CR-171166

E83-10337

SR-P1-04163
NAS9-15466

AgRISTARS

"Made available under NASA sponsorship in the interest of early and wide dissemination of Earth Resources Survey Program information and without liability for any use made thereof."

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Supporting Research

April 1980

ANALYTICAL DESIGN OF MULTISPECTRAL SENSORS

ORIGINAL PAGE IS
OF POOR QUALITY

D. J. Wiersma, D. A. Landgrebe

(E83-10337) ANALYTICAL DESIGN OF MULTISPECTRAL SENSORS (Purdue Univ.) 11 p HC A02/MF A01 CSCL 02C

N83-29748

Unclas
G3/43 00337

Rec.
OCT 22 1981
APDL

Purdue University/LARS
W. Lafayette, Indiana 47906

Lyndon B. Johnson Space Center
Houston, Texas 77058

ORIGINAL PAGE IS
OF POOR QUALITY

041080

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL GE 18, NO. 2, APRIL 1980

Analytical Design of Multispectral Sensors

DANIEL J WIERSMA, MEMBER, IEEE, AND DAVID A LANDGREBE, FELLOW, IEEE

Abstract—An analytical procedure for the design of the spectral channels for multispectral remote sensor systems is defined. An optimal design based on the criterion of minimum mean-square representation error using the Karhunen-Loeve expansion was developed to represent the spectral response functions from a stratum based upon a stochastic process scene model. From the overall pattern recognition system perspective the effect of the representation accuracy on a typical performance criterion, the probability of correct classification, is investigated. The optimum sensor design provides a standard against which practical (suboptimum) operational sensors can be compared. An example design is provided and its performance is illustrated.

Although the analytical technique was developed primarily for the purpose of sensor design it was found that the procedure has potential for making important contributions to scene understanding. It was concluded that spectral channels which have narrow bandwidths relative to current sensor systems may be necessary to provide adequate spectral representation and improved classification performance.

I Introduction

A PATTERN recognition system as used in a remote sensing system for earth resources consists of three fundamental components—the scene, the sensor, and the processor (Fig 1) The scene is that portion of the earth's surface observed by the sensor The desired information is contained in the spectral, spatial, and temporal variations of the electromagnetic energy emanating from the scene The sensor collects the energy and measures its features The processor is typically a digitally implemented classification algorithm which makes an appropriate decision based on the feature measurements provided by the output from the sensor Various types of ancillary data are also now typically used in the decision-making process

At present the design of the processor algorithms is quite advanced and provides variety and flexibility for optimal performance given a feature set [4], [5] However, the design of the best set of features is a complex matter which is not well understood There are many sensor parameters involved [11, ch 7] In the current work we limit considerations to the design of the spectral aspects of features to make the problem more tractable, leaving other aspects to later occasions

Manuscript received January 4, 1979, revised December 7, 1979 This work was supported by NASA/Johnson Space Center under Contracts NAS9-14970 and NAS9-15466

D J Wiersma was with the Department of Electrical Engineering and the Laboratory for Applications of Remote Sensing (LARS), Purdue University, Lafayette, IN He is now with the Minneapolis Honeywell Company, Minneapolis, MN

D A Landgrebe is with the Department of Electrical Engineering and the Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, IN 47907

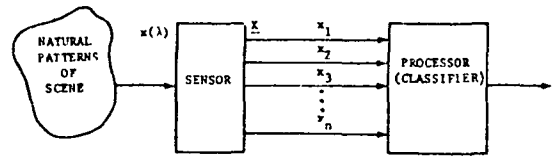


Fig 1 Pattern recognition system

II Spectral Representation and Optimum Sensor Design

In order to achieve an optimal design of a set of spectral features one must have suitable analytical representations for 1) the spectral response of the scene, 2) the sensor system, 3) the processor system, and 4) one must have a suitable, analytically expressible optimality criterion Further, we note the following factors which influence the creation of a spectral feature design procedure

1) The scene is very complex in the fashion in which it reflects and emits optical radiation Mathematical models which predict the scene radiant exitance at least to the level of accuracy and precision needed for our problem, do not yet exist As a result an empirical scene model must be used

2) Because satellite-borne sensor systems are very expensive they cannot usually be designed specifically for a certain use or user Rather they must be optimized with regard to a large number of scenes and uses (Fig 2) The feature space which the sensor defines must be adequately detailed, for example, such that in early season when agricultural crop canopies have achieved only 10-15-percent cover, both the crop species mapping user and the soils mapper can be served This fact is important in the choice of optimality criterion, as will be seen shortly

3) It is highly desirable that the spectral features be designed in such a way that they are maximally efficient in the sense that a feature set of any given size contain the maximum amount of useful information possible so that any given analysis can proceed with the smallest number of features possible There are at least three reasons for this feature efficiency in this sense tends to decrease the amount of processor computation required, it tends to decrease the processor complexity required, and it tends to reduce the amount of training sample data needed

4) There are a number of constraints on the design of a sensor, generally of a practical character, which cannot reasonably be expressed analytically Examples are those resulting from optical design considerations, sensor material sensitivity curves, cost factors, spacecraft size and weight

0196-2892/80/0400-0180$00 75 © 1980 IEEE

ORIGINAL PAGE IS
OF POOR QUALITY

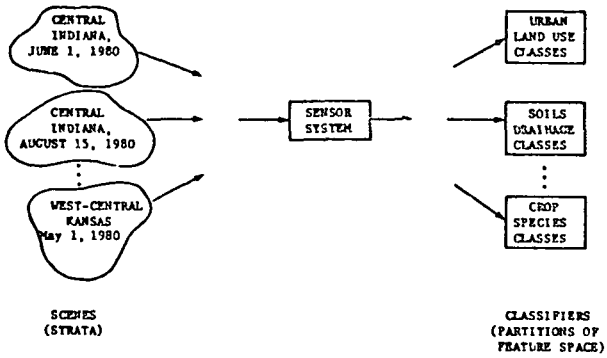


Fig 2 A sensor system must be designed to perform satisfactorily for many reasons scenes at various times of the season, i e , defining features spaces detailed enough to permit successful partitions for varying applications

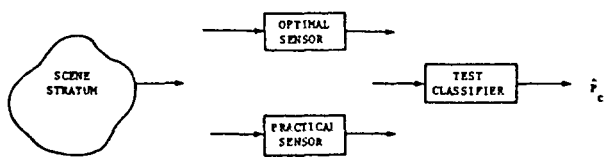


Fig 3 Sensor system design procedure

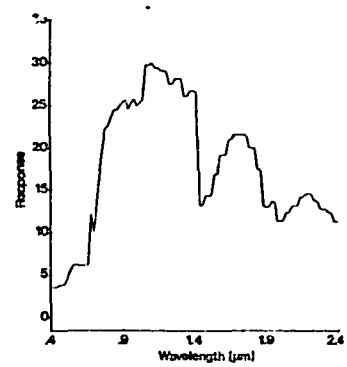


Fig 4 Spectral response function for Mature Wheat collected on August 4, 1977, over Williams County, ND

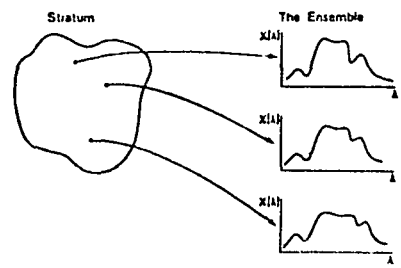


Fig 5 Realization of a stratum as the ensemble of spectral sample functions

considerations, etc , and especially from the interrelationships of these types of factors To mitigate this circumstance we will use the scheme depicted in Fig 3. That is, we will determine optimal sensor characteristics using entirely analytical means without regard for their physical realizability These characteristics will then serve as a guide by which to determine nearly equivalent but physically realizable characteristics which perform nearly as well

In order to achieve the desired spectral feature design scheme we must deal specifically with each of these lettered and numbered items above

A The Scene Model

Let us begin by considering the information bearing aspects of the spectral response function $x(\lambda)$ [8] This response function (e g , from a single pixel) is proportional to the electromagnetic energy received by the sensor as a function of wavelength λ (Fig 4) Many factors determine the spectral response function for a given observation The irradiance of the sun, the conditions of the atmosphere, and the reflectance of the surface features all have important effects on the response Since a deterministic relationship between the response function and the many factors affecting it would be very complex, the set of functions which are observed in practice are best modeled as a stochastic process

The ensemble of the stochastic process [13] will be defined in terms of the stratification necessary to apply pattern recognition methods to the earth observational problem A stratum S is defined as the largest contiguous area which can be classified to an acceptable level of performance with a single training of the classifier It is noted that the sensor must be designed to operate satisfactorily over a large number of such strata, which vary greatly with time, location and application The collection of all possible strata which a sensor may observe is denoted by S_0 Since the set S_0 is quite large, it is necessary to select a smaller subset which is representative in a statistical sense in order to perform the design

The random experiment for the stochastic process consists of the observation of a point in a stratum S Each point in the stratum is mapped into a spectral response function (Fig 5) The collection of all response functions from a stratum defines an ensemble The ensemble plus the corresponding probability measure defines the stochastic process [13] It is appropriate to assume a Gaussian probability measure for this process [3].

B The Sensor System Model

Next we choose a mathematical model for the sensor to represent the spectral response function for each observation Let the sensor be represented by a set of N filter functions or basis functions $\{\phi_i(\lambda)\}$ such that the output of each filter is given by (Fig. 6)

$$x_i = \int_\Lambda x(\lambda)\,\phi_i(\lambda)\,d\lambda \tag{1}$$

The output of the sensor model is a sequence, $\{x_1, x_2, \ldots, x_N\} = X$, which represents the spectral response by the

ORIGINAL PAGE IS OF POOR QUALITY

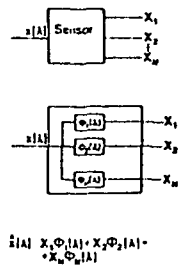


Fig 6 Sensor system model

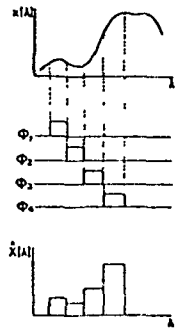


Fig 7 Approximation of the spectral response function by a set of four basic functions

approximation

$$x(\lambda) \approx x_1\phi_1(\lambda) + x_2\phi_2(\lambda) + \cdots + x_N\phi_N(\lambda)$$
$$= \sum_{i=1}^{N} x_i\phi_i(\lambda) = \hat{x}(\lambda) \tag{2}$$

A simple illustration of the concept is given in Fig 7 However, by relaxing the usual restrictions on the shape of the $\{\phi_n(\lambda)\}$, considerable advantage can be obtained There is no theoretical or practical reason, for example, for the $\{\phi_n(\lambda)\}$ to be nonoverlapping What is needed is to determine the ordered set of basis functions which are optimal with regard to a meaningful system performance criterion

C The Optimality Criterion Including the Processor Model

A key consideration is the choice of the optimality criterion It is desirable to optimize the sensor design with respect to an overall system (including the scene and processor) performance criterion The probability of correct classification is the criterion to be used here This choice is made because it is an objective indicator of desired performance in a practical sense for a large proportion of applications, and it is perhaps the best studied and understood in a theoretical sense In selecting this performance measure there must also be associated with it a processor (classifier) model, in this case we chose the maximum likelihood rule, for the same reasons

However, because of the design factor pointed out in item (2) and Fig 2, we find it desirable to define an intermediate optimality criterion Because the sensor must function over a varied collection of strata using any of a large collection of classifiers, a criterion was chosen which is a measure of the fidelity with which the output of the sensor represents the input We will choose the set $\{\phi_i(\lambda)\}$ such that for a given $x(\lambda)$ the approximation $\hat{x}(\lambda)$ is as close as possible to the true spectral response function One may think of this approach as one intended to minimize the information loss through the sensor even though it cannot be known to the sensor designer what the information is In passing from $x(\lambda)$ to $\{x_i\}$ there is no information loss if $x(\lambda)$ is recoverable from $\{x_i\}$

A common criterion for representation accuracy is the expected mean-square representation error given by

$$E\{\epsilon_r\} = E\left\{\int_\Lambda [x(\lambda) - \hat{x}(\lambda)]^2\, d\lambda\right\} \tag{3}$$

However, it is desirable at this point to generalize this criterion by introducing a weight function $w(\lambda)$ on the spectral interval As will be seen, the weight associated with each λ can be used to introduce into the analysis *a priori* knowledge concerning the spectrum Thus (1) and (3) become [16].

$$x_i = \int_\Lambda x(\lambda)\,\phi_i(\lambda)\,w(\lambda)\,d\lambda \tag{1a}$$

$$E\{\epsilon_r\} = E\left\{\int_\Lambda [x(\lambda) - \hat{x}(\lambda)]^2\, w(\lambda)\, d\lambda\right\} \tag{3a}$$

We want to choose the set of basis functions $\{\phi_i(\lambda)\}$ which is optimal with respect to the spectral representation criterion of expected mean-square error $E\{\epsilon_r\}$ More specifically, it is desired that the representation be complete in the sense that the expected mean-square error for any function in the ensemble be made arbitrarily small simply by including enough terms, that convergence of the approximation to the original response be rapid in the first few terms, and, without loss of generality, we may also ask that the basis functions be orthogonal to each other

A technique for determining the set of optimal basis functions for an ensemble which satisfies the desired properties is based on the weighted Karhunen-Loeve expansion [2], [16], [17] The solution to the homogeneous linear integral equation is

$$\gamma_i\phi_i(\lambda) = \int_\Lambda K(\lambda, \xi)\,\phi_i(\xi)\,w(\xi)\,d\xi \tag{4}$$

with the covariance function of the stochastic process, $K(\lambda, \xi)$, as kernel is a set of eigenfunctions $\{\phi_i(\lambda)\}$ with corresponding eigenvalues γ_i If the eigenvalues are arranged in descending order, the corresponding sequence of eigenfunctions can be used to form a linear combination of the eigenfunctions which converges to the original spectral response function with arbitrarily small expected mean-square error Furthermore, because of the ordering of the eigenvalues, convergence in the first few terms is very rapid This rapid convergence allows truncation of the series expansion after a finite number of terms N with mean-square error minimized over all possible choices of N basis functions The mean square error is given by

$$E\{\epsilon_r\} = \sum_{i=N+1}^{\infty} \gamma_i \tag{5}$$

ORIGINAL PAGE IS
OF POOR QUALITY

Since the unmodified ($w(\lambda) = 1$) Karhunen-Loeve expansion is a well-studied technique and satisfies the desired properties for finding the basis functions, a sound analytical method is provided for determining the optimal set of basis functions

The optimal sensor design problem may be solved on a digital computer using empirical data taken by field measurements Some approximations must be made in order to take into consideration some practical constraints First the response functions are not available as continuous functions but are obtained in the field by sampling the spectrum with an instrument that uses very narrow spectral windows Secondly, the parameters of the process are not known *a priori*, hence, it is necessary to estimate the mean and covariance functions using a representative sample from the ensemble Finally, because the data will be stored and processed digitally it is necessary to quantize the amplitude of the response at each of the spectral sample points Each of these constraints potentially can contribute to the representation error It has been shown that with reasonable care in selecting a sufficiently high spectral sampling rate, a large enough sample from the ensemble, and a large number of quantization intervals that the contribution of these factors to the representation error is small [16] The integral equation (4) becomes the matrix equation

$$\Phi\Gamma = \Lambda W \Phi \tag{6}$$

where Φ is the matrix of eigenvectors, Γ is the diagonal matrix of eigenvalues, K is the covariance matrix, and W is the diagonal matrix of weight coefficients

III Relationship Between the Spectral Representation and System Performance

The performance of the overall system is ultimately what we wish to optimize For this purpose, as previously indicated the probability of correct classification P_c has been chosen as the performance indicator to be optimized If the vector X is an observation from one of M classes C_i, $i = 1, 2, \ldots, M$ with *a priori* probabilities P_i, the probability of correct classification, using the maximum likelihood rule is given by

$$P_c = \int \max_i \{P_i p(X|C_i)\}\, dX \tag{7}$$

where $p(X|C_i)$ is the conditional (multivariant) probability density function for class i The integral in (7) is over the observation space

The analytical procedure based on the weighted Karhunen-Loeve expansion has prescribed a sensor design which minimizes the mean-square representation error One would like to know how the ability to represent a process influences the classification performance To study this relationship the graph of the probability of correct classification versus expected mean-square error is introduced (Fig 8) We will briefly discuss some of its characteristics

The addition of terms to the series expansion causes a monotonic decrease in the spectral representation error, but the effect of the additional terms on the overall system performance must be determined It can be shown that increasing the number of terms in the representation will never decrease the performance provided that the stochastic process is completely known If after N terms the improvement in performance is small compared to the reduction in representation error, then the representation is sufficient This is illustrated by case A of Fig 8 in which the threshold T indicates the minimum required $E[\epsilon_r]$ However, if the performance is showing significant improvement for a small decrease in the mean-square error, case B of Fig 8, more terms are necessary to complete the representation

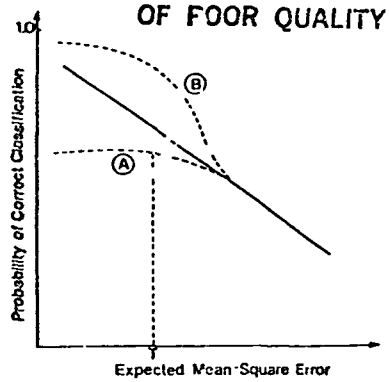


Fig 8 Probability of a correct classification as a function of expected mean-square representation error

Since the parameters of the stochastic process must be estimated from a sample of the ensemble, the effect of the size of the sample relative to the dimensionality of the system is important Hughes [10] has shown that if the sample size is too small, the classification performance may actually be degraded by adding terms to the expansion Thus it is necessary to maintain a large set of sample functions from which to estimate the statistics

The choice of information classes also influences the performance of the pattern recognition system For purposes of classifying the data into distinct classes it is required that the class list have the following properties simultaneously [11]

1) Each class must be of interest to the user, i e , of informational value

2) The classes must be separable in terms of the features available

3) The list must be exhaustive, i e , there must be a class to which it is logical to assign each pixel in the scene

The classes may be arranged in a hierarchial tree structure in which case classes deeper in the tree require different representation accuracies to achieve a given level of classification performance

The area of the ground resolution element is determined by such system characteristics as the instantaneous field-of-view (IFOV), the altitude of the sensor, the scan rate, and the velocity of the sensor These are examples of spatial representation parameters The size of the objects which can be identified and the energy available are influenced by the choice of ground resolution element size. If a typical object which one wishes to identify is smaller than the ground resolution element size, then it is very difficult to classify that object Mobasseri [12] has investigated the effect of the area of the

ORIGINAL PAGE IS
OF POOR QUALITY

resolution element on classification performance Increasing the area often improves the signal-to-noise ratio which in turn can improve the classification performance

For a given remote sensing problem the *signal* is the part of the received response which is information bearing, and the *noise* is that part which is noninformation bearing The influence of the signal-to-noise ratio where the noise is white, Gaussian and additive was demonstrated by Ready *et al* [14] Results show that overall classification performance decreased with an increase in the noise level A class which was difficult to identify under low noise level conditions suffered the most degradation when noise was added

IV Experimental System

An experimental software system has been set up to implement the analytical procedure that has been developed The software system has been implemented on an IBM 370 computer at the Laboratory for Applications of Remote Sensing (LARS) at Purdue University, Lafayette, IN

A collection of field data consisting of spectral response functions on three dates from Williams County, ND, and three dates from Finney County, KS, was available from the field measurements library at Purdue/LARS More than one thousand spectra were available from each location and collection date The response functions were sampled in wavelength using narrow windows of 0 02 μm over the range $0\ 4 \leqslant \lambda \leqslant 2\ 4\ \mu$m

The optimal set of basis functions is found numerically by estimating the covariance matrix from the sample response function Maximum likelihood estimates of the mean and covariance matrices are given by

$$\bar{X} = E\{X\} \approx \hat{\bar{X}} = \frac{1}{N_s}\sum_{i=1}^{N_s} X_i \tag{8}$$

and

$$K \approx \frac{1}{N_s}\sum_{i=1}^{N_s} (X_i - \hat{\bar{X}})(X_i - \hat{\bar{X}})^T \tag{9}$$

where N_s is the number of sample functions available and X_i is the ith sample vector The covariance is the kernel in the linear integral equation whose solutions are the optimal basis functions or eigenvectors A remarkably stable and accurate method for numerically computing the eigenvalues and eigenvectors was published by Grad and Brebner [6]

The eigenvectors are used to perform the linear transformation

$$y_i = \iota^T (X - \hat{\bar{X}}) \tag{10}$$

on the data set The class conditional statistics are computed using the transformed data

In order to compare the performances of two systems an algorithm which estimates the probability of correct classification for an M class problem, given the class conditional multivariate Gaussian statistics, was used [16] This algorithm, which is based on the stratified posterior estimator [15], was found to be accurate within one-half of 1 percent

The experimental system also included an ability to simulate (suboptimal) practical sensors Although nearly any sensor characteristic could be simulated, most of the sensors which were simulated consisted of a small set of rectangular basis functions, i e ,

$$\psi_k(\lambda) = \begin{cases} 1\ 0, & \lambda_{lk} \leqslant \lambda \leqslant \lambda_{uk} \\ 0\ 0, & \text{otherwise} \end{cases} \tag{11}$$

where the λ_{lk} and λ_{uk} are the lower and upper limits of the spectral bands The spectral bands of two suboptimal sensors which were implemented are listed in Table I

TABLE I
SPECTRAL BAND LOCATIONS FOR TWO PRACTICAL SENSOR DESIGNS*

Sensor Number 1		Sensor Number 2	
Band	Wavelength	Band	Wavelength
k	λ_{lk} to λ_{uk}	k	λ_{lk} to λ_{uk}
1	5 to 6µm	1	45 to 52µm
2	.6 to 7µm	2	52 to 60µm
3	7 to 8µm	3	63 to 69µm
4	8 to 1 1µm	4	76 to 90µm
		5	1 55 to 1 75µm
		6	2 08 to 2 35µm

*The band edges of these sensors were selected to coincide with the nominal bandwidths of the MSS of Landsats 1, 2, and 3, and those of the TM of Landsat D

V Results

It became necessary during this research to create a means for incorporating certain types of ancillary information into the optimization process The reason for this can best be illustrated with an example We first determine the optimal features, $\{\phi_k(\lambda)\}$, for a data set using the uniform weight function of Fig 9 Plots of the first four eigenvectors for this case are shown in Fig 10 Now the shape of an eigenfunction is instructive in this application because whenever an eigenfunction is large (either positive or negative) at a given λ, it implies that that wavelength is important in representing the given ensemble and, therefore, perhaps also in classifying it

It was observed in this example that the eigenvectors are dominated at least in the case of $k = 1$ and 4 by components in the bands near 1 4 and 1 9 μm These bands are dominated by water absorption in the atmosphere, and it is known *a priori* that very little solar flux reaches the earth's surface in them Thus these regions near 1 4 and 1 9 μm represent noise rather than signal, and under test, the eigenvectors which were sensitive to these bands contributed very little to the classification performance Using a second weight function which assigns a very small weight in the water absorption bands (Fig 11) the influence of the spectrum in the water absorption bands is significantly reduced, and a marked improvement in classification performance (e g , a classification error rate reduction from 19 6 percent to 11 8 percent for a six-feature classification) is observed

It is a characteristic of the Karhunen-Loeve representation that spectral intervals with the greatest ensemble variation are given greatest emphasis Some of this ensemble variation may be expected to be information-bearing (signal) and some not (noise). As can be seen by this example, the weighting function provides a convenient means to incorporate into the feature design prior knowledge about whether ensemble variations in a given range are signal or noise In the remainder of

ORIGINAL PAGE IS
OF POOR QUALITY

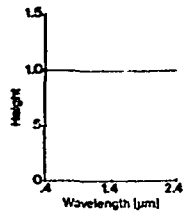


Fig. 9 Weight function number 1

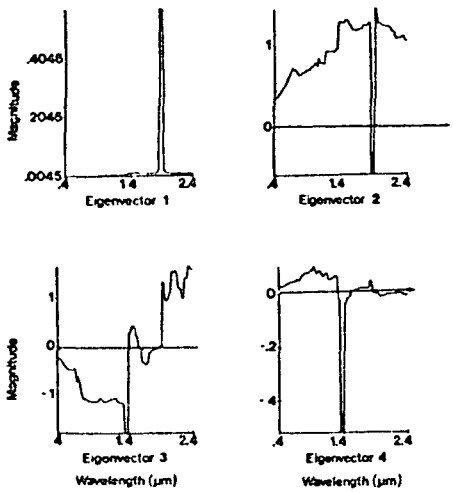


Fig 10 First four optimum basis functions using weight function number 1 over Williams County, May 8, 1977, data

Fig 11 Weight function number 2

the results presented, the weighting function of Fig 11 was used Two other weighting functions have been tested [16], the choice of which is the best weighting function is thought to depend on a number of factors not all of which are well understood

Results for the design of the optimal sensor for one of the six sets of data are presented in Figs 12-14.[1] The data from the stratum was collected over Williams County, ND, on May 8, 1977 The three classes represented in this data set were spring wheat, summer fallow, and pasture Since the data were taken relatively early in the growing season, the wheat canopy was still quite sparse which leads one to expect the wheat class to resemble spectrally bare soil or fallow In Fig 12 the expected mean square error is plotted as a function of the number of terms in the expansion Rapid convergence in the first few terms is demonstrated in this graph

The first eight weighted eigenvectors are plotted in Fig 13 Note that the magnitude of the first eigenvector has the general shape of a typical spectral response of soil, since the predominant response in two of the three classes is soil, this first eigenfunction may be thought of as the average response over the ensemble The second eigenvector tends to divide up the spectrum into three to four relatively broad regions As the number of terms in the expansion is increased, the terms that are added require higher spectral resolution to reduce mean-square error, indicating that spectral fine structure may be increasingly important. We will comment further on this later.

In Fig 14 the estimated probability of correct classification is graphed as a function of the expected mean-square error It is seen that at least the first eight basis functions contribute significantly to the classification accuracy The numbers in parentheses on Fig 13 indicate the order in which the features were chosen by a feature selection processor[2] The relatively broad bands of $\phi_2(\lambda)$ are indicated as of greatest importance to class discrimination but apparently rather quickly thereafter the finer spectral structure represented by $\phi_6(\lambda)$ becomes important

To complete the design illustration, the procedure was used to determine a set of physically realizable rectangular spectral bands Six to ten bands were required based on a study of the dimensionality An initial set of band edges was chosen based on an examination of the eigenvectors of the May 8 data set From (1a) it is observed that regions of the spectrum that are significantly different from zero for a particular eigenvector may contain important information This, together with the feature ordering determined by a standard feature selection algorithm, indicates spectral intervals important for discrimination Zero crossings of the eigenvectors may be possible indications of band edges

Based upon such a study of the optimal functions a set of rectangular spectral bands was chosen. The final set of band edges for the example design is presented in Table II In Fig 15, comparison of the two simulated sensors of Table I and the optimum set is made on the basis of classification performance for the May 8 data The first ten optimal basis functions were used from the optimal design, where 10 features provide a small mean-square error yet keep the computational difficulties involved with high dimensional systems to a reasonable level As shown in Fig 15, the four-band sensor 1 compares rather poorly with the best four[3] optimal basis functions, showing sensor 1 to be far from optimal for this set of classes Sensor 2 is an improvement over sensor 1, but still somewhat lower in performance than desirable The eight-band example design, however, demonstrates very good performance relative to the performance of the ten feature optimal set. Also, the expected mean-square error is significantly reduced over that of sensors 1 and 2 It should be

[1] The results for the other five data sets, which were from different times of the year at the North Dakota site and a Kansas site, are contained in [16]

[2] The feature selection scheme is based upon maximizing the average interclass-pair divergence The details are in [16]

[3] The term "best" refers here to those chosen by a feature selection algorithm as best for discriminating between the design set classes

ORIGINAL PAGE IS
OF POOR QUALITY

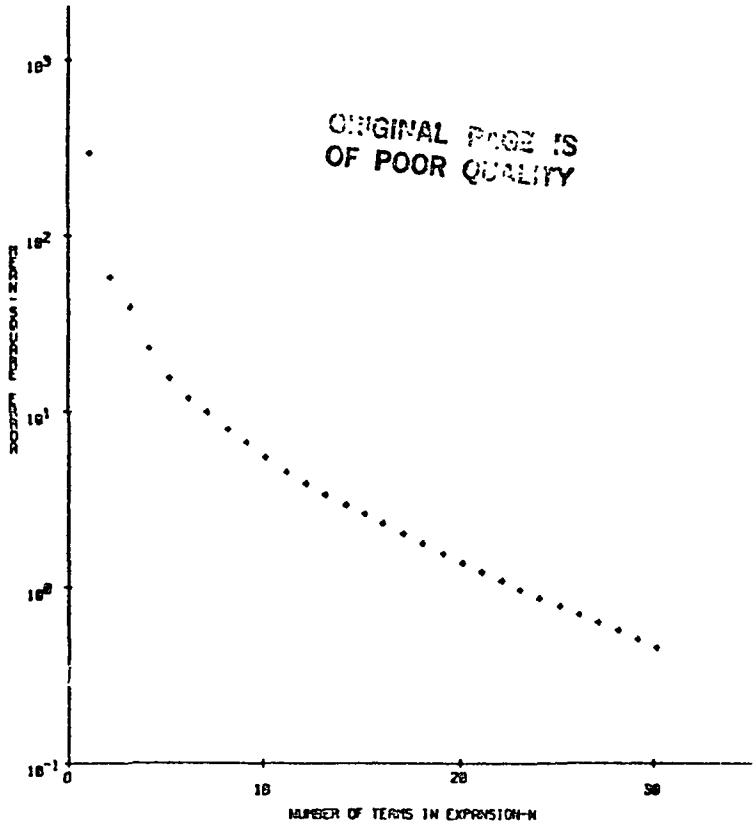


Fig 12 Expected mean-square error as a function of the number of terms in the Karhunen–Loeve expansion for Williams County, May 8, 1977, using weight function number 2

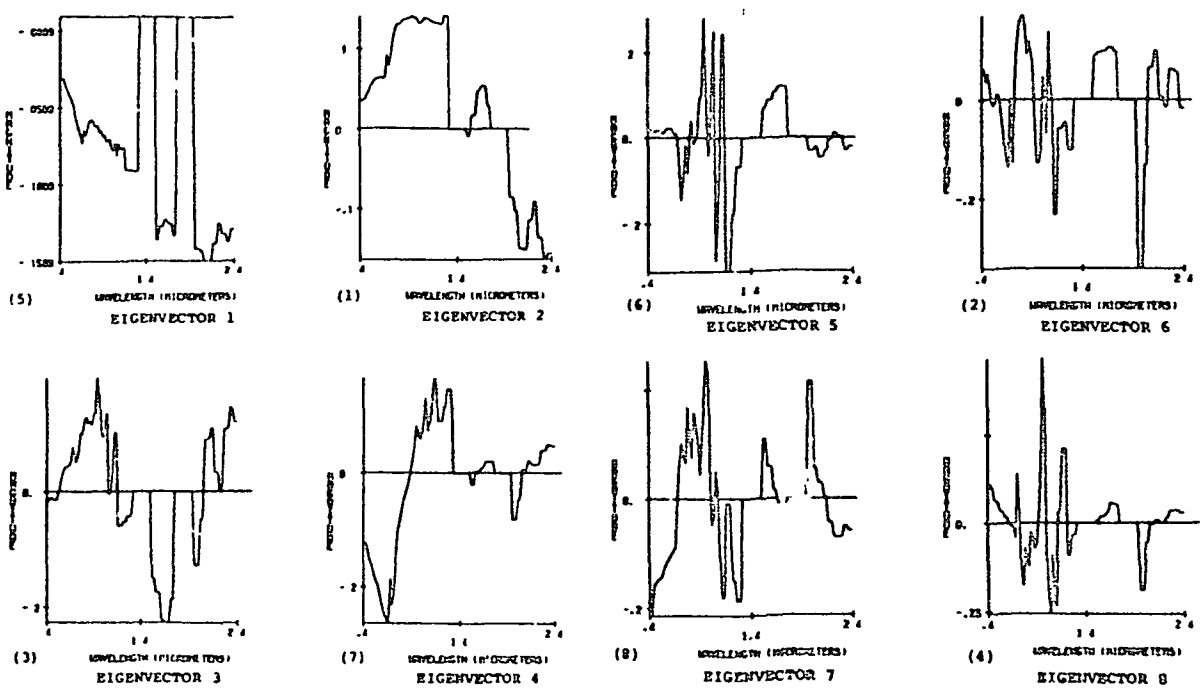


Fig 13. First eight eigenvectors for Williams County, May 8, 1977, using weight function number 2.

ORIGINAL PAGE IS
OF POOR QUALITY

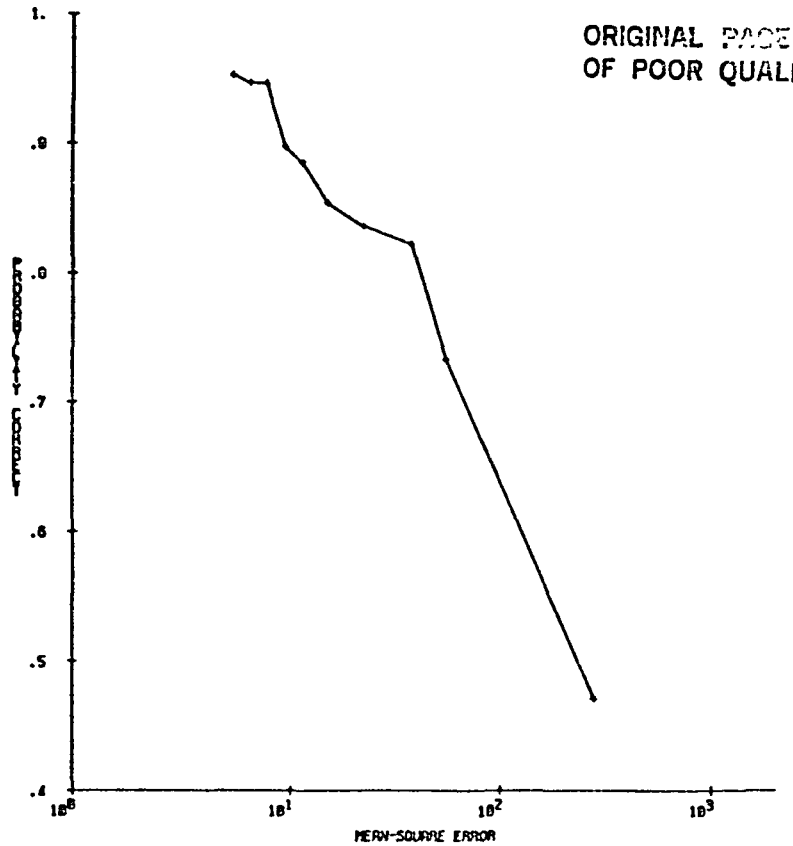


Fig 14 Estimate of probability of correct classification versus expected mean-square error Williams County, May 8, 1977, using weight function number 2

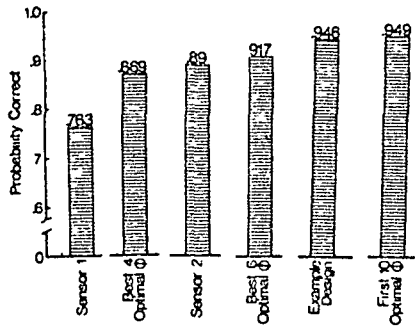


Fig 15 Comparison of sensor performance Williams County, May 8, 1977

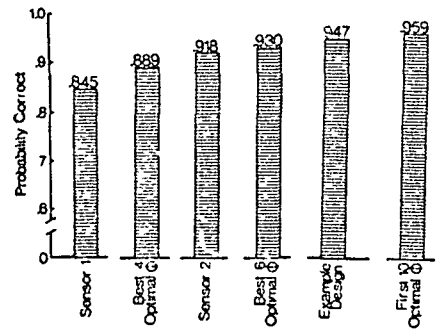


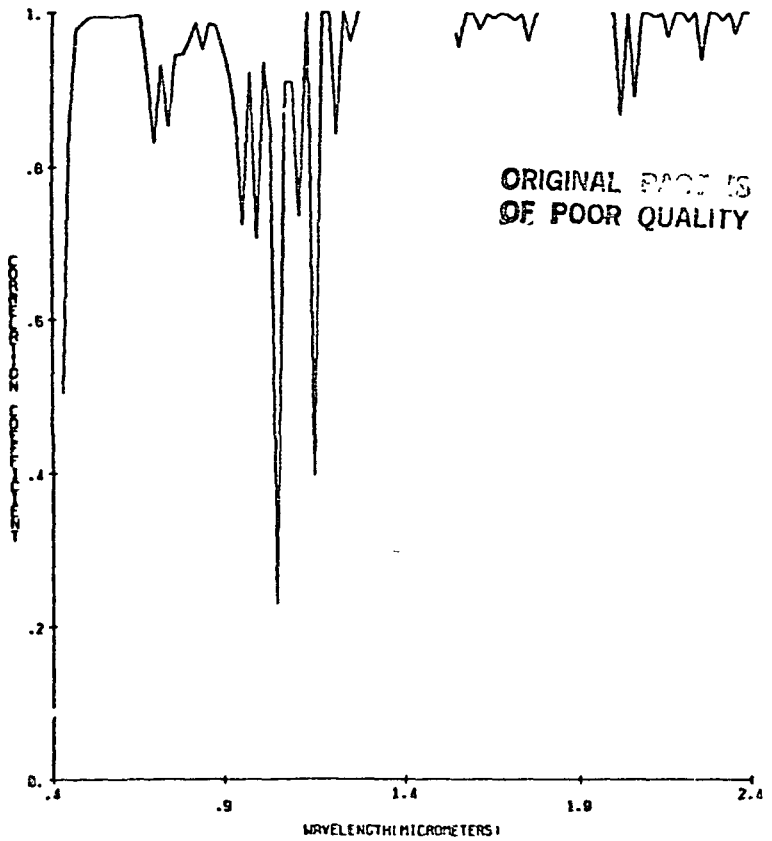
Fig 16 Comparison of sensor performance Average of six cases

TABLE II
SPECTRAL BAND LOCATIONS FOR THE EXAMPLE DESIGN

Channel	Endpoints
1	42μm - 54μm
2	56μm - 66μm
3	68μm - 70μm
4	72μm - 90μm
5	92μm - 1 00μm
6	1 02μm - 1 30μm
7	1 52μm - 1 74μm
8	1 92μm - 2 40μm

pointed out that these comparisons of classification performance are based on a particular set of information classes and the results may be different if different classes were chosen The reduced representation error may imply increased robustness in performance for differing information classes

The performance was evaluated for the six data sets and the average performance for each of the sensors is shown in Fig 16 Again the experimental sensor is quite close in performance to the optimal, a difference of about one percent.



ORIGINAL PAGE IS
OF POOR QUALITY

Fig 17 Correlation coefficients versus wavelength for Williams County, May 8, 1977

Although the primary purpose of this work was to develop a design procedure for sensors, important contributions to the understanding of the scene can be gained Properties such as signal dimensionality for a given stratum, maximum possible class separability, the spectral resolution required, and the accuracy of spectral representation required to obtain a given level of performance can be studied

The dimensionality of the ensemble can be defined as the number of terms in the expansion necessary to represent the original waveform to the desired accuracy For the information classes used on the six data sets the dimensionality was between six and eight Some of the strata required fewer terms in the expansion to obtain a good representation and good performance while others required more terms

A review of the eigenvectors of Fig 13 shows an unexpected amount fine structure, particularly in the region from 0 66 to 1 26 μm as previously noted Up to this time it had often been observed that spectral bands very near to one another tend to be highly correlated and thus redundant To investigate this structure a bit further the correlation coefficient between each band of width 0 02 μm and its nearest neighbor was calculated The results are shown in Fig 17 Based upon previously known results one might expect the graph of the figure to deviate little from unity The fact that it does deviate significantly from unity changing significantly in very narrow bands may be further evidence that there is useful information in the spectral fine structure

VI Conclusion

The design of the spectral bands for multispectral instruments has been an important matter for a number of years In the past such design has of necessity been carried out as a spectral band selection problem using in a subjective manner the direct application of experience and judgment Nonoverlapping bands were selected one-by-one based upon the current understanding of reflectance phenomena Each band was generally selected on its own individual meters because evaluating sets of bands is difficult to do without computational aids A typical example of this process was the design of the Thematic Mapper bands [7] Although this procedure has served adequately to this time, as the level of sophistication of applications and, therefore of data needs continues to grow, it seems clear that more quantitative and objective design tools are needed It was this need to which the work reported herein was directed

As has been seen the design technique which resulted utilizes an empirical stochastic scene model as the best currently feasible means for defining quantitatively the scene and its information and noise content Since not all sensor system limitations can be analytically specified, the technique utilizes an optimal design calculation but it still requires engineering judgment as the best means for resolving the tradeoff of more tangible factors with those which are less tangible It is by no means intended as an "automatic" design procedure

The optimal design calculation was established utilizing an overall performance measure, probability of correct classification, and an intermediate one, spectral representation error averaged over the ensemble. One can argue that it is overall system performance which is, after all, the most important. However, a specific performance measure calculation can only be carried out with regard to a specific set of classes Different applications frequently require the same data set to be classified into different class sets; thus ore desires robustness with regard to various class sets as well in a feature set. The hypothesis used here is that a concern for representation accuracy will tend to insure this robustness, thus the bilateral approach to performance index

The demonstration presented here of the design technique appears to warrant the conclusion that the calculation procedure provides a substantial aid to the overall design process With relatively limited effort a feature set with good overall performance was obtained On the other hand, the band set given in Table II should by no means be regarded as worthy of consideration for implementation as is The ensemble definition and sampling were not an important point in this work and would require much more detailed and careful consideration than they received here The same is true of the band set and its comparison to alternative choices

And finally, we again draw attention to the use of the tools of this design procedure for carrying out more basic research into understanding the scene itself and its spectral characteristics By being able to determine and quantitatively assess the information-bearing attributes of scene spectral characteristics, the ability to study potentially important characteristics of the scene is greatly facilitated

References

[1] R B Ash, *Information Theory* New York Wiley, 1967

[2] W B Davenport and W I Root, *An Introduction to the Theory of Random Signals and Noise* New York McGraw-Hill, 1958

[3] R B Crane, W A Malila, and W Richardson, "Suitability of the normal density assumption for processing multispectral scanner data," *IEEE Trans Geosci Electron*, vol GE-10, pp 158-165, 1972

[4] R O Duda and P E Hart, *Pattern Classification and Scene Analysis* New York Wiley, 1973

[5] K Fukunaga, *Introduction to Statistical Pattern Recognition* New York Academic, 1972

[6] J Grad and M A Brebner, "Eigenvalues and eigenvectors of a real general matrix, Algorithm 343," *Association for Computing Machinery, Comm*, vol 11, no 12, pp 820-826, 1968

[7] J L Harnage and D A Landgrebe, "Landsat-D Thematic Mapper Technical Working Group- Final Report," NASA/Johnson Space Center report no JSC-90797, Houston, Texas, 1975

[8] R A Holmes and R B MacDonald, "The physical basis of system design for remote sensing in agriculture," *Proc IEEE*, vol 57, pp 629-639, 1969

[9] R M Hoffer, "Biological and physical considerations in applying computer-aided analysis techniques to remote sensor data," in *Remote Sensing The Quantitative Approach*, P H Swain and S M Davis, eds New York McGraw-Hill, 1978

[10] G F Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans Inform Theory*, vol IT-14, pp 55-63, 1968

[11] D A Landgrebe, "Useful information from multispectral image data another look," in *Remote Sensing The Quantitative Approach*, P H Swain and S M Davis, eds New York McGraw-Hill, 1978

[12] B G Mobasseri, C D McGillem, and P E Anuta, "A parametric multiclass Bayes error estimator for the multispectral scanner spatial model performance evaluation," Ph D dissertation, LARS Technical Rep 061578, Lab for Applications of Remote Sensing, Purdue Univ, West Lafayette, IN, 1978

[13] A Papoulis, *Probability, Random Variables and Stochastic Processes* New York McGraw-Hill, 1965

[14] P J Ready, P A Wintz, S J Whitsitt, and D A Landgrebe, "Effects of compression and random noise on multispectral data," in *Proc of the 7th Int Symp on Remote Sensing of the Environment*, Ann Arbor, MI, 1971

[15] S J Whitsitt and D A Landgrebe, "Error estimation and separability measures in feature selection for multiclass pattern recognition," Ph D dissertation, LARS Inform Note 082377, LARS, Purdue, West Lafayette, IN, 1977

[16] D J Wiersma and D A Landgrebe, "The analytical design of spectral measurements for multispectral remote sensor systems," Ph D dissertation, LARS Tech Rep #122678, Purdue University, West Lafayette, IN, 1979

[17] H L VanTrees, *Detection, Estimation, and Modulation Theory—Part I* New York Wiley, 1968

ORIGINAL PAGE IS
OF POOR QUALITY

END

DATE

FILMED

SEP 15 198