from Landsat (K,U,J).
8. Correlate sample data from aerial photographs to estimate percentage of available rooftop in each land use class (J).
9. Perform polygon overlay to determine areas of land use per power station district (S).
10. Merge correlate power-station-district electricity use data into the land use table (EE, Z).
11. Convert land use area to rooftop area using rooftop percentages from step 8 (Z, AA).
12. Convert rooftop area to electric generation capacity using known characteristics of of photovoltaic arrays and assuming 50 percent utilization of rooftop (AA).
13. Compare with power-station-district electricity use and produce printed report (AA, W).
14. Inject the potential photovoltaic generating capacity (as percentages) into the power district image to produce a choroplethic map (U).

Engineers may tell us the characteristics of a single rooftop silicon array, but it is the geographer's task to provide data on widespread use. This study has shown that a surprising amount of the power needed by a suburban area can be provided by rooftop generation. Details are described by Angilici and Bryant.[9]

IV. CONCLUSION

A key methodology for the development of information processing systems for remote sensing data has been described in this report. If a system is built with primitive operations that have great generality, then widespread applications are possible using combinations of those primitives. Completeness of the set of primitives eliminates the need for programmers to fill in required steps for a new task. This method is a common one in computer engineering. One of its drawbacks is that an application can become a bewildering sequence of primitive operations, however, there are other methods such as commonality of data formats and macro processing to help tame the complexity of sequences of tasks.

V. REFERENCES

1. Bryant, N.A., and A.L. Zobrist, "IBIS: A Geographic Information System Based on Digital Image Processing and Image Raster Datatype," Proceedings of the LARS Symposium on Machine Processing of Remotely Sensed Data, Lafayette, Indiana, 29 June 1976, P. 1 A-1.

2. Tobler, W., and Lau, J., (1978), "Isopleth mapping using histosplines." Geographical analysis, 10, 3, 273-279.

3. Bauer, M.E., Cipra, J.E., Anuta, P.E., and Etheridge, J.B. (1979). "Identification and area estimation of agricultural crops by computer classification of Landsat MSS data." Remote Sensing of Environment 8, 77-92.

4. Zobrist, A.L., and Nagy, G. "Pictorial information processing of Landsat data for geographic analysis," Computer, November, 1981, pp. 34-41.

5. R.G. McLeod and H.B. Johnson, "Resource Inventory Techniques Used in the California Desert Conservation Area," presented at the Arid Lands Resource Inventory Workshop, La Paz, Mexico, Nov. 30-Dec. 6, 1980.

6. The California Desert Conservation Area Final Environmental Impact Statement and Plan, US Department of Interior, Bureau of Land Management, Oct. 1, 1980.

7. N. A. Bryant, A.J. George, Jr., and R. Hegdahl, "Tabular Data Base Construction and Analysis from Thematic Classified Landsat Imagery of Portland, Oregon," Proc., Fourth Annual Symposium on Machine Processing of Remotely Sensed Data, Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, June 21-23, 1977, pp. 313-318.

8. S.Z. Friedman and G.L. Angelici, "The Detection of Urban Expansion from Landsat Imagery," Remote Sensing Quarterly, Vol. 1, No. 1, Jan. 1979. Originally presented at spring meeting of Association of American Geographers, New Orleans, Apr. 1978.

9. G.L. Angelici and N.A. Bryant, "Solar Potential Inventory and Modelling," Data Resources and Requirements: Federal and Local Perspectives, contributed papers from the 16th Annual Converence of the Urban and Regional Information Systems Association, Washington, D.C., Aug. 6-10, 1978, pp. 442-453.

Albert L. Zobrist is presently a member of the technical staff at the Jet Propulsion Laboratory. He was an associate professor of computer sciences at the University of Arizona, Tucson, 1974-75, and from 1970-74 was a professor of electrical engineering and computer science at the University of Southern California, Los Angeles, where he was also a member of the research staff of the Information Sciences Instutute from 1973-74. Before that, he was a member of the technical staff at the Aerospace Corporation, El Segundo, California.

Zobrist received the BS degree in mathematics from Massachusetts Institute of Technology, Cambridge, in 1964, and the MS degree in mathematics and the PhD degree in computer science, both from the University of Wisconsin, Madison, in 1966 and 1970.