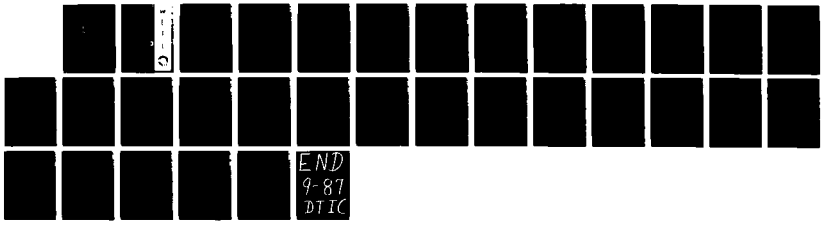
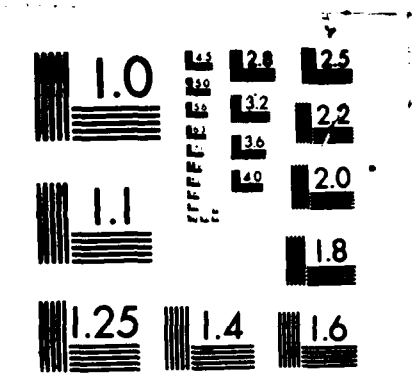
AD-A183 537 PATTERN CLASSIFICATION TECHNIQUES APPLIED TO HIGH RESOLUTION SYNTHETIC AP..(U) ARMY ENGINEER TOPOGRAPHIC LABS FORT BELVOIR VA R A HEVENOR ET AL. NOV 86 1/1

UNCLASSIFIED ETL-0443 F/G 17/9 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

DTIC FILE COPY

(12)

ETL-0443

AD-A183 537

Pattern classification techniques applied to high resolution, synthetic aperture radar imagery

Richard A. Hevenor
Pi-Fuay Chen

November 1986

DTIC
ELECTE
AUG 0 5 1987
S D
E

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED

Prepared for
U.S. ARMY CORPS OF ENGINEERS
ENGINEER TOPOGRAPHIC LABORATORIES
FORT BELVOIR, VIRGINIA 22060-5546

87

E T L

U.S. ARMY ENGINEER TOPOGRAPHIC LABORATORIES
ETL

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ETL-0443	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* Pattern Classification Techniques Applied to High Resolution, Synthetic Aperture Radar Imagery		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Richard A. Hevenor Pi-Fuay Chen		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Army Engineer Topographic Laboratories Fort Belvoir, Virginia 22060-5546		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 4A161102B52C
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Engineer Topographic Laboratories Fort Belvoir, Virginia 22060-5546		12. REPORT DATE November 1986 13. NUMBER OF PAGES 27
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Radar Imagery
Pattern Recognition

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This report describes the application of 10 pattern classification techniques to selected samples of high resolution, synthetic aperture radar imagery taken over the Huntsville, Alabama area. Sections of the radar imagery were digitized and stored on a digital disk unit. A Lexidata system 3400 image processor and a Hewlett Packard 1000 computer were used to display the images on a cathode ray tube and to take 100 samples for each of four terrain classes from the imagery. The 400 image samples were then used as training sets to derive the

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

10 classifiers. Once the classifiers were derived, the training set data were then used as input to the classifiers to see how well each would do in classifying the original training sets.

PREFACE

The authority for performing the work described in this research note is contained in Project 4A161102B52C, "Research in Geodetic, Cartographic, and Geographic Sciences."

The work described in this research note represents an application of 10 standard pattern classification techniques to samples of high resolution synthetic aperture radar imagery. The task was performed under the supervision of Dr. Frederick W. Rohde, Team Leader, Center for Physical Sciences, and Dr. Robert D. Leighty, Director, Research Institute.

COL Alan L. Laubscher, CE, was the Commander and Director, and Mr. Walter Boge was the Technical Director of the Engineer Topographic Laboratories during the report preparation.

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

DTIC
COPY
INSPECTED
6

CONTENTS

ILLUSTRATIONS

PATTERN CLASSIFICATION TECHNIQUES APPLIED TO HIGH RESOLUTION SYNTHETIC APERTURE RADAR IMAGERY

INTRODUCTION

The problem of automatic pattern classification of remotely sensed imagery has been the subject of study for many years. It is the purpose of this research note to show the results of applying 10 different pattern classification techniques to samples of synthetic aperture radar imagery. In the past, pattern classification methods have been applied to various types of optical imagery; however, little work has been done in applying these methods to high resolution radar imagery. In order to perform automatic classification of terrain features using radar imagery, the application of pattern classification methods is a necessary step. These methods are very general in nature and can be applied to any type of imagery that can be represented with a feature vector. The following sections will present a short discussion of the classification methods used and the final results obtained.

METHODOLOGY

The pattern classification methods that were used in this study were all standard methods and are described in detail in various textbooks on pattern recognition. For this reason, no attempt will be made to explain each method in detail and only the most significant equations will be presented and discussed. The imagery used consisted of samples of high resolution synthetic aperture radar imagery taken over the Huntsville, Alabama, area with the APD-10 radar system. Sections of radar imagery were digitized and stored on a digital disk unit. A Lexidata system 3400 image processor and a Hewlett Packard 1000 computer were used to display the images on a cathode ray tube and to take 100 samples for each of four terrain classes from the imagery. Each sample consisted of a 32 by 32 pixel window located in a homogeneous section of a particular terrain class. The four classes considered were (1) cities (combination of commercial and residental structures, DLMS category #504 FIC 301 and #505 FIC 401), (2) fields (agriculture used primarily for crops and pasture land, DLMS category #501 FIC 950), (3) water (rivers with smooth fresh water, DLMS category #510 FIC 940) and fresh water subject to ice, (lakes and reservoirs, DLMS category #510 FIC 943), and (4) forests (mixed trees--deciduous and evergreens--DLMS category #501 FIC 954). A feature vector consisting of 13 components for eight of the classification techniques and 15 components for two of the classification techniques was computed for each image sample. These components of the feature vector consisted of the first- and second-order histogram statistics calculated from the 32 by 32 pixel window. The equations for these histogram measures are provided in the appendix. A discriminant analysis technique was used for feature selection to reduce the dimensionality of the feature vector from 13 to 2 in such a way that the resulting components were optimized for showing class separability. This feature selection technique was the subject of a previous ETL report[1] and provides a linear transformation of the following form:

[1] Richard Hevenor, *Application of a Feature Selection Technique to Samples of High Resolution Synthetic Aperture Radar Imagery*, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, VA, ETL-0330, July 1983, AD-A 135 006

$$\underline{y} = A\underline{x} \tag{1}$$

where $\underline{x}$ is the original feature vector with dimensionality 13 X 1, A is the transformation matrix of dimensionality 2 X 13, and $\underline{y}$ is the transformed feature vector with dimensionality 2 X 1. The solution for the elements of matrix A is given in the previous ETL report. For the Bayes Classifier, 15 components of the feature vector were used to compute the decision functions. For the minimum distance classifier, the five most separable feature vector components from the originally established 15 components (shown in appendix) were selected using visual inspection. The visual inspection was used for a set of more than 50 training samples. These five feature vector components were the mean, the variance, the skewness, the autocorrelation, and the covariance. After feature selection was performed, pattern classifiers were implemented and applied to the 400 samples of radar imagery. The pattern classification techniques implemented were

1. Ho-Kashyap Algorithm.
2. Increment Correction Algorithm.
3. Least Mean Square-Error Algorithm.
4. Method of Potential Functions.
5. Fisher Linear Discriminant.
6. Pseudoinverse Technique.
7. Widrow-Hoff Procedure.
8. Relaxation Algorithm.
9. Bayes with normal distributions.
10. Minimum distance classifier.

A short discussion of each of these techniques follows.

Ho-Kashyap Algorithm

The Ho-Kashyap algorithm as explained in the text by Tou and Gonzalez[2] is a trainable, non-parametric classifier that attempts to solve for a weight vector $\underline{w}$, such that for the two class problem we have

$$\underline{w}^T\hat{\underline{y}} > 0 \text{ if } \hat{\underline{y}} \in \omega_1 \tag{2}$$

$$\text{and } \underline{w}^T\hat{\underline{y}} < 0 \text{ if } \hat{\underline{y}} \in \omega_2 \tag{3}$$

[2] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, Addison-Wesley, 1974.

where ω_1 represents class 1 and ω_2 represents class 2 and the vector $\hat{\underline{y}}$ is equal to the vector $\underline{y}$ augmented by 1.

$$\hat{\underline{y}} = \begin{bmatrix} y_1 \\ y_2 \\ 1 \end{bmatrix} \tag{4}$$

In equation (4) y_1 and y_2 are components of $\underline{y}$. For our case the dimensionality of $\hat{\underline{y}}$ and $\underline{w}$ is three. The T associated with the $\underline{w}$ vectors in the above inequalities means transpose. An iterative solution for $\underline{w}$ can be obtained as follows:

$$\underline{w}(k+1) = \underline{w}(k) + cY^{\#}[\underline{e}(k) + |\underline{e}(k)|] \tag{5}$$

$$\underline{e}(k) = Y\underline{w}(k) - \underline{b}(k) \tag{6}$$

$$\underline{b}(k+1) = \underline{b}(k) + c[\underline{e}(k) + |\underline{e}(k)|] \tag{7}$$

where $\underline{w}(1) = Y^{\#}\underline{b}(1)$ and $\underline{b}(1) > 0$ but otherwise arbitrary

$$Y^{\#} = (Y^T Y)^{-1} Y^T \qquad Y = \begin{bmatrix} \hat{\underline{y}}_1^T \\ \hat{\underline{y}}_2^T \\ \cdot \\ \cdot \\ \cdot \\ \hat{\underline{Y}}_N^T \end{bmatrix}$$

N is the total number of pattern points for the two classes. The iterative index k is used not only on $\underline{w}$ but also on the vector $\underline{b}$, so that both $\underline{b}$ and $\underline{w}$ are updated on each iterative pass. The $\hat{\underline{y}}$ data belonging to class 2 are multiplied by minus one before insertion into the matrix Y. A solution for $\underline{w}$ can be obtained when $0 < c \leq 1$ and if the classes are linearly separable in the first place.

Stochastic Approximation Methods

Stochastic approximation is a general approach to the derivation of statistical pattern classification algorithms. These methods use the training set data to obtain an approximation for the a posteriori probabilities $P(\omega_i | \hat{\underline{y}})$ where ω_i represents the i^{th} class. These methods are nonparametric and allow for the presence of noise in the training samples. There were three stochastic approximation methods used in this work, the increment correction algorithm, the least mean square error algorithm, and the method of potential functions.

Increment Correction Algorithm

The increment correction algorithm is discussed by Tou and Gonzalez[3] and assumes a linear approximation to the a posteriori probability as follows:

$$P(\omega_i|\ \hat{\underline{y}}) \approx \underline{w}^T \hat{\underline{y}} \tag{8}$$

where $\underline{w}$ is a weight vector to be determined. An iterative solution for $\underline{w}$ can be obtained by using the following equation:

$$\underline{w}(k+1) = \underline{w}(k) + \alpha_k\, \hat{\underline{y}}(k)\, \mathrm{sgn}\ \{\, r\,[\hat{\underline{y}}(k)] - w^T \hat{\underline{y}}(k)\,\} \tag{9}$$

where $\underline{w}(1)$ is arbitrary

$$\alpha_k = 1/k \text{ and } k = 1, 2, 3, \ldots.$$

$$r[\hat{\underline{y}}(k)] = \begin{cases} 1 \text{ if } \hat{\underline{y}}(k) \in \omega_1 \\ 0 \text{ if } \hat{\underline{y}}(k) \in \omega_2 \end{cases}$$

$$\mathrm{sgn}\ \{r[\hat{\underline{y}}(k)] - \underline{w}^T \hat{\underline{y}}(k)\} = \begin{cases} 1 \text{ if } r[\hat{\underline{y}}(k)] > \underline{w}^T \hat{\underline{y}}(k) \\ -1 \text{ if } r[\hat{\underline{y}}(k)] \leq \underline{w}^T \hat{\underline{y}}(k) \end{cases}$$

The form of the approximation for $\underline{w}$ given by equation (9) comes from an application of the Robbins-Monro algorithm, which is a standard method for finding the root of a regression function. Once a solution for $\underline{w}$ is found to a sufficient accuracy, the decision rule can be implemented as follows:

$$P(\omega_1|\ \hat{\underline{y}}) \approx \underline{w}^T \hat{\underline{y}} > 1/2 \text{ then } \hat{\underline{y}} \in \omega_1$$
$$P(\omega_2|\ \hat{\underline{y}}) \approx \underline{w}^T \hat{\underline{y}} < 1/2 \text{ then } \hat{\underline{y}} \in \omega_2$$

Least Mean Square Error Algorithm

The least mean square error algorithm as presented by Tou and Gonzalez[4] also approximates the a posteriori probability as equation (8). However, the solution for the $\underline{w}$ vector is quite different as shown below.

$$\underline{w}(k+1) = \underline{w}(k) + \alpha_k \hat{\underline{y}}(k)\, \{r[\hat{\underline{y}}(k)] - \underline{w}^T \hat{\underline{y}}(k)\} \tag{10}$$

where α_k, and $r[\hat{\underline{y}}(k)]$ have the same definitions as used in the increment correction algorithm. Equation (10) provides an iterative solution for $\underline{w}$, which is also a result of an application of the Robbins-Monro algorithm. After a solution for $\underline{w}$ is obtained, the decision rule is implemented in the same manner as the one given for the increment correction algorithm.

---

[3] Ibid.

[4] Ibid.

Method of Potential Functions

The method of potential functions is developed by Tou and Gonzalez[5] and is based on computing an approximation for the a posteriori probability that makes use of a series expansion as shown below:

$$P(\omega_i | \hat{\underline{y}}) \approx \sum_{j=1}^{m} c_j(k)\, \phi_j(\hat{\underline{y}}) \tag{11}$$

In this expansion the functions $\phi_j(\hat{\underline{y}})$ are a given set of orthonormal functions, and the $c_j(k)$ are unknown coefficients that must be determined. In our case the $\phi_j(\hat{\underline{y}})$ were chosen to be a set of Hermite polynomial functions. The decision rule is simply based on the value of $P(\omega_i|\hat{\underline{y}})$,

$$\text{if} \quad P(\omega_i | \hat{\underline{y}}) > P(\omega_j | \hat{\underline{y}}) \; \forall \; j \neq i \text{ then } \hat{\underline{y}} \in \omega_i \tag{12}$$

The coefficients $c_j(k)$ are determined iteratively, such that if the machine makes a correct classification for the sample pattern $\hat{\underline{y}}(k+1)$, then

$$c_j(k+1) = c_j(k) \tag{13}$$

When the machine makes a misclassification for the sample pattern $\hat{\underline{y}}(k+1)$ and $\hat{\underline{y}}(k+1) \in \omega_i$, then

$$c_j(k+1) = c_j(k) + \gamma_{k+1} \phi_j(\hat{\underline{y}}(k+1)) \tag{14}$$

and if $\hat{\underline{y}}(k+1) \notin \omega_i$,

$$c_j(k+1) = c_j(k) - \gamma_{k+1} \phi_j(\hat{\underline{y}}(k+1)) \tag{15}$$

In equations (14) and (15) γ_{k+1} plays the same role as α_k in the increment correction rule.

Fisher Linear Discriminant

The Fisher Linear Discriminant technique is presented by Duda and Hart[6] and seeks to reduce the original $\underline{x}$ vector to a scalar by multiplying $\underline{x}$ by an appropriate vector $\underline{w}$. The decision rule then becomes

$$\text{if } \underline{w}^T \underline{x} > w_0 \text{ then } \underline{x} \in \omega_1, \tag{16}$$

$$\text{if } \underline{w}^T \underline{x} < w_0 \text{ then } \underline{x} \in \omega_2 \tag{17}$$

where the constant w_0 is determined by examination of the training set data and $\underline{w}$ is determined from the following equation:

$$\underline{w} = S_w^{-1}(\underline{m}_1 - \underline{m}_2) \tag{18}$$

where S_w is the within-class scatter matrix. The vectors $\underline{m}_1$ and $\underline{m}_2$ are the mean vectors for class 1 and class 2, respectively. S_w can be computed as follows:

$$S_w = \sum_{\underline{x} \in \omega_1} (\underline{x} - \underline{m}_1)(\underline{x} - \underline{m}_1)^T + \sum_{\underline{x} \in \omega_2} (\underline{x} - \underline{m}_2)(\underline{x} - \underline{m}_2)^T \tag{19}$$

---

[5] Ibid.

[6] Duda, Richard and Hart, Peter, *Pattern Classification and Scene Analysis*, Wiley Interscience, 1973.

Pseudoinverse Technique

This technique is a linear classifier developed in detail in Duda and Hart[7], and attempts to solve for a vector $\underline{a}$, such that

$$\text{if } \underline{a}^T\hat{\underline{y}} > 0 \text{ then } \hat{\underline{y}} \in \omega_1 \tag{20}$$

$$\text{if } \underline{a}^T\hat{\underline{y}} < 0 \text{ then } \hat{\underline{y}} \in \omega_2 \tag{21}$$

The vector $\underline{a}$ for our case has three components. A solution for $\underline{a}$ can be obtained by forming a matrix H from all the training samples taken from the two classes. Each row of H will consist of a sample $\hat{\underline{y}}^T$ with the samples coming from ω_2 being multiplied by -1. An equation involving H and $\underline{a}$ is given as

$$H\underline{a} = \underline{b} \tag{22}$$

The vector $\underline{b}$ has as many components as there are training samples from the two classes. Each component of $\underline{b}$ is an arbitrarily specified positive constant. Duda and Hart[8] show how the following solution for $\underline{a}$ comes from (22).

$$\underline{a} = (H^TH)^{-1} H^T\underline{b} \tag{23}$$

Once the components of $\underline{a}$ have been calculated for each possible pair of classes, then the classifier has been completed.

Widrow-Hoff Procedure

The Widrow-Hoff procedure as explained in Duda and Hart[9] uses an iterative technique to solve for the vector $\underline{a}$ in the equation $H\underline{a} = \underline{b}$ given above. The iterative solution is derived in Duda and Hart[10] and is presented as

$$\underline{a}_{k+1} = \underline{a}_k - \rho_k H^T(H\underline{a}_k - \underline{b}) \tag{24}$$

where $\underline{a}_1$ is arbitrary

$$\rho_k = \rho_1 / k$$

and ρ_1 is any positive constant. For our problem we let ρ_1 be equal to 1. The Widrow Hoff procedure has the advantage over the pseudoinverse technique of being able to obtain a solution even when the matrix H^TH is singular.

---

[7] Ibid.
[8] Ibid.
[9] Ibid.
[10] Ibid.

Relaxation Algorithm

The relaxation algorithm is developed in detail in Duda and Hart[11] and provides a method for obtaining a solution for the vector $\underline{a}$, such that

$$\text{if } \underline{a}^T \hat{\underline{y}} > b \text{ then } \hat{\underline{y}} \in \omega_1 \tag{25}$$

$$\text{if } \underline{a}^T \hat{\underline{y}} < b \text{ then } \hat{\underline{y}} \in \omega_2 \tag{26}$$

where b is a constant. An iterative solution for $\underline{a}$ is given as

$\underline{a}_1$ is arbitrary

$$\underline{a}_{k+1} = \underline{a}_k + \rho \frac{b - \underline{a}^T_k \hat{\underline{y}}_k}{\| \hat{\underline{y}}_k \|^2} \hat{\underline{y}}_k \tag{27}$$

where $\hat{\underline{y}}_k$ is the k^{th} sample point from the training set. $\| \hat{\underline{y}}_k \|$ is the Euclidean norm or magnitude of the vector $\hat{\underline{y}}_k$.

For our problem, b was set equal to 1 and ρ was set equal to 0.5.

Bayes with Normal Distribution

If the probability density functions of the unknown patterns can be assumed to be multivariate normal (Gaussian), the Bayes classifier becomes practical for developing decision functions. The Bayes decision function for normal patterns is given by Tou and Gonzales[12] as

$$d_i(\underline{x}) = \ln p(\omega_i) - \tfrac{1}{2} \ln |\underline{C}_i| - \tfrac{1}{2} [(\underline{x} - \underline{m}_i)^T \underline{C}_i^{-1} (\underline{x} - \underline{m}_i)],$$

$i = 1, 2, \ldots, M$, and M is the total number of classes (28)

where the mean vector $\underline{m}_i = E_i\{\underline{x}\}$, and the covariance matrix $\underline{C}_i = E_i\{(\underline{x} - \underline{m}_i)(\underline{x} - \underline{m}_i)^T\}$. $E_i\{\ \}$ denotes the expectation operator over the patterns of class ω_i. The unknown pattern $\underline{x}$ is assigned to class ω_i if $d_i(\underline{x}) > d_j(\underline{x})$ for all $j \neq i$. The above decision function is derived based upon the assumption of zero loss for correct classifications and equal loss for misclassifications.

The quantity $p(\omega_i)$ represents the a priori probability for the i^{th} class. For our computations, it was assumed that all the a priori probabilities were equal to 1 M.

Minimum Distance Classifier

When all covariance matrices of equation (28) become equal $\underline{C}_i = \underline{C}$ for i = 1, 2, ..., M and also $\underline{C} = \underline{I}$, where $\underline{I}$ is the identity matrix, and $p(\omega_i) = 1$ M, for i = 1, 2, ..., M, then equation (28) reduces to

$$d_i(\underline{x}) = \underline{x}^T \underline{m}_i - 1\ 2\ \underline{m}_i^T\ \underline{m}_i,\ i = 1, 2, \ldots, M \tag{29}$$

---

[11] Ibid.

[12] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, Addison-Wesley, 1974

where a pattern $\underline{x}$ is assigned to class ω_i

if $d_i(\underline{x}) > d_j(\underline{x})$ for all $j \neq i$.

Equation (29) is recognized as the decision functions for a minimum distance pattern classifier as developed in Tou and Gonzalez[13]. The mean vector $\underline{m}_i = E_i\{\underline{x}\}$, as well as the expectation operator E_i, were defined previously.

Results

The 10 pattern classification techniques discussed above were applied to the selected 400 samples of synthetic aperture radar images taken over the Huntsville, Alabama area. The 400 samples were used as a training set to derive each classifier. The 400 samples were then submitted to the classifiers to see how well each one would classify the original training set. This section will present the results of this work for each classifier.

When using the first eight classification techniques, the four classes were considered by taking them two at a time. The final decision was made simply on the basis of which class had received the most votes after all six possibilities had been decided. When computing the second-order histogram statistics to be used as components of the original $\underline{x}$ vector, a spacing between pixels in x and y had to be chosen. This spacing was chosen as a part of the feature selection process, which resulted in a determination of the A matrix for each pair of classes. Table 1 shows the values of the x and y spacings used to compute the second-order histogram statistics. These spacings provided an optimum separation of the feature vector data for each pair of classes.

Table 1. Pixel Spacing for Computing Second-Order Histogram Statistics.

	Spacing in X	Spacing in Y
1. FORESTS and FIELDS	5	0
2. FORESTS and CITIES	-3	4
3. CITIES and FIELDS	2	4
4. FORESTS and WATER	1	0
5. FIELDS and WATER	1	0
6. CITIES and WATER	1	0

In figures 1 through 6 the plots are shown of the $\underline{y}$ data for each pair of classes when the increment correction algorithm was used. The line drawn on each figure represents the computed decision boundary. Tables 2 through 9 present the final results for the first eight classifiers. For each classifier the percentage of correct classification is presented for each class along with an overall percentage of correct classifications.

[13] Ibid.

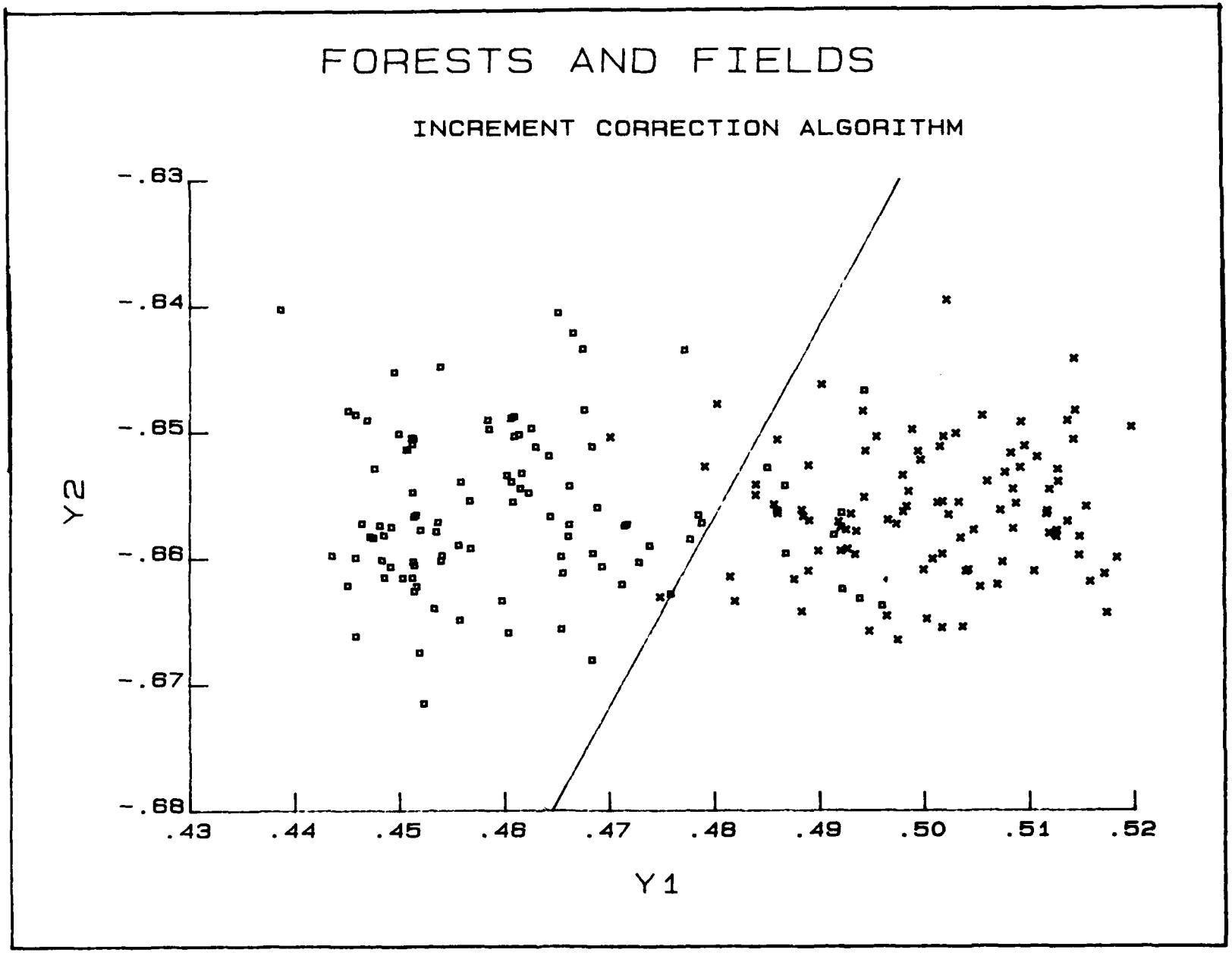


FIGURE 1. Forest and field samples.

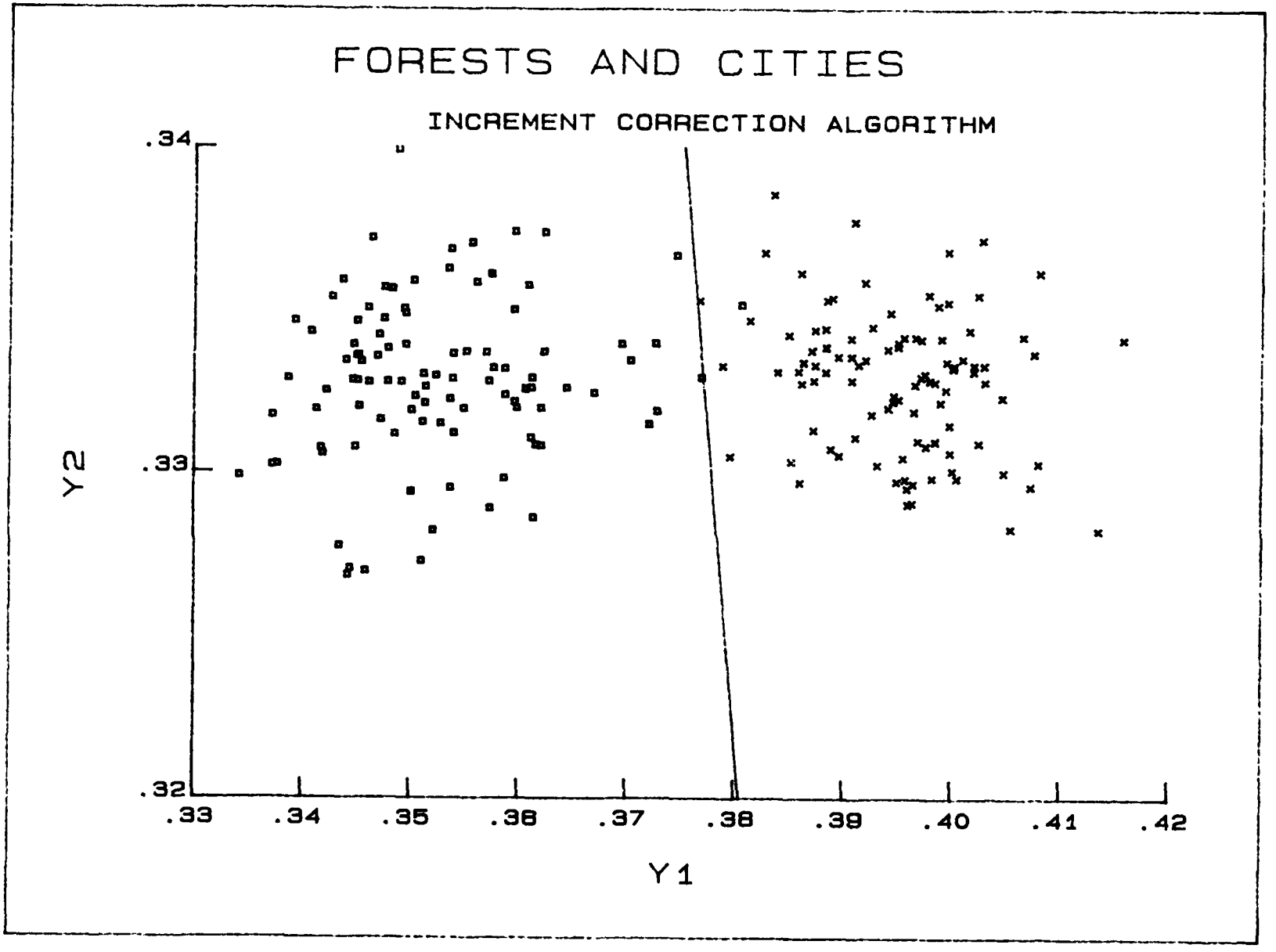


FIGURE 2. Forest and city samples.

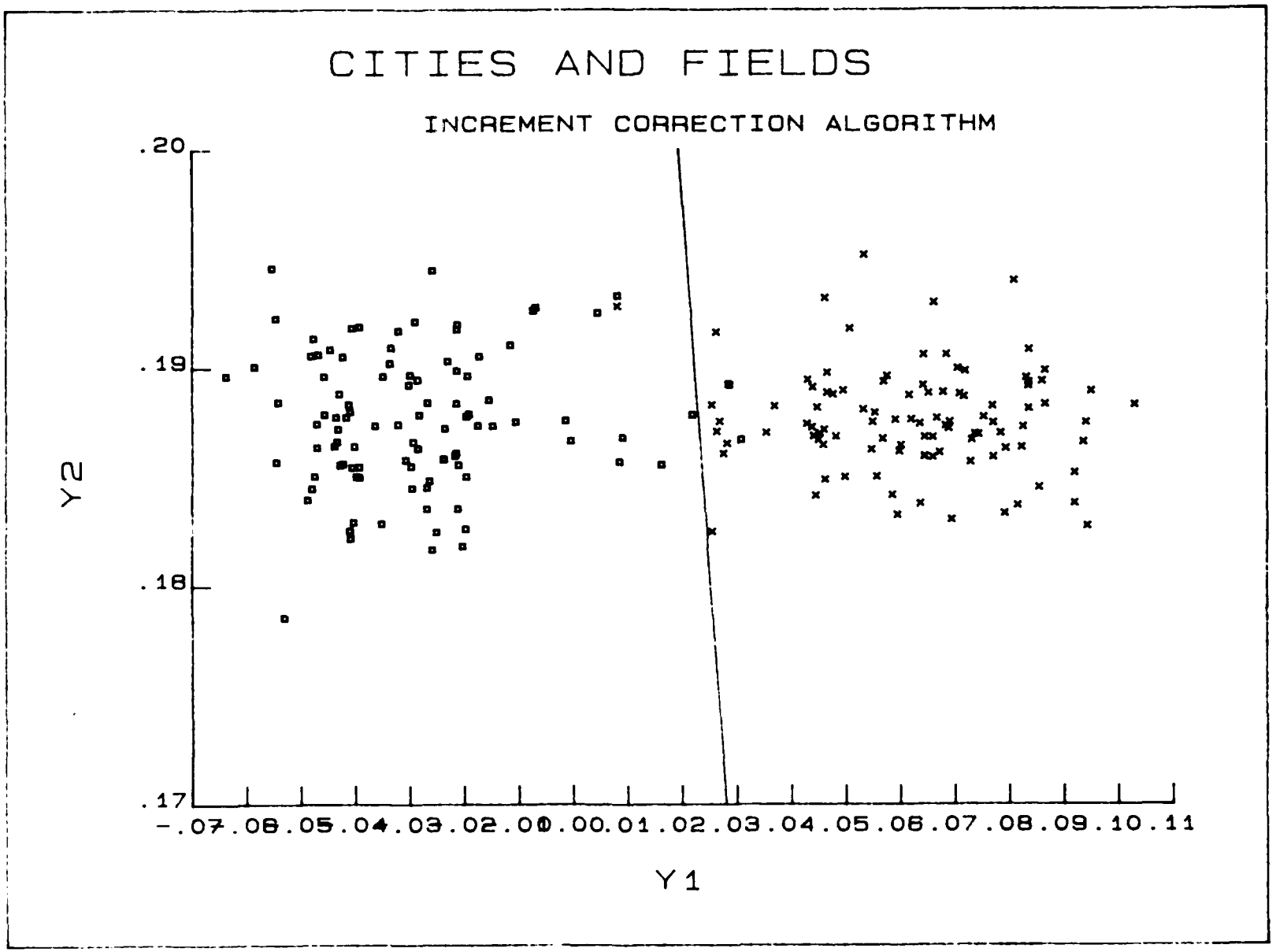


FIGURE 3. City and field samples.

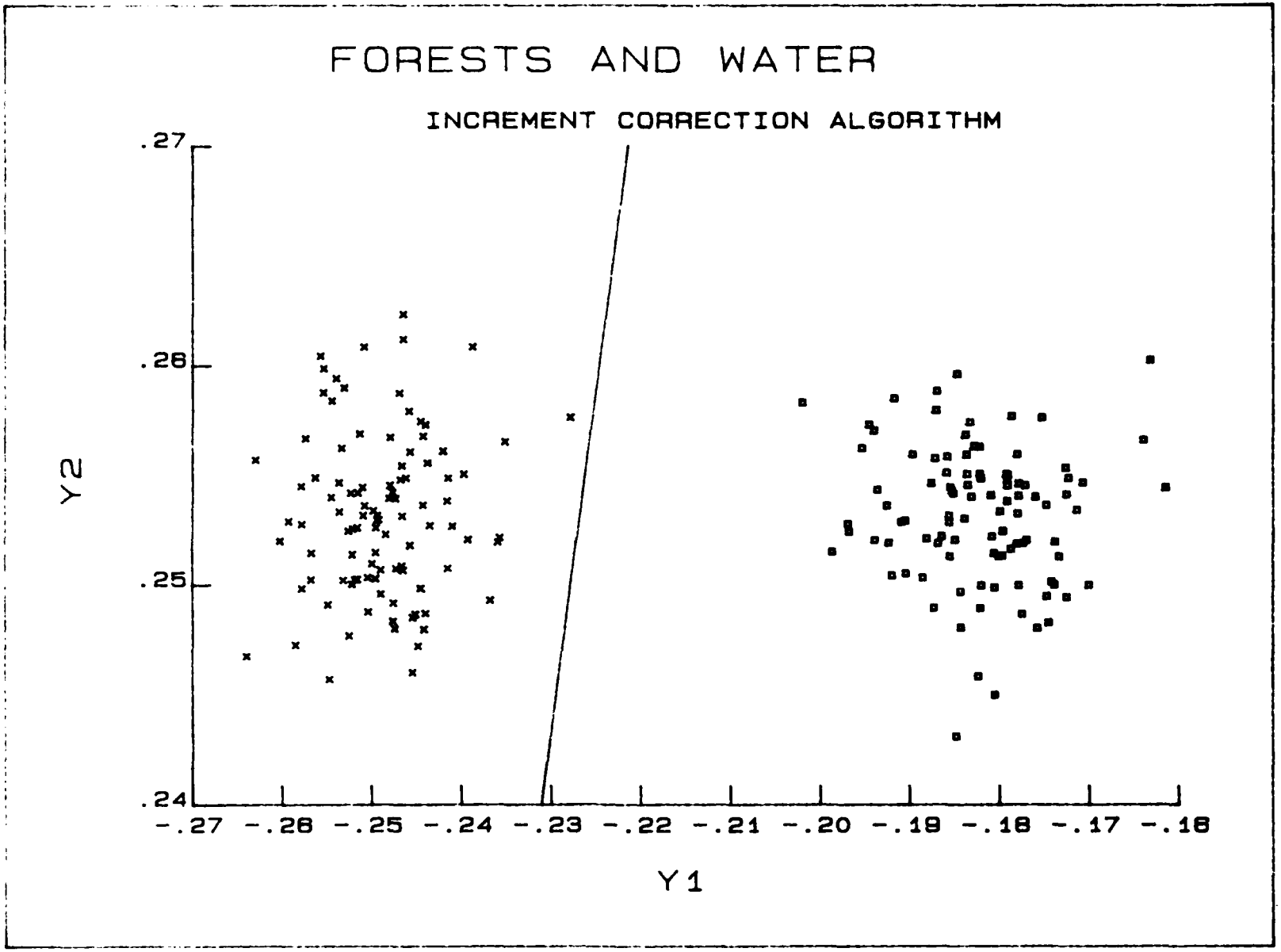


FIGURE 4. Forest and water samples.

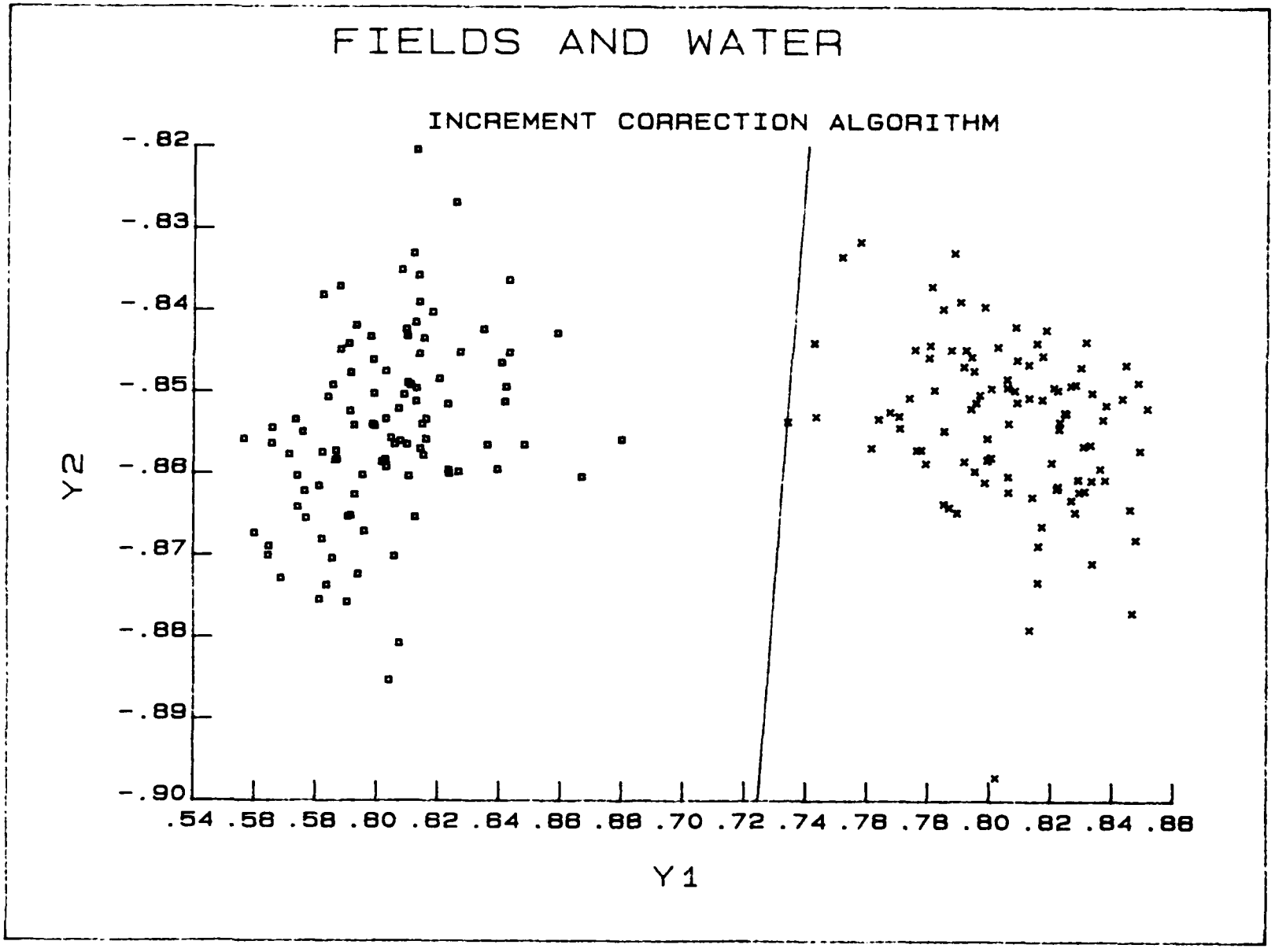


FIGURE 5. Field and water samples.

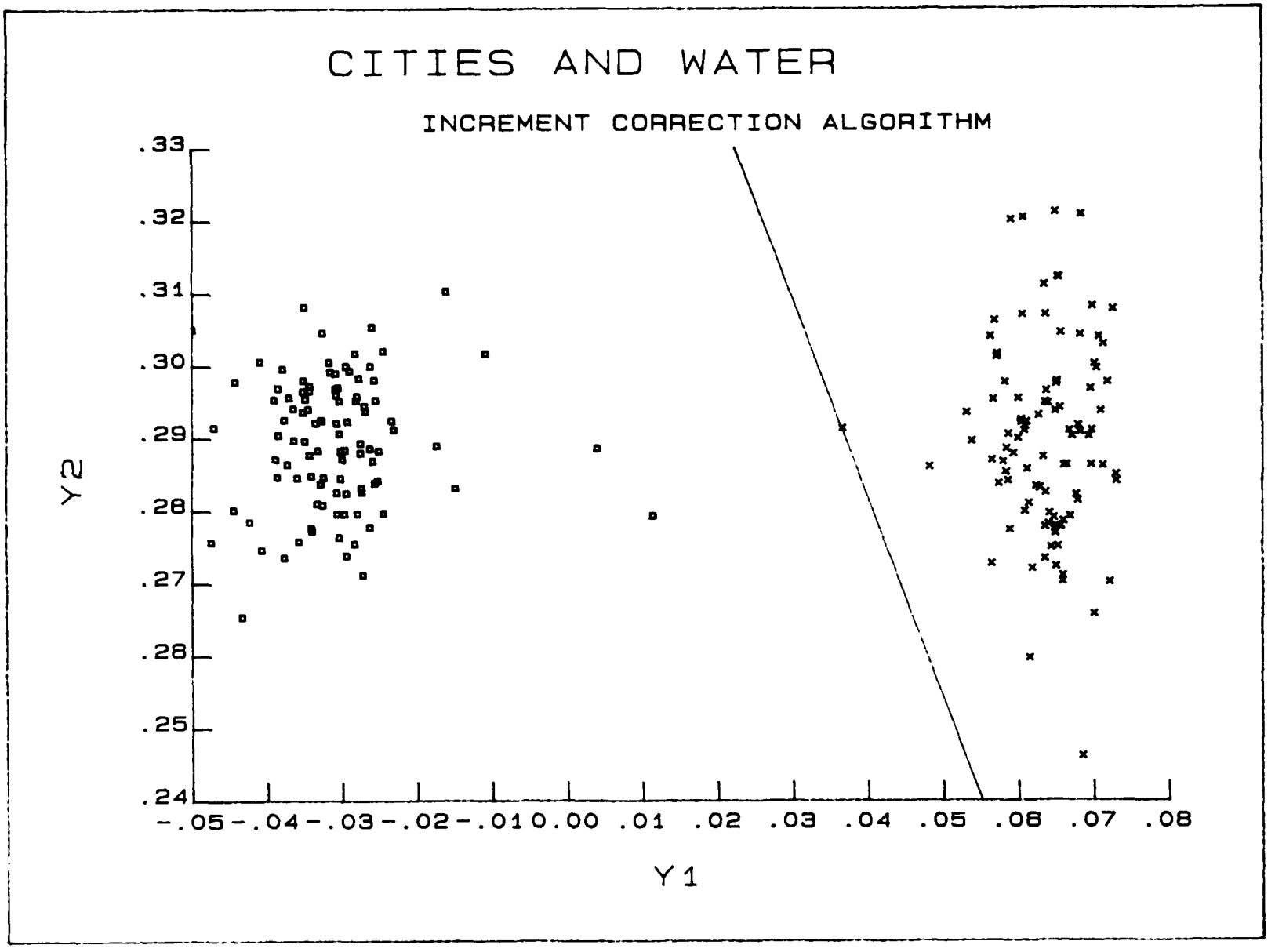


FIGURE 6. City and water samples.

Table 2 — Final Results for the Ho-Kashyap Algorithm.

	Number of Correct Classifications	Number of Incorrect Classifications	Percentage of Correct Classifications
1. FORESTS	96	4	96%
2. FIELDS	90	10	90%
3. CITIES	97	3	97%
4. WATER	100	0	100%

Overall Percentage of Correct Classifications = 95.75%

Table 3 — Final Results for the Increment Correction Algorithm.

	Number of Correct Classifications	Number of Incorrect Classifications	Percentage of Correct Classifications
1. FORESTS	96	4	96%
2. FIELDS	90	10	90%
3. CITIES	98	2	98%
4. WATER	100	0	100%

Overall Percentage of Correct Classifications = 96%

Table 4 — Final Results for the Least Mean Square Error Algorithm.

	Number of Correct Classifications	Number of Incorrect Classifications	Percentage of Correct Classifications
1. FORESTS	96	4	96%
2. FIELDS	90	10	90%
3. CITIES	98	2	98%
4. WATER	100	0	100%

Overall Percentage of Correct Classifications = 96%

Table 5 — Final Results for the Method of Potential Functions.

	Number of Correct Classifications	Number of Incorrect Classifications	Percentage of Correct Classifications
1. FORESTS	98	2	98%
2. FIELDS	87	13	87%
3. CITIES	94	6	94%
4. WATER	97	3	97%

Overall Percentage of Correct Classifications = 94%

Table 6 — Final Results for the Fisher Linear Discriminant Algorithm.

	Number of Correct Classifications	Number of Incorrect Classifications	Percentage of Correct Classifications
1. FORESTS	91	9	91%
2. FIELDS	91	9	91%
3. CITIES	98	2	98%
4. WATER	100	0	100%
Overall Percentage of Correct Classifications = 95%			

Table 7 — Final Results for the Pseudoinverse Technique.

	Number of Correct Classifications	Number of Incorrect Classifications	Percentage of Correct Classifications
1. FORESTS	96	4	96%
2. FIELDS	89	11	89%
3. CITIES	97	3	97%
4. WATER	100	0	100%
Overall Percentage of Correct Classifications = 95.5%			

Table 8 — Final Results for the Widrow-Hoff Procedure.

	Number of Correct Classifications	Number of Incorrect Classifications	Percentage of Correct Classifications
1. FORESTS	96	4	96%
2. FIELDS	89	11	89%
3. CITIES	95	5	95%
4. WATER	100	0	100%
Overall Percentage of Correct Classifications = 95%			

Table 9 — Final Results for the Relaxation Algorithm.

	Number of Correct Classifications	Number of Incorrect Classifications	Percentage of Correct Classifications
1. FORESTS	94	6	94%
2. FIELDS	90	10	90%
3. CITIES	98	2	98%
4. WATER	100	0	100%
Overall Percentage of Correct Classifications = 95.5%			

The results of applying the Bayes classifier to the four hundred samples of radar imagery will now be considered. The classification accuracy was evaluated for each class of terrain feature and for various scanning directions (IDIR)* and inter-pixel spacings (IPS). The IDIR and IPS were the two parameters used to compute the joint-probability matrices during the feature measurement stage, which precedes the classification. The overall classification accuracy was then calculated for each combination of these two parameters and illustrated in table 10. It is noticed that the best overall classification accuracy of 95.5 percent was obtained for the case where the IDIR was 0 degrees and the IPS was 2 pixels. The same accuracy was also obtained for another case where the IDIR was 135 degrees and the IPS was 3 pixels. The least accurate case was 92.75 percent which occurred when the IDIR was 90 degrees and IPS was 1 pixel.

Similar to the previous case, an overall classification accuracy for the minimum distance classifier was computed using the 400 samples of SAR images. As expected, the overall classification accuracy for all cases was inferior to that of the other classifiers. The best overall classification accuracy of 91.25 percent was obtained with a scanning direction of 0 degrees and an inter-pixel spacing of 1 pixel. The worst accuracy of 68.25 percent resulted when the scanning direction was 135 degrees and the inter-pixel spacing was equal to 3 pixels. Table 11 shows the overall classification accuracy for all cases considered.

The final results of all the classification techniques are shown together in table 12.

*IDIR is an acronym for INTEGER DIRECTION and is used as such in the computer program.

IPS / IDIR	1	2	3	4	5	6
1	93.75	95.50	95.25	94.25	94.00	93.25
2	94.75	95.25	94.00	94.00	93.25	93.00
3	92.75	93.25	93.75	94.25	94.00	94.50
4	94.50	95.00	95.50	95.25	95.00	93.75

Note 1. All entries are in percent.
2. IPS: Inter-pixel spacing.
3. IDIR: Scanning direction 1 = 0 degree, 2 = 45 degrees, 3 = 90 degrees, and 4 = 135 degrees.

Table 10. Overall classification accuracy for radar imagery consisting of fields, water, forests, and cities with a bayes classifier.

IDIR \ IPS	1	2	3	4	5	6
1	91.25	81.25	80.75	71.50	71.25	70.25
2	82.50	70.00	89.75	87.00	88.75	88.75
3	88.89	74.50	71.00	71.00	77.50	71.50
4	83.00	70.25	88.25	88.75	88.50	89.00

Note 1. All entries are in percent.
2. IPS: Inter-pixel spacing.
3. IDIR: Scanning direction 1 = 0 degree, 2 = 45 degrees, 3 = 90 degrees, and 4 = 135 degrees.

Table 11. Overall classification accuracy for radar imagery consisting of fields, water, forests, and cities with a minimum distance classifier.

Table 12. Final Results for all Classification Techniques.

Classification Techniques	Overall Percentage of Correct Classifications
Ho-Kashyap Algorithm	95.75%
Increment Correction Algorithm	96%
Least Mean Square Error Algorithm	96%
Method of Potential Functions	94%
Fisher Linear Discriminant	95%
Pseudoinverse Technique	95.5%
Widrow-Hoff Procedure	95%
Relaxation Algorithm	95.5%
Bayes with Normal Distribution	95.5%
Minimum Distance Classifer	91.25%

As can be seen from table 12, the results from all the classifiers are very close and no one technique stands out from all the rest. The method of potential functions was found to be quite complicated and computationally extensive even though the end result was similar to other techniques.

Conclusions

1. The results of applying the 10 pattern classification techniques to a limited set of radar image samples yielded a correct classification rate between 91.25 percent and 96.00 percent for the training samples used.

2. Even though all 10 classification techniques yielded similar results, they were not all of equal computational complexity.

3. The method of potential functions was found to be difficult to implement and computationally intensive.

4. No relationship was established between the percentage of correct classifications and the number of samples used for training. This was due to the fact that the number of training samples used for all phases of this work was constant.

Appendix A. Feature Vector Components

The following first-and second-order histogram measures were used to construct feature vectors. The first 13 measures were used to form a feature vector for the first eight classification techniques. The last two classifiers used all 15 measures.

Mean $\bar{b} = \sum_{b=o}^{L-1} bP(b) = x_1$

Variance $\sigma_b^2 = \sum_{b=o}^{L-1} (b-\bar{b})^2 P(b) = x_2$

Skewness $b_S = \frac{1}{\sigma_b^3} \sum_{b=o}^{L-1} (b-\bar{b})^3 P(b) = x_3$

Kurtosis $b_K = \frac{1}{\sigma_b^4} \sum_{b=o}^{L-1} (b-\bar{b})^4 P(b)-3 = x_4$

Energy $b_N = \sum_{b=o}^{L-1} [P(b)]^2 = x_5$

Entropy $b_E = - \sum_{b=o}^{L-1} P(b) \log_2[P(b)] = x_6$

Autocorrelation $B_A = \sum_{a=o}^{L-1} \sum_{b=o}^{L-1} abP(a,b) = x_7$

Covariance $B_C = \sum_{a=o}^{L-1} \sum_{b=o}^{L-1} (a-\bar{a})(b-\bar{b}) P(a,b) = x_8$

Inertia $B_I = \sum_{a=o}^{L-1} \sum_{b=o}^{L-1} (a-b)^2 P(a,b) = x_9$

Absolute Value $B_V = \sum_{a=o}^{L-1} \sum_{b=o}^{L-1} |a-b| P(a,b) = x_{10}$

Inverse Difference $$B_D = \sum_{a=o}^{L-1} \sum_{b=o}^{L-1} \frac{P(a,b)}{1 + (a-b)^2} = x_{11}$$

Energy $$B_N = \sum_{a=o}^{L-1} \sum_{b=o}^{L-1} [P(a,b)]^2 = x_{12}$$

Entropy $$B_E = - \sum_{a=o}^{L-1} \sum_{b=o}^{L-1} P(a,b) \log_2 [P(a,b)] = x_{13}$$

Mean $$\bar{b} = \sum_{a=o}^{L-1} \sum_{b=o}^{L-1} bP(a,b) = x_{14}$$

Variance $$V_b = \sum_{a=o}^{L-1} \sum_{b=o}^{L-1} (b-\bar{b})^2 P(a,b) = x_{15}$$

where L is the number of gray levels and P(b) and P(a,b) are given as

$$P(b) = \frac{Q(b)}{M}$$

M is the total number of pixels in the sample window. In this case M was equal to 1024. Q(b) is the number of pixels of gray tone b that occur in the sample window.

$$P(a,b) = \frac{N(a,b)}{M}$$

N(a,b) is the number of times gray tone a is located next to gray tone b by the displacement Δx and Δy.

GLOSSARY

This report contains a large number of symbols which tend to be easily confused unless a strict definition is held for each one. This section will explain the symbols used most frequently in the report.

Symbol	Explanation
$\underline{x}$	Original feature vector consisting of thirteen or fifteen components that are calculated from the first and second order histogram statistics of the image samples.
A	Feature selection transformation matrix used to reduce the dimensionality of $\underline{x}$ from thirteen to two and to optimize the separation between classes. This matrix has the dimensionality of two by thirteen.
$\underline{y}$	Transformed feature vector consisting of two components.
$\hat{\underline{y}}$	Transformed feature vector augmented by *1* and consisting of three components.
$\underline{w}$	Weight vector consisting of three components which are determined from the training samples. The method of solving for $\underline{w}$ is determined by the particular pattern classification technique used.
N	Total number of pattern points for the two classes.
$\underline{b}$	Vector consisting of N arbitrary positive constants.
Y	Matrix obtained from the training samples taken from the two classes. Each row of Y consists of a sample of $\hat{\underline{y}}^T$, where the samples coming from class two are multiplied by -1.
c	A positive number between zero and one.
ω_1	A representation for pattern class one.
ω_2	A representation for pattern class two.
ϵ	Belongs to or is in.
$\notin$	Does not belong to or is not in.
$P(\omega_i \mid \hat{\underline{y}})$	The a posteriori probability for class ω_i given that the vector $\hat{\underline{y}}$ has been calculated.
α_k	A member of a sequence of positive numbers which satisfy the following three conditions:

Symbol	Explanation
	1. $\lim_{k \to \infty} \alpha_k = 0$
	2. $\sum_{k=1}^{\infty} \alpha_k = \infty$
	3. $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$

The sequence used in this report which satisfies the above conditions is the harmonic sequence $(1/k) = (1, 1/2, 1/3, \ldots)$.

$\phi_j(\hat{\underline{y}})$	A given set of orthonormal functions. In this report the $\phi_j(\hat{\underline{y}})$ were chosen to be a set of Hermite polynomial functions.
$c_j(k)$	Unknown coefficients in an expansion which is used to approximate the a posteriori probability. A solution for these coefficients is obtained by the method of potential functions.
$\forall$	For all values of.
S_W	The within class scatter matrix calculated from the original vectors ($\underline{x}$) and the mean vectors.
$\underline{m}_1$	Mean vector for class one.
$\underline{m}_2$	Mean vector for class two.
$\underline{a}$	Weight vector consisting of three components. A solution for $\underline{a}$ depends on the particular pattern classification technique used.
H	Matrix obtained from the training samples taken from the two classes. Each row of H consists of a sample of $\hat{\underline{y}}^T$, where the samples coming from class two are multiplied by -1. Identical with the definition of Y.
$\left\| \hat{\underline{y}}_k \right\| = \left[\sum_{j=1}^{3} \hat{y}_j^2 \right]^{1/2}$	Euclidean norm or magnitude of the vector $\hat{\underline{y}}_k$.
$d_i(x)$	Decision function.
ln	Logrithm to the base e.
$p(\omega_i)$	A priori probability of the class ω_i.
$\underline{C}_i$	Covariance matrix for the class ω_i.

END

9-87

DTIC