N O T I C E

80-10189

NASA CR-
160610

JSC-09478
December 1974

DATA RESOLUTION VERSUS FORESTRY CLASSIFICATION

(E80-10189) DATA RESOLUTION VERSUS FORESTRY CLASSIFICATION (Lockheed Electronics Co.) 34 p HC A03/MF A01 CSCL 05B

N80-27766

Unclas
G3/43 00189

National Aeronautics and Space Administration

LYNDON B. JOHNSON SPACE CENTER

Houston, Texas

LEC-5168

DATA RESOLUTION VERSUS FORESTRY CLASSIFICATION

APPROVED

R. L. Smelser

R. L. Smelser, Project Manager
Forestry Applications Project
Southern Region
U.S. Department of Agriculture Forest Service

A. W. Patteson

A. W. Patteson, Branch Chief
Exploratory Studies and Flight Project Support Branch
NASA/JSC

R. B. MacDonald, Chief
Earth Observations Division
Science and Applications Directorate
NASA/JSC

JSC-09478

ACKNOWLEDGMENT

THIS DOCUMENT WAS PREPARED BY

Edwin P. F. Kan, Principal Scientist

David L. Ball, Associate Scientist

of

Lockheed Electronics Company, Inc., Aerospace Systems Division

under

Contract NAS 9-12200, Job Order No. 75-315

for

Forestry Applications Exploratory Study Project

of

Southern Region

U.S. DEPARTMENT OF AGRICULTURE FOREST SERVICE

in cooperation with

Science and Applications Directorate

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

December 1974

LEC-5168

ABSTRACT

The effects on classification accuracies due to changes in data resolution are examined in timber stand classifications using remotely sensed multispectral data. Such an investigation is valuable in deciding on optimal sensor and platform designs. Data resolution means the actual ground area covered by a pixel recorded on the multispectral data; high resolution implies small ground area.

To date results in the data resolution study indicate that classification accuracies for data with high resolution are actually less than the accuracies for data with lower resolution. This conclusion is supported by theoretical justifications and by experimental verification. The verification was performed on multispectral data sets over the Sam Houston National Forest.

CONTENTS

Section

ACRONYMS LIST

ERIPS	Earth Resources Interactive Processing System
FAP/TR	Forestry Applications Project on Timber Resources
LARSYS	Laboratory for Application of Remote Sensing System
MSS	Multispectral Scanner
MSS/24	The Bendix 24 Channel Multispectral Scanner
PCC	Probability of Correct Classification
PMC	Probability of Misclassification
SHNF	Sam Houston National Forest

1.0 OBJECTIVES

This data resolution investigation is designed to:

- Study the theoretical effects on classification accuracies due to changes in data resolution
- Verify the theoretical conclusions by performing a forestry classification, using real and simulated data with various reduced resolution.

2.0 JUSTIFICATION

It is often expected that multispectral data with different data resolution permits different classification accuracies for varied hierarchies of ground features. This follows from photointerpreters' experience in mapping ground features, using varied scales of imagery. In this paper, data resolution means the actual ground area covered by a pixel, or picture element, recorded on the multispectral data, e.g., as recorded by the multispectral scanner.

The question naturally arises; "What is the optimal data resolution for classification of remotely sensed data?" For specific applications, e.g., forestry applications, the same question must also be answered for the specific features of interest. In forestry applications, the features are timber stands of different species and/or condition classes. Species composition defines the timber type of the stand, while the age and/or size determines the condition class of the timber stand.

Forest scenes are particularly complex, especially when viewed from low altitudes. The complexity is due to the nonhomogeneity of the tree patterns, nonuniformity of the composition of trees in the stand, variation in the undergrowth and spacing between individual trees, texture effects due to shadows, etc. All these effects are significant for data with resolution less than perhaps 20 meters square.

Low altitude data, i.e., high resolution data, is sometimes undesirable for machine processing and recognition. This is contrary to the belief and practice in photointerpretation. The reason is machine processing of multispectral data is not completely parallel with the human interpretation process. The latter process actually extract information such as texture, shape of tree crowns, shadows of trees indicating their profiles, etc., from high resolution data, which are used for detection of details. However, the machine process has not been advanced to using these nonspectral information that photointerpreters use; the standard training field pattern recognition technique of Laboratory for Application of Remote Sensing System (LARSYS) (ref. 1) uses spectral information alone. Thus, forest scene complexities in high resolution data make stand identification difficult. Smoothing out the complexities would be expected to improve classification accuracies.

Guided by this intuition, the complexities in forest scenes are smoothed by simulating lower resolution data from high resolution data. This is achieved by an averaging process, which simulates data sets obtained at higher altitudes. With this kind of modeling, the theoretical effects on classification accuracies due to resolution reduction can be examined. And with this kind of simulation, experimental work can be performed to verify the theoretical conclusions.

3.0 SCOPE OF APPLICATIONS

Since this work is part of the Forestry Applications Project on Timber Resources (FAP/TR) (ref. 2), forest scenes over Sam Houston National Forest (SHNF) were examined. Mission M230 of the C-130 aircraft was flown over SHNF on March 21, 1973 at 10,000 feet altitude; the Bendix 24 channel multispectral scanner (MSS/24) onboard the aircraft collected MSS data. Eight edits from the MSS/24 coverage were selected for study, as discussed in section 2.6 and appendix H of reference 3. These are called Edit numbers 3, 6, 9, 12, 14, 18, 53 and 54. Each edit is approximately 11 square kilometers. The forest features of interest are listed in table I.

The data sets over the eight edits were preprocessed. Preprocessing means calibration, scan-angle correction and registration to ground (ref. 2 and 4-8). The resulting data resolution is approximately 8 meters square.

TABLE I. — THE TYPE (CLASSES) AND CONDITION CLASSES (SUBCLASSES) OF FOREST FEATURES OF INTEREST IN SAM HOUSTON NATIONAL FOREST OF TEXAS

Type No.	Type (Class)	Subclass No.	Condition Class (Subclass)
1	Shortleaf pine	1.1	Plantation - 3 years old
		1.2	Poletimber - immature
		1.3	Sawtimber - immature
		1.4	Sawtimber - mature
2	Loblolly pine	2.1	Plantation - 1 year old
		2.2	Plantation - 3 year old
		2.3	Seedling and Sapling - adequately stocked
		2.4	Poletimber - immature
		2.5	Sawtimber - immature
		2.6	Sawtimber - mature
3	Laurel oak - willow oak	3.1	Sawtimber - immature
4	Sweetgum - nuttal oak - willow oak	4.1	Sawtimber - low quality
		4.2	Sawtimber - immature
		4.3	Sawtimber - mature
5	Post oak - black oak	5.1	Sawtimber - immature
6	Loblolly pine - hardwoods	6.1	Sawtimber - immature
7	Cut-over land	7.1	Site prepared and windrowed
		7.2	Not site prepared

4.0 TECHNICAL APPROACH

4.1 Simulation of Data with Reduced Resolution

An averaging or weighted averaging relationship is assumed between data at different resolutions.

For example, data {X} has a resolution of (5 meter)^2. Data {Y} has a reduced resolution, i.e., coarser resolution, of (10 meter)^2. Thus, the same (10 meter)^2 ground area covered by one Y measurement, y, will be covered by four X measurements, x_1, x_2, x_3, and x_4. The averaging relationship is

$$y = \tfrac{1}{4}(x_1+x_2+x_3+x_4) \quad .$$

The weighted averaging relationship is

$$y = w_1x_1+w_2x_2+w_3x_3+w_4x_4 \qquad \sum w_i = 1,\ w_i \geq 0 \quad .$$

Data set {X} and set {Y} could theoretically be acquired using the same sensor, flown at two altitudes; the altitude for {X} is half the altitude for {Y}. (See section 2.5 of ref. 3.)

4.2 Classification and Evaluation Procedures

The classification technique is the widely used scheme of supervised pattern recognition. That is, training fields are selected to train the maximum likelihood classifier, such as in LARSYS (ref. 1). Normal statistical distributions are assumed on the training classes. Equivalently, the Bayes' classifier using equal a priori probability is employed (ref. 9).

The evaluation procedure is the calculation of classification accuracies of training fields/classes. The classification accuracy is a measure of the statistical probability of correct classification (PCC) (ref. 10), which is another widely accepted evaluation parameter. Also, the divergence measure (ref. 10) is calculated to convey the extent of separability between classes. In special cases, it has been established that the divergence measure has direct relationship with PCC.

5.0 THEORETICAL RESULTS

5.1 Probability of Correct Classification

The following discussion shows that there is a gain in the PCC when the data resolution is lowered. Actually, the probability of misclassification (PMC) for the 2-class classification case is computed below; PMC = 1 - PCC . Data sets {X} and {Y} are studied, where {Y} is a mX (i.e. m times) reduction of {X} ; i.e., a generic data point in {Y} is an average of m^2 data points in {X} .

Assume the following notations for the means and covariance matrices for the two classes C_1 and C_2 in the data sets {X} and {Y} :

$$C_1: \quad \mu_{X1}, \Sigma_{X1}; \ \mu_{Y1}, \Sigma_{Y1}$$

$$C_2: \quad \mu_{X2}, \Sigma_{X2}; \ \mu_{Y2}, \Sigma_{Y2} \quad .$$

These parameters can be estimated from the {X} and {Y} data sets, using the normal method of training field selection and statistics calculation. By the averaging process which simulates {Y} from {X} , it can be shown that:

$$\mu_{X1} = \mu_{Y1}$$

$$\Sigma_{X1} = m^2 \Sigma_{Y1}$$

$$\mu_{X2} = \mu_{Y2}$$

$$\Sigma_{X2} = m^2 \Sigma_{Y2} \quad .$$

Using these statistics to train the classifier, the Bayes' regions (ref. 10) for equal a priori probabilities are established and denoted by R_{X1}, R_{X2} for data set {X} and by R_{Y1}, R_{Y2} for data set {Y} . It can be shown that

$$R_{X1} = R_{Y1}$$

$$R_{X2} = R_{Y2}$$

By the definition of PMC, which is (ref. 10)

$$PMC = \tfrac{1}{2}\mathrm{Prob}(R_1/C_2) + \tfrac{1}{2}\mathrm{Prob}(R_2/C_1) ,$$

it can be seen that

$$(PMC)_X \geq (PMC)_Y$$

That is, PMC is lower for data set {Y} than for {X} , because the distributions in data set {Y} taper off quicker than in {X} ; this is a result of the relationship between the covariance matrices. Therefore,

$$(PCC)_Y \geq (PCC)_X .$$

That is, the classification accuracy will be higher for the lower resolution data {Y} than for the higher resolution data {X} .

5.2 Separability: Divergence Measure

The following establishes that the divergence between C_1 and C_2 increases with the lowering of the data resolution; the same situation as in section 5.1 is assumed. The divergence measure is used because it has been shown (ref. 11) that the divergence value relates to PCC. In fact, when $\Sigma_1 = \Sigma_2$ for C_1 and C_2, the divergence $J(C_1,C_2)$ between C_1 and C_2 has a one-to-one relationship with PCC; and, $J(C_1,C_2)$ increases if and only if PCC increases. Generally, the larger the divergence value, the more separable C_1 is from C_2.

The divergence $J(C_1,C_2)$ between C_1 and C_2 is defined as:

$$J(C_1,C_2) = \tfrac{1}{2}\mathrm{tr}\left[\Sigma_1-\Sigma_2\right]\left[\Sigma_2^{-1}-\Sigma_1^{-1}\right] + \tfrac{1}{2}\left[\mu_1-\mu_2\right]^T\left[\Sigma_1^{-1}+\Sigma_2^{-1}\right]\left[\mu_1-\mu_2\right] .$$

By the relationship established in section 5.1 between μ_{X1} and μ_{Y1}, Σ_{X1} and Σ_{Y1}, μ_{X2} and μ_{Y2}, Σ_{X2} and Σ_{Y2}, $J_X(C_1,C_2)$ and $J_Y(C_1,C_2)$ for data sets {X} and {Y} can be related by the following inequalities:

$$J_X(C_1,C_2) \leq J_Y(C_1,C_2) \leq m^2\, J_X(C_1,C_2) .$$

$J_Y=m^2J_X$ in the special case when $\Sigma_1=\Sigma_2$; and $J_Y=J_X$ when there is no averaging, i.e., when data set {Y} is identical to data set {X}.

In general, when {Y} is a weighted average of {X} ; i.e., a generic data point, y , in {Y} relates to the generic data points, x_i , in {X} in the following manner:

$$y = \sum_{i=1}^{m^2} w_i x_i \, , \; \sum w_i = 1 \quad w_i \geq 0 \; ;$$

the m^2 factors in the above discussion will be replaced by m^2 times a constant. This constant equals $\sum w_i^2$, and takes values between 1 and $1/m^2$. That is,

$$J_X(C_1,C_2) \leq J_Y(C_1,C_2) \leq k \, J_X(C_1,C_2)$$

where $1 \leq k \leq m^2$.

6.0 FORESTRY APPLICATIONS: TO DATE RESULTS

6.1 Data Set Studied: Edit 9

During this reporting period, only Edit 9 out of the 8 edits was investigated. A three-channel color rendition of the MSS data is shown in figure 1, with the timber stand and compartment boundaries delineated on the imagery. The codes for timber stand types are found in table I of section 3.0.

The unreduced data plus two simulated data sets were studied: 1X, 2X, and 3X, where 1X has a data resolution of approximately (8 meters)^2; 2X, (16 meters)^2; and 3X, (24 meters)^2. The simulation is by the data resolution reduction program discussed in section 2.5 of reference 3.

Only 12 channels on the MSS/24 were operating during M230. They are the channels numbered 1 through 11 and 13 on the scanner. However, in Edit 9 data, these channels are numbered 1 through 12. The spectral coverages of these channels are shown in table II.

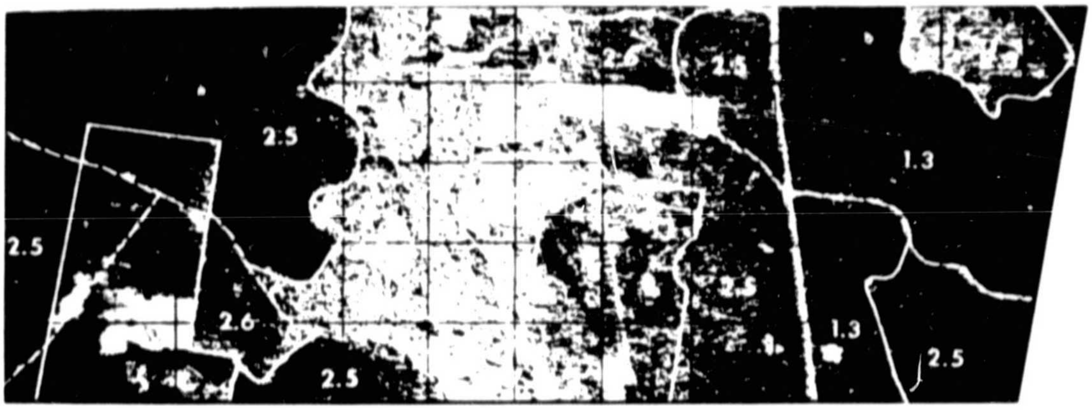
6.2 Field Selection

The fields selected for classification and divergence studies are shown in figure 2. The entire Edit 9 area is divided into 3 sections, left (L), middle (M) and right (R); hence the labels of fields, e.g., L2.5, R2.5.

TIMBER STAND AND COMPARTMENT MAP OVER

SAM HOUSTON NATIONAL FOREST MSS/24 EDIT

MISSION NO. 230 — EDIT NO. 9



APPROXIMATE SCALE 1:28,000

Figure 1. — Edit number 9.

~~6-2~~

13

TABLE II. – SPECTRAL COVERAGES OF 12 CHANNELS OF THE EDIT 9 MSS DATA

Channel No.	Spectral coverage (micrometer)
1	0.375-0.405
2	0.40-0.44
3	0.466-0.495
4	0.53-0.58
5	0.588-0.643
6	0.65-0.69
7	0.72-0.76
8	0.770-0.810
9	0.82-0.88
10	0.981-1.045
11	1.20-1.30
12	2.10-2.36

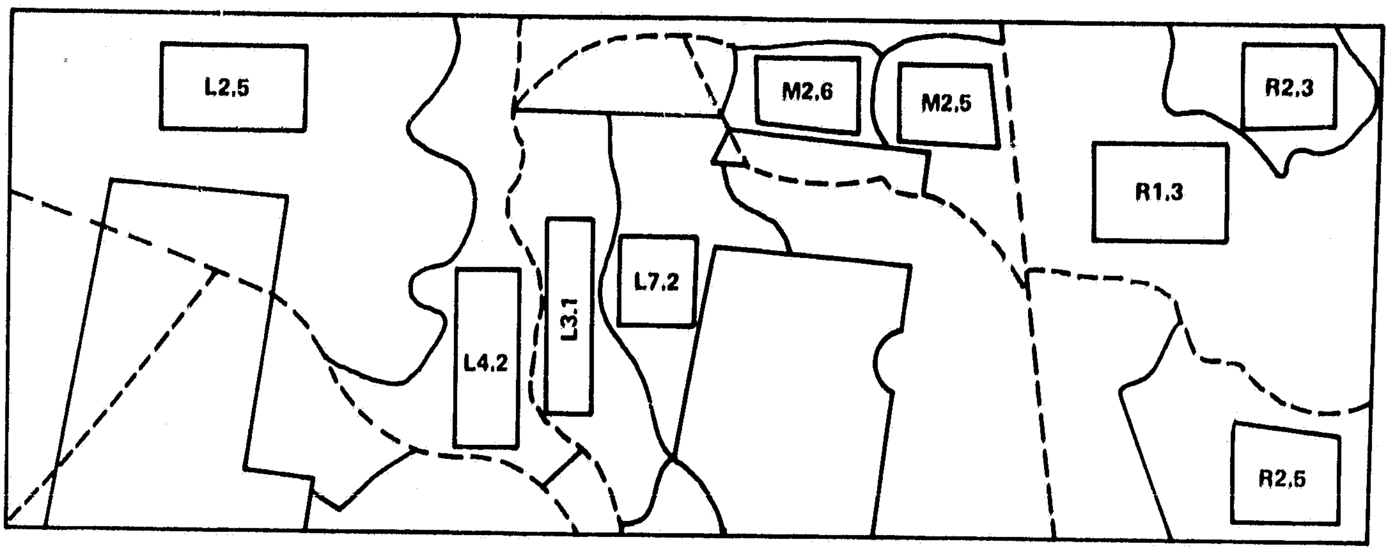


Figure 2. — Locations of fields selected on Edit 9.
(Used in classification and divergence studies)

Because it was felt that the scan-angle correction performed earlier in the project did not adequately correct for the scan-angle effects, the left fields, middle fields and right fields were studied separately. Actually, it was found that the mean values for 2.5 fields in the left differ by a few counts from the 2.5 fields in the right; hence the conclusion on inadequate scan-angle correction.

The same physical fields were selected from 1X, 2X, and 3X. Thus, the field coordinates in 1X, 2X, and 3X are directly related.

6.3 Data Processing

The 1X, 2X, and 3X data were processed on the Earth Resources Interactive Processing System (ERIPS). The left four fields, middle two fields, and right three fields were studied separately.

Statistics of these fields were generated; pairwise classification and divergence calculations were made. For example, for the right three fields R1.3, R2.3 and R2.5, there are three pairs: R1.3/R2.3, R1.3/R2.5 and R2.3/R2.5.

Classification and divergence calculations were performed using three different channel sets. (1) 8 channels: numbers 2, 4, 5, 7, 8, 9, 10, and 11. (2) 4 channels: numbers 3, 5, 8, and 11. (3) 4 best channels as dictated by the channel selection processor on ERIPS: numbers 2, 7, 10, 12 for 1X and 3X; numbers 2, 3, 7, and 11 for 2X.

In case 1, the 8-channel set was chosen arbitrarily because of the limitation of ERIPS in the divergence calculation. Channels 1 and 3 were arbitrarily dropped, because channel 2 contains very similar information (at least visually); channel 2 was retained; channel 6 was dropped because of data drop-out; channel 12 was dropped because data values were very low. In case 2, the 4-channel set was arbitrarily chosen, and spaced throughout the 12 channels. The channel set in case 3 was dictated by the channel selection processor on ERIPS.

6.4 Analysis Results

The results of performing the classification and divergence measurements are summarized in figures 3 through 5. Each figure is in bar-chart form.

Each bar-chart shows the classification accuracy for the pairwise classification (on the ordinate) versus the specific pairs of classes used in classification (on the abscissa). The classification accuracy is a measure of PCC and is given by ½(classification accuracy of C_1 + classification accuracy of C_2)*; classification accuracy of C_i = the number of points of C_i correctly classified into C_i/the number of points of C_i.

---

*Beside this definition of classification accuracy, other measures have also been commonly used; for example: number of correctly classified points of C_1 and C_2/total number of points of C_1 and C_2.

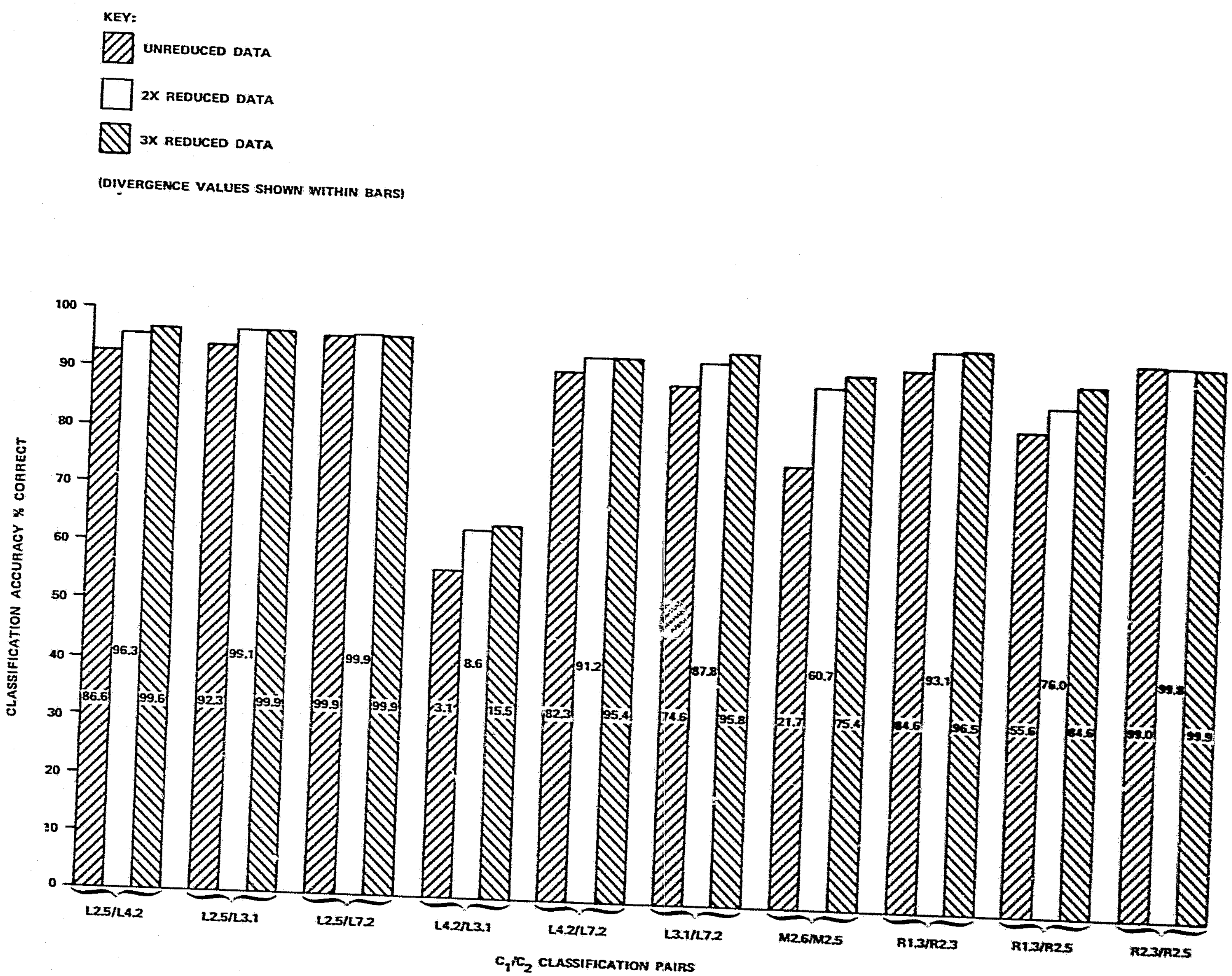


Figure 3. – Bar charts of pairwise classification accuracies: case 1.

6-7
18

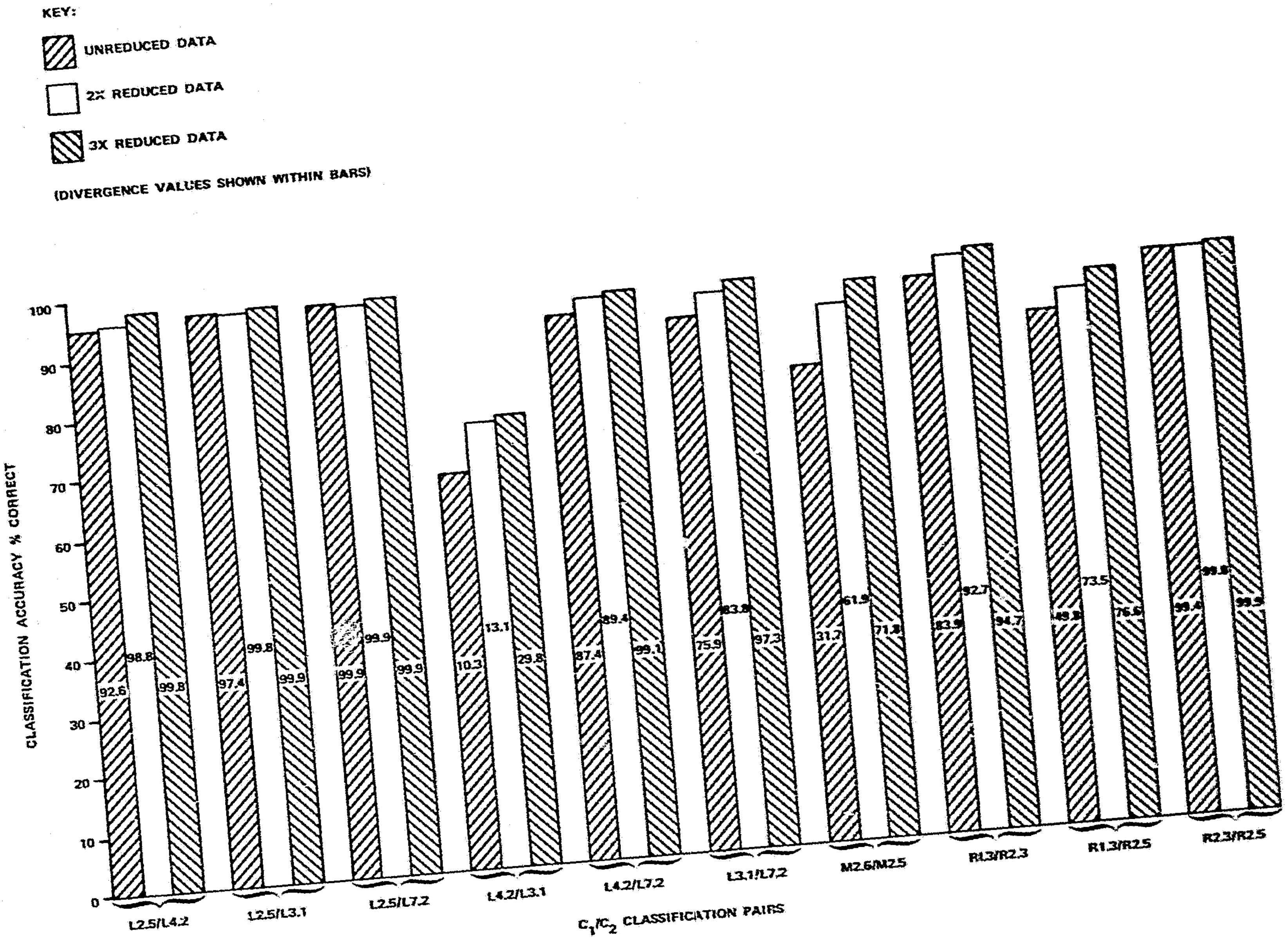


Figure 4. — Bar charts of pairwise classification accuracies: case 2.

~~8-9~~
19

ORIGINAL PAGE IS
OF POOR QUALITY

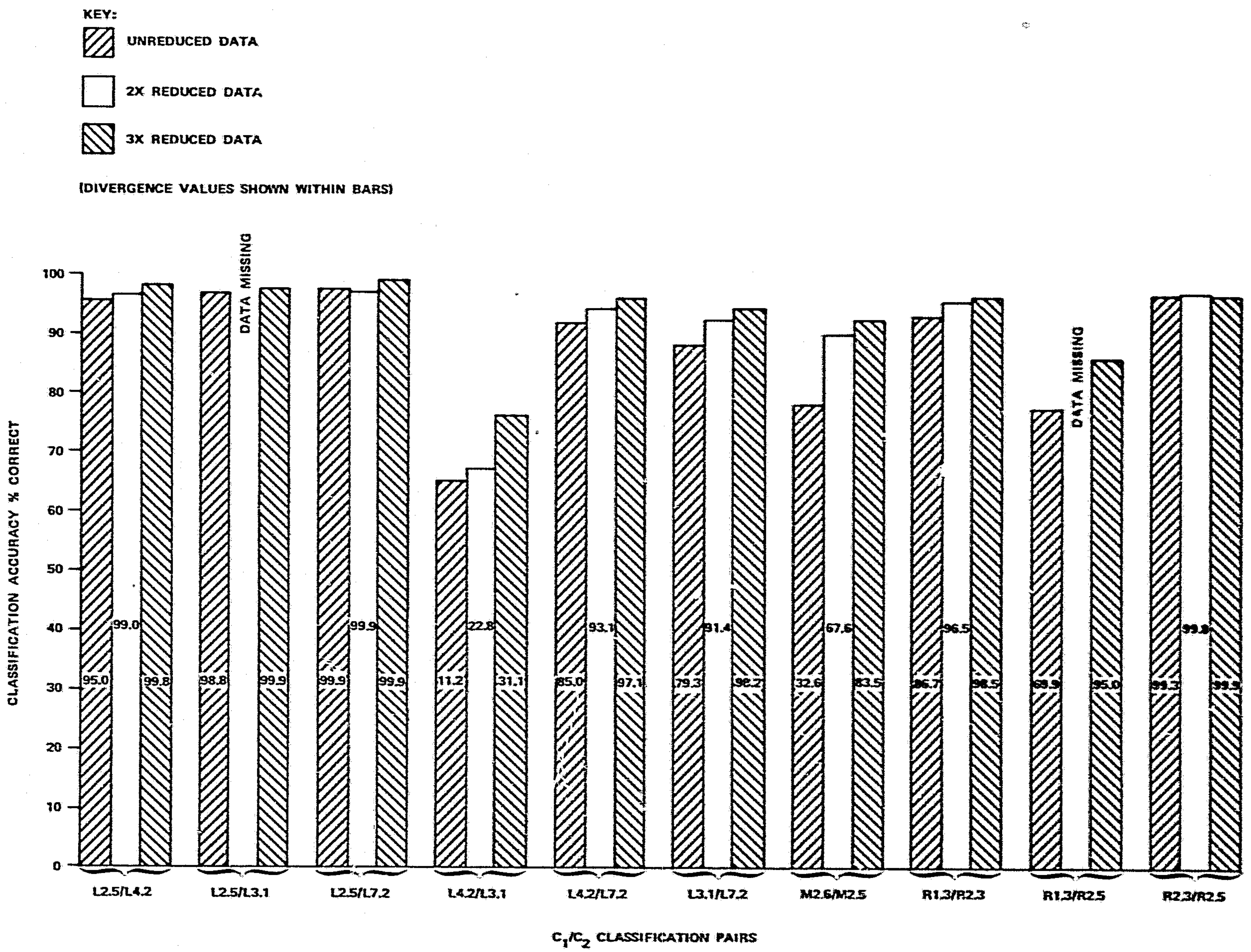


Figure 5. — Bar charts of pairwise classification accuracies: case 5.

~~6-9~~ 20

The pairwise divergence values are also indicated in the bar-charts. The values are written in the bars.

Figures 3 through 5 correspond to cases 1, 2, and 3 of data processing discussed in section 6.3. The 1X, 2X, and 3X results are shown side-by-side.

6.5 Inference from Analysis Results

Figures 3 through 5 lead to the following conclusion: Classification accuracies increase with the lowering of data resolution. Also, classes are more separable at lower resolution.

This conclusion reinforces the theory discussed in section 5.0.

7.0 THE PARADOX AND SOME REMARKS

7.1 The Paradox

The theoretical and experimental results conclude that classification accuracies increase with the reduction in fidelity of data resolution. This gives rise to the following paradox:

> "If ground features can be classified using low altitude data, they can also be classified, and even with higher accuracies, using high altitude data."

7.2 The Paradox's Intuitive Explanation

The paradox should not cause any alarm, because the statement is asserted for classification accuracies alone, and because the classification technique employs spectral information alone.

The accuracy measure used in the analysis comes from evaluating training/test data which are well defined and delineated due to prior knowledge. The loss in boundary accuracy and mensuration accuracy in the analysis of higher altitude data has not been accounted for. These two factors are most often deciding factors on optimal data resolution. Also, the gain in details at higher resolution is not an asset to the spectral classification rule. In fact, the details in texture, etc., add to the complexity in machine processing in this case.

Another explanation for the increase in classification accuracies for lower resolution data is that a "PERFIELD" classification (ref. 11) is performed on the lower resolution data, compared to a "PERPOINT" classification on the higher resolution data. A "PERFIELD" classification rule has been suggested to be superior to the "PERPOINT" classification rule. That is, nonhomogeneity in the higher resolution data is reduced by the averaging process, which gives the lower resolution data. This explanation is readily acceptable, especially for forest scenes, where complexities abound with high resolution.

7.3 A Remark on Detection

An interesting remark follows from the conclusion of the above analysis. For detection purposes, ERTS-1, for example, will outperform aircraft data analysis, as long as the features to be detected have physical sizes larger than the ERTS-1 resolution (preferably at least four times larger, in order to assure total containment of the feature in at least one pixel). Detection here means the detection of the presence of the feature, disregarding its size.

7.4 Decision on Optimal Data Resolution

An optimizing criterion can be set up where the optimal choice of data resolution is a compromise between classification accuracy, boundary accuracy and mensuration accuracy. The criterion, D, could then be written as

$$D = d_1\theta_c + d_2\theta_b + d_3\theta_m ,$$

where d_1, d_2, and d_3 are weights in the criterion, and θ_c, θ_b, θ_m are respectively the accuracies in classification, boundary location, and mensuration. An optimal solution for data resolution will be obtained by achieving maximum value of the criterion D. Different applications will call for different weights d_1, d_2, and d_3; and will produce different solutions. Other factors such as the cost of data acquisition, cost of data processing, etc., can be also incorporated into the criterion as follows:

$$D' = d_1\theta_c + d_2\theta_b + d_3\theta_m + d_4C_a + d_5C_p ,$$

where C_a and C_p are the respective costs.

An optimal decision on data resolution will lead to an optimal design of sensors and platforms.

8.0 CONCLUSIONS

This study has derived the theoretical effects on classification accuracies due to the changes in data resolution. The modeling of the data reduction scheme led to simulation of forestry data at lower resolutions. The performance of the forestry applications experiment led to analysis results that reinforce the theoretical inferences. That is, classification accuracies increase with the reduction in fidelity of data resolution using the training field and maximum likelihood classification scheme.

This apparent paradox is explained in section 7.0. Suggestions are then made concerning the optimal design of data resolutions, and thus the optimal design of sensors and platforms.

Further efforts in this data resolution study will be spent in: (1) Extension of the results and analyses to other areas in the Sam Houston National Forest; (2) Expansion of the results and analyses to obtain classification images; (3) Quantitative and qualitative evaluation of the resulting classification images.

9.0 REFERENCES

1. "LARSYS Version 3, Users' Manual" (edited by T. L. Phillips), Purdue University, Laboratory for Applications of Remote Sensing, West Lafayette, Indiana 47906. June 1973.

2. Kan, E. P.: The Forestry Applications Project on Timber Resources (FAP/TR): The FY75 Investigation Plan. Lockheed Electronics Co., Inc., Houston, Texas, Tech. Rep. #LEC-3567, June 1974.

3. Kan, E.P.: The Forestry Applications Project on Timber Resources (FAP/TR): FY75 Interim Report (7/74-12/74). Lockheed Electronics Co., Inc., Houston, Texas, Tech. Rep. #LEC-5003, Dec. 1974.

4. Hutchens, D.: Bendix Multispectral Scanner Program (BENMSS), Program Q747. Lockheed Electronics Co., Inc., Houston, Texas, DPD#0668, Project 4699D, Ref. #CAD-2709, Apr. 1972.

5. Lopez, M.: The REFORM (Reformatting and Removal of Instrument Instabilities) Program. NASA/Johnson Space Center, Memo. EOD/TF4, 1973.

6. Yao, S.: A Method for Digital Image Registration Using a Mathematical Programming Technique. Lockheed Electronics Co., Inc., Houston, Texas, Tech. Rep. #LEC 640-TR-138, Ref. EOD 2910, July 1972.

7. Sargent, J.: Data Verification Plan. NASA/Johnson Space Center, Houston, Texas, Memo. #TF4/250-73, Sept. 11, 1973.

8. Rader, M.: Geometric Registration (Mapping) of 24-Channel Scanner Bendix Digital Data. Lockheed Electronics Co., Inc., Houston, Texas, Tech. Rep. #LEC-2086, Mar. 1974.

9. Anderson, T. W.: *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York, N.Y., 1958.

10. Marill, T.; and Green, D. M.: On the Effectiveness of Receptors in Recognition Systems. IEEE Transactions on Information Theory, Vol. IT-9, Jan. 1963.

11. Gupta, J. N.; et al: Machine Boundary Finding and Sample Classification of Remotely Sensed Agricultural Data. Proceedings of the Conference on Machine Processing of Remotely Sensed Data, held at Purdue University, Lafayette, Indiana, Oct. 16-18, 1973, pp. 4B.25-35.