Geographic Considerations in Automatic Cover Type Identification[1]

ROGER M. HOFFER and FORREST E. GOODRICK
Department of Forestry and Conservation,
Purdue University, West Lafayette, Indiana 47906

Abstract

Analysis of multispectral scanner data from Indiana, Texas, and California, showed that the LARS multispectral data processing techniques could be satisfactorily used for data from different geographic locations. The problems of agricultural variability and training sample selection were similar among the three sets of data. All results showed a high degree of accuracy for classification of the various basic and agricultural cover types.

Introduction

The Laboratory for Applications of Remote Sensing (LARS) was established at Purdue University to develop techniques whereby aircraft and spacecraft could fly over large geographic areas of the earth's surface and monitor the type and condition of ground cover below (2, 3). Since large areas were involved, some type of automatic data processing technique was required. Much early effort was devoted toward agricultural problems, but since then LARS has been expanded to include problems of forestry, hydrology, geology, geography, and others.

Currently we use data collected by an airborne multispectral optical-mechanical scanner, which allows the reflected or emitted energy in many discrete wavelength bands to be detected and recorded on magnetic tape. These data are digitized and gray scale maps are printed out by the computer. The maps are similar to a very coarse-resolution photograph of the flightline area. However, the printouts represent the energy being reflected or emitted in only a very narrow portion of the electromagnetic spectrum. Using ground truth data collected at the time of the flight, we located and identified various cover types on the computer printouts, and trained the computer to recognize the spectral pattern of energy in the various wavelength bands for the particular materials of interest. The computer then examined the unknown spectral patterns for the remainder of the area overflown, and statistically classified all the unknown spectral patterns into the various cover type categories which the computer was trained to recognize.

Objectives

One of the more difficult problems involved in automatic classification of vegetative cover and other earth resource materials is the proper selection of adequate and correct training data. Much of

[1] Journal Paper No. 4269, Purdue University Agricultural Experiment Station. This Research was supported jointly by the USDA and NASA.

the difficulty lies in the spectral variability of the vegetation, soil, or water resources. To examine problems which might be encountered for data obtained from widely separated geographic areas, we analyzed data collected in Central Indiana, Southern Texas, and Central California.

The first data studied were from a 70-mile flightline south of Indianapolis (Fig. 1). The particular objectives of this study were: 1) to determine the capability to accurately identify all green vegetation, bare soil, and water in the flightline regardless of natural variability; and 2) to study the variations in spectral response for these three basic cover conditions as affected by instrumentation, atmospheric, and natural conditions of variability.

Figure 1. *Location of the 70 mile flightline along Indiana State Highway 37.*

The second set of data studied came from the area of Moon Lake in Southern Texas, and included various agricultural and forest vegetation, soils in various moisture conditions, and water. The last data analyzed were for an agricultural area in California, and included fields of immature and mature rice, safflower, other vegetation cover, water, and bare soil. The specific objectives in analyzing these sets of data were: 1) to assess the problems involved in selection of training samples from widely separated geographic locations; and 2) to determine the capability for automatic classification of various cover types by using various scanner calibration techniques.

Highway 37 Data Analysis

Multispectral scanner data had been collected in 17 discrete wavelength bands from an altitude of 3,200 feet. At that altitude, the 80° field of view of the scanner covers a swath width of just over

1 mile, thereby resulting in about 70 square miles of area to be classified. For this analysis, only the first 12 wavelength bands (0.4-1.0 μm range) were considered, because of the configuration of the scanner.

For automatic classification, approximately 10 areas for each class (*i.e.* green vegetation, bare soil, and water) were selected for training samples. These included areas of light and dark soils, river and pond water, and a variety of green vegetation conditions including pasture, winter wheat and forest cover. Since the data were collected in late April, the trees were in various stages of leafing out, and agricultural cover such as winter wheat and pasture were quite varied in terms of ground cover. Also, areas of bare soil varied considerably in moisture content. The photo mosaics in Figures 2 and 6 indicate the variability of these three cover types.

In addition to the natural variability, atmospheric conditions or instrumentation changes can cause variations in the scanner spectral measurements. In this analysis, all of the training samples were selected from a relatively small area 20 miles from the north end of the 70-mile flightline. Thus, if variability in spectral response existed throughout the flightline, the spectral measurements in the training sample area would be different from those in the southern portions of the flightline. This would cause the classification results to be less accurate for areas more distant from the training sample area.

Multispectral response graphs were obtained for each cover type and analyzed to determine which combinations of wavelength bands should be used for the classification. The 0.40-0.44, 0.58-0.62, 0.66-0.72, and 0.80-1.00 μm (micrometer) wavelength bands were selected. Figure 2 shows a photo mosaic of one segment of the flightline, along with gray scale printouts of two of the wavelength bands used in the classification.

Marked differences in spectral response may be observed in some of the wavelength bands (Fig. 2, 3). Note especially the 0.80-1.00 μm reflective infrared band, where water is very absorptive and therefore has a relatively low reflectance and low response on the graph. Green vegetation is highly reflective and has a relatively high response on the graph. Figure 3 also indicates that throughout the visible portion of the spectrum (0.40-0.72 μm), green vegetation and water have very similar spectral responses. This helps explain why some of the ponds surrounded by green vegetation are not obvious on the color photography but do appear with striking contrast on color infrared photography.

The entire 70 miles of flightline data were then automatically classified into green vegetation, bare soil, and water categories. Figure 4 shows the classification results for the same area seen in Figure 2. The photo mosaic in Figure 4 was compiled from data flown for planning purposes on April 18, ten days before the scanner data

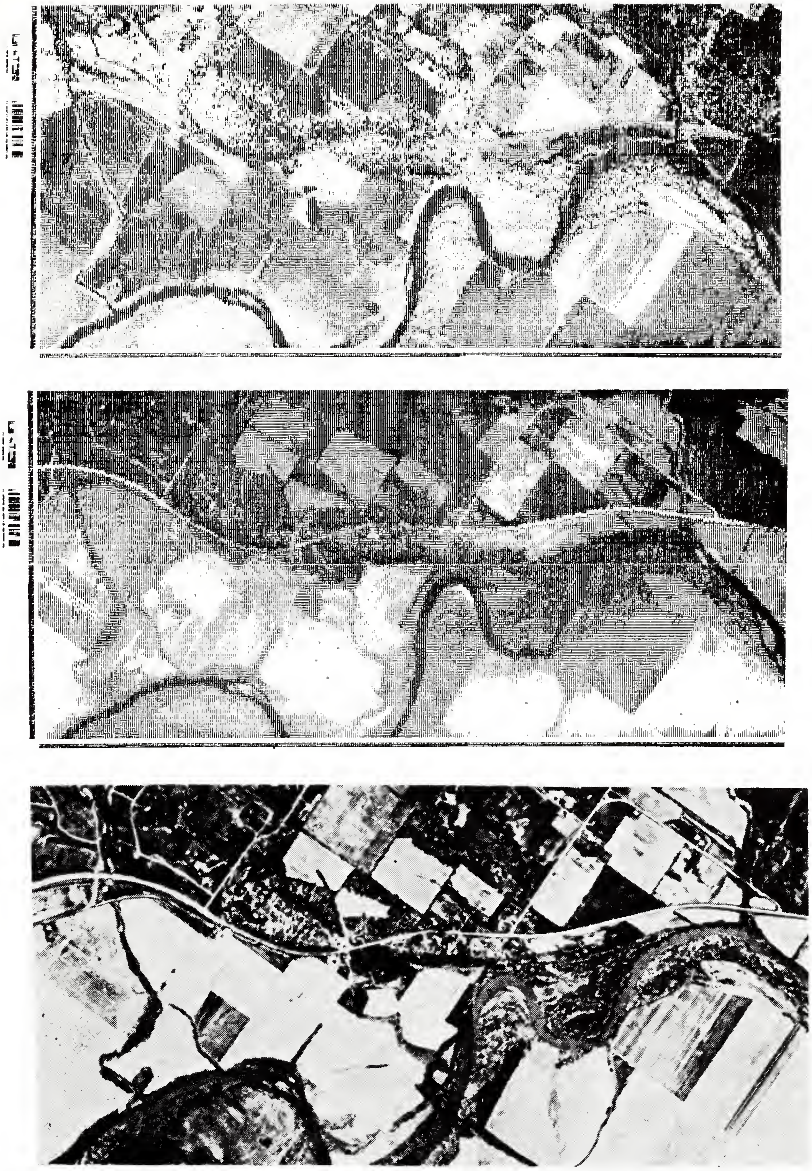


FIGURE 2. *Top, aerial photograph of a segment of Highway 37. Center, gray scale of the same area in the 0.66-0.72 μm wavelength band (visible red). Bottom, gray scale of the area in 0.80-1.00 μm band (reflective infrared). Note reversals in tone between visible and infrared, especially in green vegetation which is highly reflective in infrared.*

were collected. Several areas were covered with green vegetation when photographed, but were plowed (leaving bare soil) between April 18 and 28. Therefore, these particular areas, which appear to be green vegetation on the photos but bare soil on the printout, were indeed correctly classified.

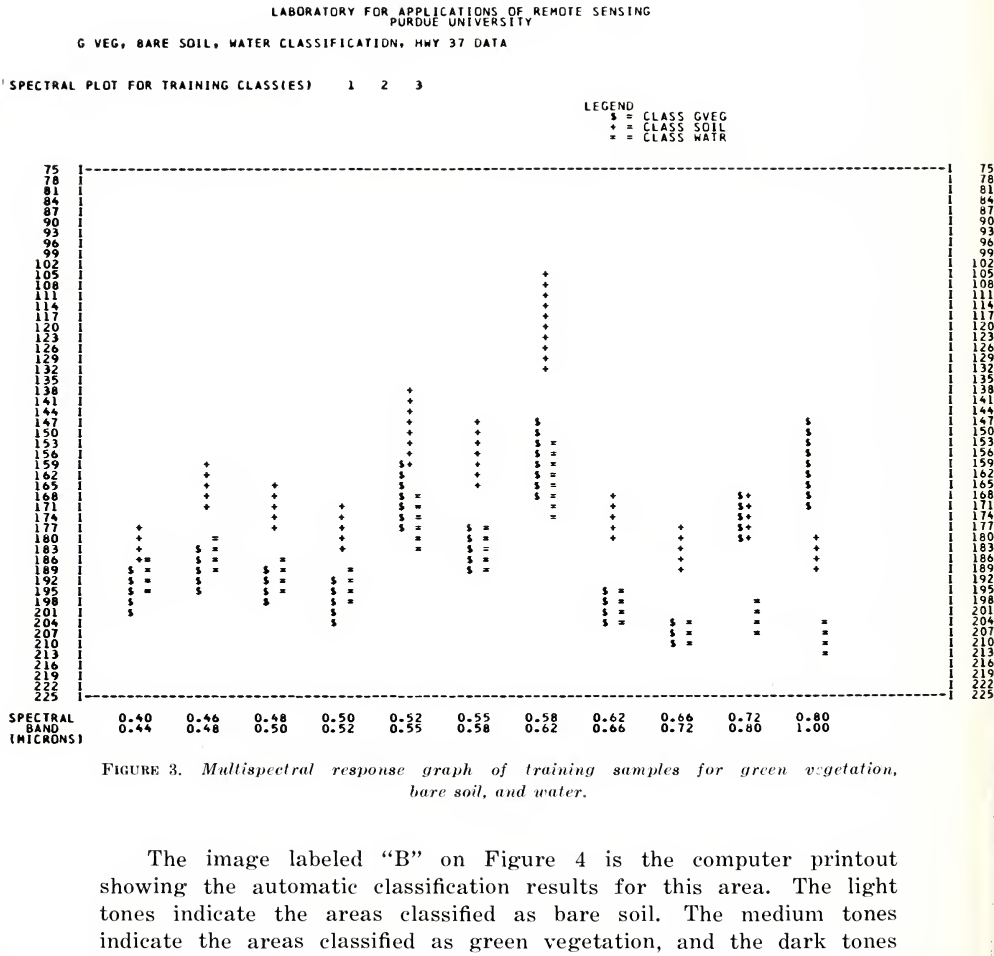


FIGURE 3. *Multispectral response graph of training samples for green vegetation, bare soil, and water.*

The image labeled "B" on Figure 4 is the computer printout showing the automatic classification results for this area. The light tones indicate the areas classified as bare soil. The medium tones indicate the areas classified as green vegetation, and the dark tones indicate water. All other ground cover conditions (such as dry, dead vegetation, roads, rooftops, etc.) have been thresholded and are shown as "blank" areas on the computer printout.

The printouts labeled "C," "D," and "E" in Figure 4 show only the areas classified into green vegetation, bare soil, and water, respectively. These are displayed in this manner to emphasize particular cover types which have been classified and to evaluate the accuracy of the automatic classification.

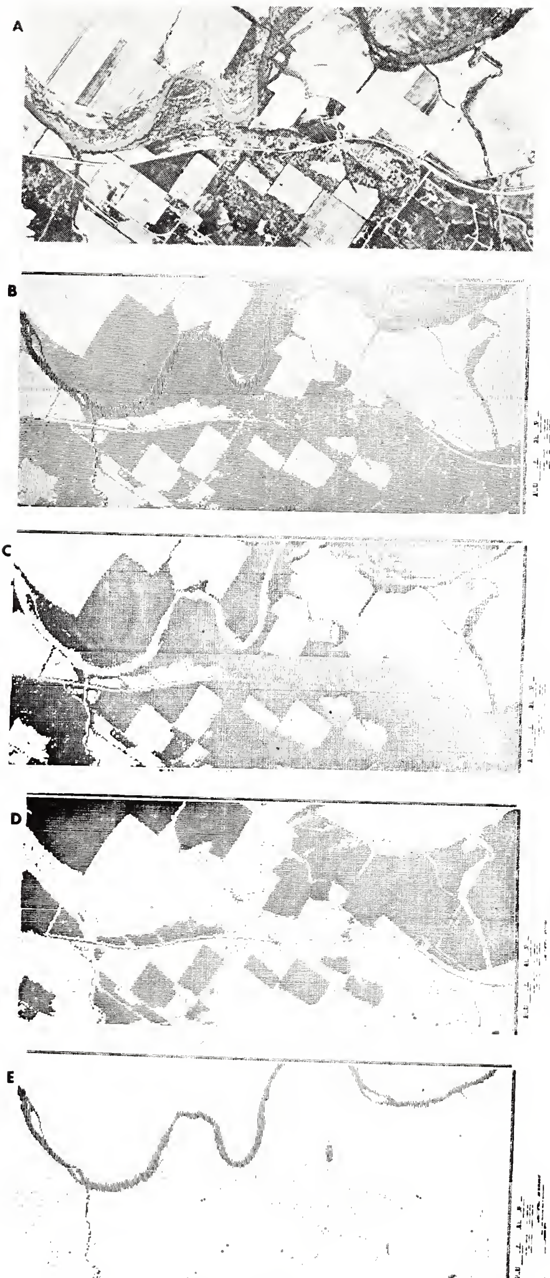


FIGURE 4. *From left: a) aerial photograph of segment of Highway 37 data. b) classification of same area into 3 cover types. Light tone is soil, medium tone is green vegetation, and dark tone is water. c) classification of green vegetation only. d) classification of bare soil only. e) classification of water only.*

Figures 5 and 6 are representative examples of the classification results which show more detail than could be seen in Figure 4. In Figure 5 a hedgerow (Arrow 1) on the photograph was correctly identified on the computer printout on the right. Arrow 2 indicates a field of winter wheat, which was very dense and at a green stage of maturation. The medium tone used to identify green vegetation on the computer printout indicates that this area was accurately classified. The dry, brown vegetation shown at Arrow 3 has been "thresholded" in the automatic classification process, which means that the multispectral response in this area was unlike the response for green vegetation, bare soil, or water. The computer program requires that no symbol be shown for such areas, and they are displayed as a "blank." The roads and houses were also thresholded and appear as "no-symbol" or "blank" areas. Arrow 4 indicates a pond which was correctly classified and indicated by the dark area on the printout.

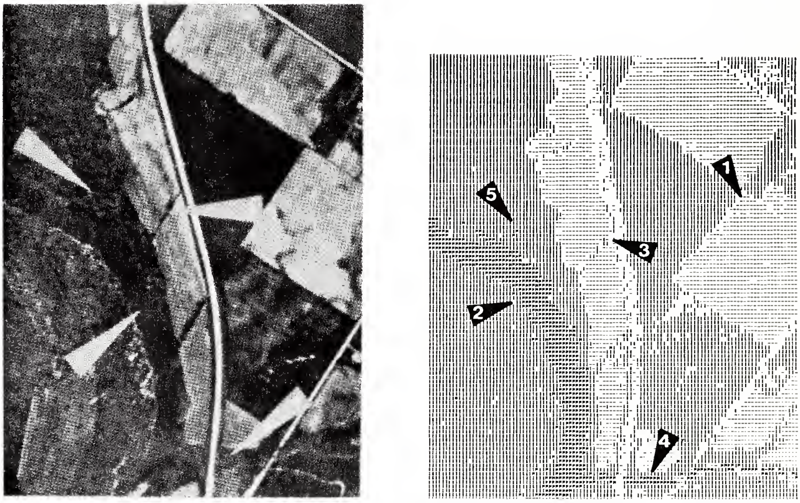
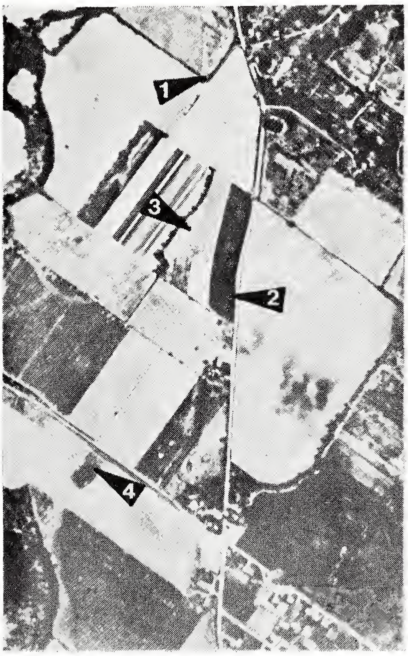


FIGURE 5. *Aerial photograph and classification showing specific features. 1) hedgerow; 2) winter wheat; 3) thresholded (data points not sufficiently like any class); 4) pond.*

In Figure 6, Arrow 1 is at the corner of a field of green vegetation and two fields of bare soil and Arrow 2 refers to the river. Arrow 3 indicates a small hedgerow between the two areas of bare soil which was accurately identified on the printout. A tributary is indicated by Arrow 4. Arrow 5 refers to a very small pond that was indistinct on the color photograph.

Many of the points identified as water appeared as scattered, individual RSU's (Remote Sensing Units or individual computer symbols on the printouts) or as very small groups of RSU's. These were at first thought to be misclassifications. Additional checking revealed ponded water or water in drainage ditches at most of these points.

In several instances, water was correctly identified automatically, but was previously overlooked on the aerial photographs. Several water areas were difficult to see on the aerial photos because of overhanging trees or a lack of distinctive color differences between water and other materials in the vicinity.



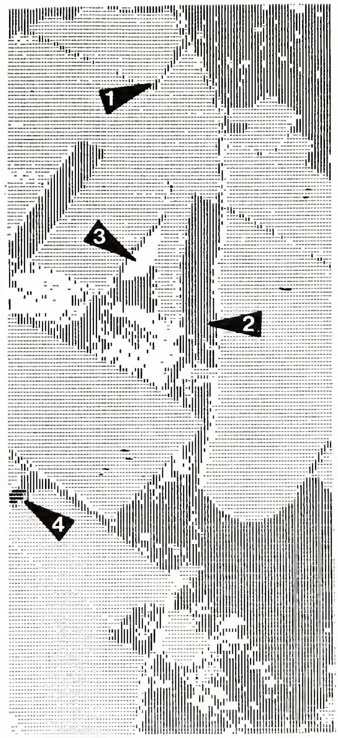


FIGURE 6. *Aerial photograph and classification showing specific features. 1) intersection of green fields and bare soil; 2) river; 3) hedgerow between bare soil areas; 4) small stream; 5) small pond.*

Some problem areas existed in the water classification, primarily because of shadows. For instance, distinct cloud shadows tended to be classified as water. This is because both water and shadows reflect little in the infrared wavelengths. In some forested areas, deep shadows between the tree crowns were also mis-classified as water. However, by placing a "moderately heavy" threshold on the water category, most shadow areas could be thresholded.

Figures 4, 5, and 6 show classification results near the northern end of the 70 mile flightline. Figure 7 shows the classification results for an area south of Bedford, near the extreme southern end of the flightline. The classification accuracies shown indicate that a set of training samples from one portion of the flightline did allow accurate classification for the entire 70 square miles.

To quantitatively check the classification results of the entire area, 89 test sample areas were randomly selected. Based upon study of the photos and the ground truth data collected at the time of the scanner flight mission, 24 of the 89 test areas were determined as belonging to the green vegetation category, 29 as soil, and 36 as water. The average classification accuracy was 99.2% for green vegetation, 97.0% for soil, and 99.7% for water. Table 1 contains the classification results for all the individual RSU's contained in the 89 test sample areas.

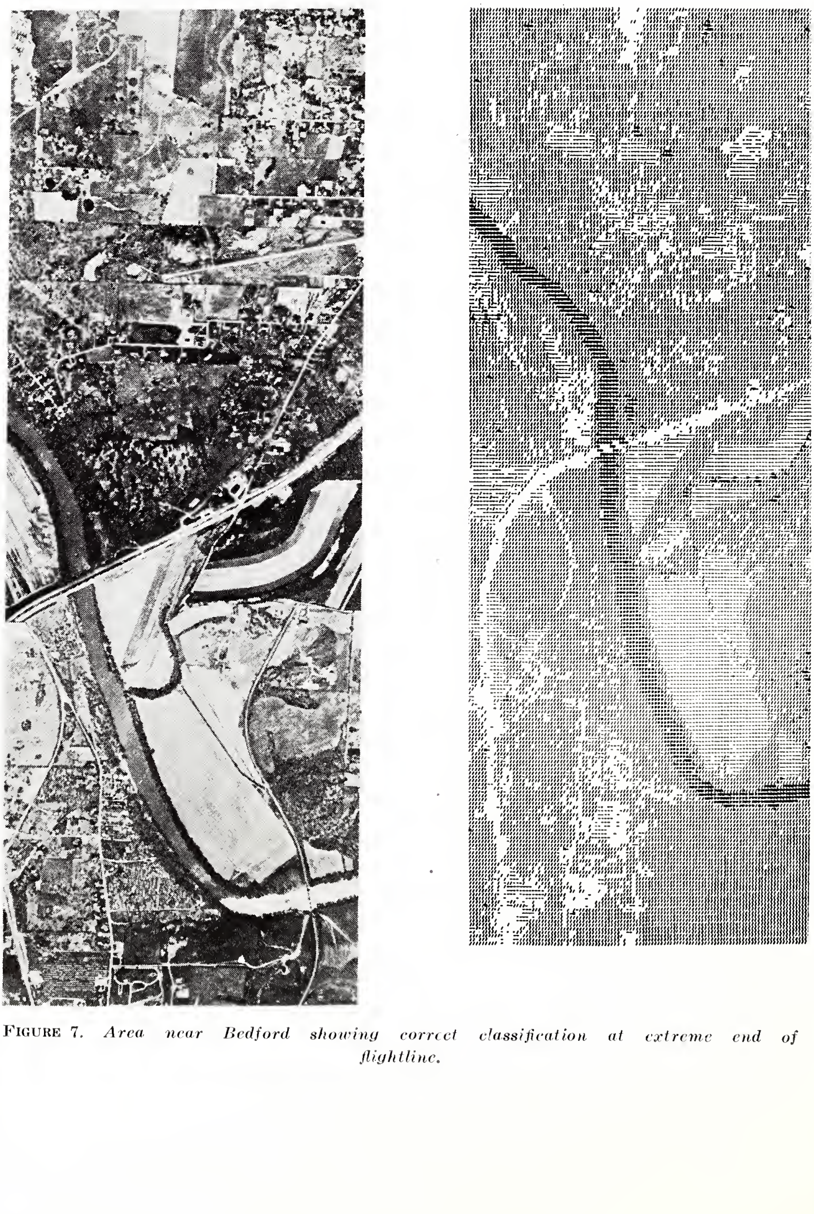


FIGURE 7. *Area near Bedford showing correct classification at extreme end of flightline.*

TABLE 1. *Highway 37 data classification summary, by test classes.*[1]

Class	*No. of RSU's*	*Pct. Correct*	*No. of RSU's Classified Into*			
			Grn. Veg.	*Soil*	*Water*	*Thresh*
Grn. Veg.	1095	99.2	1086	1	5	3
Soil	3165	97.0	2	3012	0	91
Water	887	99.7	1	1	884	1
Total	5087					

Overall Performance = 97.9; Average Performance by Class = 98.6

[1] Serial No. 302005213, (entire 70 mile flightline, using segments). Spectral bands used: 0.40-0.44, 0.58-0.62, 0.66-0.72, 0.80-1.00 μm.

Moon Lake Data Analysis

The next phase of this analysis utilized the Moon Lake data. Five groups of training samples were used to represent all cover types for the automatic classification. These were: water, bare soil, trees, and two spectrally different groups of green agricultural vegetation. The data were classified and the results indicated a 97% accuracy for the training samples.

The classification results appear qualitatively to be very accurate (Fig. 8). Note the narrow strips of vegetation in the field of bare soil near "A," and the field at "B" where vegetation only partially covers the soil. The classification of the area at "B" accurately depicts this combination of bare soil and green vegetation.

Figure 9 shows a close-up view of the classification of an orchard at Point "C." Note the pattern shown for the "T" and "/" symbols, representing the "tree" and "green agricultural vegetation" classes, respectively. The individual RSU's do not indicate individual trees due to the digitization procedures used with the scanner data. Rather, the pattern of RSU's identified as "trees" and as "green vegetation" is a distinct moire pattern (1). However, the number of RSU's identified as "trees" and as "green vegetation" appear in about the expected proportion after viewing the aerial photograph. It is important that the orchard trees were identified as such, since the "tree" training samples were obtained from a forested area beside Moon Lake. These trees had similar physiognomic characteristics but were not the same species.

California Data Analysis

Scanner data collected near Davis, California, were utilized in the last phase of this study. The data were classified, using the following classes: soil (both light and dark), immature rice, mature rice, safflower, and water. Figure 10 shows imagery of the area in the 0.8-1.0 μm wavelength band and the computer printout of the classification results for the entire area, on a crop by crop basis.

The accuracy was 98% for the training fields and 97% for the test fields (Table 2).

TABLE 2. *California data classification summary by test classes.*[1]

Class	*No. of RSU's*	*Pct. Correct*	*No. of RSU's Classified Into* Soil	Immature Rice	Safflower	Rice	Water	Misc. Green	Rotate 90° Thresh.
Soil	4127	99.3	4100	0	0	0	0	0	27
Immature Rice	7942	97.2	5	7722	0	1	35	49	130
Safflower	3316	97.2	0	0	3224	0	0	29	63
Rice	5248	95.5	0	2	0	5013	0	233	0
Total	20633								

Overall Performance = 97.2; Average Performance by Class = 97.3

[1] Serial No. 309019404, (California Test Site). Spectral Bands used: 0.46-0.48, 0.55-0.58, 0.72-0.80, 0.80-1.00 μm.

Discussion

The classification results for the Indiana data were particularly important, since this was the first time that such a large geographic area has been automatically mapped into basic cover types. These results demonstrated the capability to successfully extrapolate from training samples taken in one small segment of the flightline to the entire 70 square mile area.

Even though the classification results of the Indiana data were highly accurate (over 97%), there was a slight change in spectral response of the various materials as the distance between the area classified and the training area increased. This was indicated by a larger percentage of RSU's being thresholded near the southern end of the flightline.

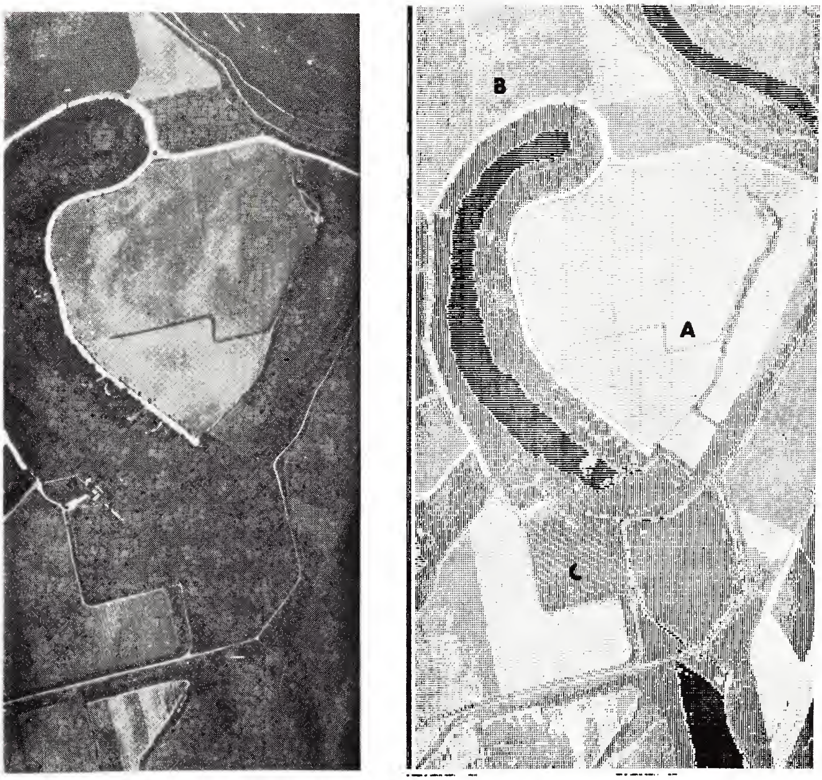


FIGURE 8. *Moon Lake test area. The classification results on the right show water, bare soil, trees, and two classes of agricultural vegetation. band. Right, imagery in 0.80-100 μm band (reflective infrared).*

Four possible reasons for spectral changes are: 1) a difference in atmospheric conditions between the northern and southern portions of the flightline; 2) slight change in the spectral characteristics of the materials due to geographic variability (*e.g.* trees being leafed out more in the south, or differing water quality between the East and West Forks of the White River); 3) adjustments in instrumentation setting, although this will normally produce dramatically different results; and (4) electronic variations in some of the data gathering, recording, or processing instrumentation. The third possibility was not believed to be a factor, but any of the other possibilities alone or in concert could have caused the observed shift in spectral response.

In this classification, the shift in response was slight, and occurred over a relatively long flightline. The results were not greatly affected since the materials being classified were spectrally very different in some wavelength bands. However, these observed changes indicate a need for further study of this type of phenomena, so that proper techniques for instrument calibration and for delineating an optimum set of training samples can be developed. If the problem cannot be corrected by instrument calibration techniques, procedures for periodically modifying the training samples may have to be developed. This might involve such procedures as using a new set of training samples every 20 miles of the flightline and weighting the new training samples more than the previously existing training samples.

FIGURE 9. *Left, classification of orchard area at Moon Lake showing a moire pattern with the rows of trees. Center is scanner imagery in 0.66-0.72 μm wavelength*

The accuracy obtained in all three sets of data indicate that the LARS automatic processing techniques for multispectral scanner data can be utilized for data from various geographic locations.

FIGURE 10. *California data imagery in the 0.80-1.00 μm band at left. The classifications of rice, safflower, and bare soil are at center and right.*

Acknowledgments

Appreciation is expressed to the Airphoto Interpretation and Photogrammetry Lab., School of Civil Engineering, Purdue Univ. for the Highway 37 scanner and photographic data.

All aircraft scanner data used were collected by the Institute of Science and Technology, University of Michigan.

Special thanks to the LARS staff for their work in the various phases of digitizing and analyzing these data.

Literature Cited

1. HOFFER, R. M. 1967. Interpretation of remote multispectral imagery of agricultural crops. Lab. for Agr. Remote Sensing, Vol. No. 1. Res. Bull. No. 831, Agr. Exp. Sta., Purdue Univ., Lafayette, Indiana. 36 p.

2 Laboratory for Agricultural Remote Sensing. 1967. Remote multispectral sensing in agriculture; Vol. No. 2 (Ann. Rep.). Res. Bull. No. 832, Agr. Exp. Sta., Purdue Univ., Lafayette, Indiana. 75 p.

3. ——————. 1968. Remote multispectral sensing in agriculture; Vol. No. 3 (Ann. Rep.). Res. Bull. No. 844, Agr. Exp. Sta., Purdue Univ., Lafayette, Indiana. 176 p.