E7.5-10.1.1.0
CR-140921

ENVIRONMENTAL RESEARCH INSTITUTE OF MICHIGAN
FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN
P. O. BOX 618 • ANN ARBOR • MICHIGAN • 48107
PHONE (313) 483-0500

101900-46-L
15 January 1975

Developing Processing Techniques for Skylab Data
Monthly Progress Report, December 1974

(E75-10110) DEVELOPING PROCESSING TECHNIQUES FOR SKYLAB DATA Monthly Progress Report, Dec. 1974 (Environmental Research Inst. of Michigan) 7 p HC $3.25 CSCL 05B
N75-16031
Unclas
G3/43 00110

EREP Investigation 456 M
NASA Contract NAS9-13280

Prepared by

Richard F. Nalepka - Principal Investigator
William A. Malila - Co-Principal Investigator
James P. Morgenstern

NASA Technical Monitor

Mr. Larry B. York/TF6
National Aeronautics and Space Administration
Johnson Space Center
Principal Investigator Management Office
Houston, Texas 77058

PRECEDING PAGE BLANK NOT FILMED

Developing Processing Techniques for Skylab Data
Monthly Progress Report, December 1974

The following report serves as the twenty second monthly progress report for EREP Investigation 456 M which is entitled "Developing Processing Techniques for Skylab Data". The financial report for this contract (NAS9-13280) is being submitted under separate cover.

The purpose of this investigation is to test information extraction techniques for SKYLAB S-192 data and compare with results obtained in applying these techniques to ERTS and aircraft scanner data.

The month of December was a shortened work month due to a blizzard at the beginning and the holidays at the end. We were able in the time available, however, to begin our processing of the set of four S-192 data tapes received at the beginning of the month.

The initial task was to find the fraction of the data sent which covered the EREP test site in southern Michigan, and to assess the quality of the data received.

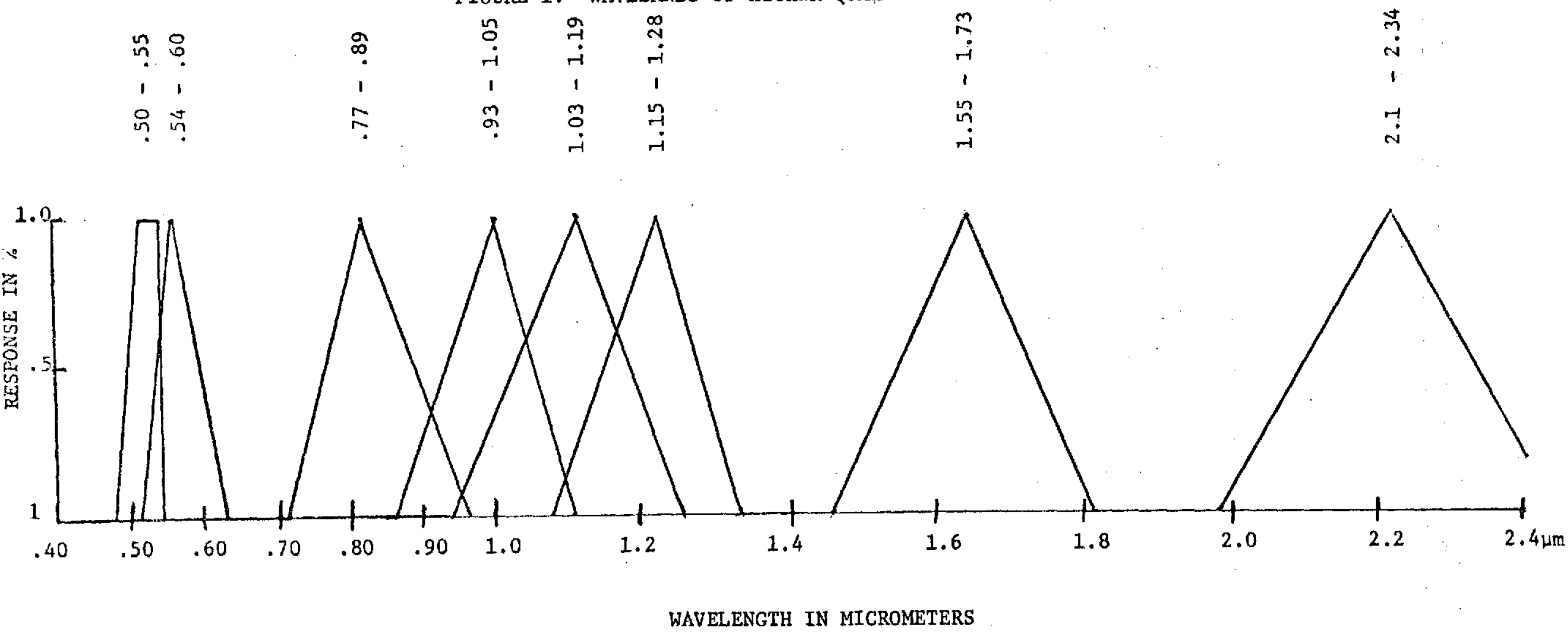
We began by examining the available three bands of screening film, and determined approximately the times of the first and last scan lines over the test site. Approximately 2 seconds of data covered the entire test area. Once the desired scan lines were identified, a broad portion of the data which included the test area was converted to ERIM format data tapes so we could continue the processing.

At the next step we generated a graymap of SDO 11, using every second line and every second pixel, and determined that we had indeed copied the desired portion of the data.

We continued checking data quality by generating a set of small graymaps, every line and point, one graymap for each SDO. Analyses of these maps showed that eight of the 13 detectors in the S-192 exhibited good signal-to-noise characteristics.

The portion of the spectrum covered by these detectors is shown in Figure 1. Of the other bands, the thermal SDO's (15, 16, 21) and the blue band (0.41 - 0.45 μm, SDO 22) displayed very low signal-to-noise ratios such that no structure could be found in the graymaps. Three other detectors, 0.45 - 0.50 μm, 0.60 - 0.65 μm, and 0.66 - 0.73 μm, (SDO's 18, 5 & 6, 7 & 8, respectively) displayed some noise, which was a function of scan frequency and intermittent loss of synchronization in digitization. It is believed at this time that use of these SDO's in future processing may degrade results of the classifier.

FIGURE 1. WAVEBANDS OF HIGHER QUALITY SKYLAB S-192 DATA



There has been some question as to whether or not we are in receipt of a final data product from JSC or an intermediate product. Until such time as this is cleared up, we will continue to process the data set at hand.

We next turned our attention to the question of registration between SDO's. By registration, we mean to what extent do all the SDO's have the same instantaneous field of view (IFOV); or to put it another way, to what extent are they all focussed on exactly the same ground area. Having all data channels in registration is certainly a requirement for all processing. However, in this study it is doubly so since we are analyzing data for an area where the dimensions of many of the object classes of interest (agricultural fields) are about the same as the resolution of the system. Maximum likelihood recognition processing based on training data statistics cannot work well if some number of the data channels are out of registration; for example, if most channels of a given pixel are focussed in one field, but some channels are focussed on an adjacent dissimilar area, the classifier will probably not work correctly. Misregistration between bands will also have serious effects on the use of a mixtures classifier; i.e., when the classifier is attempting to estimate properties of a pixel which are smaller than the IFOV. Thus, it is felt that the data must be well registered in order to meaningfully process the data.

Thus we began studying the registration between S-192 SDO's. One discrepancy turned up immediately. The EREP users handbook states that misregistration between SDO's will be no worse than 0.1 resolution element. This cannot be true since in digitizing the detectors' output, all the even numbered SDO's are sampled 0.5 resolution elements after the odd numbered SDO's, for a given pixel. Thus there are two groups of SDO's which are registered no better than 0.5 resolution element. In addition, the SDO-SDO registration may be affected during scan line straightening, since the straightening is done independently for each detector and is done on a nearest neighbor basis. If there are changes in registration due to the straightening algorithm, we would expect the SDO-SDO registration to vary quasi-randomly throughout a scan line. It is certainly a problem that we intend to look into.

During the coming month we also intend to begin the process of locating line and point coordinates of the fields in the test area for which we have ground information. These areas will then serve as training and test fields for the processing of the S-192 data of the Michigan test site.

Progress for December in the processing of the aircraft-gathered multispectral scanner data centered around the acquisition of good training signatures for later use in classification. The training focussed on an area of approximately 1.5 square miles which was located between miles 3 and 5 (miles from the beginning of the flight line) and included the area of scan $\pm 40^{o}$ from nadir. The region was chosen for training because it was the first area in the data set which contained several large contiguous areas of corn, soybeans, trees and bare soil.

The ERIM cluster program was run on this training area. The parameters for the program and the methods used to obtain them were described in last month's progress report. Further, to save time we used only every second pixel from every second line; this did not seriously impair the accuracy of the results. In all 6516 pixels were clustered into 59 different groups.

The output graymap of cluster assignments was explained and a list was developed connecting clusters to the actual ground cover. It was shown that four major object classes (corn, soybeans, trees, hay) were represented by very few clusters, while the various other ground covers, which display a wide degree of variability such as weeds, bare soil, wet bare soil, cut hay, senescent vegetation, pastures, farmsteads, etc., were represented by 85% of the clusters.

The next problem was to obtain some semblance of order from the large number of clusters of the other ground covers. First, it seemed that all the weed fields were represented by only 4 clusters. So these four were set aside. Then, in channel by channel graphs of all the cluster means, it became apparent that these other clusters stratified into three general groups. These groupings were found to be consistent from channel to channel and in fact appeared to be a function of the amount of vegetative ground cover. These three groupings were sparse vegetation, bare soil, and dark or wet bare soil. Thus, we were able to generalize most of the clusters into 8 broader groups of common ground cover; these groupings are summarized in Table 1 below.

TABLE 1. COMBINING CLUSTERS BASED ON REPRESENTING COMMON OBJECT CLASSES

	Class	No. of Clusters	Total No. of Points
1	Corn	2	2006
2	Soybeans	3	217
3	Trees	3	566
4	Hay	1	1771
5	Sparse Vegetation	8	252
6	Weeds	4	889
7	Bare Soil	9	305
8	Dark or Wet Bare Soil	6	301
	TOTAL	36	6307

As the next step, the statistics (means and standard deviations) for the clusters in each group were combined to yield one signature for use in classification processing. It was necessary to combine the clusters so as to greatly reduce the number of training signatures used in classification processing in order to reduce costs. Also, it was felt that no loss of accuracy would result since it appeared from our analyses that there was very little overlap between groups of clusters. As an additional safeguard the program which calculates the new signature first performs a χ^2 test on each signature to measure its distance (in a probability sense) to the mean of the other signatures in the group.

Seven signatures were obtained by combining clusters. For the hay signature, a full signature (mean and covariance matrix) was calculated from those pixels which had been associated with the hay cluster during clustering. Finally, since there was no water in the training area, the water signature previously calculated was added to the group of training signatures.

With the nine training signatures now fully defined, we calculated the pairwise probability of misclassification for the training signatures. These are shown in Table 2.

TABLE 2. PAIRWISE MISCLASSIFICATION PROBABILITY IN PER CENT FOR AIRCRAFT TRAINING SIGNATURES

	Soy	Trees	Hay	Weeds	Sp Veg	Dk Soil	Soil	Water
Corn	0.4	4.0	8.0	3.0	0.5	0	0	0
Soy		0.1	8.0	0.5	0.1	0	0	0
Trees			4.0	0.3	0.1	0	0	0
Hay				0.6	0.1	0	0	0
Weeds					13.0	0	0.6	0
Sparse Vegetation						4.0	11.0	0
Dark Bare Soil							5.0	0
Bare Soil								0

The large values which occur between sparse vegetation and weeds and sparse vegetation and bare soil, just accurately reflects the wide variability of such ground conditions and is not viewed as a problem. The other major confusion, between hay and corn, soy, trees, appears to arise because the hay cluster in n-space occupies a hyper-volume which is to a great extent in the interior of a hyper triangle whose vertices are the corn, soybeans and tree clusters. Thus the overlap in these signatures indicated by the probability of misclassification is readily understandable. Some degree of confusion between corn and trees usually exists in processing multispectral data; a probability of misclassification of 4% between this pair may be too large to be tolerable. Further investigation of this problem is in order.

During the next month we intend to continue the training process and finally to perform classification processing on this aircraft data set.

Submitted by: Richard F. Nalepka
Richard F. Nalepka
Principal Investigator

Approved by: Jon D. Erickson
Jon D. Erickson
Head, Information Systems
and Analysis Department

Approved by: Paul R... for
Richard R. Legault,
Director, Infrared and
Optics Division

dd