THE SYSTEM OF SEQUENTIAL CLASSIFICATION, CLUSTERING, AND COUNTING FRUIT FROM DIGITALIZED GROUND PHOTOGRAPHS

Research and Development Branch
Research Division
Statistical Reporting Service
U. S. Department of Agriculture

October 1975

THE SYSTEM OF SEQUENTIAL CLASSIFICATION, CLUSTERING, AND COUNTING FRUIT FROM DIGITALIZED GROUND PHOTOGRAPHS

by

Jack Nealon

TABLE OF CONTENTS

INTRODUCTION

The purpose of this research effort is to determine if various types of fruit at different stages of development can be counted from digitalized ground photographs by the system of sequential classification, clustering, and counting.

The fruit to be analyzed are apples and peaches. The apples have transcended from their immature stage, which is characteristic of green colored apples, to the stage when apples are yellow in color. The peaches are in a more immature stage of development and are characterized by green and yellow peaches.

It was known by examination of the sections to be analyzed that eight apples were present on the apple tree and six peaches were present on the peach tree.

If accurate fruit counting can be produced by this system of sequential classification, clustering, and counting for various fruit before maturation has been attained, this system may prove beneficial for surveys to forecast crop yield.

DATA COLLECTION

Data Source:

Data were obtained by acquiring side-view ground photographs of an apple tree and peach tree and digitalizing a representative section on each photograph using the Photometric Data System (PDS) microdensitometer.[1] Representative sections were digitalized to minimize computer costs. Digitalized results obtained from the microdensitometer were stored on magnetic tape in a form that could be processed by software in the Statistical Analysis System (SAS).[2]

Scanning Parameters:

The aperture size and shape chosen to digitalize the selected sections for the apple tree and peach tree photographs was 80 microns square, which provided an excellent representation of the data.

All filter and scanning mode combinations were utlized when digitalizing the photographs so that multivariate responses for each pixel or data point could be examined. Four filters (clear, red, green, and blue) and two scanning modes (transmission and density) were available. Therefore, intensity readings for eight filter and scanning mode combinations for each pixel were recorded.

For the apple tree photograph, 41 pixel readings were digitalized per line. Forty-six lines were scanned, and therefore 1,886 pixels with readings for eight filter and scanning mode combinations were stored on magnetic tape.

There were 55 pixels digitalized per line on the peach tree photograph, and 49 lines scanned. Therefore, 2,695 pixels were stored with multivariate responses.

Groups:

The spectrally distinct groups represented on the digitalized section for the apple tree were: dark leaves (D), light leaves (L), and yellow apples (Y).

The digitalized section for the peach tree displayed a greater variety of spectrally distinct groups. The groups were: yellow peaches (Y), green peaches (G), light leaves (L), dark leaves((D), dark branches (B), sunny objects (W), shadows (S), and light branches (X).

Training Data:

Training data, which is also referred to as labeled data, were determined by selecting from each spectrally distinct group a collection of pixels. Selection of training data was made by recording on the PDS microdensitometer's teletype a set of boundaries in which a spectrally distinct group was known to exist. Training data were collected separately for the apple tree and peach tree from the digitalized sections.

APPROACH

The approach by which fruit on ground photographs was counted is identical to the approach taken to count fruit trees on aerial photographs.[3]

For each digitalized section training data were generated and analyzed to determine suitable discriminant functions for classifying unlabeled data. After unlabeled pixels were classified, pixels classified as fruit pixels were clustered and counted. If misclassifications of unlabeled data occurred, and therefore an excess of clusters resulted, training data from the clusters were examined to determine which clusters were fruit and which were not fruit.

DATA ANALYSIS

Feature Selection:

Filter(s) and scanning mode(s) utilized for the classification procedure of unlabeled pixels from the apple and peach tree photographs were selected by examining two-dimensional plots of all combinations of variables, which are all filter and scanning mode combinations.

As illustrated in Figure 1, the red filter in density and transmission scanning modes provided the best visual separation of yellow apple pixels from dark and light leaf pixels.

Training data for the digitalized section of the peach tree demonstrated that red and green filters in the transmission scanning mode separate yellow peach pixels and green peach pixels from the remaining pixels.(Figure 2)

In summary, density-red and transmission-red were the variables selected for the apple tree for discriminant analysis of the training data. Transmission-red and transmission-green were the variables chosen for the peach tree.

Discriminant Analysis Applied To Training Data:

(1) Apple Tree:

Once discriminant functions were obtained using intensity readings for each pixel in each group in the training data, training data were treated as unlabeled data to test the discriminant functions.

Using the variables chosen from the two-dimensional plot in Figure 1, quadratic discriminant functions with equal or unequal prior probabilities stated for the three groups provided perfect classification of each pixel in the training data. The classification matrix was:

	Dark Leaves	Light Leaves	Yellow Apples
Dark Leaves	22	0	0
Light Leaves	0	20	0
Yellow Apples	0	0	18

which provided perfect classification of the training data.

(2) Peach Tree:

To reduce computer costs, pixels with transmission-green readings greater than 170 or less than 63 were deleted from further analysis. This deletion eliminated dark leaf pixels, branch pixels and most sunny pixels. The remaining sunny pixels were combined with light leaf pixels because of their similarity in intensity readings. Inspection of Figure 2 shows the justification of this action. Obviously, the removal of these pixels does not hinder any analysis to be performed.

The prior probability of each group was based upon an approximation of the relative frequency of each group in the digitalized section. Quadratic discriminant functions based upon the chosen variables provided excellent classification of the training data. The matrix obtained for classification of the training data when training data were treated as "unlabeled" data was:

	Green Peaches	Yellow Peaches	Light Leaves and Sunny Objects	Shadows	Light Branches
Green Peaches	29	0	0	0	0
Yellow Peaches	0	61	0	0	0
Light Leaves and Sunny Objects	0	1	75	0	0
Shadows	0	0	0	4	0
Light Branches	0	0	0	0	3

which contained only one misclassification, which was a light leaf or sunny pixel classified as a yellow peach pixel.

Classification Results:

(1) Apple Tree:

The classification procedure involves classifying unlabeled data by the discriminant procedure chosen to classify the training data.

Since dark leaf pixels clearly separate from yellow apple pixels in Figure 1, pixels with density-red readings greater than 130 were deleted to reduce computer costs.

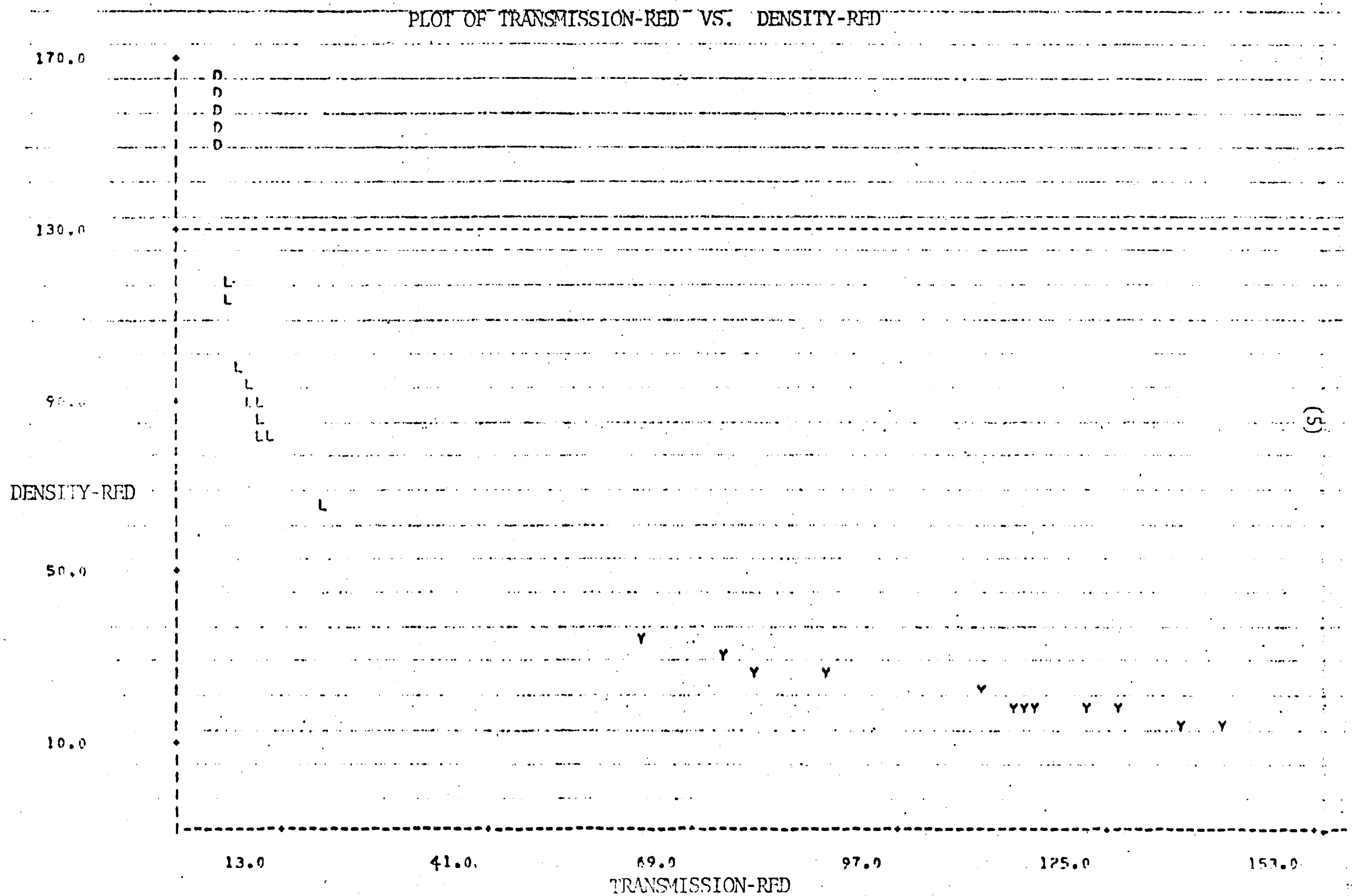
With this deletion the classification matrix obtained for unlabeled pixels digitalized on the apple tree photograph was:

	Light Leaves	Yellow Apples
Light Leaves	20	0
Yellow Apples	0	18
Unlabeled Data	1121	65

Therefore, 83 pixels have been classified as yellow apple pixels.

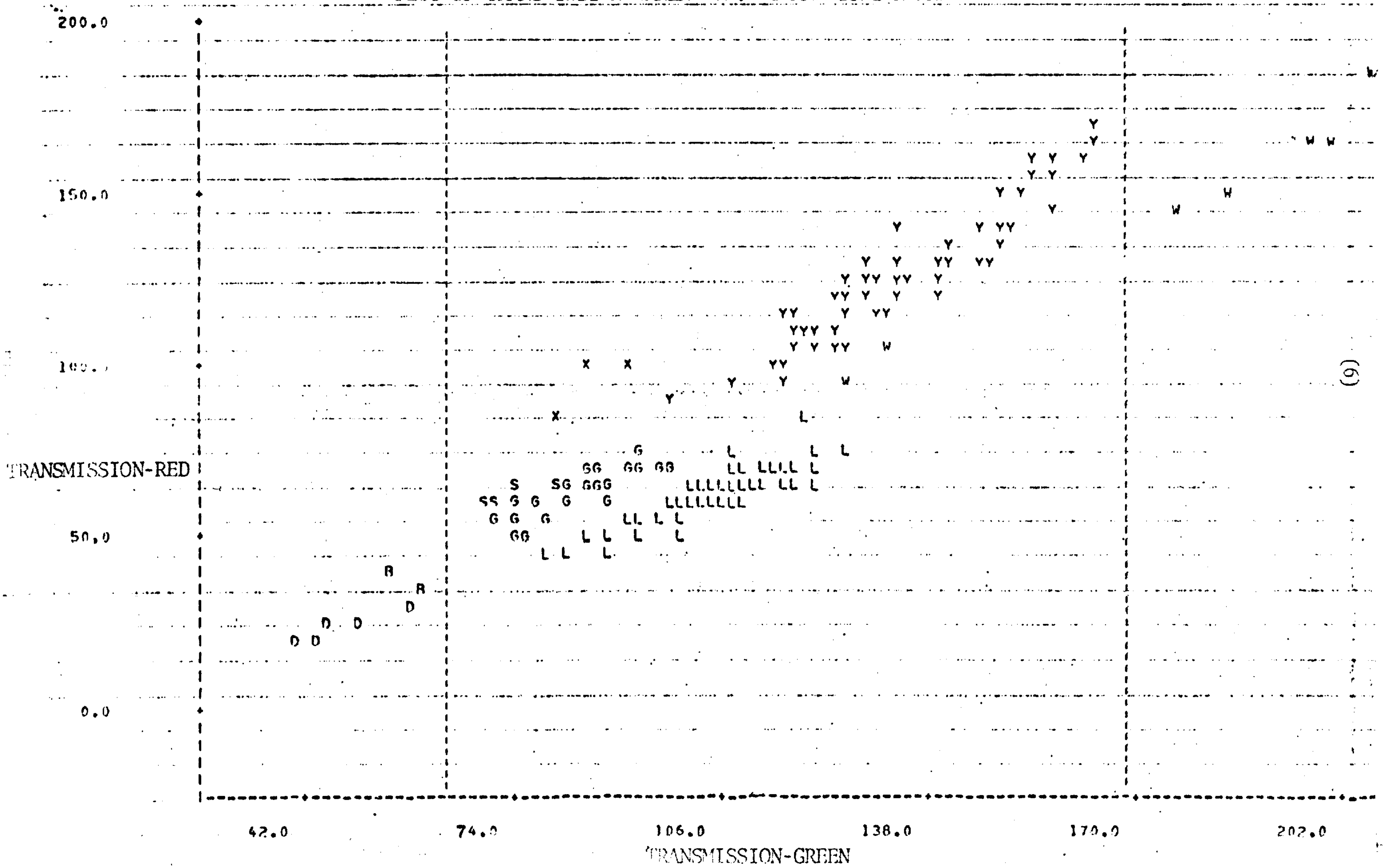
(2) Peach Tree:

The classification matrix for unlabeled pixels on the digitalized peach tree photograph was:



PLOT OF TRANSMISSION-RED VS. DENSITY-RED
170.0
130.0
90.0
50.0
10.0
DENSITY-RED
13.0
41.0
69.0
97.0
125.0
153.0
TRANSMISSION-RED
(5)

PLOT OF TRANSMISSION-GREEN VS. TRANSMISSION-RED



	Green Peaches	Yellow Peaches	Light Leaves and Sunny Objects	Shadows	Light Branches
Green Peaches	29	0	0	0	0
Yellow Peaches	0	61	0	0	0
Light Leaves and Sunny Objects	0	1	75	0	0
Shadows	0	0	0	4	0
Light Branches	0	0	0	0	3
Unlabeled Data	196	290	1887	8	15

A total of 225 pixels were classified as green peach pixels, and 352 were classified as yellow peach pixels.

Cluster Analysis:

(1) Apple Tree:

Procedure MSTCLUS in the Statistical Analysis System (SAS) was used to cluster classified yellow apple pixels and count the yellow apples. Clustering was performed on classified yellow apple pixels with respect to their x and y coordinates as recorded by the PDS microdensitometer.

Eight clusters were produced for classified yellow apple pixels by MSTCLUS, which is identical to the actual yellow apple count in the section under analysis. An examination of the two-dimensional plot with respect to the x and y coordinates of all classified pixels also shows that no misclassification occurred with respect to yellow apple pixels.

(2) Peach Tree:

As in the apple tree, MSTCLUS was executed with respect to the x and y coordinates of classified peach pixels to count the peaches. Since peaches were to be counted, yellow and green peach pixels were concatenated since some peaches contained both yellow and green peach pixels. Fifty-two peach clusters were obtained from the 577: (225 + 352) classified peach pixels. Therefore, some misclassifications had occurred since it was known a priori that only six peaches existed.

Using cluster size, which was generated by MSTCLUS, as the discriminator between peach clusters and non-peach clusters, peach clusters were easily separated from non-peach clusters. The criteria were: (1) If cluster size is greater than 35 pixels, then the cluster will be classified as a peach. (2) If cluster size is greater than 100 pixels, the cluster will be classified as two peaches.

Since over ninety percent of non-peach clusters were characterized by a cluster size of less than five pixels, peach clusters were easily detected.

Utilizing the two-fold criteria, six clusters were classified as peaches, which is identical to the number of peaches actually present on the section analyzed.

CONCLUSION

Apples, which are yellow in color, and immature peaches can successfully be counted from digitalized ground photographs by the system of sequential classification, clustering, and counting. Eight apples and six peaches were counted by this system, and there were eight apples and six peaches on the areas analyzed.

COMMENTS

Further research need be conducted to determine if this system is applicable to other types of fruit in the immature stage.

Several photographs of the same type of fruit should be analyzed to determine if spectral homogeneity exists among photographs.

Classification results using varying aperture sizes with all filter and scanning mode combinations should be examined for various fruit at different stages of development to determine the best aperture size and filter-scanning mode combinations for each fruit at various developmental stages.

Research should be performed to determine which positioning of the camera with respect to the sun will provide the best classification results of unlabeled pixels.

Since aerial photographs are more economically obtained than ground photographs in a large scale survey, aerial photographs should be analyzed to test this system for counting immature fruit.

REFERENCES

[1] Boller and Chivens, "User's Manual Microdensitometer Operating Instructions", South Pasadena, California: The Perkin-Elmer Corporation, 1973.

[2] Barr, Anthony J., and Goodnight, James H., "A User's Guide to the Statistical Analysis System", Raleigh Student Store, North Carolina State University, 1972.

[3] Nealon, Jack, "THE SYSTEM OF SEQUENTIAL CLASSIFICATION, CLUSTERING, AND COUNTING FRUIT TREES FROM DIGITALIZED AERIAL PHOTOGRAPHS," Research and Development Branch, Research Division, Statistical Reporting Service, U. S. Department of Agriculture, Washington, D. C., 1975.