NASA TECHNICAL
MEMORANDUM

NASA TM X-64793

A BRIEF DESCRIPTION OF AN EARTH RESOURCES TECHNOLOGY SATELLITE (ERTS) COMPUTER DATA ANALYSIS AND MANAGEMENT PROGRAM

By Robert R. Jayroe, Jr.
Aero–Astrodynamics Laboratory

December 6, 1973

NASA

George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama

TABLE OF CONTENTS

PRECEDING PAGE BLANK NOT FILMED.

TECHNICAL MEMORANDUM X-64793

A BRIEF DESCRIPTION OF AN EARTH RESOURCES TECHNOLOGY SATELLITE (ERTS) COMPUTER DATA ANALYSIS AND MANAGEMENT PROGRAM

INTRODUCTION

To make use of the large amounts of Earth Resources Technology Satellite (ERTS) data that are available over any particular test site, some systematic method of handling and analyzing the data must be developed. This report describes a systematic method presently being used at NASA, George C. Marshall Space Flight Center (MSFC), Huntsville, Alabama, and indicates future expansion capabilities of the analysis system.

The method of analysis initially involves acquiring digital tapes and ERTS imagery from NASA, Goddard Space Flight Center, Greenbelt, Maryland, and also any supporting aircraft imagery that is available from other possible sources. The main bulk of the analysis is performed by computer on the digital tapes with the available imagery being used to support the interpretation of the digital results. Because of the large geographic areas contained in the data, the computer programs are designed to operate on the data without prior knowledge of ground truth. The results of the computer analysis are then used to determine where ground truth information should be collected. At present, the output of the computer analysis is limited to ordinary computer printout, a Stromberg-Carlson-4020 microfilm plot frame, and Xerox copy flow. However, color display and interactive processing capabilities will be available in the very near future.

The mathematical concepts used in the computer programs were developed by or under the direction of members of the Flight Data Statistics Office, Aerospace Environment Division, Aero-Astrodynamics Laboratory at MSFC, and the operation and programming support is provided by the Data Reduction Branch, Engineering Computation Division, Computation Laboratory. Support for the color display device was obtained from the Environmental Applications Office at MSFC, and the system is to be maintained and operated by the Data Reduction Branch of the Computation Laboratory.

DATA DESCRIPTION

The ERTS provides four images of the same ground scene that cover an area of 34,299 km² (13,243 sq mi). These four images are recorded with spectroscopic bandwidths that range from 0.5 to 0.6, 0.6 to 0.7, 0.7 to 0.8 and 0.8 to 1.1 μm, and the images are in the form of a square with a side corresponding to 185.2 km (115 mi) on the ground. The digital data associated with each image contain 3240 samples per scan (in the equatorial direction on the image) and 2340 scans of data (in the polar direction of the image). Thus, each spectroscopic bandwidth or channel of data contains a total of 7,581,600 data points. In a crude sense, the resolution of each data point could be represented by a rectangular area whose ground scene dimensions are 79.15 m (259.66 ft) in the polar direction and 57.16 m (187.53 ft) in the equatorial direction. Since the four digital images are analyzed simultaneously, each position on the ground is represented by four data points, and therefore each set of four data points is treated as a four-dimensional vector. Thus, a total of 30,326,400 data points are associated with a particular ground scene for a corresponding set of multispectral ERTS images.

To obtain complete coverage of Alabama, for example, it is necessary to obtain the multispectral data from approximately 14 different ground scene images. However, because of partial overlap of the different images, the amount of data can be reduced to roughly seven to nine equivalent images. In most cases, it is also desirable to obtain ERTS imagery as a function of time. This can easily be accomplished since the satellite repeats its orbit every 18 days. When temporal aspects are added to the multispectral information, the result is a data volume which truly merits some management consideration.

To gain experience, encounter some of the problems that will arise, and be in a position to make recommendations for expanding the computer analysis area coverage, a test site in North Central Alabama is presently being analyzed that comprises 42 percent of an ERTS image. The following sections describe the analysis procedures used, existing processing capabilities, planned future processing capabilities, and comments being derived from the analysis of the test site.

RAW DATA HANDLING AND OPERATIONAL PROCEDURES

The digital data associated with each ERTS ground scene image
is contained on four separate tapes. Each tape contains four channels
of data, 810 samples/scan/channel and 2340 scans of data. Because
of computer storage requirements, each tape is analyzed as four
separate strips containing 202 samples/scan/channel and 2340 scans,
except for the last strip, which contains 204 samples/scan/channel.
Thus an entire ERTS image is analyzed in 16 separate strips.

The computer programs for analyzing the data are written in
FORTRAN IV, the dimensioned variables are limited to a total of 20K,
and the programs are run in a 32-K environment on an IBM-7094
computer. For production running, frozen versions of the computer
programs are contained on fast tape, while experimental and research
versions of the programs are maintained as computer card decks to
allow for ease in logic changing. Because of the large amounts of
data involved, there is a constant effort to reduce the running times
of the programs. The programs, however, are almost to the point
where significant additional time reduction can be accomplished only
by going to a special-purpose computer system or a faster computer.

To communicate the desired analysis for the data, run submission
forms were designed which permitted a wide variety of choice for the
data analysis options. The user provides the necessary information
to the Computation Laboratory by putting the appropriate information
on the run submission forms. All of the programs have the ability for
selecting subportions of a tape for analysis, and in certain programs
it is possible to transfer information between tapes contained within the
same or a different data set.

PRESCREENING, EXAMINATION OF DATA ANOMALIES,
AND SPECIAL DISPLAY ROUTINES

Normally, the types of programs listed in this section heading
are not used unless there is some problem or irregularity associated
with the data or unless there is a special requirement for the above type
of analysis. Typically, these programs operate only on one channel
of data at a time, and the following types of analyses are available:

1. Compute Histogram of Data.

2. Display Gray Level of Data.

3. Contour Data.

4. Isometric Display of Data (two-dimensional projection of a three-dimensional plot of data amplitude versus sample number per scan and scan line number).

An additional program is available that computes the joint histogram between any two channels of data. For additional information concerning these programs, see the author's memorandum[1] and Reference 1.

CLASSIFICATION ANALYSIS PROGRAMS

The purpose of the classification programs is to produce a map of the ground scene from the digital data showing the location and area coverage of all features that have been categorized according to the spectroscopic information derived from the multispectral data. The features, for example, may be various types of crops, forested areas, waterways, etc. From such maps it is then possible to perform inventories for determining how the land and resources are being used and possibly to locate new resources or obtain new geologic information.

At present, two classification programs are being evaluated toward their application of producing accurate feature maps. Both programs have the same end product but different initial approaches.

One of the programs is called the Sequential Clustering Program (SCP) and obtains an initial estimate of the channel mean values and variances of the candidate multispectral features in the data by simultaneously considering several consecutive (typically six to ten) data elements within a scan. If, according to a statistical test, the data elements are similar, then a population is initiated containing those data elements. By population, it is meant that the channel means and variances are computed for those data samples. If the next consecutive data sample fits into that population, according to a statistical test,

1. R. R. Jayroe, Automated Computer Programs Description for Analysis of Earth Observation Data, Office Memorandum, S&E-AERO-YF-1-73, June 21, 1973.

then the population channel means and variances are updated with that data sample. If that new data sample did not fit in the population, then an attempt would be made to establish a new population by finding several consecutive data samples that could establish a new different population. This is the manner in which candidate multispectral features are established, and each new data sample is examined to determine whether it belongs to a previously established population or whether an attempt should be made to establish an additional new population.

As the samples are placed into populations, a map of the data is produced which labels each data sample with the corresp...ding population identification. The program allows for a maximum of 30 candidate features or populations, and, if the maximum is exceeded, then two populations which are statistically the most similar are merged together. This procedure of merging is repeated as often as necessary until all of the desired data have been examined. The program then uses an iterative scheme involving the "K-means" algorithm to improve the accuracy of the channel means and variances for each population. The iterative scheme is accomplished by making additional classification passes through the data set and recomputing the channel means and variances for each population.

The end product of the analysis is a feature map of the data with the improved statistics for each population or feature. The program has several input parameters which allow for considerable flexibility in controlling the number of desired candidate features, the establishment of new populations, the merging of populations, and the number of desired iterative passes through the data. For detailed information concerning the input parameters and statistical tests used in the program, see Reference 2. Several improvements for the statistical tests used in the SCP have been reported in Reference 3, and these new tests are currently being programmed for evaluation.

The second program being used is called the Spatial and Spectral Clustering Program (SSCP). The first step in the data analysis is to locate areas of homogeneity within the ground scene coordinates, which will later be identified as belonging to a feature. The location of these areas is determined by producing a boundary map of the ground scene, which utilizes the comparison of feature vectors in n-dimensional space of each ground scene coordinate with the feature

5

vectors of neighboring ground scene coordinates. If the distance in the n-dimensional space between the feature vectors of neighboring ground scene coordinates is large enough, then a boundary coordinate is detected and indicated on the map.

The second step is to fetch and surround the n-dimensional data, contained within each homogeneous area, with an n-dimensional surface. The data from each area is fetched by using a square ground scene spatial data array. This fixed shape array is small enough to fit into most of the homogeneous areas but large enough so that it cannot pass through any holes that might exist in the boundary surrounding a homogeneous area. The boundary map is overlaid on the raw data tape, and the data array is allowed to move within a homogeneous area on the boundary map and fetch data from the raw data tape, but it is not allowed to move on coordinates occupied by a boundary. When no more data can be gathered from a homogeneous area, the array is moved to another area and the process is repeated. To surround the data from each homogeneous area with an n-dimensional surface, the general location of the data in n-dimensional space must be determined. This is accomplished by computing the average or mean value of all the n-dimensional vectors contained within a homogeneous area. Thus, a first-order statistic is used for establishing the location of the n-dimensional surface. Second-order statistics are used to calculate the equation of the surface. These statistics take the form of terms such as x^2, y^2, and xy and are the variances of each channel of data and the covariances between all channels of data for a particular homogeneous area. The form of the terms for the second-order statistics is highly suggestive of the equation of an ellipse, and therefore, the data from each homogeneous area are surrounded with an n-dimensional hyperellipse.

The third step is to decide which homogeneous areas are spectrally similar and which are spectrally different by examining the n-dimensional hyperellipses for two homogeneous areas. If the centroids of the two hyperellipses are contained in both ellipses, then there is sufficient overlap between the ellipses, and the two homogeneous areas are said to represent the same ground scene feature. The statistics from these two areas are combined, and a new n-dimensional hyperellipse is calculated to surround, in n-dimensional space, the data contained in both homogeneous areas. This process is repeated as long as there are hyperellipses from different homogeneous areas that overlap sufficiently.

6

After this process is completed, m different nonoverlapping hyper-
ellipses will remain, and the information associated with these
hyperellipses will represent m different spectral features detected
in the ground scene.

The final step is to check the end point of every feature vector
contained in the ground scene image for determining whether it is
contained in one of the m hyperellipses. From this information, a
map of the ground scene is produced showing the location of data
that were contained in each of the m hyperellipses.

The mathematical rationale for SSCP can be found in Reference 4,
and the details of the program are presented in Reference 1. There
are four main decisions used to control the entire classification
program. First, the boundary map must be examined to insure that
there are enough boundaries on the map so that data representing
different features will not be mixed during spatial clustering. This
examination also helps in deciding the proper size to use in the
spatial clustering data array. Third, the value of the spectral
merging parameter must be set so that the data from significantly
different features are not spectrally merged together. Finally, the
value of the parameter used in classifying individual data vectors
must be set large enough so that the data elements originally selected
as belonging to a particular feature are classified as belonging to
that feature, but it must also be set small enough so that there is a
minimum of misclassification of individual data elements.

In summary, the end products of both programs are maps and
tables which indicate the following information:

1.   The homogeneity and patterns of terrain features.

2.   The number of spectrally distinct features with their respective
mean spectral signatures and variances.

3.   The areal extent, location, and distribution of the feature within the
ground scene.

4.   The quantity of ground truth needed and direction for ground truth
patrols.

The main information not provided is the identification of the different
spectral features. Some identification can be inferred from aerial
photography, but, in general, a certain amount of ground truth
information must be collected by an observer. However, the amount

7

of ground truth required can be significantly reduced, in terms of amount and cost, by applying the analysis scheme previously described. The collection of ground truth can be optimized by directing ground truth patrols to a few areas that contain a maximum number of unidentified features on the computer map.

CHANGE DETECTION AND OTHER PROGRAM OPTIONS

As mentioned before, the ERTS data tapes are divided into four strips for analysis. The usual procedure is to analyze one strip of data and identify the features that are represented on the classification map. The statistics describing those features are output on a saved tape, which can be used to classify the remaining strips of data without having to go through the spatial clustering and merging routines; i.e., it is necessary only to run the classification program using the already available statistics for classifying the data. In some cases, data will be encountered which represent features that are not described by the available statistics. In this case, the existing classification map can be used, rather than a boundary map, as an input to the spatial clustering and merging program to pick up the new features. The statistics tape can then be updated to include the new features or classes and the analysis can continue on the rest of the data. The end result is that a complete set of statistics can be obtained to classify nearly all the data as belonging to one of the previously obtained features. A Manual Selection of Clusters Program is also available for manually selecting homogeneous areas from the boundary map to be used as candidate features. Rather than classifying the entire data set, this program provides the option of selecting a few features and specific areas for classification, and, if desired, provides a means of checking the Spatial Clustering and Spectral Merging programs.

For displaying the results of the classification maps, there are several smaller programs which are of considerable use. For example, consider a small mountain covered with one species of trees. There will be typically three classes representing trees whose differences are due to lighting conditions and terrain slope: one class of trees on the shady slope of the mountain, a different class of tree for the sunny slope of the mountain, and if a plateau is present on the mountain, a third class for trees is possible. Usually one is interested only in the presence of trees rather than the lighting conditions, and therefore it would be desirable to represent all three classes as one class. This

8

can be done with the use of a character enhancement program which allows
a user to assign the available single computer characters in any manner
to any of the classes. Since only black and white output is available at
this time, the program also allows one to choose characters so that some
contrast can be obtained in the map for ease in locating different features.
When it is desirable to merge two or more classes by representing them
with the same computer symbols, two other programs are used in provid-
ing information for deciding which classes can be merged. One program
plots the mean vectors for each class on a graph whose ordinate and
abscissa are any two combinations of different channel amplitudes for
visual inspection, and the other program lists in order of closeness,
vector distance wise, pairs of class mean vectors. Thus, these two
programs provide some rationale for manually merging different
classes after the original analysis has been performed.

For change detection applications, several possibilities exist.
First of all, and without having to register two data sets acquired from
the same ground scene but at different times, the classification
statistics acquired from one data set can be used to classify the other
data set, and the two classification maps can be visually inspected to
determine what changes have occurred. However, if the two data sets
can be registered, then a wide variety of options is available.

At present, a manual registration program is being used to register
different data sets. This program requires a user to select correspond-
ing boundary elements from the two data sets and to enter their scan and
column coordinates into the registration program, which then translates
and/or rotates the data sets according to the coordinate inputs. An
Automatic Registration Routine is currently being programmed which
takes advantage of the fact that the boundary map only contains -1's
and 0's, representing boundary and homogeneous data elements,
respectively. A correlation routine is used, but with binary numbers
it is possible to replace multiplication with integer addition, and the
routine computes only those correlation lag points that are effected by
the overlay of two boundary maps. On the average, the binary correlation
routine is 35 times faster than ordinary correlation routines and 7 times
faster than fast Fourier transform correlation methods. The binary
correlation routine and registration program are discussed in a NASA
Technical Note.[2]

---

2. R. R. Jayroe, J. F. Andrus, and C. W. Campbell, Digital Image Regis-
   tration Method Based Upon Binary Boundary Maps, NASA Technical Note,
   to be published.

In using registered data, it would be possible, for example, to combine 4 sets of 4 channel ERTS data acquired from the same ground scene but at 4 different seasons to produce a 16-channel data set. The spectral signatures would contain not only spectral and spatial information but also temporal information. The classification program could then be run on the 16-channel data set. Another option would be to overlay the boundary map containing the spatial clusters from one data set on another data set for fetching the data from identical ground scene coordinates at a different season and comparing the change in signature. An additional option for comparing two identical registered channels of data acquired from the same ground scene but at different times would be to compute the joint histogram between the two channels of data. If no changes have taken place in the ground scene, then one would expect the joint histogram to be in the form of a straight line. Deviations from this linear distribution would therefore indicate changes that have taken place in the data. For checking the accuracy of the registration program one could also compute the joint histogram between two registered boundary maps, and, in terms of classification maps, one could also compute the joint histogram to determine how the classifications have changed. Thus, by using the joint histogram in conjunction with two data sets, it is possible to output on over-lay map of the two data sets showing the location of the changes that have taken place, as well as areas where no change has taken place. It would also be possible to indicate the degree of change in the data on such a map.

FUTURE CONSIDERATIONS

The future considerations are mainly concerned with the incorporation of a color display device. The display device is to be driven by a small existing computer system utilizing three tape drives, card punch and reader, and high-speed line printer. It is envisioned that at a later date, the small computer system will be interfaced with an existing larger system to provide an increase in flexibility and capability. For the immediate future, however, the display system will considerably enhance the visual information contained within the analysis maps, and, in conjunction with creative software programming, will provide for on-line quality control of analysis and interactive processing capabilities.

It appears that the software programs described within this report and the listed references provide a sufficient analysis base to expand the ERTS analysis to statewide and temporal coverage. However, before any recommendations can be made, it will be necessary to complete the analysis on the small test area so that reliable estimates can be made on resources, time-liness, and quality of analysis.

10

REFERENCES

1. Phillips, M. R.: Earth Resources Data Processor. NASA CR-61399, October 1972.

2. Su, M. Y.: The Composite Sequential Clustering Technique for Analysis of Multispectral Scanne Data. NASA CR-128999, October 1972.

3. Dalton, C. C.: Notes for the Improvement of a Remote Sensing Multispectral Data Non-Supervised Classification and Mapping Technique. NASA TM X-64762, July 1973.

4. Jayroe, R. R.: Unsupervised Spatial Clustering with Spectral Discrimination. NASA TN D-7312, May 1973.

11

APPROVAL

A BRIEF DESCRIPTION OF AN EARTH RESOURCES TECHNOLOGY SATELLITE (ERTS)COMPUTER D.ITA ANALYSIS AND MANAGEMENT PROGRAM

By Robert R. Jayroe, Jr.

The information in this report has been reviewed for security classification. Review of any information concerning Department of Defense or Atomic Energy Commission programs has been made by the MSFC Security Classification Officer. This report, in its entirety, has been determined to be unclassified.

This document has also been reviewed and approved for technical accuracy.

WILLIAM W. VAUGHAN
Chief, Aerospace Environment Division

E. D. GEISSLER
Director, Aero-Astrodynamics Laboratory