2

NASA CR- 134069
ERIM 31650-155-T

A STUDY OF TECHNIQUES FOR PROCESSING MULTISPECTRAL SCANNER DATA

by

R. B. Crane, W. Richardson, R. H. Hieber, and W. A. Malila
Infrared and Optics Division

ENVIRONMENTAL
RESEARCH INSTITUTE
OF MICHIGAN
FORMERLY WILLOW RUN LABORATORIES,
THE UNIVERSITY OF MICHIGAN

prepared for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Lyndon B. Johnson Space Center
Houston, Texas

(NASA-CR-134069) A STUDY OF TECHNIQUES FOR PROCESSING MULTISPECTRAL SCANNER DATA (Environmental Research Inst. of Michigan) 56 p HC $5.00 CSCL 05B

N74-11981

Unclas
G3/08 22766

NOTICES

Sponsorship. The work reported herein was conducted by the Environmental Research Institute of Michigan for the National Aeronautics and Space Administration, Lyndon B. Johnson Space Center, under Contract NAS 9-9784, Dr. Andrew Potter (TF3), Project Manager. Contracts and grants to the Institute for the support of sponsored research are administered through the Office of Contracts Administration.

Disclaimers. This report was prepared as an account of Government-sponsored work. Neither the United States, nor the National Aeronautics and Space Administration (NASA), nor any person acting on behalf of NASA:

(A) Makes any warranty or representation, expressed or implied with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

(B) Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used above, "person acting on behalf of NASA" includes any employee or contractor of NASA, or employee of such contractor, to the extent that such employee or contractor of NASA or employee of such contractor prepares, disseminates, or provides access to any information pursuant to his employment or contract with NASA, or his employment with such contractor.

Availability Notice. Requests for copies of this report should be referred to:

National Aeronautics and Space Administration
Scientific and Technical Information Facility
P.O. Box 33
College Park, Maryland 20740

Final Disposition. After this document has served its purpose, it may be destroyed. Please do not return it to the Environmental Research Institute of Michigan.

NASA CR-
ERIM 31650-155-T

Technical Report

A STUDY OF TECHNIQUES FOR PROCESSING MULTISPECTRAL SCANNER DATA

by

R. B. Crane, W. Richardson, R. H. Hieber, and W. A. Malila
Infrared and Optics Division

ENVIRONMENTAL RESEARCH INSTITUTE OF MICHIGAN
FORMERLY WILLOW RUN LABORATORIES, THE UNIVERSITY OF MICHIGAN

prepared for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

September 1973

NAS 9-9784, Task B 2.8

Lyndon B. Johnson Space Center
Houston, Texas
Earth Observations Division

/

FOREWORD

This report describes part of a comprehensive and continuing program of research into remote sensing of the environment from aircraft and satellites and the supporting effort of recording and analyzing the data gathered by these vehicles by ground-based researchers. The research is being carried out for NASA's Lyndon B. Johnson Space Center, Houston, Texas, by the Environmental Research Institite of Michigan. The basic objective of this multidisciplinary program is to develop remote sensing as a practical tool to provide the planner and decision maker with extensive information quickly and economically.

Timely information obtained by remote sensing can be important to such people as farmers, city planners, conservationists, and others concerned with problems such as crop yield and disease, urban land studies and development, water pollution, and forest management. The scope of our program includes (1) extending the understanding of basic processes; (2) discovering new applications, developing advanced remote-sensing systems, and improving automatic data processing to extract information in a useful form; and (3) assisting in data collection, processing and analysis and ground truth verification.

The research described herein was performed under NASA Contract NAS 9-9784, Task B 2.10, and covers the period 1 November 1971 through 31 January 1973. Dr. Andrew Potter was Technical Monitor. The program is directed by R. R. Legault, Associate Director of the Institute, and J. D. Erickson, Principal Investigator. The work was done under the management of the Earth Observations Division, Lyndon B. Johnson Space Center. The Institute number for this report is 31650-155-T.

ACKNOWLEDGMENT

The authors would like to acknowledge the administrative direction for this work provided by Dr. Jon Erickson.

ABSTRACT

A linear decision rule to reduce the time required for processing multispectral scanner data is developed. Test results are presented which justify the use of the new rule for digital processing whenever both accuracy and processing time are important. A method of evaluating the performance of the rule is also developed and applied to the problem of choosing a subset of channels. A technique used to find linear combinations of channels is described. The ability to extend signatures throughout a small area of approximately fifty square miles is tested. After preprocessing, signatures derived from the first of seven overlapping data sets are applied to all data sets. The test results show that the average probability of misclassification tends to increase with an increase in the number of data sets over which the signatures are extended.

PRECEDING PAGE BLANK NOT FILMED

CONTENTS

FIGURES

TABLES

SYMBOLS

C	average cost
C_{ji}	cost of choosing Z_i when X_j is the true material
D	vector
F	scalar
I_m	identity matrix with rank m
M	measure or abstract distance as an estimate of average probability of misclassification
m	rank
$P_i(y)$	probability density of y when X_i is true
$P(X_i)$	a priori probability of X_i
$P(X_i, Z_j)$	joint probability of X_i and Z_j
$P(Z_j \mid X_i)$	conditional probability of Z_j when X_i is true
Q_0	common covariance matrix
Q_i, R_i	covariance matrix for i-th material
q	parameter of best linear decision rule
R_i	decision space for Z_i
$U(\theta)$	multiplicative correction factor
$V(\theta)$	additive correction factor
y	datum vector point
X_i	set of possible materials
Z_i	set of possible decisions
δ_{ij}	reflectance
θ_{ij}	unknown angle
λ	diagonal matrix
μ_i	mean vector for i-th material

A STUDY OF TECHNIQUES FOR PROCESSING MULTISPECTRAL SCANNER DATA

1
SUMMARY

This report describes part of a comprehensive and continuing program to investigate remote sensing of the environment from aircraft and satellites. The overall objective of this multidisciplinary program at the Environmental Research Institute of Michigan (previously the Willow Run Laboratories of The University of Michigan) is to develop remote sensing as a practical tool to provide the planner and decision maker with extensive information quickly and economically. The work described in this report covers the problems of improving multispectral discrimination techniques and extending their recognition capability in the face of changing conditions. Four main contributions are made in this report:

(1) A linear decision rule that requires less computational time on a digital computer than the conventional quadratic maximum likelihood decision rule, without an increase in the average probability of misclassification, is described and test results discussed. A second decision rule is found for use with a parallel processor, which decreases the amount of required circuitry.

(2) An accurate and extremely rapid method is shown for finding a subset of channels to use for processing.

(3) The choice of linear combinations of data channels for processing data is discussed. A systematic approach to the problem of making an intelligent choice is developed and possible implementations are presented.

(4) An evaluation is made of a specific preprocessing technique to classify accurately data from a small area with a limited amount of training information.

2
INTRODUCTION

2.1. MOTIVATION AND RATIONALE

Remote sensing by the multispectral scanner has been available for some time. Thus far the principal use of the technique has been to demonstrate the feasibility of performing useful tasks important to some segment of our society. The Corn Blight Watch of 1971 and the multispectral portion of the ERTS program can be considered as tentative steps away from the basic feasibility stage. As more confidence and experience are accumulated, we can expect to make better use of multispectral scanner capabilities.

One of the advantages of the multispectral scanner method of remote sensing is the convenience of performing automatic data processing for recognition of ground cover. We foresee heavier collection and use of data requiring a corresponding increase in data processing. New processing techniques which reduce the expense of processing without sacrificing recognition accuracy are needed because the processing of large amounts of multispectral scanner data tends to be expensive.

Processing on general purpose (or serial) digital computers can also be painfully slow. For example, to process all the data collected in 10 minutes during a 20-mile flight by our multispectral scanner, the Control Data 1604-B digital computer can take 900 hours for recognition alone. In addition to recognition, computer time is required for digitizing, formatting, preprocessing, and post-decision analysis. The usual procedure, when processing time is limited, is to reduce arbitrarily the amount of data analyzed by (1) enlarging the along-track and along-scan resolutions either by use of a subset of points (undersampling) or by smoothing, (2) choosing a subset of spectral channels for processing, and/or (3) processing only a portion of the total ground area scanned. Also, decision classes can often be combined so that a limited number of possible decision classes is used. But even with all of these methods, processing times can be excessive, or performance can be seriously degraded.

Most of the time-saving methods listed above fall into the category of reducing the amount of data processed digitally. Time can also be saved by formulating and implementing a decision rule to determine which material from a given set of materials is represented by each sampled vector datum point. The form of the decision rule determines digital computer running time.

Special-purpose and parallel computers offer some advantages over general-purpose serial computers for multispectral recognition processing. However, they too have costs related to the volume of data processed and the type of decision rule implemented. These costs are more closely linked to hardware complexity than to computation time. Perhaps the best example of a special-purpose recognition computer is the ERIM Spectral Analysis and Recognition Computer

(SPARC) which has been in operation for many years. SPARC employs parallel processing circuitry for analog computation of a quadratic likelihood ratio decision rule, but has hardware limitations in the number of material classes and spectral channels used for processing any given data set. Other examples are the hybrid analog/digital designs that have been proposed by ERIM personnel for several years, and an all-digital system with parallel processing logic that is presently being designed and constructed.

2.2. DERIVATION OF DECISION RULE

A review of the derivation of the commonly used quadratic decision rule helps to explain the function of the new decision rule and the linear approximations thereof. Let the set of possible materials be X_i, $i = 1, \ldots, m$. For an n-dimensional vector datum point y, let the possible decisions be Z_i, $i = 1, \ldots, M$. If we divide the n-dimensional hyperspace containing all possible samples of y into m regions, the decision becomes Z_i for any y in the region R_i. The decision rule is based on the Bayes formulations with a cost function C_{ij}, the cost associated with choosing Z_i when X_j is the true material [1]. The optimum decision rule is that which minimizes the average cost C, where

$$C = \sum_{i,j=1}^{m} C_{ji} P(X_i, Z_j) \tag{1}$$

When the joint probability $P(X_i, Z_j)$ is expanded, the result is:

$$C = \sum_{i,j=1}^{m} C_{ji} P(X_i) P(Z_j | X_i) \tag{2}$$

The conditional probability $P(Z_j | X_i)$ depends upon the region R_j, so that:

$$C = \sum_{i,j=1}^{m} C_{ji} P(X_i) \int_{R_j} P_i(y)\, dy \tag{3}$$

where $P_i(y)$ is the conditional density function of y given X_i. Equation (3) can be rewritten in the following form:

$$C = \sum_{j=1}^{m} \int_{R_j} \left[\sum_{i=1}^{m} P(X_i) C_{ji} P_i(y) \right] dy \tag{4}$$

From Eq. (4), the decision rule becomes: for any y choose Z_j for that j minimizing the bracketed term. The resultant choice of all of the R_j's produces a minimum cost function C. A common practice is to let

$$C_{ji} = 0, \quad i = j$$
$$= 1, \quad i \neq j \tag{5}$$

which weights all misclassifications equally and all correct decisions equally. After some simplification, Eq. (4) becomes:

$$C = 1 - \sum_{i=1}^{m} \int_{R_1} P(X_i)P_i(y)\,dy \tag{6}$$

For this case, the regions R_i contain those points y where Z_i maximizes $P(X_i)P_i(y)$. This is commonly known as a weighted maximum likelihood decision rule. If the a priori probabilities $P(X_i)$ are all equal or are assumed equal, the decision would be based only on the a posteriori probabilities (likelihood functions) $P_i(y)$. For that which follows, the (unweighted) maximum likelihood decision rule will be assumed.

If the data from each material have a normal, or Gaussian distribution, then

$$P_i(y) = \frac{1}{(2\pi)^{n/2}|Q_i|^{1/2}} \exp\left[-\frac{1}{2}(y - \mu_i)^t Q_i^{-1}(y - \mu)\right] \tag{7}$$

where μ_i and Q_i are the mean vector matrix and covariance matrix for the i-th material and n is the number of spectral channels. The decision is the i that maximizes $P_i(y)$ or, alternatively, that minimizes

$$(y - \mu_i)^t(Q_i)^{-1}(y - \mu_i) + \ln|Q_i| \tag{8}$$

This is the decision rule that we have called the quadratic decision rule. The μ_i and Q_i for each i are estimated from portions of the data called training data.

2.3. APPROXIMATIONS INVOLVED

While the derivation of the quadratic decision rule is straightforward, there are inherent assumptions that are not realized in multispectral scanner data. One assumption is that the data from each material class be distributed normally. This assumption has been shown to be erroneous [2] . Another assumption is that the μ_i and Q_i can be estimated accurately from the data; however, when inter-field variations are present in the ground cover, the estimates may not be representative of the data from the entire class. A third assumption is that the likelihood function $P_i(y)$ can be formulated for each y independently. However, with the possible exception of smoothing of the data, we are ignoring the values of neighboring data points. Finally, we have the classical objection to the Bayes formulation. We cannot hope to estimate accurately the costs C_{ij} and the probabilities $P(X_j)$. In spite of the failure of multispectral scanner data to meet these assumptions and conditions, the quadratic decision rule has been used often and successfully.

Section 3 contains a description of linear approximations to the maximum likelihood decision rule and a comparison of measured performances of the rules for recognition of several data sets. Section 4 derives the best linear rule and analyzes the performance of the rule. A rapid algorithm for finding a subset of channels is discussed, and a suggested method of finding a linear combination of channels is presented. In Section 5, recognition results are shown for a small area with the training data chosen from a small portion of the area. Finally, additional details of the preprocessing methods used for Section 4 are presented in the appendices.

3
LINEAR DECISION RULES

3.1. GENERAL DESCRIPTION

A number of possible linear decision rules could be used for classifying multispectral scanner data. We chose to evaluate rules that can be put in the form of a series of pairwise decisions; the linear discriminant method is such a decision [3]. Another common feature of the selected rules is that the same data channels are used for each series of pairwise decisions. We felt that, while choosing different channels for each pairwise decision should be advantageous when classifying the training data, the advantage would tend to be lost when non-training data were classified.

Our choice of linear decision rules was influenced by our previous study in which we found that it was not worthwhile to find likelihood functions better matched than the normal function to the training data [2]. Thus, for all rules, we are assuming the distributions to be normal and describable by mean vectors and covariance matrices measured from training data.

3.2. DESCRIPTION OF RULES TESTED

Five linear decision rules were tested and compared with the quadratic decision rule. One rule, optimistically labeled "best linear," uses the pairwise linear rule that best classifies data from two normal distributions [3]. As we will show, the best linear decision rule proved to be the best of the linear rules that we tested. Its formulation is presented in Section 4.

The maximum likelihood decision rule becomes linear when the assumption is made that all covariance matrices are the same. Therefore, we call it the equal covariance rule, although it is sometimes referred to as the linear discriminant rule. To develop another rule we tested, the modified equal covariance rule, we found the linear function and the additive constant which, when added to the quadratic function formed from the common covariance matrix, best approximated the likelihood function in the mean square sense. To be more precise, if the quadratic decision rule based its decision on the maximum of

$$g_i(y) = (y - \mu_i)^t Q_i^{-1} (y - \mu_i) + \ln |Q_i| \tag{9}$$

and we wished to use

$$h_i(y) = (y - \mu_i)^t Q_0^{-1} (y - \mu_i) + y^t C_i + D_i \tag{10}$$

where the same covariance matrix Q_0 is used for all processes, then we found the C_i and D_i that minimized

$$\int [h_i(y) - g_i(y)]^2 P_i(y)\, dy \qquad (11)$$

where the integral is n-dimensional. The use of $h_i(y)$ involves a linear function of y only, since the quadratic function $y^t Q_0^{-1} y$ is common to all processes and hence plays no part in the decision. This rule we labeled modified equal covariance. It proved to be no more accurate than the unmodified rule.

We tested two additional decision rules. In one, we used the diagonal terms of the covariance matrices, letting the remaining elements be zero. We labeled this rule the zero covariance rule. The nearest neighbor rule, derived by setting the covariance matrices equal to the unit matrix, chooses the distribution with the nearest mean. These last two rules, as well as the equal covariance rules, are convenient for both serial and parallel computers. The best linear rule is used most easily on a serial digital computer.

3.3. NULL SET

For all of these decision rules, we included the ability to reject all materials. This null decision is made when a quadratic function of the datum point exceeds a certain value. The function depends on the mean vector and covariance matrix corresponding to the decision rule in use, and represents how far, in a probabilistic sense, the datum point is from the mean.

3.4. TEST RESULTS

We performed a series of tests designed to evaluate the performance of the various decision rules [5, 6]. Our desire was to devise a quantitative comparison, and thereby to eliminate such qualitative procedures as a visual comparison of recognition maps. We also wanted to make comparisons in a manner directly applicable to the classification problem. We programmed the decision rules for a digital computer, counting the number of correct and incorrect recognitions within fields for which we had corroborative ground observations, and measuring the time taken by the computation. A probability of misclassification was computed for each field by finding the number of incorrect recognitions, and dividing by the sum of the number correct and the number incorrect. An average error rate was then found by establishing the average probability of misclassification. The time measurement included not only that time needed to make the actual decisions and test for possible rejection, but also the time to read the data from tape, construct each datum vector, and write the decision on the output tape in our normal format. Therefore, only relative processing times are presented on the graphs that follow. We varied the number of channels in order to generate the results.

The results of our classification studies are twofold: first, the quadratic and best linear decision rules are compared, after which the various linear decision rules are evaluated. In Fig. 1 we show the results with data from the Imperial Valley. The points indicate the measured results; the lines between adjacent points were drawn for convenience in identifying the trends. For the quadratic decision rule, the points from left to right were obtained with two, four, and six channels of data. For the best linear decision rule, the points correspond to two, four, six, eight, and ten channels of data. The choice of channel subsets to use was determined by our normal channel selection procedure (Section 3).

In Fig. 1, the bottom two curves were obtained from tests of the training data from which had been derived the mean vectors and covariance matrices used in the decision rules. The two decision rules may be compared by noting that the curve for the quadratic decision rule lies above and to the right of the curve for the best linear decision rule, showing that for a fixed processing time the quadratic decision rule had a higher error rate. Alternatively, for a fixed error rate, the quadratic decision rule required more processing time. When the same number of channels were used, the quadratic decision rule required more computer time and produced a slightly lower error rate.

Nontraining or test data provided the top two curves, so they are probably more representative of an actual classification of data for recognition. The most obvious difference between these curves and those taken from the training data is that here the error rates are considerably greater. The quadratic decision rule curve is again above and to the right of the linear decision rule curve. Also, it is no longer true that for the same channels, the quadratic decision rule produces a lower error rate. The same observations hold for all of the data sets that we tested.

In order to determine whether the particular choice of training data determines the test results, we decided to use one half of the fields for training data and the other half for test data. After testing in this manner, we reversed the roles of the data so that the data that had been training fields became test fields. Three of our classes had an odd number of fields, so one field from each of these classes was always used as a test field. The results of these tests can be seen in Figs. 2 and 3. Comparing these figures with Fig. 1, we see that the error rates are reduced for the test fields and increased for the training fields. One might conclude from this that for recognizing unknown data, one should use multiple training fields for each class. We can also see from the curves that the best linear decision is preferred whenever both error rate and processing time are important considerations.

To see whether our conclusions regarding the comparison of the two decision rules were only valid for the one data set, we repeated our tests for three other sets. The results can be seen in Figs. 4, 5, and 6. Again we note the same general conclusions that we found from Fig. 1. For Fig. 4, we took an additional step in processing the data. All data sets but one were collected at high aircraft altitude, enabling us to choose fields in the center of the scan and thus avoiding any scan angle effect (dependence of radiance in the scan angle). On the other hand,

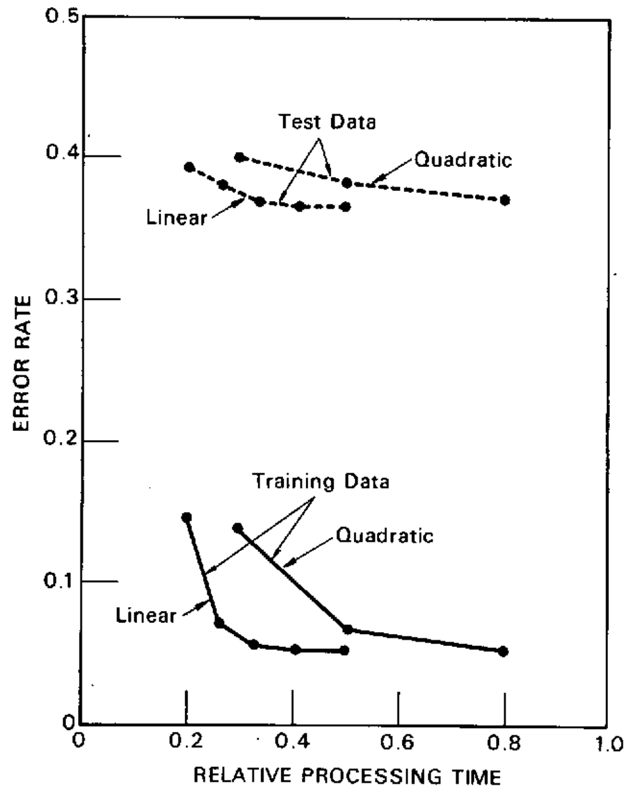

FIGURE 1. LINEAR VERSUS QUADRATIC DECISION RULES FOR SINGLE TRAINING FIELDS. Imperial Valley, 1969, 7 fields, 7 classes for training; 36 fields for testing. Quadratic data are for 2, 4, and 6 channels; linear data are for 2, 4, 6, 8, and 10 channels.

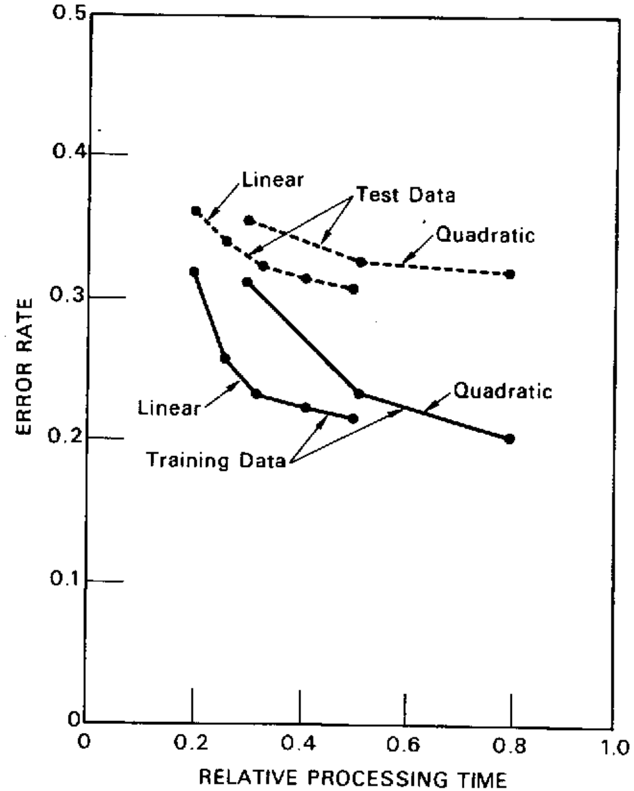

FIGURE 2. LINEAR VERSUS QUADRATIC DECISION RULES FOR COMBINED TRAINING FIELDS NO. 1. Imperial Valley, 1969; 20 fields, 7 classes for training; 23 fields for testing.

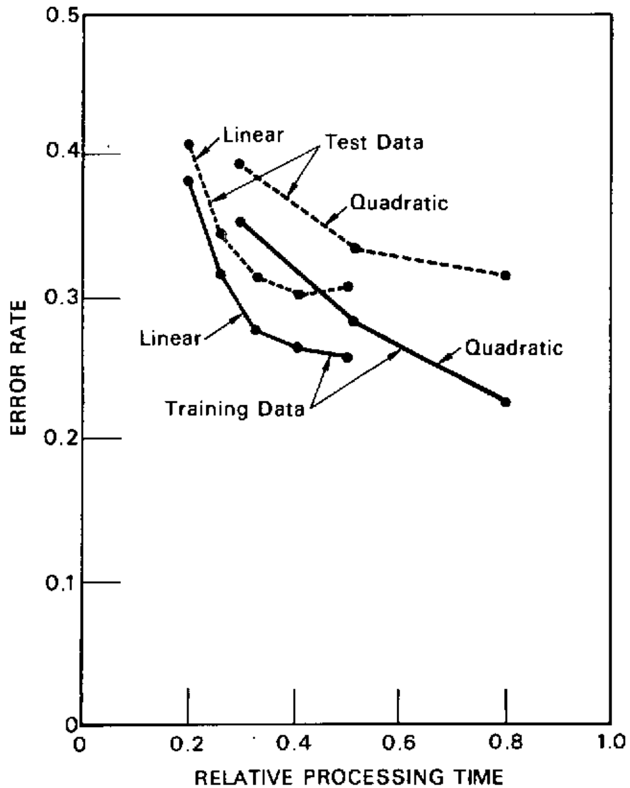

FIGURE 3. LINEAR VERSUS QUADRATIC DECISION RULES FOR COMBINED TRAINING FIELDS NO. 2. Imperial Valley, 1969; 20 fields, 7 classes for training; 23 fields for testing.

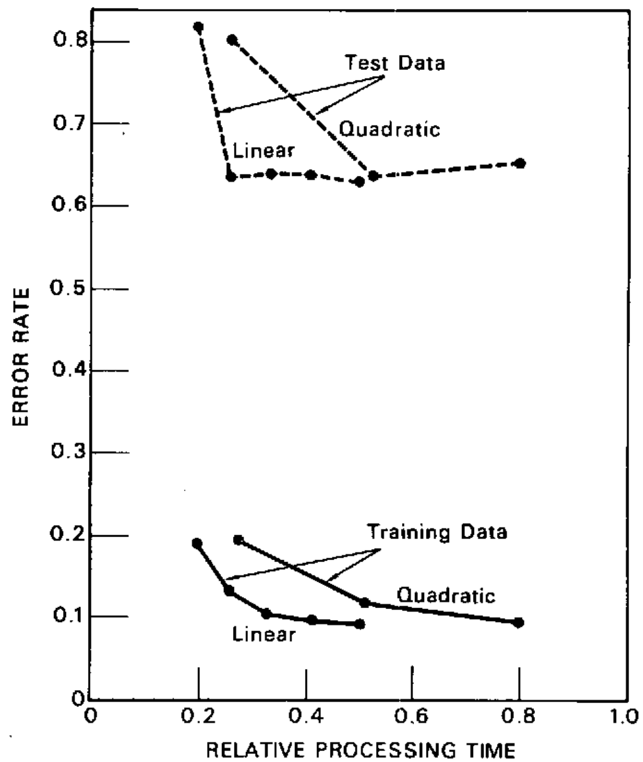

FIGURE 4. LINEAR VERSUS QUADRATIC DECISION RULES FOR HAZY CONDITIONS. Willow Road, 3 September 1969; 14 fields, 7 classes for training; 15 fields for testing.

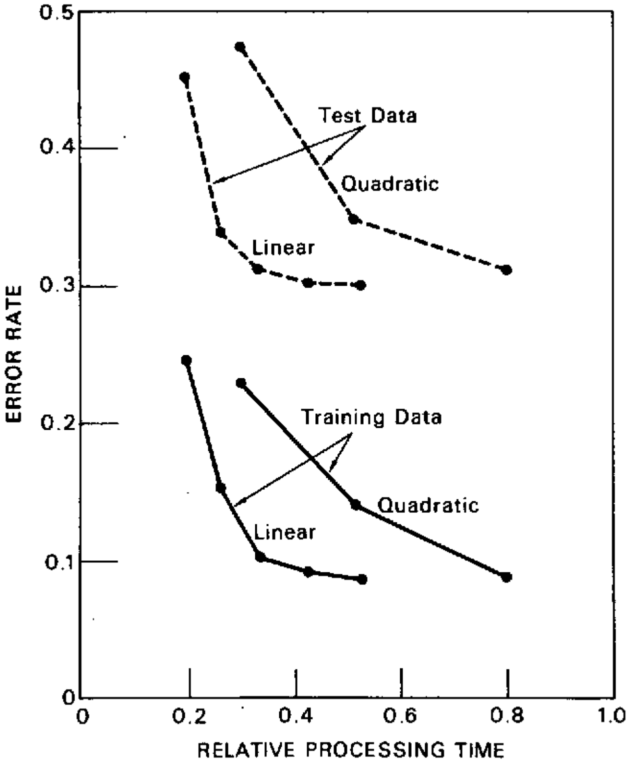

FIGURE 5. LINEAR VERSUS QUADRATIC DECISION RULES FOR CORN BLIGHT WATCH DATA NO. 1. Segment 203, 13 August 1971; 16 fields, 6 classes for training; 20 fields for testing.

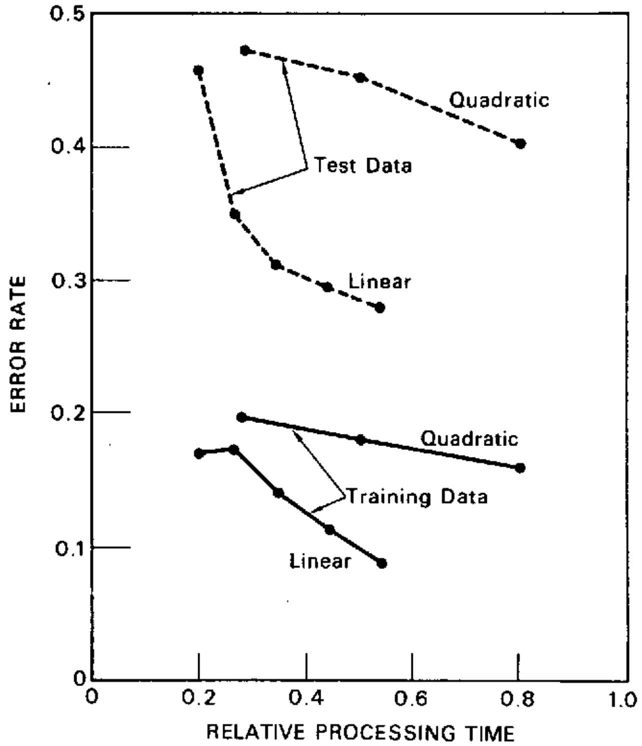

FIGURE 6. LINEAR VERSUS QUADRATIC DECISION RULES FOR CORN BLIGHT WATCH DATA NO. 2. Segment 212, 17 August 1971; 9 fields, 6 classes for training; 26 fields for testing.

the Willow Road data were collected at a low altitude under hazy conditions. Because fields often extended over half the scanning width at low altitude, the angle effect was not possible to eliminate by a suitable choice of fields. We therefore preprocessed the data before classification by making additive and multiplicative corrections [7]. This procedure reduced the scan angle effect so that there was less variation over scan angle than between different fields of the same class. The haze had the effect of reducing contrast, which preprocessing cannot restore. With decreased contrast, the signal-to-noise ratio was also decreased, contributing to the large error rates for the test fields.

To compare the performance of the different linear decision rules, we used the data sets identified in Fig. 2, 3, and 5. Actually only two data sets were used, because the Imperial Valley data was used twice with two different choices of training fields.

Results of the classifications made by the five linear and one quadratic decision rules are shown in Figs. 7 through 12. The curves obtained with the test and training data are shown separately because of the many decision rules. After examining and evaluating the six figures, we reached the following conclusions.

(1) When only two channels of data are used, there is no particular preference to be had among the linear decision rules. The quadratic decision rule required more processing time with no noticeable decrease in error rates.

(2) When four or more channels were used, the "best linear" rule had the lowest error rates of the linear rules. For equivalent processing time, it had lower error rates than did the quadratic rule. For the same number of channels processed, the linear rule did as well as or better than the quadratic rule on the test fields but slightly worse on the training fields.

(3) The nearest neighbor and zero covariance decision rules gave the largest error rates. In fact, there was no clear choice between the two.

(4) The modified equal covariance decision rule was no better than the equal covariance decision rule and, for one of the data sets, was much worse.

(5) Of the rules suitable for parallel computation, the quadratic rule gives the best performance and the equal covariance rule follows. The equal covariance rule has the advantage of requiring far fewer coefficients; however, an assessment of the relative costs of their implementation on parallel processors was not made.

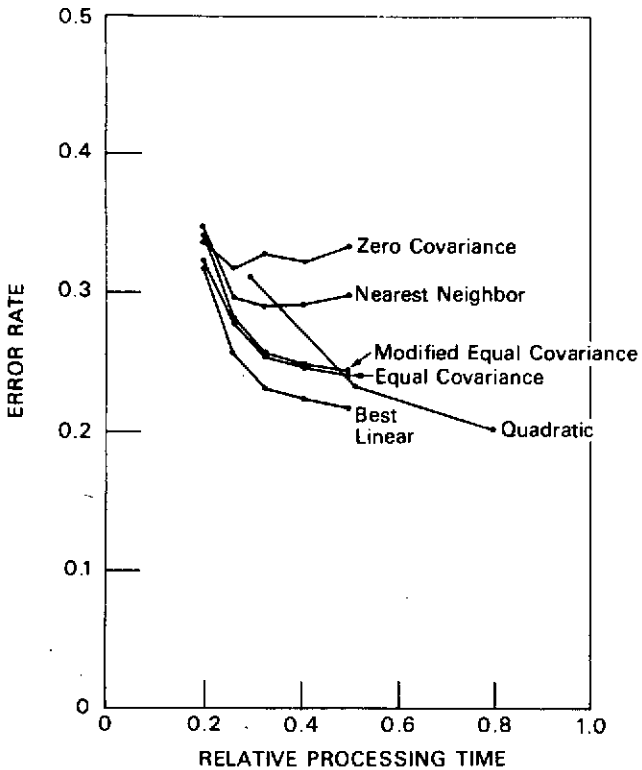

FIGURE 7. COMPARISON OF LINEAR DECISION RULES FOR NO. 1 COMBINED CHOICE OF TRAINING FIELDS, NO. 1 CHOICE OF TRAINING DATA. Imperial Valley, 1969; 20 fields, 7 classes for training; 23 fields for testing.

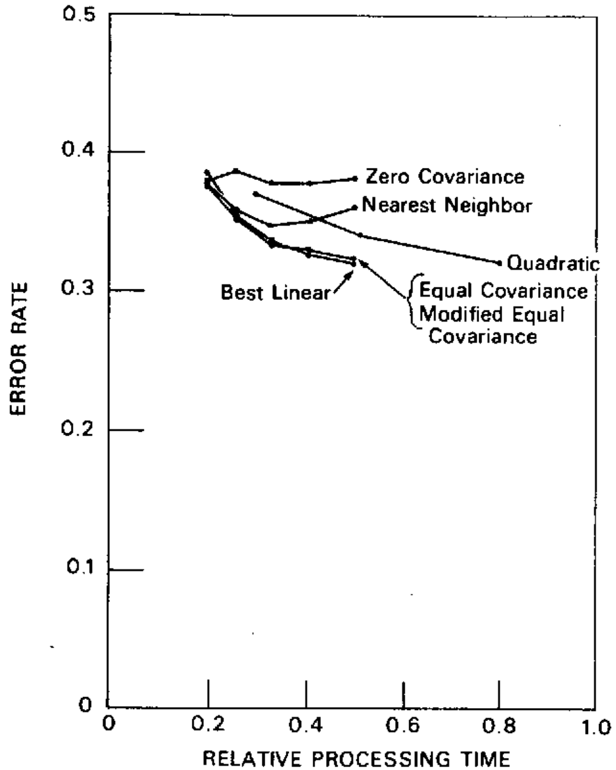

FIGURE 8. COMPARISON OF LINEAR DECISION RULES FOR NO. 1 COMBINED CHOICE OF TRAINING FIELDS, NO. 1 CHOICE OF TEST DATA. Imperial Valley, 1969; 20 fields, 7 classes for training, 23 fields for testing.

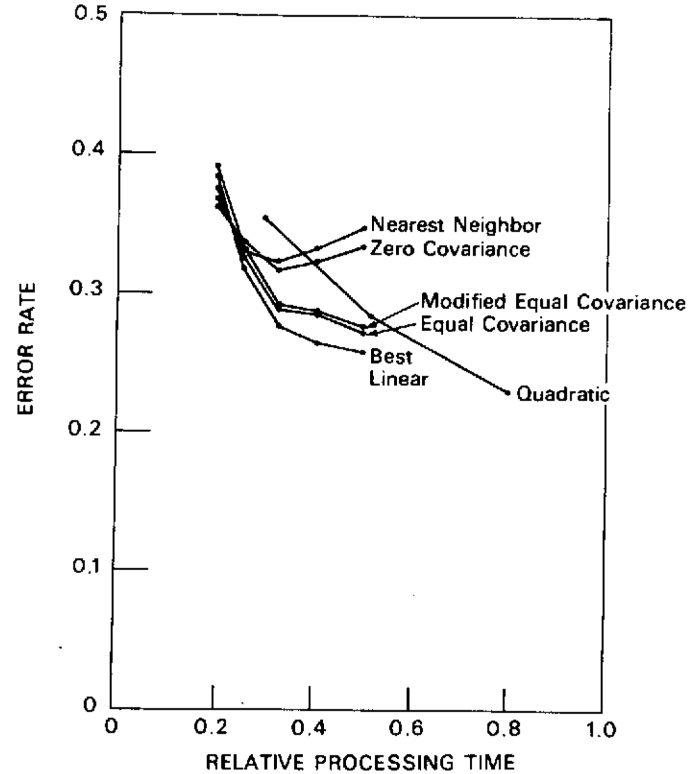

FIGURE 9. COMPARISON OF LINEAR DECISION RULES FOR COMBINED NO. 2 CHOICE OF TRAINING FIELDS, NO. 2 CHOICE OF TRAINING DATA. Imperial Valley, 1969; 20 fields, 7 classes for training, 23 fields for testing.

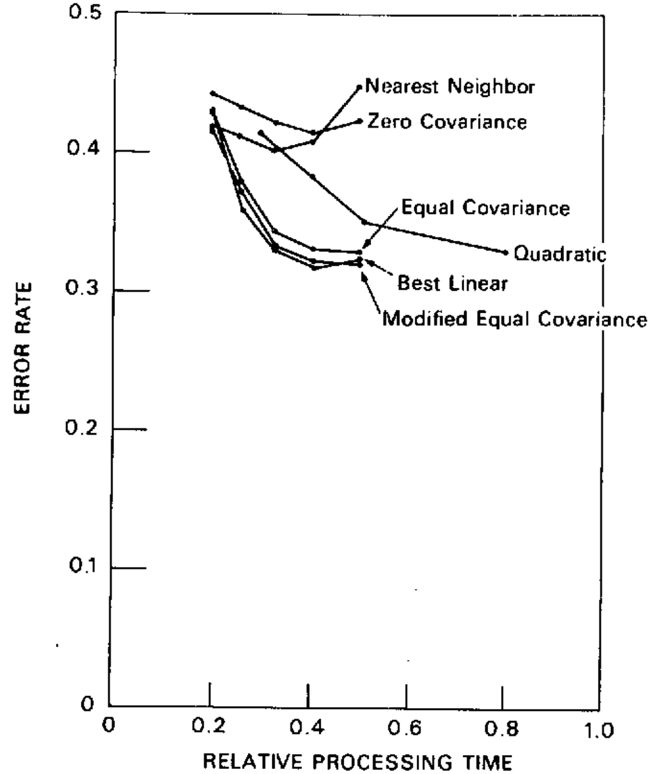

FIGURE 10. COMPARISON OF LINEAR DECISION RULES FOR COMBINED NO. 2 CHOICE OF TRAINING FIELDS, NO. 2 CHOICE OF TEST DATA. Imperial Valley, 1969; 20 fields, 7 classes for training; 23 fields for testing.

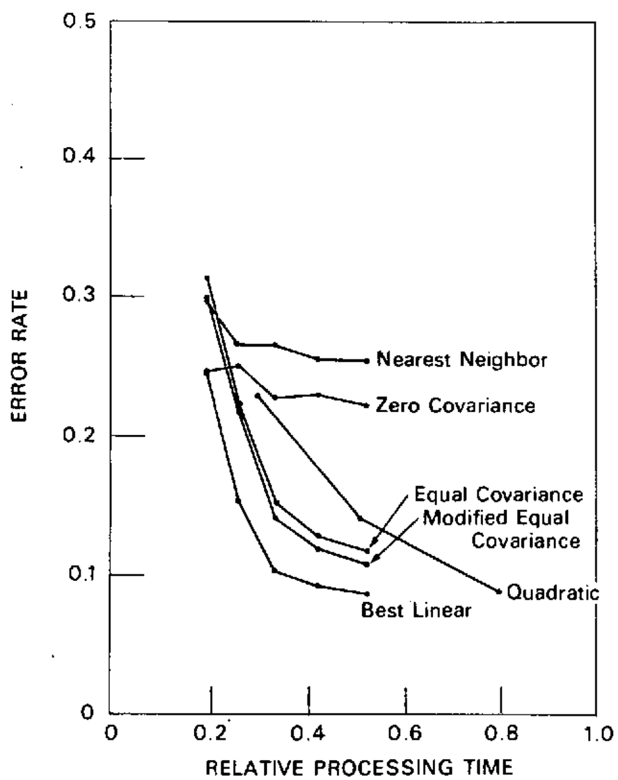

FIGURE 11. COMPARISON OF LINEAR DECISION RULES FOR CORN BLIGHT WATCH, NO. 1 TRAINING DATA. Segment 203, 31 August 1971; 16 fields, 6 classes for training; 20 fields for testing.

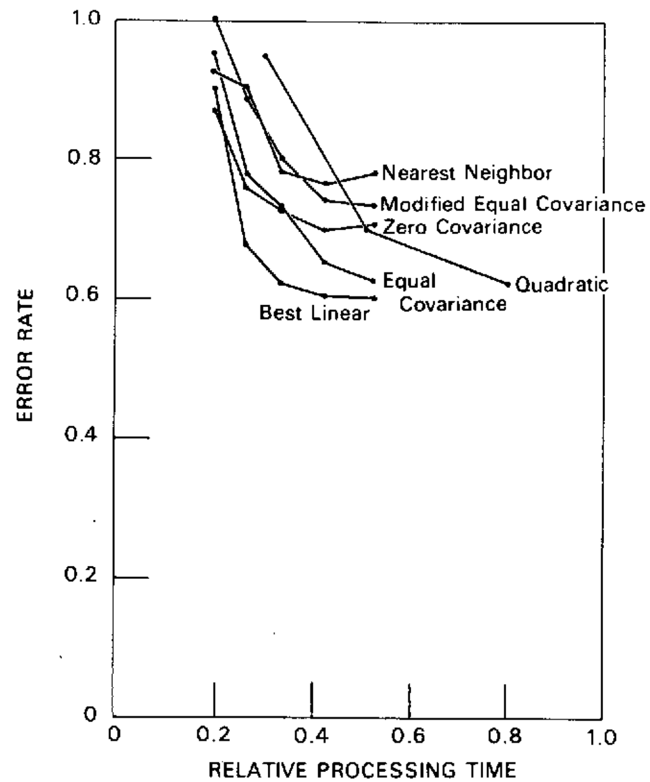

FIGURE 12. COMPARISON OF LINEAR DECISION RULES FOR CORN BLIGHT WATCH, NO. 1 TEST DATA. Segment 203, 31 August 1971; 16 fields, 6 classes for training; 20 fields for testing.

4
PERFORMANCE EVALUATION DERIVED FROM THE BEST LINEAR DECISION RULE

Section 2 demonstrated that the best linear decision rule can be used for recognition of multispectral scanner data. A second function for which the rule can be used conveniently is that of evaluating the expected performance of the recognition process. Performance evaluation can be used as a distance measure for (1) choosing a subset of channels, (2) forming linear combinations of channels, and (3) deciding which decision classes should be combined. The first two of these uses are discussed below, after a derivation of the best linear decision rule.

4.1. DERIVATION OF BEST LINEAR DECISION RULE*

The best linear decision rule is derived from normal distribution functions which are assumed to describe the data. For a pair of functions we wish to test the hypothesis, H_0, that a vector point X is a sample from a random process distribution whose probability function $P_0(X)$, rather than an alternative hypothesis, H_1, with probability function $P_1(X)$. Both $P_0(X)$ and $P_1(X)$ have the Gaussian form

$$P_0(X) = N(\mu_0, R_0) \tag{12}$$

$$P_1(X) = N(\mu_1, R_1) \tag{13}$$

Either Eq. (6) or the Neyman-Pearson lemma can be used to show that the quadratic decision rule will minimize the probability of a Type II error (choosing H_0 when H_1 is true) for any fixed probability of a Type I error (choosing H_1 when H_0 is true). A linear decision rule can be one that decides H_0 when

$$(X - \mu_0)^t D < F \tag{14}$$

and decides H_1 otherwise, with suitable choices for the vector D and scalar F. The advantages of the best linear decision rule accrue from the choice of D and F.

The function $(X - \mu_0)^t D$ is univariate normal, $N(0, D^t R_0 D)$ or $N(\mu^t D, D^t R_1 D)$, depending upon which hypothesis, H_0 or H_1, is true, where $\mu = \mu_1 - \mu_0$. The probability of a Type I error is

$$P\left[(X - \mu_0)^t D > F \middle| H_0\right] = \Phi\left[\frac{F}{\left(D^t R_0 D\right)^{1/2}}\right] \tag{15}$$

where

$$\Phi(X) = \int_X^\infty \left(\frac{1}{\sqrt{2\pi}}\right) e^{-y^2/2} dy \tag{16}$$

*The complete derivation is given in Ref. [3].

The probability of a Type II error is

$$P\left[(X - \mu_0)^t D < F \middle| H_1\right] = \Phi\left[\frac{\mu^t D - F}{\left(D^t R_1 D\right)^{1/2}}\right] \tag{17}$$

where

$$\mu = \mu_1 - \mu_0 \tag{18}$$

The best linear decision rule is derived from Eqs. (16) and (17). The vector D and scalar F are found that maximize $(\mu^t D - F)/(D^t R_1 D)^{1/2}$ when $F/(D^t R_0 D)^{1/2}$ is constant.

The maximization problem can be put into a different form by means of the following substitutions:

(1) $G^t R_0 G = I$

(2) $G^t R_1 G = \lambda^{-1}$

(3) $Z = G^t(X - \mu_0)$

(4) $\alpha = \dfrac{FG^{-1}D}{D^t R_0 D}$

(5) $Z_1 = G^t(\mu_1 - \mu_0)$

The matrix G is chosen so that λ is a diagonal matrix. The decision rule is to choose H_0 if $Z^t\alpha < \alpha^t\alpha$. The probabilities of the Type I and Type II errors are $\Phi[(\alpha^t\alpha)^{1/2}]$ and $\Phi\left[\dfrac{\alpha^t Z_1 - \alpha^t\alpha}{(\alpha^t\lambda^{-1}\alpha)^{1/2}}\right]$. For

$$\alpha = \frac{q}{1 - q}(\lambda^{-1} + KI)^{-1} Z_1 \tag{19}$$

the quantity $\dfrac{\alpha^t Z_1 - \alpha^t\alpha}{(\alpha^t\lambda^{-1}\alpha)^{1/2}}$ is a maximum when $\alpha^t\alpha$ is held fixed. The value of the constant q determines the probabilities of Types I and II errors.

By Eq. (19) and the substitutions above, the best linear decision rule is found to be: decide H_0 if

$$(X - \mu_0)^t R^{-1}\mu < q\mu^t R^{-1} R_0 R^{-1}\mu \tag{20}$$

where

$$R = qR_0 + (1 - q)R_1 \tag{21}$$

and μ is defined by Eq. (18). The probability of the Type I error is

$$P(I) = \Phi\left[q(\mu^t R^{-1} R_0 R^{-1} \mu)^{1/2}\right] \tag{22}$$

$$P(II) = \Phi\left[(1 - q)(\mu^t R^{-1} R_1 R^{-1} \mu)^{1/2}\right] \tag{23}$$

The constant q is chosen to minimize a weighted sum of Eqs. (22) and (23). There is normally a constraint placed on q,

$$0 < q < 1 \tag{24}$$

which causes the decision region for any class to include the mean value of that class.

4.2. COMPARISON WITH QUADRATIC DECISION RULE FOR PAIRS OF SIGNATURES

An analytic method was used to evaluate the usefulness of the best linear decision rule as compared to that of the quadratic decision rule. The data used for recognition were assumed to be Gaussian and the average pairwise probability of misclassification was computed. The classes of data were specified by mean vectors and covariance matrices measured from scanner data. This method provides a conservative comparison, because the quadratic rule is optimum for Gaussian data.

The first analytic comparison was a computation of the difference in performance when the two rules were used on 55 pairs of data. The average percentage of points misclassified was calculated from both decision rules. The difference of the two averages for each of the 55 pairs is shown in Table 1. With the exception of two barley fields, the largest difference is seen to be 2%. Also noted was a significant difference in the times needed to compute the performance of the two rules.

4.3. THE LINEAR DECISION RULE APPLIED TO CHANNEL SELECTION

The ability to analyze the performance of the linear decision rule rapidly is useful for channel selection. The problem of choosing a subset of channels is twofold: how many and which channels to use. In order to decide how many channels to use for the subset, a measure such as average probability of misclassification should be associated with each combination under consideration. As a preliminary step, we made a series of comparisons for pairs of classes. We decided to use 10-channel data and find subsets of 3 and 5 channels with the best linear method and two related and faster (but more approximate) methods [9]. In order to evaluate the difference in performance accruing from the use of two choices of subsets found by

TABLE 1. DIFFERENCE BETWEEN PERCENTAGE MISCLASSIFIED, WITH LINEAR AND QUADRATIC RULES ON PAIRS OF FIELDS

Field No.	29	39	75	91	179	180	190	191	202	205
21	-	-	-	-*	-	2	-	-	-	2
29		-	1	-	-	-*	2	-	1	-*
39			-	-	2*	-	1*	-	-	-
75				-	-	-	-	-	5*	-
91					-	2	-	-	-	-
179						-	-*	-	-	-
180							-	-	-	1*
190								-	-	-
191									-	-
202										-

*Denotes comparison of fields with the same ground cover; e.g., barley.

-Denotes 0%.

the linear and quadratic decision rules, we decided to rank-order each possible subset using average pairwise evaluation of the quadratic decision rule. This evaluation required numerical interrogation. The ranks correspond to optimum rankings when Gaussian data are assumed and the training sets describe the data statistics.

Seven pairs of fields were used for the comparison. For each of the pairs of signatures, Table 2 gives (1) the rank of the subset chosen by the fastest linear method; (2) the difference between the quadratic probability of misclassification (pm) of that subset and the probability of misclassification of the best subset; (3) the probability of misclassification of the best subset; and (4) the difference between the probabilities of misclassification of the poorest and the best subsets. The table shows that the linear method picked the best subset in 9 cases out of 14, did no worse than third for all but two of the cases, and, in the poorest case, chose a subset with a probability of misclassification only negligibly greater than the lowest probability of misclassification. The fourth column of the table shows how badly a poorly chosen subset of channels might perform.

The importance of the rankings is shown in Fig. 13, for a subset of 3, and in Fig. 14, for a subset of 5 channels. The pairwise average probability of misclassification is shown as a function of the ranking of a subset of channels. One can see that more than one subset can be considered usable. Some subsets, ranked near the upper end of the scale, are clearly undesirable. The break in the curve for fields 29-75 occurs because one data channel, 0.8 to 1.0 μm, when combined with any other pair of channels, provided better discrimination than all combinations which did not include this channel.

For more than two classes, the selection of a subset of channels must be made with computational time as an important consideration. We selected a stepwise procedure that successively adds the one channel which gives the lowest average probability of misclassification when used with the channels already selected. When this was tried on the seven pairs of classes, the first- or second-ranked subset was found in each case.

We applied this stepwise procedure for ten channels and nine classes, using first the quadratic, then the linear, evaluation method. For this study, the quadratic method took one hour while the linear method took 70 seconds. This difference resulted in a factor-of-50 time savings in favor of the linear method. The rankings made with the two methods can be seen in Table 3. The orderings are the same except that the linear method interchanged the last two pairs of channels. The use of Channel 2 rather than Channel 8 in a subset of seven channels increases the average probability of misclassification by only .0001, according to the quadratic calculations. Similarly, for a subset of nine channels, the interchange in the ordering of the last two ranked channels increases the average by .00003. Note that the predictions of average probabilities of misclassification are quite comparable.

TABLE 2. PERFORMANCE OF LINEAR CHANNEL SELECTION FOR SEVEN PAIRS OF SIGNATURES

Subset size 3 (120 subsets)

Rank of Subset Chosen by Linear Method	Linear pm - Lowest pm	Lowest pm	Highest pm - Lowest pm
4	0.009	0.110	0.32
1	0.0	0.081	0.27
1	0.0	0.110	0.13
1	0.0	0.023	0.08
1	0.0	0.006	0.13
1	0.0	0.010	0.25
1	0.0	0.035	0.29

Subset size 5 (252 subsets)

Rank of Subset Chosen by Linear Method	Linear pm - Lowest pm	Lowest pm	Highest pm - Lowest pm
9	0.006	0.090	0.24
1	0.000	0.072	0.23
2	0.000	0.082	0.10
1	0.000	0.013	0.04
1	0.000	0.004	0.07
3	0.000	0.006	0.04
2	0.000	0.022	0.15

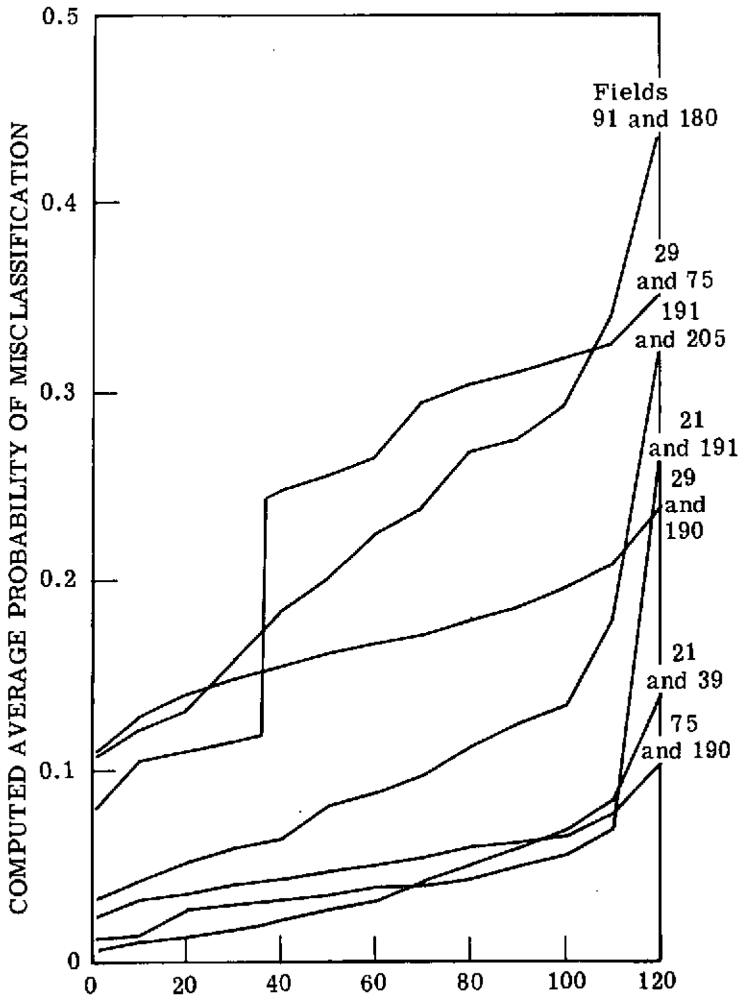

FIGURE 13. IMPORTANCE OF RANK OF 3-CHANNEL SUBSETS FOR PAIRWISE DISCRIMINATION

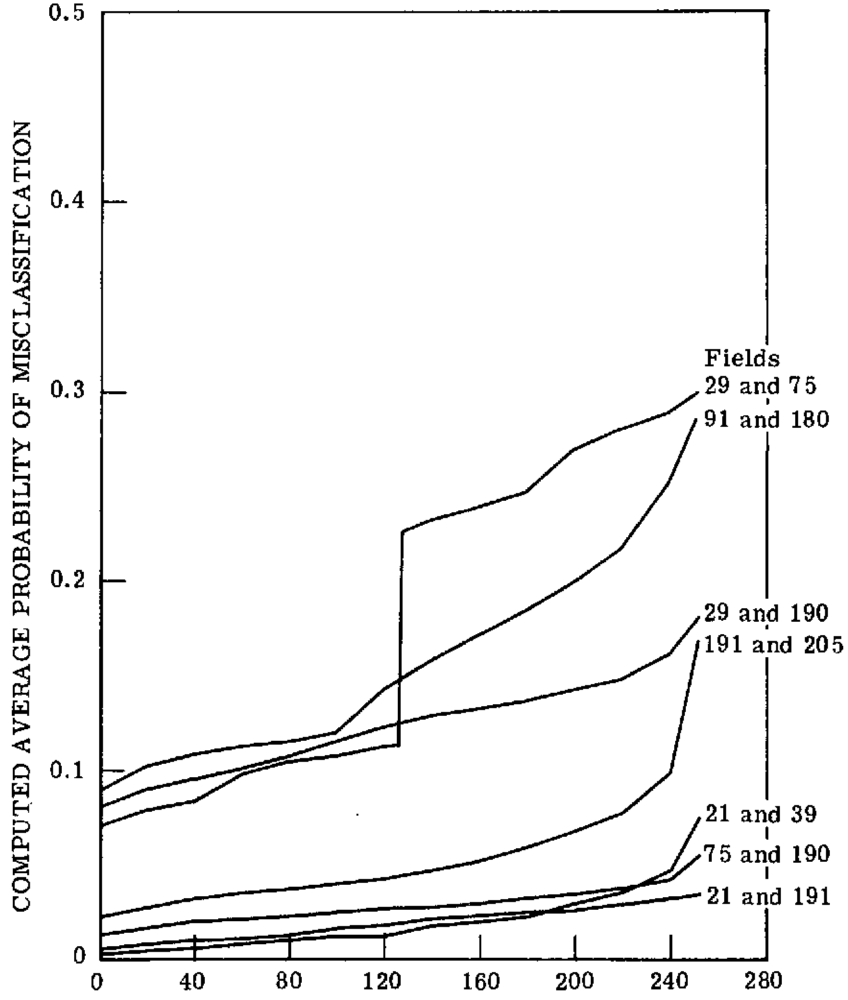

FIGURE 14. IMPORTANCE OF RANK OF 5-CHANNEL SUBSETS FOR PAIRWISE DISCRIMINATION

TABLE 3. COMPARISON OF CHANNEL SELECTION METHODS FOR NINE SIGNATURES

Quadratic Channel Selection (STEPER2)

Order of Channels	Average Probability of Misclassification
4	.119
10	.054
1	.031
9	.025
7	.023
5	.021
8	.019
2	.018
3	.017
6	.016

Linear Channel Selection (STEPLIN)

Order of Channels	Average Probability of Misclassification
4	.122
10	.059
1	.034
9	.028
7	.025
5	.024
2	.023
8	.021
6	.021
3	.020

4.4. LINEAR COMBINATIONS

The distance measure, in which the average probability of misclassification is calculated with a linear decision rule, can be used to find linear combinations of channels, rather than a subset of channels, for processing multispectral data. Linear combinations may be preferred to a subset of channels if, for the same average probability of misclassification, fewer channels are to be used. The choice between the two methods depends upon the ease with which linear combinations can be formed. With analog data available, the linear combinations can be formed at the time the data is digitized. For data in digital form, it may be convenient to form the linear combinations when the data is converted into a format suitable for recognition, or during the preprocessing operation.

The problem of finding a good method of choosing linear combinations is primarily one of finding a workable algorithm in three distinct steps: (1) develop a measure of performance; (2) develop a minimum seeking technique, and (3) find suitable starting points for initiating the minimum seeking technique. In addition, the algorithm should not require an excessive amount of computational time.

The performance measure used is similar to that employed to find a subset of channels and is derived from the linear decision rule (discussed in Section 3) now used routinely in this laboratory. The measure can be expressed as:

$$M = \sum_{i,j} \Phi\left\{1/2\left[(\mu_i - \mu_j)^t\left(\frac{R_i + R_j}{2}\right)^{-1}(\mu_i - \mu_j)\right]^{1/2}\right\} \tag{25}$$

where the summation is for all signatures, the i-th class is distributed normally with mean vector μ_i and covariance matrix R_i, and $\Phi(X)$ is as defined in Eq. (16). An advantage of Eq. (25) is that it can be developed directly from the maximum likelihood decision rule, so the approximations used can be enumerated and evaluated. In fact, Eq. (25) is approximately proportional to the average probability of misclassification that would be measured.

The second step, that of developing a minimum seeking technique, has been completed. A method has been developed to find a local minimum of a function of several variables by starting at a point and following a path of steepest descent by steps of controllable size. Both the local gradient and the local curvature are used to estimate the path of steepest descent.

Finding starting points, the third step, is more difficult. One possible starting point is the linear combination of channels which comprise the optimum subset of the original channels. Another starting point can be derived by finding the principal components of the covariance matrix that result from combining all the classes into one single class.

The measure of performance M (25), is computed with the mean vector and the covariance matrix of the linear combinations of the data. For a recorded point X, a set of linear combinations, Y, can be formed by a matrix A:

$$Y = AX \tag{26}$$

where A has as many columns as there are channels in the recorded data and as many rows as there are linear combinations. The points, Y, are the points used in the decision process, and the statistics of Y, rather than of X, are used in Eq. (25). Thus, the problem becomes one of finding a suitable matrix A, because a choice of A is equivalent to a choice of linear combinations.

The problem of finding a matrix A is one of finding the elements of A. If A is an $m \times n$ matrix, there are mn components to be determined. This number of components can be reduced to $m(n - m)$ by the choice of a suitable canonical form for A. A canonical form is possible because the value of M obtained for any A is not changed if PA is substituted for A, where P is any nonsingular matrix. The canonical form we chose is:

$$A = \left(I_m \left| \begin{matrix} \tan\theta_{11} & \tan\theta_{12} & \cdots \\ \tan\theta_{21} & \cdots & \\ \vdots & & \end{matrix} \right. \right) \tag{27}$$

where I_m is the identity matrix with rank m. There is one inherent theoretical difficulty with this particular choice: it eliminates the linear combinations which have only the value of any one of the last $n - m$ channels. This limitation should not have any practical significance, however, because it is possible to pick the θ_{ij} so that any of the channels will essentially be one of the linear combinations. As an example, if θ_{11} were set equal to $\pi/2 - \epsilon$, with ϵ arbitrarily small, and $\theta_{12}, \theta_{13}, \ldots$ were set equal to zero, then the first linear combination chosen would consist of channel $m + 1$ along with negligible contributions from the other channels.

An apparent difficulty with the given canonical form is that it eliminates any A matrices of the form $\underset{m\times n}{A} = (S_{m\times m}\ T_{m\times(n-m)})$ where S is singular, because $PA = (PS/PT)$ and $PS \neq I$ if S is singular. This apparent difficulty can be circumvented by reordering the data channels.

The canonical form with the θ_{ij} has two advantages. The first is that, in general, a minimum number of unknown scalars must be found. The second advantage is that the minimization process can be accomplished by continuously varying the θ_{ij}. It is not necessary to have large jumps in the values of the unknown scalars, which occur if the $\tan\theta_{ij}$ are considered to be the unknown scalars.

Thus far, the minimization problem has been expressed in terms of the θ_{ij}. The digital decision process to be employed at ERIM is expected to be the pairwise best linear decision process. For pairwise decisions we can consider a good point as the set of θ_{ij} which minimizes the average probability of misclassification between any pair of classes, while a bad point maximizes the average probability. There can be a large number of good and bad points. For example, with 10 classes, there can be as many as 45 of each type of points, because there are 45 combinations of 10 classes taken 2 at a time. Each point is defined modulo π [$\tan \theta_{ij} = \tan(\theta_{ij} + \pi)$] for each θ_{ij}.

Possible starting points can be found by finding linear combinations as far away as possible from all of the bad points or as near as possible to all of the good points. Euclidean distance can be used to compute the starting points, even though we are not dealing with a Euclidean space. If computer time were of no concern, average probability of misclassification would be the proper distance measure. We feel that any errors introduced by the use of Euclidean distance to find a starting point will be negated during the minimization procedure, which uses the better distance measure.

The problem of finding starting points can be visualized by considering Fig. 15. Two linear combinations of three channels of data can be determined by a choice of θ_{11} and θ_{21}. All of the possible pairs of linear combinations, including subsets, can be represented by points in the graph. A subset consisting of the first two channels is located at the point (0, 0) in the center of the graph; another, consisting of the first and third channels, is located at the points (0, $\pi/2$) and (0, $-\pi/2$); the third possible subset, channels two and three, is located at ($\pi/2$, 0) and ($-\pi/2$, 0). Channel combinations comprising possible subsets are shown in the figure as circles.

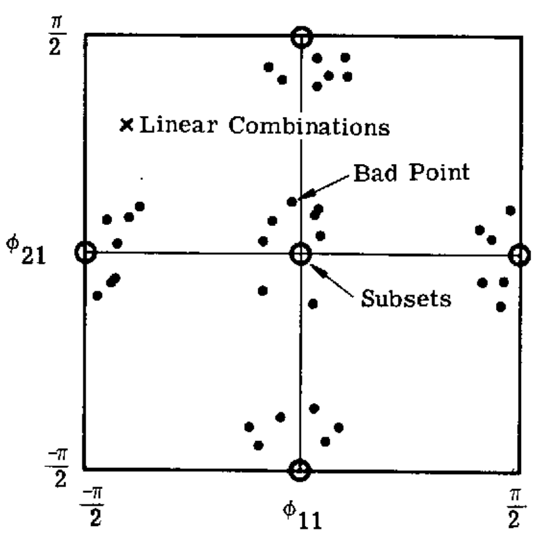

FIGURE 15. EXAMPLE OF BAD POINTS WHEN FINDING TWO LINEAR COMBINATIONS OF THE THREE CHANNELS

If ten classes are of interest, as many as 45 points of misclassification may be located within the boundaries of the graph. One can imagine an example where the 45 points were positioned near the subset locations, in which case the subsets would all be poor choices for starting points and for processing.

Location of starting points can be complicated. At this point in our program, it is not clear whether good points, bad points, or both should be used. One advantage of bad points is that they are fairly easy to find because they depend only on the mean vectors of the class signatures. However, the approximate locations of the good points can be found almost as easily by assuming a common covariance matrix shared by each pair of classes.

The work reported above serves only as a starting point for continued investigations of linear combinations of signal channel data for reducing processing costs without reducing recognition performance. The approach taken here differs from others because it not only considers more than two distributions but also employs the performance-related probability-of-misclassification criterion.

5
PROCESSING OF AREA SURVEY DATA

For the processing of remote sensing (the use of computerized recognition techniques on multispectral scanner data) to be practical in surveying large areas of ground, it is desirable to establish recognition signatures determined from a limited area of known content and then extend them to data collected from other areas. This requires the development of methods to compensate for systematic variations in the data. Prior work has demonstrated the importance of correcting variations associated with the scan angle, caused principally by differing path lengths through the atmosphere and by the angle dependence of the surface reflectance.

Additional sources of variation, which could be ignored for short flights over limited areas, become important when larger areas are covered. For data collected at altitudes up to 10,000 ft, the time required to overfly larger areas implies that changes in the sun's position cause non-negligible variations within the data set. When the area is large enough to require collecting data over a period of many hours within a day or successive days, changes in the atmosphere can also cause variations in the data. Ground differences such as elevation or terrain irregularity also affect data.

High altitude aircraft and satellite data can minimize the importance of time-dependent variations by surveying larger areas quickly, but the atmospheric effects are even more important at these higher altitudes and, over the larger areas that can be covered (10,000 sq mi in each ERTS frame), space-dependent variations cannot be ignored. Correction methods also should compensate for time-varying changes in the instrumentation such as variations in the calibration of the scanner and recording equipment from one day to the next or from one run to the next.

This section reports on results obtained from the application of one particular type of empirical preprocessing transformation to area survey data. The primary objective was to study run-to-run variations and our ability to extend the signatures from one run to the next. A discrete form of the multiplicative-additive (U-V) preprocessing transformation, described in Appendix I and Ref [7], was used to extend signatures extracted from a one-square-mile area of one run to five subsequent partially overlapping parallel runs. Secondarily, the performance of a continuous U-V preprocessing transformation was studied in correcting scan-angle dependent effects prior to the run-to-run extension.

5.1. DATA DESCRIPTION

The data used for this area survey was gathered in Ingham County, Michigan on 6 August 1972. Seven runs were used for this study, with the data gathered shortly after noon at 5000 ft over a rectangular strip one mile wide from east to west and eight miles long from north to south for which ground truth information was available.

(1) Six runs, alternately eastbound and westbound, perpendicular to the ground truth strip; the first (run 3) near the northern end of the strip and each successive run one mile farther south. Each run included a portion of the ground truth area and overlapped approximately 40% of the area covered by adjacent runs. These runs were used for the area survey studies.

(2) A southbound run over the ground truth strip. This run was examined to find how much any variation between the east-west runs was the result of differences in the conditions of the ground covers of the same crop from one field to another.

These data meet the following conditions desired for the area survey study:

(1) Data were gathered at a sufficient aircraft altitude to cover an appreciable area.

(2) Each of the east-west runs included a portion of the ground truth area necessary to establish recognition accuracy for each run.

(3) Each east-west run overlapped approximately 40% of each adjacent run, providing overlapped fields (fields included in both of two adjacent runs) needed for deriving the discrete U-V preprocessing transformation applied to extend the signatures between runs.

This data was gathered on a predominantly clear day with few cirrus clouds. The haze was fairly light, the visibility being estimated as 15+ mi, or 24 km.

The recorded radiances exhibited severe scan-angle dependent variations. The angle effects on all east-west runs showed a lack of symmetry about the nadir. On some east-west runs, the radiance increased to 60% more than at the minimum on the northern side in the shortest wavelength channel (0.46-0.49 μm), with similar but lessening effects for longer wavelengths. Yet, on the southern side, there was seldom a discernible increase over the minimum in any except the shortest wavelength channel, and the southern peak was small compared to the peak on the northern side. The minimum radiance was never near the nadir, being displaced south in this channel and near the southern edge of the scan in the other channels. Angle plots (radiance versus scan angle, averaged longitudinally over the entire length of the run) in all except the shortest wavelength channel were monotonic, increasing toward the northern edge with scanner vignetting at the extreme edges.

Figure 16 shows an example of typical scan-angle dependent variations for the 0.46 to 0.49 μm channel of eastbound run 6 (this is typical, but not the most severe of variations observed).

On the southbound run 2, the scan-angle dependencies were also fairly large but much more symmetric about the nadir.

5.2. PROCESSING

As a preliminary step, the ground truth area of one run was selected to be digitized, clamped, and scan-angle corrected. Signatures extracted for all 12 channels were then used to select the best channels for recognition. Five channels were selected having 0.46 to 0.49-, 0.52 to 0.57-, 0.72 to 0.92-, 1.0 to 1.4-, and 9.3 to 11.7-μm wavebands at 50% response. Only these five channels were digitized and used for the remaining runs. The preliminary run was also used to study the use of the continuous U-V scan-angle correction programs in more detail than possible on all runs.

The northernmost run, 3, was picked for this preliminary step because:

(1) A count of the fields covered by each east-west run and for which ground truth had been gathered, showed that this run had about the best representation of crops for signature extraction of all east or westbound runs; these signatures were to be used for recognition of all the other parallel runs.

(2) It was desirable to use either the first or last of the parallel runs for signature extraction to maximize the separation between the training fields and some of the test fields.

Extra care was devoted on this run to delineate the fields accurately and to use only uniform fields to derive the angle corrections and signatures. Several combinations of fields were tried for derivation of scan-angle corrections, though without important variations in the results. Because this procedure produced only minor improvements and was quite time consuming, no such procedure was attempted for the runs subsequently analyzed.

After the five best channels for recognition were selected, all runs were digitized with the same A-D settings. The entire length of each run was digitized, though only the ground truth areas were used in the final analysis. The data were smoothed during analog-to-digital conversion, with eight analog lines combined into one digital line and two resolution elements combined into one point. To avoid any potential sources of bias resulting from possible variations in the A-D equipment, run 3 was also redigitized. All preliminary work on run 3 was then repeated: field boundaries were rechecked, scan-angle corrections rederived and applied, and signatures again extracted.

Four distinct types of ground cover on 15 of the 31 usable fields, were selected to derive the final angle corrections for run 3. To extract the correction shown in Fig. 17, 30 fields of 10 different ground covers were used.

As noted earlier, two forms of the U-V preprocessing corrections were applied to the data:

(1) A continuous transformation to correct scan-angle dependent variations within each run,

(2) A discrete transformation to match the mean signal levels in each of the later east-west runs to that of run 3, the first east-west run used for signature extraction.

The only distinction between these two forms of transformation is the parameter θ; a continuous θ represents the scan angle within each run for the first case, and a discrete θ represents the entire run in the second case. The June 1972 annual report discussed the theory and some applications of U-V transformations [8].

5.2.1. CONTINUOUS SCAN-ANGLE CORRECTIONS

The computer programs to derive and apply the continuous U-V transformations for correcting scan-angle dependent variations were improved and expanded to incorporate the following features:

(1) It is now possible to combine an arbitrary number of fields having the same ground cover into one class of fields, this one class then being used as data for deriving the U-V corrections in the same way an individual field was used previously. Since this class of fields can cover a larger range of scan angles than an individual field, this feature should make the programs more suitable for use on data where large individual fields are not available (which is particularly the case for higher altitude data). However, this feature can only be used when ground truth is available. All scan-angle corrections on the area survey data made use of this feature (individual fields are small for this 5000-ft data, and ground truth was available on all runs).

(2) The previous program could only derive quadratic corrections—that is, the multiplicative correction of

$$U(\theta) = 1 + (\theta - \theta_0)\mu_1 + (\theta - \theta_0)^2\mu_2$$

and the additive correction of

$$V(\theta) = (\theta - \theta_0)v_1 + (\theta - \theta_0)^2 v_2$$

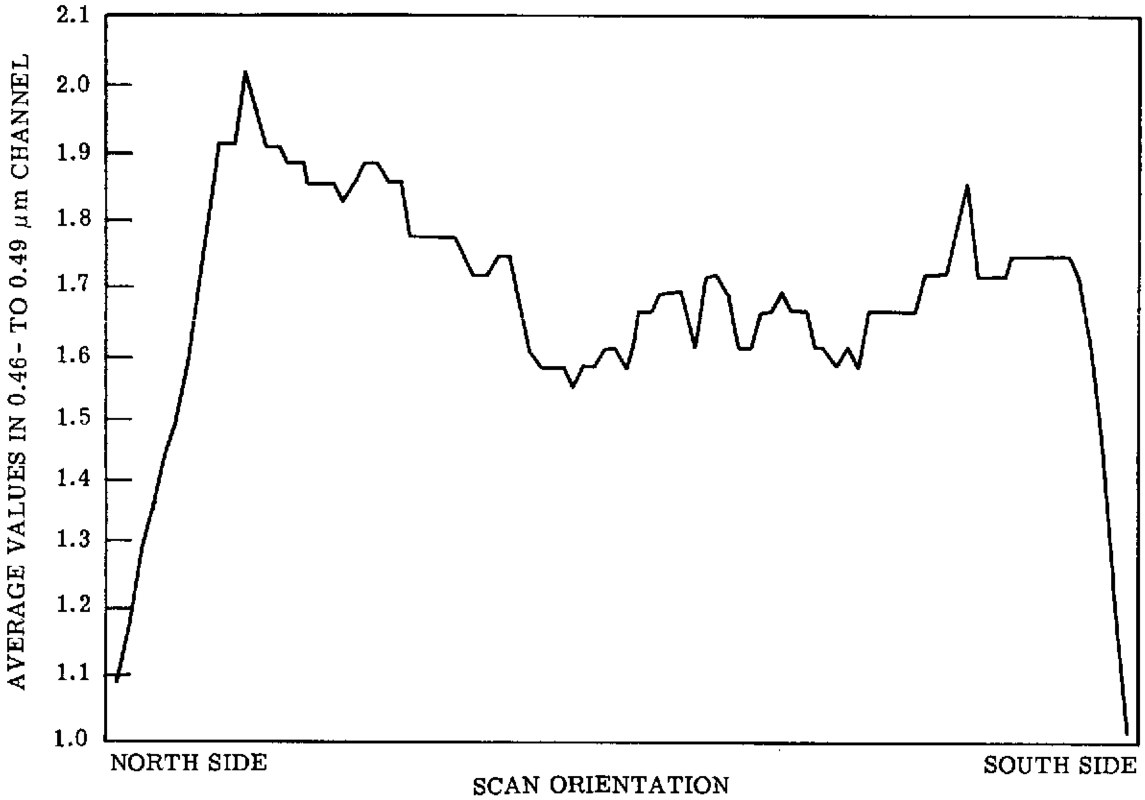

FIGURE 16. AVERAGE SCAN LINE BEFORE SCAN ANGLE CORRECTIONS

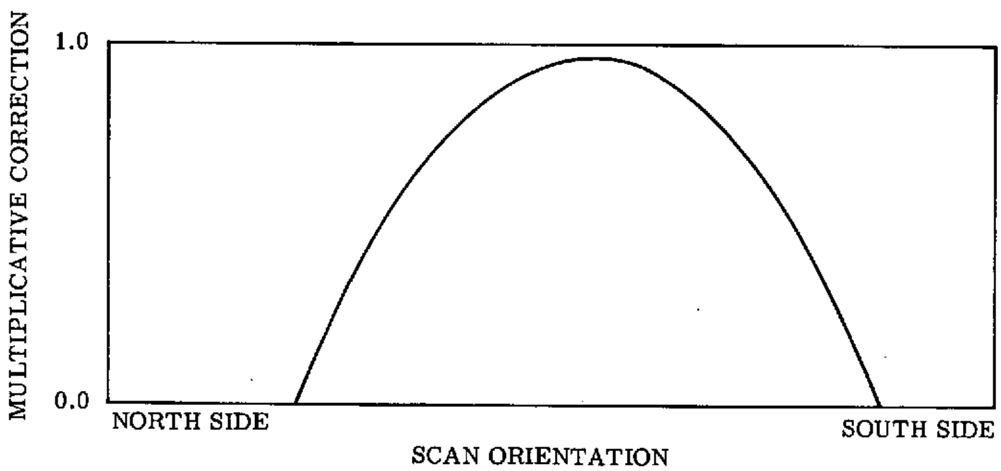

FIGURE 17. THE MULTIPLICATIVE SCAN ANGLE CORRECTION DERIVED FROM FIGURE 16

These are applied to the data as

$$(\text{corrected data value}) = (\text{original data value}) \cdot [U(\theta)] + V(\theta)$$

where θ is the scan angle, θ_0 is a reference scan angle (usually picked near the nadir angle), and a set, $U(\theta)$ and $V(\theta)$, exists for each channel. The new program has been generalized to allow deriving any order $U(\theta)$ and $V(\theta)$ up to the tenth order. Only quadratic order $U(\theta)$ and $V(\theta)$ were used for the scan-angle corrections of the area survey data. Higher order fits may be advantageous where the shape of the scan-angle dependency cannot be adequately approximated by quadratic corrections, although cumulative inaccuracies inherent in deriving corrections from real data may negate any benefits from higher order fit. Conclusive tests have not been made.

The algorithm for deriving the U-V corrections requires a minimum of two and preferably more distinct ground covers to achieve different reflectances in all channels (the channels are handled independently). In practice, we have found three or four distinct ground covers satisfactory (on both the Ingham County area survey data and several other data sets). It is, however, necessary to attain good representation of as many different scan angles as possible in the input data used to calculate the angle corrections, and this is particularly vital at scanner FOV edges. If not, the corrections rapidly degrade beyond the edges of the data supplied. The choice of particular fields appears to be less important if the angular coverage is not affected.

From a theoretical model that assumes Lambertian reflectors, we expect the quadratic multiplicative correction function $U(\theta)$ to be concave upwards, having a minimum of 1 and varying no more than approximately 20% above unity; and the additive correction $V(\theta)$ to be convex upwards and negative, having a maximum of 0 (assuming that the reference angle is properly chosen at the data minimum; the multiplicative function is made exactly 1.0, and the additive function exactly 0.0 at the reference angle).

In practice, we obtained multiplicative and additive correction functions which had the opposite curvatures from those expected and were individually far stronger than expected on the Ingham County area survey data and some other data tried. However, the curves tended to counteract each other and, except near the edges of the run, produce reasonable corrections. In about 75% of the cases the multiplicative correction function went negative near or just within the useful range of scan angles. Deviations from the theoretically expected correction functions may be attributable to bidirectional reflection. But, in practice, the U-V method should probably be considered as one for merely determining the best fit of two arbitrary curves to noisy data.

The U-V transformation is not restricted to Lambertian reflectors, being theoretically capable of completely correcting data where the reflectances are of the form

$$\delta'_{ij}(\theta) = \delta^*_{ij}K_{1j}(\theta) + K_{2j}(\theta) \tag{28}$$

in each channel j, where δ^*_{ij} depends only on the ground cover i and $K_{1j}(\theta)$ and $K_{2j}(\theta)$ depend only on the scan angle θ [2]. However, it has not been established how well any real data corresponds to this formula. Incomplete removal of scan-angle dependencies after the U-V transformation may be explained by different angle dependencies for the reflectances of the different ground covers, in contradiction to the above formula; in which case, the scan-angle corrections applied may be an average correction for the ground covers used as data in calculating the corrections.

The reductions in the scan-angle dependent variations resulting from the continuous U-V scan-angle corrections were studied in most detail in the preliminary run 3 analysis. Angle plots (radiance versus scan angle) of individual fields and the mean and standard deviation of data within individual fields were compared for all fields of the same ground cover at different scan angles before and after angle corrections. These angle plots show how well individual ground covers of interest are angle corrected, and may be more reliable in examining the angle corrections than overall angle plots (radiance versus all scan angles, averaged longitudinally over all data on the length of the run before plotting). Both types of examination were in general agreement. Investigation of individual fields showed we were able to remove 50 to 75% of the scan-angle dependent variations for individual ground covers, while the overall angle plots showed we were removing at least half the variations. But since examining individual fields this way is very time consuming and the overall angle plots generally representative, only overall angle plots were verified in the subsequent analysis.

The 60% increase in radiance noted at the northern edge of east-west runs in the shortest wavelength channel on the uncorrected data was typically reduced to 20 to 30% as viewed on overall angle plots (with correspondingly better results in the longer wavelengths).

Figures 18 and 19 show the multiplicative and additive correction functions derived from the data of Fig. 16 (the shortest wavelength channel of eastbound run 6). These corrections are fairly typical of the shorter wavelengths, illustrating how strongly corrections counteract each other over most scan angles of interest, though the Fig. 16 data has scan-angle variations only half as severe as those of some other runs. Table 4 shows the data after the corrections were applied.

To verify the feasibility of combining small regions known to be of the same material class, we repeated one of the angle corrections using exactly the same regions of ground data but

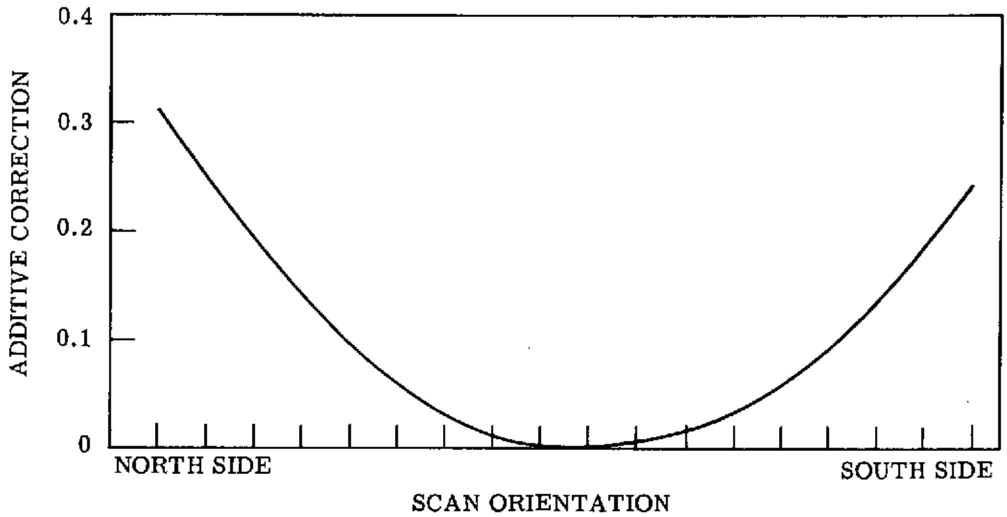

FIGURE 18. THE ADDITIVE SCAN-ANGLE CORRECTION DERIVED FROM FIGURE 16

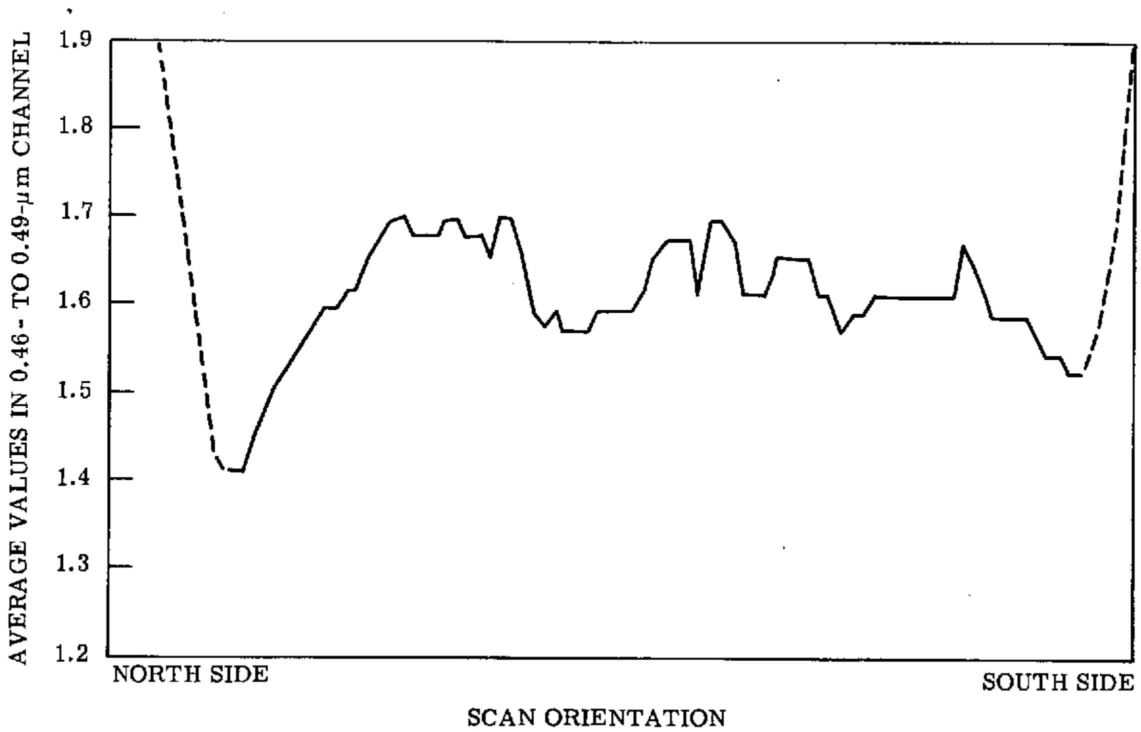

FIGURE 19. AVERAGE SCAN LINE AFTER SCAN-ANGLE CORRECTIONS

TABLE 4. THE GROUND TRUTH FIELDS USED

(a) Number of Fields used from each East-West Run

Run	Total Identified	Angle Correction	Signature Extraction	Recognition	Fields Matching Runs 3	4	5	6	7	8
3	31	15	30	30	--	10				
4	27	24	0	27	10	--	13	13		
5	18	run 7 used	0	18		8	--			
6	33	33	0	33		9		--	15	
7	35	25	0	34				15	--	12
8	35	28	0	32					12	--

(b) Number of Fields of each Ground Cover used for Recognition on the East-West Runs

Run	Corn EW Rows	Corn NS Rows	Dry Beans	Woods	Hay	Oats	Winter Wheat	Buckwheat	Rye	Pasture
3	5	2	5	2	3	4	2	1	1	5
4	10	2	2	0	3	4	1	0	0	5
5	5	3	1	1	3	0	0	0	0	5
6	6	12	0	0	8	1	0	0	1	5
7	6	9	0	5	7	1	2	0	1	3
8	5	9	0	5	4	4	3	0	2	0
Combined for 4 Classes	Corn		Dry Beans	Woods	Grains and Grasses					

without identifying the ground covers to the program; thus the program was required to consider each region as a separate field of potentially different ground cover. The angle corrections resulting from the second run were somewhat degraded. The central corrections were less uniform, and the multiplicative correction coefficient went to a negative value nearer the center. But the resulting corrections still seemed fairly satisfactory.

5.2.2. DISCRETE RUN-TO-RUN CORRECTIONS

The key feature of this area survey study was the adjustment of scanner signal levels in later runs to match those of run 3 before the signatures extracted from run 3 were applied. This level adjustment was done with the discrete U-V preprocessing transformations. Transformations between adjacent pairs of runs were derived from mean radiances for fields observed on each of the two runs. A recursion formula was used to obtain a single correction by combining the $U(\theta)$ and $V(\theta)$ corrections for each successive pair of adjacent runs between run 3 and the run being adjusted. Common fields were available from the region where each east-west run overlapped about 40% of the preceding east-west run. This method has the advantage that no ground truth information is needed to pick suitable fields (such ground truth is generally not available in typical area surveys). As with the continuous U-V transformation used for scan-angle corrections, it is necessary to use at least two different ground covers with distinct reflectances in each channel. Although suitable fields could have been selected from the overlap regions anywhere along the seven mile length of the runs, the already delineated fields from the mile of ground truth were used. Ten to thirteen such fields representing several ground covers of distinct reflectances were available between each pair of runs (except run 5, discussed below). We felt these provided sufficient data.

Run 5, for which adequate ground truth fields for deriving the angle corrections were difficult to locate, proved to be an identification stumbling block. Only 18 fields were considered to be reliably identified and delineated, in comparison with 27 to 35 on the other east-west runs. Only five of these fields were shared by runs 4 and 5, and none by runs 5 and 6. The five fields shared by runs 4 and 5 were augmented with eight other fields selected outside the ground truth area to obtain enough fields for averaging over atypical data. Two combinations of overlap fields were tried, but both gave inferior adjustments; the resulting recognition on run 5 was only about 9% correct for 10 classes; 14 of the 18 fields had no correct points.

Since the results of the overlapping fields method were already so markedly degraded by run 5, we considered run 5 to be anomalously bad, and bridged over it with an alternate method of deriving the U-V transformations. For this approach, we used the ground truth to compare non-overlap fields in runs 4, 5, and 6 (first averaging together fields of the same ground cover within each run, since we were no longer comparing identical fields).

The alternate method also was used to derive run 6 corrections in two ways: directly from run 4 data to run 6 data, and from run 4 to run 5 and then from run 5 to run 6. Both corrections gave essentially the same results, the correct recognitions on run 6 differing by less than 1%.

The basic method of using overlap fields was again used from run 6 on.

5.3. CLASSIFICATION RESULTS

Recognition results were obtained in the ground-truth area for three situations: (1) for each east-west run after scan-angle correction but before the run-to-run signature extension (as a basis for comparison), (2) for each east-west run after run-to-run signature extension by level adjustments with the discrete U-V transformation, and (3) for areas of north-south run 2 corresponding to each east-west run (to see how much variation was the result of differences in the ground cover). We did not perform recognition before the scan-angle corrections, since poor results were expected from the magnitude of scan-angle effects present and also because the improvement resulting from the angle corrections had been previously demonstrated.

Recognition was carried out with the LINMAP linear recognition program, with the ten signatures extracted from run 3 and all five channels selected for the complete digitization. The null-set limit was set at the 0.1% confidence level in all cases.

Experience proved that all ten classes originally selected on run 3 could not be reliably distinguished, so these were then combined into four classes by adding the recognitions of similar classes. The four classes are shown in Table 5. Results were analyzed for both the ten and the four classes. The overall trend on successive runs was the same in either case, although naturally the percent correct recognition was consistently higher for the four classes.

The correct recognition in an individual field was calculated as the number of points correctly classified divided by the number of points correctly classified plus the number incorrectly classified, with non-classified points being ignored (28.10% of the points were not classified on the east-west runs before signature extension, 7.44% after signature extension, and 19.59% on the north-south run).

Table 5 shows the average correct recognition on a run-by-run basis for the three situations described above, while Table 6 illustrates the individual ground cover versus individual class recognition results for signature extraction run 3. The recognition accuracy generally degrades as one progresses away from the training area of run 3.

TABLE 5. RUN-BY-RUN AVERAGE CORRECT RECOGNITIONS

(a) Before application of U-V signature extension transformations to the east-west data
(b) After application of U-V signature extension transormations to the east-west data
(c) For corresponding regions of the north-south run

10 Classes	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8
(a) Before extension (%)	78.291	35.958	22.052	17.511	24.352	20.509
(b) After extension	78.291	38.940	27.972	13.824	6.748	16.591
(c) North-South run	69.788	48.363	27.195	28.882	41.799	40.038
4 Classes						
(a) Before extension	91.477	70.707	53.629	45.741	54.090	43.669
(b) After extension	91.477	75.542	59.427	66.688	38.381	36.155
(c) North-south run	83.463	76.131	62.668	63.967	67.846	72.077

TABLE 6. INDIVIDUAL GROUND COVER VERSUS INDIVIDUAL CLASS RECOGNITION RESULTS FOR RUN 3

	Number Recognized in Each of the 10 Classes										4 Classes		Not Clas-sified	Percent Correct	
	Corn E-W	Corn N-S	Dry Beans	Weeds	Hay	Oats	Winter Wheat	Buck-wheat	Rye	Past-ure	Right	Wrong		10 classes	4 classes
Corn E-W	1600	202	29		2	42	551	1	44		1802	669	50	67.434	78.298
Corn N-S	27	493	14		1	8	82		44		520	149	34	68.014	72.089
Dry Beans	32	20	1544			9	5	1	31		1544	98	36	93.369	93.369
Woods	1			675							675	1	24	99.765	99.765
Hay	1				406	13	1			24	444	1	30	91.450	99.645
Oats		7			24	94	3		3	340	464	7	49	35.984	96.354
Winter Wheat	40	46	4			68	351		18		437	90	22	66.728	82.595
Buckwheat								140			140	0	0	100.000	100.000
Rye	2	5	4			1	2		153		156	11	1	91.617	93.413
Pasture	1				19	21				690	730	1	7	93.163	99.861
	1704	773	1595	675	452	256	995	142	293	1054	6912	1027	253	78.291	91.477

5.4. LIMITATIONS

These results do not represent a satisfactory demonstration of signature extension techniques for area survey operations. On the other hand, and for several reasons, we do not believe that the results should lead one to conclude that signature extension techniques are not adequate for processing area survey data. First, this was an initial attempt and used only one technique of several developed over a period of years; others which have been successfully tested should be tried for comparison. Second, in retrospect, the manner in which the transformations were derived and applied could be improved upon. In particular, the overlap regions used for run-to-run extension were subject to the greatest irregularities in the scan-angle corrections which preceded the run-to-run analysis, and such errors tend to accumulate. Third, there are questions about the representativeness of the signatures used and the makeup of the cover classes identified; certainly the stages of crop development at the time this data set was collected could be a complicating factor.

6
CONCLUSIONS AND RECOMMENDATIONS

The operational use of multispectral scanners for taking resource inventories depends in part on the development of techniques to reduce data processing costs. Such costs can be reduced through the development and use of more efficient decision algorithms and computers and/or, as is usually done, by reducing the volume of data processed.

A linear decision rule has been developed for general purpose (serial) digital computation and tested against the more conventional quadratic decision rule. The linear rule was found to be superior when four or more data channels were used, with its probability of misclassification lower for equal processing times, and its processing time shorter for equal probabilities of misclassification on test fields. Four other linear rules also were tested but found to be less accurate than the best linear rule. Of those suitable for implementation on parallel processors, the equal covariance rule had the best performance. We recommend that the best linear decision rule, as embodied in the ERIM computer program called LINMAP, be implemented on the multispectral data processing systems at MSC for use and further verification.

It was shown that the best linear decision rule can also be used to evaluate the expected performance of recognition processing. When we applied this rule to the problem of selecting subsets of channels, a very substantial reduction in computer time was realized as compared to results with a more exact quadratic procedure. The outputs are readily interpretable in terms of average probabilities of misclassification.

A new approach to the problem of determining linear combinations of channels to reduce processing costs has been introduced. It employs the concepts of the best linear decision rule in its performance measure and considers multiple distributions. A minimum-seeking technique has been implemented. Only the problem of selecting suitable starting points remains before we complete the first full implementation of the approach for testing and evaluation; fortunately, some promising starting point selection methods have been identified. Continued development of the procedure for determining linear combinations of channels is recommended.

It was shown that in processing area survey data, recognition performance degrades as one moves away from the training area. A study of the U-V preprocessing transformation for signature extension to improve recognition results on the area survey data did not produce satisfactory results; while we consider our early testing to be neither complete nor conclusive, the experience gained in this exercise may well enable more effective tests of the U-V preprocessing technique. Other available preprocessing and adaptive techniques for signature extension should also be explored.

Appendix I
SIGNATURE EXTENSION METHODS

The received multispectral signal from any target in any channel is described by the equation:

$$L(\theta) = \rho E(\theta)T(\theta) + L_p(\theta) \tag{29}$$

We develop from this basic equation techniques which improve processing capabilities by removing the effects of changes in $E(\theta)$, $T(\theta)$, and $L_p(\theta)$ when θ changes. If we interpret θ as a scan angle, we then remove the scan-angle effect. If we interpret θ as altitude or flight number, we remove the effects of spatial or temporal changes in the atmosphere. If θ is considered to be a position on the ground, we then extend the approach to two- or three-dimensional correction. We can make these various interpretations of θ because we need not require that θ be continuous. However, if in fact θ is continuous for some particular consideration, then this fact can be used to simplify the correction technique.

The idea of the correction scheme is that Eq. (29) can be written in the form:

$$L(\theta_0) = L(\theta)U(\theta) + V(\theta) \tag{30}$$

where multiplicative correction $U(\theta)$ and additive correction $V(\theta)$ are independent of ρ and hence apply to any ground reflector. Thus a measurement $L(\theta)$ taken under condition θ can be corrected to be the measurement $L(\theta_0)$ that would be measured from the same reflector under condition θ_0, provided $U(\theta)$ and $V(\theta)$ are known. Since Eq. (30) is a linear equation, if measurements of two reflectors with different ρ's are made under conditions θ and θ_0, then $U(\theta)$ and $V(\theta)$ can be computed.

Let us first list some of the possible applications of this technique. A thoughtful examination of this list which is by no means exhaustive, together with the discussion to follow on possible limitations, allows an estimate of the applicability of this technique to most if not all problem areas.

(1) Scan angle. The amount of atmosphere through which the reflected energy propagates depends upon scan angle. The atmosphere affects both the transmission and path radiance.

(2) Cloud shadows. Signals received from ground reflectors that are in cloud shadows may differ greatly from measurements made in adjacent sunlit areas of the same reflectors.

(3) Changes along flight line. This problem area is simply one of extending the scan-angle effect to two or three dimensions.

(4) Different data collections from the same scanner. These collections can be taken either on the same day at different geographical locations and altitudes, or on different days or times of the same day. Corrections of calibration differences are automatically provided in the course of preprocessing; the scanner is operated in its linear range.

(5) Data Bank. This is a result of (4) above. If we can extend signatures from one collection to another, we can collect measured data in a special Data Bank, and withdraw this data for use as signatures in the processing. Where signatures have been identified, they too can be stored in the Data Bank.

(6) Data collection from different scanners. In order to extend signatures from one scanner to another, we are faced not only with the problems inherent in Eq. (29), but also with the fact that there is no one-to-one correspondence between the channels in any two scanners.

Problem areas inherent in Eq. (29), although not explicitly stated, imply that the corrections cannot be made perfectly.

(1) The most obvious difficulty to be encountered occurs when there are not two in-scene references that could be used to find $U(\theta)$ and $V(\theta)$. If only one were available, then either a multiplicative or additive correction, rather than the desired combination, could be used. Many researchers have tried this type of correction, and claim varying degrees of success.

(2) The various sources of noise have not been considered. The noise will affect the precision with which $U(\theta)$ and $V(\theta)$ can be measured. The correction would be applied so that the mean value of the data measured under condition θ corresponded to the mean value under θ_0. The covariance matrix of the corrected data would not necessarily be the same as that taken under θ_0. Some investigation is needed into the measured covariance matrices in order to find a logical way to handle this problem.

(3) The reflectivity ρ may depend upon θ. To the extent that we can write $\rho = \rho_i \rho(\theta)$ where ρ_i is a constant dependent on the type of reflector and $\rho(\theta)$ is common to all reflectors, we can neglect $\rho(\theta)$. We never measure $E(\theta)T(\theta)$ directly, so the inclusion of $\rho(\theta)$ into this product would not change the correction algorithm. Even if we knew the variation of the reflectivity, ρ, with θ, this would not help since we would have to classify the data before we could utilize the knowledge. Since we want to correct the data to im-

prove classification capability, we clearly cannot do this if the correction itself depends upon the classification. An exception might be to use the measured effect of bidirectional reflectivity itself as a signature. This possibility is mentioned for academic interest only, not as a suggestion that should be implemented.

(4) Equation (29) ignores the effect known as green haze. To some extent, limited by noise and the validity of an assumed model, this effect can be eliminated.

(5) Dependence of reflectivity upon time and spatial position. The classification is based upon the assumption that measurement of the reflectivity would be sufficient to identify a ground cover. Reflectivity of a given ground cover, as before and after a rainfall, is not necessarily constant.

We can see that there are problems, not necessarily insurmountable, in applying the correction. Implementation of the technique described herein could be important in that it could extend the usefulness of the multispectral scanner.

REFERENCES

1. M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, Vol. 1, Charles Griffin and Company, Ltd., London, 1963.

2. R. B. Crane, W. A. Malila, and W. Richardson, "Suitability of the Normal Density Assumptions for Processing Multispectral Scanner Data," IEEE Transactions on Geoscience Electronics, Vol. GE-10, No. 4, October 1972, pp. 158-165.

3. T. W. Anderson and R. R. Bahadur, "Classification into Two Multivariate Normal Distributions with Different Covariance Matrices," Annals of Mathematical Statistics, Vol. 33, 1962, pp. 420-431.

4. D. W. Peterson and R. L. Mattson, "A Method of Finding Linear Discriminant Functions for a Class of Performance Criteria," IEEE Transactions on Information Theory, Vol. IT-12, No. 3, pp. 380-387.

5. R. B. Crane and W. Richardson, "Rapid Processing of Multispectral Scanner Data Using Linear Techniques," Presented at the Conference on Earth Resources Observations and Information Analysis System, The University of Tennessee Space Institute, Tullahoma, Tennessee, and published in Remote Sensing of Earth Resources, Vol. 1, ed. by F. Shahrokhi, University of Tennessee Space Institute, March 1972.

6. R. B. Crane and W. Richardson, "Performance Evaluation of Multispectral Scanner Classification Methods," Proceedings of the Eighth International Symposium on Remote Sensing of Environment, October 1972.

7. R. B. Crane, "Processing Techniques to Reduce Atmospheric and Sensor Variability in Multispectral Scanner Data," Proceedings of the Seventh International Symposium on Remote Sensing of Environment, May 1971, pp. 1345-1355.

8. W. A. Malila, R. B. Crane, and R. E. Turner, Information Extraction Techniques for Multispectral Scanner Data, Report No. 31650-74-T, Willow Run Laboratories of the Institute of Science and Technology, The University of Michigan, Ann Arbor, June 1972.

9. W. A. Malila, R. B. Crane, and W. Richardson, "Discrimination Techniques Employing Both Reflective and Thermal Multispectral Signals," Report 31650-75-T, Environmental Research Institute of Michigan, Ann Arbor, January 1973.