Reprinted from

Symposium on Machine Processing of Remotely Sensed Data

June 21 - 23, 1977

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
77CH1218-7 MPRSD

TABULAR DATA BASE CONSTRUCTION AND ANALYSIS FROM THEMATIC CLASSIFIED LANDSAT IMAGERY OF PORTLAND, OREGON*

NEVIN A. BRYANT
Jet Propulsion Laboratory

ANTHONY J. GEORGE, JR.
Oregon Department of Environmental Quality

RICHARD HEGDAHL
Columbia Region Association of Governments
(CRAG), Portland, Oregon

ABSTRACT

The creation of a land use map of the Portland, Oregon region from LANDSAT has proven to be a useful demonstration that helped substantiate the utility of satellite thematic mapping for regional planning purposes. The tabulation of land use acreage by traffic zones and census tracts, with the application of Image Based Information System (IBIS) software, has permitted a more systematic verification of the LANDSAT classification. Subsequent tabular data base constructions merged socioeconomic and environmental data with the LANDSAT-derived land use statistics. The application of IBIS software to the data bases permitted the crosstabulation of population statistics and modelling and portrayal of per capita air pollution impact.

The development and system characteristics of an image based information system (IBIS) that makes use of digital image processing techniques has been described by Bryant and Zobrist in June 1976.[1] A brief summary of the approach is described using Portland case as an example.

The generation of land use statistics by census tracts and traffic zones permitted the direct comparisons of acreage estimates derived from LANDSAT with those obtained from conventional air photo interpretation and field surveys. The systematic errors associated with the Bayesian classifier as it was applied to different LANDSAT frame dates could be noted for different portions of the scene. Thus, it was possible to note the degree of misclassification in the various suburban, industrialized, and central business district areas of the region.

The merging of the LANDSAT-derived tabular data sets with socioeconomic data was accomplished using conventional sort-merge routines. IBIS-based applications included several spatially-dependent manipulations of both kinds of data. In the first instance population statistics by census tract were allocated to traffic zones in accordance with the LANDSAT-derived distribution of residential land use. Similarly, it was possible to use residential and commercial land use acreage by traffic zone as coefficients of trip generation to map potential automobile trip generation. Finally, it was possible to model air pollution impacts through the pixel-for-pixel cross correlation of land use and socioeconomic data collected by census tract and environmental data collected by two kilometer grid cells and map the results as a contoured surface.

INTRODUCTION

Several agencies in the greater Portland area have during the past two years had the privelege of working on a task designed to demonstrate and determine the utility of LANDSAT digital imagery for urban applications. Under the general coordination of the Pacific Northwest Regional Commission, thematic classification of LANDSAT images have been executed at both NASA/Ames Research Center and the USGS-EROS Data Center, while special purpose data aggregation has been undertaken at the Jet Propulsion Laboratory's Image Processing Laboratory. The principle topic to be discussed here will be the data aggregation procedures undertaken by the JPL-IBIS (Image Based Information System) geographic information system.

Metropoliatn regions have in recent years become increasingly impacted by competing demands for natural and economic resources. As the value of land and the cost of construction materials has increased, monitoring urban change and modelling urban trends has moved from being of purely academic and long-range planning interest to becoming an important factor in economic decision-making for the near term. Scarcity of natural resources, particularly energy, water, and land, have intensified the conflict between commercial and environmental interest groups. For the growing metropolises of the South and West concerns focus on the need to accurately project future energy demand and conversion of agricultural land, while the predominant concern of stable cities of the East and North if the more efficient utilization of existing resources and discovering ways to revitalize the inner city.

---

* This paper 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.

In an effort to better plan for resource utilization and more fairly assess the impact of alternative proposals for land use, many state and local governments are expanding their use of remote sensing and incorporating the interpreted results of remote sensing into geographic information systems. For instance, changes in land use noted over a two year period may be used to help update decennial Census statistics and predict future growth trends. Characteristically, no individual data source, such as a census or survey, can provide answers to the variety of problem addressed by decision makers. The cost of comprehensive surveys is usually too expensive for local governments, whereas the combination of a variety of existing data sets (e.g., building permits) and remote sensing products frequently provide estimates that fall within the statistical error inherent in survey techniques.[2] Geographically integrated information systems, which have progressed roughly at the rate of advancement in computer technology, seek to capitalize on the synergism inherent in being able to automatically compare a variety of socioeconomic, environmental, and land use data sets for the same point on the ground (see Figure 1).

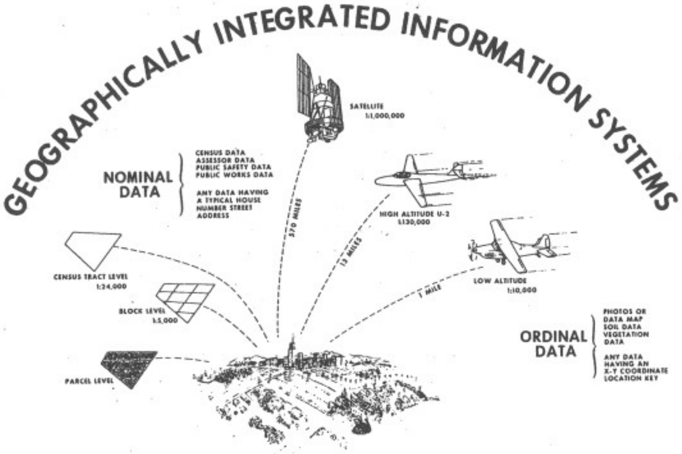


FIGURE 1

Conceptual Diagram of Data Components Used In Metropolitan Geographic Information Systems.

Practitioners of the art of geocoded systems design have progressed with varying degrees of success towards the goals outlined. As a rule, generalized systems have only rudimentary data manipulation capability (i.e., status updating and interrogation by area), while highly specific and specialized systems have progressed further with modelling applications.[3] Polygon and grid cell coded information systems have been developed to access data for selected areas. Such systems rely on the tabular formatting of the input data, a costly and time-consuming process. Often the system falls into disuse because the updating of major segments of the data base becomes prohibitively expensive. In response to the desire to provide up-to-date resource information, investigators have studied the feasibility of applying LANDSAT and high altitude photo imagery to natural resource mapping. The consensus which appears to be evolving is that, while remotely sensed imagery can provide timely coverage and sufficiently accurate maps using multispectral classification techniques, the end product is still a map that cannot interface directly with an existing geocoded information storage and retrieval system.[4]

LAND USE MAP GENERATION

The production of a map of land use from the LANDSAT digital imagery was performed at the EROS Data Center, with assistance from the NASA-Ames Research Center. Actual editing and land cover generation was performed by Oregon personnel from the Oregon Land Conservation and Development Commission, Oregon Department of Environmental Quality, Columbia Region Association of Governments, Multnomah County, and the City of Portland. While a complete description of the procedures and discussion of the classification results can be found in the U.S.G.S. open file report, a summary follows.[5]

A 1,000 square mile area, centered on metropolitan Portland, was chosen as the project area because of the diverse types of urban and non-urban forms of which the area is comprised and because it contains numerous small urban centers with high growth rates. Parts of five Oregon counties, Columbia, Washington, Multnomah, Yamhill, and Clackamas are included in the project area, as is part of Clark County, Washington.

The overall goal of the comprehensive land cover classification demonstration was to categorize land cover types within the Portland region. Another objective of the demonstration was to determine which season or combination of seasons appeared to be best for land cover inventory. Consequently, a four-date temporal overlay of 1973 LANDSAT data was prepared, including data from January, April, July, and October. Parallel classifications were developed for single dates and pairs of dates using the Image-100, LARSYS, and IDIMS systems.

To check the results of each classification iteration and to provide a quantitative means of comparing classification results, four tests sites (ranging in size from 5 to 15 square miles) were established. Land use and land cover maps and acreage summary tables were prepared for each site from interpretation of aerial photographs and local knowledge of the area. The January data was not useful, because of the low degree of contrast between land cover types. Individual classifications of the April, July, and October data revealed that October was the better month for land cover classification accuracy. Comparable results were obtained from using a combination of April and July LANDSAT data, as well as July and October. Classification results from combining April and October data were less accurate.

Classification accuracy between the three systems, Image-100, LARSYS, and IDIMS, were similar for the majority of classifications tests. For production of a final land cover map, the combined April-

July classification obtained from using the LARSYS software was selected. A number of carefully chosen plots were submitted to a clustering algorithm, using eight LANDSAT bands (four from April and four from July). Six of the eight LANDSAT bands, 4, 5, and 7 from April and 5, 6, and 7 from July, were selected to classify test areas using a Gaussian maximum likelihood function, and the resulting cluster classes matched with desired land cover classes. Residential areas, improved open space (urban), clear water, forested, and agricultural land were identified satisfactorily but confusion existed between commercial/industrial and turbid water. Manually selected training fields were developed for these two land cover classes, which resulted in successful separation. Figure 2 presents the final product.

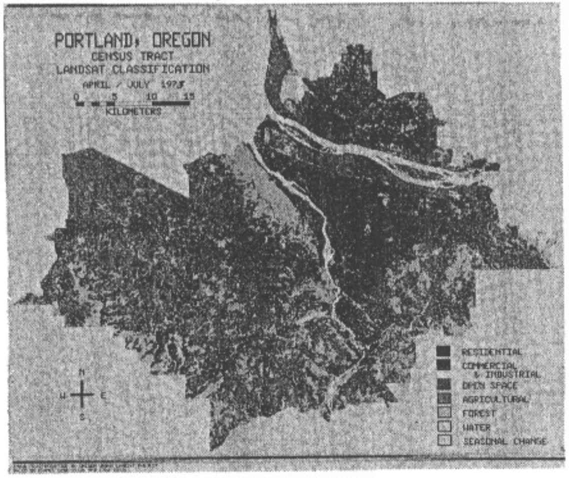


FIGURE 2

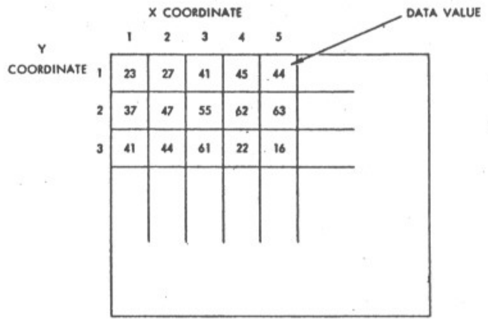
Land Use Maps of Portland Derived from LANDSAT

INTERFACING THE IMAGE BASED INFORMATION TECHNOLOGY AND THEMATICALLY CATEGORIZED LANDSAT IMAGERY

Data Management Considerations

The initial motivation for the development of an Image Based Information System (IBIS) was to introduce remotely sensed imagery into the mainstream of data processing application. A basic example is that of processing a LANDSAT thematic map showing land use or land cover in conjunction with a census tract polygon file to produce a tabulation of land use acreages per census tract. An analysis of the steps necessary to achieve this basic capability brings forth two facts. First, a large number of image processing and data manipulation capabilities would be needed for even the simplest case. Second, with the proper design the minimal system can be extended into a general information system with novel features and capabilities. The term Image-Based Information System (IBIS) has been adopted because the image datatype and image processing operations are crucial to many of the new capabilities.

Until recently, the image format has been used primarily as a computer processable equivalent of a photograph, with the value stored in each cell of the image representing a shade of grey or a color. But if the image is of a geographical area, then the value in the cell can be a datum for the area corresponding to that cell. A principle advantage of the image representation is that data for a geographical point can be accessed immediately by position in the image matrix. Figure 3 illustrates the calculation of memory address of the data value from a latitude-longitude pair.



	1	2	3	4	5
1	23	27	41	45	44
2	37	47	55	62	63
3	41	44	61	22	16

LOCATION FORMULAS:

$$X = \left[A \times LAT + B \times LONG + C\right]_{\text{NEAREST INTEGER}}$$

$$Y = \left[D \times LAT + E \times LONG + F\right]_{\text{NEAREST INTEGER}}$$

$$\text{MEMORY ADDRESS} = \text{BASE} + KX + Y$$

FIGURE 3

Image Matrix As A Data Representation.

The image datatype seems to be a powerful and general representation for spatially distributed data, and the range of uses can be divided into several broad categories:[6]

Physical Analog: The pixel value represents a physical variable such as elevation, rainfall, smog density, etc.

District Identification: The pixel value is a numerical identifier for the district which includes that pixel area.

Class Identification: The pixel value is a numerical identifier for the land use or land cover, or for any other area classification scheme.

Tabular Pointer: The pixel value is a record pointer to a tabuler record which applies to the pixel geographical area.

Point Identification: The pixel value identifies a point, or the nearest of a set of points, or the distance to the nearest set of points.

Line Identification: The pixel value identifies a line, or the nearest of a set of lines, or the distance to the nearest of a set of lines.

The applications of IBIS to the Portland case made use of the first four datatypes, physical analog, district identification, class identification and tabular pointer. These four functions apply spatial integration of data sets with reporting by district (i.e., census tract, traffic zone, two kilometer grid cells). The products derived, therefore, do not represent modelling in the true sense, but rather the product of spatially-structured tabulation and crosstabulation.

Land Use Tabulation

A thematic land use map can be tabulated only after a number of preparatory steps are performed. As the traffic zone and census tract boundary lines were supplied on a map, control points needed to be located, the boundary lines digitized onto tape, and finally output as LANDSAT row and column positions. Following the digitizing steps, the lines are digitally output (scribed) onto a blank image. This task is performed by converting the unit of measure in which the tape was digitized to the line and sample coordinates of the LANDSAT image. Inputing control points from a map featuring the unit of measure used for the tape file and identical ones on the image into a conversion routine produces an accurate transformation. The image is then inspected closely for digitizing errors, such as missing lines, gaps left between lines, or errant lines that do not represent actual boundaries. Photographic enlargements aid the analyst in finding the line and sample coordinates of these errors. Alterations to the digitized data or in software parameters used for reading the input tape finally produce an image of boundary lines ready for overlaying onto the LANDSAT image. Upon digitally summing the boundary line image and the LANDSAT image, the technician can view the overall registration quality. Registration within 50 (fifty) pixels of the accurate location is usually sufficient at this point in the process. Gross misregistration requires rechecking control point numbers or shifting the starting line and sample of the boundary line image.

The precise registration of the boundaries requires the use of additional control points and application of a continuous surface fitting algorithm. With an enlargement of the image of the boundaries overlaid onto the LANDSAT frame, corresponding points of the district lines and on the LANDSAT frame can be located and marked. When a collection of such points is found throughout the image, their row and column coordinates are listed as well as the differences between the present and the correct coordinates. After placing the points in a proper format, a computer routine readjusts the boundaries to fit as precisely as the digitizing accuracy allows. Figure 4 shows the digitized census tracts correctly overlaid onto the Portland frames, which Figure 5 shows the traffic zones. The boundary lines, once adequately registered, are then processed by a routine that fills holes left due largely to double-digitizing. Each district is then encoded with a unique digital value, and the border lines are randomly assigned to the adjoining districts. An image displaying the different regions in four of five shades of grey is created (Figure 6 and 7).[7]

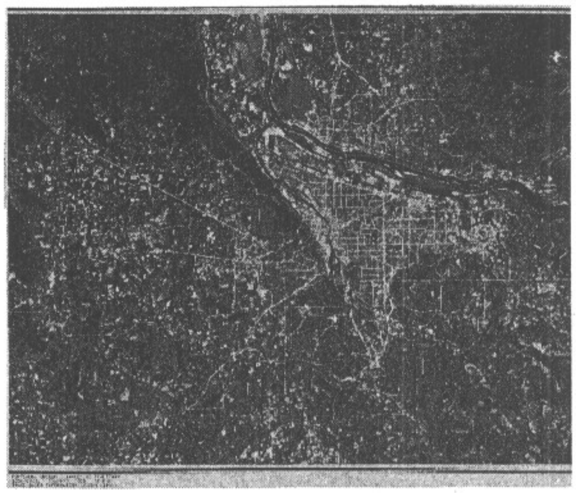


FIGURE 4

Census Tract Boundaries Registered to LANDSAT Image of Portland, Oregon.

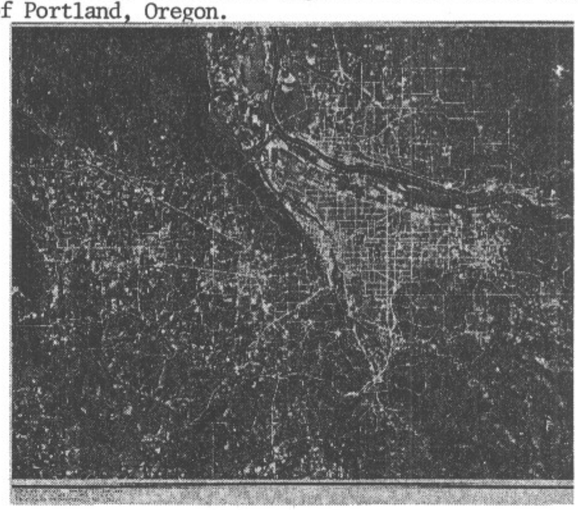


FIGURE 5

Traffic Zone Boundaries Registered to LANDSAT Image of Portland, Oregon.

At this point the classification map image and the uniquely encoded georeference image, both registered to the LANDSAT image, can be interfaced. A computer routine digitally overlays both images row by row and outputs histograms of data representing the number of pixels of each class in each administrative district. The columns are then sorted, aggregated, and transferred in accordance with the user's specifications and new columns listing area (in acres, etc.) and calculating percentages of each land use are created. After the uniquely encoded areas are properly labeled by their district names, a tabulation of land use by administrative district is produced.

The generation of land use statistics by census tracts and traffic zones permitted the direct comparisons of acreage estimates derived from LANDSAT with those obtained from conventional air photo interpretation and field surveys. The systematic errors associated with the Bayesian classifier as it was applied to different LANDSAT frame dates could be noted for different portions of the scene. Thus, it was possible to note the degree of misclassification in the various suburban, industrialized, and central business district areas of the region.

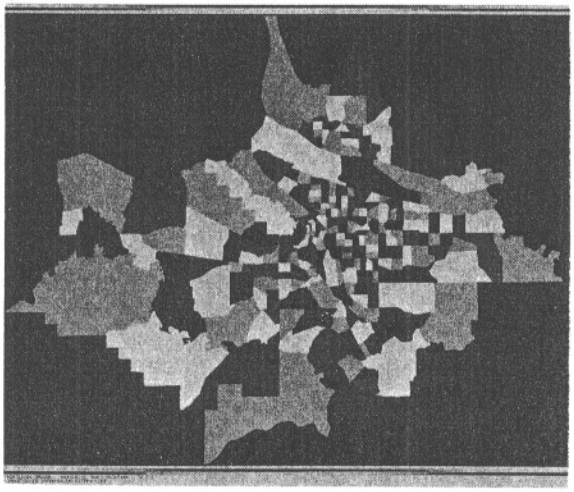


FIGURE 6

Geo-reference Image of Portland Census Tracts

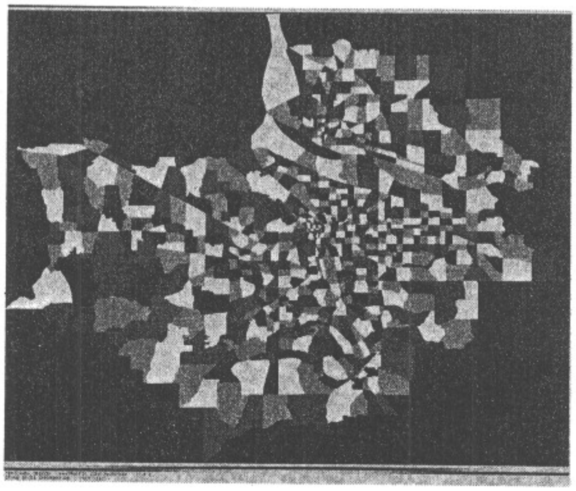


FIGURE 7

Geo-reference Image of Portland Traffic Zones

Crosstabulation:

Two crosstabulation applications were applied to the Portland data. In the first case, population statistics were allocated to traffic zones from census tracts in proportion to the distribution of residential land use. Normally such allocation tasks are performed by constructing area correspondence tables that assume an equal distribution of population over both kinds of districts. As the assumption of uniform population distribution is never valid, any distribution estimate based upon a residential land use map that is better than fifty percent accurate (i.e., equivalent to random) assures a better allocation of population. The proportionate allocation of population by traffic zone will permit CRAG to combine census and traffic data with residential and commercial land use acreages to predict automobile trip generation.

The second application was the assignment of population from the 1973 update of Census statistics to the 2 kilometer grid map of pollutants used by the Oregon Department of Environmental Quality. For each grid in 1973, the calibrated grid concentration (total annual suspended particulates) was derived from a grid of 1973 uncalibrated grid concentrations (C) using the formula:

$$\text{Conc}_{(x,y)} = [0.66\ C_{(x,y)}]\ [\text{Population}_{(x,y)}]$$

From the derived data it has been possible to calculate over the entire area the percent of population exposed to the following levels of total suspended particulates:

75 µg/M
60 µg/M
50 µg/M
40 µg/M

$$\frac{\Sigma\ \text{Pop}\ (x,y)\ \text{exposed to level}}{\Sigma\ \text{Pop}(x,y)} \times 100\%$$

In the future, it is planned to use the residential land use data as a distribtuion coefficient that maps the pollution impacts more precisely, and derives contours of impacted areas (75 µg and 60 µg levels).

ACKNOWLEDGEMENTS

Many have contributed to the results presented in this paper. Al Zobrist was responsible for the system architecture and much of the algorithm development. The programming effort was assisted by Howard Wilczynski and John Addington. Development of the case presented was performed by Thomas Logan at JPL and by William Todd of the EROS Data Center.

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. "FY76 Summary Report on the Thematic Mapper Simulation in Metropolitan Regions", Typescript. Earth Resources Applications Group, Jet Propulsion Laboratory, Pasadena, CA, 12 July 1976.

3. Kohn, C.F., "The 1960's: A Decade of Progress in Geographical Research and Instruction," Annals of the Association of American Geographers, Vol. 60, No. 2, (June 1970), p.215.

4. Eastwood, L.F., et. al., Preliminary Needs

Analysis Report: Program on Earth Observation Data Management Systems, Center for Development Technology, Washington University, St. Louis, MO, December 31, 1975.

5. Todd, W.S., R.N. Hall, and C.C. Henry, "Metropolitan Land Cover Inventory Using LANDSAT Satellite Data" Open File Report, U.S. Geological Survey, EROS Data Center, Sioux Falls, S.D. May 1977.

6. Zobrist, Albert L. "Elements of an Image Based Information System", Proceedings Caltech/JPL Conference on Image Processing Technology, Data Sources and Software for Commercial and Scientific Applications, California Institute of Technology, Pasadena, California, November 3-5, 1976, p. 3-1 - 3-5, JPL SP 43-30.

7. McLemore, B.D., and A.L. Zobrist, "Computability of Five Color Maps", Symposium on Current Mathematical Problems in Image Science, Monterey, California, November 10-12, 1976, (in press).

Nevin A. Bryant is responsible for the conceptualization and application of geographic information systems based on digital image processing technology and the use of remotely sensed imagery at the Jet Propulsion Laboratory's Image Processing Laboratory.

Dr. Bryant received his B.A. degree from McGill University in Honours Geography; his M.A. degree from the University of Hawaii in Geography, and his Ph.D. from the University of Michigan.

Dr. Bryant has conducted field research in Western Samoa and Java, Indonesia related to the interaction between population growth and regional economics and has more recently undertaken projects related to vegetation and land use mapping. His present interest include data reduction and interpretation of satellite imagery, geographic information systems, and regional economic modelling with particular emphasis on metropolitan areas.