NASA TECHNICAL REPORT

NASA TR R-434

NASA TR R-434

A PROCEDURE FOR AUTOMATED LAND USE MAPPING USING REMOTELY SENSED MULTISPECTRAL SCANNER DATA

by Sidney L. Whitley

Lyndon B. Johnson Space Center
Houston, Texas 77058

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • JANUARY 1975

NASA TR R-434 JANUARY 1975

1. Report No. NASA TR R-434	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle A PROCEDURE FOR AUTOMATED LAND USE MAPPING USING REMOTELY SENSED MULTISPECTRAL SCANNER DATA		5. Report Date January 1975
		6. Performing Organization Code
7. Author(s) Sidney L. Whitley		8. Performing Organization Report No. JSC S-406
		10. Work Unit No. 177-89-89-00-72
9. Performing Organization Name and Address Lyndon B. Johnson Space Center Houston, Texas 77058		11. Contract or Grant No.
		13. Type of Report and Period Covered Technical Report
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

A system of processing remotely sensed multispectral scanner data by computer programs to produce color-coded land use maps for large areas is described. The procedure is explained, the software and the hardware are described, and an analogous example of the procedure is presented. Detailed descriptions of the multispectral scanners currently in use are provided together with a summary of the background of current land use mapping techniques. The data analysis system used in the procedure and the pattern recognition software used are functionally described. Current efforts by the NASA Earth Resources Laboratory to evaluate operationally a less complex and less costly system are discussed in a separate section.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
·Signature Analysis ·Spectral Signatures ·Pattern Recognition ·Earth Resources ·Remote Sensors ·Data Processing ·Data Acquisition ·Cartography	STAR Subject Category: 13.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	49	$3.75

*For sale by the National Technical Information Service, Springfield, Virginia 22151

CONTENTS

TABLES

FIGURES

Figure		Page

A PROCEDURE FOR AUTOMATED LAND USE MAPPING USING REMOTELY SENSED MULTISPECTRAL SCANNER DATA

By Sidney L. Whitley*
Lyndon B. Johnson Space Center

SUMMARY

The NASA Earth Resources Laboratory has developed a computerized system for automatically generating color-coded land use maps of very large areas through use of multispectral scanner data, sensed remotely from either spacecraft or aircraft platforms. The procedure, the software, and the hardware used in the system are described, and an analogous example of the procedure is included. The system incorporates ideas from past research efforts and is based primarily on a system developed at Purdue University that uses a Bayesian maximum-likelihood ratio calculation in spectral pattern recognition. The automated system enables governmental and industrial planners to produce land use maps faster than by conventional land use mapping techniques. The color-coded land use map produced is reasonably accurate for many applications but should not be considered a high-quality metric map as produced by Government mapping agencies. Although usable, the system is complex to implement and operate, and the Earth Resources Laboratory currently is involved in an operational evaluation of a portable image display system that is expected to be simpler and less costly.

INTRODUCTION

The objective of this report is to provide to potential users of multispectral scanner data an understanding of a procedure, and the associated hardware and software, for producing land use maps by automatic computerized systems. The procedure and system described in this report are currently used by the Earth Resources Laboratory (ERL) of the NASA Lyndon B. Johnson Space Center (JSC). This report also contains a summary of research and development work that is expected to simplify the hardware, the software, and the procedure for generation of land use maps.

A land use map is a scaled projection on which the uses being made of the Earth surface and the natural conditions of the surface are delineated as a selected set of categories. Land use maps are used primarily by governmental and industrial planners to provide control, to approve activities, to direct growth patterns, and so forth.

*National Space Technology Laboratories, Bay St. Louis, Mississippi 39520.

Conventional land use mapping techniques are very slow and costly. The usual technique is to analyze photography acquired by aircraft flying at a low level and to transfer the information manually to a scaled map. Vegas (ref. 1) has demonstrated the advantages of using photography acquired by high-altitude (18.3 kilometers (60 000 feet)) aircraft in reducing the efforts and costs associated with land use mapping. Vegas (ref. 2) has also demonstrated the advantage of using spacecraft-acquired multispectral scanner system (MSS) imagery in a similar manner to produce land use maps. Multispectral scanner system technology is advancing rapidly, and a significant amount of data is now available from scanner systems.

For several years, research has been conducted in the development of various phases of automated land use mapping techniques, and the technique development effort still continues. The ERL has selected an approach to land use mapping that uses various ideas from past research efforts. The approach was stabilized, and changes were made only to improve the flow of data or to make data manipulation easier. The land use mapping approach has incorporated a set of equipment and a procedure to establish a complete system for land use mapping. After the usefulness of the procedure was evaluated with a number of surveys, the procedure and system were documented.

Although multispectral scanner data processing techniques were developed at the University of Michigan, the University of Kansas, and Purdue University, the ERL selected the approach developed at Purdue University for application evaluation because the system can be implemented on digital computers, which are widely available throughout the United States. The procedure used by Purdue University is described in reference 3.

The land use mapping system currently used by the ERL is primarily an outgrowth of the basic system developed at Purdue University; the ERL equipment used for screening data has been improved, as has the Purdue University equipment, and the ERL classification scheme has been simplified and shortened. The land use mapping output devices have also been improved so that color-coded land use maps can be generated more efficiently.

Early land use mapping systems were designed primarily to process data from relatively small test sites, with emphasis on processing a set of data by using different classification schemes and controls. The ERL processing system and procedures have been modified to enable users to produce land use maps of very large regions. Full spatial resolution of the input data is preserved, and all input data may be used if desired.

A data analysis system (DAS) is used by the ERL to screen and reformat data from a variety of multispectral scanner systems, multiband photographic systems, and microwave imager systems and is a tool for performing research in data processing and analysis techniques. Although the ERL uses this system (for screening, preprocessing, training the computer for land use mapping, recording output map products, etc.), the DAS far exceeds the minimal requirements for an effective operational land use mapping system.

The procedure described in this report makes use of the ERL DAS, but a portable image display system (PIDS) that is being developed will adequately meet the requirements for image display, screening, training sample selection, and final display for land use mapping. The PIDS, described in detail later in this report, will be priced within the budget of most users.

In this report, the steps required to plan a land use mapping survey, to acquire data, to process the data, and to produce a color-coded land use map are presented. The procedure was developed with ERL equipment and facilities but could be used with any similar set of equipment. Functional descriptions of the software and hardware systems are provided in the appendixes, and more detailed descriptions, referenced throughout the report, are available from the ERL.

As an aid to the reader, where necessary the original units of measure have been converted to the equivalent value in the Système International d'Unités (SI). The SI units are written first, and the original units are written parenthetically thereafter.

The development of the procedure and hardware and the acquisition of equipment described in this report have taken place over the past 2 years. The basic software was provided to NASA by the Laboratory for Applications of Remote Sensing at Purdue University. The software conversion was accomplished by several programers of the Lockheed Electronics Company. Program CHOICE was developed by Clay Jones of the ERL. The author is grateful for the contributions made to this report by J. D. Derbonné, Clay Jones, and Thomas Pendleton of the ERL.

PROCEDURAL STEPS

The procedural steps leading to the production of a land use map from multispectral scanner data are presented in the following section. The steps are listed under three categories: task planning, data acquisition, and data processing.

Task Planning

Task planning includes defining potential application; developing survey requirements; selecting platform, sensors, and data bands; and requesting or acquiring data.

Define potential application. - The first step in the procedure for generation of land use maps from multispectral scanner data is to define the potential application. Because a significant amount of resources — manpower, computer time, and equipment — will be required to complete the task, the chosen application should be based on identified needs.

Develop survey requirements. - After the application has been chosen, a set of survey requirements should be developed. Because the resources required to generate land use maps are directly related to the size of the survey area, the survey area chosen should be no larger than the application requires. Survey areas are usually chosen for one or more of the following three reasons.

1. No baseline land use map exists for the survey area, and a land use map is needed.

2. The survey area is dynamic, and a determination of its status at a selected time is required.

3. The survey area is not particularly dynamic, but planned activities will greatly alter its characteristics, and a land use map is needed as a baseline before and after the alterations to determine impact.

In developing the survey requirements, the planner should define the classes of materials to be identified. (The terms "classes of materials" and "land use categories" are used interchangeably throughout this report.) For aircraft surveys, it is important to select the direction, the location, and the spacings of flight lines to be overflown.

The required end products should be defined during the task planning phase, and the following factors should be considered.

1. Products needed: An example of products needed might be color-coded land use maps with county lines, major roads, town names, and so forth inscribed.

2. Scales desired: For example, map scales of 1:62,500 or 1:24,000 might be chosen.

3. Resolution required: To inventory 1012-square-meter (0.25 acre) fields, greater spatial resolution is required than is needed to inventory 16-square-hectometer (40 acre) fields.

4. Repetitive coverage needed: Whether the land use map can be constructed from a single data collection pass or whether the survey area must be sensed at some repetitive interval to detect seasonal changes, manmade changes, and so forth must be determined.

5. Ground truth required: One requirement for ground truth is to provide the spectral pattern recognition programs with training samples on which the signature for each kind of material is based. The second requirement for ground truth is to evaluate the accuracy and quality of the land use map in the analysis phase.

6. Supporting sensors defined: Sensors other than the MSS needed to provide supporting information must be identified. Cameras are frequently used as supporting sensors on land use mapping surveys to aid in locating training samples. A camera provides greater spatial and geometric resolution than a scanner system.

7. Location and related data needed: Requirements should be defined (date, time of day, latitude and longitude, county, township, etc.), and acceptable limits of accuracy should be established.

Each of the preceding task requirements should be used in the preparation of a formalized survey plan.

Select platform, sensors, and data bands. - After the requirements have been defined, they should be evaluated against the available data collection platforms, sensors, and data bands. To produce land use maps by the procedure outlined in this report, the source of data must be some type of multispectral sensor (a sensor that filters the energy from the scene being viewed into several discrete wavelength bands and records the energy as several separate images). Several multispectral scanner systems currently in use are listed in table I and briefly described in the following text. A functional description of the systems is contained in appendix A.

TABLE I. - MULTISPECTRAL SCANNER SYSTEMS

Sensor	Platform	No. of data bands	Reference
ERTS-1 MSS[a]	ERTS-1 satellite (NASA)	4	4
EREP S192 MSS[b]	Skylab spacecraft (NASA)	13	5
MSDS[c]	C-130 aircraft (NASA)	24	6
Michigan scanner	Commercial aircraft	18	--
MMS[d]	Commercial aircraft	11	7
DS-1250 multispectral scanner	Commercial aircraft	11	8

[a]Earth Resources Technology Satellite 1 multispectral scanner system.

[b]Earth resources experiments package multispectral scanner system (on board the currently inactive Skylab spacecraft).

[c]Multispectral scanner and data system.

[d]Modular multispectral scanner.

The Earth Resources Technology Satellite 1 (ERTS-1) MSS is currently in a near-polar Earth orbit at an altitude of approximately 926 kilometers (500 nautical miles) and covers the same location on the surface of the Earth every 18 days. The ERTS Project is a research and development project with approved investigators, who receive data from the NASA Goddard Space Flight Center. The ERTS MSS data are also available in the form of photographic products and computer-compatible tapes for the cost of reproduction from the Earth Resources Observation Systems (EROS) Data Center, 10th Street and Dakota Avenue, Sioux Falls, South Dakota 57198. Also, in browse files of ERTS data located in various U.S. Geological Survey offices, prints of film products can be viewed before tape recorded data are ordered.

The Earth resources experiments package (EREP) S192 MSS is on board the Skylab spacecraft, which is currently inactive and in orbital flight at an altitude of approximately 434 kilometers (234 nautical miles). The EREP S192 MSS data from three manned missions are available through the EROS Data Center.

The JSC multispectral scanner and data system (MSDS) (flown on a C-130 aircraft) was designed primarily for research and development activities in support of approved Earth Resources Survey Program investigations. Data from the MSDS are provided to approved investigators by JSC.

The Michigan scanner is flown on a medium-sized, commercial-type aircraft by the Environmental Research Institute of Michigan (Ann Arbor, Michigan) and is available to collect MSS data and supporting data under contract. The institute provides flight time, data formatting, and data processing as services.

The Bendix Aerospace Systems Division of Ann Arbor, Michigan, also provides data acquisition and data processing services. The scanner used is a modular multispectral scanner, which is mounted in a commercial aircraft.

All the available multispectral scanner systems record data on tapes that cannot be read by standard digital computers. All the scanners listed in table I record output imagery and housekeeping data in digital form except the Michigan scanner, which produces analog output data. The data from these systems must be preprocessed (including digitizing the Michigan scanner data) for conversion to computer-compatible tape formats.

Some users may desire to use multiband photography as a source of input data. Engineering experiments have been performed to show that such data can be used, and some of these attempts have been successful. Very expensive, highly specialized equipment is needed to extract the data so that the readings from the different data bands coincide. Because of deficiencies in the state of the art and available equipment, potential users are urged to investigate carefully all the steps involved in converting the film recorded data to digital tape recorded data before setting up a survey project in which camera data are used. The source of difficulty is in the reading of the film (e.g., shading effects, registration problems, exposure nonuniformity, and shutter speed variation) rather than in the procedure described in this report.

In general, if a large area is to be surveyed, fewer categories of land use would be involved and data from a satellite platform would be preferable. Aircraft platforms are usually chosen when smaller areas and a larger number of land use categories are required. Some survey objectives can be adequately met with a small number of spectral bands and gross resolution cells (spot size observed at the surface), whereas other requirements can be fulfilled only with high-resolution sensors at low altitudes. The most effective data bands for land use classification can best be determined after a study of recently published reports in which authors recorded those data bands found to be most useful in their land use studies. One excellent approach for preliminary data band selection is to consult the University of Michigan Target Signature Analysis Center, which gives nominal signatures for many classes of materials under known conditions.

Spectral signatures are not unique for a given class of material; a family of spectral "signature curves" describes each material throughout its growth cycle and under various lighting and seasonal conditions. The signatures are also modified by effects of the atmosphere, by particles in the atmosphere, and by variations in solar lighting conditions. Therefore, although a probable set of data bands can be selected before a data collection exercise, the final selection of data bands should be made as an inherent part of the data processing procedure, especially when processing data from a scanner having several data bands. It is also advisable to have a few extra data bands that are near optimum.

Data bands cannot be selected by observing the signature for a single material; all spectrally similar and all major signatures (i.e., signatures that make up a significant percentage of the survey area, whether of interest or not) in the survey area must be considered. Figure 1 is an example of a plot of the arithmetic mean scanner responses for four classes of material as recorded by a 13-spectral-band MSS. For planning purposes, potential spectral data bands should be selected from data bands for which signature curves are as widely separated as possible. For the case shown in figure 1, data bands 2, 3, and 5 should be rejected; the remainder of the data bands would be considered desirable candidates.

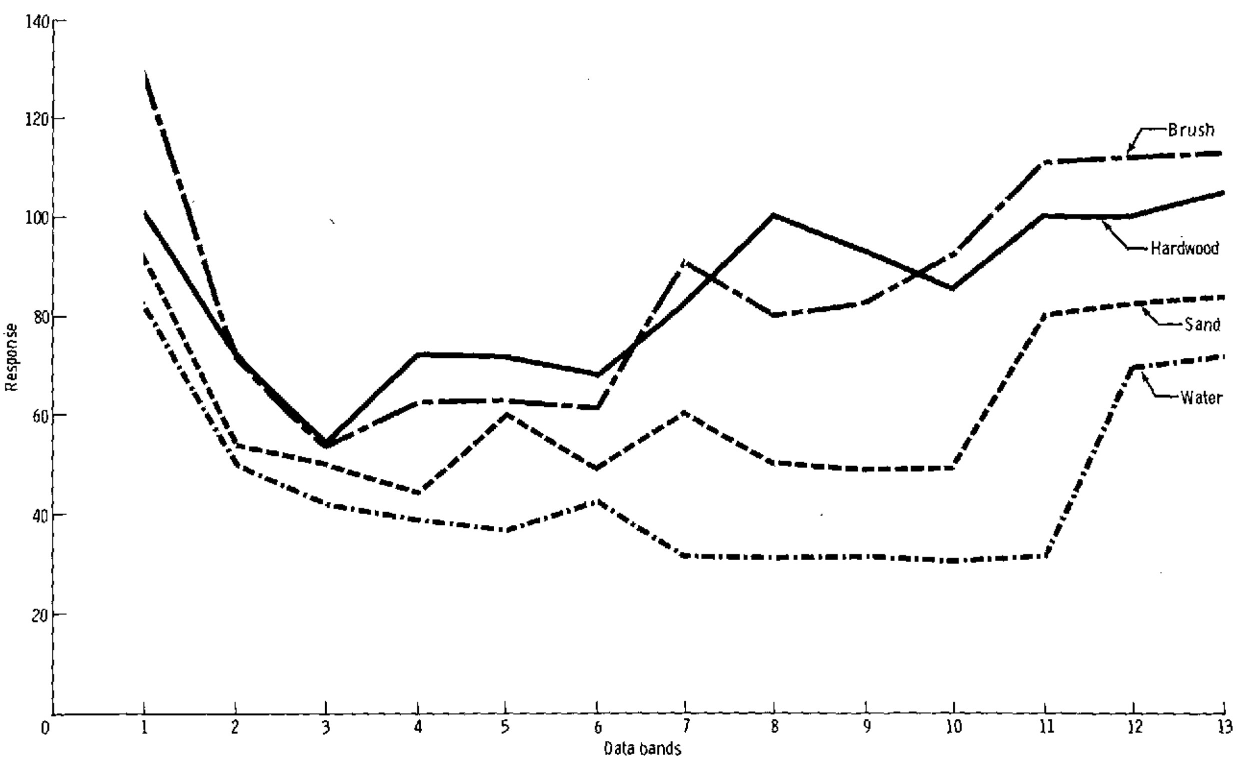


Figure 1.- Mean spectral plots for classes of material observed by a multispectral scanner.

At this point in the procedure, the data user has identified the objectives, the survey area, the desired products, the sensor platform, and the sensors required; has made a reasonable selection of data bands; and is prepared to request a mission for acquiring data or to request data that already exist in a storage bank.

Request or acquire data. - If existing imagery will fulfill the requirement for data, request such data from the EROS Data Center, described earlier. If the data required do not exist or if no ground-truth data are available, then develop a set of requirements (survey plan) and pursue the appropriate course as follows.

1. If a satellite platform is desired, plan to conduct a data collection exercise in conjunction with an ERTS overpass and to acquire the imagery from the EROS Data Center. Collect ground-truth data as required in the survey plan.

2. If the survey plan requires a set of data that would be best acquired from an aircraft platform, contract for the necessary survey flights from appropriate commercial sources.

3. If the survey plan entails a research and development activity, it may be considered as a potential investigation by the Earth Resources Survey Program. A set of requirements should be developed and submitted to the NASA Office of Applications, Washington, D.C. If approved, the investigation would be assigned to a NASA field center. Certain investigations may also be of interest to the EROS Program of the Department of the Interior.

Data Acquisition

Data can be acquired by remote sensing from space platforms or from aircraft platforms.

Remote sensing from space platform. - If the survey requires a spaceborne MSS, select an area over which and at a time when the spacecraft overflight occurs and when lighting conditions are suitable. Perform ground-truth data collection activities as required by the survey plan. Ground-truth personnel at the various preselected training sample areas should take notes on the various measurements and observations made relative to each survey site. The notes should record the identity of the material, the size of materials, the percentage of ground cover, the uniformity of coverage, and any unusual activities in the area. An example of a ground-truth data collection form used by ERL for ERTS data acquisition is shown in figure 2. In certain cases, it may also be desirable to acquire photography, spectrophotometer data, temperature, and other supporting measurements and to make observations about standing water and recent rainfall. For certain large survey areas, remotely sensed aircraft data should be used as an extension to ground-truth data. Organize and compile all field notes in an orderly manner for use in the processing and analysis of the data.

Sample No. or
Field Code ______________

NASA/ERL
GROUND TRUTH DATA

DATE: ______________________ TAKEN BY: ______________________

LEGAL DESCRIPTION: ____
County 1/4 1/4 Section Township Range

FIELD OR SAMPLE SIZE: ________ + ______ (feet) or ______________ acres

LAND USE CLASSIFICATION: ______________________________________

SPECIES OR SPECIES ASSOCIATION: ______________________________
____
____

PLANTING TECHNIQUE: ____

ROW DIRECTION: ____

ROW WIDTH: ____

SOIL CONDITION: Appearance such as: freshly cultivated, rough, smooth, wet or dry, etc.
____
____
____

SOIL MOISTURE: (e.g., moist, dry, waterlogged): _______________
____

CROP PLANT HEIGHT: ____

CROP PHYSIOLOGICAL STATE: ________________ (flowering, heading, etc.)

CROP VISUAL CONDITION: __________________ (chlorotic, wilted, etc.)

PERCENT GROUND COVER: ____________________

WEED INFESTATION (SPECIES & %): ______________________________

DISEASE INFESTATION (KIND & AMOUNT): __________________________

INSECT INFESTATION (KIND & AMOUNT): ___________________________

PHOTOGRAPHS: (1) Film Roll No. ____________ (2) Frames ______ to ______

COMMENTS: ____
____
____

Figure 2.- Ground-truth data form.

Remote sensing from aircraft platform. - If an aircraft platform is required, make necessary arrangements to acquire the services of an appropriately equipped aircraft by commercial contract or some other approach. Develop detailed plans to coordinate aircraft data collection activities with ground-truth activities. Select flight lines carefully to ensure complete coverage without gaps and to achieve the necessary resolution spot size. Flight altitude and speed should be chosen so that the MSS records contiguous scan lines of data. Because visual bands of an MSS are subject to the same lighting constraints as camera systems, Sun angle should be considered to provide optimum lighting conditions.

The acquisition of data should be planned so that it can be successfully completed without communications; however, communications should be provided between ground-truth personnel and the aircraft through a coordinator to coordinate real-time changes in the flight plan due to weather conditions, sensor anomalies, and so forth. The ground-truth operation associated with aircraft data acquisition exercises is essentially similar to that used with a spacecraft survey.

Data Processing

The procedures recommended for processing of data are presented in the following paragraphs. Figure 3 is a photograph of the ERL DAS facility, which is described in appendix B.

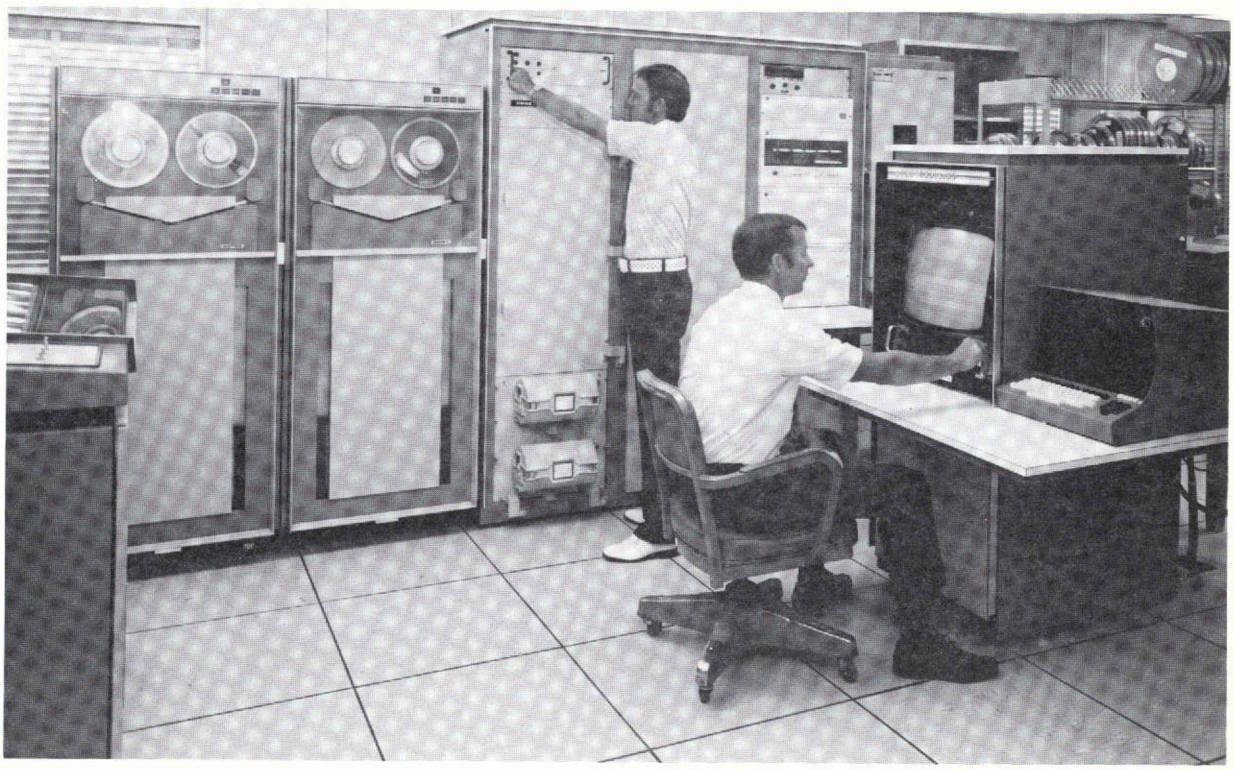


Figure 3. - Earth Resources Laboratory data analysis system facility.

Control data. - The data from all sources should be routed through a control office to be logged and routed to the appropriate facility for preprocessing or processing. The location and status of data should be known at all times.

Preprocess data. - All data should be preprocessed for conversion to the necessary format for the remainder of the data processing tasks. Field notes should be compiled and reviewed for completeness and accuracy.

Spacecraft-acquired MSS data (ERTS and Skylab) are preprocessed to convert the imagery from instrumentation tape recorded data to computer-compatible tape recorded data. During image conversion, all known corrections and calibrations are applied to provide good geometric and radiometric data. Also, the curved scan lines of the Skylab S192 MSS are converted to straight scan lines during preprocessing.

Aircraft-acquired MSS data also require preprocessing, since all known aircraft scanner data are also recorded on instrumentation tape. If necessary, the data are digitized. Corrections for calibration, geometry, gain, level, and so forth are made to the scanner data as necessary, and computer-compatible tapes are generated for further data processing.

Screen and evaluate data. - The screening function is identical for all types of multispectral scanner data. Either the data bands can be selected individually and displayed as black and white images on the DAS display screen or each of three data bands can be displayed simultaneously on the screen as red, blue, and green signals. The color guns of the display screen may be extinguished selectively so that only one or a combination of two bands can be seen as desired. Data bands can be selected as desired until all available bands have been displayed and examined. Results of the screening process are recorded in a screening report that gives the quality of the signal in each data band. The screening report is provided to the user for his consideration. If the data quality is too poor, a new set of data may be necessary before an acceptable land use map can be prepared. If the data quality is acceptable, processing can continue. The screening and evaluation steps are usually a joint effort between the operators of the data analysis system and the data user. The evaluation begins during screening and continues while the user reviews the notes taken during screening and the flight log (for aircraft-acquired data) transmitted with the data tapes.

Prepare DISPLAY tapes. - During the screening and evaluation steps, data processing personnel will select three MSS data bands to be used for training sample selection. These three data bands will be extracted from the complete original data set and formatted into a special DISPLAY tape. It should be emphasized that the DISPLAY tape is used only to recognize the location of areas within the test site that are to be used as training samples. Figure 4 is an example of the three data bands displayed as red, green, and blue images superimposed on the DAS display screen.

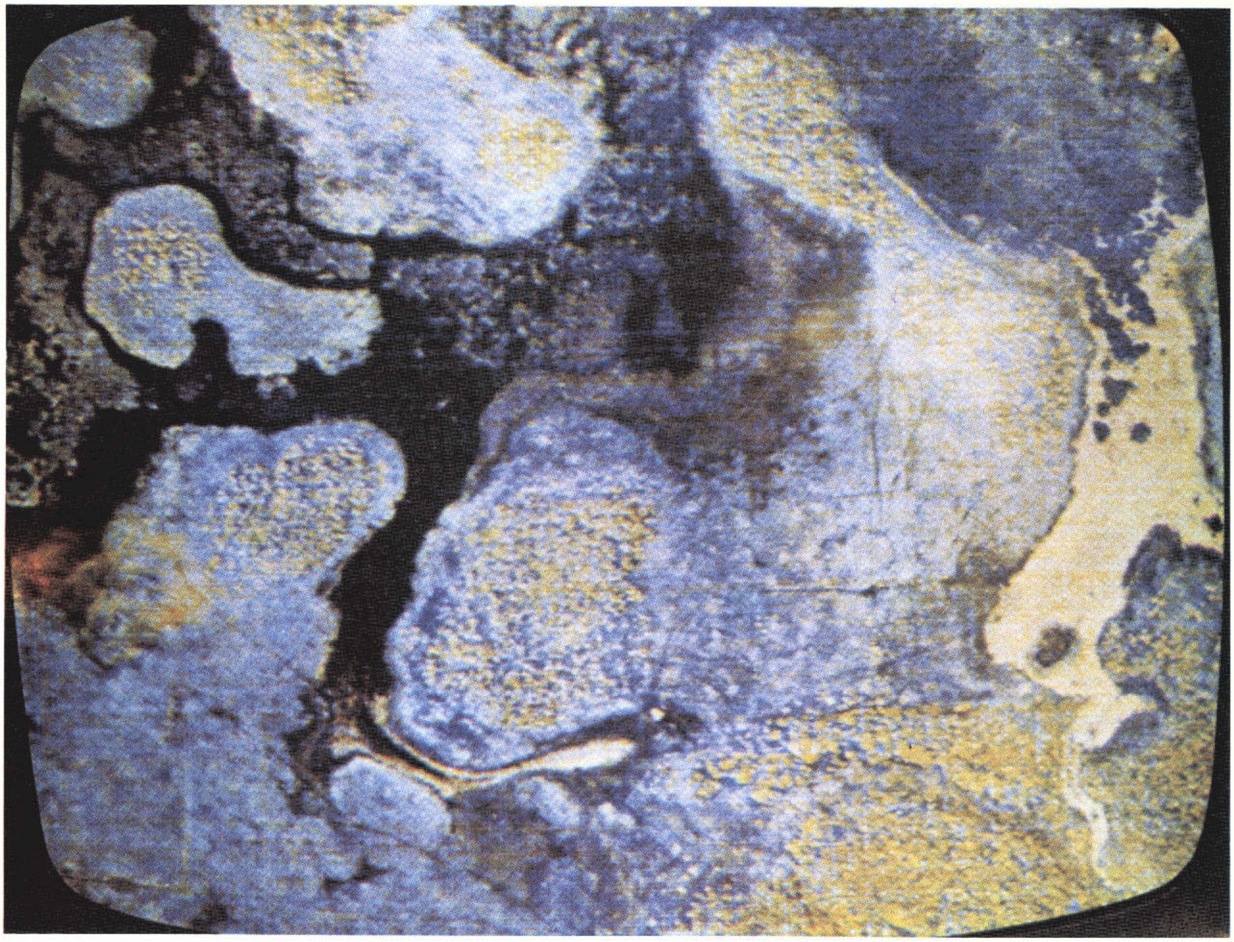


Figure 4. - DAS display screen showing three data bands (red, green, and blue) simultaneously.

Select training samples. - The most critical step in the procedure for generating land use maps from multispectral scanner data is the selection of training samples. Training samples are used by the spectral pattern recognition program as the basis for statistically computed signatures. Therefore, the training samples chosen must statistically represent each major class of material in the survey area.

Prepare for survey: Training samples should be preselected, located, and verified on the surface by ground-truth personnel before performing the survey, and these training samples should be reverified on the day of the survey. Training samples should be distributed throughout the survey area. It is more important that the training samples represent all conditions for a given class of material than it is to have very large training samples. Training sample size may range from 30 by 30 meters (100 by 100 feet) for aircraft-acquired data to 402 by 402 meters (1320 by 1320 feet) or larger for spacecraft-acquired data. It is recommended that training samples be located on either side of the center of the flightpath because lighting conditions generally differ for the two sides.

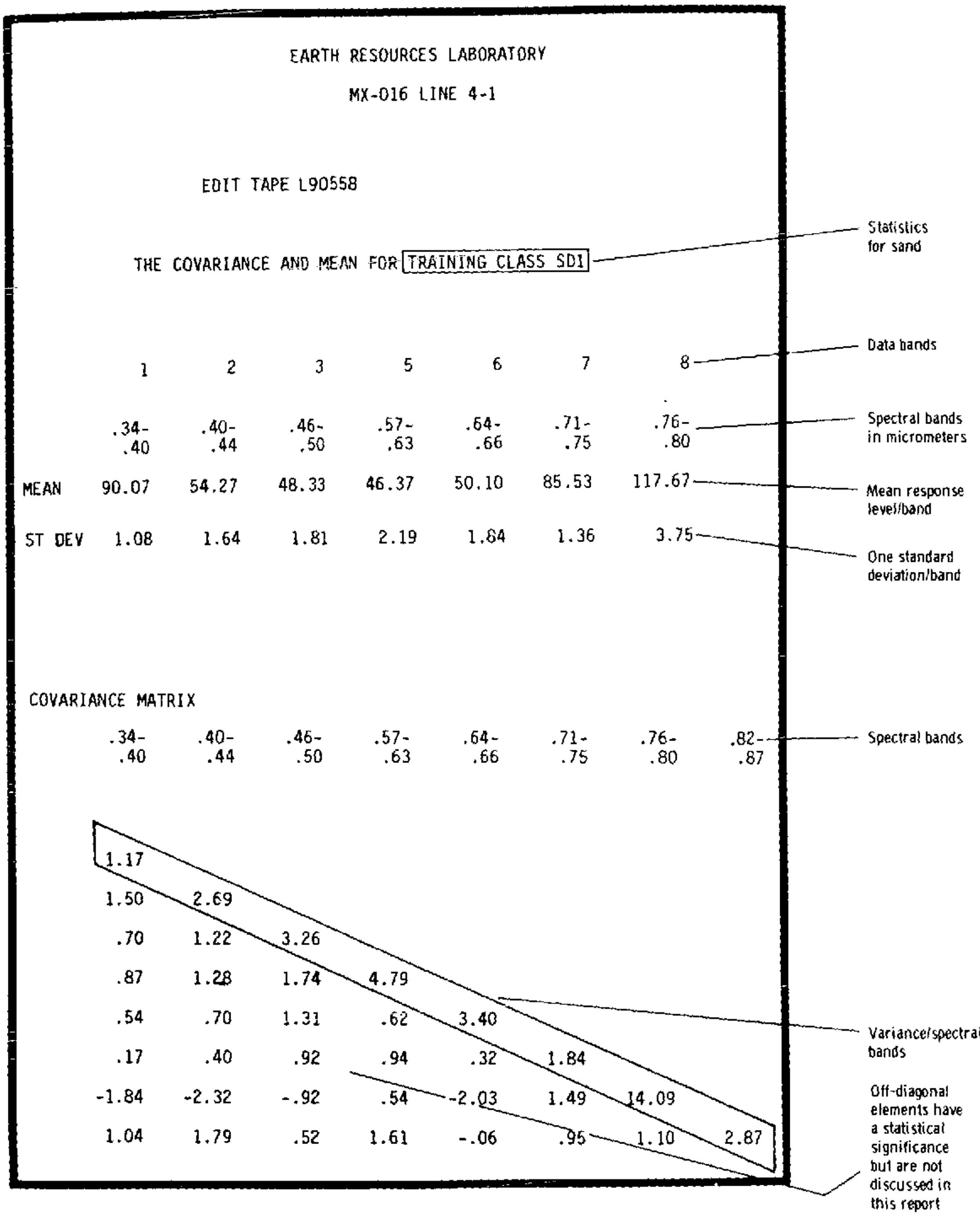
Identify training samples: The display tape prepared earlier is read by the DAS computer and displayed on the cathode-ray tube (CRT), as shown in figure 4. The DAS operator and the data user locate a training sample in the DAS color display. The DAS is equipped with a light-pen cursor and an associated light-pen target. The DAS operator enables the light-pen cursor, which causes the light-pen target to appear on the screen of the DAS display. The four corner points of the light-pen target can be guided to form a four-sided cursor figure that fits within any selected training sample. Avoid selection of training samples that obviously contain two kinds of material since the signature produced from such a dichotomous population would probably be unlike that of any class of material occurring in the survey area. When the light-pen cursor is correctly positioned, the coordinates of the training sample are recorded by the DAS computer, and a code name for the class of material, together with the coordinates of the training sample data, is entered into DAS storage by the DAS operator for later use.

The process of locating training samples in the DAS screen, positioning the cursor, and entering a coded name and coordinates for materials is repeated until a good statistical sampling has been taken of all major materials in the survey area. Some additional samples are identified by the user for use in assessing the accuracy of the survey. These samples are defined as "test fields." Test fields are selected by the same procedure as training samples but are not used to statistically compute signatures for the various classes of materials. To classify 12 surface conditions from ERTS data, the ERL selects a minimum of 8 training samples per surface condition (total of 96) for each 185- by 185-kilometer (100 by 100 nautical mile) area. Because the test area will probably be quite large, the training samples will account for a very small percentage of the total survey area. Training samples must be chosen for all major classes of materials in the survey area, not just for those of interest, so that the separability of two or more spectrally similar classes of material in the survey area can be determined.

Extract training samples. - After the training samples have been located from the display tape by use of the DAS screen, the DAS operator mounts the MSS data tape, which contains all channels of MSS data available, on the DAS. Program TAPE COPY then is used to produce a training sample tape that contains only training sample data and the code names identifying each class of material.

Analyze training samples statistically. - The training sample tape produced on the DAS is processed by program STAT on the Univac 1108 computer. (Software program modules are described in appendix C.) Program STAT produces punched cards or computer-compatible tapes and statistical tabulations for each separate training sample and for combinations of all training samples of one kind of material. For example, there may be 25 separate training samples for the agricultural crop rice. The STAT program will compute statistics for each of the 25 separate training samples and a combination of all 25 rice samples. The following statistics are provided.

1. Mean response for each data band for each training sample - An example of the mean response is shown in figure 5.



EARTH RESOURCES LABORATORY

MX-016 LINE 4-1

EDIT TAPE L90558

THE COVARIANCE AND MEAN FOR TRAINING CLASS SD1

	1	2	3	5	6	7	8
	.34-.40	.40-.44	.46-.50	.57-.63	.64-.66	.71-.75	.76-.80
MEAN	90.07	54.27	48.33	46.37	50.10	85.53	117.67
ST DEV	1.08	1.64	1.81	2.19	1.84	1.36	3.75

COVARIANCE MATRIX

.34-.40	.40-.44	.46-.50	.57-.63	.64-.66	.71-.75	.76-.80	.82-.87
1.17							
1.50	2.69						
.70	1.22	3.26					
.87	1.28	1.74	4.79				
.54	.70	1.31	.62	3.40			
.17	.40	.92	.94	.32	1.84		
-1.84	-2.32	-.92	.54	-2.03	1.49	14.09	
1.04	1.79	.52	1.61	-.06	.95	1.10	2.87

Figure 5.- Training sample mean, standard deviation, and variance in each spectral band for one class of material (sand).

2. Variance in response level in each data band for each training sample (fig. 5) - The variance provides a measure of the variation in response for each data band. This measurement provides a good clue to noisy data, which could be caused by scanner malfunction in a given data band. A large variance can also indicate that more than one type of material is included in a training sample (i.e., that the training sample is not homogeneous).

3. Spectral plots - Program STAT produces a plot of response level as a function of frequency bandwidth (or data band). The spectral plot (fig. 6) includes not only the mean value for each material but the variance as well. A spectral plot can be provided for each training sample and for groups of training samples.

4. Histograms - Frequency of occurrence of response levels for a given training sample for one data band is shown in histograms (fig. 7). The value of the histogram lies in ensuring that two materials have not been mistakenly included in one training sample (fig. 8). The histogram is valuable to data processing personnel because it frequently indicates that one class of materials should be separated into two classes because of the existence of two species of the same material or indicates that the material is in two distinct portions of a growth cycle.

The output from program STAT is important because it indicates to data processing personnel the degree to which the selected training samples statistically represent the data set. If the statistical fit is good, processing should be continued; if it is poor, the training sample should be reevaluated because, at this point, only a small amount of expensive computer time has been used.

<u>Determine best bands for pattern recognition classification.</u> - If the statistics obtained are acceptable, the training sample data will be processed through program CHOICE on the Univac 1108 computer. Through a divergence calculation performed by using program CHOICE, the best four data bands for pattern recognition classification (appendixes C and D) are determined. To determine the best four bands for ERTS, which has only four data bands, obviously is not necessary; however, if the user desires, program CHOICE can be used to identify the best one, two, or three bands. The program CHOICE capability to identify the best bands is very helpful when processing data from scanners having several bands. Program CHOICE allows data processing personnel to determine the relative separability or pairwise divergence of each class of material from all other classes of material and indicates cases in which materials are not statistically separable. In such cases, any of the pairs of statistically indistinguishable materials can be considered as one class of material in the next processing step. Although ERTS has only four data bands, it is important to process the ERTS data through program CHOICE to determine the degree to which the various training samples are statistically identifiable. Figure 9, which contains the pairwise divergence of pairs of material, is a sample of the information provided by program CHOICE.

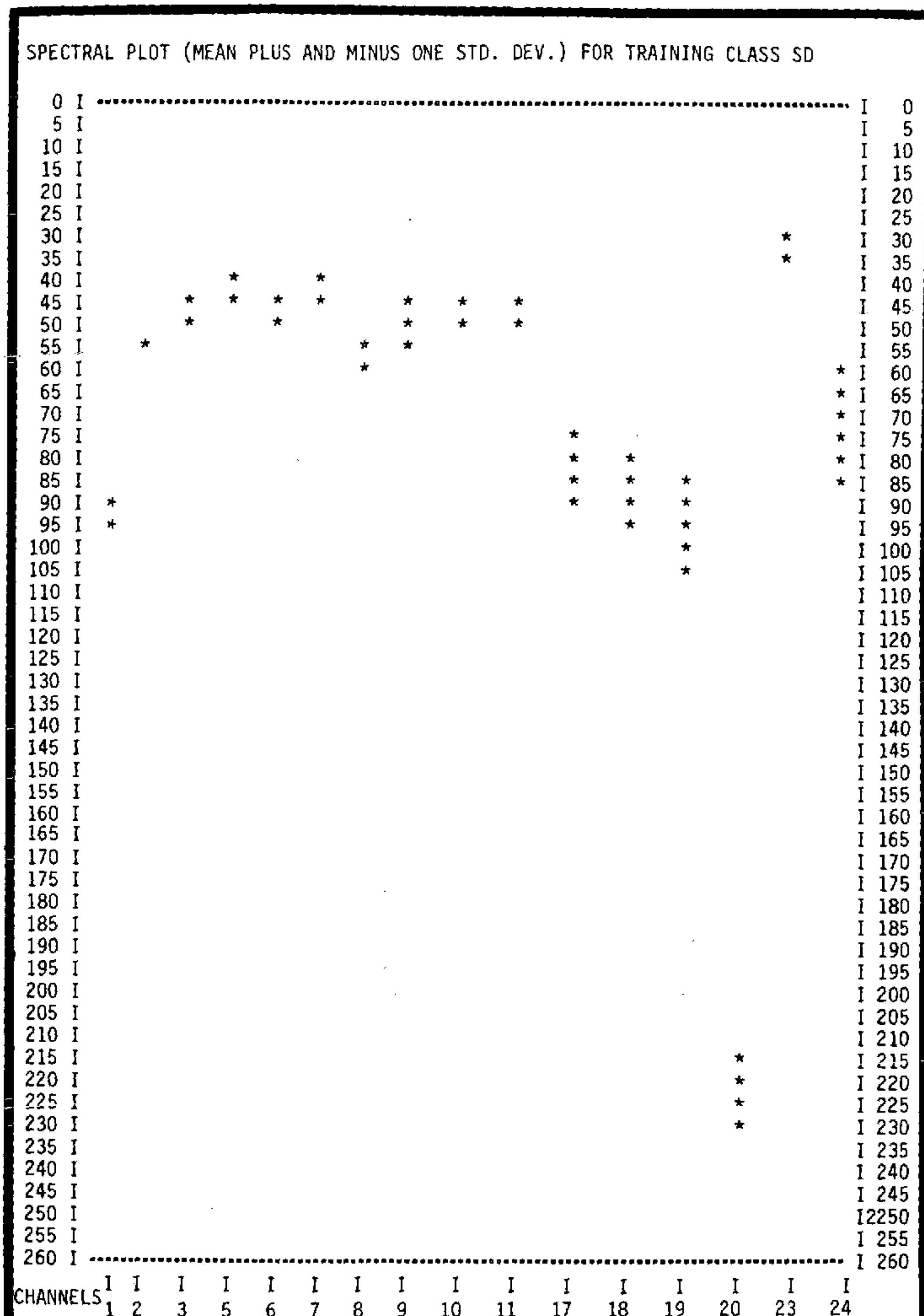


Figure 6.- Spectral plot of response levels plus and minus one standard deviation in each spectral band for a class of material.

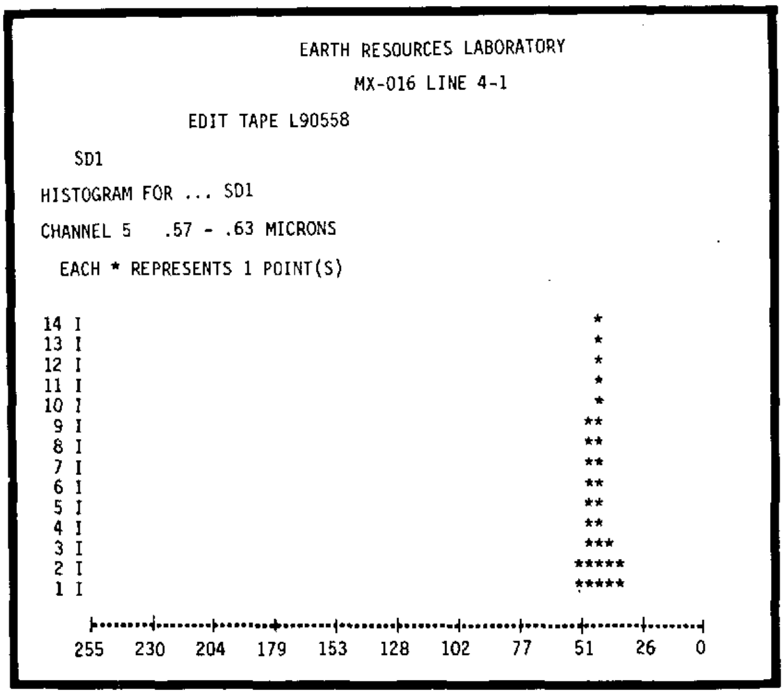


Figure 7. - Histogram showing frequency of occurrence of response levels in one data band for a given class of material (sand).

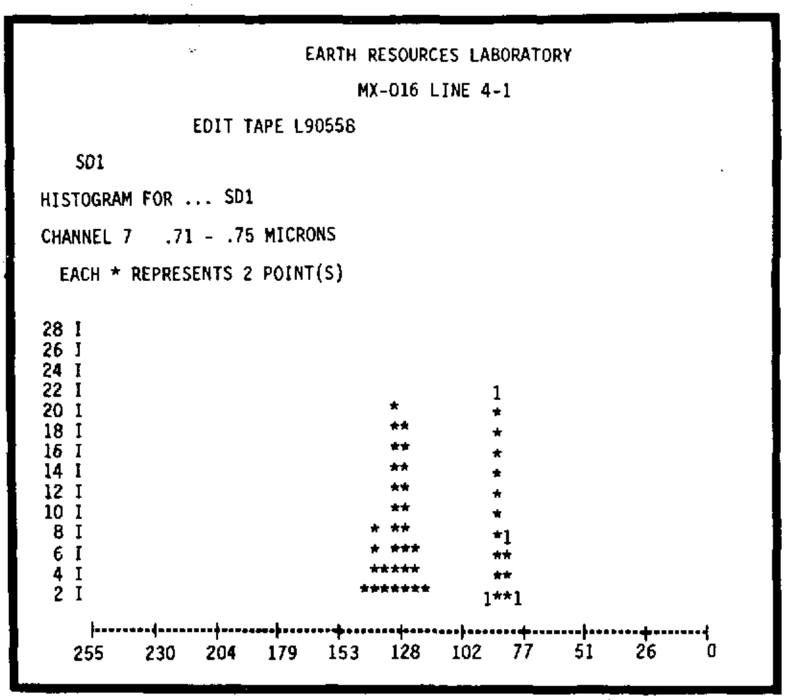


Figure 8. - Histogram showing frequency of occurrence of response levels in one data band when two different materials are identified erroneously.

```
CHANNELS=      1     9     11     12
AVERAGE PAIR-WISE DIVERGENCE=    443.0

MINIMUM PAIR-WISE DIVERGENCE=     25.8

RATIO OF THIS CHANNEL SET WITH CHANNEL SET YIELDING MAXIMUM PERCENT SEPARATION=    .33

AVERAGE DIVERGENCE BY CLASS...
AVERAGE INTERCLASS DIVERGENCE FOR CLASS S =     173.702
AVERAGE INTERCLASS DIVERGENCE FOR CLASS C =     268.075
AVERAGE INTERCLASS DIVERGENCE FOR CLASS O =     180.312
AVERAGE INTERCLASS DIVERGENCE FOR CLASS W =     303.635
AVERAGE INTERCLASS DIVERGENCE FOR CLASS R =     654.247
AVERAGE INTERCLASS DIVERGENCE FOR CLASS A =     696.094
AVERAGE INTERCLASS DIVERGENCE FOR CLASS Y =     206.235
AVERAGE INTERCLASS DIVERGENCE FOR CLASS X =    1165.065
AVERAGE INTERCLASS DIVERGENCE FOR CLASS E =     339.305

 PAIR-WISE DIVERGENCE
CS=   26.1  CS=   78.4  OC=   76.6  WS=  202.4  WC=  324.5  WO=  110.9  RS=  209.5
RC=  100.3  RO=   90.3  RW=  674.6  AS=  188.7  AC=   82.2  AO=   92.2  AW=  733.0
AR=   25.8  YS=  117.5  YC=  204.3  YO=   37.9  YW=   52.6  YR=  423.3  YA=  439.3
XS=  346.8  XC=  921.4  XO=  804.1  XW=  299.6  XR= 2962.8  XA= 3281.7  XY=  325.5
ES=  220.2  EC=  409.2  EO=  152.0  EW=   31.5  ER=  747.5  EA=  725.9  EY=   49.6
EX=  378.6
```

Figure 9.- Display of channels ranked according to average pairwise divergence.

Generate classification tables.- Program LARSYSAA (ref. 3), which was the parent program of the ERL pattern recognition programs, used a Bayesian maximum-likelihood ratio calculation to identify, on an individual resolution-cell basis, the probability that an unknown material is identical to each of the materials that the computer has been trained to recognize. The computer decides that the unknown material is actually the material with the highest calculated probability, provided that probability exceeds some minimum threshold value. The calculations are very lengthy and time consuming.

The ERL spectral pattern recognition programs are also based on the maximum-likelihood ratio, in that this principle is used to generate decision tables for each class of material. Programs SUPER-T and R-TABLE are used in sequence by ERL to generate decision tables for each class of material (as many as 30 classes) to be identified; the probabilities are determined by table look up (an approach to pattern recognition devised by Eppler (ref. 9)) rather than by calculation. The decision tables are recorded on a computer-compatible tape and are used as the basis for classification in program R-CLASS.

Extract best data bands from total data set.- Program R-CLASS, which is the heart of the land use mapping system, is limited by computer tape-reading rates. Since data input and throughput rates can be improved, it is advantageous to extract only those data bands identified by program CHOICE for use in program R-CLASS. The best data bands are extracted from the entire data set by DAS program Digital Tape Copy. This step is not necessary for the ERTS MSS data but is intended for scanners having a larger number of data bands.

Classify unknown data sets. - Program R-CLASS accepts the following inputs: control cards (input information for which the user selects program options), decision tables produced by programs SUPER-T and R-TABLE, and the data tape reflecting the best data bands. Program R-CLASS identifies classes of materials by the table look-up technique (ref. 9). The digital table look-up classification scheme currently is programed to classify as many as 12 classes of material in one computer run. If more than 12 classes of material are to be identified, program R-CLASS may be run as many times as desired with different sets of decision tables. The output tapes from the several classification runs may be combined into a single output tape containing any number of material classes.

The tape recorded outputs from program R-CLASS are characters representing identified classes of materials and are not in a form convenient for analysis. These characters are translated to other appropriate displays.

Color-code, scale, and rectify land use map. - The program R-CLASS output tape, known as MAPTAP, is processed through Univac 1108 program R-COLOR. Program R-COLOR assigns a color-coded output symbol to each resolution element of the land use map in a format compatible with the ERL DAS, applies an interpolation formula to each scan line to correct for geometric distortion of the scanner system, and applies the correct electronic magnification function to provide the desired map scale.

Spacecraft-acquired MSS data must be corrected for the rotational effects of the Earth. The along-track component of rotation may be corrected in program R-COLOR for ERTS and Skylab scanner data; the crosstrack component of rotation is corrected in the DAS program DIGITAL CONVERT just before the data are recorded on the DAS film recorder.

Prepare scorecard. - The MAPTAP tape is also processed through Univac 1108 program REPORT. Program REPORT shows how accurately program R-CLASS identified materials in the survey area by a check of two kinds of information. The classification accuracy is checked in training sample areas that were used to develop the signatures and also in "test fields" for which the identity of material classes is actually known beforehand. (Data from the test fields are not used in the development of signatures.)

Program REPORT provides a measure of the accuracy of the land use map produced. The output of program REPORT is called a "scorecard," an example of which is shown in table II. With a good selection of training samples, it is not unusual to have percentages of correct classification in the 80- to 100-percent range in training sample areas. Low percentages of correct identification can be anticipated if there is a low pairwise divergence in the output for program CHOICE. Program REPORT can provide, as an output, acreage compilation for the various classes of materials. The approach used is to multiply the total acreage of the survey area by the percentages of each class of material identified in a defined survey area. More sophisticated techniques for acreage compilation are currently under development.

TABLE II. - SCORECARD PERFORMANCE EVALUATION

(a) Classification

Class no.	Material no.	Material name
1	1	Pine
2	2	Water
3	3	Hardwood
4	4	Sand
5	5	Brush

(b) Performance by test field

Test field	No. of samples[a]	Percent correct[b]	Class no.				
			1	2	3	4	5
Pine 1	425	84.3	358	0	40	0	27
Pine 2	248	87.2	217	0	21	0	10
Water 1	125	98.0	0	122	0	0	0
Water 2	64	97.3	0	62	0	0	0
Hardwood 1	83	88.1	10	0	73	0	0
Hardwood 2	245	85.4	31	0	209	0	5
Sand 1	84	92.3	0	2	0	77	0
Sand 2	42	95.1	0	1	0	40	0
Brush 1	76	82.0	6	0	8	0	62
Brush 2	34	76.1	4	0	5	0	25

[a]Total number of resolution elements in all training samples and test fields considered in accuracy assessment.

[b]Overall performance = 88.58 percent.

Record land use map on color film. - The character-encoded map produced by program R-COLOR is converted to a color-coded land use map and recorded on either positive or negative color film. The required conversion is performed by reading the tape produced earlier by program R-COLOR into DAS program T-SAMPLE SELECT and then by recording it on color film that is 24.13 centimeters (9.5 inches) wide. The ERTS and Skylab MSS imagery must be corrected for Earth rotational effects. The along-track component of rotation is corrected in program R-COLOR, and the cross-track component is corrected in DAS program T-SAMPLE SELECT. Although the color-coded land use map has reasonable geometric accuracy for many applications, it should not be considered a high-quality metric map as produced by Government mapping agencies.

Process color film. - The recorded color film is processed as a positive or negative color transparency in an Eastman Kodak Versamat color processor. Color and all processing parameters must be tightly controlled so that different strips of color map may be mosaicked without the introduction of distracting color discontinuity.

Check film product quality and print. - The land use map color transparencies are reviewed by photographic, data processing, and user personnel for quality and usability. Color paper prints are requested as required to make a final color-coded land use map.

Prepare final product. - The color paper prints requested previously are mosaicked as carefully as possible. Figure 10 is an example of a typical final color-coded land use map. County and state lines and identification of major cities, islands, rivers, lakes, and so forth can be added to the mosaic as required. A legend can be added together with other descriptive information to aid the user in the analysis of the data.

In figure 10, which is a portion of an ERTS frame, three wide strips of film cover the major portion of the survey area because an ERTS frame representing 185 by 185 kilometers (100 by 100 nautical miles) is provided to users on four separate tapes. The narrow strips mosaicked between the wide strips were included because of color discontinuity in processing.

The following final data products may be provided.

1. Statistics from program STAT

    a. List of training samples chosen

    b. Means, per data band per training sample

    c. Covariance matrices, per class of material

    d. Correlation matrices, per class of material

    e. Variance, per training sample

    f. Spectral plots, per class of material

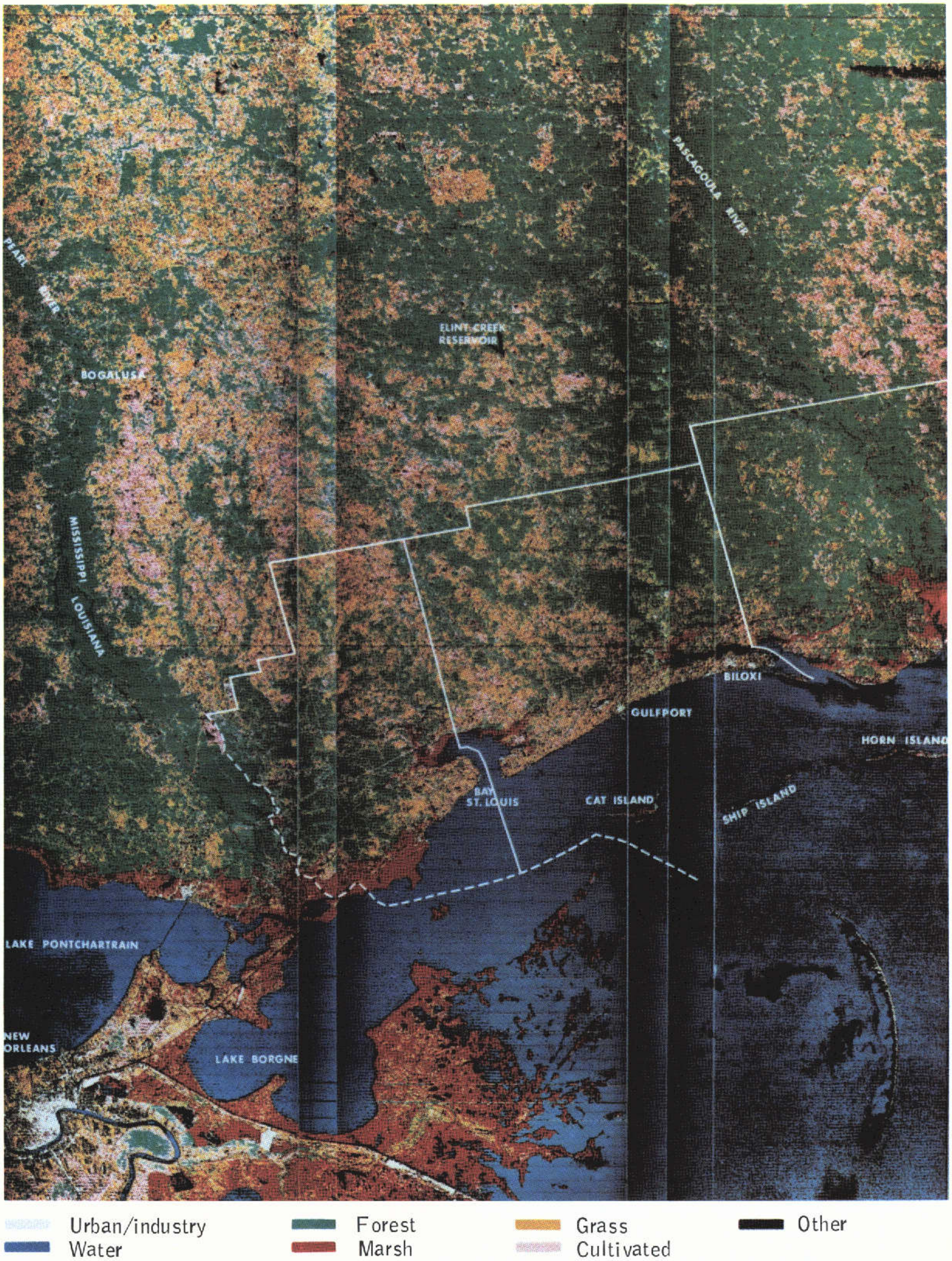


Figure 10. - Computer-generated, color-coded land use map.

2. Pairwise divergences for sets of data bands taken a number N at a time ($N \leq 36$) with best combinations identified

3. A scorecard

a. Total number of samples considered in accuracy check

b. Identification of each class in known data set

c. Percentage and number of correct classification

d. Percentage and number of incorrect classification, with confused class identified

e. Acreage compilation by class of material

4. A scaled, color-coded land use map as continuous strips of film (positive or negative transparency)

5. A scaled, color-coded land use map as continuous strips of color paper (prints)

6. A scaled, mosaicked, color-coded land use map with legends and annotation

Copies of tabular products may be obtained by using available duplicating equipment; copies of the color-coded land use map may be made by using a color-corrected copy camera.

FUTURE PROCEDURE AND HARDWARE SIMPLIFICATION

A major objective of the ERL is to greatly simplify the procedure, hardware, and software required to generate land use maps from multispectral scanner systems. Although the procedure, the software, and the hardware used by the ERL represent the state of the art in land use mapping, the system is still too complex to be considered operational.

Many approaches to simplify the automatic land use mapping system have been considered by the ERL, including elimination of the DAS and use of the photographically recorded imagery, such as ERTS-1 imagery, to estimate (by measurement on the film) the location of training samples. The uncertainty of the exact coordinates (scan line counts and picture element numbers) made training sample selection of 16-square-hectometer (40 acre) fields impractical. This level of inaccuracy may be acceptable for most applications, but for others, the capability to identify fields of smaller size may be critical. For this reason, several mathematical techniques were used to refine the estimates of the location, but these did not adequately improve the location of training samples and did not provide a sufficient degree of confidence that the training samples were located accurately.

In an informal document, Dr. Walter G. Eppler of the Lockheed Electronics Company (LEC) describes a technique for superimposing a grid over multispectral scanner imagery (as the imagery is being translated from tape to film) in such a way that coordinates of training samples can be located to the nearest picture element. Eppler's informal document is entitled "Computer-Band Coordinate Grid for Multispectral Scanner Imagery" and was published as an internal LEC document in 1973. Eppler's technique appears to be feasible but expensive to implement, since it requires some rather sophisticated equipment. It is apparent that a hardware image display device is required for effective operational land use mapping for cases (e.g., ERTS) in which training samples are likely to be smaller than 16 square hectometers (40 acres).

Requirements

Hardware.- The data analysis system (fig. 3 and ref. 10), which cost approximately $730 000, must be replaced in any operational system by a simple, inexpensive hardware image display system that can read MSS data directly from computer tape. The image display device must enable users to locate training samples to the nearest resolution element on the computer-compatible tape.

The image display system should provide enough flexibility to accept several types of MSS imagery if the data are properly preformatted. To reduce the costs of an image display system, the input data should be reformatted to one common-image format for image display. This reformatting can be accomplished on almost any small, general-purpose digital computer having at least two tape drives. The image display system should be capable of processing multispectral scanner imagery having a variable number of elements per scan line to make it compatible with the various existing MSS's. The image display system must display the image as an enlarged, false-color image to aid the data user in recognizing training samples.

The prototype image display system must be developed, evaluated, and refined so that it can provide reliable, trouble-free operation. The system must be accompanied by user documentation, and detailed specifications and drawings must be provided that are suitable for competitive manufacture in large quantities.

Software.- The software required for an operational land use mapping system must be simplified by removing the flexibilities required for technique development, but must retain the necessary features to permit data evaluation. The software must be sized to accept available multispectral scanner data but must be simple enough for use with widely available computer systems.

The software must be documented in sufficient detail to enable implementation by experienced programers on a general-purpose computer, and must be documented in such a way that user personnel with minimum mathematical or statistical training and a suitable disciplinary background can use the land use mapping system. The software must fit into a computer possessing core storage of approximately 32 000 words of 32 bits. The absolute minimum word size is 16 bits, but the system can be implemented more efficiently on a 32- or 36-bit computer system. The computer must be capable of accessing items within 1 microsecond or less and should have two tape drives of seven or nine tracks.

Simplification Approach

Software. - The ERL spectral pattern recognition programs are being rewritten to simplify the software design and to reduce the computer size. The storage arrays will be sized to handle at least ERTS-type MSS imagery. The software will be documented in sufficient detail to enable experienced computer programers to convert the programs for use on computers of the class described in the preceding paragraph.

Hardware. - The ERL has developed a portable image display system with the capability of reading several types of MSS data from properly formatted, computer-compatible tapes and of displaying the image on a flicker-free color television (TV) monitor. A selectable, 256 by 240 picture element array will be displayed on the TV screen and will be advanced in a "moving window" fashion to allow users to identify or locate preselected training samples in a desired test area. The PIDS will feature a cursor symbol (+) that can be positioned as desired at the corner points of training samples. The coordinates of the cursor location will be displayed on light-emitting diodes expressed in terms of scan line count and picture element number. The PIDS will have the capability to temporarily store four cursor coordinates (corner points for a training sample) and the material class identification code. Upon command, the PIDS will transmit the stored coordinates and the material class identification code to an external paper tape punch, a card punch, or a keypunch device.

Properly formatted, color-coded land use maps in the form of computer-compatible tapes (after classification by pattern recognition programs) can be read into the PIDS and displayed as a color-coded land use map on the TV screen. The PIDS display screen may be photographed, and frames of the displayed land use map may be mosaicked to create a very inexpensive land use map. Any of a number of tape-to-film converters could be used for making a land use map with an investment of as much as $35 000 for a high-resolution color-film recorder. It is believed that duplicate PIDS units could be produced for less than $35 000 each after current operational evaluation has been completed.

CONCLUSIONS AND RECOMMENDATIONS

A workable, semiautomated system of hardware and software components and operating procedures exist for generating land use maps from multispectral scanner data. Although, by using the system, reasonably accurate land use maps of large areas can be produced in less time (and with comparable accuracy) than by using conventional land use mapping techniques, the system is very complex to implement and operate. Current research and development is being directed toward simplification of the procedures, the hardware, and the software for a land use mapping system. A simpler and less costly system is scheduled for operational evaluation in late 1974 and early 1975.

The following actions are recommended.

1. Interested multispectral scanner data users should establish a working relationship with an organization that has an existing land use mapping system to gain an insight into the system, the procedure, the problems, and so forth.

2. Implementation of other existing software and hardware systems should be evaluated against the specifications of the planned simplified land use mapping system.

3. Users should devote sufficient time to defining their operational land use mapping requirements, planning test areas, and preselecting training sample areas of appropriate classes of materials.

Lyndon B. Johnson Space Center
National Aeronautics and Space Administration
Houston, Texas, September 3, 1974
177-89-89-00-72

REFERENCES

1. Vegas, Paul L.: A Detailed Procedure for the Use of Small-Scale Photography in Land Use Classification. NASA TN D-7542, 1973.

2. Vegas, Paul L.: Extracting Land Use Information from the Earth Resources Technology Satellite Data by Conventional Interpretation Methods. NASA TN D-7730, 1974.

3. Landgrebe, D. A.: Automatic Processing of Earth Resource Data. NASA TM X-66484, 1969.

4. Anon.: Earth Resources Technology Satellite Data Users Handbook. General Electric Corp. Document No. 71SD4249, Nov. 17, 1972.

5. Anon.: Earth Resources Production Processing Requirements for EREP Electronic Sensors. PHO-TR524, Philco-Ford Corp., Feb. 1973.

6. Anon.: Earth Resources Production Processing Requirements for Aircraft Electronic Sensors. PHO-TR523, Philco-Ford Corp., Feb. 1973.

7. Anon.: Modular Multispectral Scanner (M^2S). BSR 8039, Bendix Corp., July 1974.

8. Ory, Thomas R.: Review of Operational Applications of Infrared and Multispectral Scanners. Proceedings of the Conference on Practical Applications of Remote Sensing. NASA CR-136316, May 1973, pp. 36-39.

9. Eppler, W. G.; Helmke, C. A.; and Evans, R. H.: Table Look-Up Approach to Pattern Recognition. Proceedings of the Seventh International Symposium on Remote Sensing of Environment, Vol. II, Willow Run Laboratories Rept. 10259-1-X, the University of Michigan (Ann Arbor, Mich.), Aug. 1971, pp. 1415-1425.

10. Anon.: Earth Resources Data Analysis System. NASA CR-128693, 1972.

APPENDIX A

FUNCTIONAL DESCRIPTION OF MULTISPECTRAL SCANNER SYSTEMS

The objective of this appendix is to provide a functional description of several multispectral scanner systems. Reference is made to documents in which more detailed information may be found.

EARTH RESOURCES TECHNOLOGY SATELLITE 1 MULTISPECTRAL SCANNER

The Earth Resources Technology Satellite 1 (ERTS-1) is in a 926-kilometer (500 nautical mile) circular orbit that is nearly polar. The satellite circles the Earth every 103 minutes; every 18 days, it passes over the same point on the surface of the Earth. On each north-to-south pass, it crosses the Equator at 9:42 a.m. local standard time.

The primary sensor on the ERTS-1 satellite is the multispectral scanner system (MSS), which scans the Earth with an 88-meter (290 foot) resolution cell; scanning is normal to the groundtrack on the Earth surface. The scan swath width is 185 kilometers (100 nautical miles). The spectral energy is dispersed by filtration and sensed in four wavelength bands: green, 500 to 600 nanometers; red, 600 to 700 nanometers; near infrared, 700 to 800 nanometers; and near infrared, 800 to 1100 nanometers. The ERTS MSS scanner is so constructed that six contiguous scan lines of data are collected simultaneously as the scan mirror traverses the flightpath. The detector array consists of six almost identical detectors for each spectral band, for a total of 24 detectors.

The detectors respond to radiant-energy changes by producing output voltages. These output voltages are converted to digital data by a 6-bit pulse code modulation (PCM) encoder and, after appropriate conditioning and formatting, are transmitted to ground-based receivers. The data are subsequently transmitted to the NASA Goddard Space Flight Center for preprocessing at the National Data Processing Facility. Reference 4 contains a more detailed description of the ERTS scanner.

SKYLAB EARTH RESOURCES EXPERIMENTS PACKAGE EXPERIMENT S192 MSS

The S192 multispectral scanner, an optical mechanical scanner with a spectral dispersion system, is on board the currently inactive Skylab orbital assembly. The scanner assembly used a rotating mirror to perform conical scanning of the image being viewed, with a cone angle of 5.5° about the nadir. The spectrally dispersed electromagnetic energy received from the Earth surface simultaneously irradiated 13 detectors. Each detector responded to a different spectral region, and all 13 detectors covered spectral regions within a range of 0.41 to 12.5 micrometers. The scan pattern covered a swath of the Earth surface approximately 72 kilometers (39 nautical miles) wide and any desired length along the groundtrack of the Skylab spacecraft. Depending on the band, each scanning arc was subdivided into 1240 or 2480 resolution elements or cells (sometimes called pixels or picture elements). The instantaneous field of view was 79 meters (260 feet). Approximately 95 scan lines of data were collected each second.

The detector elements produced an electrical output signal resulting from detector response to the average radiance over a 0.182-milliradian field of view. The detector output was sampled only during the front 118° of the 360° scanning cycle. During part of the unused portion of the scanning cycle, the detectors viewed suitable calibration sources.

The output data were digitally encoded in Miller code and recorded on board the Skylab spacecraft on instrumentation magnetic tape, and the tapes were returned to Earth by returning Skylab crewmembers. Bands 3 to 7 and 11 to 13 were sampled at twice the rate of the remaining bands but in synchronization with the bands that were sampled at the lower rate. A more complete description of the system is contained in reference 5.

MULTISPECTRAL SCANNER AND DATA SYSTEM

The multispectral scanner and data system (MSDS) is a 24-band multispectral scanner designed for integration and operation in a C-130 aircraft. The MSDS is an optical mechanical scanner with a spectral dispersing system. The scanner assembly uses a wedge-shaped scanning mirror to sweep out straight scan lines normal to the longitudinal axis of the aircraft and through an 80° swath centered at the nadir of the aircraft. The scanning/sampling system is compensated for roll, but no corrections are applied for pitch or yaw or for drift angle.

The 24 detector elements produce a voltage output that is PCM encoded as a biphase level signal and, after buffering to reduce data rates, is recorded on an instrumentation magnetic tape that is 2.54 centimeters (1 inch) wide. Calibration sources are viewed by the detector elements during the part of the scan cycle in which the scanner is not viewing the Earth surface. Hence, radiometric calibrations are available in each of the 24 data bands for every scan cycle. Reference 6 contains additional details.

MICHIGAN SCANNER

The Michigan scanner was the prototype of all contemporary scanners and has been significantly improved by repeated modifications in response to scanner technology development. The Michigan scanner is flown in a C-46 aircraft. The system is equipped with a wedge-shaped scanning mirror. The radiant energy from the Earth surface enters two apertures and is dispersed onto 18 detectors covering the visual and infrared portions of the electromagnetic spectrum. The scanning system views the Earth through an 80° swath centered at the nadir of the aircraft. Signals from any 12 of the 18 detectors, together with the necessary pulses to permit the removal of track-to-track skew introduced by the onboard tape recorder, may be recorded as analog signals on magnetic tape.

It is significant to note that the data from the Michigan scanner system are analog recorded, whereas the other scanner systems described herein convert the signal, by analog to digital conversion electronics, before the signal is recorded on magnetic tape. This recording technique makes the Michigan scanner data compatible either with analog computers directly or with digital computers after the data are preprocessed by an analog to digital converter in a ground-based data processing system.

APPENDIX B

EARTH RESOURCES LABORATORY

DATA ANALYSIS SYSTEM DESCRIPTION

PHYSICAL DESCRIPTION

The Earth Resources Laboratory (ERL) data analysis system (DAS) (fig. 3) was designed with the capability to reduce a broad range of remote-sensor data. The DAS will accept multispectral remote-sensor data in three formats. The data may be read from appropriately formatted, computer-compatible nine-track tapes, from pulse code modulation (PCM) encoded analog magnetic tape, and from photographic film (either black and white or color transparencies).

The DAS output may be recorded on nine-track digital computer-compatible tape for further processing, displayed on a high-resolution color television (TV) monitor, recorded on 24.13-centimeter (9.5 inch) wide color or black and white film, or listed on a high-speed line printer. The basic system includes the following major components.

1. A 14-channel broadband instrumentation tape reproducer/recorder

2. A high-speed, general-purpose, digital miniature computer

3. Two 381-cm/sec (150 in/sec) digital tape recorder/reproducer units

4. A high-resolution, three-color TV display used to provide fast presentations of sensor data for evaluation by the system operator (Data may be presented in true color, in pseudocolor, or in black and white. A light-pen/cursor system is provided for delineation of data areas of interest.)

5. A high-resolution, 24.13-centimeter (9.5 inch) wide strip-film camera system capable of recording data presentations in color or in black and white

6. A versatile interactive operator control console (IOCC) providing complete operator-processor communication and control

7. A multispectral optical (film) input subsystem capable of high-resolution scanning and digitizing of black and white multispectral transparencies or color transparencies

These major system components are coupled to the high-speed, general-purpose digital computer through interface units specially designed for maximum data-transfer rate with minimum software monitoring required for operation.

The 14-channel broadband instrumentation tape reproducer/recorder is a 2.54-centimeter (1 inch) wide tape Ampex FR-2000 with a set of direct-playback electronics components and the associated equalizing networks to allow playback at tape speeds from 4.76 to 304.8 cm/sec (1.875 to 120 in/sec). In addition, there is one set of direct-record electronics components that may be used with the bench test unit (BTU) to generate analog tapes in selected data formats. The FR-2000 reproducer/ recorder is also equipped with a servomotor speed control to play back data tapes at continuously variable tape speeds and to allow matching of the data-input rate to the data-reduction-processing speed.

The high-speed, general-purpose computer is a Varian model 620F, 16-bit word length, equipped with two buffer interface controller (BIC) units; two block transfer controller (BTC) units; two nine-track, 381-cm/sec (150 in/sec) tape controllers; and 16 000 words of memory. An optional instruction set including a hardware multiply/ divide capability is included. A high-speed line printer and a 300-card/min card reader with required controllers are also included in the computer subsystem.

The three-color TV display system uses a Conrac 48.26-centimeter (19 inch) red-green-blue (RGB) monitor with a 48-track data disk-memory unit that is used for monitor display refresh. The display refresh memory provides a constant flicker-free presentation of the displayed data by providing the color or black and white information at standard television line and field rates. The combination of the computer and the data disk memory provides a completely flexible scan conversion system that presents the data at a standard television rate regardless of the scan rate used when the data were acquired.

The display system is implemented with an overriding cursor. The cursor information is stored on the data disk memory on disk tracks separate from the image data, thus leaving the original image data intact. Control of the cursor size and position on the screen may be obtained from input from the numeric display keyboard. In addition, movement and modification of the cursor corners may also be accomplished through a light-pen interface.

The multispectral optical system consists of a high-resolution Beta Instrument flying-spot scanner (FSS), control electronics, and a Perkin-Elmer multispectral optical subsystem. The optical subsystem accepts as many as four black and white film transparencies or one color transparency and has the optics required to separate the color transparency into RGB data. Data are digitized to 1 part in 1023 (10 bits). Figure B-1 is a block diagram of the ERL DAS system showing input/output (I/O) and subsystem signal flow.

The Varian 620F computer system consists of a main frame and two expansion chassis. The system contains the following features.

1. 16 000-word memory

2. Optional instruction set including hard-wire multiply and divide capability

3. Two BIC's providing automated I/O transfer

4. Two BTC's providing high-speed direct I/O transfer with memory

5. Two nine-track tape transports and controllers providing a tape speed of 381 cm/sec (150 in/sec) at 315 bits/cm (800 bits/in.) (12-kilohertz transfer rate)

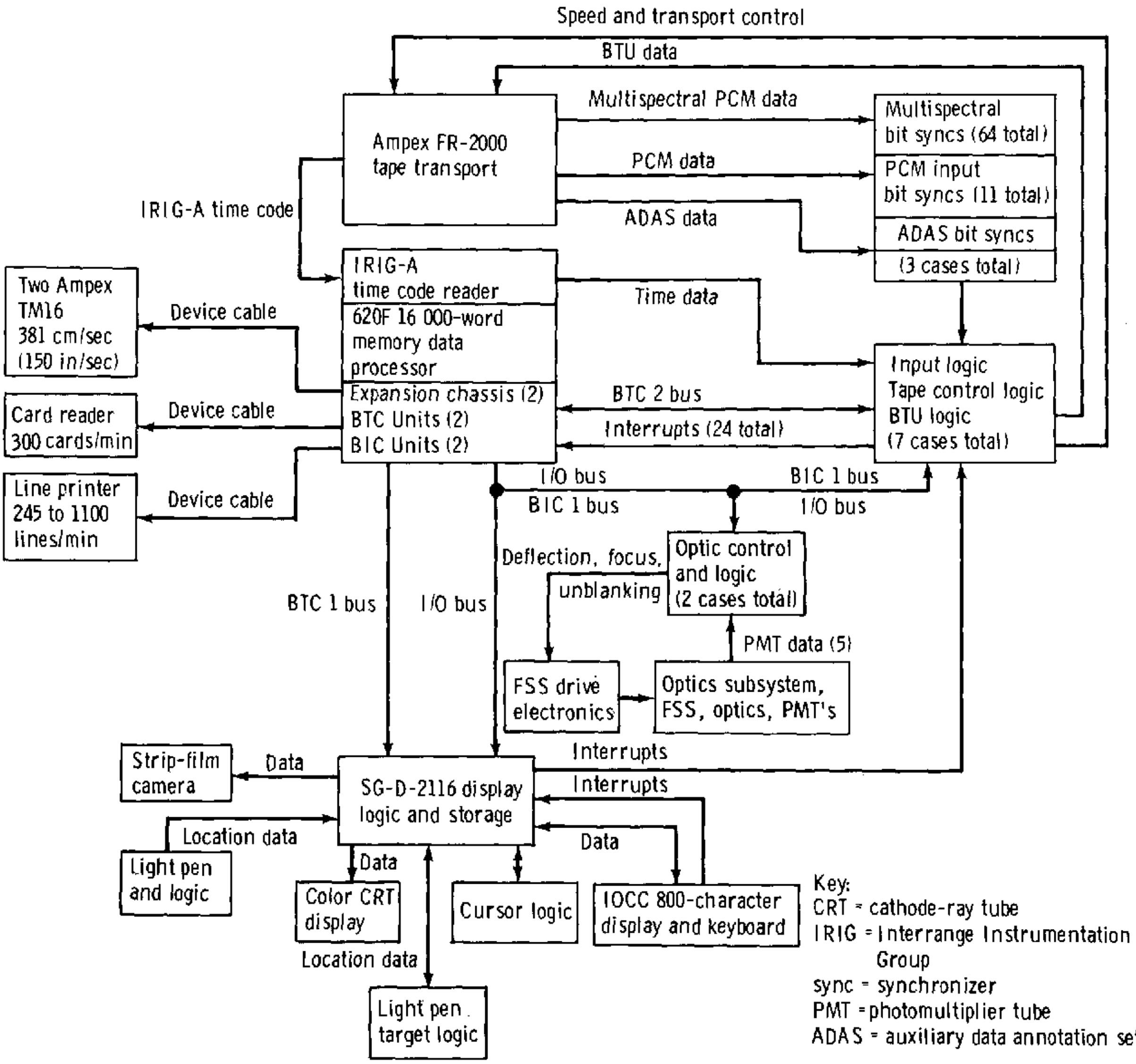


Figure B-1. - ERL DAS block diagram.

6. One 300-card/min card reader and controller

7. One high-speed line printer and controller providing 132 columns, 64 characters at a rate of 245 to 1100 lines/min

8. Three eight-level priority interrupt modules

FUNCTIONAL DESCRIPTION

As indicated by the physical description, the DAS can be used to perform a wide variety of functions. The functions related to land use mapping are basically the following.

1. Reformatting of instrumentation multispectral scanner and data system (MSDS) tape to DAS TAPE format; reformatting of Earth Resources Technology Satellite 1 (ERTS-1) multispectral scanner system (MSS) tape to DAS TAPE format

2. Display of MSDS, ERTS-1 MSS, and Skylab S192 MSS data on the DAS screen for screening and evaluation

3. Display of three selected data bands for training sample selection from the MSDS, the ERTS-1 MSS, and the Skylab S192 MSS

4. Selection of training samples by light-pen cursor and recording of the coordinates and identification codes of training samples on tape for further use

5. Extraction of training sample data from the total data set and tagging with the appropriate identification code name (Training samples are recorded on training sample tape.)

6. Extraction of selected subsets of data bands (best bands for pattern recognition analysis) from the complete data set based on inputs from program CHOICE

7. Display of color-coded land use map on display screen for review purposes

8. Recording of color-coded land use map, as produced by programs R-COLOR and R-CLASS, on color-coded film (This is a tape-to-film conversion.)

9. Correction for crosstrack component of Earth rotation for spacecraft-acquired MSS data

A complete description of the ERL DAS is contained in reference 10.

APPENDIX C

PATTERN RECOGNITION SOFTWARE FUNCTIONAL SYNOPSIS

The current operational Earth Resources Laboratory (ERL) multispectral pattern recognition system can be divided generally into operations performed by use of the data analysis system (DAS) and operations performed by use of the software designed for the Univac 1108 computer. This appendix provides a functional description of six Univac 1108 software modules, which serve to train the data processing system, to classify the data based on initial training, and to produce an output product that can be displayed by means of cathode-ray tube (CRT) or photographic film by the DAS. Fundamental to the ERL pattern recognition system is the assumption of Gaussian statistics.

MODULE STAT

The module STAT extracts the data used to train the pattern recognition system. The training data correspond to categories of known surface conditions of interest to users of the pattern recognition system. For each training sample and for each category, the module STAT computes the average and spread for each spectral band of data and the relation between the various spectral bands. That is, the STAT module computes the mean and standard deviation for each spectral band and the covariance and correlation matrix for the data set corresponding to a training sample or category. For each training sample or category, the module STAT produces a plot in histogram form of each data set, a plot based on means and standard deviations of each signature, and a plot based on means and standard deviations of the relation of each training sample to the appropriate category. Finally, the module STAT is capable of grouping prespecified training samples into categories and of providing the described plots for each grouping. A detailed description of the STAT module is available from the ERL in an informal document written by J. W. Skipworth in 1972 entitled "ERL Configuration III - Pattern Recognition." The information produced by the module STAT is used to evaluate the training sample data, to aid in grouping training samples into appropriate categories, and to produce the training information used in the channel selection module CHOICE and the training module SUPER-T/R-TABLE.

MODULE CHOICE

The fundamental quantity computed by the module CHOICE is a measure, called divergence, of the difficulty of discriminating between two training samples or between two categories. The module CHOICE was developed to choose a subset of spectral data channels that would best support data classification. This choice is necessary because of the expense involved in the use of a large number of data channels in data classification and is made by determining the channel subset that will support the smallest number of classification errors. The module CHOICE ranks the channel subsets by three different criteria: the best average divergence between all categories taken pairwise, the best ratio of the products of all pairwise divergences in a given channel subset with the products in any other channel set, and the best divergence between the two categories that are the most difficult to separate. A detailed description of the CHOICE module is available from the ERL in an informal document written by W. C. Jones in 1973 entitled "CHOICE: 36 Band Feature Selection Software with Applications to Multispectral Pattern Recognition." This document is also identified as ERL Report No. 059.

The divergence measure computed by the module CHOICE also aids in grouping training samples into potential categories. The divergence measure is an indicator of training samples that are of the same category but are more readily handled as separate categories. When such training samples are identified, the module STAT is used to produce training information based on the new categories for use, again, by the CHOICE and SUPER-T/R-TABLE modules.

MODULE SUPER-T/R-TABLE

The module SUPER-T/R-TABLE uses the information generated by the module STAT and the assumption of Gaussian statistics to produce a table for each category. The construction of the table for each category is based on the four-channel subset of all available spectral data channels indicated by module CHOICE as being optimum for discriminating between the given category and all other categories. The table for each category contains a representation of each data element that most probably belongs to the given category and that belongs to the given category at a probability level greater than some prespecified level.

MODULE R-CLASS

The module R-CLASS uses the tables generated by the module SUPER-T/R-TABLE to classify each data element presented to the module R-CLASS. The module R-CLASS performs the classification by sequentially searching the tables for each category until the representation of the data element to be classified is found. If the representation is not found, the data element in question is not classified. The result of the classification is recorded on a magnetic tape, referred to as a MAPTAP.

MODULE REPORT

The module REPORT extracts from the MAPTAP the results of the classification for prespecified regions and computes the number and percentage of data elements classified in each category.

MODULE R-COLOR

The module R-COLOR generates, from the classification results recorded on the MAPTAP, a set of color-coded classification results that can be displayed on the data analysis station CRT or film recorded on the data analysis station film recorder. When the film recorder is used, the color-coded classification is scaled, corrected for equal angular scanner sampling as opposed to equal target distance sampling, and corrected for either scanner oversampling or undersampling. Detailed descriptions of programs R-CLASS, REPORT, and R-COLOR are provided in the previously described ERL document entitled "ERL Configuration III - Pattern Recognition."

APPENDIX D

PATTERN RECOGNITION: AN ILLUSTRATIVE EXAMPLE

The objective of this appendix is to illustrate the Bayesian maximum-likelihood decision technique on data familiar to everyone. The decision technique is very general in its possible applications. The steps taken in this discussion are also taken in applying the pattern recognition computer programs for classifying multispectral scanner data.

STATEMENT OF THE PROBLEM

Height and weight data on 10 male and 10 female subjects are given in table D-I. Height and weight data on 1000 subjects of unknown gender are assumed. The task is to classify the 1000 subjects as male or female by using only one of the measurements (either height or weight) in the classification decision. The analogies in table D-II will be useful in analyzing the problem.

TABLE D-I. - HEIGHT AND WEIGHT DATA FOR 10 MALES AND 10 FEMALES

Subject no.	Height		Weight		Subject no.	Height		Weight	
	cm	in.	kg	lb		cm	in.	kg	lb
Male					Female				
1	178	70	79	175	11	165	65	54	120
2	196	77	100	220	12	158	62	57	125
3	168	66	66	145	13	170	67	64	140
4	180	71	91	200	14	170	67	73	160
5	170	67	70	155	15	165	65	56	123
6	191	75	90	198	16	160	63	48	105
7	175	69	75	165	17	158	62	63	138
8	188	74	109	240	18	160	63	50	110
9	183	72	82	180	19	178	70	64	140
10	188	74	86	190	20	160	63	54	118

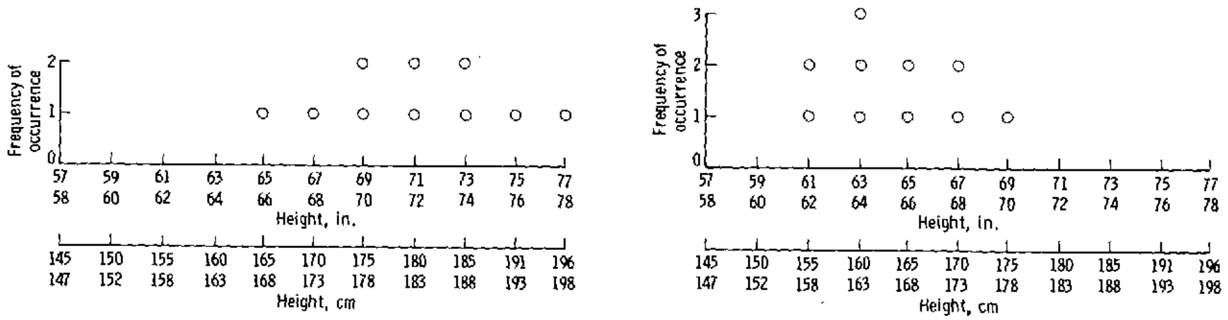
TABLE D-II. - ANALOGIES FOR PATTERN RECOGNITION ANALYSIS

Illustrative problem	Multispectral pattern recognition
Subject	Picture element or remote-sensing unit
Height measurements	Channel 1 readings
Weight measurements	Channel 2 readings
Male	Class 1
Female	Class 2
10 known males	Training field 1
10 known females	Training field 2
1000 subjects (unknown gender)	Raw data
Deciding male or female	Classification
Choosing either height or weight	Feature selection
Multiple measurement plot	Spectral plot

DATA ANALYSIS AND STATISTICS

Histograms

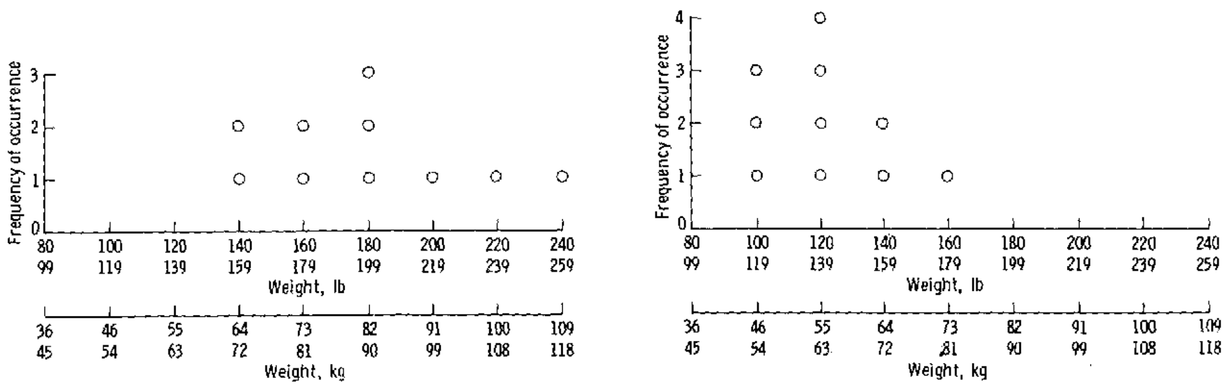
The height and weight data for the 10 males and 10 females are plotted in histograms. The frequency of occurrence of heights within intervals (i.e., a histogram) is illustrated in figures D-1(a) and D-1(b); the weight data are illustrated in figures D-2(a) and D-2(b). The histograms are useful in obtaining an overall view of how the data are distributed within a class and from one class to another.



(a) Male.

(b) Female.

Figure D-1. - Histograms of height measurements for training data.



(a) Male. (b) Female.

Figure D-2. - Histograms of weight measurements for training data.

Means and Covariance Matrix

The means and the covariance matrix (fig. 5) are computed for the 10 males and 10 females. For the 10 males, the mean height is 181.6 centimeters (71.5 inches) with a standard deviation of 9.1 centimeters (3.6 inches); the mean weight is 85.0 kilograms (187.3 pounds) with a standard deviation of 13.3 kilograms (29.3 pounds). The covariance matrix is

	Height	Weight
Height	12..72	--
Weight	90.06	857.34

For the 10 females, the mean height is 164.3 centimeters (64.7 inches) with a standard deviation of 6.6 centimeters (2.6 inches); the mean weight is 58.0 kilograms (127.9 pounds) with a standard deviation of 7.5 kilograms (16.5 pounds). The covariance matrix is

	Height	Weight
Height	6.90	--
Weight	25.41	273.66

The mean (average) height of the 10 males computes to 181.6 centimeters (71.5 inches); a reference to the histogram (fig. D-1(a)) will show that the height data for the 10 males tend to cluster around this value.

The covariance matrix indicates the extent of data scatter. In figure 5, the diagonal terms are the variances and the off-diagonal elements are the covariances. The variance of the height measurements for the 10 males is 12.72. The standard deviation is 9.1 centimeters (3.6 inches) (the square root of the variance). The histogram for male heights (fig. D-1(a)) shows that more than one-half of the height data for the 10 males falls between 172.4 and 190.7 centimeters (67.9 and 75.1 inches). These values were determined by computing the mean height minus one standard deviation and the mean height plus one standard deviation. The covariance of height with weight for the 10 males is positive (i.e., 90.06). This indicates that height and weight are positively correlated. That is, if a man is heavier than average, he is expected to be taller than average; similarly, if he is lighter, he is expected to be shorter.

Multiple Measurement Plot

The average height plus and minus one standard deviation for both male and female is plotted on the left of figure D-3; on the right, the average weight plus and minus one standard deviation is plotted. An examination of this figure shows that the two classes are slightly more separated when weight measurements rather than height measurements are used.

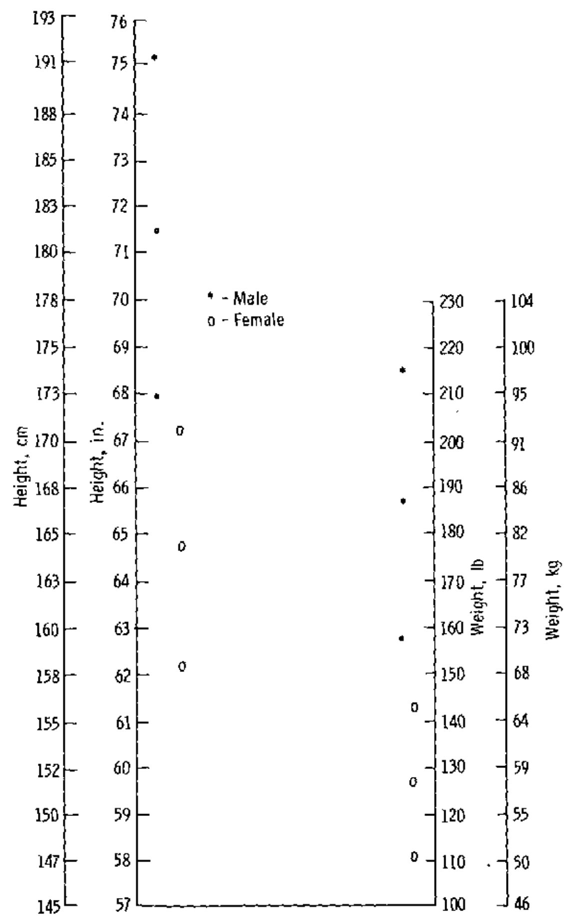


Figure D-3. - Multiple measurement plot. Values are means plus and minus one standard deviation.

Divergence

One method of choosing a measurement is interclass divergence. The assumption is usually made that the data are normally distributed, and this assumption will be made in the divergence calculation and, later, in the classification decision. The measurement that produces the highest interclass divergence is usually considered the most desirable. With the divergence criterion of 18.5 for weight and 10.7 for height, and with allowances for only one measurement, weight would appear to be the "better" of the two measurements.

Classification

Under the assumptions that the occurrence of a male or a female is equally likely and that the data are normally distributed, the Bayesian maximum-likelihood decision rule can be applied to classify the data. The value of $P(X|M)$ is calculated; this parameter can be considered to be the probability of the occurrence of a weight X, given the fact that a male is being weighed. Similarly, $P(X|F)$ can be considered to be the probability of occurrence of a weight X, given that a female is being weighed. The maximum-likelihood decision rule dictates that if $P(X|M)$ is greater than $P(X|F)$, then the subject must be classified as a male. If $P(X|F)$ is greater than $P(X|M)$, then the subject must be classified as a female. Table D-III shows how the 10 known males and females were classified, based on their weight, using this decision rule. The results could be summarized in a "report card," as shown in the following table.

Known	Were classified as —	
	Male	Female
Male	9	1
Female	1	9

Because these results are based on the classification of the training data, they can be expected to be an upper limit on the classification accuracy of the raw data.

It is interesting to note that had height data been used instead of weight data, three subjects (3, 5, and 19) would have been misclassified rather than two (3 and 14). Now, the decision rule can be applied to the 1000 subjects of unknown gender as shown in table D-IV.

TABLE D-III. - CLASSIFICATION OF TRAINING DATA USING WEIGHT MEASUREMENT

$$\left[P(X|M) = \frac{1}{\sqrt{2\pi(857.34)}} e^{-\frac{1}{2}\frac{(X-187.3)^2}{857.34}} ;\ P(X|F) = \frac{1}{\sqrt{2\pi(273.66)}} e^{-\frac{1}{2}\frac{(X-127.9)^2}{273.66}} \right]$$

Subject no.	Weight		P(X M)	P(X F)	Class
	kg	lb			
1	79	175	0.01247	0.00042	M
2	100	220	.00730	$<1 \times 10^{-8}$	M
3	66	145	.00480	.01413	F
4	91	200	.01240	$<1 \times 10^{-5}$	M
5	70	155	.00741	.00630	M
6	90	198	.01274	$<1 \times 10^{-5}$	M
7	75	165	.01019	.00195	M
8	109	240	.00270	$<1 \times 10^{-11}$	M
9	82	180	.01321	.00017	M
10	86	190	.01316	$<1 \times 10^{-5}$	M
11	54	120	.00097	.02152	F
12	57	125	.00142	.02375	F
13	64	140	.00370	.01846	F
14	73	160	.00882	.00367	M
15	56	123	.00122	.02308	F
16	48	105	.00026	.00925	F
17	63	138	.00330	.02002	F
18	50	110	.00042	.01343	F
19	64	140	.00370	.01846	F
20	54	118	.00081	.02016	F

TABLE D-IV.- CLASSIFICATION OF 1000 SUBJECTS OF UNKNOWN GENDER

$$\left[P(X|M) = \frac{1}{\sqrt{2\pi(857.34)}} e^{-\frac{1}{2}\frac{(X-187.3)^2}{857.34}} \ ; \ P(X|F) = \frac{1}{\sqrt{2\pi(273.66)}} e^{-\frac{1}{2}\frac{(X-127.9)^2}{273.66}} \right]$$

Subject no.	Weight		P(X|M)	P(X|F)	Class
	kg	lb			
21	45	98	0.00013	0.00471	F
22	74	163	.00966	.00254	M
23	62	136	.00294	.02139	F
24	67	148	.00554	.01153	F
...	...	...	...	...	...
1018	69	152	.00659	.00835	F
1019	77	170	.01144	.00095	M
1020	99	218	.00786	$<1 \times 10^{-8}$	M