Purdue University
Purdue e-Pubs

LARS Technical Reports | Laboratory for Applications of Remote Sensing

1-1-1973

Machine Boundary Finding and Sample Classification of Remotely Sensed Agricultural Data

J. N. Gupta

R. L. Kettig

D. A. Landgrebe

P. A. Wintz

Follow this and additional works at: http://docs.lib.purdue.edu/larstech

Gupta, J. N.; Kettig, R. L.; Landgrebe, D. A.; and Wintz, P. A., "Machine Boundary Finding and Sample Classification of Remotely Sensed Agricultural Data" (1973). *LARS Technical Reports.* Paper 119.
http://docs.lib.purdue.edu/larstech/119

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

LARS Information Note 102073

Machine Boundary Finding and Sample Classification of Remotely Sensed Agricultural Data

J. N. Gupta, R. L. Kettig,
D. A. Landgrebe, P. A. Wintz

The Laboratory for Applications of Remote Sensing

Purdue University, West Lafayette, Indiana

1973

Copyright © 1973
Purdue Research Foundation

MACHINE BOUNDARY FINDING AND SAMPLE CLASSIFICATION OF REMOTELY SENSED AGRICULTURAL DATA*

J. N. Gupta, R. L. Kettig, D. A. Landgrebe, and P. A. Wintz

The authors are with the Department of Electrical Engineering and the Laboratory for Applications of Remote Sensing (LARS), Purdue University, West Lafayette, Indiana.

I. ABSTRACT

Initially, methods for analyzing earth observational data involved the use of only spectral variations measured from the scene. The work reported in this paper provides a method for making some use of spatial variations as well. The results of some preliminary tests of this new method show significant improvements in accuracy.

II. INTRODUCTION

Presently, machine classification of multispectral data images is most commonly done on a point-by-point basis, as if the data vectors from one resolution element to the next were uncorrelated. Thus, no use is made of the spatial information contained in the scene. This is a useful but suboptimal approach, since in a practical situation strong correlations are certain to exist. Wacker and Landgrebe (1972, see reference 15) have demonstrated the value of using spatial information to improve classification accuracy. Using ground truth information they divided the image up into connected regions, each of which contained only members of a single class. Each region was then classified as a whole on the basis of all its members collectively, a technique known as "sample classification." This technique achieved considerably better classification accuracy than a maximum likelihood "point" classifier when the overlap of the class densities in feature space was moderate. Of course, in a practical situation, ground truth information is available only for training the classifier and evaluating its performance. So in order to realize the benefits of sample classification, another method must be employed to determine the regions to be classified. But can this be done accurately enough to preserve the natural advantage of the sample classifier? The research reported herein indicates that it can.

III. DESCRIPTION OF THE DATA SOURCE

The multispectral source to be studied is shown in figure 1. The process $a(x,y,\lambda)$ is some measure of the spectral energy at wavelength λ for ground resolution point (x,y). The multispectral data image is the set of vectors, $\{\underline{A}_{lm};\ l=1,2,\ldots,L;\ m=1,2,\ldots,M\}$, obtained from the continuous stochastic process, $a(x,y,\lambda)$, by discretizing the two spatial variables (x,y) and the spectral variable (λ). Thus

$$\underline{A}_{lm}=[a(x_l,y_m,\lambda_1),\ a(x_l,y_m,\lambda_2),\ldots,\ a(x_l,y_m,\lambda_N)]^t$$

*This paper was prepared from material extracted from references 5 and 8, by R. L. Kettig with the accord of all authors. The research was supported by NASA Grant NGL 15-005-112 and NASA Grant NGR 15-005-152.

Let $p(\underline{A}_{lm}/\omega_j)$ be the joint probability density function of the elements of vector $\underline{A}_{lm}$ conditioned on the event that the point (x_l, y_m) lies in a region whose proper classification is the jth class out of J possible classes. It will be assumed that this function is Gaussian with mean vector $\underline{\mu}_j$ and covariance matrix $\underline{\Sigma}_j$. This assumption leads to a fairly simple form of classifier which classifies a group of vectors merely on the basis of their mean vector and covariance matrix and which has proven to be fairly robust even when used with non-Gaussian distributions.

We note in passing that in order for the J classes to be "separable", there must be significant differences between their means and/or covariances. Thus the unconditional probability density $p(\underline{A}_{lm}) = \sum_{j=1}^{J} p(\underline{A}_{lm}/\omega_j)P(\omega_j)$ will be multi-modal in general.

IV. BOUNDARY FINDING

A. BACKGROUND

"Boundary finding" is the name given to the process of locating homogeneous regions within the data image which are presumed to contain only members of a single class. Most methods for boundary 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. Hueckel (1971, see reference 7) 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 (1973, see reference 4) to detect edges in simple scenes using apriori 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 (Rosenfeld, et al., 1971, 1972). Some approaches investigated at Purdue University were based on the gradient concept (Anuta, 1970), clustering (Wacker 1969), and hypothesis testing based on first order statistics (Anuta, et al., 1972). 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, and was therefore selected for the task at hand. Two versions were tested: the one based on first order statistics (Anuta, et al., 1972) and another which tests second order statistics also.

B. HYPOTHESIS TESTING

In this discussion, each vector $\underline{A}_{lm}$ of the data image is called a "pixel" (picture element), and a small, square group of pixels is called a pixel group. A "field" is a collection of connected pixel groups which have been found to be statistically similar. Hypothesis testing is used to compare a pixel group, which has not yet been assigned to a field, with a neighboring field. Call the unassigned pixel group "Sample 1" and the field "Sample 2." Consider any particular spectral channel, λ_n, and let:

X_{ij} = the data value of pixel i in sample j (in the nth channel)

N_j = the number of pixels in the jth sample

$$M_j = \frac{1}{N_j} \sum_{i=1}^{N_j} X_{ij}$$

$$V_j = \sum_{i=1}^{N_j} (X_{ij} - M_j)^2$$

$$F = \frac{N_2-1}{N_1-1} \cdot \frac{V_1}{V_2}$$

$$t = \sqrt{\frac{N_1 N_2 (N_1+N_2-2)}{(N_1+N_2)(V_1+V_2)}}\,(M_1-M_2)$$

Due to the nature of the data source (Section III), we can assume that the random variables X_{ij} come from a normal distribution whose mean (μ_j) and variance (σ_j^2) are unknown. On the basis of the two samples, it is desired to test the hypotheses in both of the following:

i) $H_{o_1}: \sigma_1^2 = \sigma_2^2$

$H_{a_1}: \sigma_1^2 \neq \sigma_2^2$

ii) $\left.\begin{array}{l} H_{o_2}: \mu_1 = \mu_2 \\ H_{a_2}: \mu_1 \neq \mu_2 \end{array}\right\}$ assuming $\sigma_1^2 = \sigma_2^2$

One can show that the maximum likelihood ratio test of the null hypothesis H_{o_2} corresponds to a two-tailed test on the parameter t, which has the tabulated "Student's t-distribution" with $(N_1 + N_2 - 2)$ degrees of freedom. The test is implemented by computing t and comparing it with a threshold value t_c. H_{o_2} is accepted only if $t^2 < t_c^2$. Likewise, the maximum likelihood ratio test of the null hypothesis H_{o_1} corresponds to a two-tailed test on the parameter F, which has the tabulated "F-distribution" with (N_1-1) and (N_2-1) degrees of freedom. For some threshold value F_c, the hypothesis H_{o_1} is accepted only if $F<F_c$ and $\frac{1}{F} < F_c$. The thresholds t_c and F_c depend upon the number of degrees of freedom and upon a significance level chosen by the user. For this purpose standard tables (Messington and Thompson, 1943; Ostle, 1963) have been used.

There is a simpler alternative to hypothesis (i) that has also been found to give good results. Essentially one assumes that all classes have the same variance in any given spectral channel. This reduces the hypothesis testing to first order statistics and eliminates the need for the F-test. But it does not eliminate the need for some sort of check on the variance to guard against violations of our implicit assumption that both samples constitute relatively homogeneous sets of pixels. (The F-test previously provided this check.) Thus hypothesis (i) is replaced by:

H_{o_1}: Both samples are "homogeneous."

H_{a_1}: Either sample is not "homogeneous."

The magnitude of the ratio (mean value/standard deviation) has been found to be a useful measure of homogeneity. Thus we say that sample j is "homogeneous" only if $V_j/N_j < (0.15\ M_j)^2$, where the constant 0.15 was derived empirically.

The goal of hypothesis testing is to determine whether Sample 1 should be incorporated into Sample 2 or assigned to another (possibly new) field. Whenever the latter action is taken, a boundary is said to exist between Samples 1 and 2. And this will be the case whenever <u>either</u> null hypothesis H_{o_1} or H_{o_2} is rejected. The boundaries may be weak in some spectral channels and strong in others, so it is necessary to perform these tests in <u>each</u> spectral channel before a decision to combine the samples can be made. This has been called a "multiple-univariate" approach. In order to take the inter-channel correlation into account, a single multivariate test could be used instead. This has been done, and it was found to be very time consuming and not very advantageous for the data available at LARS.

C. GEOMETRY OF FIELD CONSTRUCTION

The digital picture (flightline) being processed is assumed to be rectangular, usually many more rows than columns. Beginning with the first row, the boundary finder builds fields from pixel groups, usually (2 pixels) x (2 pixels). A field begins with one pixel group and expands laterally and down the flightline absorbing more pixel groups until it reaches its natural boundaries. Thus the field can be any shape. At the boundaries new fields are begun which expand in the same manner until the entire flightline has been partitioned into homogeneous regions. The actual logic by which this is accomplished is conceptually simple, but difficult to describe. Therefore, the description has been relegated to the flow chart in figure 2.

V. SAMPLE CLASSIFICATION

Once a closed field has been "found" in the data set, many methods are available for classification. For example, one could simply classify each vector in the field individually using a maximum likelihood decision rule and poll the results to determine the field classification (Huang, 1969). Or one could average the data vectors over the entire field to obtain an estimate of the mean vector and then merely classify this mean using a maximum likelihood decision rule. If the field has a sufficient number of elements to estimate its N-dimensional probability density in feature space, then one of the more elaborate "minimum distance" sample classifiers (Wacker and Landgrebe, 1972) can be employed. The number of data vectors required can be relatively few if a parametric characterization of the probability density is used.

The classification algorithm that was developed for use with the boundary finder contains both a maximum likelihood vector classifier <u>and</u> a minimum distance sample classifier. The vector classifier is used only if the field contains an insufficient number of elements to estimate its probability density. Then only the mean vector is classified and the result is assumed to apply to all elements in the field. Specifically, the maximum likelihood decision rule amounts to choosing the class (ω_j) which <u>minimizes</u> the quantity

$$(\underline{M}-\underline{M}_j)^t \, \underset{\sim}{\Sigma}_j^{-1} \, (\underline{M}-\underline{M}_j)+\ln |\underset{\sim}{\Sigma}_j|, \; j=1,2,\ldots,J, \text{ where:}$$

J = the number of classes being considered
$\underline{M}$ = mean vector of the field to be classified
$\underline{M}_j, \underset{\sim}{\Sigma}_j$ = estimated mean vector and covariance matrix of the jth class based on training samples.

The assumption is made that all classes are equiprobable. Otherwise it would not be clear how to assign individual class probabilities.

The sample classifier chosen for this experiment is based upon the Bhattacharyya distance* (Wacker and Landgrebe, 1972). Specifically, the class (ω_j) is chosen for which the following quantity is minimum:

$$\tfrac{1}{2}\ln\frac{\left|\tfrac{1}{2}(\underset{\sim}{\Sigma}+\underset{\sim}{\Sigma}_j)\right|}{\sqrt{|\underset{\sim}{\Sigma}|\,|\underset{\sim}{\Sigma}_j|}}+\tfrac{1}{4}(\underline{M}-\underline{M}_j)^t(\underset{\sim}{\Sigma}+\underset{\sim}{\Sigma}_j)^{-1}(\underline{M}-\underline{M}_j), j=1,2,\ldots J,$$

where $\underset{\sim}{\Sigma}$ is the covariance matrix of the field to be classified.

VI. EXPERIMENTAL RESULTS

Figures 3 and 4 show approximate gray-scale printouts (in distinct spectral channels) of a typical agricultural flightline having 12 channels in all. This data was collected by an airborne scanner system during the 1971 Corn Blight Watch Experiment. The strip of terrain observed is represented by 222 samples across track and 176 samples/mile along track taken at a 5000 foot altitude. For analysis purposes, a subset of 3 of the 12 available spectral channels was selected, namely

*In this case the word "distance" is not to be confused with the mathematical term "metric."

0.61-0.70 μm, 0.72-0.92 μm, and 9.30-11.70 μm. This step reduces the overall computation time and computer memory requirements. Also it requires fewer data vectors to estimate a 3 x 3 covariance matrix than a 12 x 12 covariance matrix. The criterion for selecting channels was that the minimum transformed divergence between any two training classes was larger for the above subset than for any other subset of 3 channels. The transformed divergence is a measure of class separability in feature space (Swain, et al., 1971), but maximizing it will not necessarily guarantee the highest classification accuracy. In other words, some other subset of 3 channels may give better performance than the one used here.

A set of training fields was obtained representing 5 classes, namely: corn, forage, soybeans, forest and water. On this basis all of the 222 samples across-track and 400 samples along-track were classified using the boundary finder and classification algorithms described previously. The accuracy of the classification was evaluated on the basis of test fields that are distinct from the training fields.

Presently the same data at LARS is being classified on a point classification basis using the Gaussian Maximum Likelihood Ratio decision rule (Fu, et al., 1969) with an error rate of 4.1%. The boundary finder based on first order statistics reduced this rate to 2.3% and the boundary finder using second order statistics as well achieved an error rate of only 2.0%. A classification map and tabulated results are provided in figure 5 and Tables 1-3.

The first order boundary finder was also used in a more complex classification study (Kettig and Landgrebe, 1973) involving 9 classes altogether. Without further elaboration, we merely note that the error rate was reduced from 8.8% to 3.6% by using the boundary finding technique.

The second order boundary finder was also used to classify data from the Earth Resources Test Satellite. Due to the altitude at which this data was collected, there are relatively few elements per field. Since there is less spatial information (correlation) to take advantage of, one would expect the boundary finder to provide only a small degree of improvement over the "point" classifier; and indeed, this was the case.

VII. SUMMARY

In summary, it has been shown that spatial information can be extracted from a multispectral data image, by machine, effectively enough to significantly assist the classification process. The degree of improvement will, of course, depend upon the amount of spatial information contained in the scene and upon the degree of overlap of the class densities. In addition, the concept of boundary finding seems to be a fairly robust one, insensitive to many changing variables. In spite of idealistic assumptions, differing implementations, and differing degrees of classification complexity, the boundary finder in conjunction with a sample classifier has produced consistently better results than its well-proven forerunner, the maximum likelihood Gaussian "point" classifier.

VIII. REFERENCES

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

2. Anuta, P. E.; Rodd, E. M.; Jensen, R. E.; and Tobias, P. R. 1972. Final Report for the LARS/Purdue - IBM Houston Scientific Center Joint Study Program. Purdue University, Lafayette, In.

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

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

5. Gupta, J. N., and Wintz, P. A. 1973. Closed Boundary Finding, Feature Selection, and Classification Approach to Multi-Image Modeling. LARS Information Note 062773, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, In.

6. Huang, T. 1969. Per Field Classifier for Agricultural Applications. LARS Information Note 060569, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, In.

7. Hueckel, M. 1971. An Operator Which Locates Edges in Digitized Pictures. J. ACM, 18:113-25, Jan.

8. Kettig, R. L., and Landgrebe, D. A. 1973. Automatic Boundary Finding and Sample Classification of Remotely Sensed Multispectral Data. LARS Information Note 041773, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, In.

9. Messington, M., and Thompson, C. M. 1943. Tables of Percentage Points of the Inverted Beta (F) Distribution. Biometrika, 33:73.

10. Ostle, B. 1963. Statistics in Research. Iowa State University Press, Ames, Iowa.

11. Rosenfeld, A., and Thurston, M. 1971. Edge and Curve Detection for Visual Scene Analysis. IEEE Trans. on Computers, Vol. C-20, pp. 562-9, May.

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

13. Swain P. H.; Robertson, T. V.; and Wacker, A. G. 1971. Comparison of the Divergence and B-Distance in Feature Selection. LARS Information Note 020871, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, In.

14. Wacker, A. G. 1969. A Cluster Approach to Finding Spatial Boundaries in Multispectral Imagery. LARS Information Note 122969, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, In.

15. Wacker, A. G., and Landgrebe, D. A. 1972. Minimum Distance Classification in Remote Sensing. LARS Print 030772, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, In.

IX. TABLES

Table 1. Test Field Performance of "Point" Classifier

Group	No. of Pixels	% Correct	No. of Pixels Classified Into: 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
Totals	14588		5826	3377	4648	699	38

Overall Performance: (13995/14588) = 95.9% correct

Table 2. Test Field Performance of "First-Order" Boundary Finder

Group	No. of Pixels	% Correct	No. of Pixels Classified Into: CORN	FORAGE	SOYBEANS	FOREST	WATER
Corn	5950	98.2	5844	51	49	6	0
Forage	3151	96.5	105	3042	2	2	0
Soybeans	4770	97.9	36	44	4668	22	0
Forest	680	97.8	0	11	0	665	4
Water	37	94.6	0	2	0	0	35
Totals	14588		5985	3150	4719	695	39

Overall Performance: (14254/14588) = 97.7% correct

Table 3. Test Field Performance of "Second-Order" Boundary Finder

Group	No. of Pixels	% Correct	No. of Pixels Classified Into: 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
Totals	14588		6064	3112	4677	699	36

Overall Performance: (14294/14588) = 98.0% correct

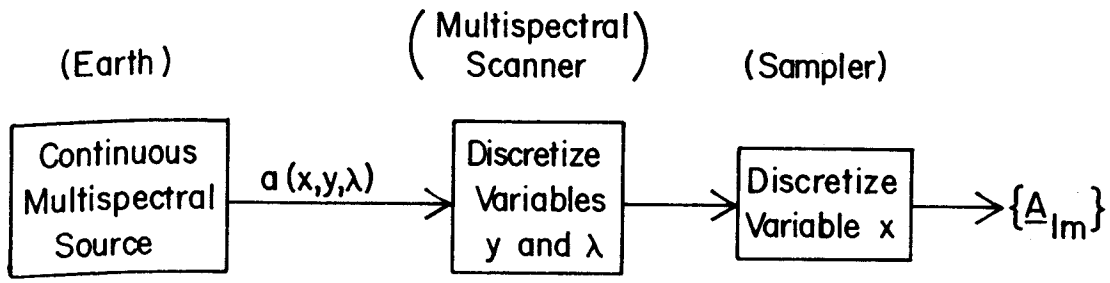
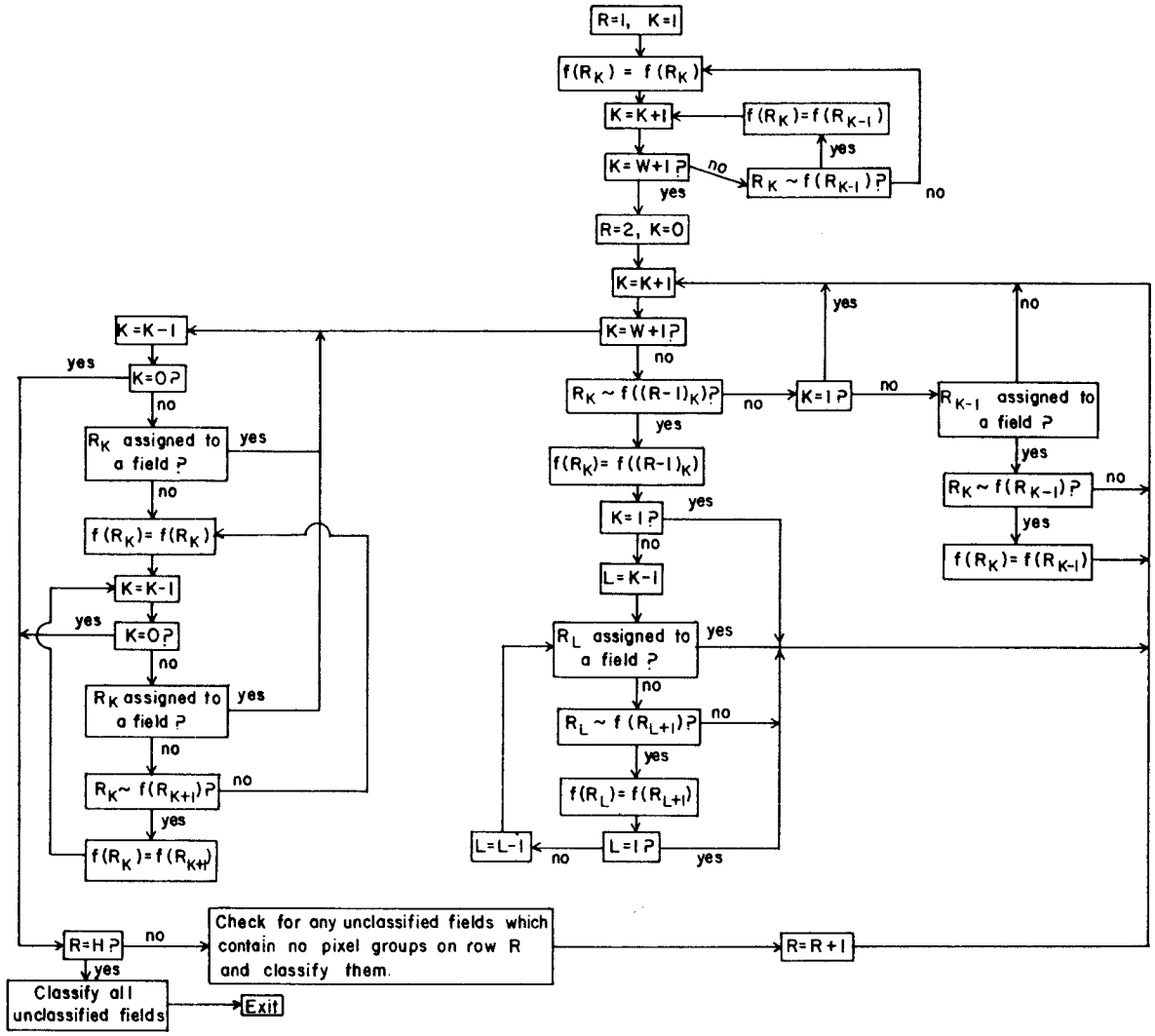


Figure 1. Model of the Multispectral Source



Legend

W = number of pixel groups per row of the digital picture.
H = number of consecutive rows of pixel groups to be processed.
R_k refers to the Kth pixel group in row R.

$f(R_k)$ refers to the field to which R_k has been assigned.

$R_k \sim f(R_{k-1})$? means "Use hypothesis test to determine if R_k is statistically similar to $f(R_{k-1})$."

$f(R_k)=f(R_{k-1})$ means "Assign R_k to same field as R_{k-1} and update the statistics of that field."

$f(R_k)=f(R_k)$ means "Begin a new field with R_k as the first entry."

Figure 2. Field Building Algorithm

Figure 3. Gray Scale Map of Flightline 210. (0.61-0.70 μm)

Figure 4. Gray Scale Map of Flightline 210. (9.30-11.70 μm)

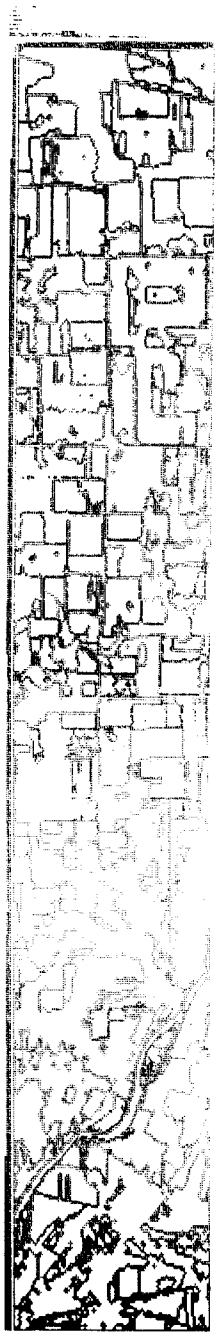
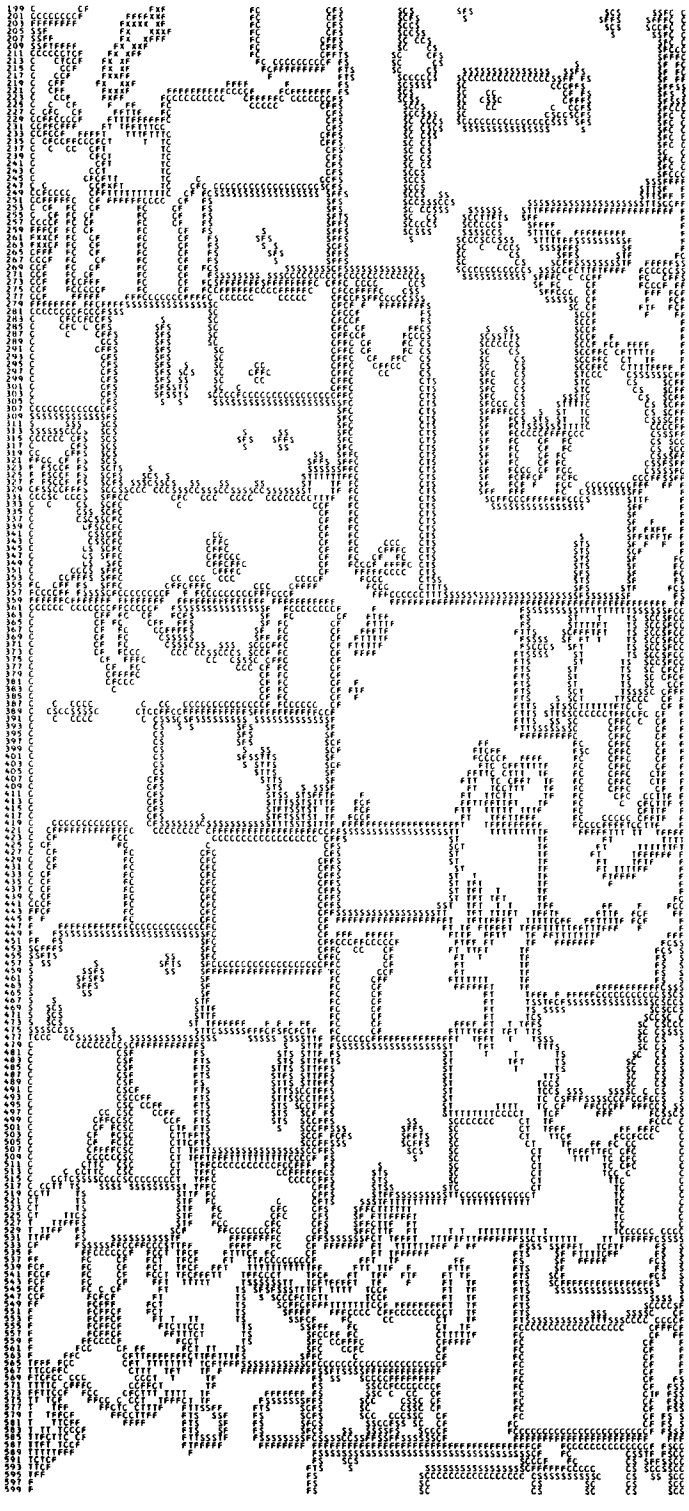


Figure 5. Classification Map of Flightline 210.



Class	Symbol
CORN	C
FORAGE	F
SOYBEANS	S
FOREST	T
WATER	X

Figure 6. Expanded View of Classification Map of Flightline 210.