General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

E77-10061

NASA CR-
ERIM 109600-70-F

NASA CR-
151001

Final Report

SIGNATURE EXTENSION USING TRANSFORMED CLUSTER STATISTICS AND RELATED TECHNIQUES

P.F. LAMBECK AND D.P. RICE
Infrared and Optics Division

MAY 1976

(E77-10061) SIGNATURE EXTENSION USING TRANSFORMED CLUSTER STATISTICS AND RELATED TECHNIQUES Final Report, 15 May 1975 - 14 May 1976 (Environmental Research Inst. of Michigan) 79 p HC A05/MF A01 CSCL 05B G3/43

N77-14560

Unclas
00061

Prepared for
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Johnson Space Center
Earth Observations Division
Houston, Texas 77058
Contract No. NAS9-14123, Task 15
Technical Monitor: Dr. A. Potter/TF3

ENVIRONMENTAL
RESEARCH INSTITUTE OF MICHIGAN
FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN
BOX 618 • ANN ARBOR • MICHIGAN 48107

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
ERIM 109600-70-F		

4. Title and Subtitle	5. Report Date
Signature Extension Using Transformed Cluster Statistics and Related Techniques	May 1976
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
P. F. Lambeck and D. P. Rice	109600-70-F

9. Performing Organization Name and Address	10. Work Unit No.
Environmental Research Institute of Michigan Infrared and Optics Division P.O. Box 618 Ann Arbor, Michigan 48107	Task 15
	11. Contract or Grant No. NAS9-14123
	13. Type of Report and Period Covered Final Technical Report May 15, 1975 through May 14, 1976
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Johnson Space Center Houston, Texas 77058	14. Sponsoring Agency Code

15. Supplementary Notes

The work was performed for the Earth Observations Division. Dr. Andrew Potter (TF3) was the technical monitor.

16. Abstract

Signature extension is a process intended to increase the spatial-temporal range over which a set of training statistics can be used to classify data without significant loss of recognition accuracy. This process is intended to help minimize the requirements for collecting ground truth and for extracting training statistics, thus allowing more timely and cost-effective surveys over large land areas. The reported effort has been primarily focussed on the problem of performing large area surveys to estimate wheat production, using MSS data from the Landsat satellites.

Current signature extension methods which have been developed or investigated at ERIM are presented. The discussion covers the underlying theory for signature extension, the development of cluster matching algorithms (specifically CROP-A and CROWN), the subsequent development of a signature extension operating system (PROCAMS), presentation of some results from signature extension tests, and conclusions and recommendations for future development in this field.

The results presented indicate some encouraging success for these techniques, but also indicate a need for improved means to estimate when and where the techniques can be applied effectively.

17. Key Words	18. Distribution Statement
Remote Sensing Multispectral Processing Signature Extension	Initial distribution is listed at the end of this document.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	ix + 82	

PREFACE

This report describes part of a comprehensive and continuing program of research in multispectral remote sensing of the environment from aircraft and satellites and the supporting effort of ground-based researchers in recording, coordinating, and analyzing the data gathered by these means. The basic objective of this program is to improve the utility of remote sensing as a tool for providing decision makers with timely and economical information from large geographical areas.

The feasibility of using remote sensing techniques to detect and discriminate between objects or conditions at or near the surface of the earth has been demonstrated. Applications in agriculture, urban planning, water quality control, forest management, and other areas have been developed. The thrust of this program is directed toward the development and improvement of advanced remote sensing systems and includes assisting in data collection, processing and analysis, and ground truth verification.

The research covered in this report was performed under NASA Contract NAS9-14123. The program was directed by R. R. Legault, Director of ERIM's Infrared and Optics Division and an Institute Vice-President, J. D. Erickson, Head of the Information Systems and Analysis Department and Project Director, and R. F. Nalepka, Head of the Multispectral Analysis Section (MAS) and Principal Investigator. The Institute number for this report is 109600-70-F.

The authors wish to acknowledge the administrative direction provided by Mr. R. R. Legault, Dr. J. D. Erickson, and Mr. R. F. Nalepka and the technical assistance given by Mr. R. F. Nalepka and Dr. R. G. Henderson. The prototype classification and mensuration system (PROCAMS) described in Section 6 was developed with the assistance of the entire MAS staff. Mr. R. Kauth is especially to be thanked for his contribution of the theory and formulation for the external effects correction algorithm presented in Section 7. Ms. D. Dickerson, E. Hugg, and J. Solosky are thanked for their secretarial assistance.

CONTENTS

CONTENTS (CONT.)

FIGURES

TABLES

viii

PRECEDING PAGE BLANK NOT FILMED

viii

PAGE INTENTIONALLY BLANK

SUMMARY

The general form of the transfer equation representing the recorded MSS signal level in each spectral band for a given material indicates that differences in recording conditions between a training scene and a recognition scene cause multiplicative and additive changes in the signal levels observed. Although changes in bidirectional reflectance are unique for each material, it is reasonable to develop signature extension algorithms which compare statistics from training and recognition scenes and which determine from this comparison coefficients for a multiplicative and additive transformation of training signature statistics. This signature transformation can then be used to change the training statistics to suit conditions in the recognition scene.

Two cluster matching algorithms, CROP-A and CROWN, which attempt to derive such a signature transformation have recently been developed at ERIM. The CROP-A algorithm imposes a linear ordering constraint on the cluster matches, while the CROWN algorithm, currently under development as an improvement upon CROP-A, takes a less stringent and more powerful approach to the cluster matching problem. Both algorithms use iterated regressions to eliminate poorly matched clusters from their computations, but still are partly limited in their application by occurrences of major dissimilarities between materials present in different scenes. Partitioning techniques (which aid in selecting optimum training and recognition scenes) offer hope for relieving this limitation.

Some supplementary procedures have been developed which can improve the performance of cluster matching algorithms. Gradient filtered clustering generates recognition clusters representing nearly pure materials, rather than recognition clusters which would include mixed materials. Reverse transform labeling inverts the signature

transformation to apply it to recognition clusters which can then be assigned labels (wheat or non-wheat) according to their classification of known materials within the training scene. This allows final classification of the recognition scene using clusters derived from that scene. The linear tasselled cap transformation (which transforms Landsat MSS data) may help to isolate the effects of soil variations, and can aid in reducing the dimensionality of the MSS data used for unitemporal or multitemporal signature extension applications.

An algorithm (CLUSTM) has been developed which aids in the definition of training and test field boundaries for use with multitemporal data. The algorithm produces a data image for locating the fields, which averages local misregistrations between time periods of the data set. The field definitions obtained are also practical for use with any subset of the time periods.

A signature extension operating system (PROCAMS), using the signature extension techniques currently available at ERIM, has been implemented. Testing is currently underway to gain information from this system regarding the partitioning requirements for these techniques.

A mathematical formulation for an external effects correction, which utilizes known physical parameters, has been defined. Incorporation of known physical information into signature extension techniques is a desirable improvement.

Current progress indicates that signature extension through the use of cluster matching algorithms appears to be a practical technique for economical and timely wheat surveys, using Landsat data, provided that the reasonable limits to its use (partitions) can be adequately determined. All aspects of the signature extension problem are continually undergoing examination, testing, and development toward the goal of attaining a practical and fully operational implementation of a signature extension capability. In particular, further improvements in dynamic partitioning, multitemporal applications, and preprocessing techniques are recommended.

2
INTRODUCTION

Signature extension is a process intended to increase the spatial-temporal range over which a set of training statistics can be used to classify data without significant loss of recognition accuracy. The training statistics which are required are extracted from multispectral scanner (MSS) data with the aid of training information (ground truth) obtained from localized surveys on the ground or from interpretation of aerial photographs or MSS data images by trained analyst interpreters (AI's). Either of these procedures for acquiring ground truth information becomes costly and time consuming even for data processing over land areas of moderate size.

The goal of signature extension is to minimize the requirements for collecting ground truth and for extracting training statistics, thus reducing the associated costs and time delays. Signature extension would then help to provide timely and cost-effective classification over extensive land areas, including remote areas for which ground truth information may not be readily available. This present signature extension effort has been primarily concerned with the problem of performing large area agricultural surveys to estimate wheat production, using MSS data from the Landsat satellites.

Many current signature extension techniques are based on a transformation of training statistics to compensate for changes in sun angle, atmospheric condition, etc., between a training area and a recognition area. Although preprocessing techniques [1,2,3] which minimize or eliminate the need for altering training statistics are also potential solutions to the problem of signature extension, the following presentation is principally concerned with those algorithms which define signature transformations based on associations between training and recognition area statistics. Specific topics to be discussed include:

1. The underlying theory for the signature transformation
2. The algorithms used to determine and to apply the transformation
3. Improvements in signature extension which can be effected through procedures which are peripheral to the transformation itself
4. Methods used to test and evaluate signature extension performance
5. A prototype signature extension operating system (PROCAMS)
6. An external effects correction algorithm.

3
THEORY

The general form of the transfer equation representing the recorded MSS signal level within a specific spectral band when viewing a given material α is expressed by

$$S_\alpha = G\ E\ T\ \rho_\alpha + G\ L_p + \delta \qquad (1)$$

G and δ represent gain and offset changes, respectively, in the response of the multispectral scanner instrument. E represents the irradiance through the atmosphere on the material, T represents the transmittance of the atmosphere over the path from the material to the scanner aperture, and L_p represents the path radiance along this viewing path due to atmospheric scattering. The bidirectional reflectance of the material α is given by ρ_α. All these variables are directly dependent on the wavelength of the signal being recorded, hence there is no significant interaction between signals at different wavelengths, in principle, and each spectral band can be treated separately from the others.

Note that whenever the bidirectional reflectance of each material remains constant, the signals recorded are related to the reflectance of each material by a simple multiplicative and additive relationship, although to determine these multiplicative and additive factors by trying to estimate values for each variable in the transfer equation is by no means simple. If one postulates a reference condition in which the above multiplicative factors all equal unity and the additive factors all equal zero, and if one realizes that the inverse of a multiplicative and additive transformation (MAT) is itself multiplicative and additive and that the concatenation of two MAT's is likewise, overall, multiplicative and additive, one can conclude that the data transformation needed to compensate for any or all of the effects above (with bidirectional reflectance held constant) will also be multiplicative

and additive. Furthermore, since there is no significant interaction between signals for different wavelengths, the required transformation may be determined separately for each spectral band.

One should be aware, however, that bidirectional reflectance does not, in general, remain constant for each material throughout a scene. Rather, reflectance is to be expected to vary differently for each material according to changes in illumination conditions (sun angle, relative intensities of direct and diffuse illumination), viewing angle, topography, crop or soil conditions (health of crop, density of ground cover, soil type, soil moisture content), crop orientation (due to wind), and cropping practice (methods of planting or harvesting). These effects, having a unique influence on the reflectance of each material, and varying sometimes from field to field or other times from county to county, cannot be fully compensated by a transformation applied indifferently to data from any and all materials in a scene. At best one can devise a general transformation or means for data manipulation which treats these disparate effects only in an average way, or which takes advantage of some salient characteristic of the major materials of interest. (An example of the latter approach would be a classifier which takes advantage of multitemporal information and a knowledge of the characteristic growth cycle of a particular crop, e.g., winter wheat [2].) Variations in bidirectional reflectance should be recognized as one of the major potential stumbling blocks for signature extension. Other potential stumbling blocks are enumerated in the following discussion.

4
SIGNATURE TRANSFORMATIONS

4.1 DERIVATION

Scanner data from a given material is usually assumed to be represented by the multivariate Gaussian probability density function

$$P_\alpha = \frac{\exp[-\frac{1}{2}(x-\mu_\alpha)^T \theta_\alpha^{-1} (x-\mu_\alpha)]}{(2\pi)^{n/2} |\theta_\alpha|^{1/2}} \qquad (2)$$

P_α is the probability that a given signal x corresponds to the material α, exclusive of any competing probabilities associated with other materials. The data vector x represents the recorded signal levels in each spectral band of the MSS for a single measurement. The vector of mean values for the signature of material α is μ_α, and θ_α is the variance-covariance matrix; together they form the "signature" of material α. All the vectors have n omponents and the matrix has n×n components, with n being the number of spectral bands used in signature.

As a means to compensate for changes in bidirectional reflectance in an average way and to compensate for the multiplicative and additive effects arising from changes in the other variables of the transfer equation (Eq. (1)), a signature transformation may be proposed which alters signatures derived from one scene to match, at least approximately, the conditions present within a second scene. If one assumes that the differences between observed signal levels in the two scenes are purely multiplicative and additive, then the signals are related by

$$x' = A\,x + B \qquad (3)$$

in which x' represents the observed signal from the second scene, while x represents a corresponding signal from the first scene. The quantity A is a diagonal n×n matrix whose non-zero components represent the multiplicative changes to the signals in each spectral band, and

B is a vector with n components, representing the additive changes. The signature transformation corresponding to this multiplicative and additive change in signal levels is given by

$$\mu'_\alpha = A\,\mu_\alpha + B \tag{4}$$

and

$$\theta'_\alpha = A^T\,\theta_\alpha\,A \tag{5}$$

One should note that Equation (5) applies for data which excludes measurement noise inherent in the scanner instrument. When a signature is extracted from a scene, this measurement noise becomes a part of the variance-covariance statistics for the signature, changing those statistics from their purely scenic values in a strictly additive fashion. Ordinarily signature extension is attempted between scenes recorded with the same MSS instrument, hence the measurement noise for each scene should be nearly the same, regardless of any changes in the variables of Equation (1). Equation (5) should only apply to that portion of the variance-covariance statistics which excludes measurement noise. Depending on the source of the measurement noise, some other form of transformation may or may not be appropriate for the noise statistics. Since the nature of the measurement noise for Landsat data is uncertain, and since, to date, we have found that transforming the variance-covariance matrix produces little change in the results of signature extension applications, the approach at ERIM so far has been not to use Equation (5), leaving the variance-covariance statistics unchanged, and to use only Equation (4) for signature transformations.

4.2 IMPLEMENTATION

Given that a signature transformation is desired to compensate for multiplicative and additive differences between two scenes, the task is next to determine the appropriate coefficients, A and B, for

Equation (4). In general one needs for this purpose some effective way to compare the data from the two scenes. One method for accomplishing this is to compare cluster statistics for the scenes.

Clusters are usually represented by multivariate Gaussian probability density functions which, when weighted according to the amount of data in a scene which generated the statistics of each cluster, and when summed together, approximate the multivariate histogram distribution for the scene. Clusters are generally assumed to be equivalent to signatures for more or less unknown but spectrally distinct materials, which represent modes of the data distribution from which the clusters were generated. The extent to which clusters actually represent modes of the data distribution depends to a great extent on the nature of the clustering algorithm which is used, however, regardless of the algorithm used, the clusters when taken together generally do represent adequately the variability to be found in the scene. The advantage in using cluster statistics for comparing data from scenes recorded under different conditions is that distinct materials by their presence give rise to representative clusters, and the cluster statistics (mean values, variance, or covariance) are not particularly sensitive to the frequency of occurrence of the materials within the scene. Hence, a valid comparison of recording conditions for two scenes requires only that clusters for similar materials be compared, rather than that the frequency of occurrence of the materials compared between scenes also be similar.

Once one has obtained a valid association between pairs of clusters from two scenes, a least squares estimate may be determined for the coefficients A and B of Equation (4) by solving the following two equations once for each spectral band to be used.

$$\frac{\partial}{\partial A}\left[\sum_i \left(\mu_i' - A\,\mu_i - B\right)^2\right] = 0 \qquad (6)$$

$$\frac{\partial}{\partial B}\left[\sum_i \left(\mu_i' - A\,\mu_i - B\right)^2\right] = 0 \qquad (7)$$

The index i identifies each cluster pair. The summations are over all cluster pairs. The mean value for the ith training scene cluster in the spectral band being considered is represented by μ_i, while μ_i' represents the mean value for the ith recognition scene cluster in the same spectral band. These equations lead to a pair of simultaneous linear equations which can be solved for the coefficients A and B in each spectral band, yielding

$$A = \frac{N \sum_i \mu_i \mu_i' - \sum_i \mu_i \sum_i \mu_i'}{N \sum_i \mu_i^2 - \left(\sum_i \mu_i\right)^2} \tag{8}$$

$$B = \frac{\sum_i \mu_i^2 \sum_i \mu_i' - \sum_i \mu_i \sum_i \mu_i \mu_i'}{N \sum_i \mu_i^2 - \left(\sum_i \mu_i\right)^2} \tag{9}$$

in which N is the total number of cluster pairs used in the regression. Again it should be realized that Equations (8) and (9) produce scalar values for A and B which are appropriate for the specific spectral band chosen. These equations need to be solved again for each additional spectral band used, to obtain the final A and B coefficient matrix and vector, respectively, indicated in Equation (4).

Since the clusters which are paired in the regression to calculate A and B must be finite in number, there is a practical limit to the accuracy with which the A and B coefficients can be determined, even with all cluster pairs being valid. Of course the multiplicative and additive transformation sought cannot compensate perfectly for all the real physical causes of the difference between the training scene and the recognition scene anyway, however in principle it is best to try to use as many valid cluster pairs in the regression as possible. Current signature extension tests at ERIM have tended to use between

10 and 20 cluster pairs for obtaining the A and B coefficients, out of a maximum of from 15 to 30 cluster pairs which were possible.

The first basic cluster matching algorithm, called MASC (for Multiplicative and Additive Signature Correction) [1], was developed at ERIM to test the cluster regression approach to determining the A and B coefficients. While this algorithm achieved some occasional successes at signature extension, it did not include a means to adequately select only valid cluster pairs, a serious requirement for achieving practical results. The task was then to automate a procedure for selecting those few valid cluster pairs which might exist among the great many arbitrary pairs which were possible.

The difficulty involved in identifying valid cluster pairs may be appreciated by considering Fig. 1, which shows one set of cluster pairings from a matrix representing all 100 possible cluster pairs between a set of 10 training scene clusters and a set of 10 recognition scene clusters.

		Training Scene Clusters									
		1	2	3	4	5	6	7	8	9	10
Recognition Scene Clusters	1										
	2	0									
	3					0					
	4			0							
	5						0				
	6										
	7							0			
	8									0	
	9								0		
	10										0

FIGURE 1. POTENTIAL CLUSTER PAIRS

For the purpose of better illustrating a point to be brought up later, an equal number of training clusters and recognition clusters has been assumed, although the number of clusters obtained from each scene in

practice turns out to be equal only occasionally. Also, for simplicity, a smaller than usual number of clusters has been assumed. The 0's in the matrix represent a hypothetical set of valid cluster pairs for this illustration. By ordering the sequence of the training scene and recognition scene clusters appropriately, these valid pairs may be made to fall close to the diagonal of the matrix, about as shown. If one tries to examine all possible sets of 10 cluster pairs to find which is best, one finds that there are 10! (=3,628,800) sets of pairs to be considered, assuming that there are no multiple pairings with the same cluster.

Obviously there are two basic difficulties to be dealt with in finding the valid cluster pairs from which to derive the required signature transformation. The first is to reduce the number of different sets of cluster pairs which need to be examined, and the second is to determine which among those several candidate sets of cluster pairs are most likely to be valid. The first attempt at ERIM toward solving the first of these two difficulties was to sort the training scene and recognition scene clusters according to their mean values in some designated spectral band, then to consider only those sets of cluster pairs which preserved that linear ordering. This procedure occasionally led to situations such as that shown in Figure 2. The X's indicate the one set of 10 cluster pairs that is permitted, subject to the cluster ordering constraint, when there is an equal number of training and recognition clusters from which to choose. The 0's again indicate the hypothetical set of valid cluster pairs specified in Figure 1. When the number of clusters in the training set differs from the number in the recognition set, the linear ordering constraint becomes less restrictive, as will be shown below. Note that of the 8 valid cluster pairs available, only two are within the candidate match indicated in Figure 2.

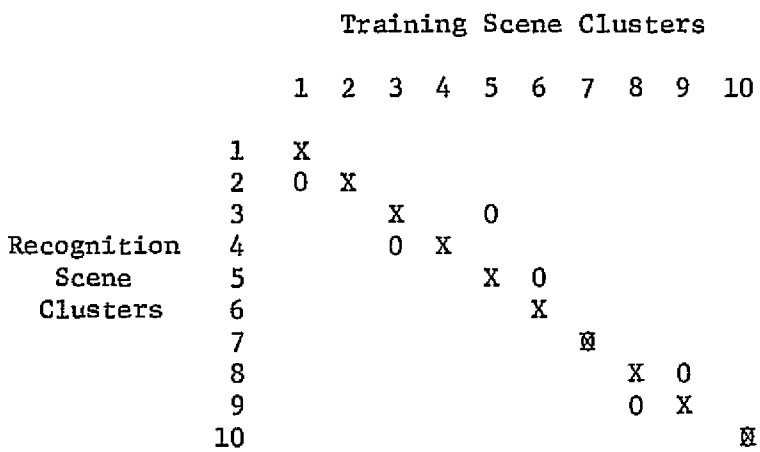

FIGURE 2. LIMITED POTENTIAL CLUSTER PAIRS AFTER LINEAR ORDERING CONSTRAINT (EXAMPLE)

An improved cluster matching algorithm, called CROP-A (for Cluster Regression Ordered on Principal-Axis), was developed at ERIM and has evolved to include a partial remedy for the linear ordering constraint difficulty indicated in Figure 2. The name for this algorithm comes from its choice of the principal eigenvector of the covariance of the training signature means as the linear direction for the cluster ordering constraint. Cluster positions along this ordering axis are determined from an apparent mean value for each cluster, given by a dot product between the cluster mean vector and a unit vector aligned with the principal eigenvector. Improvements in signature extension performance due to using this cluster ordering direction instead of using a particular spectral band appear to be mostly inconsequential, however the other new features contained in the algorithm appear to reap substantial benefits. In particular, the algorithm contains a provision to force a difference to occur in the number of training clusters and recognition clusters which are to be paired. For this purpose the algorithm keeps track of the number of data values used to generate each cluster. First, clusters generated from less than 1% of the data

used to generate all clusters in the same set are excluded from being paired at all. This eliminates some of the "false alarm" clusters derived from minority constituents of a scene, which may be less likely to have counterparts in another scene. The percentage threshold for excluding clusters is then increased above 1% for one of the two sets of clusters (whichever requires the least number of additional exclusions) until a desired difference in the number of clusters remaining in the two sets is reached. Ordinarily the increased threshold is less than 2% when this condition is obtained. For cluster sets of between 15 and 30 clusters, a forced difference of 4 in the number of clusters is currently used, producing between 1000 and 30,000 candidate sets of cluster pairs. This situation is simulated in miniature in Figure 3.

		Training Scene Clusters									
		1	2	3	4	5	6	7	8	9	10
	1E										
	2	⊠	X	X	X						
	3		X	X	X	⊠					
Recognition	4E			0							
Scene	5			X	X	X	⊠				
Clusters	6E										
	7				X	X	X	⊠			
	8					X	X	X	X	0	
	9						X	X	⊠	X	
	10							X	X	X	⊠

FIGURE 3. LESS LIMITED POTENTIAL CLUSTER PAIRS AFTER CROP-A FORCED DIFFERENCE

Recognition clusters eliminated by the requirement for a forced difference of 3 in the number of clusters in the two sets are designated (hypothetically) by the letter "E". The candidate cluster matches available from Figure 3, subject to the cluster ordering constraint, consist of sets of pairs designated by X's, one from each row,

such that the chosen X's can be joined in sequence by a monotonic broken line segment. This requirement is equivalent to matching all possible subsets of 7 training clusters with the 7 retained recognition clusters, in sequence. In this simple case one obtains 120 (10!/7!/3!) candidate sets of 7 cluster pairs, rather than the single candidate (with 10 pairs) indicated in Figure 2. Also, one of the available candidates (in this case, with 7 pairs) now contains 5 valid cluster pairs, compared to only 2 for the candidate (with 10 pairs) in Figure 2. This new candidate is shown in Figure 4.

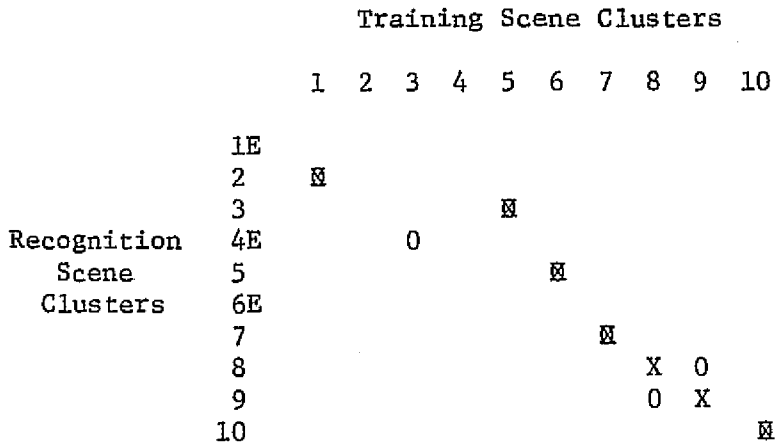

FIGURE 4. OPTIMUM CANDIDATE CLUSTER MATCH AFTER CROP-A FORCED DIFFERENCE

Note that the pairing of recognition cluster #9 with training cluster #8, although potentially allowed by the CROP-A forced difference (Figure 3), would by its choice in a candidate exclude from that candidate, due to the ordering constraint, the valid pairings with recognition clusters #3, #5, and #7. Hence, at best this alternate candidate could only contain 3 valid pairs. This sort of limitation is not uncommon when a linear ordering constraint is used. The result is that not all of the valid cluster pairs can be selected by the algorithm at one time.

As a potential solution to the somewhat severe restrictions occasionally imposed by the CROP-A linear ordering constraint, another cluster matching algorithm, called CROWN (for Cluster Regression Ordered With N channels), is currently undergoing development and testing at ERIM. This algorithm uses a matrix of merit figures, one figure for each possible cluster pair, to allow apparent optimum cluster associations to be chosen one by one until a specified number of candidate sets of a fixed number of cluster pairs become available. The merit figures for the matrix are determined on the basis of similarities in the location of each training and recognition cluster within its respective overall cluster distribution. This technique appears to be satisfactory for reducing the complexity of the cluster matching problem without excluding any significant number of valid pairs from consideration.

Having devised a means to select a practical number of candidate cluster matches, one next needs to find the best candidate among those chosen and to determine which of the cluster pairs from that candidate are most likely to be valid. Both CROP-A and CROWN use the regression procedure itself to perform this selection. Presuming that invalid cluster pairs will tend not to match as closely as the valid pairs, these algorithms delete from the regressions performed for each candidate match those cluster pairs which appear to match the most poorly. This is accomplished by comparing the transformed training cluster mean values to the untransformed recognition cluster mean values for each cluster pair. The mean values are first compared within the individual spectral bands as each separate regression is performed (Equations (8) and (9)), since this is computationally the earliest opportunity to delete a cluster pair from the subsequent calculations. The cluster pair deleted after each iteration through the regression is the one among those with a difference in mean values in excess of a specified threshold (currently 4.0 counts), which has the largest

difference in mean values. This iterative procedure continues until a stable situation is reached, with the regression in each spectral band updated to reflect deletions caused by the thresholding in any of the spectral bands. The RMS distance between the remaining cluster mean values is then tested, using an average over all spectral bands. If the greatest RMS distance is more than a second threshold (currently 6.0 counts), all cluster pairs with RMS distances greater than the average of the greatest RMS distance with this second threshold are deleted. The regressions are then updated accordingly and the test is repeated until once again the situation becomes stable. If at this point any of the deleted pairs now matches with an RMS distance less than a third threshold (currently 6.0 counts), the pair is restored and the regressions are updated just once more. This procedure has seemed to be quite effective. Candidate matches, with poorly matching cluster pairs deleted, are then compared to select the final result. The final result selected is that which has the minimum RMS mismatch between clusters, comparing averages over a specific fixed number of the "best" pairs from each candidate. Typically for CROP-A this final selection is based on the best 67% of the cluster pairs in each match (whether deleted or not), while for CROWN it is based on the best 90%. Note, however, that the CROWN algorithm contains a provision to automatically select the number of cluster pairs which are reasonable to constitute a candidate, and that this number may sometimes be less than the number of pairs required for a CROP-A candidate, although the CROWN algorithm generally retains numerically more cluster pairs in its final result than does CROP-A.

Although the above candidate selection procedures and the subsequent iterated regressions with step by step deletions of poorly matched cluster pairs have seemed to be quite effective, it has for some time been apparent that the performance of cluster matching algorithms is limited by a fundamental difficulty somewhat allied with

the problems caused by variations in bidirectional reflectance, mentioned earlier. This limitation occurs when there are an insufficient number of valid cluster pairs to be found, as happens when scenes contain dissimilar major constituents. Such major differences between scenes may arise simply from differences in crop varieties grown (different rates of growth), or from differences in crop treatment (fertilization or irrigation), as well as from more fundamental differences (different crops). Major differences between scenes constitute another potential stumbling block for signature extension. A method (partitioning) for partially alleviating this problem will be briefly discussed later (in Section 5.5).

5
PERIPHERAL PROCEDURES

The manner in which a signature extension module, such as CROP-A or CROWN, is embedded in an overall signature extension system has been identified as an important consideration in determining its performance and value as a signature extension tool. In this regard, research is currently underway at ERIM to define an optimum signature extension system, utilizing the current state of the art. Some particular techniques being tested are discussed below.

5.1 BOUNDARY PREPROCESSING FOR CLUSTERING

The effectiveness of many signature extension techniques, as well as the quality of recognition, depends on the manner in which clusters are produced. This section describes ways in which clustering can be improved by restricting pixels used in forming clusters to those which are very likely to be field center pixels.

Since Landsat data is made up of many pixels, each representing an instantaneous field of view of approximately 79 meters square, these pixels often contain a mixture of signals from more than one material. For typical scenes in the Great Plains, often 50% or more of the Landsat pixels straddle field boundaries and hence contain mixed signals. Since such mixture pixels can have adverse effects on cluster statistics, which ideally should describe the true ground cover classes that occur in a scene, we seek to generate the statistics using pixels which represent only pure materials. Within a training scene, where the training field boundaries are known, one can cluster over pixels within field boundaries and obtain relatively clean statistics. However, within a recognition scene, information on field boundaries is not available.

Any edge detecting technique can be of value in eliminating boundary pixels from clustering. In this section, the use of a gradient edge detector (GRAD) will be discussed. Another possible

edge detecting technique is a spectral-spatial cluster technique, wherein pixels consist not only of spectral bands, but also of bands containing line and point information [4].

The technique called GRAD works as follows. For each pixel, differences between opposing pairs of its eight neighbors are formed to give an estimate of the spatial rate of change in signal value, or gradient. The gradient value is a measure of the nonuniformity, and thus of the likelihood that the pixel is a mixture pixel.

Consider a pixel E from line n, with point number m, and its eight neighbors A,B,... as shown below.

		Point	
Line	m - 1	m	m + 1
n - 1	A	B	C
n	D	E	F
n + 1	G	H	I

The calculation of gradient measure g for pixel E is

$$g = \sum_i |\ell_i| + |p_i| \tag{10}$$

where i is the channel number, and the line rate of change ℓ_i is

$$\ell_i = 2(G_i - A_i) + 3(H_i - B_i) + 2(I_i - C_i) \tag{11}$$

and the point rate of change p_i is

$$p_i = 2(C_i - A_i) + 3(F_i - D_i) + 2(I_i - G_i). \tag{12}$$

(There are other possible gradient estimators, such as the largest difference between pixel E and any of its neighbors.)

Once gradient values are established, a threshold is used to reject a specified fraction of pixels as being probable boundary pixels. Typically 75% is used as a compromise between accepting too many pixels, some of which will be mixtures, and accepting too few pixels so as to run the risk of rejecting high texture, field center pixels, or of using insufficient data to get sound statistics. Rejected pixels are coded so that they will not be used in clustering.

The question of evaluating the performance of a boundary pixel detector is an interesting one. In the first place, it is difficult to define which pixels should be called pure, even with full information. For example, if two similar wheat fields adjoin each other, pixels on the field boundary are boundary pixels, but are not mixtures of different materials. A gradient method would call them field centers. If the two wheat fields differ somewhat, but not radically, it is not clear whether a "perfect" algorithm should flag the pixels along their boundary as mixtures. How one should detect and flag almost pure pixels is a difficult question.

Aside from the conceptual problem of defining which pixels should be called mixtures, one can still get some indication of performance by examining the fraction of pixels, known to be well inside field boundaries, which GRAD detects as being pure, using a variety of detection thresholds. For one scene, this information is shown in Figure 5.

To compare the gradient technique to the procedure of accepting the given percentage of pixels from the scene at random, we mark the dashed 45 degree line on the figure. The results from GRAD are visibly better than random.

In order to assess performance more quantitatively, we consider the two kinds of error -- type I error, where field center pixels are mistakenly called probable mixture pixels; and type II error, where pixels called probable pure pixels are in fact mixture pixels. Type I error is not believed to be serious unless so many pixels are rejected

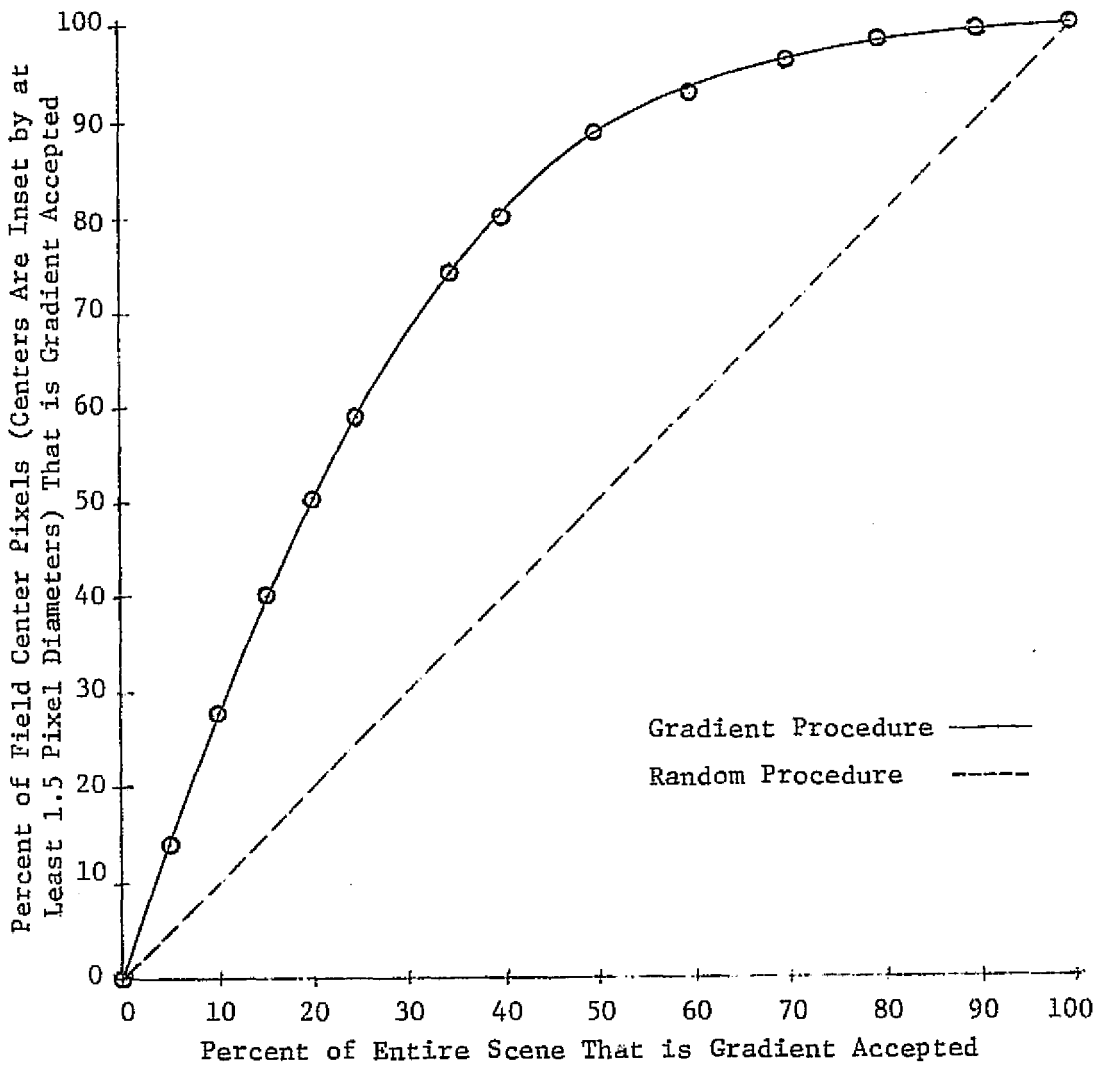

FIGURE 5. THE EFFECTIVENESS OF THE GRAD PROCEDURE IN ACCEPTING FIELD CENTER PIXELS, USING FINNEY ITS DATA FOR 26 MAY 1974

that the variability of the important pure classes in the scene is not adequately represented. Type II error is contrary to the purpose of gradient filtering in that mixtures of pure classes may be present in unwanted abundance.

For various gradient thresholds, type I error can be inferred directly from Figure 5. In the scene illustrated, the threshold which accepts 25% of the pixels in a scene (a number less than the number of actual field center pixels, which is estimated variously around 50% of the scene), accepts a full 60% of the field center pixels whose centers are conservatively within a 1.5 pixel diameter inset from the field boundaries. This inset is sufficient to ensure that the eight neighbor pixels used in the gradient computation are generally within the same field. As the inset is decreased to the value 0.5, which does not assure that the eight neighbors are all in the same field, but still assures that the pixel is within the field, only 35% of this larger class of field centers are accepted by the same 25% gradient threshold. This latter percentage is of little concern, since we prefer to handle only pixels conservatively inset from actual field boundaries.

Type II error concerns us somewhat more. To determine the number of pixels which are called pure by the gradient threshold when they in fact fall on a boundary, we tabulate gradient accepted pixels falling inside a field boundary inset by 0.5. Because this inset at least approximately distinguishes field center pixels from boundary pixels, gradient accepted pixels not tabulated are considered to have fallen on boundaries. In the scene used for Figure 5, one in three gradient accepted pixels falls on a boundary for the linear, low threshold (up to approximately 30% of scene accepted) portion of the curve. This is an upper bound on the type II error, since many of the boundaries separate fields with the same type of crop, in which case the technique is not expected to find the boundary. The actual type II error is believed to be much less, maybe 10 or 20 percent, but it has not been measured at this time.

Figure 6 shows the other scene tested. Here results are better than the first scene. Over 65% of conservative (1.5 pixel inset) field centers are accepted with a 25% scene threshold, and approximately one gradient accepted pixel in five is on a boundary, even though the second scene has a slightly smaller average field size and an estimated slightly larger percentage of boundary pixels in the scene.

From the above analysis, it seems likely that for typical scenes pixels passed to the clustering algorithm would be concentrated so as to produce a proportion of pure pixels on the order of 80-90%, a substantial improvement from a typical proportion of 40-60%.

Next we turn to the effect gradient filtering has on clustering. By virtue of clustering primarily non-mixture pixels, modes of impure classes should be somewhat suppressed. This effect is illustrated in Figure 7. Here, a smoothed histogram of Landsat Band 6 was formed over the scene, tabulating only pixels with a Landsat Band 5 value of 26. This was done first for the full scene, and second for the gradient filtered scene, with 75% of the pixels excluded by the gradient threshold. The latter curve was scaled upward by a factor of 4 so that the curves have approximately the same area beneath them. Each curve then represents one strip out of a smoothed, two dimensional histogram, and the particular strip was chosen to pass through one major mode (Band 6 value near 33), corresponding to wheat, and to fall on or near another mode (Band 6 value near 50), corresponding to pasture. The wheat mode was clearly accentuated by the technique, the pasture mode was sharpened, and the region between modes was depressed.

The effect of this on clustering is twofold. First, clusters which occur between modes would have a relatively smaller number of pixels, so that they can be given proportionately small consideration in subsequent operations. The second effect relates to how clusters are formed. In most clustering techniques, once a cluster is started, it receives points which are nearby in signal space, but not points

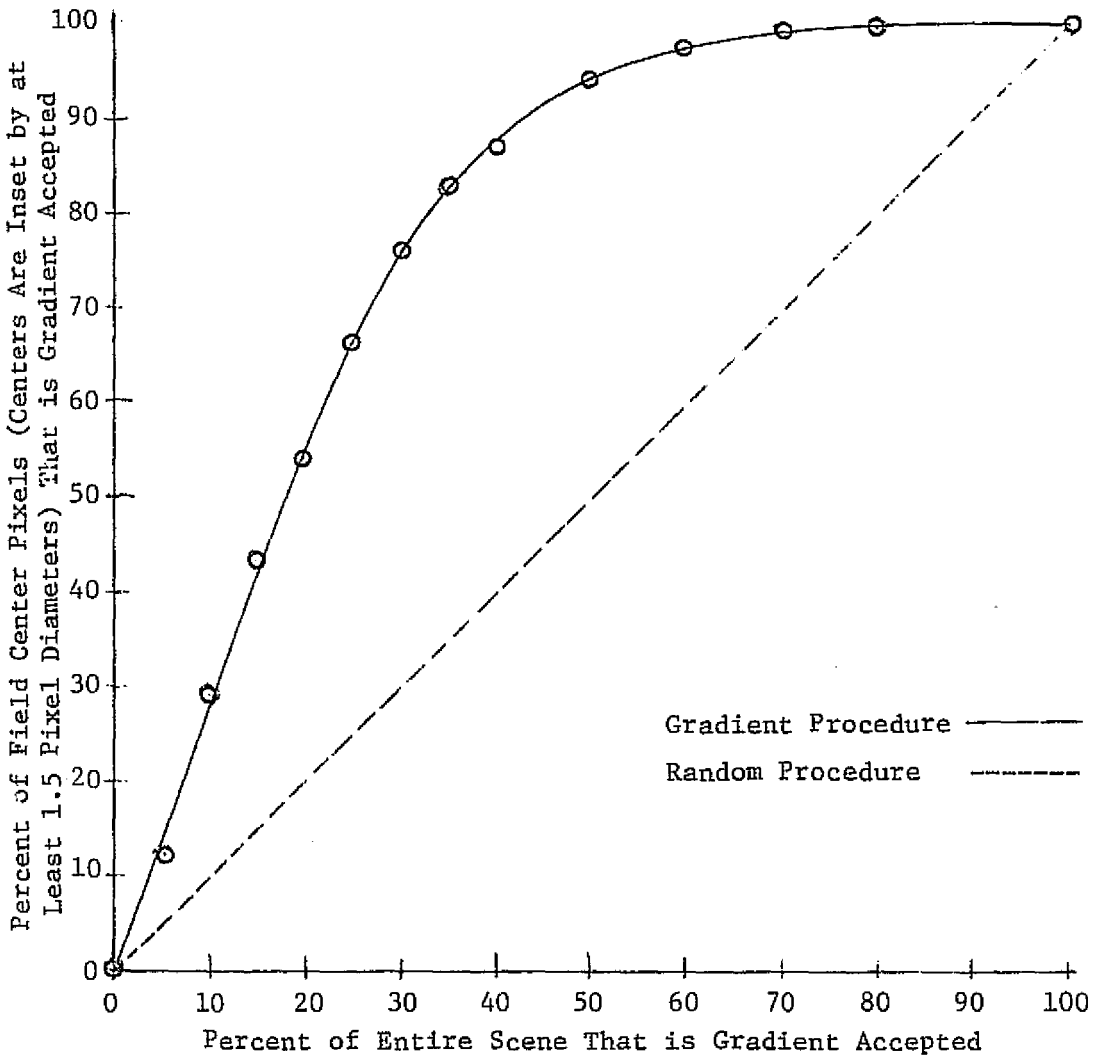

FIGURE 6. THE EFFECTIVENESS OF THE GRAD PROCEDURE IN ACCEPTING FIELD CENTER PIXELS, USING ELLIS ITS DATA FOR 12 JUNE 1974

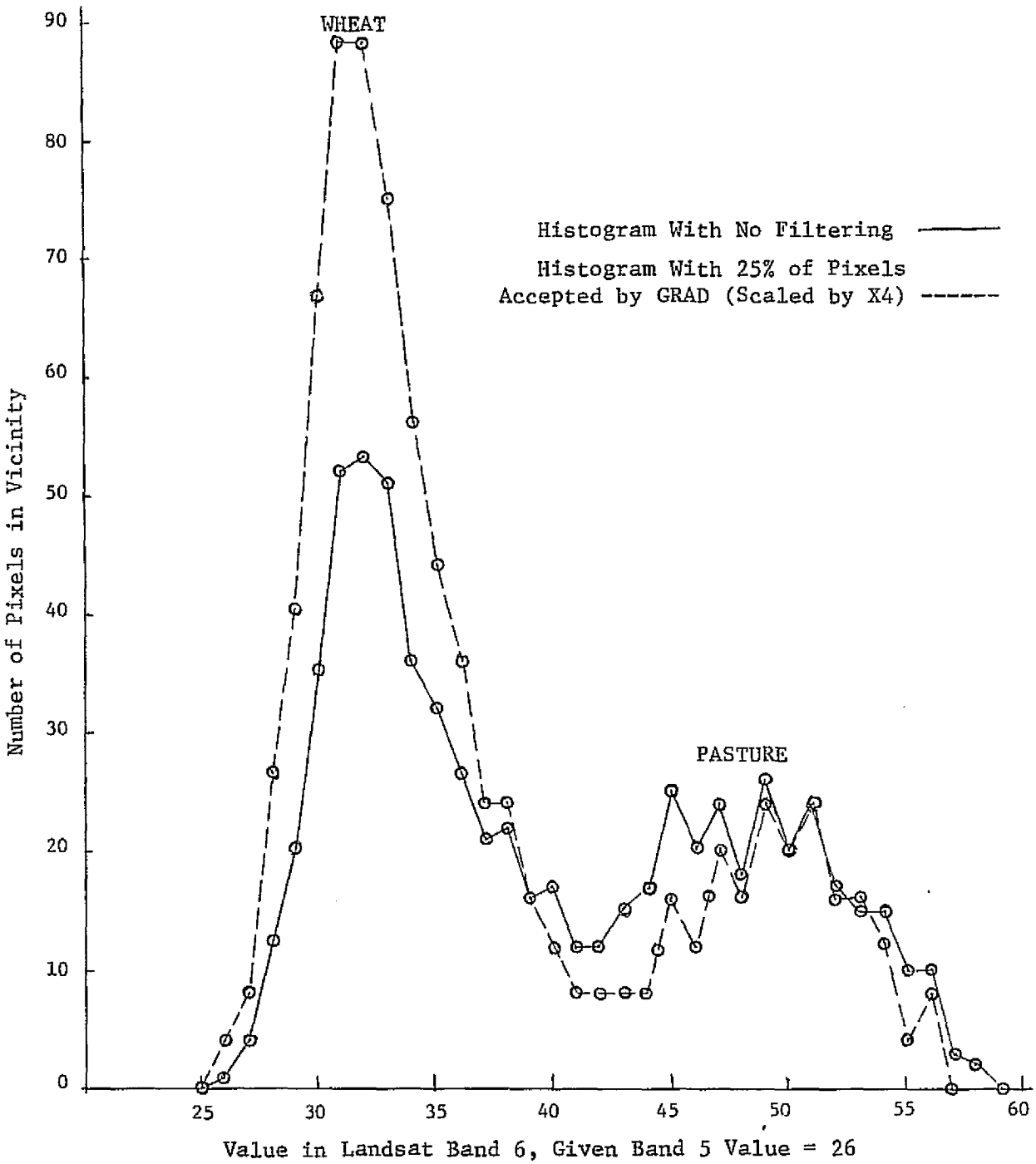

FIGURE 7. COMPARISON OF HISTOGRAM MODES WITH AND WITHOUT GRADIENT FILTERING

which are near other clusters that are forming. Thus, there is a tendency for clusters, once started, to grow away from each other, and to more or less uniformly cover the populated regions of signal space. By depopulating the regions which do not correspond to pure classes, we expect to limit the growth of clusters toward such regions, and concentrate the final clusters nearer true modes.

5.2 REVERSE TRANSFORM LABELING

Still more improvement in signature extension performance might be expected to result from optimizing the way in which the transformed and untransformed clusters are used. With this in mind, ERIM has developed a technique called reverse transform labeling. This technique, rather than transforming training scene clusters to match the recognition scene, transforms the recognition scene clusters to match the training scene. The local classification of the training scene by the training clusters is then compared, pixel by pixel, to the classification of the training scene by the transformed recognition clusters. The number of pixels classified locally as wheat or non-wheat and also classified by each transformed recognition cluster are tabulated to be used as votes for labeling the clusters as wheat, non-wheat, or undecided. A cluster is labeled undecided whenever fewer than 10 votes are obtained, or whenever fewer than two thirds of the votes favor the majority. The untransformed recognition clusters, with these labels, but with undecided clusters excluded, can then be used to classify the recognition scene. This technique depends only on determining a signature transformation accurate enough to produce proper recognition cluster labels from the training scene information, and may be especially effective if, due to gradient filtering, the recognition scene clusters can be made to represent mostly pure materials. Thus, this approach is less sensitive to minor inaccuracies in the signature transformation coefficients and to minor variations in bidirectional reflectance between training and recognition scenes than the transformation

of training signatures to classify the recognition scene.

5.3 MULTITEMPORAL TECHNIQUES

Some other improvements in signature extension performance should result from using additional sources of information. A generalized mathematical technique which can, in principle, utilize virtually any form of additional information, provided that appropriate mathematical relationships are known, is discussed in Section 8. A more restricted technique is to use MSS data from additional time periods as supplementary sources of information. Such a use of multitemporal data is an especially reasonable means to augment current unitemporal techniques, since analyst interpreters attempting to define training fields by examining Landsat imagery routinely use not only the most recently acquired data, but preceding data as well, in order to ensure the best choice of training fields. Since signature extension is intended to reduce the requirements for extensive interpretation of imagery (or ground based surveys), without significant loss of classification accuracy, it is reasonable that signature extension techniques incorporate the same information that is used by the analyst interpreters. Since clouds often obscure training or recognition scenes during Landsat data acquisitions, multiple or alternate training scenes are required in order to most effectively implement multitemporal signature extension techniques. The details of a multitemporal training procedure (which incorporates the reverse transform labeling technique) are discussed in Sections 7.5 and 7.6. Of course multiple training scenes may also be used for unitemporal signature extension applications, as discussed in Section 7.6. Multitemporal signature extension carries with it the requirement for multitemporally registered data. Although proper registration of the data from separate time periods to form single multitemporal data sets adds to the processing time for multitemporal signature extension, the benefits obtained through this procedure should still warrant the additional processing effort.

Another use of multitemporal data is in signature extension from one year to the same time period(s) in a following year. In such cases, when signature extension is attempted from one area to the same area a year later, it is probable that the cropping practices and types of materials present in that scene will not have changed substantially. Thus, the chances for signature extension success are improved. Since extensive training data has been generated for past years, when signature extension was less well developed, much of the large data base required for such year to year applications already exists.

5.4 THE TASSELLED CAP DATA TRANSFORMATION [5]

Feature extraction techniques are also useful in signature extension. Extensive work at ERIM with cluster statistics from numerous Landsat scenes has led to the identification of a basic four dimensional region in the Landsat data space, shaped somewhat like a tasselled cap, which delimits data distributions for typical agricultural scenes. Specific portions of this tasselled cap distribution correlate with meaningful features within these scenes: the base of the cap representing variations in scene brightness or soil type and color, the peak of the cap representing mature green vegetation, and the tassels tracing the process of senescence in the various crops. This visualization of the Landsat data space has been used to devise transformations which change the four standard bands of Landsat data into new data channels more closely associated with meaningful features of agricultural scenes. Although the optimum data transformation of this type would be nonlinear, and would need to be derived separately for each scene, a fixed linear representation of the data transformation (essentially a four dimensional rotation matrix applied to the four Landsat bands) has been derived which is expected to be quite satisfactory for most scenes. (The coefficients for this transformation are specified in Section 7.2.) This transformation is useful in concentrating the most meaningful, and the least useful or most confusing, information

from the original Landsat bands into separate data channels. This enables exclusion of the least useful or most confusing channels from processing and the retention of only the most useful channels, producing a potential for improved signature extension performance and a simultaneous saving in the processing efforts which follow the data transformation. This saving applies especially to multitemporal applications, since processing effort for some of the signature extension procedures increases in an accelerated manner with an increase in the number of data channels.

5.5 PARTITIONING

Another potential improvement in signature extension performance can be derived from developing the wisdom to know when and when not to try to use signature extension techniques. Earlier, the problem of training and recognition scenes with dissimilar major constituents was mentioned. An obvious attempted solution to this problem is to use only training and recognition scenes which are sufficiently similar so that the signature extension algorithms used can handle them. The act of selecting appropriate associations of training and recognition scenes for signature extension is called partitioning. To effect this partitioning procedure, one may first define spatial-temporal domains, based on knowledge of physiographic information (soil class, type, and color, climate, topography, or past history of rainfall, cropping practice, etc.), which have nearly uniform physiographic characteristics. These spatial-temporal domains would be called strata. Subsequently, these domains could be further subdivided, based on knowledge gained from the MSS data itself, into smaller spatial-temporal domains called partitions. If this procedure were done appropriately, the final partitions obtained would determine the training scene to recognition scene associations that would be reasonable to use for signature extension. Note that the strata, defined according to information on long term effects, would tend to be fixed, or "static", for appreciable periods of time, although

they may change slowly as a function of time throughout the crop growing season. The partitions, however, would be more variable, or "dynamic", since in principle they would change with each data acquisition (according to where rain fell recently, or to where atmospheric haze might be especially dense). The partitioning problem at present is highly complex and of course can vary substantially, depending on the signature extension techniques which are to be employed. Research is currently underway to determine to what extent the signature extension algorithms themselves can help to identify the spatial-temporal boundaries of a partition.

6
EVALUATION OF PERFORMANCE

The development of signature extension techniques has been slowed by the requirement for a great amount of data preparation and testing in order to properly judge signature extension performance. The following discussion outlines the basic test and evaluation procedures currently in use at ERIM, and describes some of the results which have been obtained.

6.1 DATA PREPARATION

The first requirement for testing practical applications of a signature extension technique is for ground truth information. This information is needed not only to generate training statistics, but to evaluate classification accuracy in the recognition areas as well. The training and test fields identified and used for the current signature extension effort have come from detailed ground surveys conducted by the U.S. Department of Agriculture. The training or test areas identified have been of two types: (1) Intensive Test Sites (ITS's), which generally cover between 20 and 80 square kilometers and (2) Statistical Reporting Service (SRS) sites, which usually cover between 1/2 and 10 square kilometers. Since the SRS sites are in general too small to provide adequate training information (e.g., the Stafford SRS site in Kansas contains no wheat fields), the SRS ground truth has been used only for testing classification accuracy within recognition areas. Training statistics have been extracted only from ITS's. To date, ground truth information determined by analyst interpreters, looking at MSS data images, has not been available to ERIM, although an effort is currently underway to enlarge the ground truth coverage for the SRS sites through such assistance.

The original field definitions for each ITS or SRS site, when received, are designated by outlines drawn on a clear plastic overlay

which matches an aerial photograph. One then needs to generate equivalent field definitions (specified as polygons) which match the Landsat data for each date of interest for each site. The conversion from x,y coordinates measured on the plastic overlay to L,P (line and point) coordinates which match each Landsat scene is performed using a mapping transformation of the form

$$L = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 x y \qquad (13)$$

$$P = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 y^2 + b_5 x y \qquad (14)$$

The coefficients $a_0,\ldots,a_5$ and $b_0,\ldots,b_5$ are determined by a least squares procedure which minimizes the mean square error of matching transformed x,y control points to corresponding L,P control points measured from a lineprinter map of the Landsat scene. The control points chosen are usually either field corners or road intersections which show up clearly in the MSS image, and which are selected at more or less equal intervals throughout the area covered by the overlay. Typically, for an SRS site between 10 and 20 control points are selected, while for ITS's between 16 and 42 control points are used. The regressions to determine the mapping transformation coefficients are iterated so that control points which appear to have been measured in error can be detected by the regression algorithm and can be deleted from subsequent iterations. A procedure somewhat like the iterated regressions in CROP-A and CROWN is followed, with control points deleted which misregister by more than one line or point after the transformation, and with control points restored which subsequently misregister by less than one line or point. After this deletion and restoration process stabilizes, and if the RMS misregistrations averaged over all control points retained is greater than one half line or one half point, the control point with the largest misregistration in lines or points is deleted, and the procedure is repeated

until stability is obtained once again. For regressions using as many as 42 control points it is sometimes nearly impossible to obtain both a maximum misregistration less than one line or point and an average RMS misregistration less than one half line or one half point using all control points. In such a case one or two deletions, if they appear to be distributed randomly throughout the scene, may be tolerated. For scenes with fewer control points, deleted points are remeasured, and the regressions are rerun until all the control points are acceptable. The mapping transformation is then applied to the x,y coordinate representation of the polygon field definitions to produce the corresponding field definitions in the line and point coordinate system.

The task of locating each 1/2 to 10 square kilometer SRS site and each 20 to 80 square kilometer ITS in each 185 x 185 kilometer Landsat frame, and of then determining appropriate control points for obtaining the polygon mapping transformations, is both arduous and time consuming. Since there is a need for polygon field definitions for each potential multitemporal signature extension application as well as for each unitemporal application, recent efforts toward generating field definitions have concentrated on obtaining transformed polygons which can satisfy both multitemporal and unitemporal needs. This effort saving approach has become possible with the availability of multitemporally registered Landsat data for the ITS's and SRS sites, which generally covers an area between 13 kilometers square (for SRS sites) and 39 kilometers square (for ITS's), making the sites easier to locate as well. However, the data is multitemporally registered only to the nearest pixel, hence one finds that a set of field definitions which optimally match one time period will sometimes differ from some of the field definitions needed to match another time period by almost as much as two pixels. What is needed then is a means to estimate the mapping transformation which produces the best overall set of

field definitions, averaging the misregistrations between the separate time periods in each multitemporal data set.

6.2 THE CLUSTM ALGORITHM

For a time, multitemporal field mapping transformations had been estimated by simply averaging the transformation coefficients derived for each unitemporal time period in the data set. Since a multitemporal classification map is probably one of the best estimators for field boundaries which would be applicable both multitemporally and unitemporally for the time periods in a multitemporal data set, an unsupervised classification algorithm, called CLUSTM, has been developed. Prior to running CLUSTM, a subset of time periods is taken, comprising all times of interest, multitemporally or unitemporally (e.g., October through mid June), and discarding the others (e.g., late June through August, which are after winter wheat has been harvested in Kansas). Next, the linear tasselled cap transformation is applied to the data. This is done to allow a minimum number of channels to be used for the succeeding steps, while retaining as much scenic information as possible. The gradient filtering algorithm is then run on the data set, using all channels except the non-such channel from each time period, to identify probable multitemporal field center pixels. CLUSTM then generates clusters in the normal ERIM manner, using only the brightness and green stuff channels from each time period, and using only gradient filter accepted pixels for input to the clustering. However, CLUSTM classifies all pixels throughout the scene, whether or not they were identified as input by the gradient filtering algorithm. Each pixel data value from the input data tape is then replaced on the output data tape by the vector of mean values for the cluster which was its least Euclidean distance neighbor (using mean and variance statistics, but excluding the covariance statistics). Two additional channels are also added to the output tape, recording the cluster identification number and the

χ^2 value for the cluster which classified each pixel. Specific channels or combinations of channels can then be displayed from the output tape, e.g., as lineprinter graymaps. One then uses the map showing the best definition of features to locate the required control points for the regression procedure.

A single channel map from the CLUSTM algorithm, according to its current use, depicts a brightness or green stuff channel from the selected time period, but with the displayed gray level for each pixel representing the mean value, in the chosen channel, of the gradient filtered cluster which classified that pixel. By using a threshold on the χ^2 output channel to edit the display of the map, one can blank out the mapping of pixels which were distant from the cluster mean values and hence are likely to represent mixtures. Thus the CLUSTM map tends to show blank areas which trace the road network and some of the field boundaries in the scene, and to indicate fairly clearly, with uniform gray levels, the larger fields that are present. Using the CLUSTM map, training or test sites with large fields, and which are surrounded by a recognizable road network, can be located quite satisfactorily. Sites with many small fields, because there are misregistrations of up to plus or minus one pixel between time periods, are often quite difficult to locate by any method. Using field definitions derived with the aid of the CLUSTM procedure, and restricting training or testing to pixels whose centers are inset more than 1.5 lines or points from the field boundaries, one can generally be assured of having a good correlation between the training or test pixels and the available ground truth.

6.3 TEST PROCEDURES

The goal for the present signature extension effort has been to generate reasonably accurate proportion estimates for wheat acreage vs. non-wheat acreage. However, wheat proportion estimates alone cannot give a clear indication of success or failure for signature

extension procedures, nor are they fully adequate for determining the causes whenever success or failure can be ascertained. Some additional insight can be gained by examining performance matrices, indicating the classification accuracy for different ground truth categories. For instance, sometimes it is found that when a good proportion estimate is obtained, the performance matrix indicates that the apparently good result was only fortuitous, since the classification accuracy for some ground truth categories was poor. Other times the performance matrix may look encouraging, but the proportion estimate may be poor, possibly due to having too few pixels within the portion of the recognition scene for which ground truth was known. In general it is assumed that a reasonable proportion estimate, together with a good performance matrix, is a sufficient indicator of signature extension success, although it may still be an insufficient indicator of the reasons for that success.

When reverse transform labeling is used, some additional insight can be gained by comparing the recognition cluster labels obtained through signature extension from the training scene to the labels that would have been obtained from ground truth in the recognition scene. Many times, however, there is insufficient ground truth within the recognition scene even to perform local training, hence in such cases the comparison of reverse transform labels to locally derived labels cannot be made.

Another technique for interpreting signature extension performance is to compare the cluster distributions for the training and recognition scenes. If labels are assigned to the clusters, the reasons for classification errors or successes can become quite apparent, especially when clusters from one scene, after being transformed to match another scene, are compared to the untransformed clusters from the other scene. This technique is especially useful for interpreting the performance of signature transformation algorithms. The recent practice at ERIM has

been to display the clusters as two dimensional ellipse plots, after first transforming both sets of clusters according to the linear tasselled cap transformation, so that the brightness and green stuff channels can be plotted. This provides as much useful information as possible in a single ellipse plot. The ellipses, centered about the mean value of each cluster, indicate the variance-covariance statistics for the clusters in the two channels plotted. Analyses of these ellipse plots have lead to two conclusions: (1) cluster matching algorithms are rapidly approaching the limit of how accurately one cluster distribution can be transformed to overlay and match another cluster distribution, excluding consideration of the physical significance or reality of the transformation obtained, as is indicated by generally good performance under carefully controlled test conditions (e.g., simulated data, or extension from one scene to the same scene on a consecutive day), and (2) the partitioning problem is more difficult than had at first been expected, as is indicated by the discouragingly few instances in which cluster distributions from different scenes actually appear to be closely similar.

As an aid to the partitioning problem, attention is beginning to be focussed on performance measures which are available within the cluster matching algorithms themselves, which may indicate whether a particular signature extension attempt is reasonable. Since all cluster matching algorithms use some sort of merit figure to select the apparent best set of cluster pairs, this figure of merit (e.g., the RMS mismatch between the paired clusters after one set has been transformed) is one potential indicator of the validity of a particular cluster association. Within algorithms such as CROWN, which assign a merit figure to each possible cluster pair, the nature and distribution of these merit figures may provide some useful feedback. Also, it has been common practice to examine and compare the multiplicative and additive coefficients of the cluster transformation in the separate

spectral bands in order to gain a general impression of the validity of a result. Such information may also be a useful source of feedback. Feedback of this sort could lead to a means for cluster matching algorithms to specify an approximate dynamic partitioning on their own.

6.4 RESULTS

A partially controlled test of signature extension has been run, comparing the performance of two cluster matching algorithms, CROP-A and ROOSTER [6] (Rank Order Optimal Signature Transformation Estimation Routine). The CROP-A algorithm has been described in Section 4.2, and uses a linear ordering constraint and an iterated regression procedure. The ROOSTER algorithm (developed by the Lockheed Electronics Company, Inc.) uses merit figures associated with all possible cluster pairs, derived from the cluster rank orderings in each spectral band, and picks one candidate with 10 pairs, from which the best 5 pairs are selected to perform the final regression. As tested, the ROOSTER algorithm was in its original form, as documented in Reference [6].

Training data for the test was derived from the Ellis, Kansas ITS, using Landsat-1 data for 13 June 1974. Training statistics were extracted by three separate clustering operations, one using the five largest wheat fields, one using the five largest grass or pasture fields, and one using the five largest summer fallow fields. These three major crop categories included all the significant materials within the ITS boundaries. All pixels whose centers fell more than 1.5 pixel diameters within a polygon boundary were clustered. (This inset guarantees that pixels near field boundaries, which may have mixed signals from the neighboring fields due to slight boundary mislocations, along scan oversampling, or other effects, will be avoided.) This clustering procedure produced six clusters from each of the categories, giving a total of 18 training clusters. The clusters were then labeled according to their training field designations. An ellipse plot of these training clusters, plotting channels 2 and 3 (Landsat

Bands 5 and 6), is shown in Figure 8. The clusters plotted are identified by two numbers separated by a hyphen, the first of which is a cluster identification number (1 through 18), and the second of which indicates the percentage of pixels within the training scene (comprising 7770 pixels) classified by each cluster. Clusters 1 through 6 are wheat (indicated with shading), clusters 7 through 12 are grass or pasture, and clusters 13 through 18 are summer fallow. When applied to classify the entire ITS, these clusters achieved 92.5% correct classification of wheat as wheat (tabulated over known fields using a 1.5 pixel inset) and 97.1% correct classification of non-wheat as non-wheat (tabulated over known fields with the same inset). The wheat proportion estimate was 43.0%, compared to 44.2% based on the ground truth information. Note that the 13 June time period is at or near the time of wheat harvest in this area, a time at which the three major crop categories (wheat, grass or pasture, and summer fallow) occupy separate corners of the triangular data distribution depicted in Figure 8.

The recognition scenes for this test were chosen from the central and southwest regions of Kansas. Information from a Lockheed Electronics Company, Inc., Interdepartmental Communication [7] which grouped Kansas ITS's and SRS sites according to similarities in rainfall, physiographic features, major crops planted, crop production, and soil type, texture, and color was used to partition the recognition scenes. This led to one group of three scenes (the Ellis ITS, the Ellis SRS site, and the Barton SRS site for 12 June 1974) which was judged to be in the same stratum as the training scene. Another group of three scenes (the Haskell SRS site, the Grant SRS site, and the Kearny SRS site for 14 June 1974) was judged to be in a separate stratum from the training scene. An additional scene (the Rush SRS site for 12 June 1974) was not described in the Interdepartmental Communication, although it was geographically close to the training

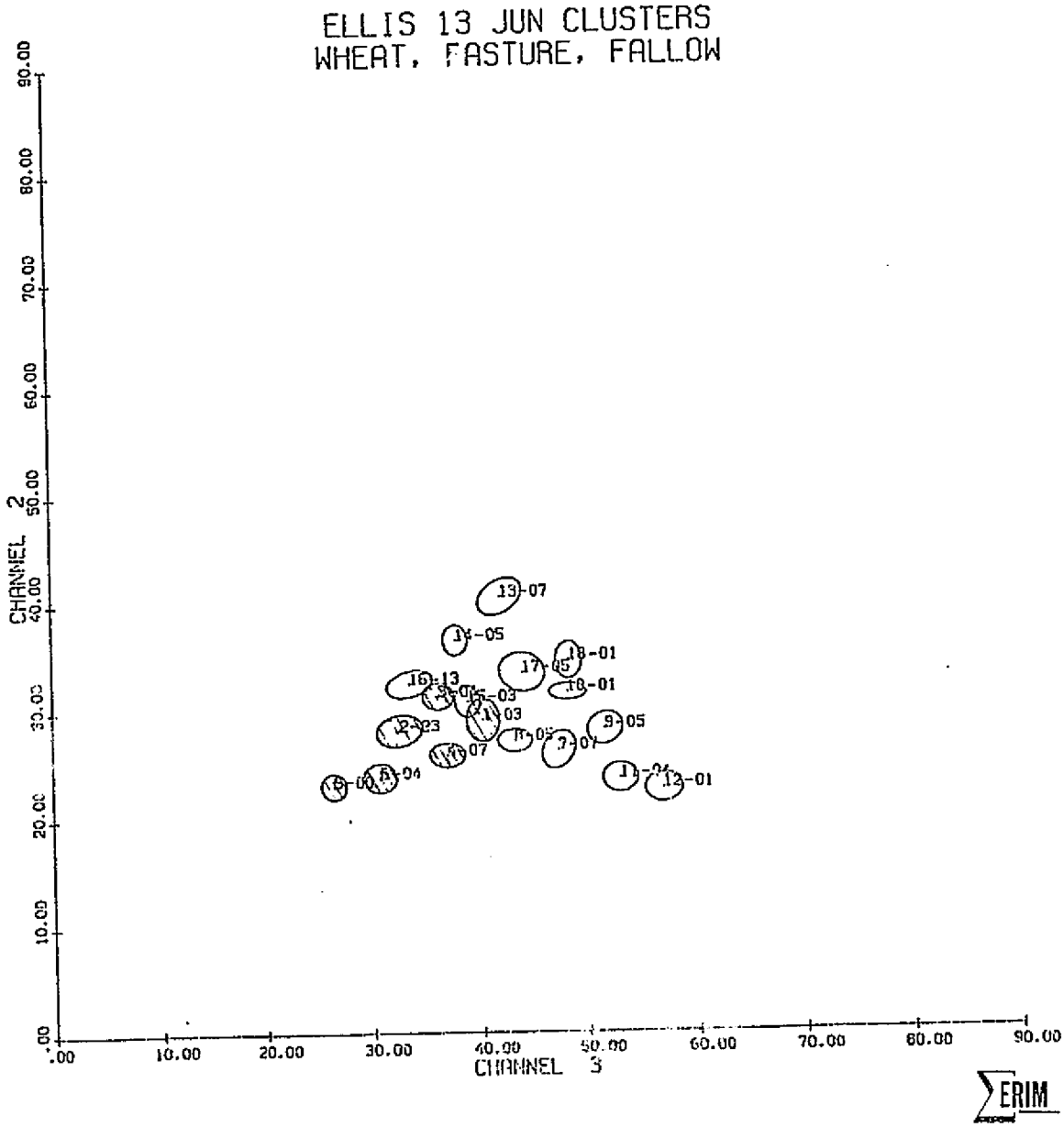

FIGURE 8. ELLIPSE PLOT OF TRAINING CLUSTERS USED TO GENERATE RESULTS FOR TABLE 1

site and to the recognition areas judged to be in the same stratum. Ellipse plots for each of the chosen scenes were then examined, after labeling the wheat and non-wheat clusters with the aid of available ground truth information, to verify that at least some chance for successful signature extensions existed for each of the proposed tests. This partitioning procedure may have been more elaborate than what is presently practical for fully operational signature extension applications.

The results of these tests are tabulated in Table 1. Gradient filtered clustering and reverse transform labeling were not used in generating these results, hence the clustering for each recognition scene was performed using all pixels, representing pure materials or mixtures, within the rectangular data region specified for the clustering. The areas clustered within the recognition scenes were small, varying in size from 1000 to 8000 pixels. The local classification results tabulated in Table 1 were derived by using the unsupervised recognition clusters to classify each recognition scene after assigning optimum labels to the clusters, based on the available ground truth within the scene (with 1.5 pixel insets from field boundaries). Consequently, these results may be somewhat pessimistic compared to local classification using training within known fields, which would generate clusters representing only pure materials. The RMS errors quoted for the proportion estimates in Table 1 are RMS values of the difference between each proportion estimate and the actual proportion. The actual proportions were calculated from the field acreages given in the ground truth information, and hence should not be assumed to be exact values but only close estimates of the actual wheat proportions. The performance matrix terms tabulated in Table 1 were derived from pixels within known fields, using a 1.5 pixel inset from the polygon field boundaries. The RMS errors quoted for the performance matrices are RMS values of the difference between each tabulated matrix value

TABLE 1. SIGNATURE EXTENSION FROM ELLIS ITS 13JUN74
(SUPERVISED TRAINING)

WHEAT PROPORTIONS (%) AND PERFORMANCE MATRICES*

Recognition Scene	Comments	Actual Proportion	Local Result		Untransformed Result		ROOSTER Result		CROP-A Result	
Ellis ITS 12JUN74	Consecutive Day Data Set	44.2	43.9	93.1 99.7	23.9	58.5 100.0	33.1	71.2 97.4	40.8	85. 97.4
Ellis SRS 12JUN74	Within Stratum	40.4	35.2	100.0 100.0	19.2	86.7 100.0	17.6	0.0 80.0	29.5	100.0 100.0
Barton SRS 12JUN74	Within Stratum	60.5	65.5	99.5 83.8	54.4	91.4 91.2	57.1	89.8 79.4	59.7	96.8 82.4
Rush SRS 12JUN74	Stratum Unknown	20.0	27.9	100.0 58.8	21.6	100.0 67.7	38.4	64.7 29.4	45.3	82.4 23.5
Haskell SRS 14JUN74	Across Strata	29.7	30.2	91.1 90.8	6.3	11.5 95.8	42.2	61.7 64.9	15.9	28.1 92.1
Grant SRS 14JUN74	Across Strata	32.1	29.9	71.0 86.3	0.1	0.0 99.8	33.9	19.8 51.8	38.8	67.9 64.9
Kearny SRS 14JUN74	Across Strata	33.9	46.2	90.9 91.2	1.1	0.0 100.0	34.4	60.6 76.5	41.9	75.8 88.2
RMS Error	Overall		6.2	12.3 18.2	22.5	65.3 12.8	12.8	55.6 37.7	12.4	32.3 33.0
RMS Error	Within Stratum		4.2	4.0 9.4	17.3	25.6 5.1	14.8	60.4 16.6	6.6	8.7 10.3

*Only the diagonal terms of the performance matrix are given, specifying percentage of wheat classified as wheat, and percentage of non-wheat classified as non-wheat.

(i.e., the diagonal terms) and 100%, which is equivalent to an RMS value for the off diagonal term of each matrix, calculated within the same row. (A row of the performance matrix represents one ground truth category, while the elements within the row represent the percentage of pixels within that category which were classified as each recognition class, i.e., wheat or non-wheat.)

Note that the first recognition scene tabulated in Table 1 is the same as the training scene, but for the preceding day. Current cluster matching signature extension algorithms tend to do reasonably well with extensions of this sort, which guarantee that the materials present within the two scenes are in fact similar. Analyst interpreters, looking at the pertinent Landsat MSS imagery, have indicated that it appears to have rained in the Ellis, Kansas area between the 12 June and the 13 June data acquisitions, hence this first test attempts to compensate for differences in soil moisture, as well as for any atmospheric, illumination, or viewing angle differences which occurred. The Ellis SRS site and the Rush SRS site are both especially small. The Ellis SRS site covers about 1.9 square kilometers (386 pixels) and has small fields, so that a total of only 20 pixels were available from which to compute the performance matrices for the site (after applying the 1.5 pixel inset). Of these 20 pixels, 15 were wheat and only 5 were non-wheat. The Rush SRS site covers about 1.3 square kilometers (305 pixels) but has larger fields than the Ellis SRS site, hence a total of 51 pixels (only 17 wheat and 34 non-wheat) were available to compute the performance matrices for the site (after applying the inset). Similarly, the Kearny SRS site covers about 3.9 square kilometers (907 pixels), but has many small fields so that only 67 pixels (33 wheat and 34 non-wheat) were available for the performance matrix computations. The other sites, however, each had a few hundred pixels available from which to generate performance matrices. All of the scenes contained at least the three basic materials -- wheat, grass

or pasture, and summer fallow -- with the exception of the Barton SRS site (no summer fallow) and the Haskell SRS site (no grass or pasture). Some of the scenes contained additional materials as well: sorghum (in the Barton SRS site and the Grant SRS site), corn (in the Haskell, Grant, and Kearny SRS sites), alfalfa (in the Grant and Kearny SRS sites), and hay and rye (in the Kearny SRS site). Those sites in the southwest region of Kansas (the Haskell, Grant, and Kearny SRS sites) would also be expected to have drier growing conditions than the other sites.

Note that the RMS errors for signature extension within a stratum for the CROP-A results (Table 1) are very encouraging. The ROOSTER results within this stratrum would have been somewhat encouraging as well, although still not as good as the CROP-A results, had it not been for the result for extension to the Ellis SRS site. As noted above, the Ellis SRS site was one of the smallest available, hence it may not be a very accurate indicator of signature extension performance. Similarly, the Rush SRS site, being small and containing considerably less variety of wheat than the training site, led to difficulties for both algorithms. The results for the recognition scenes which were in the separate stratum are surprisingly good, considering the differences in major crops present in the scenes, which are indicated above.

The signature extension results in Table 1 indicate that within an adequate partitioning scheme current signature extension algorithms (specifically CROP-A) can achieve reasonable and useful results, however there is growing evidence that not all of the necessary partitioning techniques have as yet been recognized and developed. In addition, the partitioning requirement is complicated by another potential stumbling block for signature extension, namely limited data acquisitions caused by clouds covering the scenes needed for processing. With the present Landsat satellites, clouds obscure preselected training or recognition areas roughly 50% of the time. Consequently, Landsat data

for the Ellis, Kansas ITS training area for the 1973 to 1974 growing season is only available for the late October, late May, mid June, and mid July time periods. At other times an alternate training site would have to be available (at least for up to date unitemporal signature extension). Also, since the partitioning scheme may have to change for different portions of the growing season, still other training areas might be needed. Future developments in signature extension will require some remedies for these problems and for the other problems mentioned earlier.

7

A PROTOTYPE SIGNATURE EXTENSION OPERATING SYSTEM

7.1 OVERVIEW

A signature extension operating system, called PROCAMS (Prototype Classification and Mensuration System), has been developed at ERIM to provide a capability for crop recognition and area estimation. PROCAMS is capable of performing local recognition and signature extension, using multitemporal or unitemporal data, and can use multiple training scenes. In this respect PROCAMS is designed to operate within the constraints of the LACIE while simultaneously taking advantage of all the information which may be available to optimize wheat proportion estimation accuracy.

A major feature of the system is the way in which signature extension is accomplished. For reasons of economy 80 percent of the operational LACIE sites are planned to have no field identifications available, therefore, a viable means to extend training information to these sites is important. PROCAMS features a cluster matching algorithm (CROP-A), used with a reverse labeling approach, which can use information from more than one training scene to assist non-local recognition.

The ability to handle multitemporal data is another means incorporated in PROCAMS for bringing additional information to bear on the discrimination and identification process. Crop proportion estimates based on multitemporal data can be expected to show lower variance than similar estimates based on unitemporal data. However, as stated earlier, the system is also capable of operating more conventionally on unitemporal data.

Another feature of the system is the ability to operate either in observation space or in a space transformed by a linear data transformation. The purpose of the transformation is to extract just the spectral information which is relevant for crop discrimination.

In Figure 9, the overall organization of the system is shown. Both local (training) and non-local (recognition) scenes undergo the steps in preprocessing, clustering, and recognition. Non-local processing differs from local processing only in that labeled field center clusters are obtained from the local scene, while the non-local clusters are obtained in an unsupervised and unlabeled manner and obtain the requisite labels from the signature extension procedure. The use, in non-local recognition, of clusters from the recognition scene rather than modified clusters from a training scene is seen as an advantage in that the effect is very similar to local recognition.

7.2 PREPROCESSING STEPS

The preprocessing steps consist of data quality checks to identify, flag, and remove cloud, cloud shadow, and bad line pixels from processing, an external effects correction, a linear (tasselled cap) data transformation, and a means for taking subsets of time periods from multitemporal data.

Clouds

Cloud pixels are identified by comparing a specific linear combination of the four original Landsat bands to a threshold which is scaled according to the cosine of the sun's zenith angle. Specifically, a pixel is identified and flagged as cloud if

$$(1.26201\ P_4 + .11004\ P_5 + .62471\ P_6 + .07028\ P_7) > 145\ \frac{\cos(Z)}{\cos(29^o)} \quad (15)$$

where P_i is the pixel value in the i^{th} Landsat band, and Z is the sun's zenith angle at frame center.

In the language of the tasselled cap transformation (see below) the linear combination represents the "brightness" (soil) direction, minus the "yellow stuff" (senescent vegetation) direction. Thus clouds are "bright", but not "yellow".

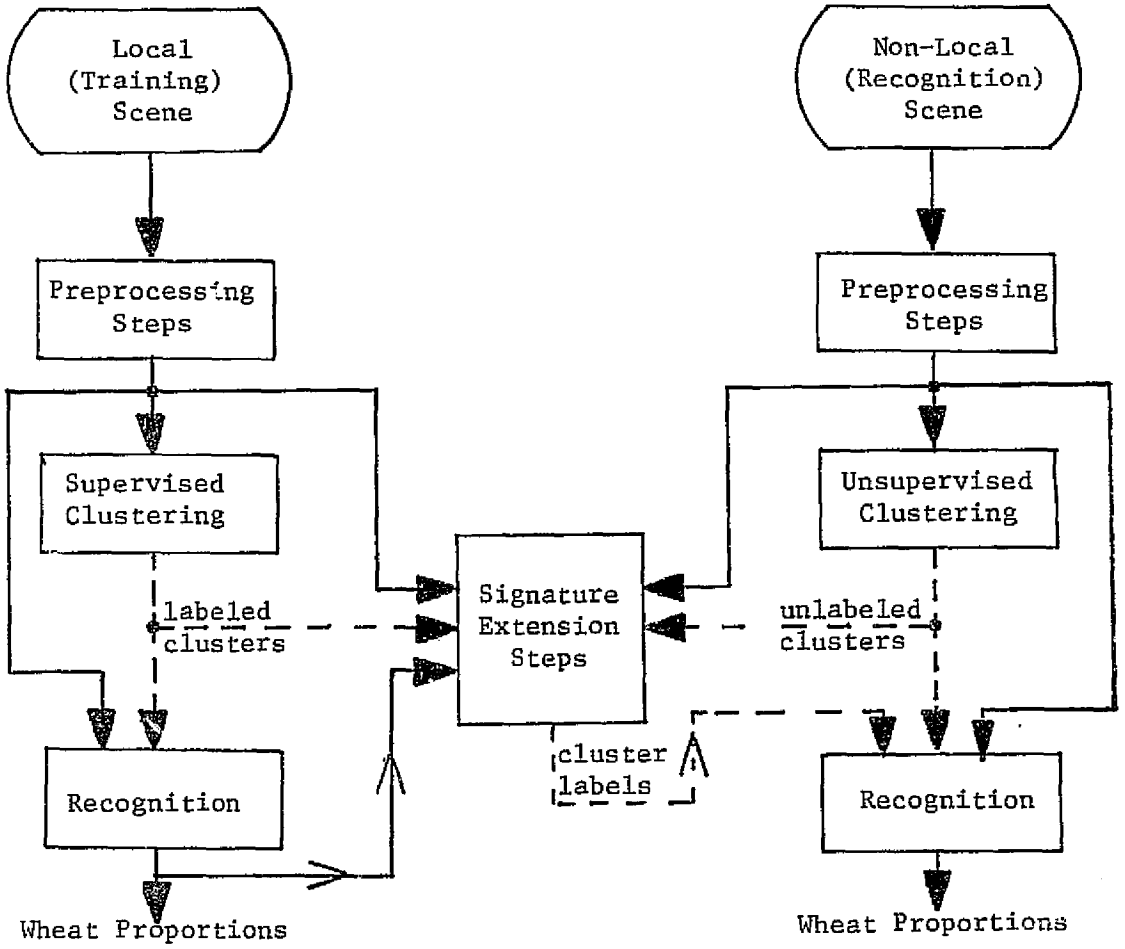

FIGURE 9. OVERVIEW OF THE PROCAMS DATA PROCESSING SYSTEM

Cloud Shadows

A pixel that has a Landsat Band 7 value of 10 or less is identified as being either water or cloud shadow. Since water is a definite class, while a cloud shadow might hide either wheat or non-wheat, one must distinguish between the two. If Landsat Band 5 exceeds Band 6, or if the bands are equal and Band 6 exceeds the value 14, then the pixel is identified and flagged as cloud shadow.

Bad Lines

These are detected by a two pass process. On the first pass, an overall data set signature is computed by sampling the entire data set. On the second pass, pixels are compared to this signature distribution, and a count is made of those pixels rejected by a chi-squared threshold which would accept 90% of all points if the distribution were Gaussian. If over 50% of the pixels in any scan line are rejected, all pixels in the line are flagged as being bad. Otherwise, no pixels are flagged.

All three data quality checks are made independently for each time period, so as not to unnecessarily discard data bad in one time period in processing steps which use only the other time periods.

External Effects Correction Algorithm (EXTEC1)

This algorithm is designed to remove the effects of different sun position, different viewing angles, and atmospheric differences. This is done by forming an affine transformation of the form

$$P' = A P + B$$

where A is a diagonal matrix,
B is a column vector,
P is the original pixel, a column vector, and
P' is the corrected pixel,

so as to normalize each data set to a data set with standardized geometric and atmospheric conditions. The method for determining the transformation is covered in Appendix A. Implementation of this step has not been completed.

Tasselled Cap Linear Transformation

The optional transformation now in use is the fixed linear tasselled cap transformation which has nearly orthogonal axes. Each pixel P is multiplied by a matrix R^T, and the result is added to a constant vector r to ensure positive values, as shown in Eq. (16).

$$P' = R^T P + r \tag{16}$$

where

P = original pixel (4-component column vector)

P' = new pixel

$$r = \begin{bmatrix} 32. \\ 32. \\ 32. \\ 32. \end{bmatrix}$$

$$R^T = \begin{bmatrix} .43258 & .63248 & .58572 & .26414 \\ -.28972 & -.56199 & .59953 & .49070 \\ -.82943 & .52244 & -.03899 & .19386 \\ .22303 & .01170 & -.54250 & .80982 \end{bmatrix}$$

Each row of the matrix represents a linear combination or feature of data, each with an interpretation which will now be described.

- Row 1 identifies the linear combination "brightness". Larger values of this feature represent generally brighter signals from the scene, and in particular, it is the direction of the major axis of soil variation.
- Row 2 identifies "green stuff". Larger values in this new channel are related to greater presence of green vegetation.
- Row 3 identifies "yellow stuff". Values in this feature correspond to the amount of senescent material in the pixels.
- Row 4 is called "non-such". Given the above three nearly orthogonal directions, this is a fourth direction orthogonal to the others. As such, it contains everything else not

accounted for by the other features. Often, banding or other data quality artifacts show up in this channel.

Normally, this step also serves as a data reduction step by using only a specific subset of the four features for the new pixels, so as to leave out information which does not aid crop discrimination or which might be confusing.

For multitemporal data, the transformation is applied independently to the set of four channels from each time period.

Subset of Time Periods

A subset of data channels may be taken, so that only channels from specified time periods are retained. This procedural step is performed for either of the following reasons:

1. To process a desi ed specific set of time periods with the system.
2. To select a matching set of time periods for training and recognition scenes to be used in performing signature extension. A time period in one data set matches a time period in another data set if the time periods represent similar states of crop development.

7.3 CLUSTERING STEPS

As an enhancement to the cluster routine, the boundary pixel excluder GRAD is used to increase the proportion of field center pixels used for clustering, as described in Section 5.1. This step tends to improve the quality of the recognition clusters, as well as their effectiveness for signature extension. Another advantage is that fewer pixels are used in clustering, so that the cost of processing is reduced. The step is optional.

Clustering can be done in either of two modes. For the training scene, clusters are formed within known field boundaries in such a manner as to label each cluster as representing either wheat or non-wheat

(supervised clustering). For the recognition scene, no training field boundaries are defined. The procedure in this case is to cluster all gradient accepted pixels in the recognition scene (unsupervised clustering). The resulting clusters are not labeled. Unsupervised clustering requires only the preprocessed multispectral data. Supervised clustering also requires polygon designations of training fields.

7.4 RECOGNITION STEPS

Recognition is performed on the data using the likelihood ratio rule to determine wheat versus non-wheat proportions. The decision rule is the same as that used in the LACIE CAMS system. Required inputs are the data (after preprocessing), and either labeled clusters (for local recognition), or unlabeled clusters plus labels determined by the signature extension procedure (for non-local recognition).

When the recognition results are tabulated, cloud, cloud shadow, and bad line pixels are not counted. Thresholded points are counted as non-wheat. The threshold is a chi-squared level chosen for a .001 probability of false rejection.

7.5 SIGNATURE EXTENSION STEPS

The heart of the signature extension procedure is CROP-A. This routine takes as input one set of clusters representing the training scene and one set representing the recognition scene. CROP-A then forms a transformation which will be applied to recognition scene clusters so that they will match training scene statistics, as discussed in Section 4.

It is necessary for the training and recognition clusters input to CROP-A to match in the number and order of time periods and in the number of channels. The corresponding time periods should be sufficiently similar with regard to the crop calendar (or have matching biophases), so that wheat is at a similar state of development in both scenes. To accomplish this, one may need to take a subset of channels in one or both scenes such that the subset represents those channels

from the compatible time periods. If such a subset is required for either the training or recognition scene, unsupervised (gradient assisted) clusters for input to CROP-A must be generated for the subset, rather than the respective local or non-local recognition clusters. Furthermore, if a subset of the recognition scene time periods is required, the transformation obtained must be applied, not to clusters from the temporally subsetted scene, but rather to clusters which are an identical temporal subset of statistics from the clusters that will be used for recognition. This is required because the latter clusters are the ones which need to be labeled by signature extension. Figures 10 and 11 will help to clarify the required steps. Once CROP-A has produced recognition clusters transformed to match the training scene, the clusters are used to classify the training scene, using the standard quadratic rule classifier, with one recognition code for each cluster. The reverse labeling program, RLAB, takes the classification results just produced, and the local scene recognition results as inputs. The local recognition results are treated as ground truth. For each cluster class in the classification just performed, this "ground truth" is consulted for every pixel classified into the class. The number of wheat versus the number of non-wheat votes is tabulated. If there are more wheat than non-wheat votes, a cluster is labeled wheat, and vice versa. If the vote is close, or there were not many votes, the cluster is labeled ambiguous (see Section 5.2).

The labels determined for each cluster are then input to the non-local recognition steps.

7.6 SIGNATURE EXTENSION FROM MULTIPLE TRAINING SCENES

The use of multiple training scenes is motivated by the desire to train on more information than may be available from one training scene; for example, when one training scene has a limited number of biophases available or may be missing some of the scene components that are present in the recognition scene.

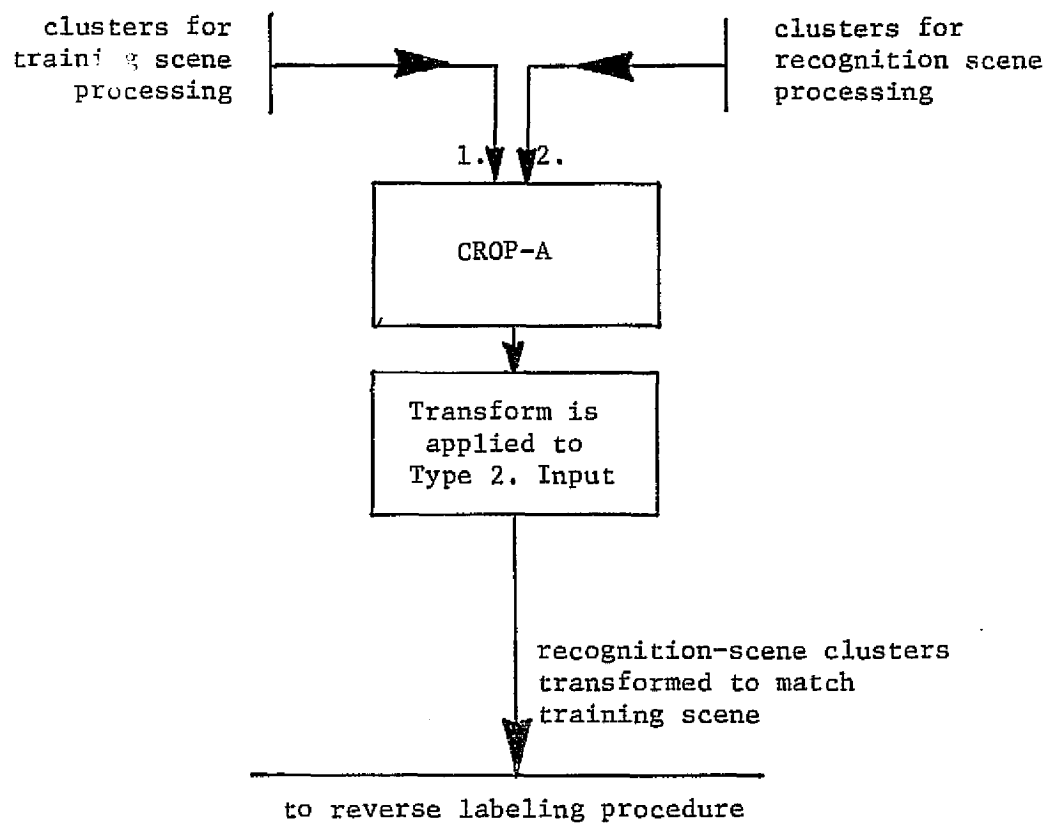

FIGURE 10. CROP-A INPUTS, IF AVAILABLE TRAINING AND RECOGNITION BIOPHASES MATCH COMPLETELY

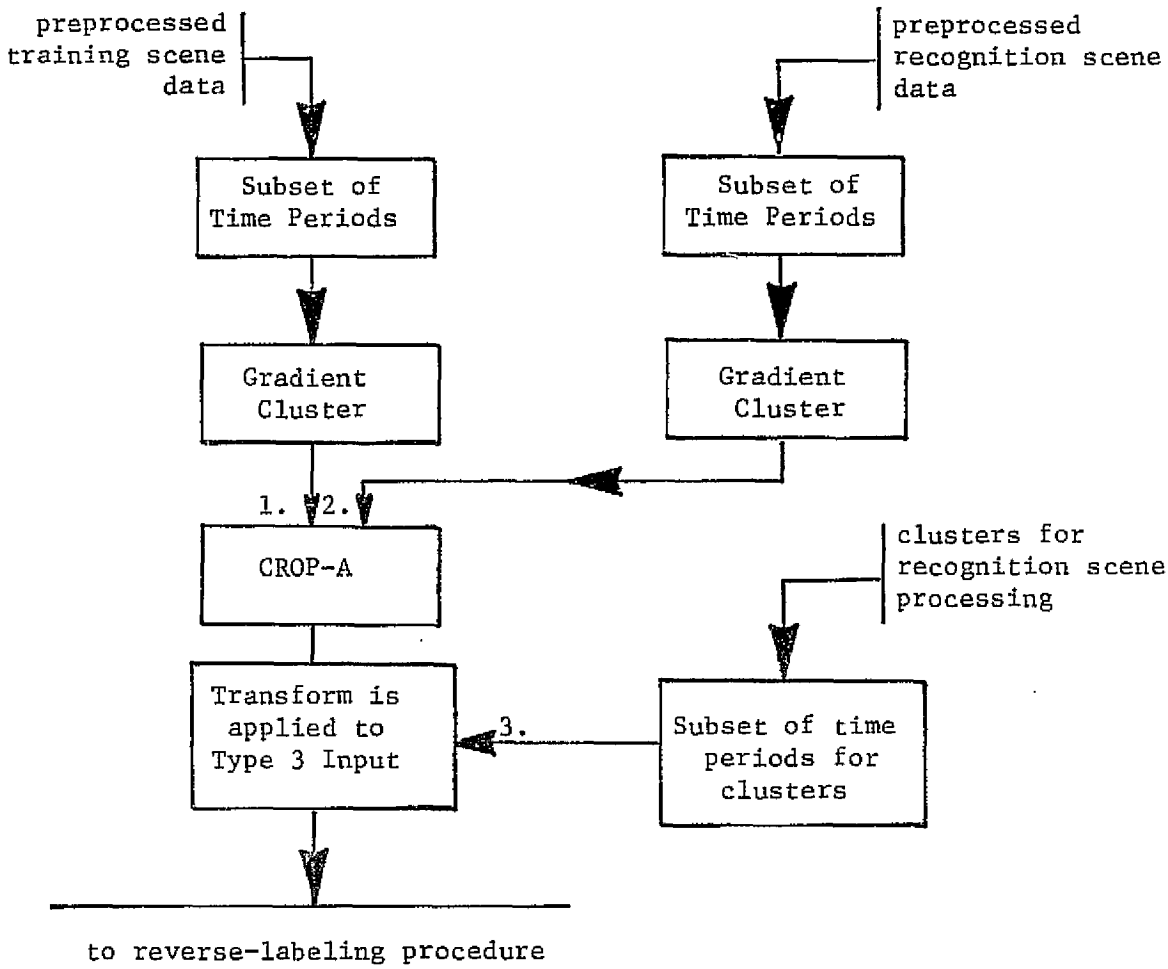

FIGURE 11. CROP-A INPUTS, IF AVAILABLE TRAINING AND RECOGNITION BIOPHASES FAIL TO MATCH COMPLETELY

The system is used as follows: For each (of an arbitrary number) of the training scenes, the signature extension steps described in Section 7.5 are performed for the same recognition scene, through the CROP-A and quadratic rule classification steps. The set of signatures used in each classification is obtained from the CROP-A transformation of the original recognition scene clusters (with subset of time periods, as appropriate), a common source. Thus for every classification the N^{th} class refers to the same N^{th} recognition class, i.e., to the same N^{th} original recognition scene cluster.

This quadratic rule classification (and the local recognition result) for each training scene, are then input to RLAB. RLAB then tabulates for each recognition cluster the number of wheat and non-wheat recognitions from the corresponding local recognition result, summing these votes for all local scenes. The labels for the recognition clusters are determined exactly as for single scene signature extension. The only difference is that there are 2 or more times as many votes added into the final tally.

8
CONCLUSIONS

The preceding discussion has generally followed the historical development of cluster matching techniques for signature extension at ERIM. An attempt has been made to indicate the theoretical boundaries which circumscribe signature extension efforts, and to indicate the step by step progress which has been achieved in cluster matching algorithms and in their use toward realizing the potential for timely, lower cost surveys over large areas, which the theory seems to offer. At this stage of its development, signature extension through the use of cluster matching algorithms, specifically the CROP-A algorithm, appears to be a practical technique for contributing to more economical and timely wheat surveys, using Landsat data, and for other uses as well, provided that the reasonable limits to its use (partitions) can be adequately determined. All aspects of the signature extension problem are of course continually undergoing examination, testing, and development toward the goal of attaining a practical and fully operational implementation of a robust signature extension capability.

Three major stumbling blocks for signature extension have been mentioned:

1. Variations in bidirectional reflectance (including variations due to changes in soil type or color)
2. Differences between major constituents in the training and recognition scenes
3. Limited data acquisitions due to cloud cover during critical time periods.

Further development in signature extension requires improved remedies for these basic problems. Some suggestions for how these remedies might be found are discussed below.

Although the need for detailed partitioning can and should be alleviated through further improvements in signature extension techniques, it is apparent that partitioning techniques themselves must be augmented and improved, particularly with respect to satisfying the needs of existing signature extension algorithms as they develop. In particular, dynamic partitioning procedures based on information available within the signature extension algorithms themselves should be investigated. Preprocessing techniques could help to enlarge the extent of partitions and hence should also be developed and tested.

Additional information should be utilized, when it is available, through techniques such as the external effects correction (Appendix I), through the use of multitemporal data, or by using multiple training sites. Techniques such as the Delta Classifier [2], which show promise for generating training information without access to any ground truth, provided that a suitable multitemporal data set is available, should be actively pursued. Also, year to year signature extension techniques, which reduce partitioning requirements, should be developed. Finally, to aid in testing and evaluating signature extension results, the available data and associated ground truth should be expanded in coverage.

APPENDIX I

EXTERNAL EFFECTS CORRECTION ALGORITHM (EXTEC1)

I.1 INTRODUCTION

While signature extension algorithms which derive coefficients for multiplicative and additive transformations to data or to signatures have generally not, to date, used as input any direct physical measurements (other than information inherent in the recorded MSS signals), techniques can be devised to incorporate direct physical measurements, or knowledge of empirical relationships between MSS data characteristics and physical effects, into the signature extension process as physical constraints on the transformation which is derived. As a result, the signature extension techniques may be made to correlate more closely with the real physical effects responsible for changes in recording conditions between scenes. Two particular physical causes for changes in recording conditions, i.e., atmospheric effects such as haze and geometric effects such as sun angle and viewing angle, are particularly amenable to being incorporated in a signature extension technique as physical constraints. The EXTEC1 algorithm is a first order attempt to transform data sets so that the signals match a standard, possibly hypothetical, scene with a typical set of atmospheric and geometric conditions, in order to largely eliminate data set differences due to those causes, and to do so without the use of atmospheric measurements, which are usually not readily available for large area surveys.

This algorithm serves as one of the preprocessing options called for by PROCAMS (see Section 7). If successful, possibly after future development, the concurrent use of other signature extension techniques may not be required in many cases.

A present limitation to EXTEC1 is that the data set being processed must contain a sufficient number of pixels representing key features of green crop development (in the sense of the tasselled cap). The algorithm may not do well if the scene is so small that those features are

not represented in sufficient detail, or if the data were gathered too early or late in the growing season. As a data normalizer, however, it may still succeed in matching two data sets, if they are both similarly devoid of, say, fields with minimum vegetative ground cover.

The basic idea embodied in the algorithm is to form a transformation which is multiplicative and additive in each of the four Landsat bands and which normalizes data of a scene to simulate what would have been observed under the conditions of a reference scene. This is done by using a physical atmosphere and canopy model [8,9,10] to first find a transformation for the geometric effects of solar zenith angle and nadir viewing angle, and second, to correct for atmospheric state after matching a set of features (namely a specific position in the tasselled cap transformed signal space) from the scene in question to a corresponding set of features, transformed for geometric effects, from the reference scene.

The reference scene we choose is hopefully a scene which has typical values for geometric parameters and for atmospheric conditions. This will improve the chances that the basic model assumptions in the following presentation will hold up well enough to make acceptable corrections.

The EXTEC1 algorithm can be thought of as:

- a haze correction algorithm.
- a general external effects correction technique.
- a scene to scene normalizer (since each scene can be normalized to the same standard).
- a signature extension technique which incorporates supplementary physical information.
- a step toward universal training (if successful in handling most data sets of interest).
- a physical model accepting some empirical inputs.

I.2 GENERAL PROCEDURE

The geometric parameters for the EXTEC1 algorithm can be calculated as accurately as required, while the atmospheric variables must be estimated indirectly from the structure of the data itself. Therefore the estimation procedure is accomplished in several discrete steps:

1. An imaginary scene is defined which is at an intermediate stage, such that the scene has the same standard atmospheric properties as the reference scene, but has the same geometric parameters (e.g., nadir viewing angle) as the recognition scene (see Figure I-1). Specific diagnostic features in the reference scene are transformed to the values they would have if observed under the same geometric conditions that occur in the intermediate scene.

2. These transformed diagnostic features are then compared to the measured features in the recognition data set, and an estimate is made of the deviation of the atmospheric state in the recognition scene from the atmospheric state in the intermediate scene (reference atmospheric state).

3. An atmospheric effects correction is determined which would transform data from the intermediate scene to match the conditions of the recognition scene.

4. Finally, the geometric effects correction (Step 1) is combined with the atmospheric effects correction to determine the required transformation of the reference scene to match the conditions of the recognition scene.

I.3 DEVELOPMENT OF MODEL

In developing the model, we consider the signal X_{ti} from a hypothetical pixel in the recognition data set. Let the signal corresponding to that pixel that would have been measured under the conditions of the intermediate scene (standard haze, but recognition segment geometry)

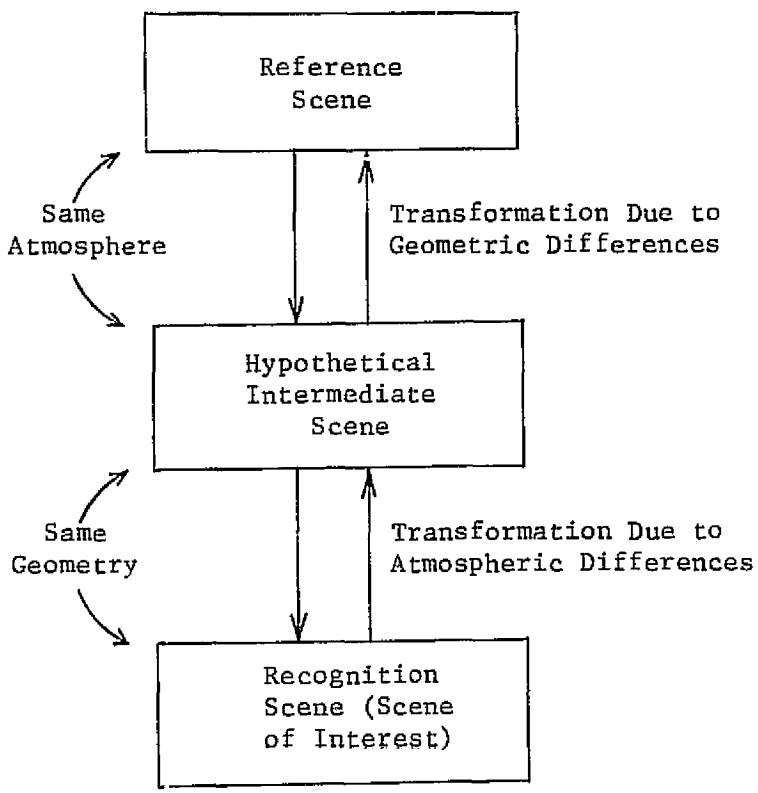

FIGURE I-1. GENERAL ORGANIZATION OF EXTECI TRANSFORMATIONS

be called X_m. Also, let the signal corresponding to that same pixel, under the reference scene conditions, be denoted X_o. Then X_n can be expressed in terms of X_o as follows:

$$\begin{aligned} X_n &= A\, X_m + B \\ &= A\,(A'\, X_o + B') + B \\ &= (A\, A')\, X_o + (A\, B' + B) \\ &= A_n\, X_o + B_n \end{aligned} \qquad \text{(I-1)}$$

The purpose of the algorithm is to determine A_n and B_n, and then, for example, to apply them to the recognition scene inversely:

$$X_o = A_n^{-1}\,(X_n - B_n) \qquad \text{(I-2)}$$

so that the recognition scene data is transformed to match the conditions in the reference scene. In these equations, A is a 4x4 diagonal matrix and B is a 4x1 vector. The fact that A is diagonal allows us to interpret, whenever it is convenient, the operation A^{-1} as component by component division.

In Equation (I-1) the terms A' and B' are functions only of the illumination and viewing geometric parameters, and we know these as accurately as we please. We have an atmosphere and canopy model [10] which predicts A and B' as a function of geometric parameters for a standard haze condition, and we use it to calculate the functions A' and B'.

In general A' and B' may be nonlinear functions of the geometric parameters, but for this correction we simply expand about the reference condition and take the first order terms. If θ is the solar zenith angle and if ϕ is the nadir viewing angle,

$$A' = I + (\theta - \theta_o)\,\alpha_1 + (\phi - \phi_o)\,\alpha_2 \qquad \text{(I-3)}$$

$$B' = (\theta - \theta_o)\,\beta_1 + (\phi - \phi_o)\,\beta_2 \qquad \text{(I-4)}$$

where α_1 and α_2 are 4x4 diagonal matrices, β_1 and β_2 are 4x1 vectors, and $(\theta-\theta_o)$ and $(\phi-\phi_o)$ are scalars.

We now need to calculate or estimate A and B in order to use Equation (I-1). These are functions only of the haze and other atmospheric conditions for the (fixed) recognition scene geometry.

Let h be the vector describing the atmospheric condition, such that it is a vector containing k parameters. As a first trial, there will be only one parameter (k = 1), namely the mass of haze material. Further, let $\gamma = h - h_o$ be the deviation from the reference atmospheric condition h_o.

Since we have no direct observation of γ available, we estimate it by observing certain features extracted from the structure of the Landsat data. The features we can observe are the components of the fixed linear transformed feature vector of a certain special point in or near the line of soils [11].

We will return to the topic of extracting and using these features later. First we develop the model for A and B in terms of γ. We expand A and B about h_o and retain the first order terms of $h - h_o = \gamma$.

$$A = I + a\,\gamma \qquad \text{(I-5)}$$

$$B = b\,\gamma \qquad \text{(I-6)}$$

Here, I is the 4x4 identity matrix, γ is a column vector of length k, and the quantity a is a known 4x4xk array. The multiplication means that term i,j of the product $a\gamma$ is

$$\sum_{\ell=1}^{k} a_{ij\ell}\,\gamma_\ell \,. \qquad \text{(I-7)}$$

The 4x4 product matrix must be diagonal. Also, b is a 4xk array so that the product $b\gamma$ is a vector of length 4.

The functions a and b depend on the recognition scene geometry, and are expanded around (θ_o, ϕ_o), keeping only first order terms, as follows:

$$a = a_o + (\theta - \theta_o)\, a_1 + (\phi - \phi_o)\, a_2 \tag{I-8}$$

$$b = b_o + (\theta - \theta_o)\, b_1 + (\phi - \phi_o)\, b_2 \tag{I-9}$$

All the a's have dimension 4x4xk, and each layer (specific value of the third subscript) must be diagonal.

We return now to the extraction of data features. The features we wish to use are the tasselled cap fixed linear transform (i.e., brightness, green stuff, yellow stuff, non-such) representation of some special point in or near the line of soils. Let us call this special point y_n, y_m, or y_o, depending upon the condition in which we observe it, i.e., recognition scene, intermediate scene, or reference scene. The transformed feature vector for the recognition scene is,

$$v_n = R^T y_n + r \tag{I-10}$$

where R^T is an orthonormal transformation such that the components of v_n are the tasselled cap components referred to above, and r is an offset vector used to keep all data values positive (see Section 7.2).

The vector v_n in the tasselled cap transformed space is what we actually measure. For the first trial of the EXTEC1 algorithm, we use

$$v_n = \begin{bmatrix} \text{the average of "brightness"} \\ \text{the } q^{\text{th}} \text{ percentile of "green stuff"} \\ \text{the } q^{\text{th}} \text{ percentile of "yellow stuff"} \\ \text{the median of "non-such"} \end{bmatrix} \tag{I-11}$$

where q is a low percentage such as 5%.

We will need the inverse transformation $y_n = R(v_n - r)$ for the recognition scene in order to have the measured point in the required signal space. (Since R^T is orthonormal, $(R^T)^{-1} = R$.)

Equation (I-1) can be rewritten in terms of the special diagnostic feature, y, as

$$y_n = A\, y_m + B \tag{I-12}$$

Using Equations (I-5) and (I-6) we can write

$$y_n = y_m + a\, \gamma\, y_m + b\, \gamma \tag{I-13}$$

The middle term on the right represents a 4-component vector resulting from the following double summation (with the m on y_m removed for clarity).

$$a\, \gamma\, y = \sum_{j=1}^{4} \left(\sum_{\ell=1}^{k} a_{ij\ell}\, \gamma_\ell \right) y_j \tag{I-14}$$

Because the summations can be interchanged, this can be written

$$a\, \gamma\, y = \sum_{\ell=1}^{k} \left(\sum_{j=1}^{4} a_{ij\ell}\, y_j \right) \gamma_\ell = a\, y\, \gamma \tag{I-15}$$

The quantity ay represents an array multiplication summed over the second subscript of a. Therefore, Equation (I-13) can be written

$$y_n = y_m + (a\, y_m + b)\, \gamma = y_m + G\, \gamma \tag{I-16}$$

where G is the 4xk array $(a\, y_m + b)$.

Equation (I-16) is the final form of the physical model describing the effects of atmospheric conditions on the feature vector y. In order to solve for γ we need to form $y_n - y_m = G\gamma$. But we also know that there will be noise in our observation of y_n. All we obtain is an estimate $\hat{y}_n = y_n + \text{noise}$. Hence we will be working with the

reduced observation,

$$z = \hat{y}_n - y_m = y_n + \text{noise} - y_m$$

$$= G\, y + \text{noise} \tag{I-17}$$

The vector $\hat{y}_n$ must be estimated. In the recognition scene we measure

$$\hat{v}_n = v_n + \varepsilon \tag{I-18}$$

and calculate,

$$\hat{y}_n = R\,(\hat{v}_n - r)$$

$$= R\, v_n + R\, \varepsilon - R\, r \tag{I-19}$$

where ε is the noise of measurement. In order to obtain y_m we use

$$y_m = A'\, y_o + B' \tag{I-20}$$

from Equation (I-1), with A' and B' determined from Equations (I-3) and (I-4). In order to obtain y_o we measure v_o and calculate

$$y_o = R\,(v_o - r) \tag{I-21}$$

The vector v_o is measured and y_o is calculated once and for all when we choose the reference scene.

We now form a reduced observation vector z

$$z = \hat{y}_n - y_m \tag{I-22}$$

Using Equation (I-19) we obtain

$$z = R\, v_n + R\, \varepsilon - R\, r - y_m \tag{I-23}$$

From Equation (I-10) we obtain

$$z = R\,(R^T y_n + r) + R\,\varepsilon - R\,r - y_m$$
$$= y_n - y_m + R\,\varepsilon \qquad \text{(I-24)}$$

But from Equation (I-16) we can substitute for y_n and obtain

$$z = G\,\gamma + R\,\varepsilon \qquad \text{(I-25)}$$

Thus, we have obtained an expression for the reduced observation vector z in terms of the underlying atmospheric state γ and the noise of measurement of the observation vector.

We assume ε is distributed as a multivariate normal density with mean and covariance,

$$\bar{\varepsilon} = E(\varepsilon) = 0 \qquad \text{(I-26)}$$

and

$$C(\varepsilon) = E(\varepsilon\varepsilon^T) = \text{diagonal with known values for the diagonal terms.} \qquad \text{(I-27)}$$

Here E means "expectation of".

The maximum likelihood estimate of γ is obtained by maximizing the normal density

$$p(z|\hat{\gamma}) = \frac{1}{K}\, e^{-\frac{1}{2} Q(z,\hat{\gamma})} \qquad \text{(I-28)}$$

with respect to $\hat{\gamma}$. This is equivalent to minimizing the quadratic form

$$Q(z,\hat{\gamma}) = (z - \bar{z})^T\, C^{-1}(z)(z - \bar{z}) \qquad \text{(I-29)}$$

Taking the expectation of Equation (I-25) we have

$$\bar{z} = G\hat{\gamma} \qquad \text{(I-30)}$$

Also from Equation (I-25), the covariance of z is given by

$$C(z) = R\, C(\varepsilon)\, R^T \tag{I-31}$$

Hence we maximize the quadratic form

$$Q(z,\gamma) = (z - G\hat{\gamma})^T\, R\, C^{-1}(\varepsilon)\, R^T\, (z - G\hat{\gamma}) \tag{I-32}$$

Taking the derivative with respect to $\hat{\gamma}^T$,

$$-2\, G^T\, R\, C^{-1}(\varepsilon)\, R^T\, (z - G\hat{\gamma}) = 0 \tag{I-33}$$

(This is a standard maximum likelihood procedure, and is described, e.g., in Reference [13]. Equation (I-33) indicates that the derivative of the scalar $Q(z,\gamma)$ with respect to each component of γ is separately set equal to zero.) Solving for $\hat{\gamma}$,

$$[G^T\, R\, C^{-1}(\varepsilon)\, G]\, \hat{\gamma} = G^T\, R\, C^{-1}(\varepsilon)\, R^T\, z \tag{I-34}$$

$$\hat{\gamma} = [G^T\, R\, C^{-1}(\varepsilon)\, R^T\, G]^{-1}\, (G^T\, R\, C^{-1}(\varepsilon)\, R^T)\, z \tag{I-35}$$

Now, having obtained the estimate for γ, we use Equations (I-5) and (I-6) to calculate A and B. Therefore, we have from Equation (I-1)

$$A_n = A\, A' \tag{I-36}$$

$$B_n = A\, B' + B \tag{I-37}$$

and since these are the quantities we had set out to obtain, we are done.

REFERENCES

1. Henderson, R. G., G. S. Thomas, and R. F. Nalepka, Methods of Extending Signatures and Training Without Ground Information, ERIM 109600-16-F, Environmental Research Institute of Michigan, Ann Arbor, Mich., May 1975.

2. Engvall, J. and J. Tubbs, "A Non-Parametric Approach to Classifying Remotely Sensed Data", Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, West Lafayette, Ind., June 1976.

3. Kriegler, F. J., W. A. Malila, R. F. Nalepka, and W. Richardson, "Preprocessing Transformations and Their Effects on Multispectral Recognition", Proceedings of the Sixth International Symposium on Remote Sensing of Environment, Ann Arbor, Mich., October 1969.

4. Kauth, R. J. and G. S. Thomas, Systems for Analysis of Landsat Agricultural Data: Automatic Computer Assisted Proportion Estimation of Local Areas, ERIM 109600-67-F, Environmental Research Institute of Michigan, Ann Arbor, Mich., May 1976.

5. Kauth, R., "The Tasselled Cap Revisited", JSC Memorandum TF3-75-5-190, Houston, Texas, May 13, 1975 (also see [4]).

6. Finley, D. R., "A New Algorithm to Correct MSS Signatures for Atmospheric Effects", Technical Memorandum, No. 6246, Lockheed Electronics Company, Houston, Texas, June 1975.

7. Villanueva, M. R., "Signature Extension for Kansas SRS and ITS Sites", Interdepartmental Communication, No. 644-516, Lockheed Electronics Company, Houston, Texas, July 1975.

8. Turner, R. E., Atmospheric Effects in Multispectral Remote Sensor Data, ERIM 109600-15-F, Environmental Research Institute of Michigan, Ann Arbor, Mich., May 1975.

9. Suits, Gwynn H., "The Calculation of the Directional Reflectance of a Vegetative Canopy", Remote Sensing of Environment 2, 117-125, 1972.

10. Malila, W. A., R. C. Cicone, and J. M. Gleason, Wheat Signature Modeling and Analysis for Improved Training Statistics, ERIM 109600-66-F, Environmental Research Institute of Michigan, Ann Arbor, Mich., May 1976.

11. Kauth, R., "An Approach to Correction of Landsat Data for Observational Effects", JSC Memorandum, No. TF3-75-5-215, Houston, Texas, May 29, 1975.

12. Potter, J. F. and M. A. Mendlowitz, "Some Results and Methods for Signature Extension", Proceedings of the Tenth International Symposium on Remote Sensing of Environment, Ann Arbor, Mich., October 1975.

13. Luenberger, D. G., Introduction to Linear and Non-Linear Programming, Addison Wesley, Reading, Massachusetts, 1973.