LARS Information Note 033174

A COMPARISON BETWEEN DIGITIZED COLOR INFRARED PHOTOGRAPHY AND MULTISPECTRAL SCANNER DATA USING ADP TECHNIQUES

M.E. COGGESHALL
R.M. HOFFER
J.S. BERKEBILE

The Laboratory for Applications of Remote Sensing

Purdue University, West Lafayette, Indiana

1974

Information Note 033174

A Comparison Between Digitized Color Infrared Photography and Multispectral Scanner Data, Using ADP Techniques[1]

By M. E. Coggeshall[2]
R. M. Hoffer[3]
J. S. Berkebile[3]

ABSTRACT

Computer classification results derived from digitized color infrared photography were compared with similarly derived results for three corresponding wavelength bands of multispectral scanner data. Classification results for 158 test fields indicated 47.5 percent overall correct identification with the digitized color infrared photography as compared to 80.5 percent for the corresponding three channels of multispectral scanner data (and 95.1 percent where all of the available twelve multispectral scanner channels were utilized). Forest cover was particularly difficult to separate from agricultural cover types using the digitized photography, but good separation was obtained with the multispectral scanner data.

Introduction

Previous work with both aircraft and spacecraft data had indicated a potential to use digitized color infrared photography for accurate computer classification of crop species (1). Later work with aircraft data, however, indicated some difficulty in spectrally separating crop cover from forested land (2,3). Thus, the question arose as to the degree of spectral differentiation possible between forest and agricultural cover types using digitized color infrared photographic data as compared with multispectral scanner data.

The possibilities for conducting this type of test have been very limited in the past. Multispectral scanner data are usually gathered from relatively low altitudes. Photographic data obtained simultaneously would yield a large number of individual frames, and the necessary densitometry operations would be difficult and costly. The solution to this difficulty would be to obtain high-altitude photography at the same time the scanner data are obtained from lower altitudes. Since the scanner data and photographic data would need to be gathered simultaneously to preclude as many variables as possible (such as sun angle changes, atmospheric attenuation, and temporal

---

[1]This work was sponsored by the National Aeronautics and Space Administration (NASA) under Grant No. NGL 15-005-112. Journal Paper #5451, Agricultural Experiment Station, Purdue University.

[2]Service Forester, Department of Natural Resources, State of Indiana.

[3]Assoc. Professor and Graduate Student respectively, Department of Forestry and Conservation, Purdue University.

variations in spectral response of the materials), suitable data sets are not easily obtained. Obviously, the difference in altitude from which the two data sets are derived does raise some questions concerning the characteristics of the data used in such comparisons. Hopefully, when the SKYLAB data becomes available, questions concerning the characteristics of data gathered from different altitudes will be answered.

Materials and Techniques

During the 1971 Corn Blight Watch experiment (4), a nearly ideal data set was obtained. Color infrared photography was collected by the NASA WB-57 aircraft from an altitude of 60,000 ft., resulting in a 1:120,000 scale photograph which contained 90 percent of the scanner test site area. The photograph's high altitude also afforded a minimal look angle, thereby decreasing resultant geometric distortion. The multispectral scanner data were obtained by the ERIM aircraft from an altitude of 5,000 ft. over the same area within 22 minutes of the same time.

The test site was a 6,400-acre area in south-central Indiana that is approximately 60 percent forested with mixed hardwoods and a few, small, pine plantations. The remainder of the area is agricultural land consisting largely of forage-type vegetation and also a fairly large number of corn and soybean fields.

The color infrared transparency was subjected to a color separation procedure which resulted in three black-and-white transparencies, each representing one of the photograph's three emulsion layers. However, the color infrared imagery supplied by NASA was on duplicate film, and the filters used in the color separation process did not correspond precisely to the sensitometric characteristics of the three color emulsion layers. Consequently, precise <u>spectral</u> interpretation of the density characteristics of the color separations is not feasible.

The color separation images were digitized and recorded on magnetic tape, using a scanning microdensitometer having a 50-micron aperture. These digitized photographic data were then overlayed onto a single data tape so that each ground resolution element would be in registry for each of the three emulsion layers.

The multispectral scanner utilized obtains data in twelve separate wavelength bands as indicated in Figure 1. Since channels four, seven, and eight of the scanner most closely correspond to the wavelength sensitivities of the three emulsion layers of the color infrared photography (Figure 2), these three channels were utilized in the comparisons for the study to be described.

The basic analysis procedures used during this study have been previously described in detail (2,5). In essence, one displays a single wavelength band of either the multispectral scanner data or the digitized photographic data on a television-like display unit attached to the computer. A light-pen allows the researcher to select rectangular samples of the various cover types, the coordinates of which are automatically punched on computer cards for later processing. In this study, 241 such areas were designated from the multispectral scanner data. Of these, 62 were used to train the classifier algorithm, and the remaining 179 were used to test the machine's performance (see Table 1). The size and number of test areas were selected to represent in proper proportion the cover types present in the test segment. In analyzing the digitized photographic data, the same training and test areas were designated as precisely as possible, although only 58 training and 158 test fields could be utilized because one square mile at the southern end of the scanner test segment was not included on the photograph.

A set of statistics is then derived from the training samples for each class using all the available channels of data. On the basis of these statistical definitions of the represented cover types, a classifier algorithm establishes decision boundaries in n-dimensional space (n=3 in this case because 3 channels were used) and assigns each data point in the test area to one of the defined classes. The classification results may then be displayed qualitatively as a map or image, and quantitatively in a test class performance table.

Results

A classification using multispectral scanner channels four, seven, and eight was obtained which gave an overall accuracy of 80.5% (Table 2). Examination of the classification results indicated that a scattering of individual points throughout the forested areas were being classified into the other categories. This is due to the large amount of spectral variability within a deciduous forest area (e.g. sunlit versus shaded sides of tree crowns being classified into different spectral categories). It is this spectral variability within a deciduous forest area which gives the forest its textural characteristics (thereby making a forest area easily identified and delineated by the human interpreter) but which also makes accurate classification of individual small resolution elements by computer quite difficult.

Classification of the digitized color infrared photography yielded an overall performance accuracy of only 47.5% (Table 3). It is important to note that the same locations designated as test areas for the scanner data analysis were also designated as test areas on the digitized photography. The average performance by cover type class was much higher than the average by land area (overall performance) because some of the classes with a relatively small number of data points, such as water and soybeans, had a much higher classification accuracy than the classes with a large number of data points, particularly forage and deciduous forest. The deciduous forest class had an especially low classification accuracy (only 36.1% for 34,970 sample points), with the misclassifications being distributed among all other categories, as shown in Table 3. It was thought that because the classification of the digitized photography involved the same training and test locations that had previously been used in the scanner data classifications, perhaps an unknown bias had been introduced. This concern was amplified by the observation that, in addition to scattered individual points misclassified on the photographic data, a large area of known forest cover was almost totally misclassified, primarily as corn. An examination of the photograph revealed the problem to be one of uneven illumination across the photograph. Thus, because of potential bias in the analysis of the photographic data set, another classification was conducted. The results of the new classification showed no significant change in accuracy over the previous classification. A large amount of confusion still existed between deciduous forest cover and the other agricultural cover types present. Apparently, the spectral characteristics of the forest cover and corn in this area were recorded by the film as being almost identical. This was shown to be the case by the statistical comparison depicted in Figure 3. Table 4 summarizes the above results and also provides a comparison with the classification performance derived from the use of all twelve scanner channels.

The overall performance was approximately the same for the two color infrared photo comparisons although some differences did occur between the two tests for individual cover types. The difference between the two tests with digitized photographic data and the classification of multispectral scanner data using three similar wavelength bands is striking. Table 4 also emphasizes that the three channels of scanner data which approximated color infrared photography resulted in a poorer classification than using the "best" three channels of multispectral scanner data. This is because the "best" three channels included one wavelength band in the middle infrared portion of the spectrum (1.5-1.8 μm) and one wavelength band in the thermal portion of the spectrum (9.3-11.7 μm). For purposes of comparison, one sees from Table 4 that a rather significant increase in accuracy was obtained when more than three channels of data were utilized. However, one should not conclude that simply adding additional channels of data will necessarily allow greater accuracy to be achieved.

Discussion and Conclusions

The classification results for the digitized color infrared photography revealed that much of the deciduous forest cover was classified as corn and that a fair amount of the corn was misclassified as deciduous forest. Examination of the histograms for corn and deciduous forest showed that the spectral characteristics of these two categories were indeed very similar, thereby making accurate separation and classification impossible when using the digitized color IR photographic data (Figure 3). Similar histograms of the same areas in the multispectral scanner data showed that a distinct separation could be obtained in the visible wavelength bands, as shown in Figure 3. Examination of the multispectral response graphs (Figure 4) allows a similar comparison between the two data sets for all of the various cover types involved.

These results suggest a much greater potential for accurate identification and classification of various important earth surface cover types using multispectral scanner data as opposed to digitized, multiemulsion, color infrared photography. Apparently this difference is due largely to the greater dynamic range and higher spectral resolution of the scanner system. It should be reemphasized, however, that the photographic data used in this study were second generation, which may have resulted in the loss of some information content.

Studies with SKYLAB data involving comparisons of multispectral scanner data and the multiband and multiemulsion photographic data will offer additional insight into the capabilities and limitations of computer-aided data processing techniques used in conjunction with the various data sources available.

REFERENCES

1. Anuta, P. E. 1970. "Multispectral Classification of Crops in the Imperial Valley, California, from the Digitized Apollo 9 Photography," Laboratory for Applications of Remote Sensing (LARS). Information Note 070770. Purdue University, W. Lafayette, Ind. 11 pp.

2. Hoffer, R. M. 1972. "ADP of Multispectral Scanner Data for Land Use Mapping," LARS Information Note 080372. Purdue University, W. Lafayette, Ind. 24 pp.

3. Coggeshall, M. E., and R. M. Hoffer. 1973. "Basic Forest Cover Mapping Using Digitized Remote Sensor Data and ADP Techniques," LARS Information Note 030573. Purdue University, W. Lafayette, Ind. 131 pp.

4. R. A. MacDonald, et. al. 1972. "Results of the 1971 Corn Blight Watch Experiment," LARS Information Note 100272. Purdue University, W. Lafayette, Ind. 33 pp.

5. Hoffer, R. M., P. E. Anuta, and T. L. Phillips. 1972. "ADP Multiband and Multiemulsion Digitized Photos." Photogrammetric Engineering, 38 (10): 989-1001.

Table 1. Summary of training and test classes for analysis of Seg. 218 MSS data.[1]

Cover Type	Training Classes: Number of Points	Training Classes: % of Total Seg.	Training Classes: % of Total Train.	Test Classes: Number of Points	Test Classes: % of Total Seg.	Test Classes: % of Total Test	Test/Train
decid	161	0.05	4.33	32252	9.08	64.77	200.3
conifer	147	0.04	3.95	88	0.02	0.18	0.6
water	321	0.09	8.62	339	0.10	0.68	1.1
forage	1870	0.53	50.24	11760	3.31	23.62	6.3
corn	784	0.22	21.06	2679	0.75	5.38	3.4
soybean	439	0.12	11.79	2676	0.75	5.37	6.1
TOTAL	3722	1.05	100.00	49794	14.02	100.00	

[1] Seg. 218 = 1600 lines x 222 columns = 355,200 points

Table 2. **Classification results: multispectral scanner data using three channels corresponding to color infrared film bands.**

SERIAL NUMBER------- 1215206803 CLASSIFIED- DEC 15,1972

CHANNELS USED

CHANNEL 4	SPECTRAL BAND 0.52 TO 0.57 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 7	SPECTRAL BAND 0.61 TO 0.70 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 8	SPECTRAL BAND 0.72 TO 0.92 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0

CLASSES

	CLASS		CLASS
1	DECID	4	FORAGE
2	CONIFER	5	CORN
3	WATER	6	SOY

TEST CLASS PERFORMANCE

GROUP	NO OF SAMPS	PCT. CORCT	DECID	CONIFER	WATER	FORAGE	CORN	SOY
			NUMBER OF SAMPLES CLASSIFIED INTO					
1 DECID	32252	83.9	27061	3133	4	433	161	1460
2 CONIFER	88	89.8	9	79	0	0	0	0
3 WATER	339	96.2	1	9	326	1	2	0
4 FORAGE	11760	64.1	5	10	12	7533	2405	1795
5 CORN	2679	97.5	0	2	0	17	2613	47
6 SOY	2676	92.5	43	0	0	44	114	2475
TOTAL	49794		27119	3233	342	8028	5295	5777

OVERALL PERFORMANCE(40087/ 49794) = 80.5

AVERAGE PERFORMANCE BY CLASS(523.9/ 6) = 87.3

Table 3. **Classification results: digitized color infrared photography using all three channels.**

SERIAL NUMBER------- 1130206803 CLASSIFIED- NOV 30,1972

CHANNELS USED

CHANNEL 1	SPECTRAL BAND 0.47 TO 0.61 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 2	SPECTRAL BAND 0.59 TO 0.71 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0
CHANNEL 3	SPECTRAL BAND 0.68 TO 0.89 MICROMETERS	CALIBRATION CODE = 1	CO = 0.0

CLASSES

CLASS		CLASS	
1	DECID	4	FORAGE
2	CONIFER	5	CORN
3	WATER	6	SOY

TEST CLASS PERFORMANCE

	GROUP	NO OF SAMPS	PCT. CORCT	NUMBER OF SAMPLES CLASSIFIED INTO DECID	CONIFER	WATER	FORAGE	CORN	SOY
1	DECID	34970	36.1	12607	7367	700	951	10389	2956
2	CONIFER	127	63.8	33	81	5	0	7	1
3	WATER	461	97.6	0	7	450	4	0	0
4	FORAGE	11985	68.2	345	601	27	8172	1957	883
5	CORN	2018	75.6	301	133	0	43	1525	16
6	SOY	1758	87.1	83	7	0	31	106	1531
	TOTAL	51319		13369	8196	1182	9201	13984	5387

OVERALL PERFORMANCE(24366/ 51319) = 47.5

AVERAGE PERFORMANCE BY CLASS(428.3/ 6) = 71.4

Table 4. Classification Results of Test Areas, Comparing MSS and Color IR Photographic Data

Category	12 Channels, MSS[1]	3 Channels, MSS[2]	3 Channels, MSS[3] (Color IR Comparison)	3 Channels, Color IR Photo[4] Test #1	3 Channels Color IR Photo[4] Test #2
Deciduous Forest	97.2	88.8	83.9	36.1	32.8
Coniferous Forest	95.5	95.5	89.8	63.8	73.2
Water	98.5	98.2	96.2	97.6	98.0
Forage	88.6	80.6	64.1	68.2	74.3
Corn	96.9	84.8	97.5	75.2	73.8
Soybeans	96.3	95.2	92.5	87.1	76.8
Overall Performance	95.1	87.1	80.5	47.5	46.8

[1]Spectral wavelength regions: visible (0.46-0.49, 0.48-0.51, 0.50-0.54, 0.52-0.57, 0.54-0.60, 0.58-0.65, 0.61-0.70 micrometers), near infrared (0.72-0.92, 1.00-1.40 micrometers), middle infrared (1.50-1.80, 2.00-2.60 micrometers), and thermal infrared (9.30-11.70 micrometers).

[2]Bands 0.58-0.65, 1.50-1.80, and 9.30-11.70 micrometers.

[3]Bands 0.52-0.57, 0.61-0.70, and 0.72-0.92 micrometers.

[4]Bands 0.47-0.61, 0.59-0.71, and 0.68-0.89 micrometers.

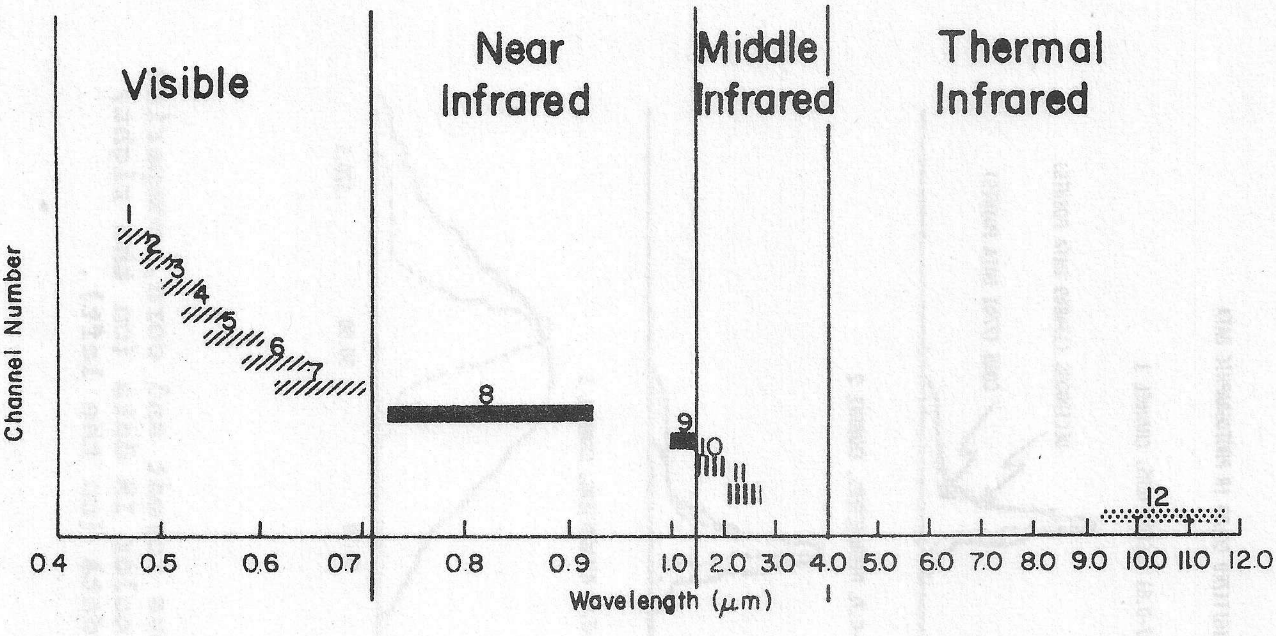


Figure 1. Relationship of MSS data channels to the electromagnetic spectrum.

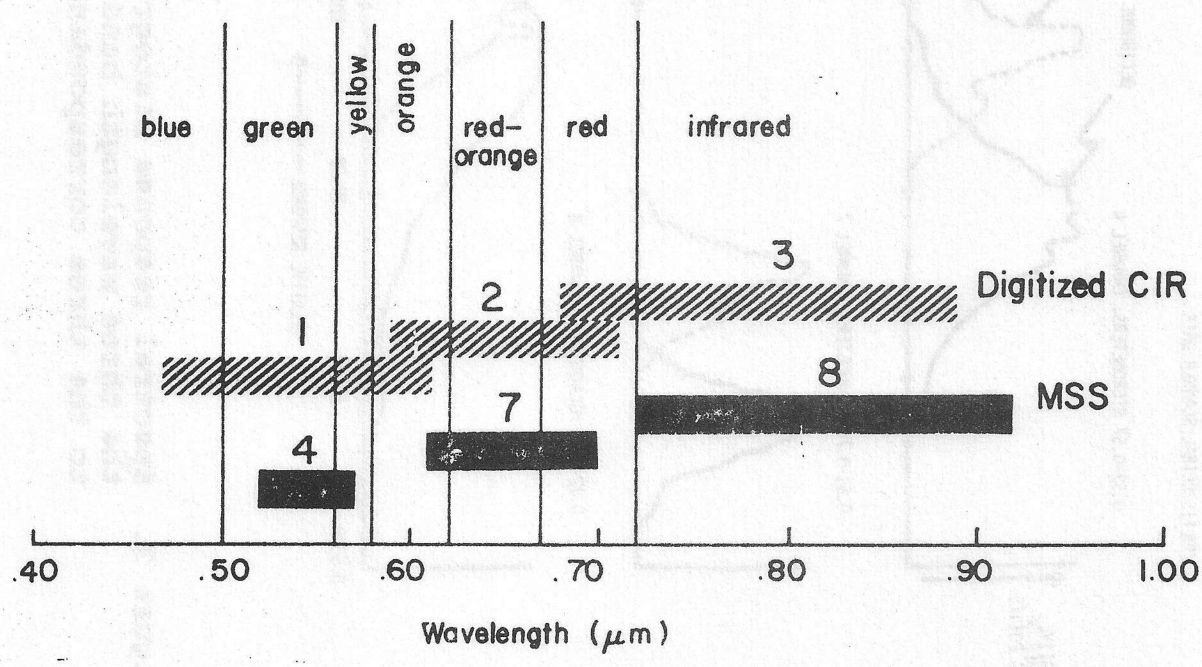


Figure 2. Relationship of the digitized color infrared photographic channels and the most closely corresponding MSS channels.

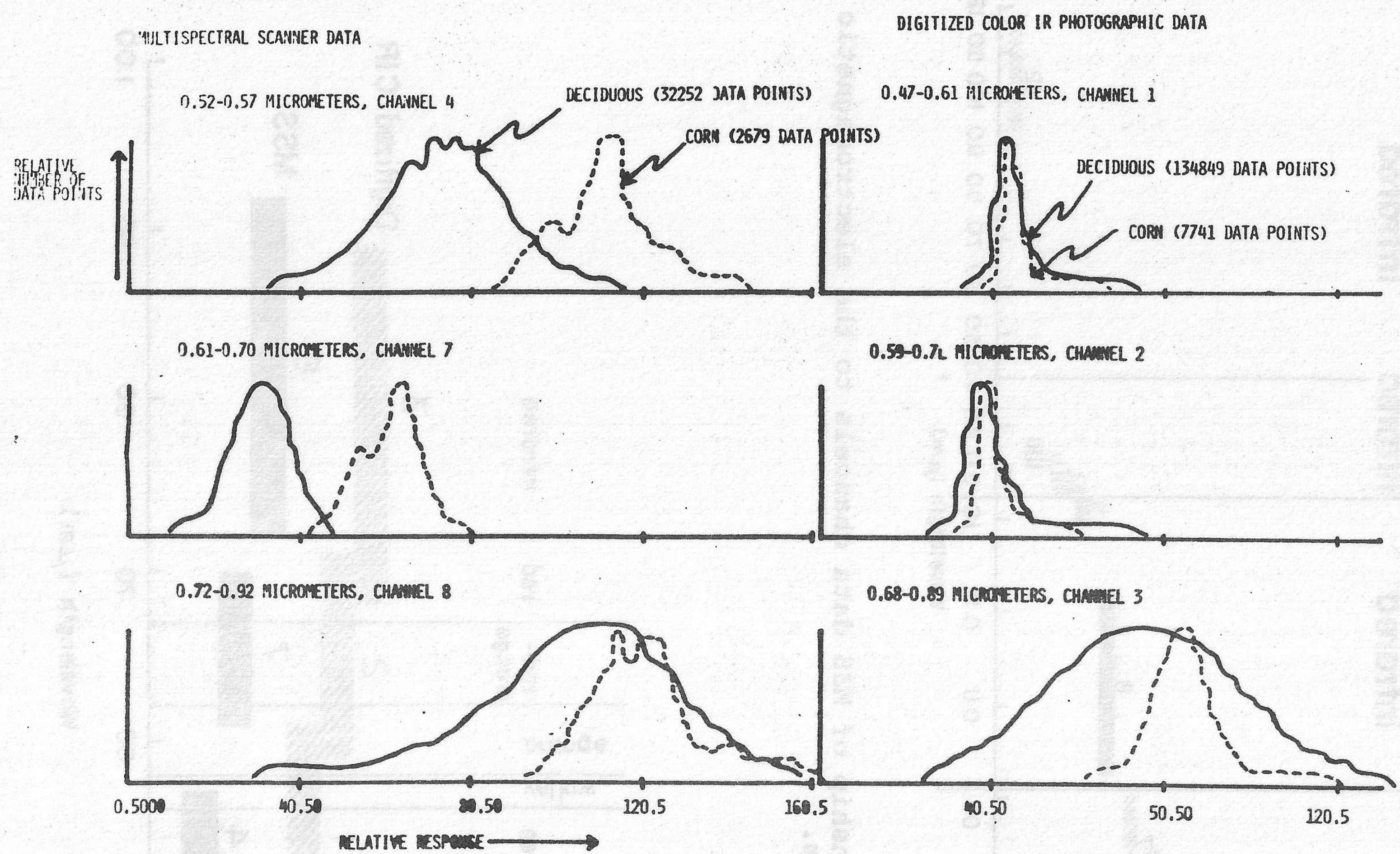
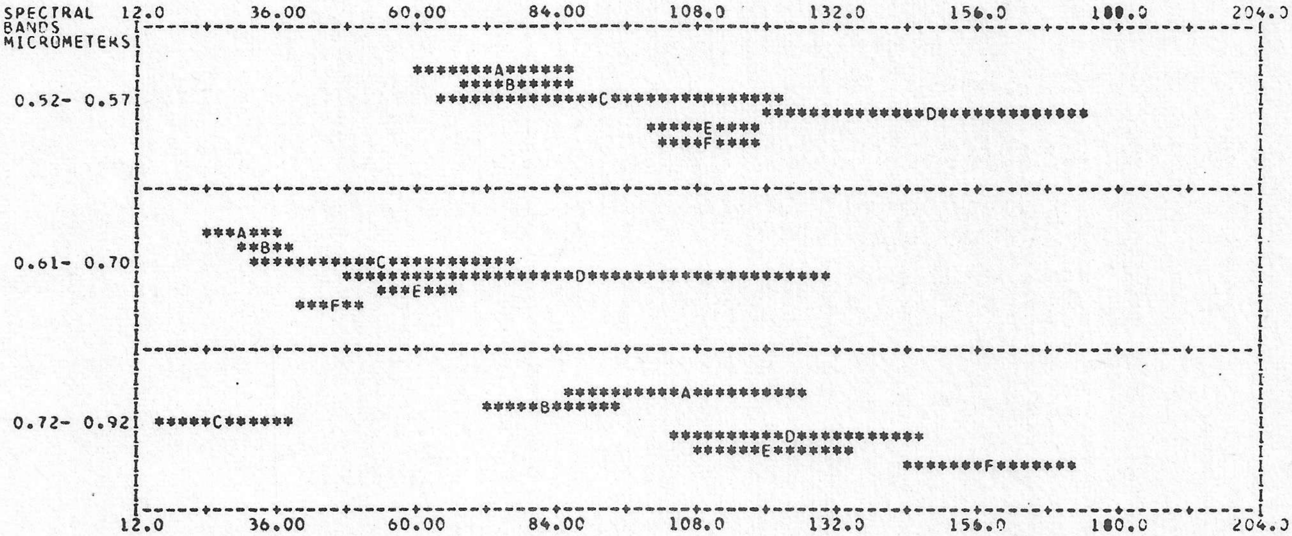


Figure 3. Spectral response histograms for deciduous forest and corn, comparing the three wavelength bands of digitized color IR data (on the right) to the three corresponding bands of MSS data (on the left).

A = CLASS 1 DECID
B = CLASS 2 CONIFER
C = CLASS 3 WATER
D = CLASS 4 FORAGE
E = CLASS 5 CORN
F = CLASS 6 SOY

MULTISPECTRAL SCANNER DATA



DIGITIZED COLOR IR PHOTOGRAPHIC DATA

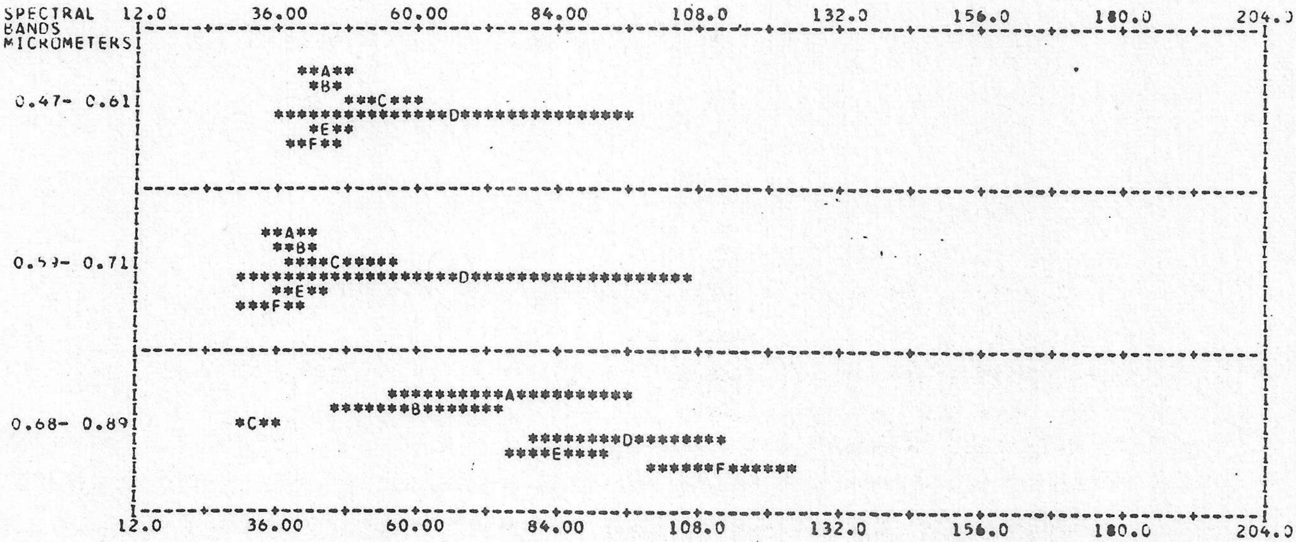


Figure 4. Spectral response graphs of the various cover type classes, comparing digitized color IR data to multispectral scanner data. The letters indicate the mean spectral response of the various classes, and the asterisks represent the ± 1 standard deviation range.