IMPLEMENTATION AND EVALUATION OF ILLIAC IV ALGORITHMS FOR MULTISPECTRAL IMAGE PROCESSING:

FINAL REPORT

Philip H. Swain
Principal Investigator

July 1974

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

*Report by the subcontractor, the Laboratory for Applications of Remote Sensing, Purdue University, to the subcontractee, the Center for Advanced Computation, University of Illinois. This project was supported by NASA and the U.S. Department of the Interior under NASA Grant NGR 14-005-202.
IMPLEMENTATION AND EVALUATION OF ILLIAC IV
ALGORITHMS FOR MULTISPECTRAL IMAGE PROCESSING:
FINAL REPORT*

Philip H. Swain
Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, Indiana 47907

ABSTRACT

This is the final report concerning one phase of a multi-
disciplinary and multi-organizational effort to implement
multispectral data analysis algorithms on a revolutionary
computer, the ILLIAC IV, and to demonstrate the effectiveness
and efficiency of the implementation in the context of a
specific application, land use classification. The team
effort, involving personnel from the Laboratory for Applica-
tions of Remote Sensing at Purdue University, the Center for
Advanced Computation at the University of Illinois, the
Geography Department at California State University (San Jose),
and the Geographic Applications Program of the U. S. Geological
Survey, has been notably successful in the phase reported
herein, viz., demonstrating the utility of the digital multi-
spectral data analysis techniques for producing useful land use
classifications from satellite-collected data.

INTRODUCTION

During the period March 1973 through October 1973, a study
was carried out jointly by the Center for Advanced Computation
(CAC) of the University of Illinois and the Laboratory for
Applications of Remote Sensing (LARS) of Purdue University for
the purpose of evaluating the application of certain advanced
data processing systems to multispectral remote sensing data
processing. It was proposed that the ILLIAC IV computer, its
associated archival laser store, and the ARPA network be
utilized for the interpretation of multispectral scanner data
from the ERTS-1 satellite. Experiments were to be conducted
in the following areas:

*Report by the subcontractor, the Laboratory for Applications
of Remote Sensing, Purdue University, to the subcontractee,
the Center for Advanced Computation, University of Illinois.
This project was supported by NASA and the U. S. Department
(1) Parallel algorithms would be implemented and documented for definition of most separable land use and natural resource categories using multispectral cluster analysis techniques. Parallel algorithms would also be implemented for classification of image elements into aggregate categories using multivariate Gaussian statistical decision theory.

(2) ARPA network linkages between ILLIAC IV and existing interactive statistical, graphical and information management systems of CAC and LARS would be explored to allow supervised training of the ILLIAC IV processor in accordance with ground truth information at hand. In this same context, various forms of hard copy output of interpreted information would be explored.

(3) A frame of ERTS-1 imagery corresponding to the San Francisco region would serve as an example multispectral image for ILLIAC IV algorithm development. An early analysis of this frame by LARS would serve as a reference by which the validity of the ILLIAC IV analysis could be checked. This early analysis at LARS would also serve to guide proposed system developments in those directions potentially most beneficial.

(4) Experiments in digital image analysis for land use classification would be conducted to assist the evaluation of alternative natural resource and land use classification systems with respect to compatibility with the capabilities of digital multispectral data analysis methods and the information needs of resource management agencies. Representing the potential client-users of these classification systems, USGS/EROS personnel would participate in the selection, testing, and evaluation of classification system alternatives. Initial investigations in this area would focus on evaluation, with respect to digital multispectral classification techniques, of the Level I and Level II classification systems outlined in USGS Circular 671: "A Land Use Classification System for Use with Remote Sensor Data."

(5) Portions of two ERTS multispectral images corresponding to the San Francisco Bay Area would be analyzed using a common land use classification system to reflect land use changes apparent from the classifications. Alternative strategies would be explored for automatic detection, analysis and display of the type and location of land use transitions within the contexts of a more comprehensive ILLIAC IV/ERTS data processing facility.

(6) Based on the research outlined above, alternative designs would be proposed for a more comprehensive ILLIAC IV/ERTS data processing facility. Alternatives would reflect detailed analysis of the efficiencies of the laser archival memory for storage and retrieval of both raw ERTS data and interpretive information. Alternatives would be outlined for interfacing geographically dispersed user groups to the ILLIAC IV system via the ARPA network.
The advantages and costs associated with alternative designs would be specified and implementation schedules outlined.

The role of the Laboratory for Applications of Remote Sensing was specifically to:

(1) Provide technical consultation concerning the computer analysis of digital multispectral remote sensing data.

(2) Recommend specific algorithms for processing multispectral data, provide sample computer subroutines in which these algorithms were implemented, and provide sample processing results with which to check results of the ILLIAC IV implementation.

(3) Provide personnel and computer facilities for the analysis of up to two frames of ERTS-1 multispectral scanner data in order to provide a base-line result for comparison with the ILLIAC IV processing results.

(4) Provide technical guidance and computer facilities to assist in the evaluation of digital multispectral data analysis methods for natural resource and land use classification. Alternative land use classification systems would be investigated to assess their compatibility with both the multispectral data analysis capability and the information needs of resource management agencies.

(5) Participate in the formulation of a follow-on proposal based on a critical evaluation of the results achieved under the present investigation and a further assay of the needs of the potential user agencies.

ACCOMPLISHMENTS

Implementation of Algorithms on ILLIAC IV

Implementation of the data analysis algorithms on ILLIAC IV and investigation of the potential utility of ARPA network linkages were primarily the responsibility of CAC and the details of this work are reported separately by CAC.* LARS assisted in these activities to the extent of providing detailed descriptions and copies of the software for the algorithms as implemented in LARSYS (the Purdue/LARS multispectral data processing system) for a general purpose computer. CAC personnel were subsequently successful in implementing on an ILLIAC IV simulator both the clustering and maximum likelihood classification algorithms. The ILLIAC IV implementation produces results which are identical with the LARSYS results to the

precision observed. Timing comparisons have been impressive, showing the ILLIAC IV would run at two to three orders of magnitude faster than the general purpose system. [At this writing, essentially similar results have been obtained on the ILLIAC IV machine itself.]

Evaluation of Digital Multispectral Data Analysis for Land Use Classification

Efforts to produce urban land use maps through the application of LARSYS to multispectral data from ERTS-1 have been highly successful. Many of the results are reported in a symposium paper, a copy of which is attached as an appendix to this report.* Additional results and some technical details of significance are included here.

It has been shown that, using relatively broad land-use classes, a remarkable level of classification accuracy is attainable despite the coarse resolution of the satellite data and the inherent complexities of man-made land cover.

The ability to apply digital techniques for geometric rectification of the MSS data has proved invaluable. In the training phase of the analysis (classifier design), the analyst's job of correlating locations in the data with ground information, whether maps or aerial photography or other forms of supporting data, is made far easier once the rectification operations (described briefly in the attached paper) have been performed. After classification is completed, of course, the printout product is much closer in terms of geometric accuracy to a map. The interaction of LARS and USGS personnel participating in this project have considerably advanced the technology and improved the achievable accuracy obtainable from the geometric rectification process.

Another achievement has been the demonstration that "rectangular" grid systems (e.g., based on the UTM system) can be used as a basis for the aggregation and tabulation of analysis results. This has significant implications for the direct utilization of these results as well as for the efficient storage of the results in computerized information systems.

A LOOK AHEAD

This project has demonstrated that digital techniques hold great promise for the effective, efficient, and timely production of land-use information from satellite-collected multispectral scanner data. The implications of the availability of the parallel processing capabilities of ILLIAC IV are enormous in this respect.

However, the surface of the available or nearly available technology has scarcely been scratched with respect to its application to land-use classification. It is felt that the land-use classification system can be much improved beyond its present state. In addition, the procedures for carrying out the analysis can clearly be further refined. The utilization of temporal (time-varying) information available from the multiple passes of the satellite over the same area has not yet been investigated. And the range of useful products which the machine-assisted analysis can be expected to yield has yet to be determined.

Future efforts along these lines will further demonstrate the advantages of speed, accuracy, cost-effectiveness, timeliness (frequency of monitoring), and versatility which the machine-assisted analysis of airborne multispectral data make available for land-use classification. Proposals recommending specific directions for these efforts have been prepared and submitted to the sponsors of this investigation.
APPENDIX

URBAN LAND-USE MAPPING BY MACHINE PROCESSING OF ERTS-1 MULTISPECTRAL DATA:
A SAN FRANCISCO BAY AREA EXAMPLE*

Richard Ellefsen, Philip Swain and James Wray
California State University, San Jose, California; Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana; and U.S. Geological Survey, Washington, D.C.

ABSTRACT

Results are promising of an experiment to map land use in an urban area by automatic digital processing of ERTS-1 data. Computer-maps of a large segment of the San Francisco-Oakland and San Jose Urbanized Areas have been produced at a scale of 1:24,000 using a segment of an ERTS-1 frame reformatted to correct skewness and scale. An area of some 6,500 square kilometers was also mapped at 1:48,000 (a one-fourth sample). For both scales, urban areas were separated from rural -- using a photo interpretation procedure -- to solve problems of the spectral similarity of functionally different land uses and land covers. Classification was achieved by grouping twenty-eight spectral classes into eleven functional classes. Reliability was checked by comparing computer results to contemporaneous high-altitude color air photographs on a pixel-by-pixel basis. Performance results are high considering the grossness of the data and the complexity of the urban landscape.

INTRODUCTION

Discussed here are the results of attempts to create computer-produced urban land-use maps using multispectral scanner data from a satellite. The study is an outgrowth of research questions posed by individuals connected with LARS/Purdue, NASA, the Earth Resources Observation Program of the Department of the Interior, and the Geographic Applications Program of the U.S. Geological Survey.

Specific study objectives have been: (1) by LARS to test the applicability of the LARSYS pattern recognition software to urban land-use studies in an area where contemporaneous ground truth was available; and (2) by the Census Cities Project of the Geographic Applications Program (Wray, 1972) to attempt to utilize ERTS-1 data as a support or possible replacement for land-use mapping achieved through conventional air-photo interpretation. Further Geographic Applications Program goals are to utilize maximally the ERTS-1 data to: (1) produce print-out maps of large (1:24,000) scale; (2) aggregate digitized land-use data which may be used in conjunction with such reported ground-collected data as census reports and parcel ownership; and (3) monitor urban change on a regular basis.

*This research was supported in part by NASA grants NGL 14-005-202 and NGL 15-005-112. Support also was provided by the U.S. Department of the Interior EROS Program.

2a-7
For the land-use mapping, the goal of this work has been to attempt to de-
lineate with maximum accuracy as many functional land-use classes as possible
given the existing limitations of data and processing techniques. The term func-
tional must be emphasized as it pertains to usage of land, and not to naturally
occurring land cover. Further, these functional land uses are selected to corre-

Workers facing this translation problem have proposed varying solutions. At
LARS, a land-use map of Milwaukee County (Todd, 1973) was prepared with thirteen
classes some of which were statements of broad urban patterns such as "inner city"
and "suburban" while others were of such discrete land cover as grass and five
classes of water. Work at Johnson Space Center (Erb et al., 1973) took a more
general tack in defining six broad classes which combined spectral and functional
characteristics. Included were "large buildings and building complexes with high-
reflective roofs" and "open grass-covered fields with few trees."

A further consideration throughout this research effort has been to approxi-
mate as closely as possible (1) the land-use classes employed in the Geographic
Applications Program's mapping of the San Francisco nine-county test site
(Ellefsen and Peruzzi, 1972) and (2) the proposed system of the Geological Survey
(Anderson et al., 1972). The attempt has thus been made to determine the limits
of inferring land uses from spectral information alone. This approach was select-
ed realizing that some urban planners (Grey, 1973) argue that intensive, accurate
land-use determination from remote-sensed imagery of any sort is not possible and
that the plotting of such discrete uses as retail, office, and many multi-family
residential units are possible only with parcel-wise data secured from ground
sources.

DATA PROCESSING

For this analysis, data were selected from the July 26, 1972 ERTS-1 frame
(system corrected digital data from the multispectral scanner) which includes the
heavily urbanized area on the east, south, and west sides of San Francisco Bay.
The multispectral imagery for this area was first subjected to a geometrical pre-
processing operation on the computer which (1) rotated the orientation to north-
south, (2) removed the skew due to the earth's rotation, and (3) rescaled the data
so that computer line-printer maps would have a scale of 1:24,000. In the pre-
processed data, each multispectral data point represented a rectangular area on the
ground, 61.0 m by 76.2 m (.465 hectare).

The data were then analyzed using the pattern recognition techniques imple-
mented in LARSYS, the remote sensing data processing system developed by LARS
(Swain, 1972). In particular, an "unsupervised classification" approach was
employed: cluster analysis was used to isolate spectrally distinct classes in the
multispectral imagery and available ground truth data were then used to associate
a ground-cover description with the resulting spectral classes. A systematic ten
percent sample of multispectral data from an urban area in the vicinity of San
Jose was subjected to the cluster analysis. A total of 5226 resolution elements
were clustered into 30 spectrally different classes. Two of the resulting classes
were later deleted because they contained too few points for computing second order
statistics. A printout of the classes was then compared against ground truth, viz.
contemporaneously taken high-altitude color infra-red air photographs (1:130,000
scale RC-10 diapositives flown by U-2 aircraft by NASA in support of GAP's Census
Cities Project). The photographs were superimposed over the computer print-out
with the aid of a Bausch and Lomb Zoom Transfer Scope. Precise determination
of character of land use, on a pixel-by-pixel basis was thus possible. Results were
tallied and then analyzed to determine the best grouping of the spectral classes
into functional classes. A total of eight categories of urban land uses and
three rural land uses were finally selected as presented in Table 1.
A further concern throughout the experiment has been to determine how well automatic processing of ERTS-1 data interacts with the scheme for the classification of land use based on remotely sensed data proposed by the Geological Survey (Anderson et al., 1972). That system contains levels of generalization I and II. More discrete level III uses, collapsible into the more general classes, may be developed by users. Comparison with the classification system developed in this experiment (see Table 2) demonstrates, at least for the studied urban sample, that machine processing of satellite scanner data is capable of a much finer classification than simply Level I. Level II is achieved in the delimitation of residential use, open space, and a combined commercial and industrial class. In addition, the very fact that we are dealing with spectral characteristics permits (even requires in order to maximize the available data) the classifying of parking lots and mobile home parks and differentiating between improved and unimproved open space and thus enters the realm of the discrete uses treated in Level III.

In short, the combination of the satellite-borne scanner and machine processing provides a different tool than either conventional air-photo interpretation or surface and statistical unit mapping. Each produces a somewhat similar but yet different product and each has advantages and disadvantages. While these have been documented for air photography and ground methods, the characteristics of machine-processed satellite data as applied to urban mapping are not well known and deserve presentation.

The advantages are:
1. High speed processing
2. Frequently obtained new data
3. Unbiased and uniformly repetitive classification
4. Production of print-out maps at a large map scale at relatively low cost (once the system becomes operational)
5. The inherent digitizing of land-use data retrievable in virtually any form or combination of forms

The disadvantages are:
1. The inability of the system to discriminate with consistent success between functionally dissimilar but spectrally similar land uses
2. The impossibility of detecting parcel ownership
3. Generalization by resolution element: at 60 meter resolution the complexity of the urban landscape cannot be shown fully
4. Identifications dependent on vegetation vary seasonally
5. Uncontrollable incidence of cloud cover

Review of the above suggests that for many potential users, the satellite/machine-processing system has advantages which outweigh disadvantages and will be welcomed as a new, powerful tool in spatial analysis work. Other users may have to await refinements in the system which will surely come with subsequent developments in scanner and data processing capabilities.

TEST AREA

Computer maps were created at three scales. For the largest -- at full size (all pixels) for a map scale of approximately 1:24,000 -- a total area of about 1,125 square kilometers was mapped. Corporate units include Oakland, Alameda, San Leandro, Hayward, Fremont, San Jose, Santa Clara, Mountain View, Sunnyvale, Palo Alto, Menlo Park and several lesser suburban municipalities.

A larger area, some 6,500 square kilometers, was also mapped at a scale of 1:48,000. A one-fourth sample was achieved by instructing the computer to classify and print out alternating pixels per column and per line. Only one-fourth of the data were treated; they were not averaged.

A similar approach was followed in preparing a computer map at a scale of 1:72,000 for an even larger segment of the area of the nine San Francisco Bay Counties by using the device of a one-ninth sample (every third pixel by column and by line). Though skew and scale were uncorrected, the product served to demonstrate the possibility of mapping large urban and urban related areas; a total of approximately 21,000 square kilometers were mapped.
The test area has many general characteristics in common with rapidly growing cities throughout the United States. Development has been more horizontal than vertical with: large areas of new single-family residences built on converted agricultural land; several large clusters of new industrial complexes; shopping centers; and various institutions. The original core areas of the nuclei cities, from which growth spread in the past twenty years, remain as islands of old within the new, but many of these have been significantly altered. Connecting ribbons of commercial development are visible even to the unaided eye on the ERTS-1 imagery. Small patches of unbuilt-upon agricultural remnants remain as enclaves while ex-claves of the expanding city are found in the rural-urban interface area.

CLASSIFICATION PROBLEMS

A key question to be examined at this stage in the experimentation with mapping urban land uses by automatic digital processing of satellite imagery is how reliably functional land-use classes can be derived from spectral data. The approach to an answer requires exploring in depth the spectral characteristics of urban features. Considering that many components of the urban scene are smaller than the 80 meter square pixel and that great spectral diversity occurs from place to place within an urban area, an examination of the structural components of each urban spectral signature is necessary. For example, residential land use is composed of such spectrally diverse features as asphalt streets, concrete drives and patios, shake roofs, varying levels of maturity of landscaping, "corner" grocery stores, churches, and schools. In addition, these vary regionally with different environmental conditions and local varieties of building and paving materials.

In another example, commercial-industrial use, following its general tendency to be conducted in a specialized urban environment, yields a unique spectral signature because it is nearly always found in buildings or clusters of buildings with flat roofs with sizable parking and storage areas adjacent. The surfaces are spectrally quite distinct from a residential area with its pitched roofs, landscaping, and a full network of accompanying streets.

Where a confusion of symbols is seen on the print-out map, ground-truth examination reveals that such areas are indeed quite diverse and present classification problems even to the land-use mapper on the ground. Cases of incorrect identification are often simply cases of recent or on-going construction, areas which have not had sufficient time to weather into a more typical spectral signature.

A number of identification problems are common to all of the classes involving man-made cover of the land. Of greatest concern is that functional use is not consistently reflected in building shape (seen in photographs) or spectral characteristics (recorded by the scanner). The user of urban land-use data has a real need to distinguish between such diverse functions as retail, education, wholesaling, and transportation. When all these functions are found in spectrally similar settings, discrimination using spectral information alone is impossible. Attempts to determine reliable signatures for commercial versus industrial have proven inconclusive: while commercial establishments, such as along arterials, and in shopping centers, usually have asphalt roof surfaces, industries exhibit bright and dull surfaces in about equal quantity.

An added problem is the differential weathering of all types of man-made surface materials. Old and new paving and roofing materials are spectrally distinct enough for the computer to classify them differently even though functional land use is the same throughout. The ability of the computer to make this distinction suggests a potential to differentiate newer from older developments, but this potential has not yet been exploited.

EXPERIMENT RESULTS

The experiment has so far produced several computer maps of varying scales. First, all of the major urban complexes of the San Francisco Bay Region (on the
single ERTS-1 frame employed) not covered by cloud have been mapped at 1:24,000. All of these lie within the segment of the frame corrected for skewness and scale and even though further adjustments are required, the registration of the computer map to the 7¼ minute U.S.G.S. quadrangle map is remarkably accurate with perfect registration falling off only some one to two pixels over a distance of about 16 kilometers. Overlaying the computer maps over the quadrangle maps on a light table shows an immediate high correspondence with the quadrangle's limited land-use information. A temporary limit has been reached with the eight urban and three rural land-use classes selected. Isolating additional classes by determining the best sub-groupings from within those now comprising several spectral classes must await the application of new experimental techniques.

For all of these maps (plus the 1:48,000 one-fourth sample), it has been necessary to separate rural from urban uses in order to overcome the persistency of the classifier to throw into the same class certain rural (usually agricultural) land uses with certain urban ones (see Figure 1). A common confusion is caused when both urban residential land and cropland occupy the same spectral class. It is a matter of coincidence that the combination of ambient soil cover -- especially if somewhat moist -- and an immature crop, is spectrally similar (within a total of twenty-eight classes) to a single-family residential area with its asphalt streets, dark shake roofs and a given amount of landscaping. Fellow experimenters at LARS, Houston, and the University of Illinois have encountered the same problem.

The method employed to date to by-pass the problem requires the preliminary step of delimiting the urbanized area; subsequent groupings of urban land uses lie inside the boundary line and rural groupings are outside the line. A procedure was followed (Peruzzi, 1973) which adapted the Census Bureau's rules for Urbanized Area delimitation to a one-quarter kilometer grid system. The example (in Figure 1) uses a one kilometer cell generalization. Kilometer squares were given UTM addresses and the corresponding computer coordinates were then entered into the computer and delimited on the print-out. Two separate groupings of the twenty-eight cluster classes, one urban one rural, were then printed out. Manual cut-and-paste techniques were then used to make a single map; these functions will be performed by the computer in future work. The one kilometer unit of generalization is also a little too gross; refinement to the quarter kilometer will further improve the product.

The introduction of the kilometer grids into the system also provided a basis for aggregating land uses by a standard areal unit. Table 3 illustrates for a typical few kilometers and for the average of 250 square kilometers the percentages of each land use for the area around San Jose. The figure of 62.0 percent for residential uses compares favorably with the 63.4 percent for the same area as measured by planimeter from air-photo interpreted uses in the work of the Geographic Applications Program's Census Cities Project.

Knowing the reliability of automatic machine classifying is of greatest importance at this early stage of the work. Testing is hindered, however, since a precise definition of reliability and the development of a method of measurement are not universally agreed upon. Others have measured accuracy quite generally on an area basis and against the ground-truth of published land-use maps. In this experiment, accuracy of classification figures are based on a comparison of the classified individual pixels to contemporaneous air photographs. The procedure employed was tedious and time-consuming but simple in its method. Upon superimposing the photo over the computer map on the Zoom Transfer Scope, the question was asked for each pixel (in a 13 square kilometer sample) if the real land use matched that given by the classifier. Score was kept and the results presented in Table 4.

Reasons for mis-classification were readily recognized in the process of checking. First, the grossness of the 80 meter resolution elements vis-a-vis the size of urban features causes considerable unavoidable error. A common occurrence is where a row of symbols of an adjacent use covers a linear feature such as a highway arterial. Being linear phenomena highways appear to the eye viewing a photograph to be wider than they actually are; a four-lane highway has only some 20 meters of asphalt or concrete surface. Part of a roadway's signature is median strip and shoulder and in cities a commercial arterial is visible on a constructed ERTS-1 image only because of the distinctive reflectivity of the
flat-roofed buildings facing it.

Other frequent mistakes include confusion between the bright dried grass of a vacant lot and a bright factory roof or the unusually heavy tree canopy of an older, well-developed residential neighborhood and a naturally wooded open space area.

It is possible to view in Figures 2, 3, 4, and 5 the degree of visual correspondence between computer map and photograph. The general land-use patterns of Hayward are demonstrated in the first two illustrations while in Figures 4 and 5 it is possible to see on a pixel basis just where classification is correct and where it fails. In reliability checking with the Zoom Transfer Scope, where the view is of a similar scale, the operator begins to "think" like the computer classifier. More of this intensive work should lead to classification refinement.

Figure 6 presents the one-fourth sampling computer land-use map along with the AVV image and a point-line identification map. In addition to general physical features visible on the map, many broad urban patterns are also recognizable such as commercial land uses along arterials, central business districts, commercial-industrial concentrations, residential areas, and open space areas. Urban land uses outside the delimited urbanized area, in accordance with the simplistic classification employed, render an adequate representation of the grass and tree landscape of the non-urbanized hills and valleys lying between San Francisco Bay and the Central Valley. To date, attention in this experiment has been directed mainly at solving urban use classification problems. Classification of rural land uses and land cover will be attempted later.

CONCLUSIONS

Results of the experiment to date demonstrate that producing land-use maps of a large scale by machine processing of ERTS-1 scanner data is feasible. By keeping land-use classes fairly broad, a remarkable level of accuracy is attained despite the relatively coarse resolution and the inherent complexities of man-made land cover.

Much follow-on work is required. One constant challenge lies in improving the reliability of land-use recognition and classification. A finer sorting of the spectral information is one road to follow. Another would be the development of algorithms using context to solve certain classification problems. These techniques--somewhat analogous to identification by association procedures used in photo interpretation--are required for the computer to make distinctions between such functionally different but spectrally similar land uses as factories and shopping centers.

A second major thrust must be made toward developing and fitting optimal classification schemes to the capabilities of scanner-produced data from Skylab, ERTS-B, and aircraft from varying altitudes. The larger map scale (and the smaller area of each resolution element) of aircraft-borne scanners may increase the reliability of use identification and lend insight to work with ERTS-1 data. The inclusion of a thermal band, as planned for ERTS-B, would add a useful variable.

An additional effort is required to utilize computer graphic techniques to improve the visual quality of the land-use maps to promote their greater usage. A tailoring of products could be achieved to accommodate the entire range of users from local to national.

Data aggregation could be expanded from the kilometer squares in this paper to include census tracts, corporate units, transportation zones, and any other areal groupings desired. The ability to aggregate land uses by such areas and to monitor change with great frequency holds enormous promise for such valuable measurements as intercensal population estimates.

Essential too is the study of a temporal series of ERTS-1 passes for the purpose of monitoring and detecting change. Many of the chores presently done by hand lend themselves to machine processing. Important products would be precise measurement of incremental growth of subdivision housing, commercial, and trans-
portation uses. Summary statements of the type of change from one use to another could be facilitated by machine processing. A further advantage would be the ability to detect the fairly small-area changes within the body of the old city at the large map scale (1:24,000) of the computer print-out.

The solution of several other practical problems could be furthered by the use of a scanner/computer-analysis system. A relatively simple task would be the frequent updating of the boundary of a city's urbanized area. Commercial applications are also possible in such common jobs as selecting optimum locations for stores, banks, and service stations. The location and measurement of open space, a matter of key general concern, would also be easily handled by such a system. Another practical problem which could be dealt with is the required measurement of land use, present and projected, as a basis for mass-transit planning. Also, careful work may yield a method for measuring housing quality.

In sum, the advantages of speed, data handling, relative low cost, and frequent synoptic monitoring could be of extreme value in helping to solve many land-use problems.

ACKNOWLEDGEMENTS

The authors wish to acknowledge significant contributions to this work by Mr. Leonard Gaydos, California State University at San Jose.

USGS participation in publication approved by Director, U.S. Geological Survey.
REFERENCES


Duilio Peruzzi, "Urbanized Area Delimitation from High-Altitude Aircraft Imagery by Quarter Kilometer Grid," paper delivered at Standardized Land Use Classification System Workshop # 3, South San Francisco, California, June 7-8, 1973.


Table 1. Functional Land-Use Classes Employed on Computer Maps

<table>
<thead>
<tr>
<th>URBAN</th>
<th>Spectral Classes Comprised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional Land-Use</td>
<td></td>
</tr>
<tr>
<td>Commercial-Industrial</td>
<td>1, 2, 3, 14</td>
</tr>
<tr>
<td>Mobile Homes</td>
<td>5</td>
</tr>
<tr>
<td>Residential</td>
<td>6, 9, 10, 13, 15, 16, 17, 18, 19, 20, 21</td>
</tr>
<tr>
<td>Parking Lots</td>
<td>8, 22</td>
</tr>
<tr>
<td>Unimproved Open Space (bare)</td>
<td>11</td>
</tr>
<tr>
<td>Improved Open Space (irrigated)</td>
<td>12</td>
</tr>
<tr>
<td>Unimproved Open Space (with trees)</td>
<td>23, 24, 25, 26, 28, 29, 30</td>
</tr>
<tr>
<td>Water</td>
<td>27</td>
</tr>
<tr>
<td>RURAL</td>
<td></td>
</tr>
<tr>
<td>Grazing and Cropland</td>
<td>1, 2, 3, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23</td>
</tr>
<tr>
<td>Tree Covered</td>
<td>24, 25, 26, 28, 29, 30</td>
</tr>
<tr>
<td>Water</td>
<td>27</td>
</tr>
</tbody>
</table>

Table 2. Comparison of a Land-Use Classification Derived from Automatic Machine Processing of ERTS-1 MSS Data with a U.S.G.S. Proposed Land-Use Classification System for Use with Remote Sensor Data

<table>
<thead>
<tr>
<th>Machine-processed ERTS-1 Data</th>
<th>U.S.G.S. Proposed System</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level I</td>
</tr>
<tr>
<td>URBAN</td>
<td></td>
</tr>
<tr>
<td>Commercial-Industrial</td>
<td>01. Urban and Built-up Land</td>
</tr>
<tr>
<td>Mobile Homes</td>
<td>01. Urban and Built-up Land</td>
</tr>
<tr>
<td>Residential</td>
<td>01. Urban and Built-up Land</td>
</tr>
<tr>
<td>Parking Lots</td>
<td>01. Urban and Built-up Land</td>
</tr>
<tr>
<td>Unimproved Open Space (bare)</td>
<td>01. Urban and Built-up Land</td>
</tr>
<tr>
<td>Improved Open Space (irrigated)</td>
<td>01. Urban and Built-up Land</td>
</tr>
<tr>
<td>Unimproved Open Space (with trees)</td>
<td>03. Rangeland</td>
</tr>
<tr>
<td>Water</td>
<td>05. Water</td>
</tr>
<tr>
<td>RURAL</td>
<td></td>
</tr>
<tr>
<td>Grazing and Cropland</td>
<td>02. Agricultural Land</td>
</tr>
<tr>
<td></td>
<td>03. Rangeland</td>
</tr>
<tr>
<td></td>
<td>06. Nonforest Wetland</td>
</tr>
<tr>
<td></td>
<td>07. Barren Land</td>
</tr>
<tr>
<td>Tree Covered</td>
<td>04. Forest Land</td>
</tr>
<tr>
<td>Water</td>
<td>03. Rangeland</td>
</tr>
<tr>
<td></td>
<td>05. Water</td>
</tr>
</tbody>
</table>

2A-15
Table 3. Land-Use Aggregations by Kilometer Squares for a Segment of the San Jose, California Area

<table>
<thead>
<tr>
<th>UTI Grid Designation</th>
<th>Comm-Indus</th>
<th>Mobile Parking</th>
<th>Homes</th>
<th>Lots</th>
<th>Res</th>
<th>Bare</th>
<th>Trees</th>
<th>Irrig</th>
<th>Water</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>135-603</td>
<td>32.5</td>
<td>1.7</td>
<td>1.7</td>
<td>59.0</td>
<td>4.3</td>
<td>0.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>134-590</td>
<td>12.1</td>
<td>1.3</td>
<td>11.2</td>
<td>72.3</td>
<td>0.4</td>
<td>1.8</td>
<td>0.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>134-591</td>
<td>9.8</td>
<td>0.0</td>
<td>14.7</td>
<td>72.3</td>
<td>0.0</td>
<td>0.9</td>
<td>1.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>134-592</td>
<td>13.3</td>
<td>0.5</td>
<td>9.0</td>
<td>75.7</td>
<td>0.0</td>
<td>0.0</td>
<td>1.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.9</td>
</tr>
<tr>
<td>134-593</td>
<td>18.8</td>
<td>0.4</td>
<td>14.7</td>
<td>62.1</td>
<td>0.0</td>
<td>3.1</td>
<td>0.4</td>
<td>0.0</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>134-594</td>
<td>32.1</td>
<td>6.7</td>
<td>33.5</td>
<td>13.8</td>
<td>0.0</td>
<td>1.3</td>
<td>11.6</td>
<td>0.0</td>
<td>0.0</td>
<td>0.9</td>
</tr>
<tr>
<td>134-595</td>
<td>42.9</td>
<td>4.5</td>
<td>18.5</td>
<td>27.7</td>
<td>3.6</td>
<td>0.0</td>
<td>1.8</td>
<td>0.0</td>
<td>0.0</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Average for 250 square kilometers: 12.9 1.3 8.2 62.0 4.5 6.2 2.2 0.2 2.5

Table 4. Reliability Test of Land-Use Classification

<table>
<thead>
<tr>
<th>Functional Land-Use Class</th>
<th>Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial-Industrial</td>
<td>62.7</td>
</tr>
<tr>
<td>Residential</td>
<td>84.6</td>
</tr>
<tr>
<td>Parking Lots</td>
<td>77.8</td>
</tr>
<tr>
<td>Unimproved Open Space (bare)</td>
<td>94.2</td>
</tr>
<tr>
<td>Improved Open Space (irrigated)</td>
<td>97.1</td>
</tr>
</tbody>
</table>
Figure 1. Computer-Classified Land-use Map of the Fremont Area, California. Map is derived from scanner digital tapes of ERTS-1 scene 1003-18175, 25 July 1972. It demonstrates use of separate classifications for Urban and Nonurban. Classification uses LARSYS pattern recognition algorithms, and was produced at Purdue University, Laboratory for Applications of Remote Sensing (LARS). Urban area is defined by one-kilometer UTM grid cell (zone 10) from USGS Census Cities ERTS experiment 1970 land-use map and NASA aircraft photography. Land use areas are aggregated by class and kilometer grid cell. Each pixel represents 0.465 hectares (1.1 acre). Urban classes: Commercial-Industrial (I); Mobile Homes (V); Residential (M); Parking Lots (-); Unimproved Open Space, Bare (-); Unimproved Open Space, Trees (H); Improved Open Space-Irrigated (+); Water (O). Nonurban: Grazing and Cropland (+); Trees (K); Water (O). Large unclassified areas (blank) are salt evaporation ponds.
Figure 2. Black-and-white Photo of the Hayward Area, California. Photo by NASA for USGS Census Cities ERTS experiment: Zeiss camera, 12 in. focal length; 1:50,000 from RB-57 at 50,000 feet; color infrared film (S-117), filter B; frame 252, 15 May 1970. Approximate orientation of one-kilometer UTM grid (zone 10) indicates scale, and facilitates comparison with corresponding computer land-use map.
Figure 3. Computer-classified land-use of the Hayward Area, California. Map is from ERTS-1 scanner digital data for frame 1003-18175, 26 July 1972. Urban classes: Commercial-Industrial (I); Mobile Homes (V); Residential (M); Parking Lots (-); Unimproved open Space, Bare (-); Unimproved Open Space, Trees (/); Open Space-Irrigated (*); Water (O). Approximate orientation of one-kilometer UTM grid (zone 10) indicates scale and direction, and facilitates comparison with corresponding air photo and other ground truth.
Figure 4 and Figure 5. Enlarged Photo-map Pair, Hayward Area, California. Photo is part of one high altitude color infrared scene acquired at 1:50,000, 15 May 1970, by NASA for USGS Census Cities ERTS experiment. Map is by Purdue/LARS and is based on ERTS-1 scanner digital data for frame 1003-18175, 26 July 1972. Scan lines have been reformatted so that they are nearly parallel to east-west UTM one-kilometer grid lines (zone 10). One printout symbol represents one ERTS scanner picture element (pixel), 0.465 hectare, or 1.1 acres. Urban classes: Commercial-Industrial (C); Mobile Homes (V); Residential (R); Parking Lots (L). Note correspondence between photo and map in mobile home area, lower left corner. Expansion of mobile home park (symbol V on ERTS computer map) is confirmed by NASA 1972 air photo.
Figure 6a. Portion of ERTS-1 RBV View, San Francisco Bay Region. View is from Band 3 (red), frame 1003-18175, 26 July 1972, three days after launch. This is same view for which digital data from the four-channel multispectral scanner (MSS) are used to classify land use by computer-aided techniques (Figure 6b). Twenty-kilometer UTM grid (zone 10) and thirty-minute geographic grid fitted by U.S. Geological Survey. Position of grids meet Federal map accuracy standards. Fog and San Francisco Bay at left (west); San Joaquin Valley and large irrigated fields at right; Hayward and Freemont areas at southeast side of San Francisco Bay.
Figure 6b. Portion of computer-classified land use map of San Francisco Bay Region. Map is derived from ERTS-1 scanner digital data, frame 10C3-16175, 26 July 1972. Classification by Purdue/LARS, uses eight Urban classes, and three Nonurban classes (Figures 4 and 5). Urban area, defined by one-kilometer UTM grid (Zone 10), is from USGS Census Cities ERTS experiment and NASA aircraft photography. The grid facilitates comparison with corresponding ERTS-1 RGB scene (Figure 6a). Map is produced at 1:48,000 by classifying every other pixel in every other scan line. About 6,500 square kilometers (or just under one-tenth of one percent of U.S. land area) were classified on LARS IBM 360-67 in about thirty minutes' computer time. Aggregation of areas by land use class and kilometer grid square can also be generated. It may soon be operationally and economically feasible to compile manuscript land use maps for large areas by this method, using additional Nonurban classes. Then edit and adapt to more conventional functional classes. Perhaps, draw use boundaries by conventional cartography or computer graphic methods, and publish maps at 1:50,000 to 1:250,000. Area measurement and land use change data by grid cell, or user jurisdiction area, would be valuable by-products. (-Ellefson, Swain, and Wray, Figure 6b)