LARS INFORMATION NOTE 062773

CLOSED BOUNDARY FINDING, FEATURE SELECTION AND CLASSIFICATION APPROACH TO MULTI-IMAGE MODELING*

by

J. N. Gupta and P. A. Wintz

Remote sensing is the science and art of acquiring information about material objects from measurements made at a distance without coming into physical contact with the objects. The information may be transmitted to the observer through electromagnetic fields, in particular, through the spectral, spatial, and temporal variations of these fields [1,2]. The data is gathered from sensors mounted on an aerospace platform and provides information on existing conditions of physical objects. The sensors measure target radiance in several regions of the electromagnetic spectrum; the data is usually called multispectral data.

In order to derive information from these data sets (Multi-images), one must be able to relate these measurements to those of known objects or materials. The quantity of data is typically so large that machine oriented analysis techniques must be devised. At the same time to enhance image processing activities, an understanding of the pictorial aspects of the data to be processed is required. Also to further progress in image analysis and image processing techniques such as image coding, registration, deblurring, noise removal, etc., better image models need to be developed.

In this paper a specific multi-image analysis and processing technique is studied to develop a good image model. A boundary finding algorithm is developed involving the image characteristics such as grey level edges and neighboring second order statistical properties of these data sets. The

---

*The research reported in this paper is supported by NASA Grants NGL 15-005-112 and NGR 15-005-152. The authors are with the Laboratory for Application of Remote Sensing and the Department of Electrical Engineering, Purdue University, West Lafayette, Indiana 47907.

technique of boundary finding narrows the method of feature selection to Karhunen-Loeve transformation and the evaluation of the boundaries has been done based on test field classification.

Introduction

Natural scenes usually consist of objects with structures unless the picture contains objects out of focus. Structure is a departure from randomness. This structure is not usually homogeneous—different parts of a picture usually contain different kinds of structure. The study of this structure is called Image Modeling.

One approach to image modeling is to characterize what we see in the image. Generally there exists structure in the relationship between areas—an area can be defined as the largest connected set of picture elements each of which has a property P. P can represent grey level, texture, a statistic etc. Adjacent areas share boundaries so that the shape of each is determined by the shape of its neighbors. For multi-spectral images these "areas" become "volumes". For future reference these will be referred to as "blobs" rather than "areas" or "volumes". Grey level blobs can be characterized in terms of their boundaries which are related to gray level discontinuities.

Most methods for foundary finding reported in the literature are based on extensions of the classical digital gradient and laplacian operators. Several investigators have studied mathematical methods of defining edge detectors that are optimum in various senses. Heuckel [3] finds the perfect step edge that best matches the given digital picture in a certain disc shaped neighborhood of each point. Similar work has been done recently by Griffith [4] to detect edges in simple scenes using a-priori information. Another approach to edge detection involves the use of both coarse and fine difference operators at each point. This approach

is suitable for detecting steps in average grey levels. Coarse operators detect the steps, while the fine ones locate them sharply [5,6]. Some approaches investigated at Purdue University were based on the gradient concept [7], clustering [8] and hypothesis testing based on first order statistics [9]. The gradient approach is inherently noisy and produces borders that are discontinuous of varying width and also produces spurious isolated points. Clustering is more stable and less noisy but very time consuming; and also closed boundaries are not guaranteed. The hypothesis testing technique guarantees closure.

The boundary finding algorithm discussed here is also based on hypothesis testing but takes into account second order statistics as well as first order statistics.

Description of the Data Source

The multispectral source to be studied is shown in Figure 1. The process $a(x_1,x_2,\lambda)$ is some measure of the spectral energy at wavelength λ for ground resolution point (x_1,x_2). The output sample space is the set of all vectors $\underset{\sim}{A}$ in R^N that are obtained from the continuous stochastic process $a(x_1,x_2,\lambda)$ by discretizing the two spatial variables (x_1,x_2) and the spectral variable (λ).

The elements of $\underset{\sim}{A}$ are assumed to be jointly normal. This assumption is made for the mathematical simplification it allows. In addition it has been found experimentally that the normal distribution [10] is a reasonable approximation for multispectral data for several applications in pattern recognition.

The output vector $\underset{\sim}{A}$ is then

$$A_{\ell m} = [a_1, a_2, \ldots, a_N]^t$$

where $a_i = a[x_{1\ell}, x_{2m}, \lambda_i]$ and $x_{1\ell}$, x_{2m} and λ_i are the ℓth, mth, and i-th sample in (x_1, x_2, λ)

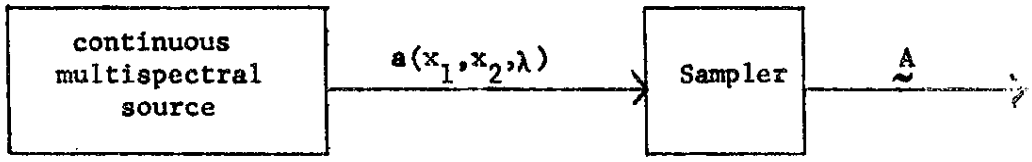


Fig. 1 Model of the Multispectral Source

The Joint density function is given by

$$g(\underset{\sim}{A}) = \frac{1}{(2\pi)^{N/2}} \cdot \frac{1}{|\underset{\sim}{\Sigma}|^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2}(\underset{\sim}{A} - \underset{\sim}{\mu})^t \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{A} - \underset{\sim}{\mu})\right]$$

where $\underset{\sim}{\Sigma}$ is the N x N covariance matrix for $\underset{\sim}{A}$ and $\underset{\sim}{\mu}$ is the mean vector.

In the following discussion each a_i will be called an element and a group of elements will be called simply a "group". The groups are squares containing equal number of elements on each side. The elements of each blob also are assumed to be Gaussian.

Hypothesis Testing

The general problem is this: we have two normal populations. One with variate x_1 which has mean μ_1 and variance σ_1^2 and one with variate x_2 which has mean μ_2 and variance σ_2^2. On the basis of two samples, one from each population, it is desired to test the hypotheses in both of the following:

i, $H_{o1} : \sigma_1^2 = \sigma_2^2$

$H_{a1} : \sigma_1^2 \neq \sigma_2^2$

ii, $H_{02} = \mu_1 = \mu_2,\ \sigma_1^2 = \sigma_2^2 = \sigma^2$

$H_{a2} = \mu_1 \neq \mu_2,\ \sigma_1^2 = \sigma_2^2 = \sigma^2$

In hypothesis i, the parameter space Ω is four dimensional; a joint distribution of x_1 and x_2 is specified when values are assigned to the four quantities $(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)$. The subspace ω is three dimensional because values for only three quantities (μ_1,μ_2,σ^2) need be specified in order to specify completely the joint distribution under the hypothesis that $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Let there be m observations $(X_{11,\lambda_\rho}, X_{12,\lambda_\rho}, \ldots, X_{1m,\lambda_\rho})$ in the sample from the first population and n observations $(X_{21,\lambda_\rho}, X_{22,\lambda_\rho}, \ldots, X_{2n,\lambda_\rho})$ from the second. Let us also assume for this analysis that λ_ρ is fixed and determines the spectral band from which the samples are being considered.

The likelihoods in parameter space Ω and subspace ω are:

$$L(\Omega) = \left(\frac{1}{2\pi\sigma_1^2}\right)^{m/2} \exp\left[-\frac{1}{2\sigma_1^2}\sum_{i=1}^{m}\left(X_{1i,\lambda_\rho}-\mu_1\right)^2\right]\left(\frac{1}{2\pi\sigma_2^2}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_2^2}\sum_{j=1}^{n}\left(X_{2j,\lambda_\rho}-\mu_2\right)^2\right]$$

and

$$L(\omega) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{m+n}{2}} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{m}\left(X_{1i,\lambda_\rho}-\mu_1\right)^2 + \sum_{j=1}^{n}\left(X_{2j,\lambda_\rho}-\mu_2\right)^2\right\}\right]$$

Differentiating $L(\Omega)$ w. r. t. μ_1, μ_2, σ_1^2, σ_2^2 the estimates of μ_1, μ_2, σ_2^2 respectively are as:

$$\hat{\mu}_1 = \frac{1}{m}\sum_{i=1}^{m} X_{1i,\lambda_\rho} = \bar{X}_{1,\lambda_\rho}$$

$$\hat{\mu}_2 = \frac{1}{n}\sum_{j=1}^{n} X_{2j,\lambda_\rho} = \bar{X}_{2,\lambda_\rho}$$

$$\hat{\sigma}_1^2 = \frac{1}{m} \sum_{i=1}^{m} \left(X_{1i,\lambda_\rho} - \bar{X}_{1,\lambda_\rho} \right)^2$$

$$\hat{\sigma}_2^2 = \frac{1}{n} \sum_{j=1}^{n} \left(X_{2j,\lambda_\rho} - \bar{X}_{2,\lambda_\rho} \right)^2$$

Substituting these values the maximum of $L(\Omega)$ is given by

$$L(\hat{\Omega}) = \left[\frac{m}{2\pi \sum_{i=1}^{m} \left(X_{1i,\lambda_\rho} - \bar{X}_{1,\lambda_\rho} \right)^2} \right]^{\frac{m}{2}} \exp\left[-\frac{m}{2}\right] \left[\frac{n}{2\pi \sum_{j=1}^{n} \left(X_{2j,\lambda_\rho} - \bar{X}_{2,\lambda_\rho} \right)^2} \right]^{\frac{n}{2}} \exp\left[-\frac{n}{2}\right]$$

Similarly the maximum of $L(\omega)$ can be derived. The maximum likelihood Ratio is then

$$R_{\lambda_\rho} = R\left(X_{11,\lambda_\rho}, X_{12,\lambda_\rho}, \cdots, X_{1m,\lambda_\rho}, X_{21,\lambda_\rho}, \cdots, X_{2n,\lambda_\rho} \right)$$

$$= \frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

$$= \frac{(m+n)^{\frac{m+n}{2}}}{m^{\frac{m}{2}} n^{\frac{n}{2}}} \cdot \frac{\sum_{i=1}^{m} \left(X_{1i,\lambda_\rho} - \bar{X}_{1,\lambda_\rho} \right)^2}{\sum_{j=1}^{n} \left(X_{2j,\lambda_\rho} - \bar{X}_{2,\lambda_\rho} \right)^2}}{\left[1 + \frac{\sum_{i=1}^{m} \left(X_{1i,\lambda_\rho} - \bar{X}_{1,\lambda_\rho} \right)^2}{\sum_{j=1}^{n} \left(X_{2j,\lambda_\rho} - \bar{X}_{2,\lambda_\rho} \right)^2} \right]^{\frac{m+n}{2}}}$$

Defining $F_{\lambda_\rho} = \frac{n-1}{m-1} \frac{\sum_{i=1}^{m}\left(X_{1i,\lambda_\rho} - \bar{X}_{1,\lambda_\rho}\right)^2}{\sum_{j=1}^{n}\left(X_{2j,\lambda_\rho} - \bar{X}_{2,\lambda_\rho}\right)^2}$ one obtains

$$R_{\lambda_\rho} = \frac{(m+n)^{\frac{m+n}{2}}}{m^{\frac{m}{2}}\, n^{\frac{n}{2}}} \frac{\left(\frac{m-1}{n-1} F_{\lambda_\rho}\right)^{\frac{m}{2}}}{\left(1+\frac{m-1}{n-1} F_{\lambda_\rho}\right)^{\frac{m+n}{2}}}$$

So if the distribution of F_{λ_ρ} is known, the distribution of R_{λ_ρ} is also known. Since F_{λ_ρ} has F-distribution with (m-1) and (n-1) degrees of freedom, therefore on plotting R_{λ_ρ} as a function of F_{λ_ρ} it is apparent that the critical region $0 < R_{\lambda_\rho} < C$ corresponds to a two tailed test on F.

In the second hypothesis the parameter space Ω becomes three dimensional with co-ordinates (μ_1, μ_2, σ^2), while, ω for the null hypothesis $\mu_1 = \mu_2 = \mu$ (say) becomes two dimensional with co-ordinates (μ, σ^2).

The maximum likelihood ratio in this case is

$$R_{\lambda_\rho} = \left[1 + \frac{\frac{mn}{m+n}\left(\bar{X}_{1,\lambda_\rho} - \bar{X}_{2,\lambda_\rho}\right)^2}{\sum_{i=1}^{m}\left(X_{1i,\lambda_\rho} - \bar{X}_{1,\lambda_\rho}\right)^2 + \sum_{j=1}^{n}\left(X_{2j,\lambda_\rho} - \bar{X}_{2,\lambda_\rho}\right)^2}\right]^{-\frac{m+n}{2}}$$

Again defining a parameter

$$t_{\lambda_\rho} = \frac{\sqrt{\frac{mn}{m+n}}\left(\bar{X}_{1,\lambda_\rho} - \bar{X}_{2,\lambda_\rho}\right)\sqrt{m+n-2}}{\sqrt{\sum_{i=1}^{m}\left(X_{1i,\lambda_\rho} - \bar{X}_{1,\lambda_\rho}\right)^2 + \sum_{j=1}^{n}\left(X_{2j,\lambda_\rho} - \bar{X}_{2,\lambda_\rho}\right)^2}}$$ which

has a student t-distribution with m + n - 2 degrees of freedom, the likelihood ratio is

$$R_{\lambda_\rho} = \left[\frac{1}{1+\frac{t^2_{\lambda_\rho}}{m+n-2}} \right]^{\frac{m+n}{2}}$$

Its distribution is determined by the t-distribution. The test is done in terms of t_{λ_ρ} rather than R_{λ_ρ}. This is also a two-tailed test. A 5 percent critical region for t_{λ_ρ} is $t^2_{\lambda_\rho} > (t_{0.25})^2$.

In short the first hypothesis is based on F-test and the second on student t-test.

Testing Algorithm

Let the group to be compared be called sample 1 and the group to be compared with it sample 2. The goal is to calculate the value of F_{λ_ρ} to compare against a critical value to determine whether both of the samples have the same variance irrespective of their means and then finally to calculate the value of t_{λ_ρ} to compare against another specified critical value to determine whether both of the samples belong to the same population.

The boundary is assumed to exist if the null hypothesis H_{01} fails for the critical value already specified. For this purpose, standard tables [11,12] have been used. The boundaries may be weak in some spectral bands and strong in others. To take this into account the first hypothesis is tested for all values of λ_ρ, $\rho = 1,\ldots,$ number of spectral bands considered. After the null hypothesis H_{01} is accepted in all the spectral bands, the algorithm proceeds to test for the second hypothesis. Again the failure

of the null hypothesis H_{02} for any value of λ_ρ indicates that the two populations are not the same and hence a boundary exists between them.

A block diagram of testing algorithm is shown in Figure 2. In this block diagram, λ_ρ is not fixed anymore, while calculating the covariance matrix.

Multivariate tests can take all the spectral bands into account simultaneously but they are very time consuming. To cut down on processing time the algorithm determines the boundaries in all the spectral bands separately. Also, the experiments conducted show that for the multispectral data available at the Laboratory for Application of Remote Sensing (LARS), multispectral tests are not very advantageous.

Building of Blobs

Let us suppose g columns and g rows of digital data are from the same distribution. The g rows of the data (one row of groups) are processed at a time. Let there be W groups in a row and k'th group of the previous row is designated as group k_ρ. Let the blob to which group k_c is assigned be given by b_{k_c} and the blob to which group k_ρ is assigned b_{k_ρ}.

In the beginning of the data to be processed group 1_c is arbitrarily assigned to blob 1_c. Group 2_c is compared with group 1_c according to the Testing Algorithm. If they have the same distribution, their statistics are added and both groups 1_c and 2_c are assumed to be in the same blob. Next group 3_c is compared with group 2_c and in general the group W is compared with group W-1 at the end of the row. At every step an account is kept for the number of groups in all the blobs.

From second row of groups onward the following steps are followed:

Group 1_c is compared to the blob containing group 1_ρ. If they are the same then this group is also added to b_1 otherwise processing continues

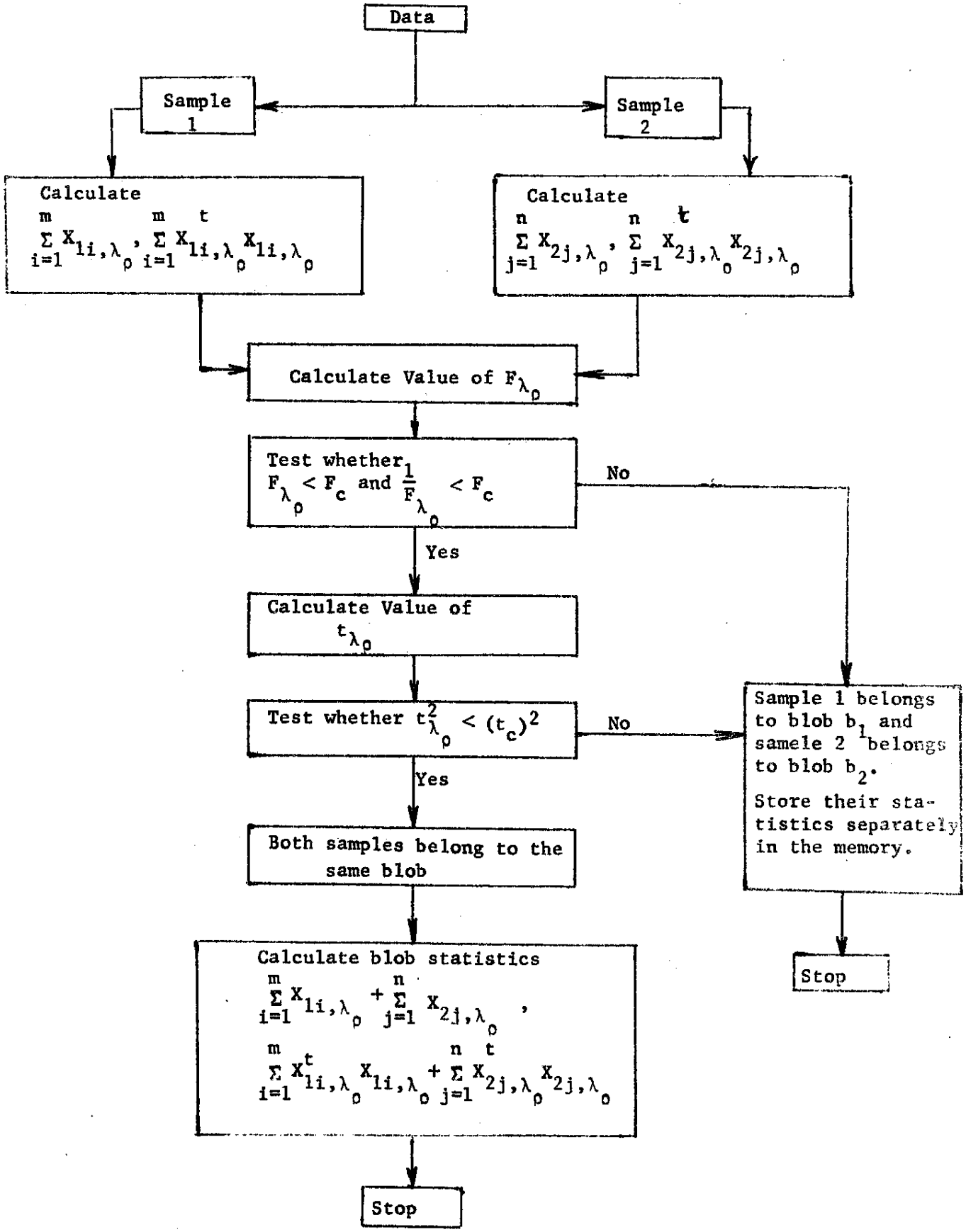


Fig. 2 Testing Algorithm

to group 2_c. Group k_c is compared to the blob which contains group k_ρ. If they are the same, group k_c is added to b_{k_ρ} and a check is made if group k_c-1 is assigned to any blob. If it is then it continues to k_c+1 otherwise group k_c-1 is tested against blob b_{k_ρ}. If they are different then the processing continues to k_c+1. Here again if group k_c-1 is the same it is added to blob b_{k_ρ} and a check is made on group k_c-2 and so on.

On the other hand if group k_c and blob b_{k_ρ} are different then group k_c is compared to group k_c-1. If they are the same groups k_c and k_c-1 are added otherwise the processing continues to group k_c+1. In the end of the second row group W_c is processed in the same manner as group k_c except that there is no other group k_c+1 to continue with.

After group W_c is processed a backward scanning is done on this row. The scanning starts with group W_c. Whenever the group J_c is encountered which has not been assigned to any blob, group J_c is assigned to a new blob and looks for group J_c-1. If group J_c-1 is assigned already the scanning continues further otherwise group J_c-1 is compared to blob b_{J_c}. If they are the same the group J_c-1 is added to b_{J_c} and a check is made whether group J_c-2 is assigned or not and so forth.

As the blob building continues to the subsequent group rows and it is determined that no group in the current row belongs to the blob, the blob is said to be closed and it is assigned to one of the pattern classes based on pattern classification techniques.

Fidelity of Reconstruction

When the blob is determined to be closed the original data elements of this blob are replaced by the mean value averaged over all the original elements of this blob and an estimate of the percent mean square error is calculated.

If $\underset{\sim}{A}$ is the original data and $\hat{\underset{\sim}{A}}$ is the reconstructed data after replacing the blobs by their mean values, then the mean square error is defined as

$$MSE = E\left[\|\underset{\sim}{A} - \hat{\underset{\sim}{A}}\|\right]^2$$

Fidelity of reconstruction can be measured by the MSE between $\underset{\sim}{A}$ and $\hat{\underset{\sim}{A}}$ or by % MSE w. r. t. the variance of the data.

If σ_i^2 is the data variance in the i-th channel and there are ρ-channels in all considered for processing purposes then

$$\% \text{ MSE} = \frac{E\left\{\|\underset{\sim}{A} - \hat{\underset{\sim}{A}}\|^2\right\}}{\sum_{i=1}^{\rho} \sigma_i^2} * 100$$

% MSE has been calculated for various significance levels of student 't' and F-tests and a plot is given in Fig. 3. Pictures using digital display system taken on the original data and the reconstructed data are also given in Fig. 4.

Classification of Data

One immediate advantage of finding blob boundaries is in classification of the data in already known classes.

Presently the classification is usually done on the basis of a per point classification (spectral information only) and the advantage of spatial information is not taken into account. To use the spatial information it becomes absolutely necessary to define spatial boundaries which separate the regions containing the classes of interest. The boundary finding algorithm provides such a means.

Let $\{W_i\}$ be the set of pattern classes with their distributions $\{F_i\}$,

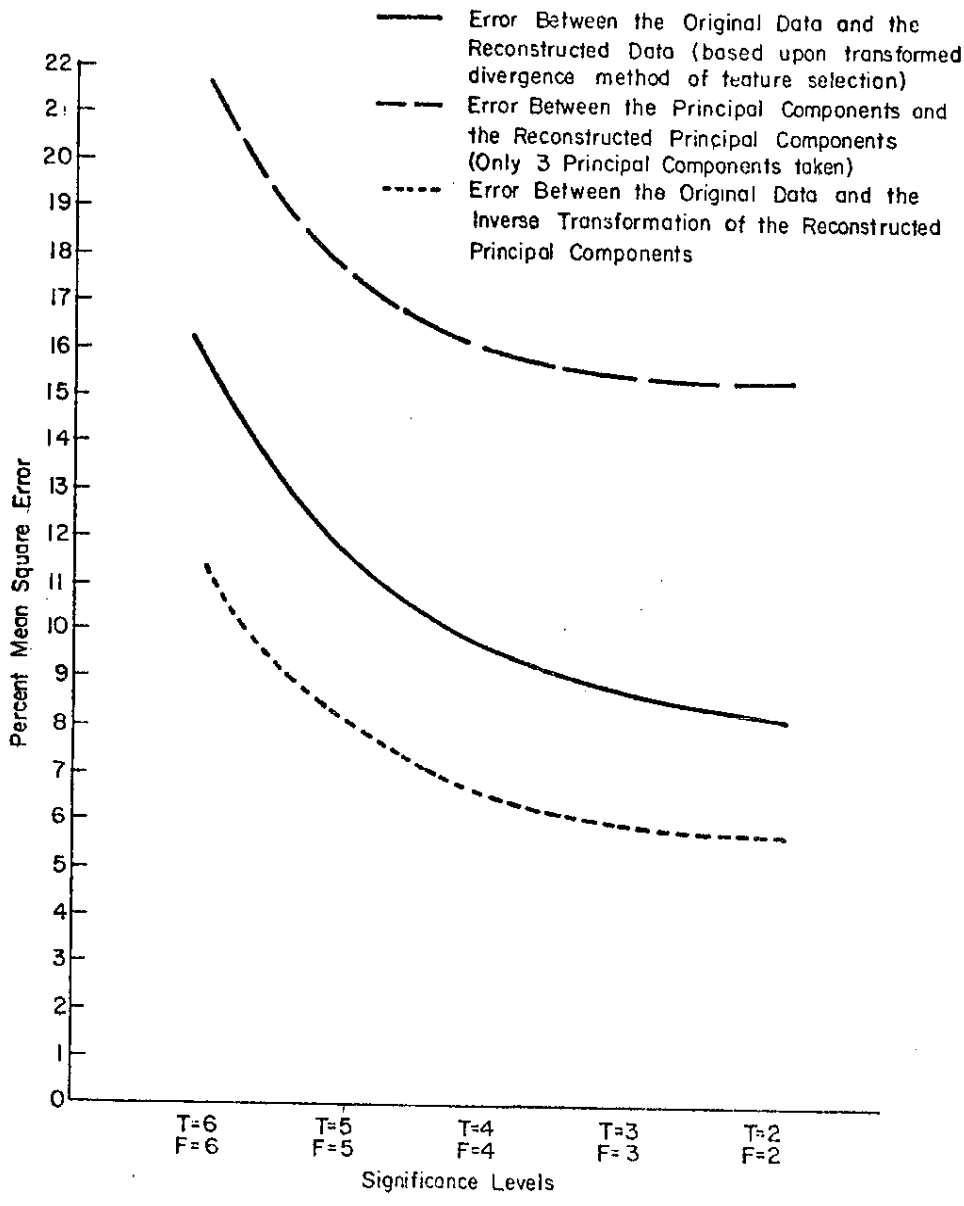
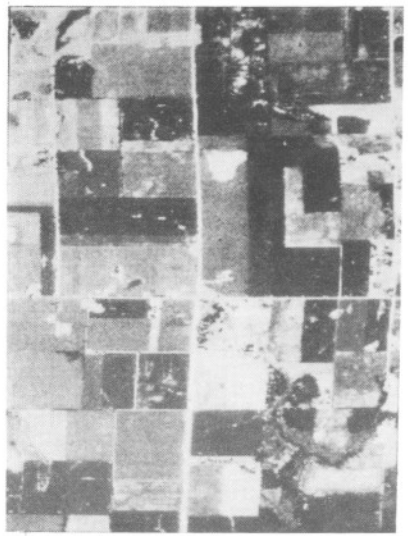
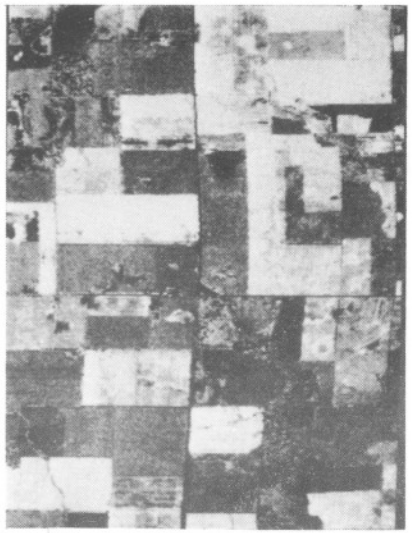

Error Between the Original Data and the Reconstructed Data (based upon transformed divergence method of feature selection)
Error Between the Principal Components and the Reconstructed Principal Components (Only 3 Principal Components taken)
Error Between the Original Data and the Inverse Transformation of the Reconstructed Principal Components
Percent Mean Square Error
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
T=6 F=6
T=5 F=5
T=4 F=4
T=3 F=3
T=2 F=2
Significance Levels

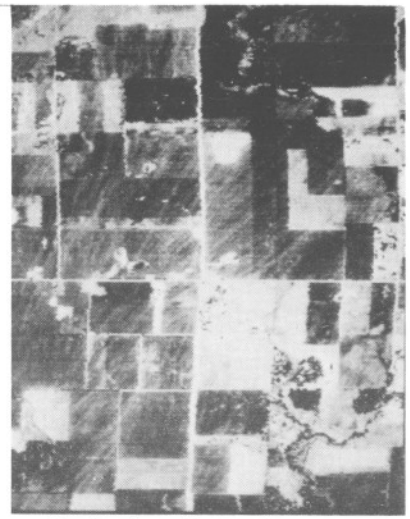
Figure 3.



Original Data
from Channel 7



Original Data
from Channel 8



Original Data
from Channel 12

Reconstructed Data
from Channel 7

Reconstructed Data
from Channel 8

Reconstructed Data
from Channel 12

Figure 4

and let the set of points $\{y\}$ be all the elements of the closed blob. To decide to which F_i they belong, a distance measure $d(F_i, F_j)$ is used as a measure of separation between any two distributions F_i and F_j.

In a mathematical sense the term "distance" and "metric" are sometimes (incorrectly) used interchangeably. A metric on a set S is, of course a real valued δ defined on S x S (x indicates cartesian product) such that for arbitrary F_i, F_j, F_k in S

a, $\delta(F_i, F_j) \geq 0$

b, $\delta(F_i, F_j) = \delta(F_j, F_i)$

c,i, $\delta(F_i, F_i) = 0$

ii, If $\delta(F_i, F_j) = 0$ then $F_i = F_j$

d, $\delta(F_i, F_j) + \delta(F_j, F_k) \geq \delta(F_i, F_k)$

The distance measure in this report should not be confused with metric. It obeys properties a, b, and c, only; not property d.

Let G(y) be a distribution governing $\{y\}$. The magnitudes of $\{d(G, F_i)\}$ are compared; the distribution F_i which minimizes the distance $d(G, F_i)$ is presumed to be the distribution which contains G(y).

In other words assign $\{y\}$ to the ith class in case

$$d(G, F_i) = \min_{j=1, \ldots, k} d(G, F_j)$$

where k is the total number of classes considered.

Let $F_i(y)$ and $F_j(y)$ be the distributions of pattern classes W_i and W_j respectively, and let $p_i(y)$ and $p_j(y)$ are their probability density functions defined in Ω_y. Then a statistical distance measure appearing in the work of Jeffreys Matusita [13,14] is defined for two distributions $F_i(y)$ and $F_j(y)$ as follows.

$$B = d^2(F_i, F_j) = \int_{\Omega} \left(\sqrt{\rho_i}(y) - \sqrt{\rho_j}(y) \right)^2 dy$$

For convenience, B will be referred to herein as the Bhattacharyya distance. Defining

$$P = \int_{\Omega} \sqrt{\rho_i}(y) \cdot \sqrt{\rho_j}(y) \; dy$$

$$d^2(F_i, F_j) = 2(1-P)$$

In multispectral data one is concerned with multivariate distributions and as already stated that the blobs have multivariate Gaussian distributions, ρ for two such densities $\{\rho_i(y) = N(\mu_i, \Sigma_i), i=1,2\}$ can be reduced to the following expressions [15]

$$P = \exp[-\sigma]$$

where

$$\sigma = \frac{1}{8} (\mu_1-\mu_2)^t \left[\frac{\Sigma_1+\Sigma_2}{2} \right]^{-1} (\mu_1+\mu_2) + \frac{1}{2} \ln \frac{\det\left(\frac{1}{2}(\Sigma_1+\Sigma_2)\right)}{\sqrt{\det(\Sigma_1)\det(\Sigma_2)}}$$

Then for whichever class P is maximum, the vector $\{y\}$ is decided to be a member of that class.

The multispectral data used for classification is taken from 1971 Corn Blight Watch Experiment [16]. It has eleven spectral bands in the wavelength region of 0.4 to 3 micrometers and one thermal infrared band from 9.7 to 11.3 micrometers. The strip of terrain observed is represented by 222 samples across track and 176 samples per mile along track taken at a 5000 foot altitude.

A set of training fields was obtained representing 5 classes namely corn, forage, soybeans, forest and water. On the basis of these 5 pattern classes, all the 222 samples across track and 400 samples along the track

were classified using the boundary finding algorithm defining blob bound and the above mentioned distance measure.

Note that σ in B-distance involves the inversion of a covariance matrix and for some of the blobs the determinant of the covariance matrix is very insignificant. Such blobs are classified according to the Euclidean distance to the mean of the pattern classes.

The classified data is written on a separate data storage tape and is evaluated on the basis of given test fields.

Presently the same data at LARS is being classified on the point classification basis using Gaussian Maximum likelihood Ratio decision rule [1] and is getting 95.9% overall classification accuracy as compared to 98% obtained by this algorithm. A summary of the results is presented in Figure 5.

Feature Selection

The number of spectral bands (channels) poses a problem in the sense that the computation time and computer memory requirements are increased as the number of channels increases. Also, to minimize the overall probability of misclassification the best set of features should be selected.

The results obtained in the previous section were obtained by using a subset chosen on the basis of minimum transformed divergence. The minimum transformed divergence between any two training classes was larger for this particular set of channels than all other combination of three channels. The transformed divergence is a measure of class separability in feature space [17].

Per point classification results (Based on test fields)

Channels (7,8,12)

Class	No. of Points	Percent Correct	Classified as Corn	Forage	Soybeans	Forest	Water
Corn	5950	95.6	5686	203	38	23	0
Forage	3151	97.3	78	3065	0	8	0
Soybeans	4770	96.0	59	95	4578	38	0
Forest	680	92.6	3	13	32	630	2
Water	37	97.3	0	1	0	0	36

Overall performance (13995/14588) = 95.9% correct

Classification results obtained by Boundary Finding Algorithm

Channels (7,8,12)

Class	No. of Points	Percent Correct	Classified as Corn	Forage	Soybeans	Forest	Water
Corn	5950	99.0	5888	50	4	8	0
Forage	3151	96.1	111	3027	4	9	0
Soybeans	4770	97.9	63	30	4669	8	0
Forest	680	99.1	2	4	0	674	0
Water	37	97.3	0	1	0	0	36

Overall performance (14294/14588) = 98% correct

Figure 5 Classification Results Using 3 Best Features Selected according to the Minimum Transformed Divergence Method of Feature Selection

The boundary finding algorithm is based on variance analysis therefore the important consideration in selecting such a feature set should be such that it contains the maximum variance in it.

To overcome these problems, the Karhunen-Loeve (K-L transformation technique of selecting the best feature can be implemented.

K-L or Eigenvector transformation is an NxN transformation matrix $\underset{\sim}{T}$ (reconstruction transformation matrix $\underset{\sim}{T}^t$) and is used to map $\underset{\sim}{A}$ down to n<N dimensions. It is shown in [18] than n rows of $\underset{\sim}{T}$ are the first n orthonormal solutions to the characteristic equation

$$\underset{\sim}{\Sigma}\,\underset{\sim}{t} = \lambda\,\underset{\sim}{t}$$

corresponding to the n largest eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_n$ of the NxN covariance matrix $\underset{\sim}{\Sigma}$ of $\underset{\sim}{A}$. The transformation to n<N dimensions is given by

$$\underset{\sim}{y} = \underset{\sim}{T}\,\underset{\sim}{A}$$

where $\underset{\sim}{T} = [\underset{\sim}{t}_1, \underset{\sim}{t}_2, \ldots, \underset{\sim}{t}_n]$

It is also shown in [18] that the elements of $\underset{\sim}{Y}$ are uncorrelated and the variance of the n elements of $\underset{\sim}{Y}$ are given by the n λ_i's, that is

$$\text{Var}\ (Y_1) = \lambda_1$$

$$\text{Var}\ (Y_2) = \lambda_2$$

$$\vdots$$

$$\text{Var}\ (Y_n) = \lambda_n$$

In other words this transformation rotates the source output $\underset{\sim}{A}$ in the N-space to a favorable orientation with respect to the coordinate systems. The favorable orientation is so-called because the average energy (variance) of the source is re-distributed over the coordinates such that a larger

percentage of the source variance is distributed over fewer coordinates. This packing of the source variance provides a means of reducing the co ordinates to determine blob boundaries and implement sample classification.

The 12 x 12 spectral covariance matrix $\underset{\sim}{\Sigma}$ was estimated for a Corn Blight Watch Experiment flight by averaging over the entire flight line. The number of data points available for processing were 2.5×10^5. The resulting covariance matrix is shown in Figure 6. The eigenvectors and normalized eigenvectors for $\underset{\sim}{\Sigma}$ were computed using the IBM Scientific Subroutine Package subroutine "EIGEN". Figure 7 lists the eigenvalues ranked such that

$$\lambda_1 > \lambda_2 > \cdots\cdots > \lambda_n$$

The eigenvector transformation $\underset{\sim}{T}$ is constructed from the first 3 eigenvectors of $\underset{\sim}{\Sigma}$. The first Principal components contain 93.9% of the total variance contained in all the twelve features. The blob boundaries were located in the transformed data and then the blobs were classified using already mentioned pattern classification techniques. The overall classification accuracy is 98.2% as compared to 98% using minimum transformed divergence method of feature selection. The blobs data was replaced by their mean values and then an inverse transformation operation was performed on this data. An estimate of mean square error was determined between the reconstructed Principal components and original Principal components and also between the inverse transformed reconstructed Principal components and the original data. These values are also plotted in Figure 3. The digital display pictures of first three principal components and the reconstructed Principal components are given in Figure 8. The summary of the classification results is detailed in Figure 9.

CHANNEL	1	2	3	4	5	6	7	8	9	10	11	12
1	434.3											
2	338.9	299.9										
3	320.4	265.0	266.1									
4	391.7	325.0	327.9	487.2								
5	447.5	375.1	366.2	491.2	552.3							
6	466.6	411.8	370.1	497.0	559.3	678.9						
7	539.7	459.8	421.3	528.9	622.3	688.4	791.7					
8	-71.6	-59.8	-27.7	21.8	-37.7	-112.9	-147.7	382.1				
9	-27.0	-22.3	-3.1	37.3	1.0	-38.6	-53.5	257.0	231.3			
10	295.7	250.1	249.7	370.7	385.0	406.2	454.5	159.2	203.2	694.5		
11	337.3	289.7	270.1	376.2	408.4	465.0	513.9	66.5	136.6	602.1	651.1	
12	441.0	374.2	337.1	430.8	516.3	631.4	707.8	-294.3	-88.8	527.2	590.0	1062.3

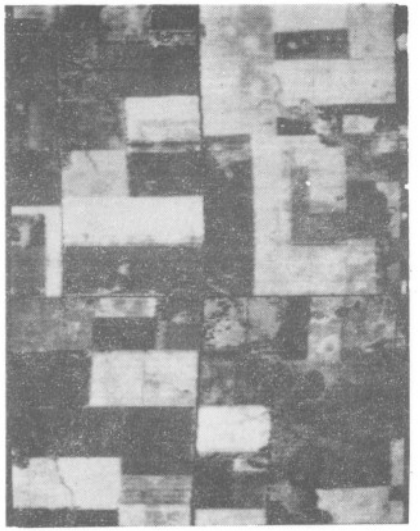
The Total Variance is 6531.66

Fig. 6. Spectral Covariance Matrix for Data for Corn Blight Watch Experiment Flight

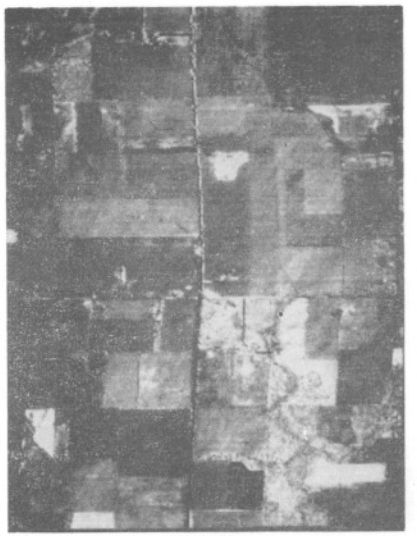
Eigenvalue	Percent of Variance	Cumulative percent
4640.70	71.05	71.05
940.85	14.40	85.45
554.47	8.49	93.94
112.28	1.72	95.66
79.93	1.22	96.89
61.00	0.93	97.82
49.26	0.75	98.57
29.72	0.45	99.03
26.51	0.41	99.43
15.79	0.24	99.68
12.42	0.19	99.87
8.77	0.13	100.00

Fig. 7. 12 Spectral Eigenvalues for CBWE Flight

1st Principal Component

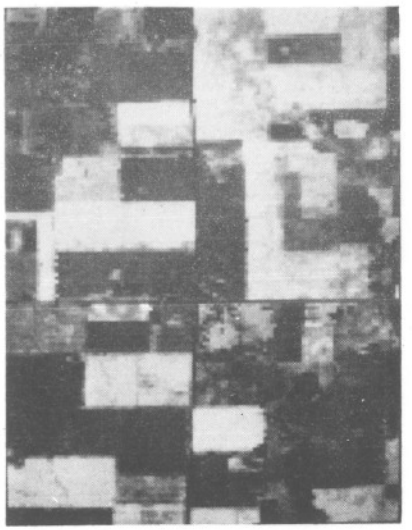


2nd Principal Component

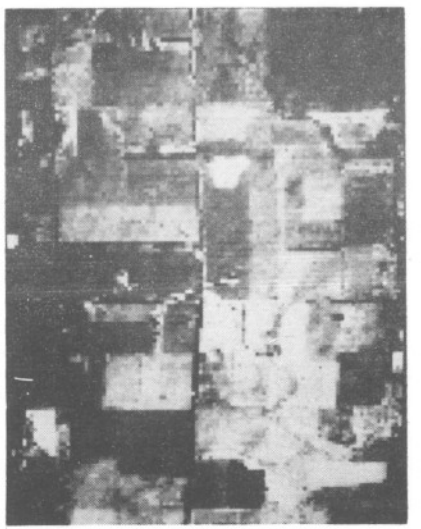


3rd Principal Component

1st Reconstructed Principal Component



2nd Reconstructed Principal Component



3rd Reconstructed Principal Component

Figure 8

Figure 10 shows the block diagram of the different operations involved to calculate the mean square error between the inverse transformed reconstructed Principal components and the original data.

The boundaries of the classified blobs were printed on a line printer of IBM 360/67 and is presented in Figure 11. An aerial photograph related to the ground truth is also presented for comparison purposes.

Class	No. Points	Percent Correct	Classified as Corn	Forage	Soybeans	Forest	Water
Corn	5950	99.7	5932	14	4	0	0
Forage	3151	94.6	160	2980	2	9	0
Soybeans	4770	98.7	21	40	4709	0	0
Forest	680	98.2	12	0	0	668	0
Water	37	97.3	0	1	0	0	36

Overall performance (14325/14588) = 98.2% correct

Fig. 9 Classification Results Based on First Three Principal Components

Summary

In this paper it has been shown that spatial information can be extracted from a multispectral data image by machine and effectively used in the classification process. The degree of improvement in classification accuracy depends upon the amount of spatial information contained in the data. An experiment was conducted on data from the Earth Resources Technology Satellite (ERTS).

For ERTS data there are relatively few elements per blob. Since there is less spatial information to take advantage of, the boundary finder provides only a small degree of improvement over the per point classifier as illustrated by figure 12.

The concept of boundary finding seems to be a fairly robust one. For the aircraft data [16] any three spectral bands were chosen at random and the boundary finder in conjunction with the sample classifier was implemented to evaluate the classification results. These results are presented in Figure 13 and 14. It is evident that though this procedure involves more complex computations but it produces consistently better results than the per point classifier.

The variance analysis used in boundary finding in a way restricts our choice to the Karhunen Loeve Transformation technique of feature selection. From figures 5 and 9 we note that the Maximum likelihood Decision Rule classifies the data with an error rate of 4.1%. The boundary finder in conjunction with Minimum Transformed Divergence method of feature selection reduces the rate of 2% while the boundary finder implemented on the transformed data according to k-L Transformation reduces this rate even further to 1.8%.

The identification of boundaries is important for other reasons also. If the boundaries are easily defined and show a repetitive pattern, the possibility of data compression exists (19). This will help in storing the data without losing much of information by storing the field boundaries and their classification. A measure of mean square error was estimated and it is found in Figure 3 that the k-L Transformation technique of feature selection reduces the percent mean square error quite significantly over the Minimum Transformed Divergence method of feature selection.

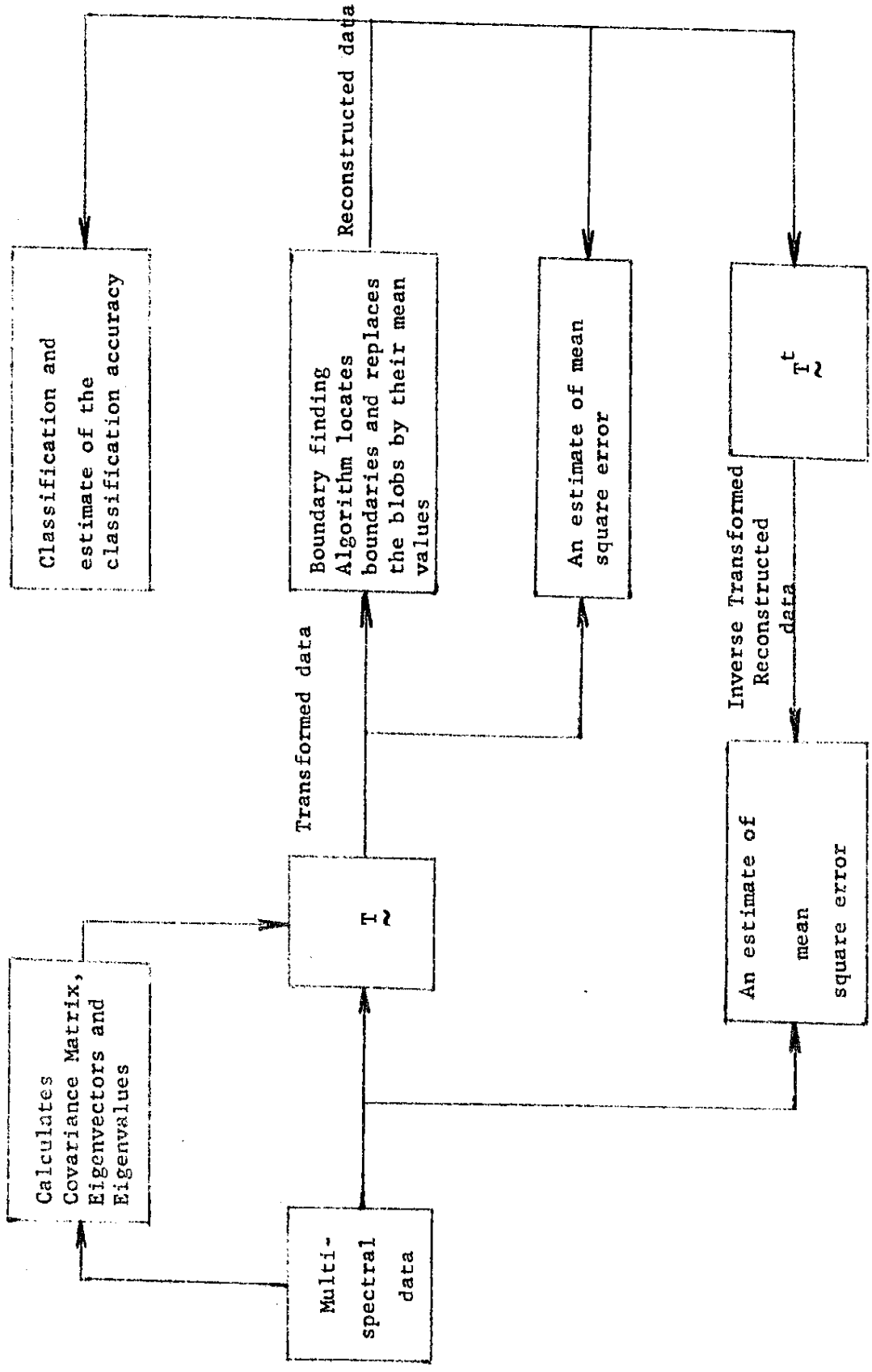
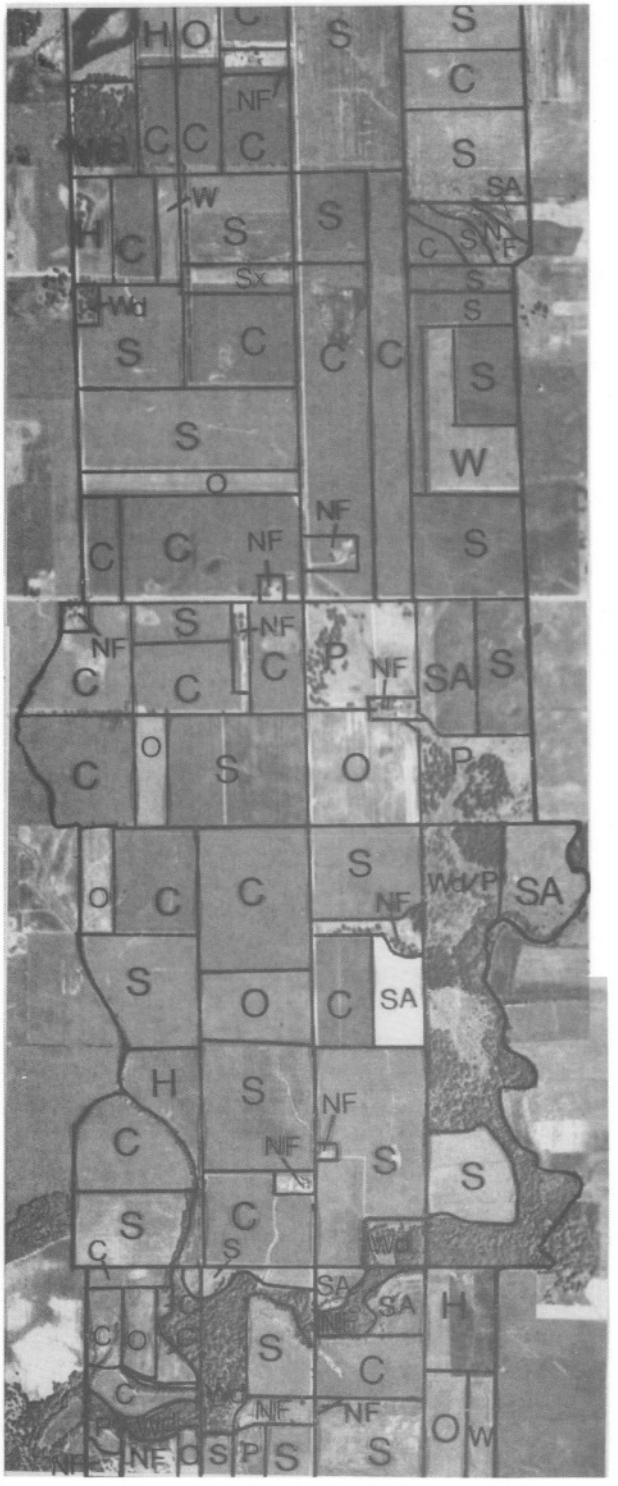
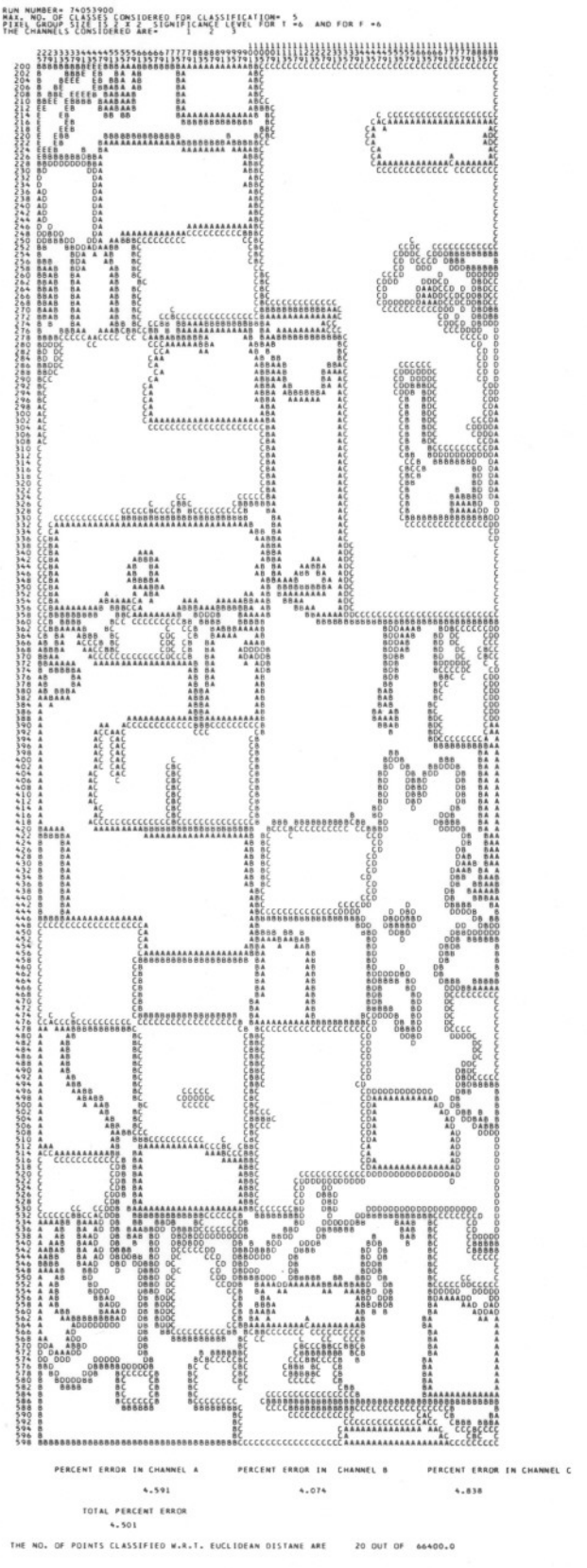


Fig. 10. Different Steps Involved in the Classification and the Estimate of the Fidelity Criterion



Aerial photograph and ground observations for agricultural area (Seg 210 - 1971 CBWE)



On-line printer results after boundary finding and classification of the data

Figure 11

Per point classification results

(0.5-0.6, 0.6-0.7, 0.7-0.8 Micrometer bands)

Class	No. of Points	Percent Correct	Classified as Corn	Soybeans	Others
Corn	3315	88.7	2941	128	246
Soybeans	593	83.8	28	497	68
Others	680	41.2	230	170	280

Overall performance (3718/4588) = 81.0 correct

Classification results based on boundary finding algorithms

(0.5-0.6, 0.6-0.7, 0.7-0.8 Micrometer bands)

Class	No. of Points	Percent Correct	Classified as Corn	Soybeans	Others
Corn	3315	91.2	3023	128	164
Soybeans	593	83.6	40	496	57
Others	680	35.4	285	154	241

Overall performance (3760/4588) = 82% correct

Figure 12 Classification Results of the ERTS Data

Per point classification results (Based on test fields)

Channels (1,2,3)

Class	No. of Samples	Percent Correct	Classified as Corn	Forage	Soybeans	Forest	Water
Corn	5950	60.1	3577	654	762	0	957
Forage	3151	69.7	570	2195	319	1	66
Soybeans	4770	63.5	635	112	3028	648	347
Forest	680	65.0	30	1	181	442	26
Water	37	94.6	0	1	1	0	35

Overall performance (9277/14588) = 63.6% correct

Classification results based on Boundary finding algorithms

Channels (1,2,3)

Class	No. of Samples	Percent Correct	Classified as Corn	Forage	Soybeans	Forest	Water
Corn	5950	68.6	4078	1740	36	4	92
Forage	3151	78.4	639	2471	41	0	0
Soybeans	4766	65.6	1556	76	3132	0	0
Forest	680	42.4	139	0	253	288	0
Water	37	97.3	0	1	0	0	36

Overall Performance (1005/14588) = 68.6% correct

Figure 13 Classification results using 3 original spectral bands of the aircraft data.

Per point classification results (Based on test fields)
Channels (5,7,9)

Class	No. Points	Percent Correct	Classified as Corn	Forage	Soybeans	Forest	Water
Corn	5950	79.0	4700	1170	2	71	0
Forage	3151	87.1	403	2744	0	4	0
Soybeans	4770	90.7	8	157	4325	280	0
Forest	680	96 5	4	6	8	656	6
Water	37	100	0	0	0	0	37

Overall performance = (12462/14588) = 85.4%

Classification results based on Boundary finding algorithm
Channels (5,7,9)

Class	No. Points	Percent Correct	Classified as Corn	Forage	Soybeans	Forest	Water
Corn	5950	97.1	5775	167	0	8	0
Forage	3151	80.7	607	2542	0	2	0
Soybeans	4770	94 7	23	44	4516	187	0
Forest	680	99.4	4	0	0	676	0
Water	37	97.3	0	1	0	0	36

Overall performance = (13545/14588) = 92.9%

Figure 14 Classification results using 3 original spectral bands of the aircraft ~~data~~ data

REFERENCES

[1] K. S. Fu, D. A. Landgrebe, and T. L. Phillips, "Information Processing of Remotely Sensed Agricultural Data," Proc. IEEE, Vol. 57, No. 4, April 1969.

[2] Proceedings of the Seventh International Symposium on Remote Sensing of Environment, University of Michigan, May 1971.

[3] M. Huecksel, "An Operator Which Locates Edges in Digitized Pictures", J. ACM 18, pp. 113-125, January 1971.

[4] A. K. Griffith, "Edge Detection in Simple Scenes Using A Priori Information", IEEE Trans. on Computers, Vol. C-22, No. 4, April 1973.

[5] A. Rosenfeld, and M. Thurston, "Edge and Curve Detection for Visual Scene Analysis," IEEE Trans. Computers, Vol. C-20, pp. 562-569, May 1971.

[6] A. Rosenfeld, M. Thurston, and Y. H. Lee, "Edge and Curve Detection: Further Experiments," IEEE Trans. Computers, Vol. C-21, No. 7, pp. 677-715, July 1972.

[7] P. E. Anuta, "Spatial Registration of Multispectral and Multitemporal Imagery Using Fast Fourier Transform Techniques," IEEE Trans. on Geoscience Electronics, Vol. GE-8, No. 4, Oct. 1970.

[8] A. G. Wacker, and D. A. Landgrebe, "Boundaries in Multispectral Imagery by Clustering", 1970 IEEE Symposium on Adaptive Processes (9th), University of Texas at Austin, December 1970.

[9] P. E. Anuta, E. M. Rodd, R. E. Jensen, and P. R. Tobias, "Final Report for the LARS/Purdue - IBM Houston Scientific Center Joint Study Program," Purdue University, Lafayette, Indiana.

[10] K. S. Fu and P. J. Min, "On Feature Selection in Multiclass Pattern Recognition," Technical Report TR-EE 68-17, Purdue University, Lafayette, Indiana July 1968.

[11] M. Messington, and C. M. Thompson, "Tables of Percentage Points of the Inverted Beta (F) Distribution," Biometrika, 33, pp. 73, 1943.

[12] B. Ostle, "Statistics in Research," Iowa State University Press, Ames, Iowa, 1963.

[13] H. Jeffreye, "An Invariant for the Prior Probability in Estimation Problems," Proc. Roy. Soc. A., Vol. 186, 1946.

[14] K. Matusita, "On the Theory of Statistical Decision Functions," Ann Instit. Stat. Math. (Tokyo), Vol. 3, pp. 17-35, 1951.

[15] T. Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection," IEEE Trans. on Comm. Tech., Vol. COM-15, pp. 52, Feb. 1967.

[16] M. E. Bauer, P. H. Swain, R. P. Mroczynski, P. E. Anuta, and R. B. Macdonald, "Detection of Southern Corn Leaf Blight by Remote Sensing Techniques", Proceedings of the 7th International Symposium on Remote Sensing of Environment, Volume 1, 17-21, May 1971.

[17] P. H. Swain, T. V. Robertson, and A. G. Wacker, "Comparison of the Divergence and B-distance in Feature Selection," LARS Information Note 020871, Purdue University, Lafayette, Indiana.

[18] H. P. Kramer and M. V. Mathews, "A Linear Coding for Transmitting a Set of Correlated Signal," IRE Trans. on Information Theory, Vol. IT-2, pp. 41-46, 1966.

[19] L. C. Wilkins and P. A. Wintz, "A Contour Tracing Algorithm for Data Compression for Two Dimensional Data", School of Electrical Engineering Purdue University, Lafayette, Indiana, Technical Report TR-EE 69, 1-14.