Marilyn Hixson
Donna Scholz
Nancy Fuhs
Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, IN 47907
Tsuyoshi Akiyama*
Ministry of Agriculture, Forestry, and Fishery
Nishinasuno, Tochigi, Japan

Evaluation of Several Schemes for Classification of Remotely Sensed Data

Development of representative training statistics is relatively more important for obtaining accurate classifications than selection of the classification algorithm.

Abstract: *Several approaches to machine analysis of Landsat* MSS *data have been developed over the past decade, and the data analyst must select which analysis approach might perform best for a given problem. The overall objective of this study was to apply and evaluate several classification schemes for crop identification. The schemes examined were (1) per point Gaussian maximum likelihood classifier, (2) per point sum-of-normal-densities classifier, (3) per point linear classifier, (4) per point Gaussian maximum likelihood decision tree classifier, and (5) texture sensitive per field Gaussian maximum likelihood classifier. Seven agricultural data sets were selected for use to sample variability in major crops and agricultural practices.*

Test site location (embodying effects of soils, climate, and agricultural practices) and classifier both had significant effects on the classification accuracy of small grains. Neither the corn and soybean accuracies, nor the overall accuracy (considering all cover types) of the classifiers differed significantly when the same training method was used. A different training method used with one of the classifiers, however, did produce results of significantly lower accuracy. The results suggest that development of representative training statistics is relatively more important for obtaining accurate classifications than selection of the classification algorithm. The complexity of use and computer costs for the classifiers also varied significantly.

Introduction

Over the past several years, the potential utility of remotely sensed data to survey and monitor agricultural crops has been increasingly recognized. The use of a per point maximum likelihood classifier in the Corn Blight Watch Experiment during 1971 was the first attempt at large scale application of digital classification of remotely sensed multispectral data[1]. This was followed in 1973-75 by the Crop Identification Technology Assessment for Remote Sensing (CITARS) project for corn and soybeans in Indiana and Illinois using Landsat MSS data[2]. Since then extensive research has been devoted to wheat inventory with the Large Area Crop Inventory Experiment (LACIE) during 1973-78[3]. Currently, interest has been directed toward analysis of multicrop areas, with corn and soybeans being the major crops of interest.

To support these efforts utilizing satellite remotely sensed data, several numerical analysis schemes have been developed and implemented

* Dr. Akiyama was a Visiting Scientist at the Laboratory for Applications of Remote Sensing at the time this work was conducted.

Photogrammetric Engineering and Remote Sensing,
Vol. 46, No. 12, December 1980, pp. 1547-1553.

at numerous university, business, and government facilities in the United States and abroad. The remote sensing data analyst, therefore, must determine which analysis approach or algorithm might perform best for a given problem. Numerous studies have evaluated the performance of a given classifier, but relatively few studies have objectively compared the performance of several approaches for a specific analysis problem.

The overall objective of this study was to apply and evaluate several available classification schemes on agricultural data sets. The data sets were selected to include corn, soybeans, winter wheat, and spring wheat as major crops. Classification accuracy for test fields, ease of analyst use, and required computer time were compared.

Experimental Approach

Test sites were selected from three major data sets: CITARS data from 1973 over Illinois and Indiana[2]; LACIE data from 1976 over the U.S. Great Plains[3]; and multicrop data from 1978[4]. An 8- by 24-kilometre area in Fayette County in south central Illinois was used from the CITARS data set. A 9.3- by 11.1-kilometre area was selected in each of Foster County, North Dakota, and Grant County, Kansas, from the LACIE data. Four segments, each 9.3 by 11.1 km, were selected from the multicrop data: Pottawattamie and Shelby Counties in west central Iowa; Tippecanoe County in west central Indiana; and Iroquois County in east central Illinois. The locations of these segments are shown in Figure 1.

The segments sampled several major crops: winter wheat in Kansas; spring wheat in North Dakota; and corn and soybeans in Indiana, Illinois, and Iowa. The Corn Belt segments were located in two distinct regions (Figure 1) to sample variability in soils, climate, and agricultural practices. Cloud-free multitemporally registered digital Landsat MSS data were available over all sites. Reference data were obtained by ground observers who identified cover types of fields and recorded these on acetate overlays of current season, color infrared aerial photographs.

For each segment, Landsat MSS data acquired on four different dates were selected for analysis. Acquisitions were selected to temporally sample the corn, soybean, and wheat crop calendars so that maximum crop development differences would be apparent (Tables 1a and 1b). For all segments, a spring acquisition was selected to aid in separating winter small grains, trees, and permanent pasture from row crops. For the Corn Belt segments, an acquisition after corn had tasseled was included to separate corn and soybeans. An acquisition after heading and/or harvest of wheat was used to aid in separating wheat from other cover types. To optimize classification performance versus computer costs, a subset of four of the 16 available wavelength bands (Table 2) was selected to maximize the average transformed divergence between pairs of classes[5].

Five classifiers, implemented on an IBM 370/148 computer at the Laboratory for Applications of Remote Sensing (LARS), Purdue University, were selected for study:

- CLASSIFYPOINTS is a per point Gaussian maximum likelihood classifier. It is a processor in LARSYS, a remote sensing data analysis system developed at LARS[6].
- CLASSIFY implements a sum-of-normal-densities

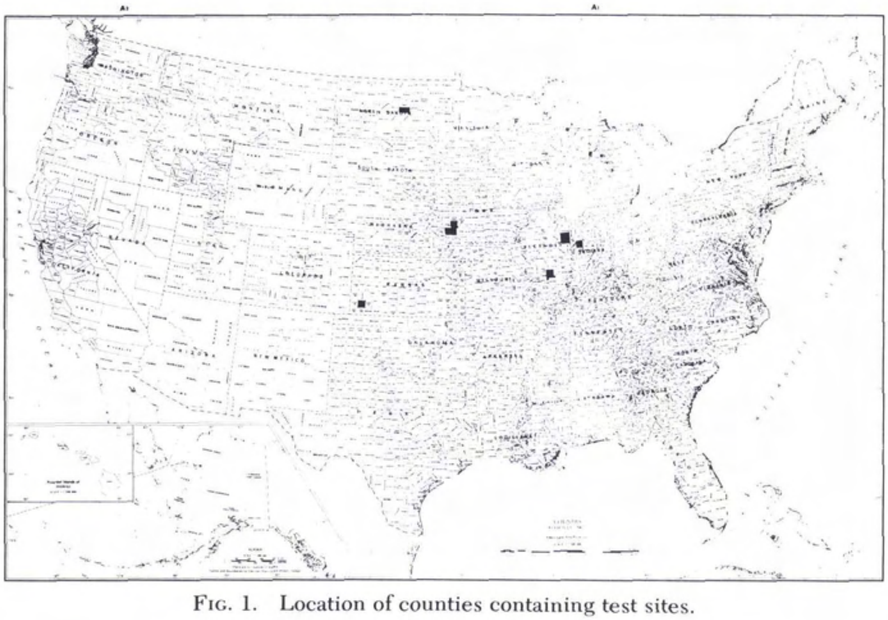

Fig. 1. Location of counties containing test sites.

TABLE 1a. MULTITEMPORAL DATA SETS FOR THE CORN BELT SITES

Corn Development Stage	Test Site				
	Illinois		Indiana	Iowa	
	Fayette	Iroquois	Tippecanoe	Pottawattamie	Shelby
	Date of Landsat Acquisition				
Emergence	6/10	6/12	6/10	6/16	6/16
Pretassel	6/29, 7/17	—	—	—	—
Tasseling	8/21	8/5	7/26	7/23	7/23
Blister	—	—	—	—	8/9
Dough	—	—	8/21	—	—
Dent	—	8/31	—	9/6	—
Mature	—	9/28	9/26	9/24	9/24

TABLE 1b. MULTITEMPORAL DATA SETS FOR THE WINTER (GRANT, KS) AND SPRING (FOSTER, ND) WHEAT TEST SITES

Wheat Development Stage	Test sites	
	Grant, KS	Foster, ND
	Dates of Landsat Acquisition	
Emergence	3/13	5/26
Heading	5/15	6/30
Soft Dough	6/2	7/19
Harvest	7/8	8/24

maximum likelihood classification rule which first assigns each pixel into an information category and then assigns the pixel to a spectral subclass within that category. It is a processor in EODLARSYS, developed at NASA/Johnson Space Center[7].

- MINIMUM DISTANCE is a linear classification rule which assigns each pixel to the class whose mean is closest to Euclidean distance[8]. It is a processor in LARSYS.
- The LAYERED classifier is a multistage decision procedure[9]. It utilizes decision tree logic with an optimum subset of features at each tree node to

TABLE 2. SPECTRAL BANDS USED IN CLASSIFICATION

Test Site	Landsat Acquisition Date	Spectral Bands Selected
		(μm)
Fayette	6/10	0.6–0.7
	6/29	None
	7/17	0.6–0.7, 0.8–1.1
	8/21	0.6–0.7
Pottawattamie	6/16	0.8–1.1
	7/23	0.6–0.7, 0.8–1.1
	9/6	0.7–0.8
	9/24	None
Shelby	6/16	0.6–0.7
	7/23	0.8–1.1
	8/9	0.8–1.1
	9/24	0.8–1.1
Tippecanoe	6/10	0.6–0.7, 0.7–0.8
	7/26	0.8–1.1
	8/21	0.7–0.8
	9/26	None
Iroquois	6/12	0.7–0.8
	8/15	0.8–1.1
	8/31	0.8–1.1
	9/28	0.6–0.7
Grant	3/13	0.8–1.1
	5/15	0.6–0.7
	6/2	0.6–0.7
	7/8	0.6–0.7
Foster	5/26	0.7–0.8
	6/30	0.7–0.8
	7/19	0.6–0.7
	8/24	0.8–1.1

classify each pixel using a Gaussian maximum likelihood decision rule. LAYERED is also a processor in LARSYS.

- ECHO (Extraction and Classification of Homogeneous Objects) utilizes both spectral and local spatial information[10]. Statistical tests are used to partition the image into homogeneous regions and each region is then classified using a Gaussian maximum likelihood sample classification rule. It was also developed at LARS and is part of LARSYS.

In order to insure that differences in classification accuracies were the result of classifier differences and not training methods, the same set of training statistics was used for all classifiers. The ground reference data were sampled to define training and test data sets. A 10- by 10-pixel grid was used to locate a systematic sample of points. If a grid intersection point fell into an agricultural field, then the field center pixels of that field were selected. Fields were selected to represent the classes of interest: corn, soybeans, and other cover types in the Corn Belt segments; and small grains and other cover types in the Great Plains segments. These selected fields were then randomly divided into a training set and a test set. The training fields were clustered to develop means and covariances to define spectral subclasses for each of the classes of interest. Training was based on 1.6 percent of the area in the Fayette site (as reference data were available on only about one-quarter of that site) and between 3.5 percent and 7.5 percent in the other sites.

Since CLASSIFY was designed as part of an automated analysis procedure without analyst intervention, a training method (referred to as ISOCLS) using a random selection of individual pixels to define initial cluster seeds for clustering the entire area is generally used in conjunction with that algorithm. Both training methods were used with CLASSIFY.

Accuracies were based on a pixel-by-pixel comparison of the test field classifications with ground inventory observations. Test fields were selected in the same manner as training fields and contained about the same number of pixels. The same set of test data was used with both training methods.

The results of the classifications were analyzed to assess the effects of segment and classifier on classification accuracy. The accuracies were transformed using arcsin $\sqrt{p}$ to increase homogeneity of variance[11]. For the Corn Belt segments, a two-way analysis of variance (ANOVA)[11] was run separately on the transformed accuracies for corn, soybeans, and other cover types and on the overall accuracy considering all cover types. For the wheat segments, ANOVA was run on wheat, other cover types, and overall accuracies. When classifier was found to have a significant effect on accuracy, the Duncan multiple range test was used to assess which classifiers were significantly different[12].

Results and Discussion

Classification Accuracy

The results of the classifications are shown in Table 3. Segment-to-segment variability was highly significant ($\alpha < 0.01$). Segment-to-segment variability embodies several potential sources of variation. The segments are located in different geographic areas having different soils, climate, and agricultural practices. The dates of Landsat MSS data acquisition are somewhat different for the segments. The wavelength band selection was carried out independently for each of the data sets. These effects cannot be quantitatively separated in the statistical analysis. It is likely, however, that the major source of variability is the location-dependent aspect since Landsat acquisitions from similar time periods were utilized and accuracies of the best subset of bands were believed not to differ greatly from use of all 16 wavelength bands.

Several factors contributed to the lower classification accuracies obtained in Fayette County: (1) the quality of multitemporal registration was only marginal, (2) the dates of Landsat data acquisition for Fayette were not as well distributed throughout the growing season as in the other counties, (3) less training data were available for the Fayette site, and (4) the training data were not well distributed geographically.

Pottawattamie and Tippecanoe Counties had larger field sizes, accounting in part for the relatively accurate classifications. Shelby County contained more confusion crops, including sorghum and spring oats, and had smaller field sizes than the other counties. Iroquois County was almost entirely corn and soybeans, making it difficult to obtain training for other cover types.

The climate of the Grant county site is semiarid and either irrigation or a summer fallow-wheat-sorghum crop rotation may be used. Some circular fields are present. Wheat, sorghum, and corn are the major crops in this region. The North Dakota site was more complex, including water bodies and some irregular field shapes. Spring wheat was the primary crop; other major agriculture included hay, pasture, and sunflowers.

Classifier differences were examined in separate analyses for the Corn Belt and wheat segments due to the differing cover types identified in the sites. The results for the Corn Belt segments will be discussed first followed by the results for the wheat segments.

Analysis of variance on the five Corn Belt segments was used to assess the classifier effect on percent correct classification of corn, soybeans, other cover types, and overall (Table 4). There was a significant effect of classifier ($\alpha = 0.05$) on all the accuracy variables. The Duncan multiple range

TABLE 3. COMPARISON OF CLASSIFIER PERFORMANCE (PERCENT CORRECT CLASSIFICATION) BY TEST SITE.

TEST SITE	CLASS	CLASSIFIER: MINIMUM DISTANCE	CLASSIFY POINTS	LAYERED	ECHO	CLASSIFY Using ISOCLS Stats[1]	CLASSIFY Using LARSYS Stats[2]	TEST SITE AVERAGE
Fayette, IL								
	Corn	81.9	81.2	63.9	77.3	77.3	78.9	76.8
	Soybeans	82.0	77.0	76.8	70.7	49.7	79.0	72.5
	Other	85.5	88.6	91.3	87.8	58.8	85.6	82.9
	Overall	83.5	83.0	80.5	79.5	61.1	81.6	78.2
Pottawattamie, IA								
	Corn	98.7	97.2	95.7	98.2	93.0	98.4	96.9
	Soybeans	92.0	89.8	92.3	90.2	86.5	89.3	90.0
	Other	85.3	98.0	97.5	97.1	92.1	98.4	94.7
	Overall	94.9	94.7	94.7	95.4	90.6	95.3	94.3
Shelby, IA								
	Corn	97.1	95.1	94.5	96.1	82.8	95.9	93.6
	Soybeans	89.3	92.9	98.2	95.4	98.0	98.0	95.3
	Other	75.5	83.7	88.2	79.4	78.7	79.7	80.9
	Overall	90.0	91.7	93.3	91.5	83.9	92.1	90.4
Tippecanoe, IN								
	Corn	93.7	89.9	91.5	86.4	99.4	93.1	92.3
	Soybeans	97.6	98.2	94.9	98.0	95.1	98.4	97.0
	Other	94.3	96.7	100.0	96.7	69.9	96.7	92.4
	Overall	95.5	94.3	94.0	92.7	94.2	95.9	94.4
Iroquois, IL								
	Corn	88.1	79.5	91.0	79.3	89.9	92.8	85.1
	Soybeans	82.8	85.2	78.1	83.6	78.8	86.3	82.5
	Other	76.4	72.7	0.0[3]	72.7	74.5	75.0	61.9
	Overall	84.9	82.1	80.5	81.2	83.6	84.2	82.8
Foster, ND								
	Small Grains	96.1	95.4	94.6	94.8	93.6	97.3	95.3
	Other	73.3	77.1	77.0	77.6	70.5	82.3	76.3
	Overall	82.7	84.7	84.3	84.8	81.3	89.3	84.5
Grant, KS								
	Small Grains	96.9	96.7	97.6	96.5	94.6	98.7	96.8
	Other	91.8	83.2	89.3	79.2	92.0	80.2	86.0
	Overall	93.1	86.5	91.4	83.5	92.6	84.8	88.6

[1] Training method generally used with CLASSIFY. Uses a random selection of individual pixels to define initial cluster seeds for clustering the entire area.
[2] Training method used with all other classifiers. Training fields were clustered to develop means and covariances to define spectral subclasses for each of the classes of interest.
[3] No pixels were classified into the class "other" in this segment using the LAYERED classifier.

TABLE 4. COMPARISON OF AVERAGE PERCENT CORRECT CLASSIFICATION FOR SEVERAL CLASSIFICATION APPROACHES

MAJOR CROPS	NO. SITES	CLASS	CLASSIFIER: MINIMUM DISTANCE	CLASSIFY POINTS	LAYERED	ECHO	CLASSIFY Using ISOCLS Stats[1]	CLASSIFY Using LARSYS Stats[2]
Corn/Soybeans	5	Corn	91.9	88.6	87.3	87.5	88.5	89.8
		Soybeans	88.7	88.6	88.1	87.6	81.6	90.2
		Other	85.4	87.9	75.4	86.7	74.8	87.1
		Overall	89.8	89.2	88.6	88.1	82.7	89.8
Small Grains	2	Small Grains	96.5	96.0	96.1	95.6	94.1	98.0
		Other	82.6	80.2	83.2	78.4	81.3	81.3
		Overall	87.9	85.6	87.8	84.2	87.0	87.0

[1] Training method generally used with CLASSIFY. Uses a random selection of individual pixels to define initial cluster seeds for clustering the entire area.
[2] Training method used with all other classifiers. Training fields were clustered to develop means and covariances to define spectral subclasses for each of the classes of interest.

test was used to assess which of the methods were different. For each variable, the five classifiers using the same training method resulted in accuracies which were not significantly different while the sixth method (CLASSIFY using ISOCLS statistics) resulted in significantly lower accuracies.

In the two wheat study sites, analysis of variance on measures of wheat, other cover types, and overall accuracies showed significant ($\alpha = 0.05$) classifier differences (Table 4). As in the Corn Belt segments, CLASSIFY using ISOCLS statistics produced accuracies which were significantly lower than any of the other classifiers. CLASSIFY using LARSYS statistics, however, gave significantly higher small grain classification accuracy (about 2 percent classification improvement over the other classifiers).

The performance of the ECHO classifier was not as high as anticipated (Table 4). This was probably due to the fact that the ECHO classifier required the analyst to set parameters defining cell size and homogeneity factors, and the optimal settings have not been determined.

In conclusion, given a representative set of training statistics, the choice of classification algorithm for differentiation of corn and soybeans from one another and from other cover types made relatively little difference in accuracy. Small grains, however, were more accurately identified using CLASSIFY than the other algorithms, as long as the same set of training statistics was used.

CLASSIFIER EASE OF USE

The classification schemes varied considerably in ease of use. Ease of use as defined in this study is a function of (1) analyst expertise required for setting classification parameters, (2) amount of time required for analyst interaction with the computer, and (3) available documentation. In increasing order of complexity the classifiers were found to be (1) MINIMUM DISTANCE, (2) CLASSIFYPOINTS, (3) CLASSIFY, (4) ECHO, and (5) LAYERED. The MINIMUM DISTANCE and CLASSIFYPOINTS classifiers were almost identical in ease of use.

CLASSIFY was designed as part of a total analysis scheme in which participation of the analyst is minimized in the clustering and definition of training statistics with control provided by a predefined set of analysis parameters. Although the classifier itself is not extremely complex, the training procedure typically used in this scheme involves a large number of parameters controlling the cluster split and combine sequences about which little is known.

ECHO utilizes both temporal and spatial information. The complexity of use for ECHO arises from the necessity of setting the parameters for cell homogeneity testing and cell size. The expertise of the analyst is essential in setting the parameters with regard to the data set used. The ECHO classifier is, however, one of the few available classifiers that utilizes spatial as well as spectral information in the classification process.

LAYERED implements a per point Gaussian maximum likelihood decision tree logic which requires the additional step of designing the decision tree. The decision tree is designed by obtaining class means and covariance matrices for all classes and using a feature selection algorithm to determine an optimal subset of features to be used at each node of the decision tree. The time needed by the analyst to design the tree can be significant if many spectral classes and features are needed to characterize the scene of interest. Although the decision tree can become very complicated and awkward to use, this classifier is particularly well suited for use with multitemporal or multitype data sets.

CLASSIFIER COST

Another important consideration in selecting a classification scheme is the computational cost per classification. In order of increasing cost per square kilometre for classification, not including cost for developing training statistics, were (1) MINIMUM DISTANCE (1.7 seconds), (2) ECHO (2.3 seconds), (3) LAYERED (2.3 seconds), (4) CLASSIFYPOINTS (3.8 seconds), and (5) CLASSIFY using ISOCLS statistics (11.3 seconds). These times were for execution on an IBM 370/148 computer at the Laboratory for Applications of Remote Sensing.

CONCLUSIONS

The results of this study show that, given a representative set of training statistics, the choice of classification algorithm for differentiation of corn and soybeans from one another and from other cover types made relatively little difference in accuracy. Small grains were more accurately identified using CLASSIFY than by other algorithms, as long as the same set of training statistics was used. However, the results for the CLASSIFY algorithm using a different training method did show a significant difference, indicating the major variable affecting classification accuracy was not the classifier but the training method used in generating the class statistics. A key aspect of training was that all cover types in the scene be adequately represented by a sufficient number of samples in each spectral subclass.

The ISOCLS training algorithm was a method designed for machine automation of a large portion of the training procedure. The statistical sampling method used for selection of training data is theoretically sound, so it is possible that the lack of analyst refinement of the training statistics seriously limited the performance. The clusters produced by this method were often of mixed cover types which may have adversely affected performance.

Additional variables of interest in the study were ease of use of the classifier and CPU cost per classification. Among the classifiers yielding similar classification accuracies, MINIMUM DISTANCE was the easiest for the analyst to use and cost the least per classification.

In conclusion, given Landsat MSS data sampled from the complete growing season and given a fixed method of obtaining training statistics, no difference among classifiers was found for Corn Belt segments while small grains were more accurately identified by CLASSIFY. The significant (α = 0.05) difference between CLASSIFY using the ISOCLS training method and all the other classification schemes suggests that development of representative training statistics was relatively more important for obtaining accurate classification than selection of the classification algorithm.

Acknowledgments

This work was supported by the NASA Johnson Space Center, Contract NAS9-15466.

References

1. MacDonald, R. B., M. E. Bauer, R. D. Allen, J. W. Clifton, J. D. Erickson, and D. A. Landgrebe, 1972. Results of the 1971 Corn Blight Watch Experiment, *Proc. Eighth Int. Symp. Remote Sens. Envir.*, Ann Arbor, Michigan, October 2-6, pp. 157-190.
2. Bizzell, R., F. Hall, A. Feiveson, M. E. Bauer, B. Davis, W. Malila, and D. Rice, 1975. Results from the Crop Identification Technology Assessment for Remote Sensing (CITARS) Project, *Proc. Tenth Int. Symp. Remote Sensing Envir.*, Ann Arbor, Michigan, October 6-10, pp. 1189-1196.
3. MacDonald, R. B., and F. G. Hall, 1978. LACIE: An Experiment in Global Crop Forecasting. *Proc. Plenary Session, The LACIE Symp.*, Houston, Texas, October 23-26, pp. 17-28.
4. Hixson, M. M., B. J. Davis, and M. E. Bauer, 1978. *Stratification and Sample Selection for Multicrop Experiments.* Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana. LARS Contract Report 112278.
5. Swain, Philip H., and Shirley M. Davis, ed., 1978. *Remote Sensing: The Quantitative Approach.* McGraw-Hill, Inc., New York.
6. Phillips, T. L., ed., 1973. *LARSYS User's Manual.* Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana.
7. Stewart, J., and P. J. Aucoin, 1978. *Earth Observations Division Version of the Laboratory for Applications of Remote Sensing System (EOD-LARSYS) User Guide for the IBM 370/148*, Volume I. System Overview. NASA, Johnson Space Center, Houston, Texas. JSC-13821.
8. Nilsson, N. J., 1965. *Learning Machines: Foundations of Trainable Pattern-Classifying Systems.* McGraw-Hill Book Co., New York.
9. Swain, P. H., C. L. Wu, D. A. Landgrebe, and H. Hauska, 1975. Layered Classification Techniques for Remote Sensing Applications. *Proc. Earth Resources Survey Symp.*, Houston, Texas, June 9, Vol. I-b, pp. 1087-1097.
10. Kettig, R. L., and D. A. Landgrebe, 1976. Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects. *IEEE Trans. Geos. Elect.* 14:19-25.
11. Anderson, Virgil L., and Robert A. McLean, 1974. *Design of Experiments: A Realistic Approach.* Marcel Dekker, Inc., New York.
12. Staff, 1979. *SAS User's Guide.* SAS Institute Inc., Raleigh, North Carolina.

(Received 21 March 1980; revised and accepted 4 July 1980)

CALL FOR PAPERS

Technical Symposium East '81

Washington, D.C.
20-24 April 1981

The Symposium, sponsored by the Society of Photo-Optical Instrumentation Engineers, will include sessions on atmospheric transmission, electro-optical instrumentation for resources evaluation, shuttle optical environment, techniques and applications of image understanding, laser spectroscopy for sensitive detection, ultraviolet and vacuum ultraviolet systems, technical issues in focal plane development, NASA-ESA Spacelab systems and programs, 3-D machine perception, infrared detector materials, and infrared astronomy—scientific/military thrusts and instrumentation.

Those interested in presenting a paper should submit four copies of a brief professional biography and four copies of a one-paragraph abstract (double-spaced on 8½ by 11 paper; 200 words maximum) by 15 December 1980 to

Technical Symposium East '81
SPIE Technical Programs Committee
P.O. Box 10
Bellingham, WA 98227
Phone: (206) 676-3290