THE APPLICATION OF AUTOMATIC RECOGNITION TECHNIQUES IN THE APOLLO IX SO-65 EXPERIMENT

by

R. B. MacDonald
Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, Indiana

INTRODUCTION

Photographic materials can be viewed as a chemical detector of electromagnetic energy over certain wavelength regions. While film emulsions are notoriously difficult to control and calibrate, they do offer a means of measuring the spectral variations of reflected energy from a scene in addition to providing measurements of the spatial distribution of these energies. Either multiple emulsion films such as conventional color and color infrared or multiple camera and multiple lens camera systems employing different film filter combinations can be utilized to record energies in many multispectral bands over a spectral interval from approximately .4 to 1.0 micrometers.

A four-camera 70mm sensor system was carried by NASA Apollo IX. Researchers at Purdue University's Laboratory for Applications of Remote Sensing selected a simultaneously exposed four-frame set of photographic frames collected by Apollo IX of the Imperial Valley, California region. The primary objective of the analysis of these data was to study problems in analyzing multiband satellite photography of Earth's surface features using digital machine processing techniques. Specific objectives included the development of techniques for getting quantitative spectral measures from such photographic sensing materials and for data handling, storage and retrieval required to handle the four to five million numbers per band contained in each frame. The analysis described required various steps which fall into three major phases:

- data preprocessing
- training class and signature analysis
- automatic classification and result evaluation

The application objectives for this research were to determine which surface features could be identified with "ERTS-like" measurements collected with these photographic materials. Three categories of surface features in the southern California region were studied. These were crops, soils, and geological features. The spectral identifiability of

surface features is, in this case, dependent on the uniqueness of a spectral response of the various features in the three wavelength bands available from the photographs.

DATA PREPROCESSING

The Apollo IX SO-65 experiment utilized four 70mm Hasselbland cameras having 80mm focal length lenses mounted together to view the same area. Film filter combinations were selected to obtain exposed frames in the green, visible red and reflective infrared portions of the spectrum. Three of the cameras contained black and white film and used filters to discriminate between the spectral bands. A fourth camera contained color infrared film which had dyes sensitive to approximately the same bands as the three black and white cameras. Thus, the experiment gathered image data in three bands by two methods. The approximate band sensed and the film data are listed in Table I.

Densitometric measurements were made over each frame with a 25 micron aperture at sampling intervals of 25 microns. This 25 micron area corresponded to a resolution element on the ground at nadir of 200 feet by 200 feet. The film density was measured using a linear scale, digitized, and recorded on magnetic tape for computer processing. The three multispectral black and white films were scanned using white light. The color infrared film was scanned sequentially using blue, green, and red filters and no filter in the white light path. In this fashion, seven representations of the frames were obtained on tape and consisted of a total of over 35 million density readings. A computer gray-scale printout of the white light scan of the color infrared frame is presented in Figure 1.

In order to analyze these data using multispectral pattern recognition programs, each of the raster scans must be registered in such a manner that each scene point is in geometric coincidence in all the scans. This image registration problem has been studied at LARS and a software system has been developed to register or "overlay" multiple images of the same scene. Following these procedures, the seven scans were brought into registration and a film data storage tape was compiled for subsequent analysis.

TRAINING SAMPLE SELECTION

Several methods are in use at LARS for studying imagery to determine separable classes in given data. The statistical multispectral pattern analysis of LARSYSAA developed at LARS supplies several methods of class

selection. The system can compute the multidimensional first and second order statistics for up to 30 wavelength bands. These statistics are printed out in the form of histograms, correlation matrices, and coincident 3 sigma spectral plots. Any of these forms can be used to compute the pairwise separability between all classes. The analyzer can also be used to select the preferred set of wavelength bands to be used for classifications based on this divergence criteria. A third method of class separation utilizes clustering techniques to group image points around cluster centers in the multidimensional space such that the overall variance of the resulting sets is minimized. Visual histogram analysis and clustering were used to determine separable classes in the SO-65 imagery.

AUTOMATIC CLASSIFICATION AND ANALYSIS

A synoptic feature analysis was made on a global scale over the entire 10,000 square mile area initially to develop a computer derived map of the following features: clouds, cloud shadow, water, rock, bare soil, sand, green vegetation and salt cliffs. Training data was obtained for the classes by visual inspection of the gray-scale pictorial printouts and large prints of the film transparencies. Clustering was used to find subclasses of the global classes. Also, geological maps of the area were helpful in choosing training "fields" for rock and alluvial deposits. One can appreciate the accuracy of this analysis by inspecting the synoptic feature map shown in Figure 2. Every 6th point was processed and classified in this effort. At this resolution the only water discernible was in a portion of the Salton Sea. Also note that large deposits of salt were discernible on this synoptic map.

Next, SO-65 Test Site 15-A, commonly referred to as the Dogwood Area in the Imperial Valley, California region was classified into categories of vegetation, bare soil and water. The results of this classification are shown in Figure 3. Note the large 1/2 x 1 mile body of water correctly classified in the upper central portion of the figure. A quantitative assessment of the accuracy of this classification is shown in Table II.

The SO-65 Imperial Valley experiment included sufficient ground truth to permit LARS investigators to make a "greater detail" analysis of two selected areas. These were areas 15A, the Dogwood area and 15D, the McCabe Road area. Certain crop types such as carrots, lettuce and onions have very large percentages of soil background which were present in very small amounts and deleted for the study. The major cover types in the agriculture test site are barley, alfalfa, sugar beets, bare soil and salt flats. Water is present in very small quantities, but is easily distinguished and was included in the study.

Samples of data from each of the classes defined were selected and used for training the pattern classifier. The number of training samples used for each class is selected on the basis of previous experience performing similar classifications on other agricultural data. Test site 15A was represented by 66,600 data points and the training samples selected totaled 1132, or 1.7% of the total. The data run for Test Site 15B contained 27,300 points and 788 training samples for 2.8% of the total taken for this area. The statistics processor of the pattern classification system was used to compute the training statistics for each of the classes selected. These statistics consist of the means and covariance matrices to be used by the Gaussian maximum likelihood classifier.

Two methods of statistical pattern recognition were used to classify these points. One method classified each image point into one of the defined classes. The other classified an entire data field with one decision. A great advantage of the "per field classification scheme" is speed. A disadvantage is that the field coordinate must be determined and fed to the classifier before any classification can be performed. At least this is true unless an automatic boundary-finding algorithm can be derived. Such an algorithm has not yet been perfected. The per point classifier can classify any data without field boundaries being specified, but it is a time consuming process since every resolution element and image is classified individually. A color coded computer printout of the analysis results of the Dogwood Area is shown in Figure 4. The areas included in the small computer designated rectangles are the areas used for either training or test purposes. These particular printouts are the results of the per point classification of the Dogwood area. Inspection of the Dogwood Road results reveals that the accuracy for certain classes is relatively insensitive to the channels used. Bare soil and barley show relatively constant accuracy for all cases where two or more channels are used. The accuracy for sugar beets recognition was good for all four channels and the alfalfa performance was poor for all cases. Salt flats and water were recognized with high accuracy for all cases. An interesting result was the high accuracy achieved for some of the classifications using only one or two channels. The highest accuracy for barley, 82.1%, was achieved using only one channel. The peak for bare soil was near 100% using the green channel, the infrared channel and the visible red channel. Equal maximums for several channels were seen for alfalfa.

A color coded computer printout of the analysis of Test Site 15D, the McCabe Road area, is shown in Figure 5. It is interesting to note that in nearly all cases the per field classifier performed better than the per point classifier.

It was generally concluded on the basis of both classification results that general cover types are readily recognizable by three-band

multispectral photographic sensor and pattern classification system. More specific breakdowns of crop types are more difficult and in some cases impossible. These results were obtained, however, with uncalibrated photography having considerable overlap of the spectral bands, and probably containing other distortions in the spectral data. Improved photographic fidelity with more and better separated spectral bands could result in significantly better accuracies.

Soil scientists at LARS were interested in attempting to identify surface soils with these data. Soil type identification was pursued in an area near El Centro in the Imperial Valley. A 50-year-old soils map of the area is shown in Figure 6. The soils in this area are generally young, the surface soils and subsoils are very much alike. The material which forms the soil consists primarily of sediments deposited by the Colorado River, and local alluvium from the California coastal range of the Chocolate Mountains. There are 26 soils mapping units, most of these are named for major soil series that occur within each unit. The 26 mapping units are grouped into five major groups based on soil characteristics and qualitites. These are:

- areas dominated by fine texture, very young soil
- areas dominated by medium and course texture, very young soils
- areas dominated by course and very course, very young soils
- areas dominated by young soils with line segregation in the subsoil
- areas dominated by miscellaneous land types

The following four soils types were distinguishable with the film data available:

- dune sand and sand dune designated by the letter D
- mellow land fine sandy loam designated Ml
- Rositas fine sand designated Rf
- Holtville silty clay designated Hc
- Holtville silty clay loam designated Hs

A computer map of the El Centro region is shown in Figure 7. The areas of different soils are designated by the arrows.

Geologists at LARS are currently conducting investigations with these data to determine how well machine processing techniques work with multiband spectral data in producing geological maps.

In the area east of the Imperial Valley, clustering techniques were used to identify nine different geological materials:

- pre-Cambrian
- metasedimentary rocks
- granitic rocks
- tertiary volcanic rocks
- pleistocene volcanic rocks
- nondifferentiable slope material; this is unconsolidated slope material within each rock type as well as some older alluvium in the form of terraces and fans, all of which cannot be distinguished from each other

- alluvium; this is unconsolidated clay, silt, sand and gravel occuring primarily as valley fill and stream wash
- sand; windblown sand from nearby sand dunes
- river

Figure 8 shows a conventional geologic map of the area west of the Imperial Valley. A computer derived geological map is shown in Figure 9. Arrow 1 points out a clearly delineated contact between pre-Cambrian and tertiary rock. (Pre-Cambrian--color coded green, tertiary volcanic rocks--orange.) Arrow 2 indicates a contact between metasedimentary rocks (purple) and tertiary volcanic rock (orange). Since contacts are often places of weakness, they erode faster producing a valley along the contact. The printout indicates a valley (alluvium--gold and yellow) at the approximate location of contact. This could be used in distinguishing the contact itself. Arrow 3 points to where the printout indicates mostly slope material and alluvium with only a few classifications of actual rock. These classifications of rock, however, line up quite well with the ridge of mountain tops indicated on geologic maps. Here is where the actual rock outcrops occur. Arrow 4 indicates a sharp contact with metasediments and tertiary volcanics caused by salt. The printout shows sharp boundaries of tertiary volcanic. Arrow 5 indicates a very good classification between two volcanics. Both are mainly salt. The older is of the tertiary age (orange) and the younger is of pleistocene age (red). It should be noted that while these results indicate that the geologic features of the range are separable, they do not indicate that a specific rock type can be identified in any frame. Recognition of a specific rock type could be complicated by variability of service conditions, atmospheric conditions and other factors. The results of the study are generally encouraging. Good separability of many surface features were achieved and it is expected that large scale automatic surface features surveys can be performed using multiband photography.

ACKNOWLEDGEMENTS

The film data studied was scanned and digitized by Optronics International of Chelmsford, Massachusetts and by Fairchild Space and Defense Systems of Syosset, New York at no expense to Purdue University or NASA. Optronics scanned the black and white photographs and Fairchild performed color separation of the color infrared films as well as scanning the black and white photos. The cooperation of these industrial firms is greatly appreciated. The multiband Apollo IX photos of the SO-65 test site ground truth were obtained by NASA, Earth Resources Division, Manned Spacecraft Center, Houston, Texas. The U. S. Department of Agriculture supplied initial soils information and crops information for the study.

Code	Film	Filter	Band
A	SO-180 Ektachrome Infrared	Photar 15	.51-.89 micrometers
B	3400 Panatomic-X	Photar 58B	.47-.61 micrometers
C	SO-246 B/W Infrared	Photar 89B	.68-.89 micrometers
D	3400 Panatomic-X	Photar 25A	.59-.715

Table I. Film filter combinations used in SO-65 experiment

Classification Summary by Test Classes

Class	No. of Samples	Percent Correct	No. of Samples Classified Into		
			Vegetation	Bare Soil	Water
Vegetation	544	96.3	524	20	0
Bare Soil	192	91.7	16	176	0
Water	72	100.0	0	0	72

Overall performance = 95.9%

Table II. Preliminary Classification Results

Figure 1

Apollo 9 Computer Map of Imperial Valley Region

Figure 2

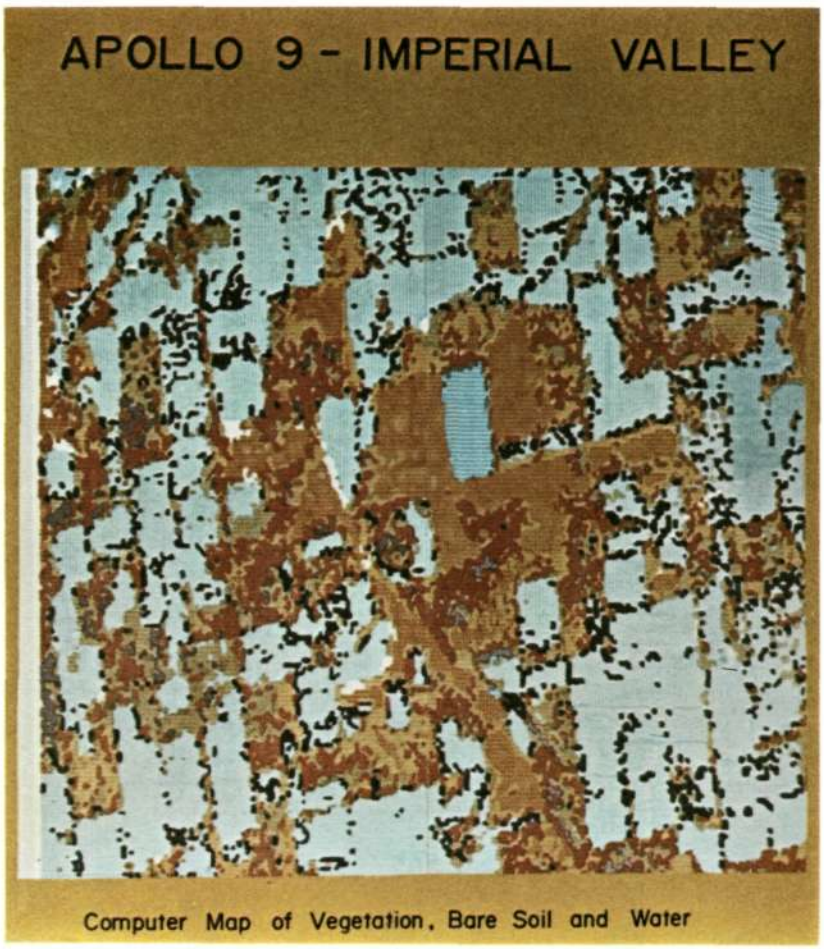
APOLLO 9 - IMPERIAL VALLEY
Computer Map of Vegetation, Bare Soil and Water

Figure 3

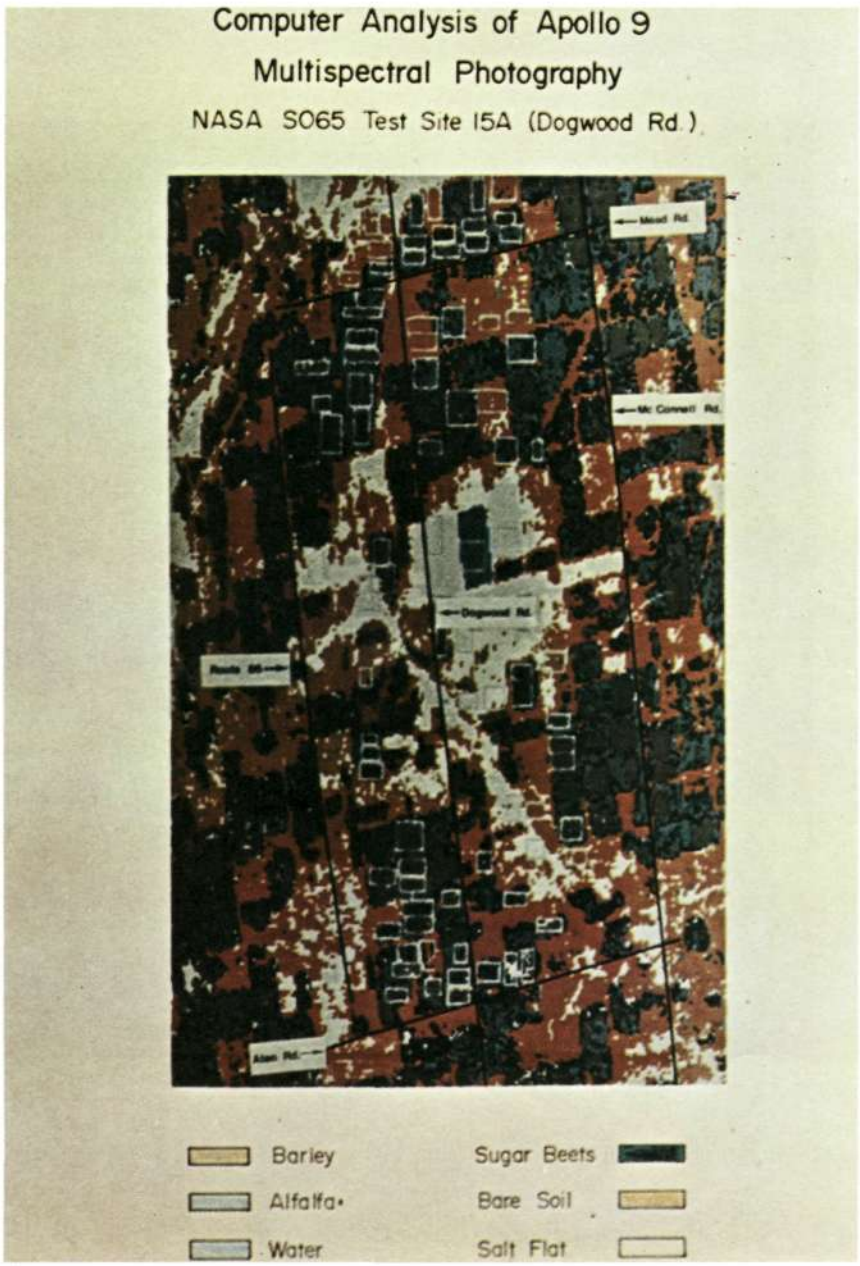
Computer Analysis of Apollo 9
Multispectral Photography
NASA S065 Test Site 15A (Dogwood Rd.)
Barley
Alfalfa
Water
Sugar Beets
Bare Soil
Salt Flat

Figure 4

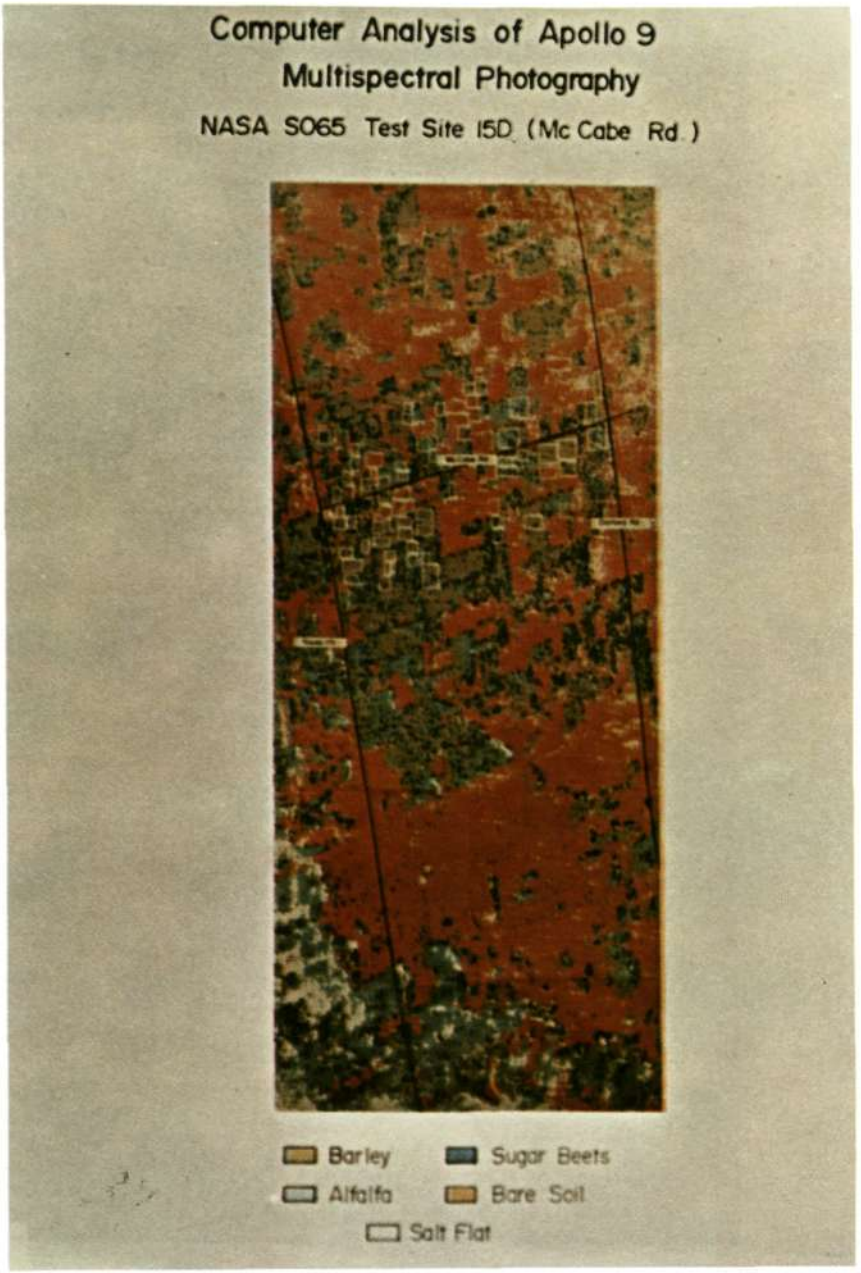
Computer Analysis of Apollo 9
Multispectral Photography
NASA S065 Test Site I5D (Mc Cabe Rd.)
Barley
Sugar Beets
Alfalfa
Bare Soil
Salt Flat

Figure 5

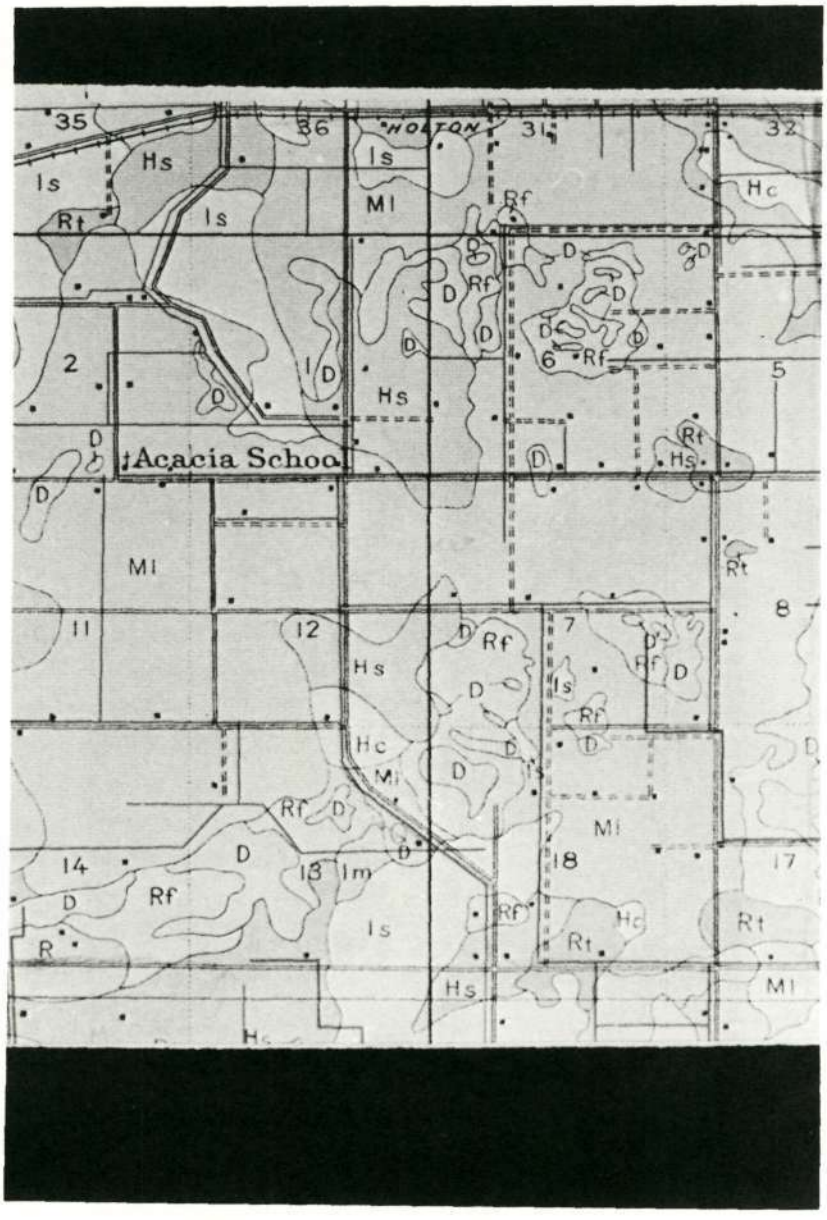
35
36
HOLTON
31
32
Hs
Is
Rt
Is
Is
Ml
Rf
Hc
D
Rf
6
Rf
2
1
D
Hs
5
Acacia Schoo
Rf
Hs
Ml
Rt
11
12
7
8
Hs
Rf
Is
Rf
Hc
Ml
Rf
Ml
14
18
17
Rf
Is
Rf
Hc
Rt
Rt
Hs
Ml

Figure 6

Figure 7

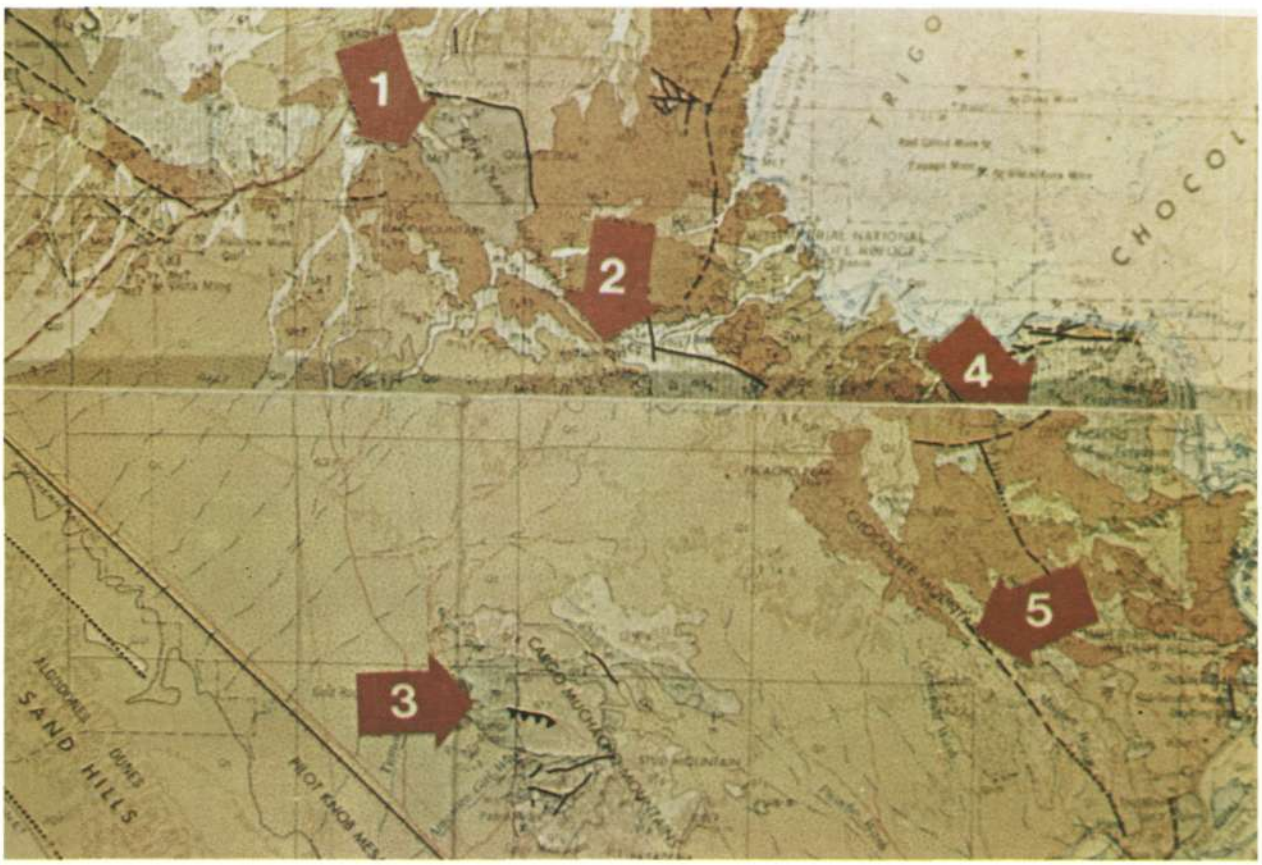

Figure 8

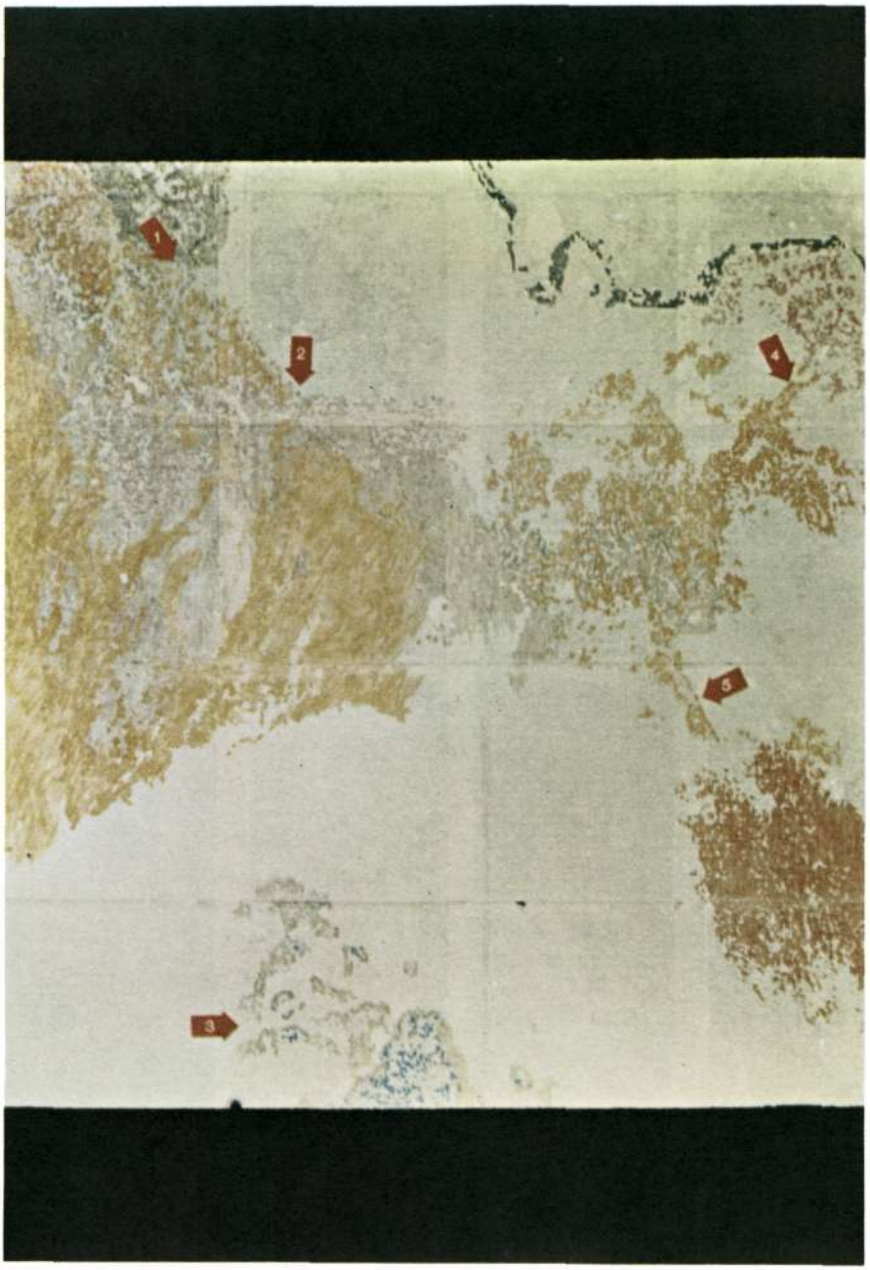

Figure 9