SEQUENTIAL CLASSIFICATION AND CLUSTERING METHODS APPLIED TO DIGITIZED PHOTOGRAPHY

Statistical Reporting Service
U.S. Department of Agriculture
Washington, D. C.

SEQUENTIAL CLASSIFICATION AND CLUSTERING METHODS APPLIED TO DIGITIZED PHOTOGRAPHY

BY

CHAPMAN P. GLEASON

Statistical Reporting Service
U.S. Department of Agriculture
Washington, D. C.

September 1976

TABLE OF CONTENTS

SEQUENTIAL CLASSIFICATION AND CLUSTERING METHODS APPLIED TO DIGITIZED PHOTOGRAPHS

By

Chapman P. Gleason
Statistical Reporting Service
U.S. Department of Agriculture 1/

ABSTRACT

This paper presents a system for object detection and counting from transparencies digitized using a scanning microdensitometer. Classification and clustering methods are applied sequentially to digitized transparencies of an orange tree and a low altitude color infrared aerial photograph of a citrus grove. The objective is to count oranges on the digitized ground transparency and to count fruit trees on the digitized infrared aerial photograph. This general flow can be applied to any type of spectral data for object detection and counting.

An 11 x 1 vector $\mathbf{z}$ is measured on each point. The vector $\mathbf{z}$ is composed of eight spectral variables, two spatial variables, and a label variable.

$$\mathbf{z} = \begin{bmatrix} \text{8 x 1, spectral component vector} \\ \text{2 x 1, spatial component vector} \\ \text{1 dimensional label variable} \end{bmatrix}$$

Discriminant analysis is first performed on the eight spectral variables using the groups identified by the label variable. Once each point has been classified, only those points classified as oranges (or trees) are clustered using the spatial variables. A graph theoretic clustering method was applied to the classified data. The graph theoretic clustering yields a measurement vector which characterizes the cluster shape. The cluster measurement vector is:

1/ Prepared for: Symposium on Machine Processing of Remotely Sensed Data
Laboratory for Applications of Remote Sensing
Purdue University
West Lafayette, Indiana June 29, 1976 - July 1, 1976

$$W = \begin{bmatrix} \text{Cluster diameter} \\ \text{Cluster size} \\ \text{Cluster average edge length} \\ \text{Cluster standard deviation} \end{bmatrix}$$

This vector W is then used as the measurement vector to classify clusters as either tree or non-tree (or orange and non-orange). The summary from the discriminant analysis gives an estimate of the number of trees (or oranges).

Keywords: Clustering, classification, pattern recognition, multivariate analysis.

ACKNOWLEDGEMENTS

Charles T. Zahn of Stanford University provided the source program for the clustering algorithm and Jack Nealon of the Statistical Reporting Service performed most of the analysis for Data Set 2 in this report and reported his findings in "The System of Sequential Classification and Clustering" [1].

INTRODUCTION AND BACKGROUND

The Statistical Reporting Service (SRS) of the U.S. Department of Agriculture has been investigating the use of remote sensing data for improvements to crop and livestock estimating programs since the early 1960's. A detailed synopsis of some of these efforts is given by Huddleston [2] and Huddleston and Wigton [3]. Some of the most promising uses of remote sensing are:

(1) double sampling techniques for tree or crop yields,

(2) tree census for use as a sampling frame,

(3) livestock inventories,

(4) quality control on operating programs.

Early efforts used photo interpreters to count objects of interest (fruit or trees, livestock) from ground and aerial photography. However, one of the sources of error in photo interpretation techniques is in counting introduced by different photo interpreters. Previous investigations have found these errors to be of sufficient magnitude that randomized plans were performed to make adjustments to the photo interpreted counts [4, 5, 6]. This paper proposes and demonstrates computerized multivariate methods of counting objects of interest that are consistent and therefore, only adjustments for bias need be estimated.

The earliest methods within SRS of computerized counting using digitized transparencies was by Himmelberger [7], where counting oranges on selected trees was the objective. His method used "shape" of an object (orange) as the primary method to count them. This was done by approximating the shape by a sphere. These methods made no use of discrimination or pattern recognition techniques. Steele [7] improved the methodology by using discriminant functions to classify the resolution elements, hereafter, called "pixel" (for picture element), into apriori defined groups. However, no process was ever developed to count objects of interest (oranges) in his paper. Steele did show (empirically), that the digitizing instrument (a scanning microdensitometer) was capable of quantitative point discrimination using discriminant analysis. It was these earlier methods that led us to improve the methodology offering a complete systematic approach to the problem of computerized counting of digitized transparencies (either aerial or ground).

DATA ACQUISITION

A color transparency of an orange tree taken from the ground with mature fruit and a color IR aerial transparency of a citrus grove were selected to demonstrate these methods. Figure 1 and 2 are black and white reproductions of the transparencies. In Figure 1, the object of interest is the oranges and in Figure 2 the object of interest is the trees. Specifically, we would like to count these objects of interest using a computer with a minimum number of manual tasks.

An area on each transparency was selected to be digitized by a Photometric Data Services' Microdensitometer [8]. This is identified as Data Set 1 and 2, respectively. For the ground transparency, there are four groups which are spectrally identifiable and which must be properly identified for possible training data for a classifier. The groups in Data Set 1 are:

1. Orange (O) -- Objects of interest
2. Foliage (F)
3. Ground (G)
4. Sky (S)

For the aerial photograph seven groups are identifiable:

1. Trees in orchard (T) -- Objects of interest
2. Shadows of tree (S)
3. Hedges (H)
4. Bushes (B)
5. Canal water (C)
6. Lake (L)
7. Road and soil (R)

The location of each of these groups on each transparency was recorded to later locate the "training data" for analysis. "Training data" is information supplied to derive estimates of the classification parameters to develop a rule that will classify an unlabeled point into one of several groups.

Each of the selected areas on the transparencies was digitized using four different color filters and two different scanning modes, which represent respectively optical density and transmission. An effective aperture of 100 microns was used for the orange transparency and an effective aperture of 240 microns was used for the aerial transparency. Calibration after each scan was performed on clear glass. Thus, for each pixel, an 8x1 vector of observations is available.

The microdensitometer outputs gray levels, representing the optical density (or transmission), onto magnetic tape. A program (Hopkins [9]) was written to reformat the tape into a form for input to the Statistical Analysis System [10]. All the data handling and statistical methodology was performed using this system.

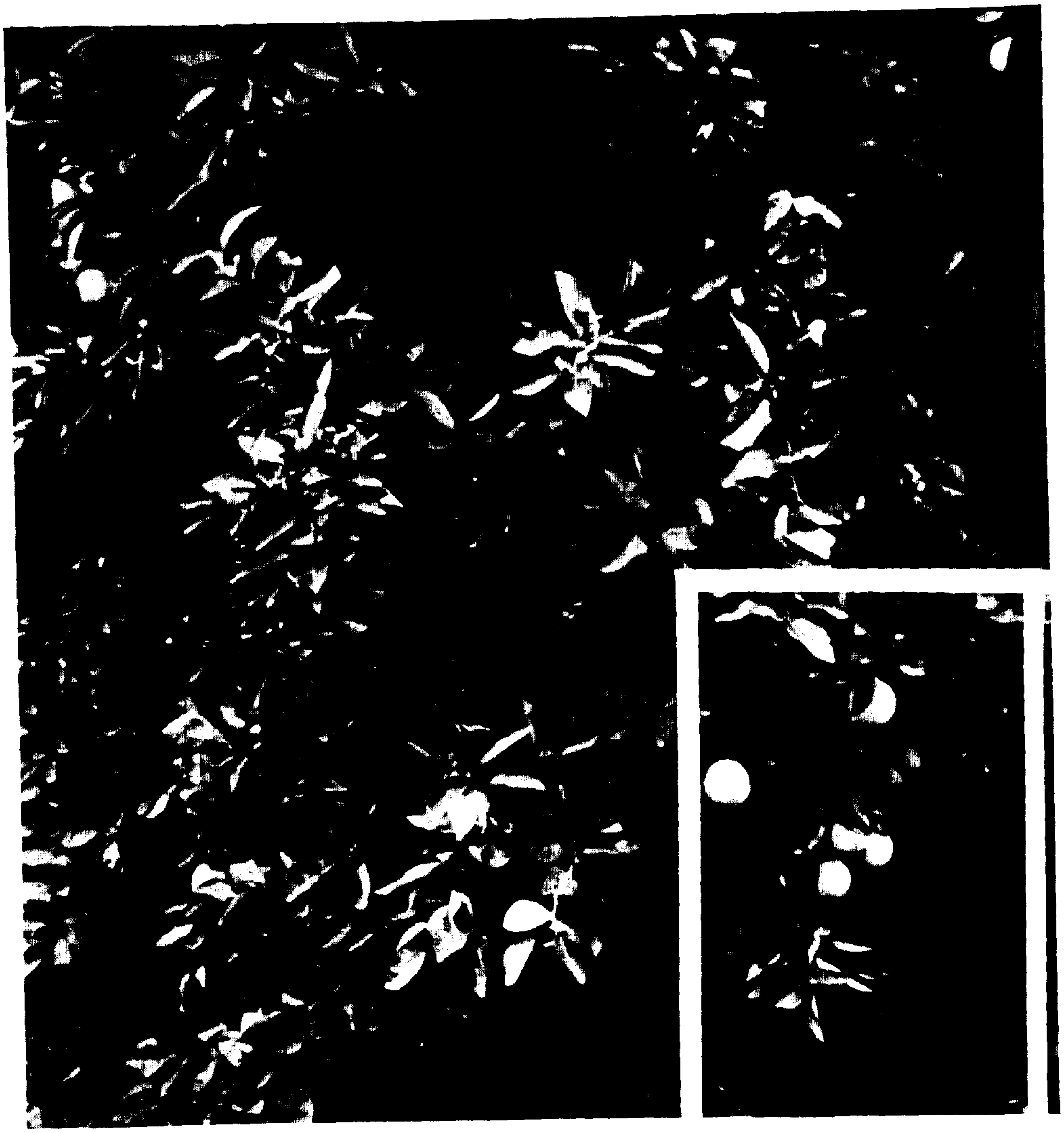


Figure 1.--Black and white reproduction of 35mm transparency of an orange tree which was digitized using a 100 micron effective aperture on a PDS Microdensitometer. Data set 1 is outlined in white.

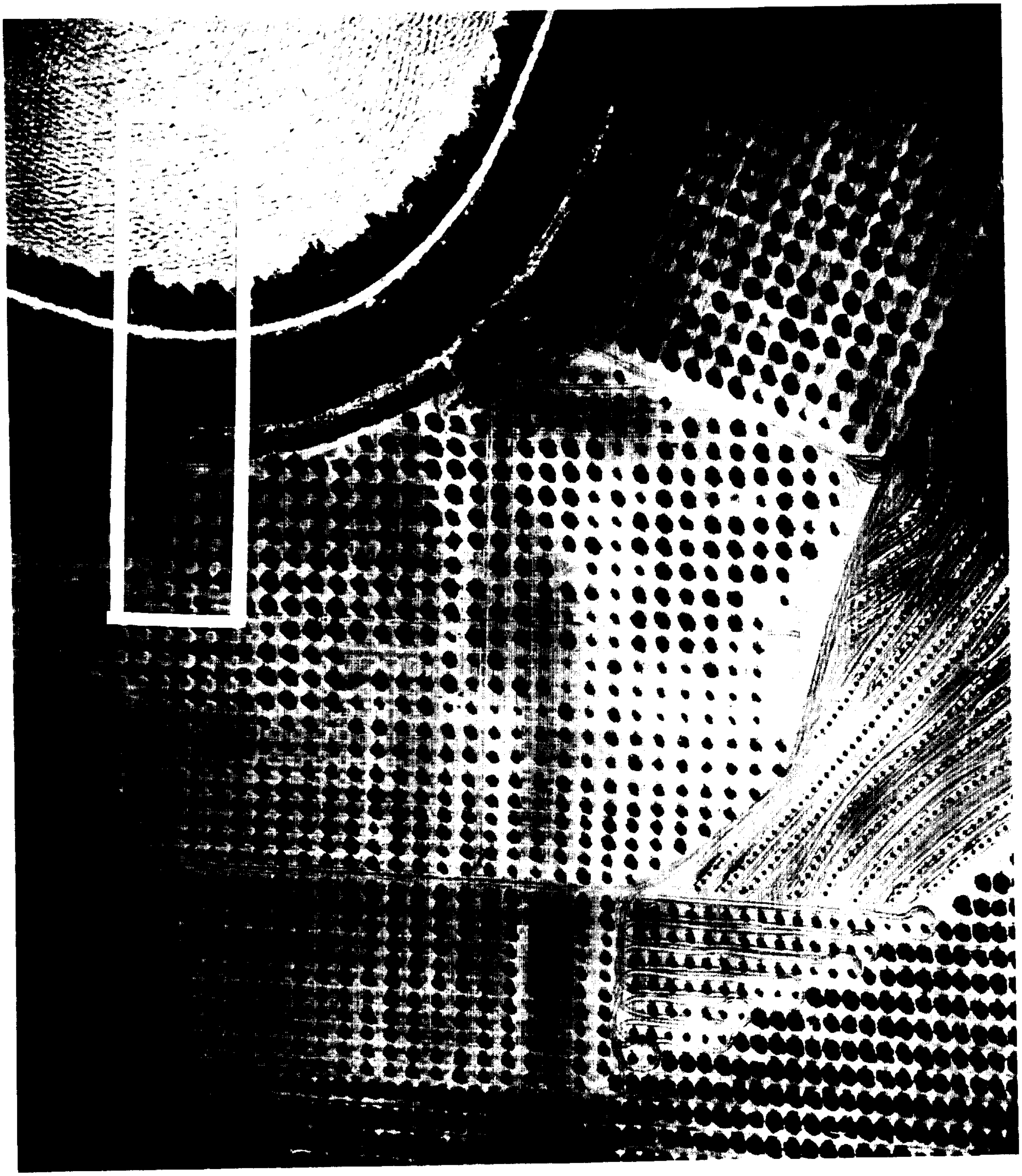


Figure 2.--Black and white reproduction of 35mm transparency of an Orchard which was digitized using a 240 micron effective aperture on a PDS Microdensitometer. Data set 2 is outlined in white.

The following gives a description of the measurement vector made on each pixel in each data set.

$$Z = \begin{bmatrix} X \\ Y \\ DCLEAR \\ DRED \\ DGREEN \\ DBLUE \\ TCLEAR \\ TRED \\ TGREEN \\ TBLUE \\ GROUP \end{bmatrix} \quad (1)$$

Where for each pixel:

X is distance (in pixels) from a fixed origin in the horizontal direction.

Y is distance (in pixels) from a fixed origin in the vertical direction.

DCLEAR is the optical density using a clear filter.

DRED is the optical density using a red filter.

DGREEN is the optical density using a green filter.

DBLUE is the optical density using a blue filter.

TCLEAR, TRED, TGREEN, TBLUE are the transmission values recorded for the corresponding filter.

GROUP is the label variable with values assigned to the pixel if it is training data for the classifier. No group label identifier is attached to any pixel which is not training data.

We see that there are three different types of variables measured on each pixel. X and Y are <u>spatial</u> variables and give the relative position of the pixel element from a fixed origin which is determined when the transparency is digitized. DCLEAR, TBLUE, etc., are <u>spectral</u> variables and each variable represents the optical density or transmission relative to a calibration point. The group variable (GROUP) gives the <u>label</u> of each pixel for training the classification procedure.

In summary, the 11x1 vector $\mathfrak{z}$ can be represented by:

$$\mathfrak{z} = \begin{bmatrix} \text{Spatial variables, two dimensions} \\ \text{Spectral variables, 8 dimensions} \\ \text{Label variable, one dimensional} \end{bmatrix}$$

THE ANALYSIS APPROACH

There are two data sets containing gray levels of a quantized transparency. Our eye can immediately detect the oranges in Figure 1 and trees in Figure 2 by their spectral properties in relation to background in each transparency. A measurement vector (1) was made on each pixel representing the gray level of the particular spot which was quantized. This vector will be used to discriminate between different spectral groups in each of the data sets using the spectral variables.

Once each of the data sets have been classified only the "group of interest" in each data set will be saved for further analysis and processing. The "group of interest" in Data Set 1 is the oranges and the group of interest in Data Set 2 is the orange trees. The pixels classified into the groups of interest will then be clustered for "possible oranges" and "possible trees" in Data Set 1 and Data Set 2 respectively using the spatial properties of the classified pixels. If we discriminated perfectly between groups (a highly unlikely turn of events), we could just count the clusters in each data set and this would be the number of objects which we were trying to detect in each data set. However, this will not likely be the case and some clusters will not be oranges and some will not be trees. However, we will determine a "conservative" classification procedure which will detect at least all objects of interest and then determine which objects to discard in a later analysis stage. Thus, we seek a classification rule which minimizes the omission error in the classification procedure.

Saving the results of the classified data in clustered form, we notice that in Figure 2 the trees have some "regularity" in regard to shape. Characteristics of the clusters will be determined to measure the shape properties of the trees. Clusters will then be identified as "objects of interest" and "not of interest" and a discriminant analysis will be performed on the cluster characteristics. The clusters classified into the group "object of interest" will then be accumulated to estimate the number of objects present in the data set.

The analysis procedure outline above will be discussed in detail on the two data sets in the following phases of the analysis. Due to the discussion presented above, the phases will be organized as follows:

- Spectral classification
- Spatial clustering
- Cluster classification

A detailed flow diagram is given in Appendix 1.

Spectral Classification

Much has been written on discriminant analysis as a technique for statistical analysis and pattern recognition [11-22]. Anderson [23], Kshirsagar [24], and a recent book by Lachenbrach [25] gives an introduction to the topic from a Bayesian statistical point of view. The earliest statistical writings on the topic was by Fisher who developed the linear discriminant function. Some advanced problems in discriminant analysis can be found in Caloullos [20]. Recent articles by Fisher and Von Ness [22], Lachenbrach [26], and Weiner [16] indicate this area of statistical analysis is being actively reviewed and researched with respect to statistical distribution, estimation, and decision/hypothesis testing theory. Almost all the theoretical work has assumed normal theory and known population parameters. Large sample theory and "faith" are the usual properties of sample estimates of the mean vector and covariance matrix when the parameters are not known and must be estimated.

Some research on nonparametic discrimination has been investigated. Cover and Hart [27], and Bennett, et al [28] give some theoretical approaches to estimating the group density functions (their form as well as their parameters) by nearest neighbor methods and spline functions.

In the field of pattern recognition, survey articles can be found in Hall, et al [11], Nagy [13], Ho and Agrawala [14]. Much of the pattern recognition work which the author is familiar with has been in the field of remote sensing using satellite data collection systems. Pattern recognition methodology has been applied extensively in remotely sensed data. Swain [12] gives a discussion for the theoretical and practical basis of using pattern recognition methods on remotely sensed data.

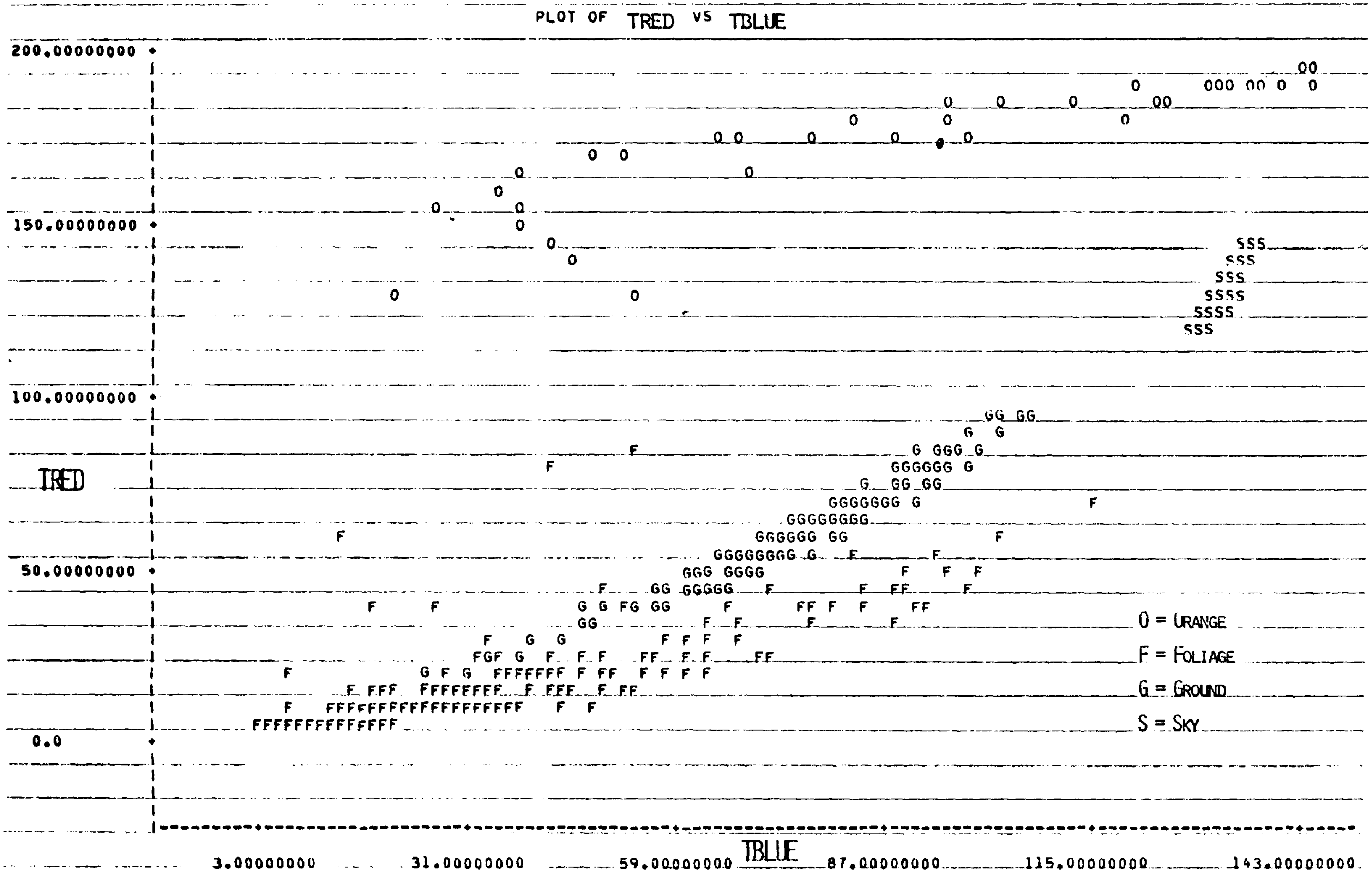
Turning now to the specific problem at hand, we have multivariate measurements made on several groups and we wish to know how to optimally (minimize the probability of misclassification) assign pixels to groups. However, no classification rule will be capable of good decisions if the groups are not separable. Figure 3 and 4 give plots using two variables for the training data in Data Sets 1 and 2.

In the case of Data Set 1, we can see (Figure 3) that the orange pixels are visually separable from the other groups using just two of the eight spectral variables. A similar situation occurs for the tree pixels in Data Set 2 (Figure 4). These plots demonstrate that the data sets are spectrally separable for the group of interest. In other fields of pattern recognition, for example, crop species identification using low resolution satellite multispectral scanner data, determination of spectrally separable groups is a time consuming and difficult undertaking involving several analytic techniques. In this case, we have solved this by visually clustering the two variable plots for each data set.

The number of groups to use in the discriminant analysis and the variables which best separate the group(s) of interest must now be determined for each data set. Thus, we must attack the group separability and variable selection (sometimes called feature or channel selection in pattern recognition) as a joint process to determine the discriminant function and hence, the classification rule. Several articles have appeared in the recent literature concerning variable selection in discriminant analysis. Weiner and Dunn [16], Davies [17], Cochran [18], and Rao [19] give discussion on variable selection for two group linear discriminant function with 0,1 loss and equal apriori probabilities. There is no general solution to the variable selection problem with $K>2$ groups and finite sample sizes. For infinite populations it can be proven that all dimensions (variables) should be used as discriminators [23]. However, this is not the

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M

SCENE=100



LEGEND: A = 1 OBS , B = 2 OBS , ETC.

FIGURE 3 - PLOT OF TRED AND TBLUE FOR DATA SET 1.

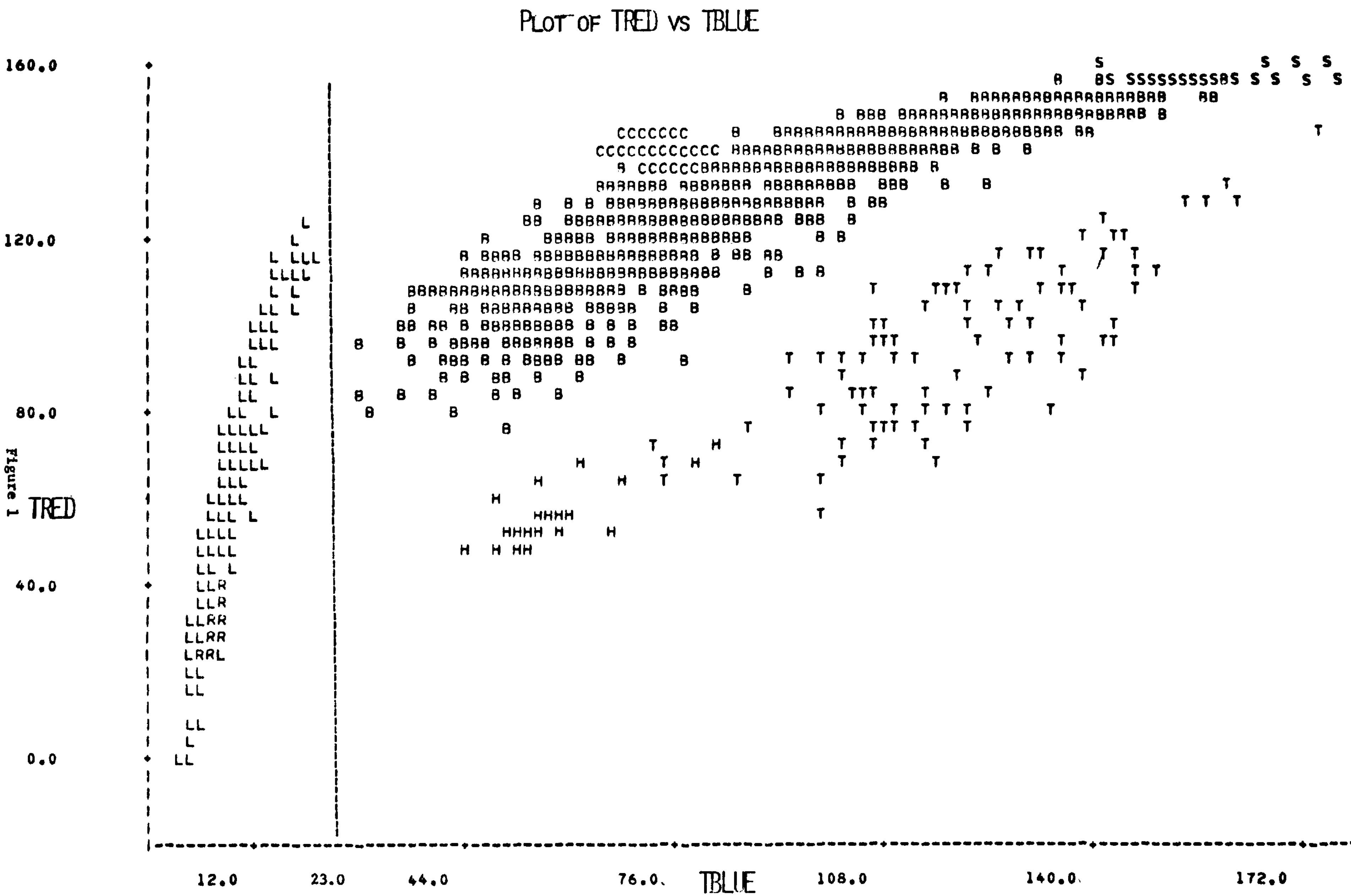


FIGURE 4 - PLOT OF TRED VS TBLUE FOR DATA SET 2.

case for finite sample and is sometimes referred to as the dimensionality problem [29, 30].

There are several classification strategies observable in the data in Figures 3 and 4. Figures 5-7 gives several different strategies for the two data sets. The strategies are referred to as "Decision Tree Classifiers" or "Layer Classifiers" in remote sensing pattern recognition [29, 30]. This methodology was developed independently and we will call these strategies "hierarchial classifiers." Classification Strategy B in Figure 5 is the usual single stage classifier or one step discriminant analysis with all groups used. Strategy A is a clustered classification strategy with group foliage, ground and sky merged into one group for the classification. Which classification strategy is "best" is a function of several variables some of which are: (a) the separability of the groups, (b) the variables used at each step of the hierarchy, and (c) how well the "training data" represents all the data types in the data set. Note that at each step of the hierarchy variables can be selected which best discriminate between the groups at that particular stage of the classification. Several different classifications using different combinations of groups and variables were tried on each data set. Figures 8-10 summarize the percent correct classification on each of the groups of interest in each data set. The percent correct classification reported was estimated using the "training" data as "test" data or as Lachenbrach calls it, the "apparent" error rate [26]. It is well known, that this measure of the true performance of a classifier underestimates the true error rate [26].

Comparing the performance of the classification strategies in Figures 8 and 9 we see that for point discrimination on DATA SET 1;

1) unequal priors degraded the classification performance with the optical density variables (DCLEAR, DRED, etc.); however, it had little or no effect for transmission variables.

2) in all cases two variables had all the discriminating power of four variables.

For DATA SET 2 each classification strategy analyzed has nearly identical performances. The use of prior probabilities in DATA SET 1 had very little effect on overall performance for transmission variables, since the orange pixels were very separable spectrally.

FIGURE 5

CLASSIFICATION STRATEGIES

DATA SET 1

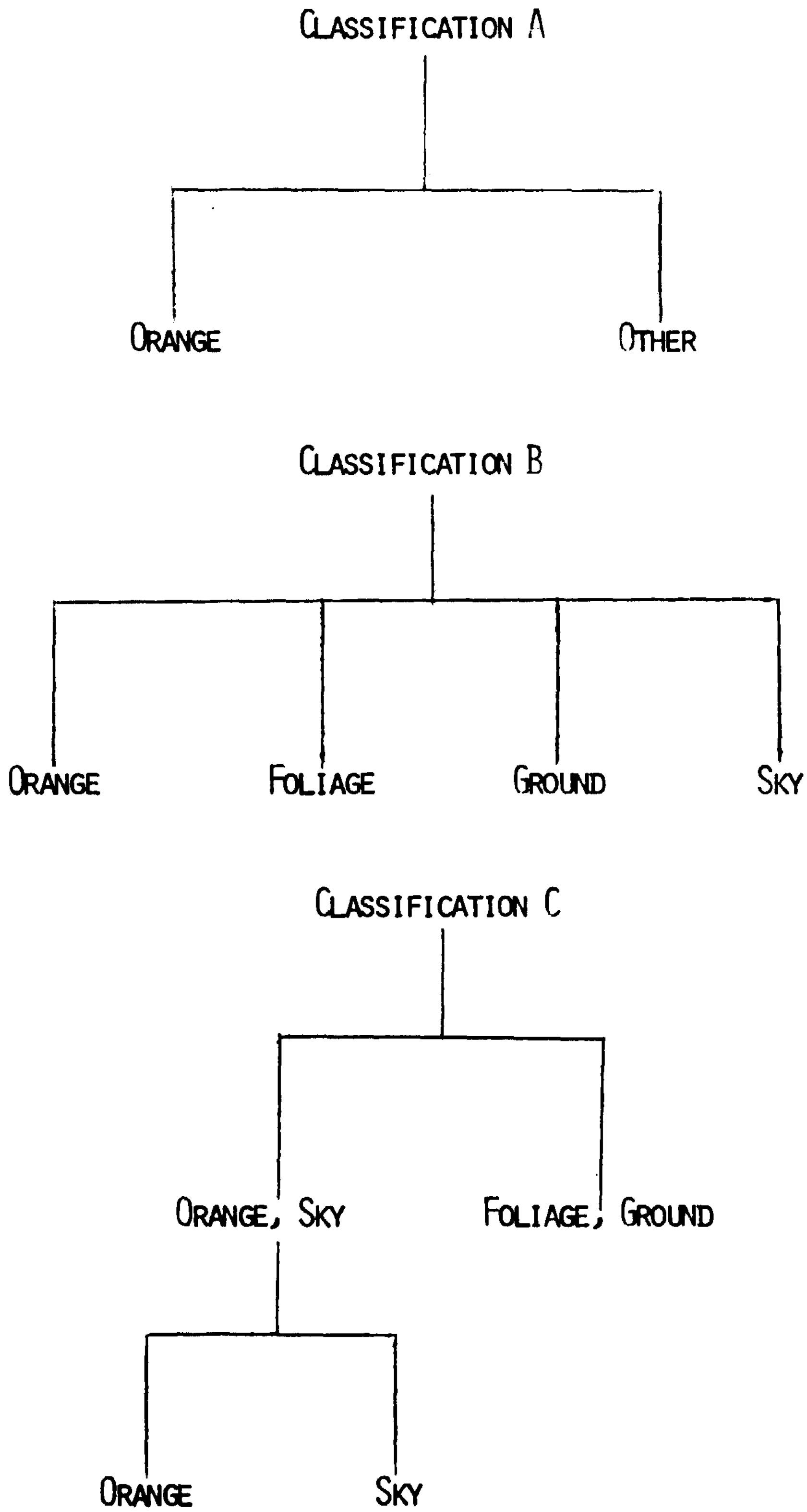


FIGURE 6

CLASSIFICATION STRATEGIES

DATA SET 2

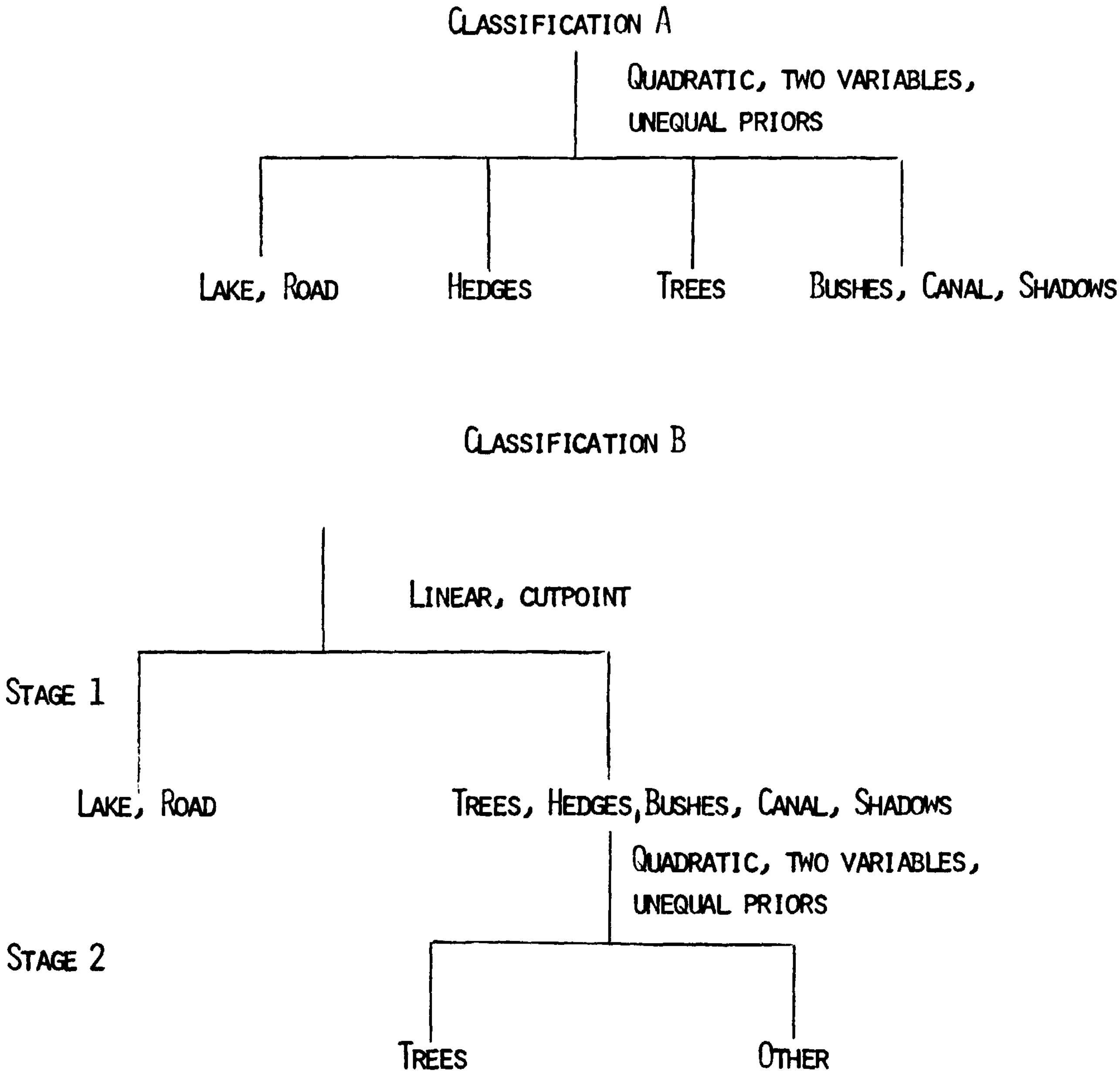
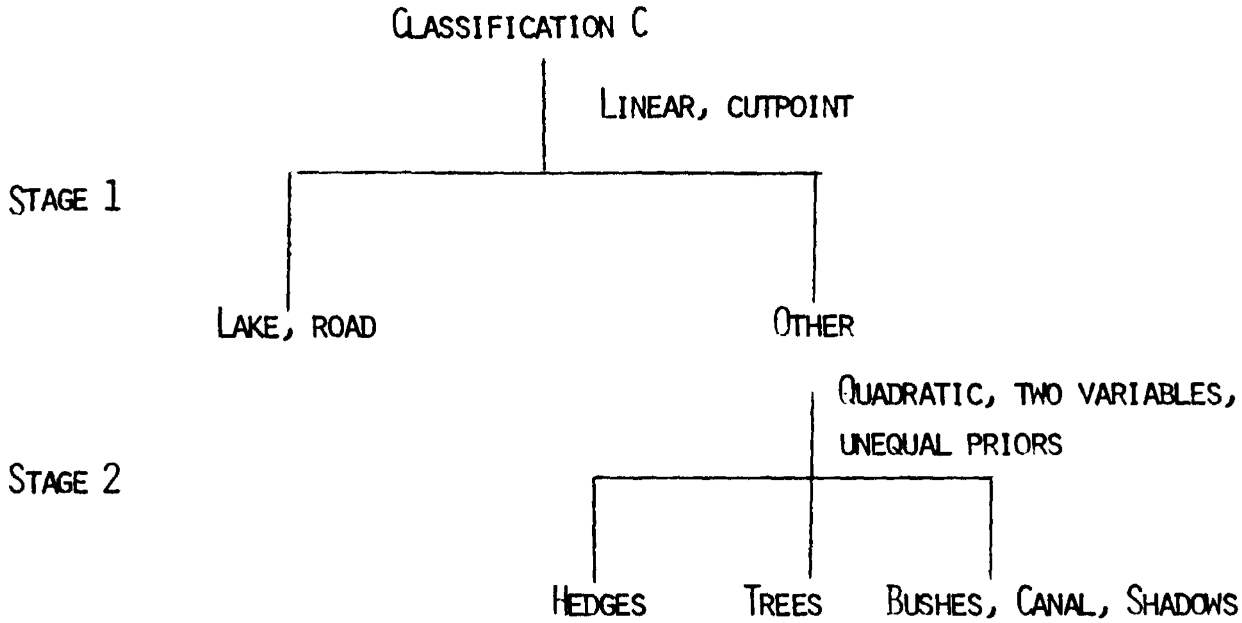
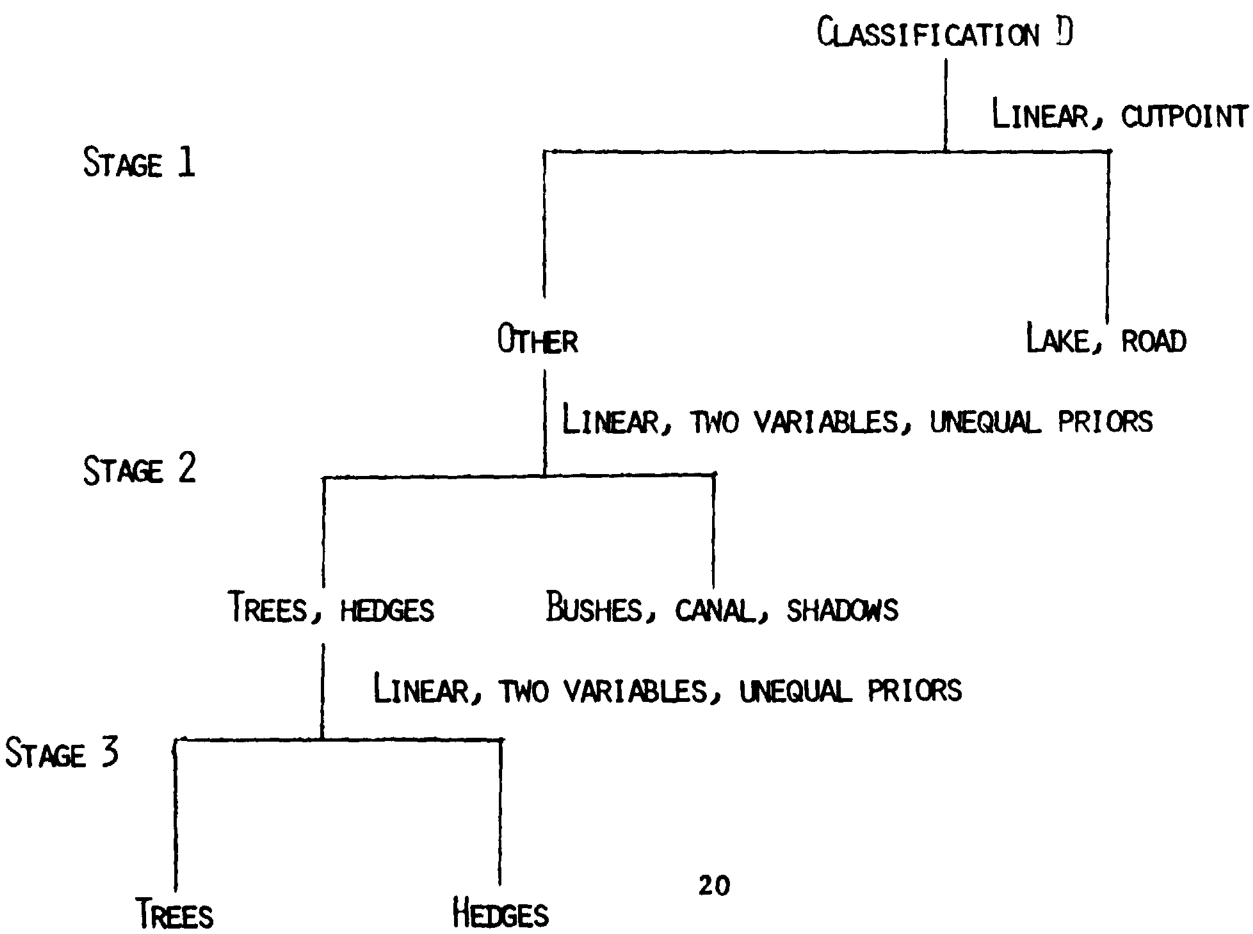


FIGURE 7

CLASSIFICATION STRATEGIES (CONT.)

DATA SET 2





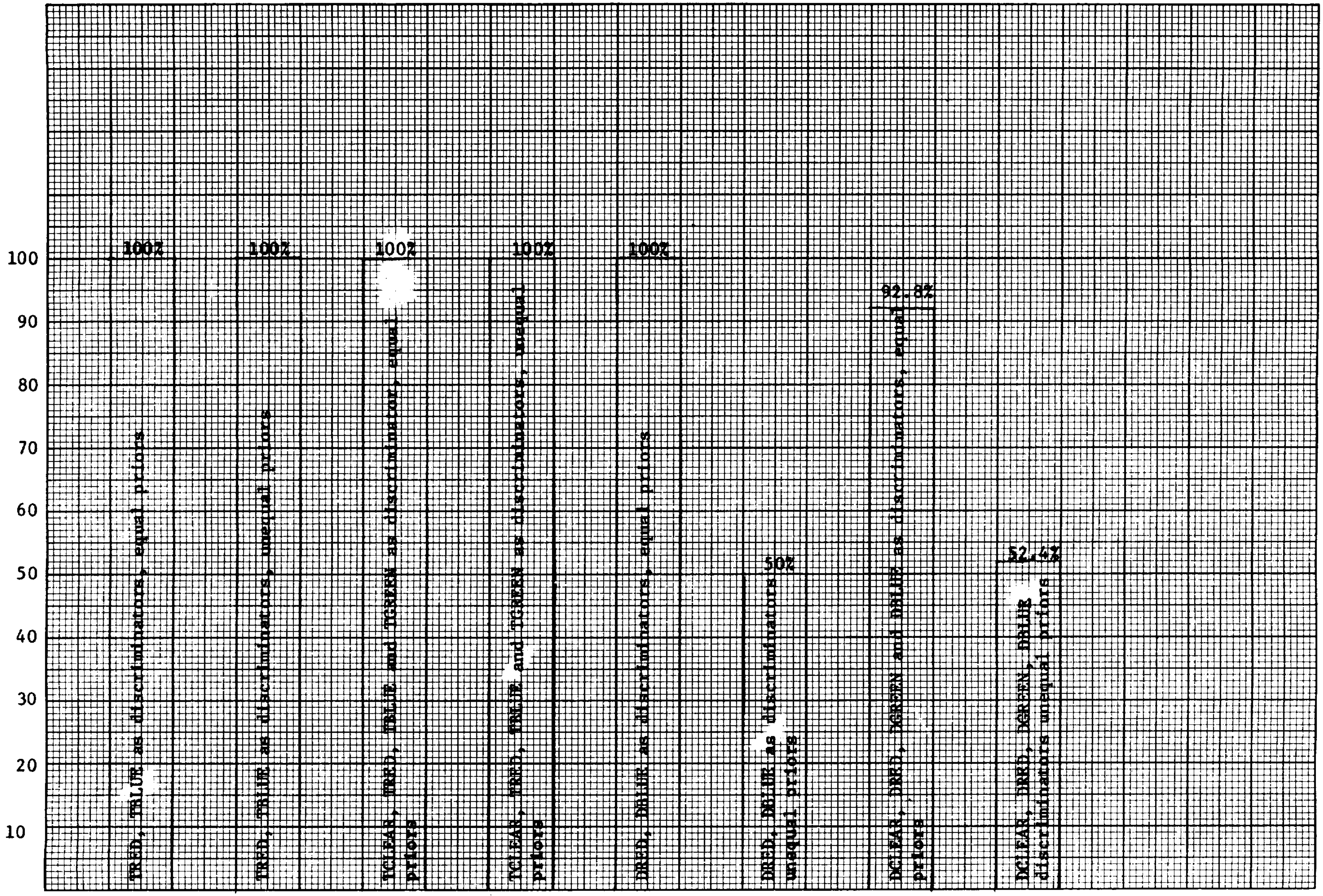


Figure 8.--Percent correct recognition of orange pixels for Classification Strategy A using training data of DATA SET 1 by equal/unequal priors and combination of variables.

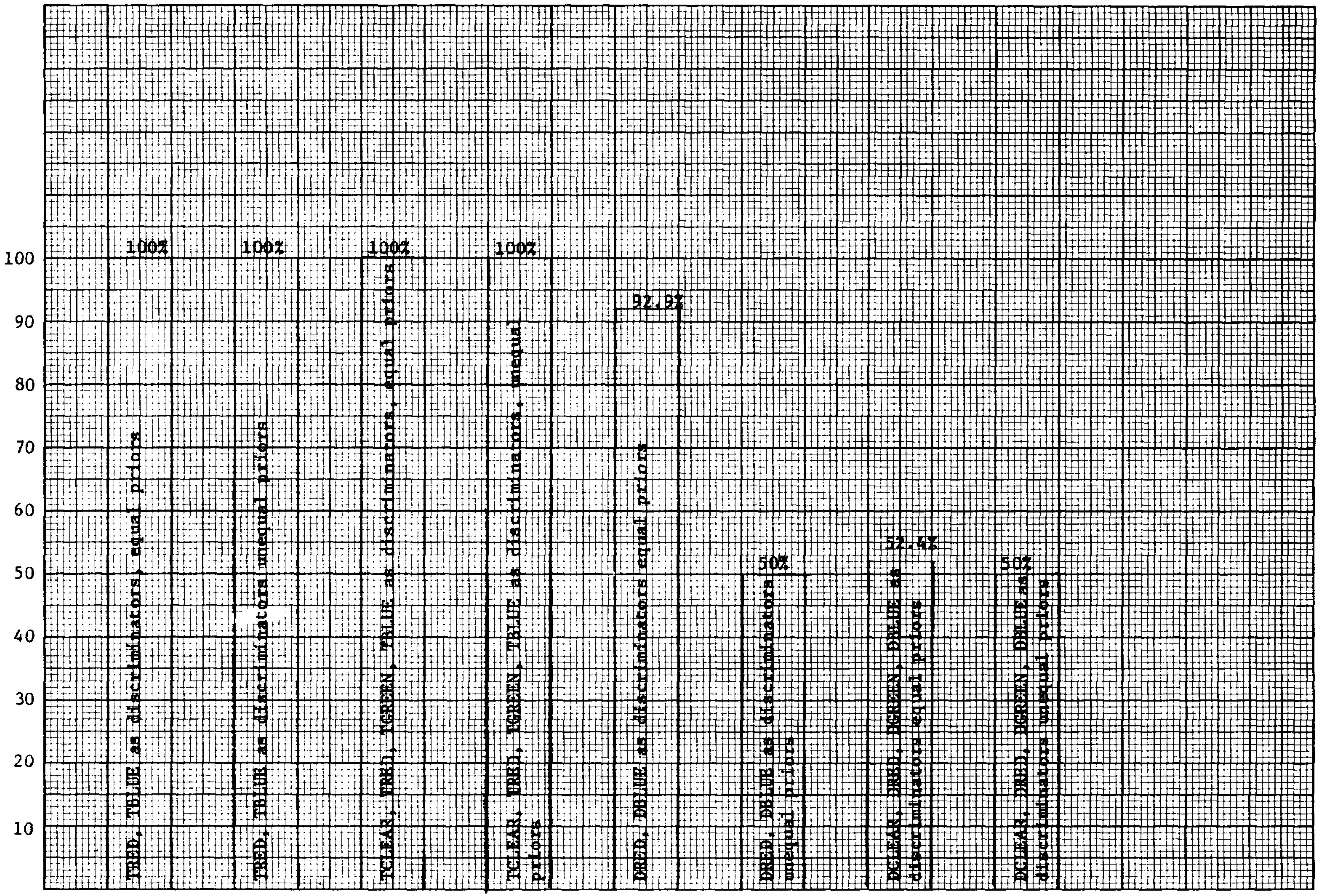


Figure 9.--Percent correct recognition of orange pixels for Classification Strategy B using training data of DATA SET 1 by equal/unequal priors and combination of variables.

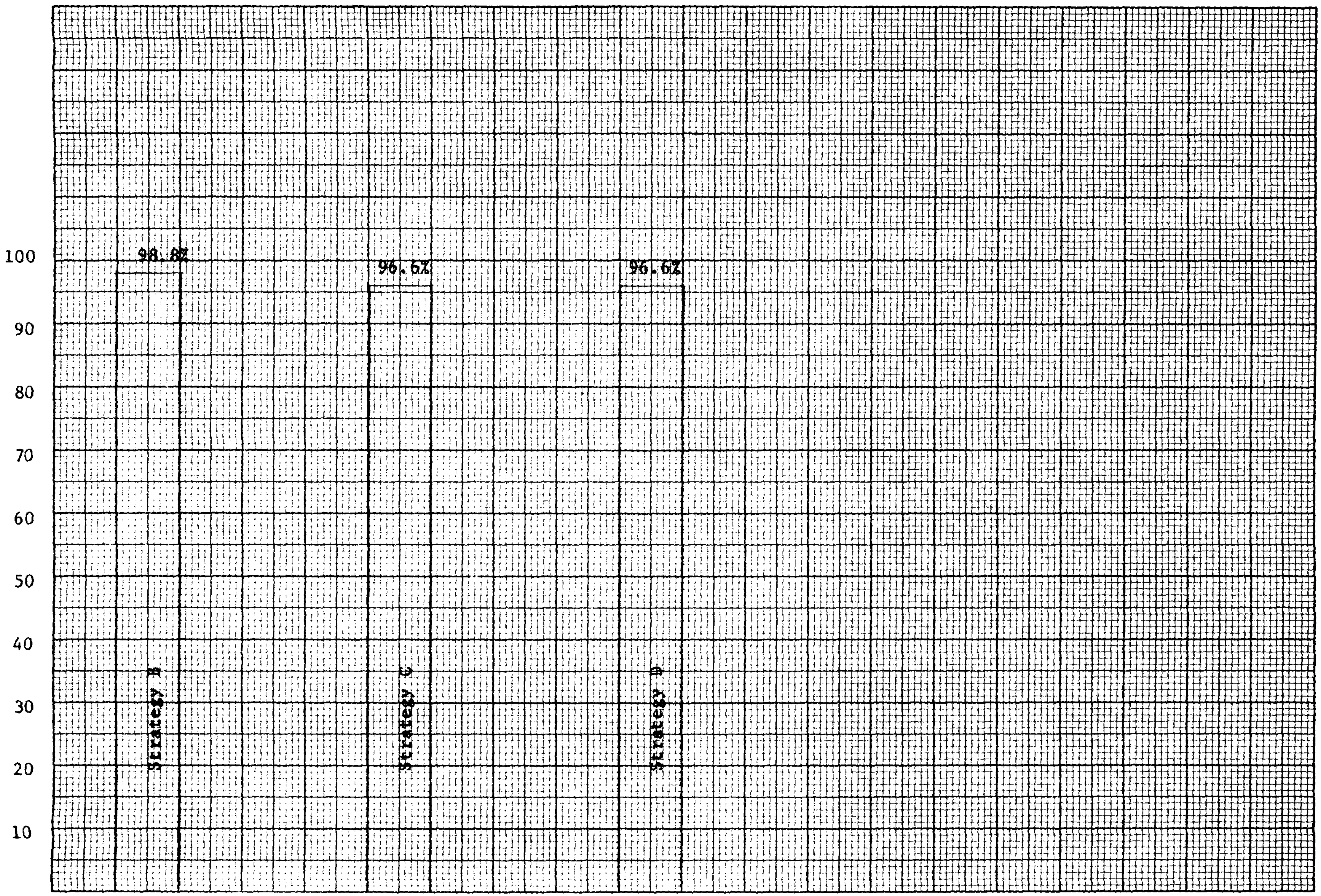


Figure 10.--Conditional percent correct recognition of tree pixels for some classification strategies in Figures 6 and 7.

Appendix 2 gives the tables of the classification results using training data as test data for some of the different classification strategies developed and summarized in Figures 8-10.

One of the considerations on which classification strategy to use among the ones presented was computer software development. Most of the software used in this research was done using the Statistical Analysis System [10]. However, the original SAS discriminant analysis routine did not have the capability to save on a file the classification group of individual pixels. Nevertheless, the classification results were printed. The relationship of two group discriminant analysis on a dummy variable and regression analysis is well known (see for example Kshirsagar [24]). Appendix 3 gives the correspondence between these two techniques. Using the SAS regression program (PROC REGR), the predicted value of dummy Y variate given the values on the discriminant variables can be saved in a file for further processing. Since there were nearly 8000 points in Data Set 1, and 30,000 points in Data Set 2 which must be classified and then punched up again, if one used the original SAS discriminant analysis procedure, regression on a dummy variate was a quicker and cheaper alternative. Later, the SAS discriminant analysis program was modified to output the classification results onto a file for further processing. This allowed for the capability of using quadratic discriminant functions with unequal priors. Table 1 below summarizes the pixels classified into the group of interest in each of the data sets using a particular classification strategy.

DATA SET 1

CLASSIFICATION A USING TWO VARIABLES TRED, TBLUE, EQUAL PRIORS

NUMBER OF POINTS CLASSIFIED INTO

ORANGE	OTHER = (FOLIAGE, GROUND, SKY)
250	7776

DATA SET 2

NUMBER OF POINTS CLASSIFIED INTO VARIOUS
GROUPS USING CLASSIFICATION SCHEME C

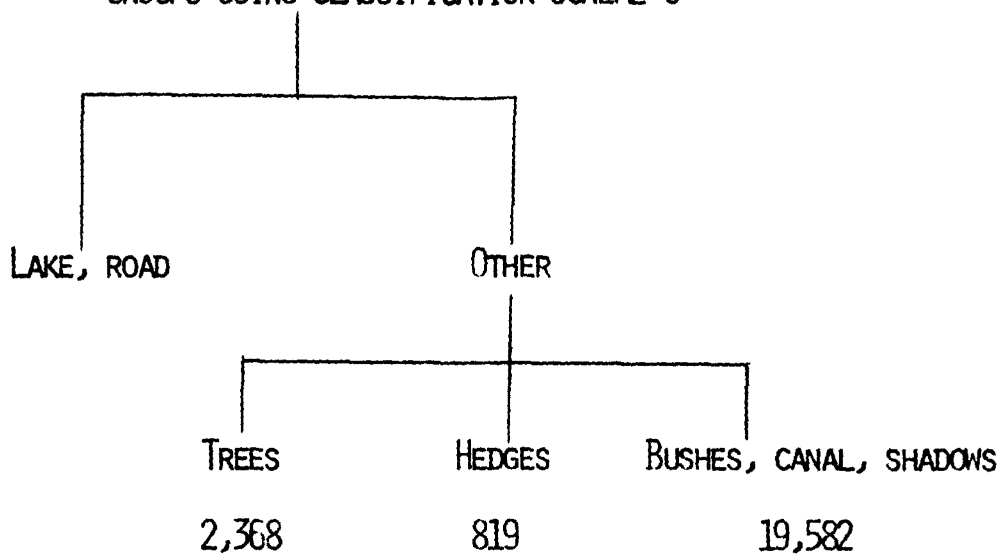


TABLE 1.--CLASSIFICATION OF DATA SET 1 AND 2 USING "BEST" CLASSIFICATION STRATEGY.

The data in each data set has been classified into groups using discriminant analysis on a subset of the eight spectral variables. We now need to cluster the pixels together to form "possible" fruit and "possible" trees. This will be discussed in the next section on Spatial Clustering.

<u>Spatial Clustering</u>

Cluster analysis is an analysis technique which is undergoing continuous refinement and research interest in recent years in both the statistical and pattern recognition community [31-41]. Clustering has been widely used as a data analysis technique in numerical taxonomy and the behavioral sciences by Jardine and Sibson [42], Sokal and Sneath [44] and Johnson [43] and others. Two recent books, one by Anderberg [45] and another, which is more software orientated, by Hartigan [46] give good introductory discussions on clustering. Other works by Ball [47] and Hall and Duda [11] also give excellent overviews.

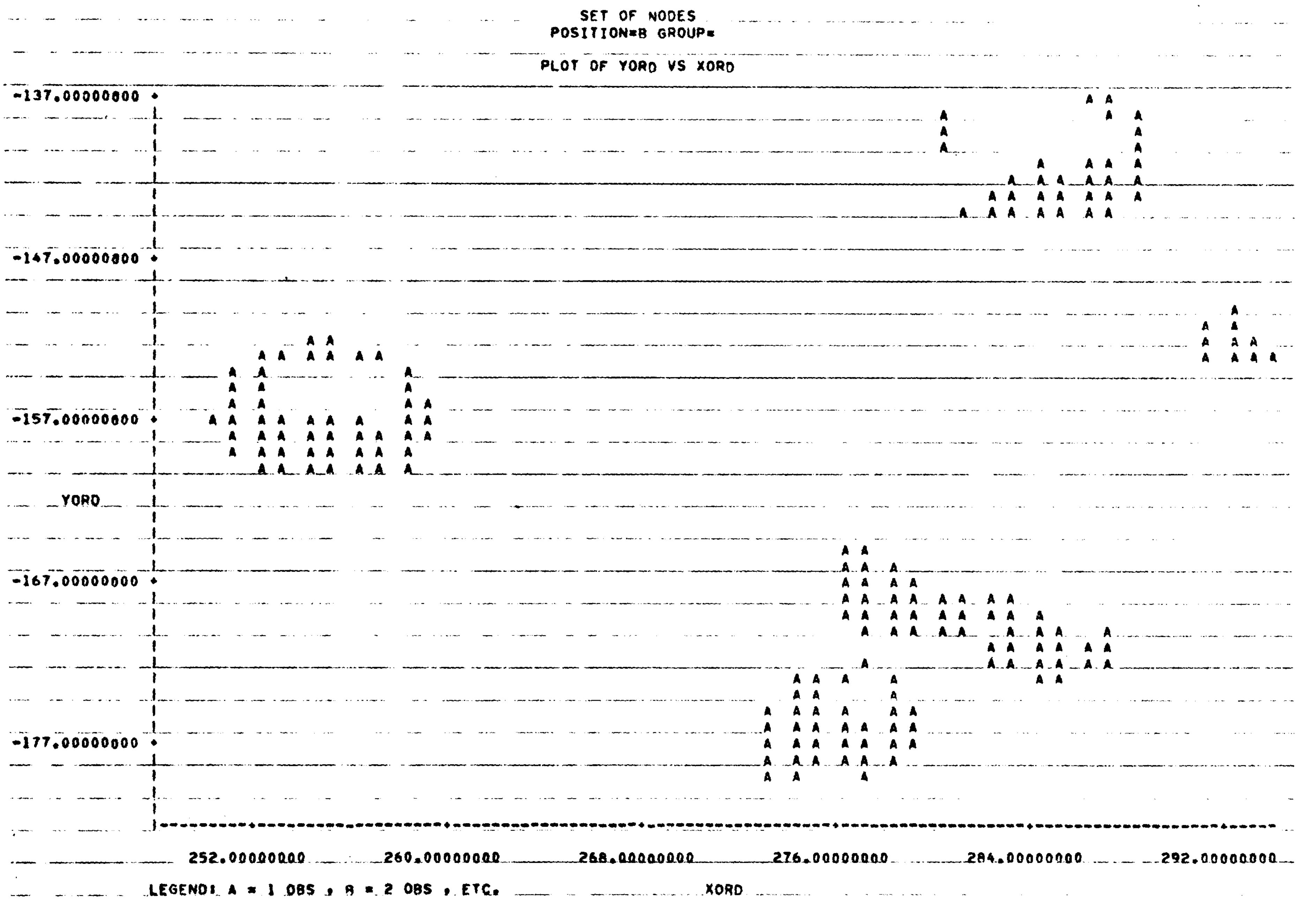


Figure 11: Spatial plot of Pixels Classified as Oranges in Data Set 1 using Strategy A.

The basic objective of clustering is to find structure or "natural groupings" in multivariate data when no apriori information is known concerning the structure of the data set(s) being investigated. For example, in an earlier section, we visually clustered the "sky" and "ground" and "foliage" pixels in Data Set 1 together into one merged group for classification. The process of determining spectral groups for classification is a cluster analysis problem.

Several different approaches have been suggested for clustering data in the pattern recognition community. One of the first such algorithms was ISODATA by Ball and Hall [48, 49]. This clustering procedure has been implemented on several computer systems for grouping multivariate remote sensor data. Our problem is somewhat different, but the ideas of grouping are similar, we wish to cluster data together based on spatial relationships. Figure 11 is the result of classifying Data Set 1 using classification Stragegy A in Figure 5.

The data in Figure 11 was plotted using the two spatial variables (X and Y), the problem is how many clusters are there in this data? Of course we can see five distinct clusters in the data. We would like to program a computer to detect these clusters. One way to represent the points in Figure 11 is to view this data as a graph. A graph G as defined by Harary [50], is a set S of points (in any dimension) and set of unordered pairs o(S) defined on the set S such that each pair x = [u, v] of o(S) is a line of G. Figure 12 is an example of a graph.

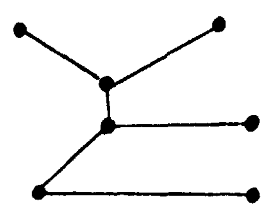


Figure 12.--Example of a graph

If the two dimensional points in Figure 11 are connected together then the problem of clustering the points becomes one of finding the connected components of a graph. Figure 13 represents the minimal spanning tree of the completely connected graph of Figure 11.

A spanning tree of a connected graph of n nodes (points or vertices) is a set of n-1 edges which provide one and only one path between any pair of nodes. If the edges have lengths associated with them (a weighted graph), then the minimal spanning tree (MST)is the shortest of all spanning trees that can be constructed from the given set of edges [46]. For our application, Euclidean distance is the length associated with an edge connecting two adjacent nodes. However, other metric distance functions can be used, as indicated by Zahn [51].

To determine the number of clusters, the MST is partitioned into connected components based upon a local edge "inconsistency" rule. An edge XY is "inconsistent" if within a neighborhood of δ nodes on both sides, of the edge, the following are both true:

(1) The edge length deviates by more than a constant number (say ST) of sample standard deviations from the average edge length in the neighborhood.

(2) If the edge's relative size to the average edge length in the neighborhood is greater than a constant (say FT).

An algorithm written in PL/I by Zahn [51], was used to construct the MST and partition the MST into cluster as defined by the inconsistency rule above. The algorithm was modified and implemented into a procedure in the Statistical Analysis System.[52]. The constants δ, ST, and FT are parameters which are specified for the clustering procedure to detect the clusters. Several other algorithms are available to compute a minimal spanning tree [53 and 54] of an arbitrary graph.

Notice, that this clustering algorithm does not require the number of cluster to be guessed at apriori and used as a clustering parameter. Most clustering procedures require an initial guess as to the number of clusters, and then use some criteria to either split or combine clusters to arrive at a final partitioning after an initial paritioning of the data is done into the "guessed number of clusters." The final number of clusters for these algorithms, that require the number of clusters to be guessed apriori, are very vulnerable to bad guesses as to the number of clusters to expect and the initial partitioning of the data into the "guessed number of clusters." The algorithm by Zahn [51] outlined above requires only the subtree depth to compute the local average length and standard deviation to delete "inconsistent" edges. No guess need be made as to the number of clusters to expect in the data. It has been shown by Gower and Ross [55] that there is a one-to-one correspondence between the MST and single linkage cluster analysis. Thus, the MST clustering procedure has all of the properties of single linkage schemes which are described by Fisher and Von Ness [56].

For the two data sets, Table 2 gives the number of clusters detected using MST clustering procedure. Figure 14 is a plot of the clusters detected in Data Set 1. The inconsistent edges are annotated in Figure 13, this is where the cut was made in the graph to detect the connected components.

From Figure 14, we can see that the clustering procedure detected one whole orange (cluster 1), a group of three oranges (cluster 5), a partial orange (cluster 4), and a split of a whole orange into clusters 2 and 3, this orange has a leaf partially covering it (See Figure 1). A procedure which will detect "whole" fruit clusters, partial fruit clusters, and groups of whole and partial fruit and count each fruit within this grouping was not done for this data set, since there was a very limited amount of data available for each type of grouping. However, graph-theoretic methods of characterizing the cluster shape could be applied to these data to count the number of either whole or partial fruit. These graph-

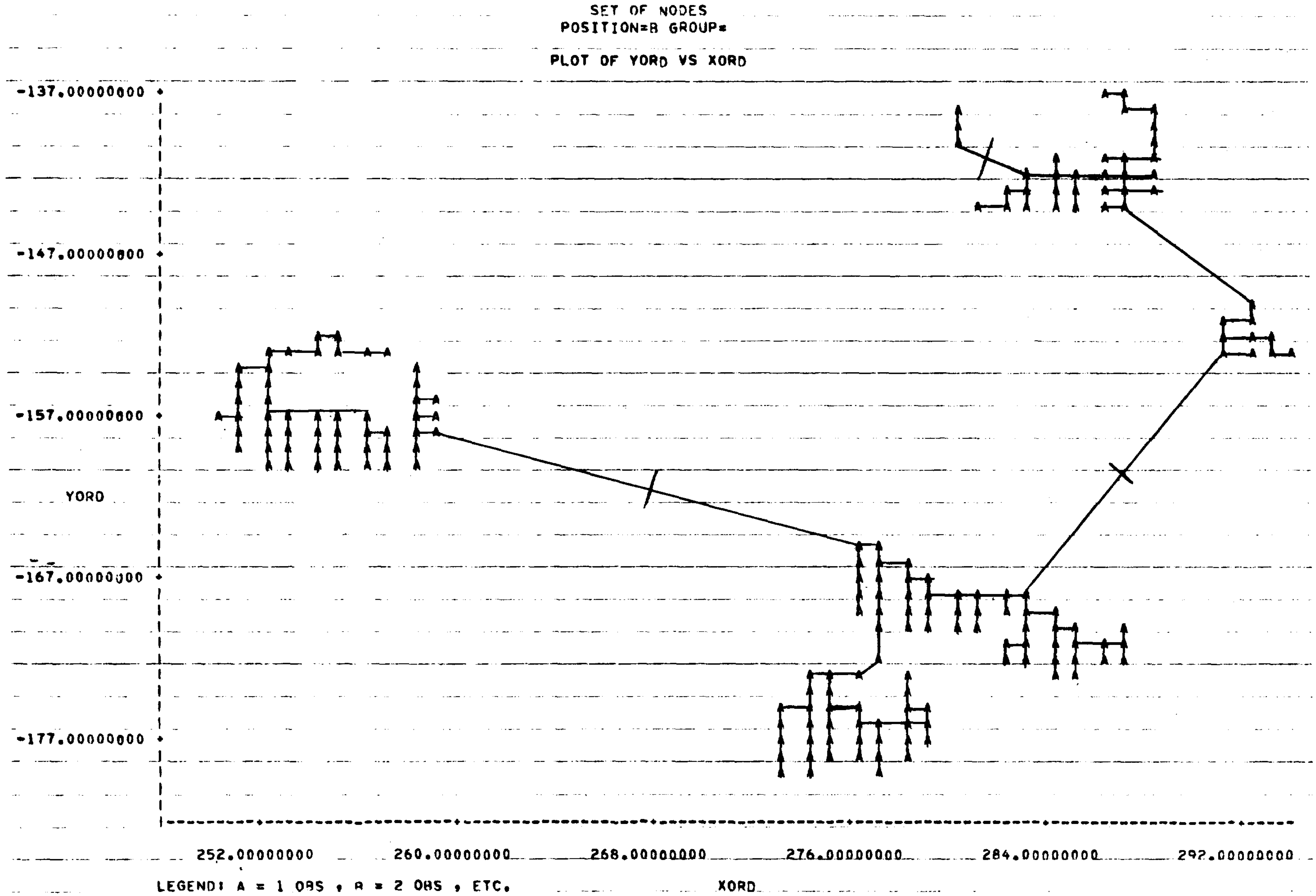


Figure 13: Minimal Spanning Tree of Pixels Classified as 'Oranges' in Data Set 1 using Strategy A.

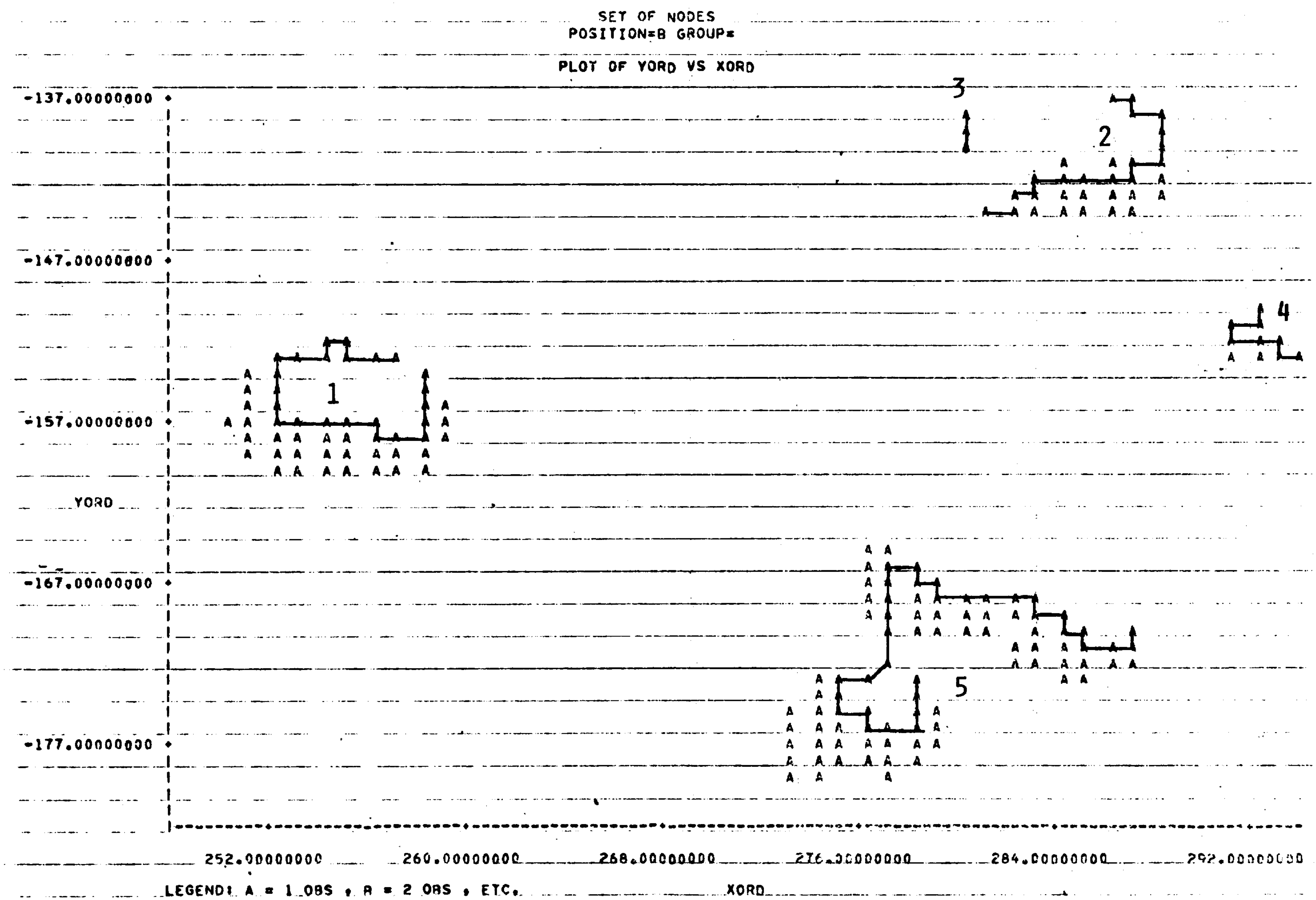


Figure 14: Clusters Detected and Counted by the MSTCLUS Algorithm for Data Set 1.

theoretic methods for shape characterization will be demonstrated with the trees in Data Set 2.

Data Set 1	Data Set 2
5 possible orange clusters	44 possible tree clusters

Table 2.--Clusters detected

We know that not all of the clusters in Data Set 2 represent trees, since there are spectral classification errors for bushes and other "tree" types which were confused with the trees in the orchard. We now wish to classify the clusters as either tree or non-tree in Data Set 2. This will be presented in the next section on Cluster Classification.

Cluster Classification

Non-tree clusters (or connected components) in Data Set 2 will tend to be long and serpentine and not circular, whereas tree clusters will tend to be circular. A measure of the shape characteristics of a graph is made on each cluster and is outputted as part of minimal spanning tree clustering procedure [51, 52]. These measures are as follows:

$$W = \begin{bmatrix} \text{Cluster average edge length} \\ \text{Cluster standard deviation} \\ \text{Number of points in the cluster} \\ \text{Cluster diameter} \end{bmatrix} \qquad (1)$$

The cluster edge length is a measure of the "compactness" of the cluster. The cluster standard deviation is a measure of the "spread" or "tightness" of the cluster. The number of points is a measure of the "size" of the cluster and the cluster diameter is a measure of the "breadth" of the cluster. These measurements characterize the cluster with respect to shape. These four measures were used as the measurement vector to perform a discrimianat analysis on the clusters.

First, two groups of clusters which are "trees" and "non-trees" are identified in the data. These clusters are then used as "training" clusters to perform another discriminant analysis using the characteristics measured in (1) above. After investigating two dimensional plots of the variables in (1) on the training clusters, only three variables in (1) proved to have all the discriminating ability.

They were:

Cluster average length

Cluster size (2)

Cluster diameter

These three measurements completely characterize the shape of the clusters in Data Set 2. The classification results of the training clusters are given in Table 3 for Data Set 2.

Class	Percent Correct	Number of clusters classified	
		Tree	Non-Tree
Tree	96.4	27	1
Non-Tree	96.5	1	28
Overall	96.4	Total Tree 28	Total Non-Tree 29

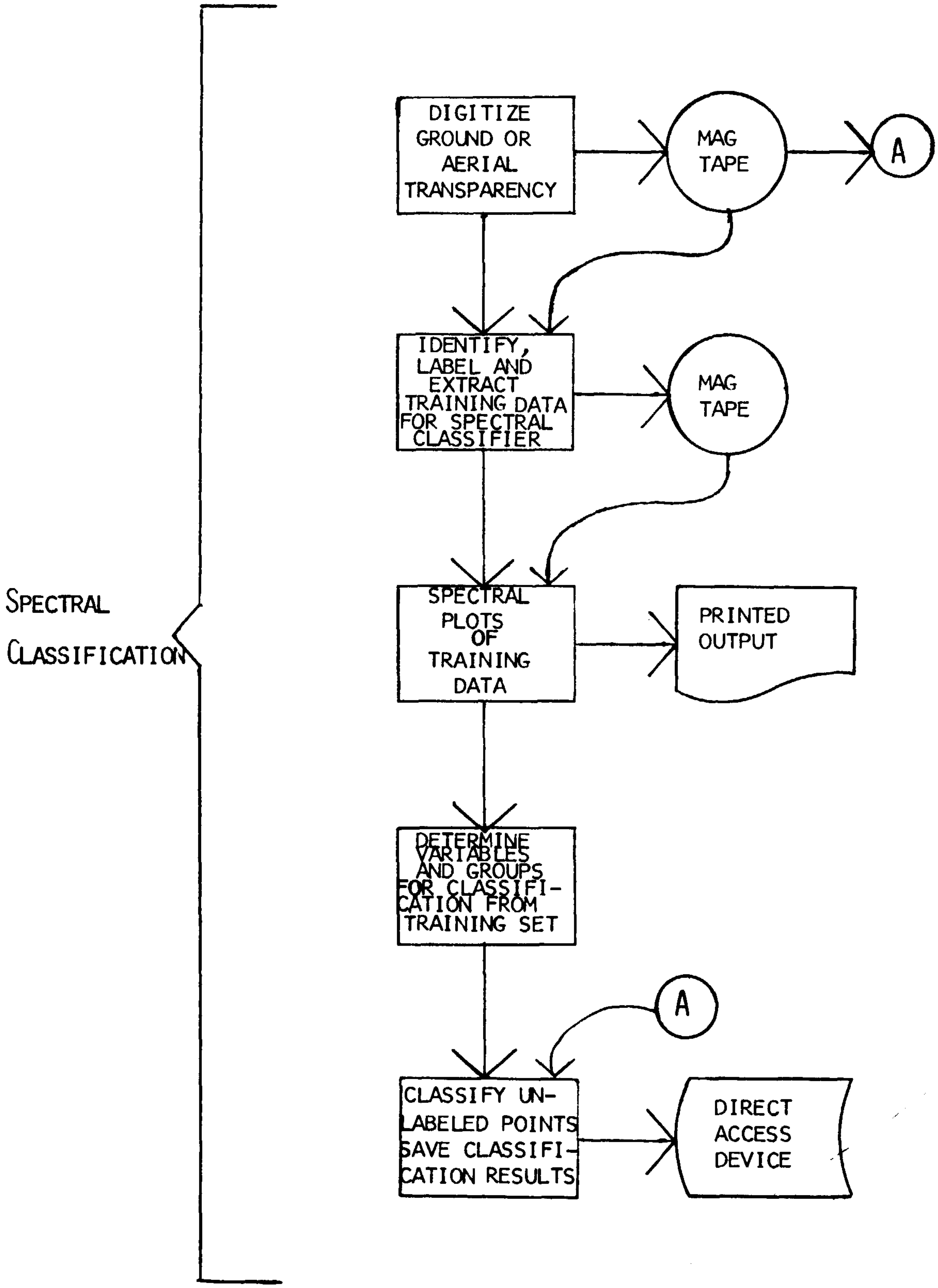
Table 3.--Classification results of tree/non-tree clusters in Data Set 2.

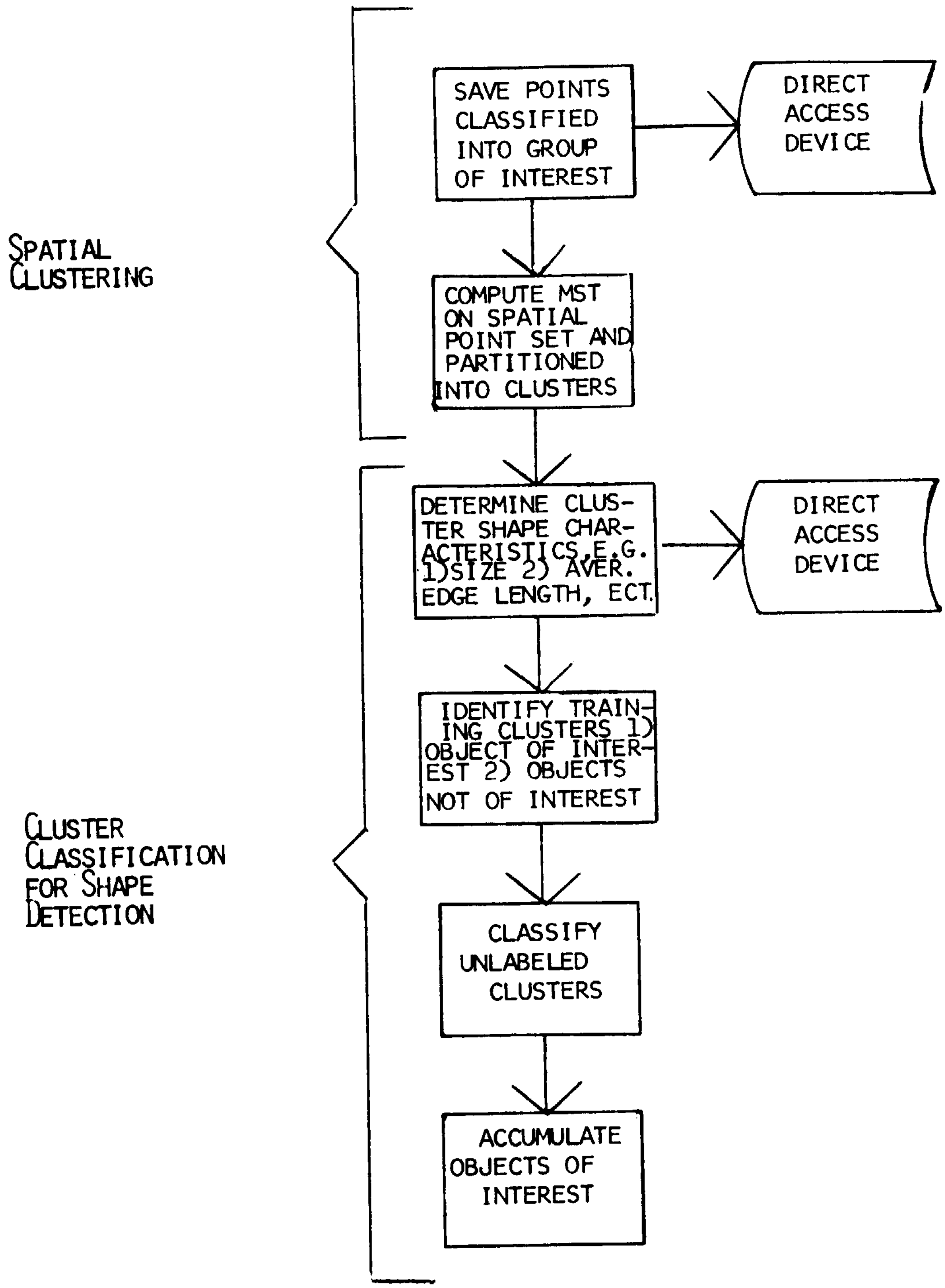
In this particular instance, compensating errors occurred and the number of trees counted was 28. As one can see in Figure 2, 28 was the number of trees present in Data Set 2. The tree cluster which was misclassified in Table 3 was the replanted seedling in the last row of trees in Data Set 2 (see Figure 2).

SUMMARY

A system was given which detects and counts objects of interest from digitized transparencies using a scanning microdensitometer. The components of the system are: (1) Spectral Classification, (2) Spatial Clustering, and (3) Cluster Classification. It was found that prior probabilities and the choice of measurement variables used in the classification procedure affected the point by point classification accuracy. Hierarchial classification techniques were outlined and applied to the two data sets. A graph theoretic clustering was demonstrated which requires no initial guess as to the number of clusters in the data. Finally, cluster shape characteristics were defined in graph theoretic terminology and applied to the cluster to detect the number of objects of interest.

Appendix 1: Sequential Classification and Clustering System Flow




SPATIAL CLUSTERING
SAVE POINTS CLASSIFIED INTO GROUP OF INTEREST
DIRECT ACCESS DEVICE
COMPUTE MST ON SPATIAL POINT SET AND PARTITIONED INTO CLUSTERS
CLUSTER CLASSIFICATION FOR SHAPE DETECTION
DETERMINE CLUSTER SHAPE CHARACTERISTICS,E.G. 1)SIZE 2) AVER. EDGE LENGTH, ECT.
DIRECT ACCESS DEVICE
IDENTIFY TRAINING CLUSTERS 1) OBJECT OF INTEREST 2) OBJECTS NOT OF INTEREST
CLASSIFY UNLABELED CLUSTERS
ACCUMULATE OBJECTS OF INTEREST

CLASSIFICATION RESULTS OF TRAINING DATA

DATA SET 1

CLASSIFICATION A USING TWO VARIABLES DRED, DBLUE

PRIOR PROBABILITIES	CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO	
			ORANGE	OTHER
EQUAL	ORANGE	100	42	0
	OTHER	99.1	15	1645
	OVERALL	99.1		
JTH PRIOR = RELATIVE FREQUENCY	ORANGE	50	21	21
	OTHER	99.8	4	1656
	OVERALL	98.5		

PRIOR PROBABILITIES	CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO	
			ORANGE	OTHER
EQUAL	ORANGE	92.8	39	3
	OTHER	98.5	25	1635
	OVERALL	98.4		
JTH PRIOR = RELATIVE FREQUENCY	ORANGE	52.4	22	20
	OTHER	99.6	7	1653
	OVERALL	98.4		

CLASSIFICATION RESULTS OF TRAINING DATA

DATA SET 1

CLASSIFICATION A USING TWO VARIABLES TRED, TBLUE

PRIOR PROBABILITIES	CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO	
			ORANGE	OTHER
EQUAL	ORANGE	100	42	0
	OTHER	99.9	1	1659
	OVERALL	99.9		
JTH PRIOR = RELATIVE FREQUENCY	ORANGE	100	42	0
	OTHER	100	0	1660
	OVERALL	100		

CLASSIFICATION A USING FOUR VARIABLES, TCLEAR, TRED, TBLUE, TGREEN

PRIOR PROBABILITIES	CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO	
			ORANGE	OTHER
EQUAL	ORANGE	100	42	0
	OTHER	99.8	3	1657
	OVERALL	99.8		
JTH PRIOR = RELATIVE FREQUENCY	ORANGE	100	42	0
	OTHER	99.8	3	1657
	OVERALL	99.8		

CLASSIFICATION RESULTS OF TRAINING DATA

DATA SET 1

CLASSIFICATION B USING TWO VARIABLES TRED, TBLUE

PRIOR PROBABILITIES	CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO FOLIAGE	GROUND	ORANGE	SKY
EQUAL	FOLIAGE	98.7	1204	14	0	2
	GROUND	95.5	9	210	0	1
	ORANGE	100.0	0	0	42	0
	SKY	100.0	0	0	0	220
	OVERALL	98.5				
JTH PRIOR = RELATIVE FREQUENCY	FOLIAGE	98.4	1200	18	0	2
	GROUND	96.4	7	212	0	1
	ORANGE	100.0	0	0	42	0
	SKY	100.0	0	0	0	220
	OVERALL	98.4				

CLASSIFICATION B USING FOUR VARIABLES TCLEAR, TRED, TGREEN, TBLUE

PRIOR PROBABILITIES	CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO FOLIAGE	GROUND	ORANGE	SKY
EQUAL	FOLIAGE	99.5	1214	6	0	0
	GROUND	98.6	3	217	0	0
	ORANGE	100.0	0	0	42	0
	SKY	100.0	0	0	0	220
	OVERALL	99.5				
JTH PRIOR = RELATIVE FREQUENCY	FOLIAGE	99.5	1214	6	0	0
	GROUND	98.2	4	216	0	0
	ORANGE	100.0	0	0	42	0
	SKY	100.0	0	0	0	220
	OVERALL	99.4				

CLASSIFICATION RESULTS OF TRAINING DATA

DATA SET 1

CLASSIFICATION B USING TWO VARIABLES DRED, DBLUE

PRIOR PROBABILITIES	CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO			
			FOLIAGE	GROUND	ORANGE	SKY
EQUAL	FOLIAGE	93.8	1144	70	5	1
	GROUND	94.5	0	208	0	12
	ORANGE	92.9	0	0	39	3
	SKY	100.0	0	0	0	220
	OVERALL	94.6				
JTH PRIOR = RELATIVE FREQUENCY	FOLIAGE	95.2	1161	52	4	2
	GROUND	94.5	0	208	0	12
	ORANGE	50.0	0	0	21	21
	SKY	100.0	0	0	0	220
	OVERALL	94.6				

PRIOR PROBABILITIES	CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO			
			FOLIAGE	GROUND	ORANGE	SKY
EQUAL	FOLIAGE	97.5	1189	22	6	3
	GROUND	99.6	0	219	0	1
	ORANGE	52.4	0	0	22	20
	SKY	100.0	0	0	0	220
	OVERALL	96.9				
JTH PRIOR = RELATIVE FREQUENCY	FOLIAGE	98.4	1200	14	3	3
	GROUND	99.6	20	219	0	1
	ORANGE	50.0	0	0	21	21
	SKY	100.0	0	0	0	220
	OVERALL	97.5				

CLASSIFICATION RESULTS OF TRAINING DATA

DATA SET 2

STAGE 1

USING DRED AS DISCRIMINATOR

CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO LAKE, ROAD	OTHER
LAKE, ROAD	100	432	0
OTHER	99+	1	1761

OVERALL = 99.9, OTHER= TREES HEDGES BUSHES CANAL SHADOWS

STAGE 2

USING DRED AND DBLUE AS DISCRIMINATORS

CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO LAKE, ROAD	OTHER
LAKE, ROAD	98.9	89	1
OTHER	98.8	20	1651

OVERALL = 98.8, OTHER= HEDGES BUSHES CANAL SHADOWS

CLASSIFICATION RESULTS OF TRAINING DATA

DATA SET 2

CLASSIFICATION C OF FIGURE 7

STAGE 1
USING DRED AS DISCRIMINATOR

CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO: LAKE, ROAD	OTHER
LAKE, ROAD	100	432	0
OTHER	99+	1	1761

OVERALL = 99.9, OTHER = TREES, BUSHES, CANAL, HEDGES, SHADOWS

STAGE 2
USING DRED AND DBLUE AS DISCRIMINATORS, UNEQUAL PRIORS

CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO: TREES	HEDGES	OTHER
TREES	96.7	87	3	0
HEDGES	90.0	2	19	0
OTHER	100	0	0	1650

OVERALL = 99.7, OTHER = BUSHES, CANAL, SHADOWS

CLASSIFICATION RESULTS OF TRAINING DATA
DATA SET 2
CLASSIFICATION D OF FIGURE 8

STAGE 1

CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO LAKE, ROAD	OTHER
LAKES, ROAD	100	432	0
OTHER	99+	1	1761

OVERALL = 99.9, OTHER = TREES, HEDGES, BUSHES, CANAL, SHADOWS

STAGE 2
USING DRED AND DBLUE AS DISCRIMINATORS

CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO TREES, HEDGES	BUSHES, CANAL, SHADOWS
TREES, HEDGES	100	111	0
BUSHES, CANAL SHADOWS	99+	1	1649

OVERALL = 99.9

STAGE 3
USING DRED AND DBLUE AS DISCRIMINATORS

CLASS	PERCENT CORRECT	NUMBER OF SAMPLES CLASSIFIED INTO TREES	HEDGES
TREES	96.7	87	3
HEDGES	95.2	1	20

OVERALL = 96.4

APPENDIX 3

Discriminant Analysis for Two Groups Using Multiple Regression on a Dummy Variate

Let $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_{n_1}$ be a sample of pxl component vectors observed on population π_1, and let $\underline{Y}_1, \underline{Y}_2, \ldots, \underline{Y}_{n_2}$ be a sample of pxl component vector observed on π_2.

Assume $\underline{X}_i \sim N(\mu_1, \Sigma)$, $\quad \underline{Y}_i \sim N(\mu_2, \Sigma)$, $\quad \mu_1 \neq \mu_2$

Let $\underline{d}' = (\overline{\underline{X}} - \overline{\underline{Y}})'$ where $\overline{\underline{X}} = \Sigma \underline{X}_i / n_1$, $\overline{\underline{Y}} = \Sigma \underline{Y}_i / n_2$

$$S = S_X + S_Y; \; S_X = \sum_{i=1}^{n_1} (\underline{X}_i - \overline{\underline{X}})(\underline{X}_i - \overline{\underline{X}})', \; S_Y = \sum_{i=1}^{n_2} (\underline{Y}_i - \overline{\underline{Y}})(\underline{Y}_i - \overline{\underline{Y}})'$$

Then it can be proven (See Anderson [23] or Kshirsagar [24]) that when the population parameters μ_1, μ_2, Σ are known that the linear function $f(X) = (\mu_1 - \mu_2)\Sigma^{-1}X$ is the best linear function to discriminate between π_1 and π_2.

To classify an unlabeled point $\underline{X}$ into π_1 or π_2 based on sample estimates of μ_1, μ_2, Σ the procedure is:

(1) Compute the linear function $z = \underline{\ell}'\underline{X}$, where $\underline{\ell} = (n_1 + n_2 - 2)\, S^{-1}\, \underline{d}$

(2) Calculate the decision boundary C_D where $C_D = 1/2\, \underline{\ell}'\, (\overline{\underline{X}} + \overline{\underline{Y}})$ (1)

(3) Classify $\underline{X} \epsilon \pi_1$ if $z - C_D \geq 0$

$\epsilon \pi_2$ if $z - C_D < 0$

This rule is Anderson's classification procedure.

Now consider the dummy variable

$$W = \begin{cases} 1 \text{ if a point is in } \Pi_1 \\ 0 \text{ if a point is in } \Pi_2 \end{cases}$$

It can be proven (See Kshirsagar [24] pg.206-209) that the regression of W on the p variables with both groups combined in a single sample yields estimates of the regression parameters β which are proportion to $S^{-1}\underline{d}$. Thus, the discriminant function and the regression function are the <u>same</u>, apart from a constant of proportionality, and this regression will also lead to the <u>same</u> classification procedure outlined in Rule A above. We will now determine this constant of proportionality.

Let $b^* = \begin{bmatrix} b \\ b_o \end{bmatrix}$ be a (p + 1) vector of the estimates of the β parameters in the regression of W on the p variables. Then the weights on the p variables $\underline{b}$ are given as follows:

$$\underline{b} = \frac{\dfrac{n_1 \cdot n_2}{n_1 + n_2}}{1 + \dfrac{n_1 \cdot n_2}{n_1 + n_2} \underline{d}' S^{-1} \underline{d}} \quad S^{-1}\underline{d} = k\ S^{-1}\underline{d} \tag{2}$$

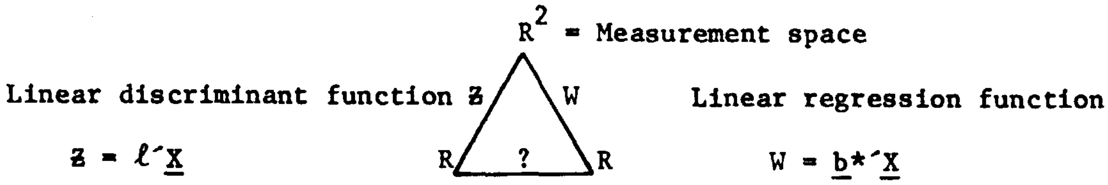
where

$$k = \left(\frac{n_1 \cdot n_2}{n_1 + n_2}\right) \Big/ \ \left(1 + \frac{n_1 \cdot n_2}{n_1 + n_2}\ \underline{d}' S^{-1} \underline{d}\right)$$

We recall that $\underline{\ell}$ in the linear discriminant function was $\underline{\ell} = f\ S^{-1}d$, $f = n_1 + n_2 - 2$. So as claimed above, the only difference between the regression weights $\underline{b}$ in (2) and the discriminant function weights $\underline{\ell}$ in (1) is the constants f and k.

This relationship allows one to save classification results in a file on a point by point basis using regression. This relationship is important since using SAS discriminant analysis procedure, the classification results cannot be saved on file. The procedure does print out the classification results point by point, but this would involve punching of the observations that were classified as "object of interest" for further analysis. However, the regression procedure in SAS allows the predicted value of an unlabeled point to be [illegible] in a file for later use, thus eliminating the manual search and punching of those points classified as an orange or a tree pixel in Data Set 1 and 2. This is a very important point from a data handling standpoint when several thousand pixels are classified into the group of interest. Corresponding to the decision boundary C_D in (1) the only thing we need to show is how to specify the decision bound C_R in discrimination space using the regression function.

The figure below depicts the situation.



We need to define the mapping from R onto R indicated by the question mark. We know the decision boundary for the linear discriminant function is

$$C_D = 1/2 \cdot \underline{\ell}' (\underline{\bar{X}} + \underline{\bar{Y}}), \text{ from (1).}$$

Now,

$$z = \underline{\ell}'\underline{X} = f(S^{-1}\underline{d})'\underline{X} = \frac{f}{k}\,\underline{b}'\underline{X} \quad (\text{since } \underline{b} = kS^{-1}\underline{d})$$

But,

$$\underline{b}'\underline{X} = \hat{W} - b_0 = \hat{W} - (W - \underline{b}' \left[\frac{n_1\underline{\bar{X}} + n_2\underline{\bar{Y}}}{n_1 + n_2}\right])$$

So,

$$z = \frac{f}{k}(W - (W - \underline{b}' \left[\frac{n_1\underline{\bar{X}} + n_2\underline{\bar{Y}}}{n_1 + n_2}\right]))$$

Thus, classify

$$\underline{X}\epsilon\pi_1, \text{ if } z - C_D = \frac{f}{k}(\hat{W} - \underline{b}' \left[\frac{n_1\underline{\overline{X}} + n_2\underline{\overline{Y}}}{n_1 + n_2}\right]) - 1/2\, f(S^{-1}\underline{d})'(\underline{\overline{X}} + \underline{\overline{Y}}) > 0$$

$$\text{iff } \frac{f}{k}(\hat{W} - (\overline{W} - \underline{b}' \left[\frac{n_1\underline{\overline{X}} + n_2\underline{\overline{Y}}}{n_1 + n_2}\right])) - \frac{f}{2k}\,\underline{b}'\,(\underline{\overline{X}} + \underline{\overline{Y}}) > 0$$

$$\text{iff } \hat{W} - (W - b' \left[\frac{n_1\underline{\overline{X}} + n_2\underline{\overline{Y}}}{n_1 + n_2}\right]) - \frac{\underline{b}'(\overline{X} + \overline{Y})}{2} > 0$$

$$\text{iff } \hat{W} - \overline{W} + \underline{b}' \left[\frac{n_1\underline{\overline{X}} + n_2\underline{\overline{Y}}}{n_1 + n_2}\right] - \frac{\underline{b}'(\overline{X} + \overline{Y})}{2} > 0$$

$$\text{iff } \hat{W} - \overline{W} + b' \left(\left[\frac{n_1\underline{\overline{X}} + n_2\underline{\overline{Y}}}{n_1 + n_2}\right] - \frac{\underline{\overline{X}} + \underline{\overline{Y}}}{2}\right) > 0$$

$$\text{iff } \hat{W} - (\overline{W} - b' \left[\frac{n_1\underline{\overline{X}} + n_2\underline{\overline{Y}}}{n_1 + n_2} + \frac{\underline{\overline{X}} + \underline{\overline{Y}}}{2}\right]) > 0$$

Thus,

$$C_R = \overline{W} - \underline{b}' \left(\frac{n_1\underline{\overline{X}} + n_2\underline{\overline{Y}}}{n_1 + n_2} - \frac{\underline{\overline{X}} + \underline{\overline{Y}}}{2}\right) \qquad (3)$$

is the decision boundary for the regression classifier, and Rule A is equivalent to the following rule:

$$
\text{Rule B}\begin{cases}
(1) & \text{Regress } W \text{ on the p discriminatory variables and predict } \hat{W} = \underline{b}'\underline{X} \\
(2) & \text{Calculate the decision boundary } C_R \text{ given in (3).} \\
(3) & \text{Classify } \underline{X}\ \epsilon\pi'_1, \text{ if } \hat{W} - C_R > 0 \\
 & \qquad\qquad\quad \epsilon\pi_2, \text{ if } \hat{W} - C_R \leq 0
\end{cases}
$$

Note that $\frac{n_1\underline{\bar{X}} + n_2\underline{\bar{Y}}}{n_1 + n_2}$ is the vector of means for all variables from the sample without any grouping and $\frac{\underline{\bar{X}} + \underline{\bar{Y}}}{2}$ is the mean of the two group means.

We have shown that the two rules are equivalent and thus one can classify unlabeled pixels into groups using regression procedure of SAS which will save classification results on a file.

REFERENCES

[1] Nealon, J.P., "The System of Sequential Classification and Clustering", Research and Development Branch, Research Division, Statistical Reporting Service, Washington, D.C., June 1975.

[2] Huddleston, H.F., "Use of Photography in Sampling for Number of Fruit per Tree", Agricultural Economics Research, Volume 23, No. 3, page 63.

[3] Huddleston, H.F. and Wigton, W.H., "Use of Remote Sensing in Sampling for Agricultural Data", Invited Paper at Joint Meeting of the International Statistical Institute of Survey Statisticians, Warsaw, Poland, September 1975.

[4] Wilson, W.W., Von Steen, D.H. and Hurt, P.V., "An Evaluation of Remote Sensing Data for Estimating Livestock Inventories", Research and Development Branch, Research Division, Statistical Reporting Service, Washington, D.C., January 1972.

[5] Ridgely, J.E., "Livestock Counts on Negative and Positive Transparencies", Research and Development Branch, Research Division, Statistical Reporting Service, Washington, D.C., September 1973.

[6] Bosecker, R.R., Ridgely, J.E. and Sestak, J., "Estimation and Forecasting of Citrus Fruit Production in the Lower Grande Valley", Research and Development Branch, Research Division, Statistical Reporting Service, Washington, D.C., November 1973.

[7] Himmelberger, D.V. and Steele, R.J., "Preliminary Report on Automated Fruit Counting for Photographs", Research and Development Branch, Research Division, Statistical Reporting Service, Washington, D.C., February 1974.

[8] Photometric Data Systems Microdensitometer System Installation, Operation and Maintenance Manual, Perkin-Elmer Corporation,: Pasedena, California.

[9] Hopkins, P.D., "PDSCMS, A Program to Convert PDS Microdensitometer Scan Lines into a SAS Compatible Form", Research and Development Branch, Research Division, Statistical Reporting Service, Washington, D.C., January 1974.

[10] Barr, A.J. and Goodnight, J.H., Service J. (Editor), "A Users Guide to the Statistical Analysis System", Department of Statistics, North Carolina State University, Raleigh, North Carolina, August 1972.

[11] Hall, D.J. Duda, R.O., Hoffman, D.A. and Wolf, D.E., "Development of New Pattern-Recognition Methods", Aerospace Research Laboratories, Air Force Systems Command, United States Air Force, Wright-Patterson Air Force Base, Ohio 45433, November 1973.

[12] Swain, P., Pattern Recognition: Laboratory for Applications of Remote Sensing, "A Basis for Remote Sensing Data Analysis", Purdue University, West Lafayette, Indiana, November 1973.

[13] Nagy, G., "State of the Art in Pattern Recognition", Proc. of the IEEE: Volume 56, No. 5, pages 836-862, May 1968.

[14] Ho, Yu-Chi and Agranala, Ashok, K., "On Pattern Classification Algorithms Introduction & Survey", Proc. of the IEEE: Volume 56, No. 12, pages 210-214, December, 1968.

[15] Wee, W.G., "On Feature Selection in a Class of Distribution Free Pattern Classifiers", IEEE: Transactions on Information Theory, Volume IT-16, No. 1, pages 47-55, January 1970.

[16] Weiner, J.M. and Dunn, O.J., "Elimination of Variates in Linear Discriminant Problems", Biometrics, Volume 22, pages 268-275, 1966.

[17] Davies, M.G., "The Performance of the Linear Discriminant Function in Two Variables", British Journal of Mathematical and Statistical Psychology, Volume 23, Part 2, pages 165-176, November 1970.

[18] Cochran, W.G., "On the Performance of the Linear Discriminant Function", Technometrics, Volume 6, No. 2, pages 179-190, May 1964.

[19] Rao, C.R., "Covariance Adjustment and Related Problems in Multivariate Analysis", Multivariate Analysis, Edited by P.R. Krishnaiah, pages 87-103, Academic Press: New York, 1966.

[20] Cacoullos, T., Discriminant Analysis and Applications, Academic Press: New York, 1973.

[21] Dixon, W.J. and Jenrich, R.I., "Computer Geographical Analysis and D s-crimination", Discriminant Analysis and Applications, Edited by T. Cacoullos, Academic Press: New York, 1973.

[22] Fisher, L. and Van Ness, J.W., "Admissible Discriminant Analysis", JASA, Volume 68, No. 343, pages 603-607, September 1973.

[23] Anderson, T.W., An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, Inc.: New York, 1958.

[24] Kshirsagar, A.M., Multivariate Analysis, Deeker, 1972.

[25] Lachenbrach, P.A., Discriminant Analysis, Hafner Press: New York,1975.

[26] Lachenbrach, P.A. and Mickey, M.R., "estimation of Error Rates in Discriminant Analysis", Technometrics, Volume 10, No. 1, pages 1-11, February 1968.

[27] Cover, T.M. and Hart, P.E., "Nearest Neighbor Pattern Classification", IEEE: Transactions on Information Theory, Volume IT-13, No. 1, pages 21-27, Jauuary 1967.

[28] Bennett, J.O., Rui, J.P., De Figueiredo and Thompson, J.R., "Classification by Means of B-Spline Potential Functions with Application to Remote Sensing", Proceedings of the Sixth South Eastern Symposium on System Theory, Baton Rouge, Louisiana.

[29] Wu, C.L., Langrebe, D.A. and Swain, P.H., "The Decision Tree Approach to Classification", Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, April 1975.

[30] Wu, C.L., Swain, P.H. and Langrebe, D.A., "Two Approaches to a Decision Tree Classification", Laboratory for Applications of Remote Sensing, Purdue University, West Lafayette, Indiana, April 1975.

[31] Bonner, R.E., "On Some Clustering Techniques", IBM Journal of Research and Development, Volume 8, pages 22-32, January 1964.

[32] Fleiss, J.L. and Zuben, J., "On the Methods and Theory of Clustering", Multivariate Behavioral Research, Volume 4, pages 235-250, April 1969.

[33] Gower, J.C., "A Comparison of Some Methods of Cluster Analysis", Biometrics, Volume 23, No. 4, pages 623-637, December 1967.

[34] Fortier, J.J. and Solomon, H., "Clustering Procedures", Multivariate Analysis, Edited by P.R. Krishnaiah, pages 493-506, Academic Press, New York, 1966.

[35] King, B., "Step-wise Clustering Procedures", Journal American Statistical Association, Volume 62, pages 86-101, March 1967.

[36] Sebestyen, G. and Edie, J., "An Algorithm for Non-Parametric Pattern Recognition", IEEE: Transactions on Electronic Computers, Volume EC-15, pages 908-915, December 1966.

[37] Maronna, R. and Jacorkis, P.M., "Multivariate Clustering Procedures with Variable Metrics", Biometrics, Volume 30, pages 499-505, September 1974.

[38] Rand, W.M., "Objective Criteria for the Evaluation of Clustering Methods", Journal of the American Statistical Association, Volume 66, No. 336, pages 846-850, December 1971.

[39] Eigen, D.J., Fromm, F.R. and Northouse, R.A., "Cluster Analysis Based on Dimensional Information with Applications to Feature Selection and Classification", IEEE: Transactions on Systems, Man and Cybernetics, Volume SMC-4, No. 3, pages 284-294, May 1974.

[40] Eigen, D.J. and Northouse, R.A., "Initial Considerations of Unsupervised Discrete Clustering", TR-CS72-1, The University of Wisconsin, Milwaukee,

[41] Duda, R.O. and Hart, P.E., Pattern Classification and Scene Analysis, John Wiley and Sons: New York, 1973.

[42] Jardine, N. and Sibson, R., "The Construction of Hierarchial and Nonhierarchic Classifications", Computer Journal, Volume II, No. 2, pages 177-184, February 1968.

[43] Johnson, S.C., "Hierarchial Clustering Schemes", Psychometrics, Volume 32, No. 3, pages 241-254, September 1967.

[44] Sokal, P.R. and Sneath, P.H., Principles of Numerical Taxonomy, San Francisco, Freeman, 1963.

[45] Anderberg, M., Cluster Analysis for Applications, Academic Press, New York, 1973.

[46] Hartigan, J.A., Cluster Algorithms, John Wiley and Sons, New York, 1975.

[47] Ball, G.H., "Classification Analysis", Tech. Note AD 716482, Stanford Research Institute, Menlo Park, California 1970.

[48] Ball, G.H. and Hall, D.J., "A Clustering Technique for Summarizing Multivariate Data", Behavorial Science, Volume 12, No. 2, pages 153-155, 1967.

[49] Ball, G.H. et al., "Isodata, A Novel Method of Data Analysis and Pattern Classification", AD 699616, Stanford Research Institute, Menlo Park, California, April 1965.

[50] Harary, F., Graph Theory, Addison Wesley, Reading, Mass, 1969.

[51] Zahn, C.T., "Graph Theoretic Methods for Detecting and Describing Gesalt Clusters", IEEE: Transactions on Computers, Volume C-20, No. 1, January 1971.

[52] Gleason, C.P., MSTCLUS 1.0.0 PROCEDURE Minimal Spanning Tree Cluster Analysis, Research and Development Branch, Research Division, Statistical Reporting Service, Washington, D.C.

[53] Rohlf, J.F., "Hierachical Clustering Using the Minimal Spanning Tree", Computer Journal, Volume 16, pages 93-95, February 1973.

[54] Whitney,M. and Kevin, V., "Algorighm 422 Minimal Spanning Tree [H]", Communications of the ACM, Volume 15, No. 4, pages 273-274, April 1972.

[55] Gower, J.C. and Ross, G.J.S., "Minimal Spanning Trees and Single Linkage Cluster Analysis", Applied Statistics, Volume 18, No. 1, pages 54-64, 1969.

[56] Fisher, L. and Von Ness, J.W., "Admissible Clustering Procedures", Biometrika, Volume 58, No. 1, pages 91-104, 1971.