N O T I C E

THIS DOCUMENT HAS BEEN REPRODUCED FROM MICROFICHE. ALTHOUGH IT IS RECOGNIZED THAT CERTAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RELEASED IN THE INTEREST OF MAKING AVAILABLE AS MUCH INFORMATION AS POSSIBLE

AgRISTARS

SR-LO-00419
JSC-16335
JUN 2 6 1980 NASA CR-
160737

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Supporting Research

June 1980

UTILIZATION OF SPECTRAL-SPATIAL INFORMATION
IN THE CLASSIFICATION OF IMAGERY DATA

C. B. Chittineni

LOCKHEED ENGINEERING AND MANAGEMENT SERVICES COMPANY, INC.
1830 NASA Road 1, Houston, Texas 77058

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
JSC-16335; SR-L0-00419		

4. Title and Subtitle	5. Report Date
Utilization of Spectral-Spatial Information in the Classification of Imagery Data	June 1980
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
C. B. Chittineni Lockheed Engineering and Management Services Company, Inc.	LEMSCO-14310
	10. Work Unit No.

9. Performing Organization Name and Address	11. Contract or Grant No.
Lockheed Engineering and Management Services Company, Inc. 1830 NASA Road 1 Houston, Texas 77058	NAS 9-15800

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
National Aeronautics and Space Administration Lyndon B. Johnson Space Center *Odenwelder* Houston, Texas 77058 Tech. Monitor: ~~Dr. R. Heydorn~~	Technical Report
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

In this paper, the problem of incorporating spatial or contextual information into the classification of imagery data is considered. Simple models which describe the spatial dependencies between the neighboring picture elements with a single parameter θ are presented. Techniques are developed for obtaining the optimal value of θ as a maximum likelihood estimate from the local neighborhood of the pattern under consideration. Expressions are derived for updating the a posteriori probabilities of the states of nature of the pattern under consideration, using information from the neighboring patterns, both for spatially uniform context and for Markov dependencies in terms of θ. Furthermore, the results of applying these techniques in the classification of remotely sensed multispectral scanner imagery data are presented.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Classification map, contextual classification, Davidon-Fletcher-Powell method, local neighborhood, Markov dependence, maximum likelihood estimation, pattern recognition, spatial information, transition probabilities model	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	69	

UTILIZATION OF SPECTRAL-SPATIAL INFORMATION IN THE CLASSIFICATION OF IMAGERY DATA

Job Order 73-302

This report describes Classificat .1 activities of the Supporting Research project of the AgRISTARS program.

PREPARED BY

C. B. Chittineni

APPROVED BY
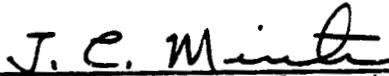
J. C. Minter

T. C. Minter, Supervisor
Techniques Development Section

J. E. Wainwright

J. E. Wainwright, Manager
Development and Evaluation Department

PREFACE

The techniques which are the subject of this report were developed to support the Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing program. Under Contract NAS 9-15800, Dr. C. B. Chittineni, a principal scientist for Lockheed Engineering and Management Services Company, Inc., performed this research for the Earth Observations Division, Space and Life Sciences Directorate, National Aeronautics and Space Administration, at the Lyndon B. Johnson Space Center, Houston, Texas.

CONTENTS

PRECEDING PAGE BLANK NOT FILMED

TABLES

FIGURES

1. INTRODUCTION

Recently, considerable interest has been shown in the development of techniques for the classification of imagery data such as remote sensing data obtained using the multispectral scanner (MSS) on board the Landsat. Classification of multichannel imagery data is typically done by applying a decision rule to each resolution element or picture element (pixel) and classifying it based on spectral information. This procedure ignores spatial information. Most of the imagery data contain much spatial information which can be used to improve computer-assisted classification.

The use of contextual information in pattern classification has attracted the attention of many researchers, mainly in the area of character recognition (refs. 1, 2). Generally, one of two basic approaches has been used, the table lookup method or the Markov approach. The table lookup method is based on the assumption that every word to be recognized is selected from a known finite table. A word is classified by comparing it with every word of the same length in the table and finding the best match.

The Markov approach is based on the assumption that the true category of a character is related in a probabilistic manner to the true categories of a small number of surrounding characters. Its use requires the estimation of the probability of occurrence of all possible pairs, triplets of characters, etc., from the sample text. Abend (ref. 3) derived optimal procedures when a Markov dependence exists between the states of nature of neighboring characters, and Raviv (ref. 4) gives the results of applying such procedures for the recognition of English text.

The use of contextual information in speech recognition is considered by Alter (ref. 5). Chow (ref. 6), using a nearest neighbor dependence method, obtained the structure and parameters of a recognition network for patterns represented by binary matrices.

Several researchers attempted to use spatial information in the classification of imagery data. Kettig and Landgrebe (ref. 7) developed a technique called

1-1

1

Extraction and Classification of Homogeneous Objects (ECHO), which segments a scene into homogeneous objects and uses sample classification to assign each object as a whole, rather than by its individual pixels. Haralick et al. (ref. 8) used textural features based on gray-tone spatial dependence matrices to characterize a local scene texture and experimentally showed them to be useful for classification purposes. Swain (ref. 9) developed a cascade model for classifying a pattern based on multiple observations in a time-varying environment. Welch and Salter (ref. 10) presented a method for the contextual classification of imagery data. Chittineni (ref. 11) discusses the use of context with linear classifiers. Toussaint (ref. 12) gives a brief review of the use of context in pattern recognition and presents an extensive list of references on the subject.

All of the approaches proposed in the literature either use arbitrarily selected transition probabilities or estimate them from a sample and treat them as global. For imagery data such as those obtained in remote sensing, the transition probabilities very often not only vary from one image to the other but also vary from one local neighborhood to the other in the same image. It is difficult to obtain global estimates of transition probabilities because of the varying nature of imagery and the nonavailability of true classes of pixels of images.

It is the purpose of this paper to develop methods for locally estimating transition probabilities and to use these estimates in contextual classification. It is assumed that the classifier is trained on representative data from the image and, for every pixel of the image, the a posteriori probabilities of the classes are estimated from spectral information. Thus, the incorporation of contextual information into classification is treated as a postprocessing operation.

The number of transition probabilities to be estimated increases as the square of the number of classes. Mathematical expressions for contextual classification become complex with the increase in the size of the local neighborhood. Thus, making the estimation of transition probabilities is computationally expensive. In this paper, the transition probabilities are modeled in terms

of a single parameter θ, under reasonable assumptions, and methods are developed for the estimation of θ. The estimated θ is then used for the incorporation of spatial information into classification.

The paper is organized as follows. Models for transition probabilities in terms of a single parameter θ are developed in section 2. Techniques for locally estimating the parameter θ of transition probabilities using the maximum likelihood method are developed in section 3. Section 4 presents expressions for using the contextual information in classification. Section 5 presents the results of contextual classification of remotely sensed agricultural imagery data, using techniques developed in the paper. Conclusions are presented in section 6. Appendix A presents an extension of spatially uniform context to large neighborhoods. In appendix B, expressions are developed for estimating transition probabilities for a two-class, three-sequential-neighborhood case without using models. The results of estimating the parameters of transition probabilities under different transition probabilities models in different directions in the local neighborhood are presented in appendix C. Appendix D presents a multitemporal interpretation of the techniques developed in the paper for applications such as in remote sensing.

2. MODELING TRANSITION PROBABILITIES

The models for the transition probabilities of the classes of the neighboring pixels, in terms of a single parameter θ, are developed under reasonable assumptions. Let i and j be the neighboring pixels as shown in figure 2-1.

$$
\begin{array}{cc}
X_i & X_j \\
\omega_i & \omega_j
\end{array}
$$
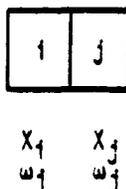
Figure 2-1.- Neighboring pixels i and j.

Let X_i and X_j be the pattern vectors and ω_i and ω_j be the labels (classes) of pixels i and j, respectively. Let ω_i and ω_j take values of 1, 2, $\cdots$, M; where M is the number of classes.

A linear model describing the dependency between the neighboring pixels in terms of a single parameter θ for different r and s is given in equation (1).

$$
\left.
\begin{aligned}
P(\omega_i = r | \omega_j = s) &= (1 - \theta)P(\omega_i = r) \\
P(\omega_i = r | \omega_j = r) &= (1 - \theta)P(\omega_i = r) + \theta \\
0 &< \theta < 1
\end{aligned}
\right\}
\tag{1}
$$

For $\theta = 0$, equation (1) becomes

$$
\left.
\begin{aligned}
P(\omega_i = r | \omega_j = s) &= P(\omega_i = r) \\
P(\omega_i = r | \omega_j = r) &= P(\omega_i = r)
\end{aligned}
\right\}
\tag{2}
$$

and

Equation (2) is the case where the labels of neighboring pixels are independent. For $\theta = 1$, equation (1) becomes

$$
\left.
\begin{aligned}
P(\omega_i = r | \omega_j = s) &= 0 \\
P(\omega_i = r | \omega_j = r) &= 1
\end{aligned}
\right\}
\tag{3}
$$

and

Equation (3) is the case where the labels of the neighboring pixels are completely dependent. Notice that the linear transition probabilities model of equation (1) is a linear interpolation in terms of a single parameter θ between the extremes of equations (2) and (3). It can be easily shown that the model of equation (1) satisfies the postulates of probabilities. That is,

$$\left. \begin{array}{c} 0 < P(\omega_i = r | \omega_j = s) < 1 \\[2mm] \sum_{r=1}^{M} P(\omega_i = r | \omega_j = s) = 1 \end{array} \right\} \qquad (4)$$

and

Using a quadratic interpolation between the extremes of equations (2) and (3), a quadratic model describing the dependencies between the labels of neighboring pixels can be written in terms of a single parameter θ as

$$\left. \begin{array}{c} P(\omega_i = r | \omega_j = s) = (1 - \theta)^2 P(\omega_i = r) \\[2mm] P(\omega_i = r | \omega_j = r) = (1 - \theta)^2 P(\omega_i = r) + \theta(2 - \theta) \\[2mm] 0 < \theta < 1 \end{array} \right\} \qquad (5)$$

The model of equation (5) also satisfies the postulates of probabilities.

However, it is to be noted that the dependencies between the neighboring pixels can be modeled through some other parameter. For example, by replacing θ with $\dfrac{\alpha}{1 + \alpha}$, the dependencies are described in terms of α, $0 < \alpha < \infty$; by replacing θ with $\dfrac{e^{-\beta}}{1 + e^{-\beta}}$, the dependencies are described in terms of β, $-\infty < \beta < \infty$.

The transition probabilities between the classes of the neighboring pixels i and j also can be modeled to satisfy the following characteristics of dependencies, resulting in a nonlinear model. Some of the general characteristics of dependencies between neighboring pixels i and j can be written as follows.

5

a. If the label $\omega_i = r$ of pixel i frequently occurs concurrently with the label $\omega_j = s$ of pixel j, then

$$P(\omega_i = r | \omega_j = s) > P(\omega_i = r) \qquad (6)$$

and, if they always occur concurrently, then

$$P(\omega_i = r | \omega_j = s) = 1 \qquad (7)$$

b. If the label $\omega_i = r$ of pixel i rarely occurs concurrently with the label $\omega_j = s$ of pixel j, then

$$P(\omega_i = r | \omega_j = s) < P(\omega_i = r) \qquad (8)$$

and, if they never occur concurrently, then

$$P(\omega_i = r | \omega_j = s) = 0 \qquad (9)$$

c. If the label $\omega_i = r$ of pixel i occurs independently of the label $\omega_j = s$ of pixel j, then

$$P(\omega_i = r | \omega_j = s) = P(\omega_i = r) \qquad (10)$$

A model satisfying characteristics a, b, and c can be written in terms of a single parameter θ for different r and s as

$$
\left.
\begin{aligned}
P(\omega_i = r | \omega_j = s) &= \frac{(1 - \theta)P(\omega_i = r)}{(1 - \theta) + \theta P(\omega_j = s)} \\[2em]
P(\omega_i = r | \omega_j = r) &= \frac{P(\omega_i = r)}{(1 - \theta) + \theta P(\omega_j = r)} \\[1em]
&\quad -\infty < \theta < 1
\end{aligned}
\right\} \qquad (11)
$$

It can be easily shown that the transition probabilities described by equation (11) satisfy the postulates of probabilities. Also, notice that requirements a, b, and c on the transition probabilities correspond to the cases where $\theta < 0$, $\theta > 0$, and $\theta = 0$, respectively. The model of equation (11) is referred to in this paper as the nonlinear transition probabilities model. Figure 2-2 illustrates the linear, quadratic, and nonlinear transition probabilities models.

2-3
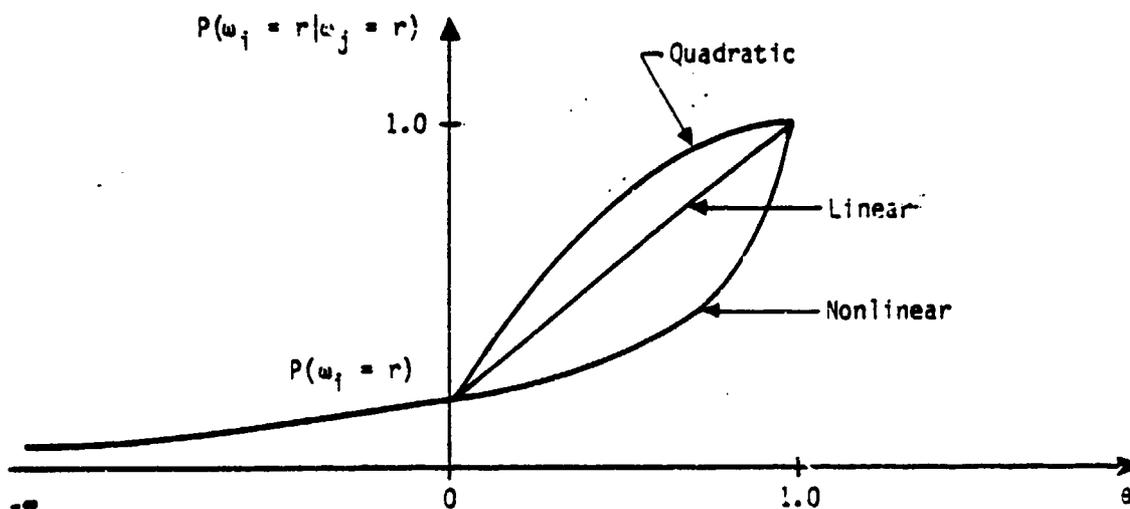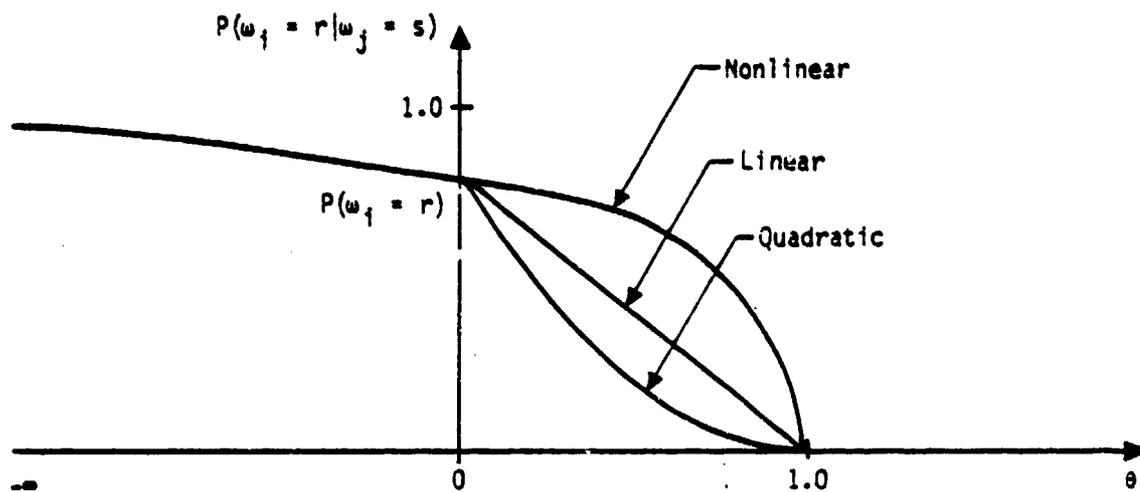
Figure 2-2.- Illustration of spatial dependency models.

In the remainder of the paper, only the linear model of equation (1) and the nonlinear model of equation (11) are considered.

2-4

7

3. LOCAL NEIGHBORHOOD ESTIMATION OF TRANSITION PROBABILITIES

In this section, techniques are developed for the estimation of transition probabilities in the local neighborhood of the pixel under consideration for use in its contextual classification. The criterion used for their estimation is the likelihood function. That is, the transition probabilities are estimated as those that maximize the likelihood function of observed spectral vectors, if their spatial relationships are as given in the local neighborhood.

3.1 A GENERAL EXPRESSION FOR THE LIKELIHOOD FUNCTION

An expression for the likelihood function of N patterns in a general local neighborhood is developed in the following. Let X_1, X_2, $\cdots$, X_N be the patterns in the general neighborhood. The likelihood function of these patterns can be written as

$$L' = p(X_1, X_2, \cdots, X_N)$$

$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_N=1}^{M} p(X_1, \omega_1 = i_1; X_2, \omega_2 = i_2; \cdots; X_N, \omega_N = i_N)$$

$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_N=1}^{M} p(X_1, X_2, \cdots, X_N | \omega_1 = i_1, \omega_2 = i_2, \cdots, \omega_N = i_N)$$

$$P(\omega_1 = i_1, \omega_2 = i_2, \cdots, \omega_N = i_N) \tag{12}$$

where M is the number of classes. In the following, it is assumed that (a) the probability density function of a pattern, given its label, is independent of other patterns and their labels and (b) the labels of the patterns are independent of the labels of their nonneighbors. By repeatedly using assumption (a), the following is obtained. Consider

$$p(X_1, X_2, \cdots, X_N | \omega_1 = i_1, \omega_2 = i_2, \cdots, \omega_N = i_N)$$

$$= p(X_1 | X_2, \cdots, X_N; \omega_1 = i_1, \cdots, \omega_N = i_N) p(X_2, \cdots, X_N | \omega_1 = i_1, \cdots, \omega_N = i_N)$$

$$= p(X_1 | \omega_1 = i_1) p(X_2 | X_3, \cdots, X_N; \omega_1 = i_1, \cdots, \omega_N = i_N) p(X_3, \cdots, X_N | \omega_1 = i_1, \cdots, \omega_N = i_N)$$

$$= p(X_1 | \omega_1 = i_1) p(X_2 | \omega_2 = i_2) p(X_3, \cdots, X_N | \omega_1 = i_1, \cdots, \omega_N = i_N)$$

$$= \prod_{j=1}^{N} p(X_j | \omega_j = i_j) \tag{13}$$

Using equation (13) in equation (12) results in

$$L' = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_N=1}^{M} \left[\prod_{j=1}^{N} p(X_j | \omega_j = i_j) \right] P(\omega_1 = i_1, \cdots, \omega_N = i_N) \qquad (14)$$

Since $\left[\prod_{j=1}^{N} p(X_j) \right]$ is independent of the transition probabilities, dividing both sides of equation (14) by it yields the criterion L to be used for estimating the transition probabilities. That is,

$$L = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_N=1}^{M} \left[\prod_{j=1}^{N} \frac{p(\omega_j = i_j | X_j)}{P(\omega_j = i_j)} \right] P(\omega_1 = i_1, \omega_2 = i_2, \cdots, \omega_N = i_N)$$

$$(15)$$

$P(\omega_1 = i_1, \cdots, \omega_N = i_N)$ depends on the particular local neighborhood and will be considered in detail in the following.

3.2 SPATIALLY UNIFORM CONTEXT — FOUR NEIGHBORS

The pixel under consideration, pixel 0, and its four neighbors in a two-dimensional local neighborhood are shown in figure 3-1.
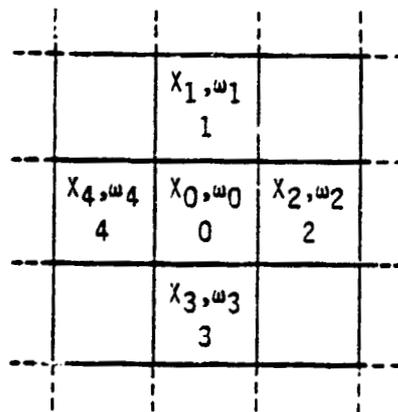
Figure 3-1.- Four neighbors of pixel 0.

By repeatedly using assumption (b), equation (16) is obtained. From equations (15) and (16), an expression for L for the local neighborhood of figure 3-1 is obtained, as shown in equation (17).

$$P(\omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4)$$

$$= P(\omega_0 = i_0)P(\omega_1 = i_1, \cdots, \omega_4 = i_4 | \omega_0 = i_0)$$

$$= P(\omega_0 = i_0)P(\omega_1 = i_1 | \omega_0 = i_0, \omega_2 = i_2, \cdots, \omega_4 = i_4)P(\omega_2 = i_2, \cdots, \omega_4 = i_4 | \omega_0 = i_0)$$

$$= P(\omega_0 = i_0)P(\omega_1 = i_1 | \omega_0 = i_0)P(\omega_2 = i_2 | \omega_0 = i_0, \omega_3 = i_3, \omega_4 = i_4)P(\omega_3 = i_3, \omega_4 = i_4 | \omega_0 = i_0)$$

$$= P(\omega_0 = i_0)\prod_{j=1}^{4} P(\omega_j = i_j | \omega_0 = i_0) \tag{16}$$

$$L = \sum_{i_0=1}^{M}\sum_{i_1=1}^{M}\cdots\sum_{i_4=1}^{M}\left[\prod_{j=0}^{4}\frac{P(\omega_j = i_j | X_j)}{P(\omega_j = i_j)}\right]\left[P(\omega_0 = i_0)\prod_{\ell=1}^{4}P(\omega_\ell = i_\ell | \omega_0)\right]$$

$$= \sum_{i_0=1}^{M}P(\omega_0 = i_0 | X_0)\left\{\prod_{j=1}^{4}\left[\sum_{i_j=1}^{M}\frac{P(\omega_j = i_j | X_j)}{P(\omega_j = i_j)}P(\omega_j = i_j | \omega_0 = i_0)\right]\right\} \tag{17}$$

3.2.1 AN EXPRESSION FOR L WITH LINEAR TRANSITION PROBABILITIES MODEL

Since a priori probabilities are position independent, when a linear model of
equation (1) is used for transition probabilities in equation (17), L becomes

$$L = \sum_{i_0=1}^{M} p(\omega_0 = i_0|X_0) \prod_{j=1}^{4} \left[\sum_{\substack{i_j=1 \\ i_j \neq i_0}}^{M} \frac{p(\omega_j = i_j|X_j)}{P(\omega_j = i_j)} P(\omega_j = i_j|\omega_0 = i_0) \right.$$

$$\left. + \frac{p(\omega_j = i_0|X_j)}{P(\omega_j = i_0)} P(\omega_j = i_0|\omega_0 = i_0) \right]$$

$$= \sum_{i_0=1}^{M} p(\omega = i_0|X_0) \left\{ \prod_{j=1}^{4} \left[(1 - \theta) \sum_{i_j=1}^{M} p(\omega = i_j|X_j) + \theta \frac{p(\omega = i_0|X_j)}{P(\omega = i_0)} \right] \right\}$$

$$= \sum_{i_0=1}^{M} p(\omega = i_0|X_0) \left\{ \prod_{j=1}^{4} \left[(1 - \theta) + \theta \frac{p(\omega = i_0|X_j)}{P(\omega = i_0)} \right] \right\}$$

$$= (1 - \theta)^4 + \theta(1 - \theta)^3 A + \theta^2(1 - \theta)^2 B + \theta^3(1 - \theta)C + \theta^4 D \qquad (18)$$

where

$$A = \sum_{i_0=1}^{M} \frac{p(\omega = i_0|X_0)}{P(\omega = i_0)} [p(\omega = i_0|X_1) + p(\omega = i_0|X_2)$$

$$+ p(\omega = i_0|X_3) + p(\omega = i_0|X_4)]$$

$$B = \sum_{i_0=1}^{M} \frac{p(\omega = i_0|X_0)}{P^2(\omega = i_0)} [p(\omega = i_0|X_1)p(\omega = i_0|X_2) + p(\omega = i_0|X_1)p(\omega = i_0|X_3)$$

$$+ p(\omega = i_0|X_1)p(\omega = i_0|X_4) + p(\omega = i_0|X_2)p(\omega = i_0|X_3)$$

$$+ p(\omega = i_0|X_2)p(\omega = i_0|X_4) + p(\omega = i_0|X_3)p(\omega = i_0|X_4)]$$

$$C = \sum_{i_0=1}^{M} \frac{p(\omega = i_0 | X_0)}{P^3(\omega = i_0)} [p(\omega = i_0 | X_1) p(\omega = i_0 | X_2) p(\omega = i_0 | X_3)$$

$$+ p(\omega = i_0 | X_1) p(\omega = i_0 | X_2) p(\omega = i_0 | X_4)$$

$$+ p(\omega = i_0 | X_1) p(\omega = i_0 | X_3) p(\omega = i_0 | X_4)$$

$$+ p(\omega = i_0 | X_2) p(\omega = i_0 | X_3) p(\omega = i_0 | X_4)]$$

$$D = \sum_{i_0=1}^{M} \frac{p(\omega = i_0 | X_0)}{P^4(\omega = i_0)} [p(\omega = i_0 | X_1) p(\omega = i_0 | X_2) p(\omega = i_0 | X_3) p(\omega = i_0 | X_4)]$$

3.2.2 AN EXPRESSION FOR L WITH NONLINEAR TRANSITION PROBABILITIES MODEL

Using the nonlinear transition probabilities model of equation (11) in equation (18) gives the following expression for L.

$$L = \sum_{i_0=1}^{M} p(\omega_0 = i_0 | X_0) \prod_{j=1}^{4} \left\{ \frac{1}{(1 - \theta) + \theta P(\omega_0 = i_0)} \right.$$

$$\left[(1 - \theta) \sum_{\substack{i_j=1 \\ i_j \neq i_0}}^{M} p(\omega_j = i_j | X_j) + p(\omega_j = i_0 | X_j) \right] \right\}$$

$$= \sum_{i_0=1}^{M} p(\omega = i_0 | X_0) \left\{ \prod_{j=1}^{M} \left[\frac{(1 - \theta) + \theta p(\omega = i_0 | X_j)}{(1 - \theta) + \theta P(\omega = i_0)} \right] \right\} \tag{19}$$

3.3 SEQUENTIAL NEIGHBORHOOD — GENERAL CASE

A general N-pixel sequential neighborhood is shown in figure 3-2.

Figure 3-2.- A general N-pixel sequential neighborhood.

The transitions for which the transition probabilities are applied in the sequential neighborhood are indicated in figure 3-2. An expression is developed here for the likelihood function L of the patterns in a general sequential neighborhood of N pixels. Only the pixels immediately adjacent to each pixel are treated as its neighbors. Consider

$$P(\omega_1 = i_1, \cdots, \omega_N = i_N)$$

$$= P(\omega_2 = i_2)P(\omega_1 = i_1, \omega_3 = i_3, \cdots, \omega_N = i_N | \omega_2 = i_2)$$

$$= P(\omega_2 = i_2)P(\omega_1 = i_1 | \omega_2 = i_2, \omega_3 = i_3, \cdots, \omega_N = i_N)P(\omega_3 = i_3,$$

$$\cdots, \omega_N = i_N | \omega_2 = i_2)$$

$$= P(\omega_2 = i_2)P(\omega_1 = i_1 | \omega_2 = i_2)P(\omega_3 = i_3, \cdots, \omega_N = i_N | \omega_2 = i_2) \qquad (20)$$

Assumption (b) is used in obtaining equation (20). The Bayes rule is used to obtain the following.

$$P(\omega_3 = i_3, \omega_4 = i_4, \cdots, \omega_N = i_N | \omega_2 = i_2) = \frac{P(\omega_2 = i_2, \cdots, \omega_N = i_N)}{P(\omega_2 = i_2)} \qquad (21)$$

Proceeding in a manner similar to equation (20), the numerator of equation (21) can be written as

$$P(\omega_2 = i_2, \omega_3 = i_3, \cdots, \omega_N = i_N)$$

$$= P(\omega_3 = i_3)P(\omega_2 = i_2 | \omega_3 = i_3)P(\omega_4 = i_4, \cdots, \omega_N = i_N | \omega_3 = i_3) \qquad (22)$$

Continuing in a similar manner obtains the following result.

$$P(\omega_{N-1} = i_{N-1}, \omega_N = i_N) = P(\omega_{N-1} = i_{N-1})P(\omega_N = i_N | \omega_{N-1} = i_{N-1}) \qquad (23)$$

The following is obtained from the Bayes rule.

$$P(\omega_2 = i_2 | \omega_3 = i_3) = \frac{P(\omega_2 = i_2)}{P(\omega_3 = i_3)} P(\omega_3 = i_3 | \omega_2 = i_2) \qquad (24)$$

Equation (25) is obtained from equations (20) through (24).

3-6

13

$$P(\omega_1 = i_1, \omega_2 = i_2, \cdots, \omega_N = i_N)$$

$$= P(\omega_1 = i_1)P(\omega_2 = i_2|\omega_1 = i_1)P(\omega_3 = i_3|\omega_2 = i_2)\cdots P(\omega_N = i_N|\omega_{N-1} = i_{N-1})$$

$$(25)$$

Substitution of equation (25) into equation (15) results in an expression for the criterion L for a general sequential neighborhood. That is,

$$L = \left\{ \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_N=1}^{M} \left[\prod_{j=1}^{N} \frac{p(\omega_j = i_j|X_j)}{P(\omega_j = i_j)} \right] [P(\omega_1 = i_1)P(\omega_2 = i_2|\omega_1 = i_1) \right.$$

$$\left. P(\omega_3 = i_3|\omega_2 = i_2)\cdots P(\omega_N = i_N|\omega_{N-1} = i_{N-1})] \right\} \qquad (26)$$

3.3.1 THE LIKELIHOOD FUNCTION L OF PATTERNS IN A SEQUENTIAL NEIGHBORHOOD WITH THE LINEAR TRANSITION PROBABILITIES MODEL

In this section, equation (26) is expressed in a polynomial form in terms of θ for a four-pixel sequential neighborhood, using the linear transition probabilities model of equation (1). The four-pixel sequential neighborhood considered is shown in figure 3-3.

Figure 3-3.- A four-pixel sequential neighborhood.

The likelihood function of equation (26) for the neighborhood of figure 3-3 becomes

$$L = \sum_{i_2=1}^{M} p(\omega_2 = i_2|X_2) \sum_{i_3=1}^{M} \frac{p(\omega_3 = i_3|X_3)}{P(\omega_3 = i_3)} P(\omega_3 = i_3|\omega_2 = i_2)$$

$$\sum_{i_4=1}^{M} \frac{p(\omega_4 = i_4|X_4)}{P(\omega_4 = i_4)} P(\omega_4 = i_4|\omega_3 = i_3) \sum_{i_5=1}^{M} \frac{p(\omega_5 = i_5|X_5)}{P(\omega_5 = i_5)} P(\omega_5 = i_5|\omega_4 = i_4)$$

$$= \sum_{i_2=1}^{M} p(\omega_2 = i_2 | X_2) \left\{ (1 - \theta)^3 + \theta(1 - \theta)^2 \left[a_{45} + a_{34} + \frac{p(\omega_3 = i_2 | X_3)}{P(\omega_3 = i_2)} \right] \right.$$

$$+ \theta^2(1 - \theta) \left[a_{345} + a_{45} \frac{p(\omega_3 = i_2 | X_3)}{P(\omega_3 = i_2)} + \frac{p(\omega_3 = i_2 | X_3)p(\omega_4 = i_2 | X_4)}{P(\omega_3 = i_2)P(\omega_4 = i_2)} \right]$$

$$\left. + \theta^3 \frac{p(\omega_3 = i_2 | X_3)p(\omega_4 = i_2 | X_4)p(\omega_5 = i_2 | X_5)}{P(\omega_3 = i_2)P(\omega_4 = i_2)P(\omega_5 = i_2)} \right\}$$

$$= (1 - \theta)^3 + \theta(1 - \theta)^2(a_{23} + a_{34} + a_{45}) + \theta^2(1 - \theta)(a_{23}a_{45}$$

$$+ a_{234} + a_{345}) + \theta^3 a_{2345} \tag{27}$$

Since the a priori probabilities are position independent in the local neighborhood, the different quantities in equation (27) can be shown to be

$$a_{23} = \sum_{i_2=1}^{M} \frac{p(\omega_2 = i_2 | X_2)p(\omega_3 = i_2 | X_3)}{P(\omega = i_2)}$$

$$a_{34} = \sum_{i_3=1}^{M} \frac{p(\omega_3 = i_3 | X_3)p(\omega_4 = i_3 | X_4)}{P(\omega = i_3)}$$

$$a_{45} = \sum_{i_4=1}^{M} \frac{p(\omega_4 = i_4 | X_4)p(\omega_5 = i_4 | X_5)}{P(\omega = i_4)}$$

$$a_{234} = \sum_{i_2=1}^{M} \frac{p(\omega_2 = i_2 | X_2)p(\omega_3 = i_2 | X_3)p(\omega_4 = i_2 | X_4)}{P^2(\omega = i_2)}$$

$$a_{345} = \sum_{i_3=1}^{M} \frac{p(\omega_3 = i_3 | X_3)p(\omega_4 = i_3 | X_4)p(\omega_5 = i_3 | X_5)}{P^2(\omega = i_3)}$$

$$a_{2345} = \sum_{i_2=1}^{M} \frac{p(\omega_2 = i_2 | X_2)p(\omega_3 = i_2 | X_3)p(\omega_4 = i_2 | X_4)p(\omega_5 = i_2 | X_5)}{P^3(\omega = i_2)}$$

$$\tag{28}$$

Using a linear model of equation (1) for the transition probabilities and the definitions in equation (28) in equation (26), expressions for the likelihood function for different sizes of sequential neighborhoods can be easily written and are listed in table 3-1.

3.3.2 EXPRESSIONS FOR THE LIKELIHOOD FUNCTION L OF A SEQUENTIAL NEIGHBORHOOD WITH NONLINEAR TRANSITION PROBABILITIES MODEL

Using the nonlinear transition probabilities model of equation (11) in equation (26), expressions for the likelihood function for several sequential neighborhoods can be easily derived. These are illustrated for three-, four-, and five-pixel sequential neighborhoods in the following expressions. In order for the transition probabilities model to hold true, the transitions in the neighborhood must be as indicated in figures 3-4, 3-5, and 3-6. Define

$$
a_5(i_4,\theta) = \frac{(1 - \theta) + \theta p(\omega = i_4 | X_5)}{(1 - \theta) + \theta P(\omega = i_4)}
$$

$$
a_{45}(i_3,\theta) = \frac{(1 - \theta)\left[\sum_{i_4=1}^{M} p(\omega = i_4 | X_4) a_5(i_4,\theta)\right] + \theta p(\omega = i_3 | X_4) a_5(i_3,\theta)}{(1 - \theta) + \theta P(\omega = i_3)}
$$

$$
a_{345}(i_2,\theta) = \frac{(1 - \theta)\left[\sum_{i_3=1}^{M} p(\omega = i_3 | X_3) a_{45}(i_3,\theta)\right] + \theta p(\omega = i_2 | X_3) a_{45}(i_2,\theta)}{(1 - \theta) + \theta P(\omega = i_2)}
$$

$$
a_{2345}(i_1,\theta) = \frac{(1 - \theta)\left[\sum_{i_2=1}^{M} p(\omega = i_2 | X_2) a_{345}(i_2,\theta)\right] + \theta p(\omega = i_1 | X_2) a_{345}(i_1,\theta)}{(1 - \theta) + \theta P(\omega = i_1)}
$$

$$(29)$$

The likelihood function for the three-pixel sequential neighborhood of figure 3-4 is given by

$$
L_3(\theta) = \sum_{i_3=1}^{M} p(\omega = i_3 | X_3) a_{45}(i_3,\theta) \tag{30}
$$

TABLE 3-1.- EXPRESSIONS FOR THE LIKELIHOOD FUNCTION FOR DIFFERENT SIZES OF SEQUENTIAL NEIGHBORHOODS
WITH THE LINEAR TRANSITION PROBABILITIES MODEL

Example	Neighborhood size, pixels	Illustration	Expression for the likelihood function
1	2	[1][2]	$L(\theta) = (1 - \theta) + \theta a_{12}$
2	3	[1][2][3]	$L(\theta) = (1 - \theta)^2 + \theta(1 - \theta)(a_{12} + a_{23}) + \theta^2 a_{123}$
3	4	[1][2][3][4]	$L(\theta) = (1 - \theta)^3 + \theta(1 - \theta)^2(a_{12} + a_{23} + a_{34})$ $\quad + \theta^2(1 - \theta)(a_{12}a_{34} + a_{123} + a_{234}) + \theta^3 a_{1234}$
4	5	[1][2][3][4][5]	$L(\theta) = (1 - \theta)^4 + \theta(1 - \theta)^3(a_{12} + a_{23} + a_{34} + a_{45})$ $\quad + \theta^2(1 - \theta)^2(a_{123} + a_{234} + a_{345} + a_{12}a_{34} + a_{12}a_{45} + a_{23}a_{45})$ $\quad + \theta^3(1 - \theta)(a_{1234} + a_{2345} + a_{12}a_{345} + a_{45}a_{123}) + \theta^4 a_{12345}$
5	6	[1][2][3][4][5][6]	$L(\theta) = (1 - \theta)^5 + \theta(1 - \theta)^4(a_{12} + a_{23} + a_{34} + a_{45} + a_{56})$ $\quad + \theta^2(1 - \theta)^3[a_{123} + a_{234} + a_{345} + a_{456} + a_{12}(a_{34} + a_{45} + a_{56})$ $\quad + a_{23}(a_{45} + a_{56}) + a_{34}a_{56}] + \theta^3(1 - \theta)^2[a_{1234} + a_{2345} + a_{3456}$ $\quad + a_{123}(a_{45} + a_{56}) + a_{234}a_{56} + a_{12}(a_{345} + a_{456}) + a_{23}a_{456}$ $\quad + a_{12}a_{34}a_{56}] + \theta^4(1 - \theta)(a_{12345} + a_{23456} + a_{12}a_{3456}$ $\quad + a_{56}a_{1234} + a_{123}a_{456}) + \theta^5 a_{123456}$

The likelihood function for the four-pixel sequential neighborhood of figure 3-5 is given by

$$L_4(\theta) = \sum_{i_2=1}^{M} p(\omega = i_2|X_2)a_{345}(i_2,\theta) \tag{31}$$

The likelihood function for the five-pixel sequential neighborhood of figure 3-6 is given by

$$L_5(\theta) = \sum_{i_1=1}^{M} p(\omega = i_1|X_1)a_{2345}(i_1,\theta) \tag{32}$$

Figure 3-4.- A three-pixel sequential neighborhood.

Figure 3-5.- A four-pixel sequential neighborhood.
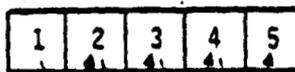
Figure 3-6.- A five-pixel sequential neighborhood.

3.4 COMPUTATION OF θ BY THE MAXIMIZATION OF LIKELIHOOD FUNCTION

With both linear and nonlinear transition probabilities models, the likelihood function is a continuous function of the parameter θ. The parameter θ that maximizes the likelihood function with the nonlinear transition probabilities model can be obtained using a one-dimensional bounded search, since the parameter θ is bounded and the likelihood function is nonlinear. With the linear transition probabilities model, the likelihood function is a polynomial in the parameter θ. The flow diagram (refs. 13-16) of figure 3-7 can be used to find the optimal θ (θ_{opt}) for the linear transition probabilities model in the range $0 < \theta < 1$, which gives the global maximum for the likelihood function.
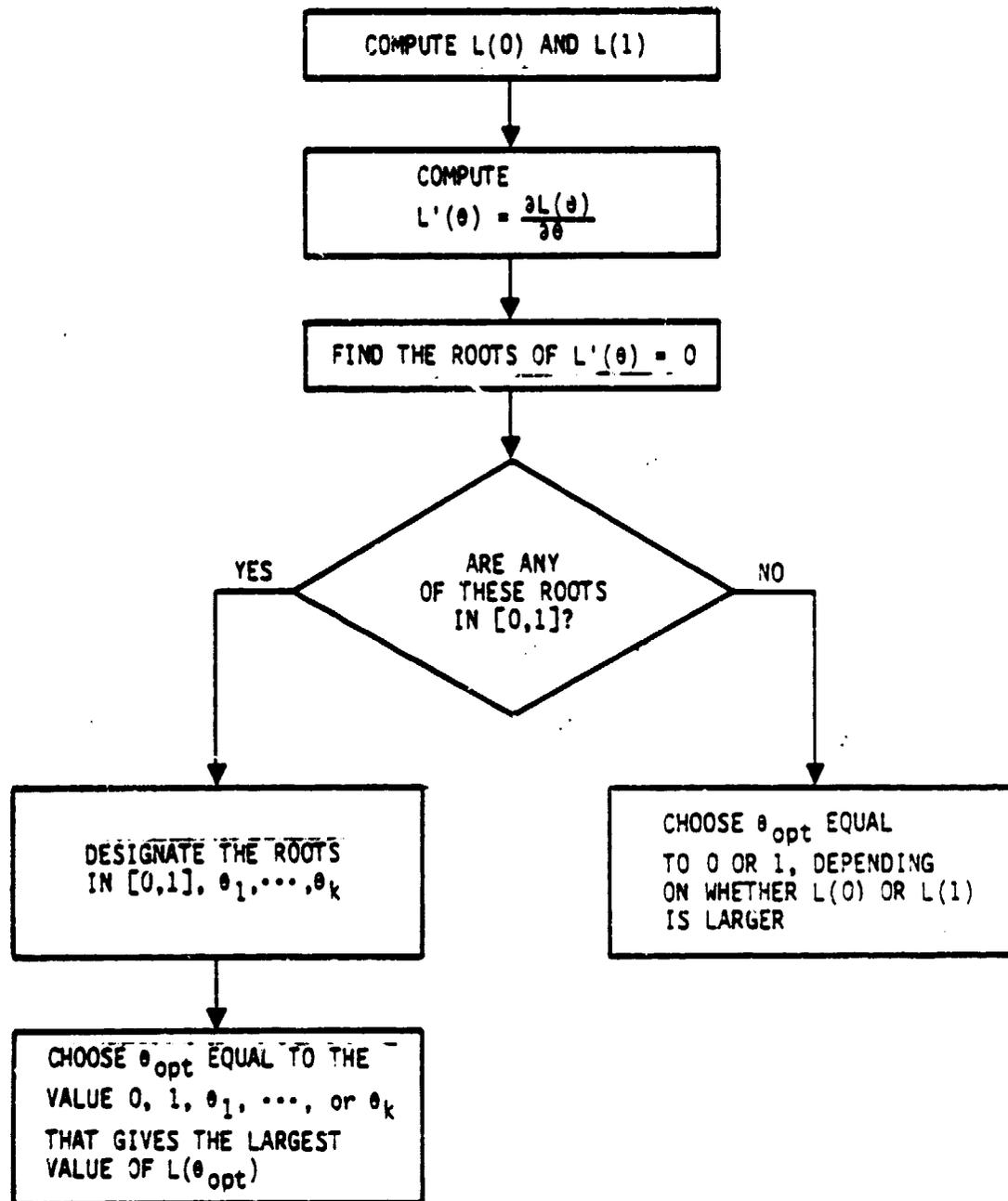
Figure 3-7.- Procedure for finding θ_{opt} in the range $0 < \theta < 1$, which gives global maximum for $L(\theta)$.

Optimal transition probabilities that maximize the likelihood function for some typical sequential neighborhoods, with both linear and nonlinear transition probabilities models, are given in appendix C.

4. UPDATING A POSTERIORI PROBABILITIES

Using the transition probabilities models of section 2, methods are developed in this section for incorporating contextual information into the classifier decision process.

4.1 UPDATING THE A POSTERIORI PROBABILITIES OF A PIXEL USING INFORMATION FROM A SINGLE NEIGHBOR

Expressions are developed for updating the a posteriori probabilities of the labels of a pixel using information from its single neighbor. These are used to exploit contextual information from large local neighborhoods. Let the pixel under consideration be X_n and its neighbors be X_{n-1} and X_{n+1}. Figure 4-1 shows the positions of these pixels.
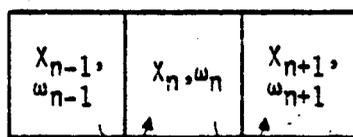
Figure 4-1.- Illustration of pixel n under consideration and its neighbors.

The assumptions used for updating the a posteriori probabilities are the same as those made in section 3. Namely: (a) The probability density function of a pattern, given its label, is independent of other patterns and their labels; (b) The labels of the pixels are independent of the labels of their nonneighbors. These assumptions are used in the rest of the section. The information contained in the pattern X_{n-1} regarding the label of the pattern X_n can be written in terms of transition probabilities as

$$p(\omega_n = k | X_{n-1}) = \sum_{i=1}^{M} p(\omega_n = k, \omega_{n-1} = i | X_{n-1})$$

$$= \sum_{i=1}^{M} P(\omega_n = k | \omega_{n-1} = i) p(\omega_{n-1} = i | X_{n-1}) \qquad (33)$$

Similarly, the following is obtained.

$$p(X_n|X_{n-1}) = \sum_{i=1}^{M} p(X_n, \omega_n = i|X_{n-1})$$

$$= \sum_{i=1}^{M} p(X_n|\omega_n = i)p(\omega_n = i|X_{n-1}) \tag{34}$$

Now, the a posteriori probabilities of the labels of the pattern X_n are updated using the information from the patterns X_n and X_{n-1} and their spatial relationship as follows, using the assumptions (a) and (b) above.

$$p(\omega_n = k|X_{n-1}, X_n) = \frac{p(X_n|\omega_n = k)p(\omega_n = k|X_{n-1})}{p(X_n|X_{n-1})}$$

$$= \frac{\dfrac{p(\omega_n = k|X_n)}{P(\omega_n = k)} \sum_{i=1}^{M} p(\omega_n = k|\omega_{n-1} = i)p(\omega_{n-1} = i|X_{n-1})}{\sum_{k=1}^{M} \dfrac{p(\omega_n = k|X_n)}{P(\omega_n = k)} \sum_{i=1}^{M} P(\omega_n = k|\omega_{n-1} = i)p(\omega_{n-1} = i|X_{n-1})}$$

$$\tag{35}$$

Using the linear transition probabilities model of equation (1) in equation (35) yields

$$p(\omega_n = k|X_{n-1}, X_n) = \frac{(1 - \theta)p(\omega_n = k|X_n) + \theta \dfrac{p(\omega_n = k|X_n)p(\omega_{n-1} = k|X_{n-1})}{P(\omega_n = k)}}{(1 - \theta) + \theta \sum_{k=1}^{M} \dfrac{p(\omega_{n-1} = k|X_{n-1})p(\omega_n = k|X_n)}{P(\omega_n = k)}}$$

$$\tag{36}$$

The information in the pattern X_n, in obtaining the label of pattern X_{n+1}, can be written as follows.

$$p(\omega_{n+1} = j|X_n) = \sum_{i=1}^{M} p(\omega_{n+1} = j, \omega_n = i|X_n)$$

$$= \sum_{i=1}^{M} P(\omega_{n+1} = j|\omega_n = i)p(\omega_n = i|X_n) \tag{37}$$

Similarly, the following is obtained.

$$p(X_{n+1}|X_n) = \sum_{j=1}^{M} p(X_{n+1}, \omega_{n+1} = j|X_n)$$

$$= \sum_{j=1}^{M} p(X_{n+1}|\omega_{n+1} = j)p(\omega_{n+1} = j|X_n) \tag{38}$$

Using the patterns X_n and X_{n+1} one has

$$p(\omega_n = k|X_n, X_{n+1}) = \sum_{j=1}^{M} p(\omega_n = k, \omega_{n+1} = j|X_n, X_{n+1})$$

$$= \frac{\sum_{j=1}^{M} p(X_{n+1}|\omega_{n+1} = j)P(\omega_{n+1} = j|\omega_n = k)p(\omega_n = k|X_n)}{p(X_{n+1}|X_n)}$$

$$= \frac{p(\omega_n = k|X_n) \sum_{j=1}^{M} P(\omega_{n+1} = j|\omega_n = k) \frac{p(\omega_{n+1} = j|X_{n+1})}{P(\omega_{n+1} = j)}}{\sum_{i=1}^{M} p(\omega_n = i|X_n) \sum_{j=1}^{M} P(\omega_{n+1} = j|\omega_n = i) \frac{p(\omega_{n+1} = j)|X_{n+1}}{P(\omega_{n+1} = j)}}$$

$$\tag{39}$$

Using the linear transition probabilities model of equation (1) in equation (39) yields the following.

$$p(\omega_n = k|X_n, X_{n+1}) = \frac{p(\omega_n = k|X_n)\left[(1 - \theta) + \theta \frac{p(\omega_{n+1} = k|X_{n+1})}{P(\omega_{n+1} = k)}\right]}{\left[(1 - \theta) + \theta \sum_{j=1}^{M} \frac{p(\omega_{n+1} = j|X_{n+1})}{P(\omega_{n+1} = j)} p(\omega_n = j|X_n)\right]} \tag{40}$$

4.2 USE OF SINGLE-NEIGHBOR UPDATING EQUATIONS FOR LARGE LOCAL NEIGHBORHOODS

This section shows how single-neighbor updating equations can be used repeatedly to exploit spatial information in large local neighborhoods.

4.2.1 SPATIALLY UNIFORM CONTEXT — FOUR NEIGHBORS

Consider the pixel under consideration, pixel 0, and its neighbors in the local neighborhood shown in figure 3-1. In this section, expressions are developed for obtaining the a posteriori probabilities of the classes of pixel 0, using information from its local neighborhood. Consider equation (41), where $f = p(X_0, X_1, \cdots, X_4)$. Using equations (13), (16), and (17) in equation (41) yields equation (42).

From equations (39) and (42), the following is easily understood. Updating the a posteriori probabilities of the classes of pixel 0, using information from its neighbors as shown in figure 3-1, is equivalent to using the single-neighbor updating equation (39) repeatedly, taking one neighbor at a time. The sequence in which the neighbors are used is immaterial.

4.2.2 SEQUENTIAL NEIGHBORHOOD — GENERAL CASE

This section considers the problem of updating the a posteriori probabilities of the classes of the pixel under consideration, pixel j, in a general sequential neighborhood. The location of pixel j in a general sequential neighborhood of N pixels is shown in figure 4-2.

Figure 4-2.- The pixel under consideration, pixel j, and its general sequential neighborhood.

The transitions for which the estimated transition probabilities apply in the whole sequential neighborhood are indicated in figure 4-2. Consider equation (43) where $NU(i_j)$ is the numerator of the first expression in equation (43). Using equations (13) and (25) in the numerator of equation (43) results in equation (44).

If the numerator and denominator of equation (43) are divided by $\left[\prod_{\ell=1}^{N} p(X_\ell) \right]$, the numerator of equation (43) can then be written as shown in equation (45).

4-4

$$p(\omega_0 = i_0 | X_0, X_1, \cdots, X_q) = \frac{p(\omega_0 = i_0, X_0, X_1, \cdots, X_q)}{p(X_0, X_1, \cdots, X_q)}$$

$$= \frac{\sum_{i_1=1}^{M} \cdots \sum_{i_q=1}^{M} p(\omega_0 = i_0, X_0, \omega_1 = i_1, X_1, \cdots, \omega_q = i_q, X_q)}{f}$$

$$= \frac{\sum_{i_1=1}^{M} \cdots \sum_{i_q=1}^{M} p(X_0, X_1, \cdots, X_q | \omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_q = i_q) p(\omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_q = i_q)}{f} \qquad (41)$$

$$p(\omega_0 = i_0 | X_0, X_1, \cdots, X_q) =$$

$$\frac{P(\omega_0 = i_0) p(X_0 | \omega_0 = i_0) \prod_{j=1}^{q} \left[\sum_{i_j=1}^{M} p(X_j | \omega_j = i_j) P(\omega_j = i_j | \omega_0 = i_0) \right]}{f}$$

$$= \frac{p(\omega_0 = i_0 | X_0) \prod_{j=1}^{q} \left\{ \sum_{i_j=1}^{M} \left[\frac{p(\omega_j = i_j | X_j)}{p(\omega_j = i_j)} \right] P(\omega_j = i_j | \omega_0 = i_0) \right\}}{\sum_{i_0=1}^{M} p(\omega_0 = i_0 | X_0) \prod_{j=1}^{q} \left\{ \sum_{i_j=1}^{M} \left[\frac{p(\omega_j = i_j | X_j)}{p(\omega_j = i_j)} \right] P(\omega_j = i_j | \omega_0 = i_0) \right\}} \qquad (42)$$

$$P(\omega_j = i_j | X_1, \ldots, X_j, \ldots, X_N) = \frac{P(\omega_j = i_j, X_1, \ldots, X_N)}{P(X_1, \ldots, X_N)}$$

$$= \frac{\sum_{i_1=1}^{M} \cdots \sum_{i_{j-1}=1}^{M} \sum_{i_{j+1}=1}^{M} \cdots \sum_{i_N=1}^{M} P(X_1, \ldots, X_N | \omega_1 = i_1, \ldots, \omega_N = i_N) P(\omega_1 = i_1, \ldots, \omega_N = i_N)}{P(X_1, \ldots, X_N)}$$

$$\qquad \qquad (43)$$

$$= \frac{NU(i_j)}{\sum_{i_j=1}^{M} NU(i_j)}$$

$$NU(i_j) = \sum_{i_1=1}^{M} \cdots \sum_{i_{j-1}=1}^{M} \sum_{i_{j+1}=1}^{M} \cdots \sum_{i_N=1}^{M} \left[\prod_{\ell=1}^{N} P(X_\ell|\omega_\ell = i_\ell)\right]\left[P(\omega_1 = i_1)P(\omega_2 = i_2|\omega_1 = i_1)\cdots P(\omega_N = i_N|\omega_{N-1} = i_{N-1})\right]$$

$$= P(X_j|\omega_j = i_j)\left\{\sum_{i_1=1}^{M} \cdots \sum_{i_{j-1}=1}^{M} \left[\prod_{\ell=1}^{j-1} P(X_\ell|\omega_\ell = i_\ell)\right]\left[P(\omega_1 = i_1)P(\omega_2 = i_2|\omega_1 = i_1)\cdots P(\omega_j = i_j|\omega_{j-1} = i_{j-1})\right]\right\}$$

$$\left\{\sum_{i_{j+1}=1}^{M} \cdots \sum_{i_N=1}^{M} \left[\prod_{\ell=j+1}^{N} P(X_\ell|\omega_\ell = i_\ell)\right]\left[P(\omega_{j+1} = i_{j+1}|\omega_j = i_j)\cdots P(\omega_N = i_N|\omega_{N-1} = i_{N-1})\right]\right\}$$

$$= P(X_j|\omega_j = i_j)\left[\sum_{i_{j-1}=1}^{M} P(X_{j-1}|\omega_{j-1} = i_{j-1})P(\omega_j = i_j|\omega_{j-1} = i_{j-1}) \cdots \sum_{i_1=1}^{M} P(X_1|\omega_1 = i_1)P(\omega_1 = i_1)P(\omega_2 = i_2|\omega_1 = i_1)\right]$$

$$\left[\sum_{i_{j+1}=1}^{M} P(X_{j+1}|\omega_{j+1} = i_{j+1})P(\omega_{j+1} = i_{j+1}|\omega_j = i_j) \cdots \sum_{i_N=1}^{M} P(X_N|\omega_N = i_N)P(\omega_N = i_N|\omega_{N-1} = i_{N-1})\right]$$

$$\qquad \qquad (44)$$

$$NU(i_j) = \frac{p(\omega_j = i_j | X_j)}{P(\omega_j = i_j)} \left[\sum_{i_{j-1}=1}^{M} P(\omega_j = i_j | \omega_{j-1} = i_{j-1}) \frac{p(\omega_{j-1} = i_{j-1} | X_{j-1})}{p(\omega_{j-1} = i_{j-1})} \right.$$

$$\left. \cdots \sum_{i_1=1}^{M} P(\omega_2 = i_2 | \omega_1 = i_1) p(\omega_1 = i_1 | X_1) \right]$$

$$\left[\sum_{i_{j+1}=1}^{M} P(\omega_{j+1} = i_{j+1} | \omega_j = i_j) \frac{p(\omega_{j+1} = i_{j+1} | X_{j+1})}{P(\omega_{j+1} = i_{j+1})} \right.$$

$$\left. \cdots \sum_{i_N=1}^{M} P(\omega_N = i_N | \omega_{N-1} = i_{N-1}) \frac{p(\omega_N = i_N | X_N)}{P(\omega_N = i_N)} \right] \tag{45}$$

The term in the first set of brackets of equation (45) is the contribution from pixels to the left of pixel j (see fig. 4-2), the term in the second set of brackets is the contribution from pixels to the right of pixel j, and the first term is the contribution from pixel j to the a posteriori probabilities of the classes of pixel j. These contributions appear in multiplicative form in equation (45).

An examination of equations (35), (39), and (45) reveals that the single-neighbor updating equations (35) and (39) can be used repeatedly to update the a posteriori probabilities of the classes of pixel j, using information from its sequential neighborhood as follows. Equation (39) is used to update the a posteriori probabilities of pixel (N - 1), using the a posteriori probabilities of pixels (N - 1) and N. The updated a posteriori probabilities of pixel (N - 1) and the a posteriori probabilities of pixel (N - 2) are used to update those of pixel (N - 2). Proceeding in a similar manner, the updated a posteriori probabilities of pixel (j + 1) and the a posteriori probabilities of pixel j are used to update those of pixel j. Similarly, equation (35) is used to update the a posteriori probabilities of pixel j, using information from pixels to the left of pixel j. The a posteriori probabilities of pixels 1 and 2 are used to update those of pixel 2. The updated a posteriori probabilities of pixel 2 and the a posteriori probabilities of pixel 3 are used to update the a posteriori probabilities of pixel 3. The process is

repeated until the updated a posteriori probabilities of pixel (j - 1) and the previously updated ones of pixel j are used to update those of pixel j.

4.2.3 APPLICATION OF SEQUENTIAL CONTEXT TO TWO-DIMENSIONAL NEIGHBORHOODS

The expressions for the likelihood function and updating equations become complex with the increase in tne size of the local neighborhood. Hence, it is proposed to use sequentially the sequential context for two-dimensional local neighborhoods. It is desirable that the updating be independent of the sequence of the sequential neighborhoods in which the updating is done. From equation (45) it is seen that, with the use of sequential neighborhoods (centering on the pixel under consideration), the updating is independent of the sequence of the sequential neighborhoods in which the updating is done. The sequential neighborhoods to be used in updating, then, are the ones centering on the pixel under consideration in four directions: 0°, 45°, 90°, and 135°. A few typical two-dimensional local neighborhoods composed of these sequential neighborhoods are illustrated in figure 4-3.
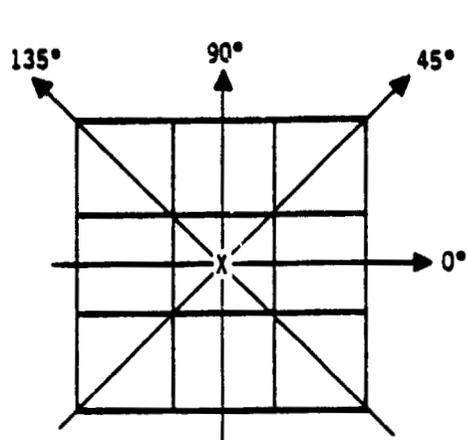
(a) 3-by-3 neighborhood.
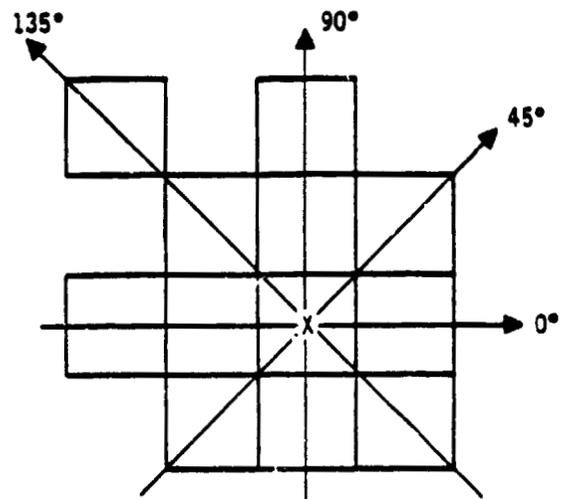
(b) 4-by-4 neighborhood.
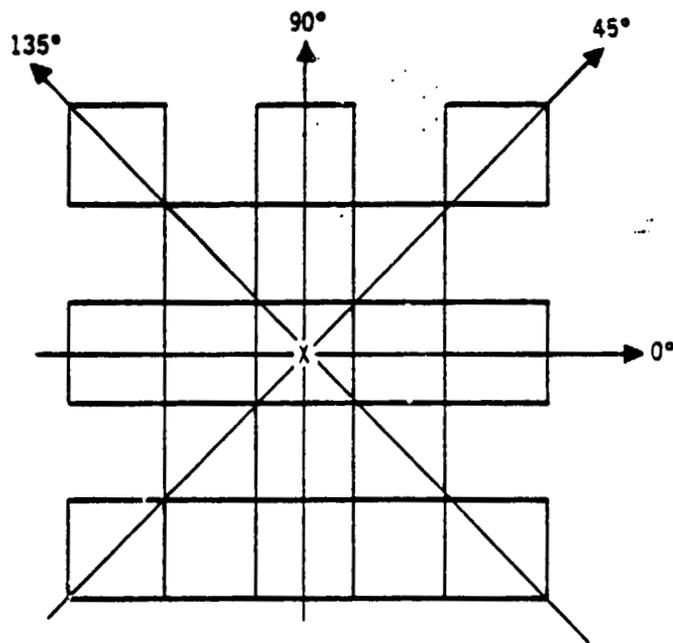
(c) 5-by-5 neighborhood.

Figure 4-3.- Some typical neighborhoods and updating directions. The pixel
under consideration is marked by X.

5. EXPERIMENTAL RESULTS

In this section, some results are obtained by applying the theory developed in
the previous sections to the classification of the remotely sensed Landsat MSS
data. Several segments[1] were processed in the following manner. The image
was overlaid with a rectangular grid of 209 grid intersections, and the labels
of the pixels or dots corresponding to each grid intersection were acquired.
Two classes are in the image: Class 1 is wheat, and class 2 is nonwheat
designated "other." A linear classifier is trained on one-half of the labeled
data. The remaining one-half of the labeled data is used as a test set. The
a posteriori probabilities of the classes of the pixels are estimated by
normalizing the discriminant function values of the classes.

5.1 COMPUTATIONAL RESULTS FOR A TYPICAL 5-BY-5 NEIGHBORHOOD

The a posteriori probabilities of the classes of the pixels in a typical
5-by-5 neighborhood from an MSS image of segment 1739 are listed in
table 5-1. This segment is in Teton County, Montana.

TABLE 5-1.- THE A POSTERIORI PROBABILITIES OF THE CLASSES
IN A 5-BY-5 NEIGHBORHOOD

[The first entry is $p(\omega = 1|X)$ and the second entry is $p(\omega = 2|X)$]

(0.716,0.284)	(0.322,0.678)	(0.820,0.180)	(0.569,0.331)	(0.326,0.674)
(0.829,0.171)	(0.899,0.101)	(0.897,0.103)	(0.762,0.238)	(0.886,0.114)
(0.625,0.375)	(0.158,0.842)	(0.285,0.715)	(0.757,0.243)	(0.117,0.883)
(0.087,0.913)	(0.062,0.938)	(0.060,0.940)	(0.080,0.920)	(0.090,0.910)
(0.125,0.875)	(0.089,0.911)	(0.132,0.868)	(0.157,0.843)	(0.127,0.873)

---

[1]A segment is a 9- by 11-kilometer (5- by 5-nautical-mile) area for which the
MSS image is divided into a 117-row by 196-column rectangular array of
pixels.

The pixel under consideration is the central pixel of the neighborhood. The a priori probabilities are estimated as an average of the a posteriori probabilities in the reighborhood. Consider the following.

$$p(\omega = 1) = \int p(\omega = 1,X)dX$$

$$= \int p(\omega = 1|X)p(X)dX$$

$$= \underset{p(X)}{E} [p(\omega = 1|X)]$$

$$= \frac{1}{N} \sum_{j=1}^{N} [p(\omega = 1|X_j)] \qquad (46)$$

where X_j ($j = 1,2,\cdots,N$) are the pixels in tne local neighborhood. The a posteriori probabilities of the classes of the pixel under consideration are updated using sequential context and the procedure described in section 4.2.3. This procedure is repeated for five iterations, and the computational results are listed in table 5-2.

TABLE 5-2.- COMPUTATIONAL RESULTS OF UPDATING THE A POSTERIORI PROBABILITIES OF THE CENTRAL PIXEL IN A 5-BY-5 NEIGHBORHOOD (USING THE LINEAR TRANSITION PROBABILITIES MODEL)

Iter- ation	A posteriori probabilities before updating	A priori probabilities in the neighborhood	Estimates of parameter θ for different sequential neighborhoods				A posteriori probabilities after updating
			0°	45°	90°	135°	
1	(0.285,0.715)	(0.4087,0.5913)	0.0	0.2904	0.4	0.4	(0.3574,0.6426)
2	(0.3574,0.6426)	(0.4130,0.5870)	0.0	0.2656	0.4	0.4	(0.4315,0.5685)
3	(0.4315,0.5685)	(0.4173,0.5827)	0.0	0.2416	0.4	0.4	(0.5034,0.4966)
4	(0.5304,0.4966)	(0.4216,0.5784)	0.0	0.2194	0.4	0.4	(0.5699,0.4301)
5	(0.5699,0.4301)	(0.4255,0.5745)	0.0	0.1995	0.5	0.4	(0.6363,0.3637)

The true class of the central pixel is wheat; and, without using the contextual information, the central pixel will be misclassified into class "other." Table 5-2 shows that, using contextual information from the local neighborhood, after the third iteration the central pixel is correctly classified.

5.2 CONTEXTUAL CLASSIFICATION RESULTS

Comparative results with and without using contextual information in classification are presented in this section. Classification maps for segment 1739 are shown in figures 5-1 through 5-3. It is observed from the independent test set that the classification accuracy for this segment increased by 5 percent with the use of contextual information from the 3-by-3 neighborhood and by 7 percent with the use of contextual information from the 5-by-5 neighborhood (over the accuracies obtained without using contextual information). While generally preserving the boundaries, contextual classification corrected the misclassifications of many pixels and did this more accurately with data from the 5-by-5 neighborhood than with data from the 3-by-3 neighborhood.

Accuracies in the classification of MSS images of a few segments with and without the use of contextual information are listed in table 5-3.

In general, an examination of the classification maps of full images and classification accuracies on the independent test set shows considerable improvement in the classifications with the use of contextual information. The improvement is greater with the increase in size of the neighborhood. The contextual classification of a full segment with a 5-by-5 neighborhood using the methods developed here took approximately 12 minutes of total time on the Purdue University Laboratory for Applications of Remote Sensing (LARS) IBM 3031 computer system.

TABLE 5-3.- CLASSIFICATION ACCURACIES (PERCENTAGES) WITH AND WITHOUT CONTEXTUAL INFORMATION

Segment	Location (county, state)	Without context	With sequential context (a)			With spatially uniform context
			NS = 5	NS = 4	NS = 3	
[b]1005	Cheyenne, Colorado	85.88	88.46	88.46	90.38	86.54
[b]1060	Sherman, Texas	80.77	85.58	82.69	81.73	81.73
[b]1231	Jackson, Oklahoma	89.42	91.35	91.35	90.38	91.35
[c]1520	Big Stone, Minnesota	84.62	87.50	85.58	86.54	84.62
[c]1604	Renville, North Dakota	60.58	63.46	60.58	59.62	60.58
[c]1675	McPherson, South Dakota	68.27	71.15	73.08	68.27	67.31
[c]1739	Teton, Montana	68.27	75.00	72.22	73.08	70.19

[a]NS = Neighborhood size.
[b]Segments in which class 1 is winter wheat.
[c]Segments in which class 1 is spring wheat.
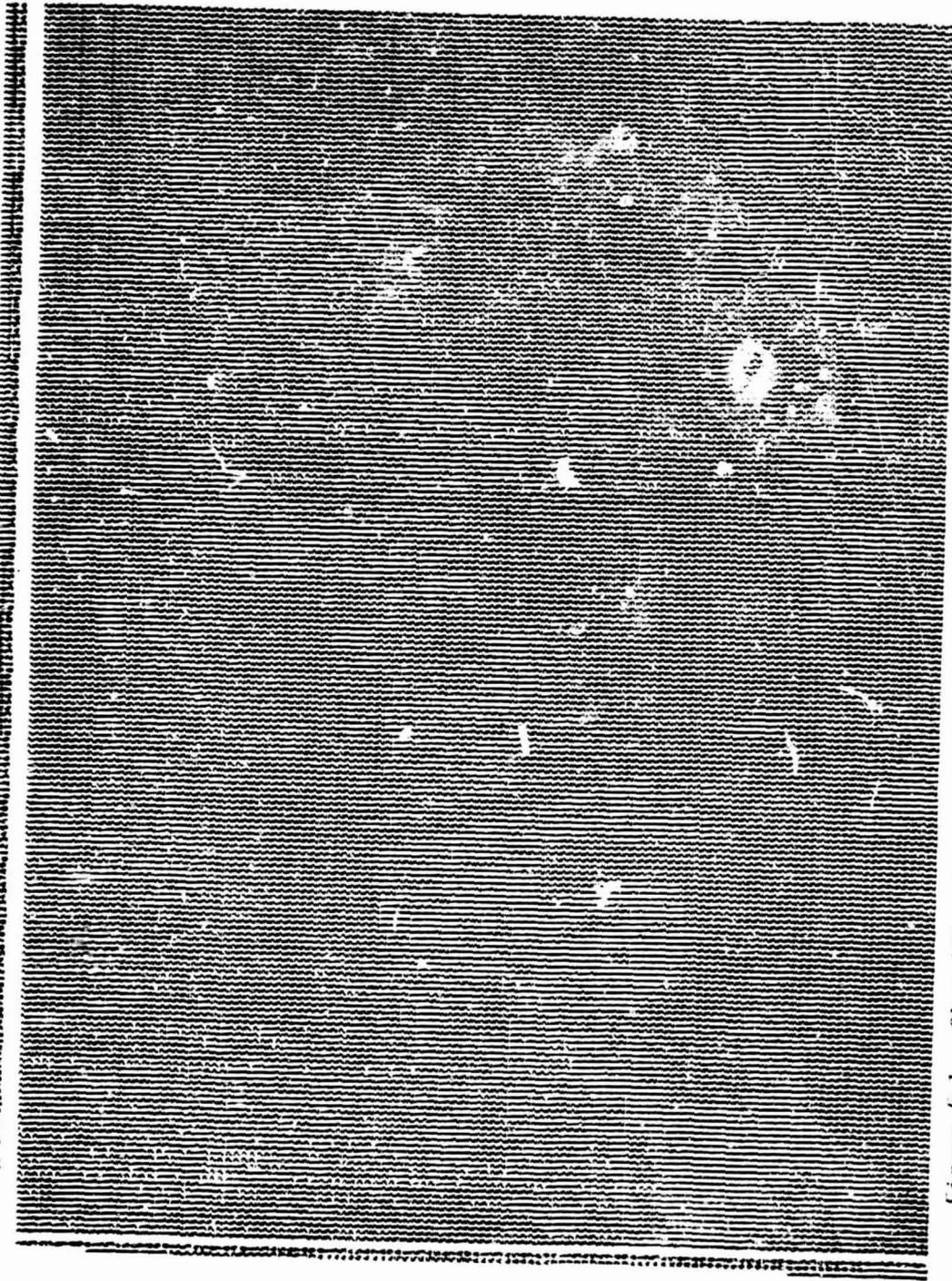
Figure 5-1.— Classification map of segment 1739 without contextual information.

$\frac{5-5}{34}$
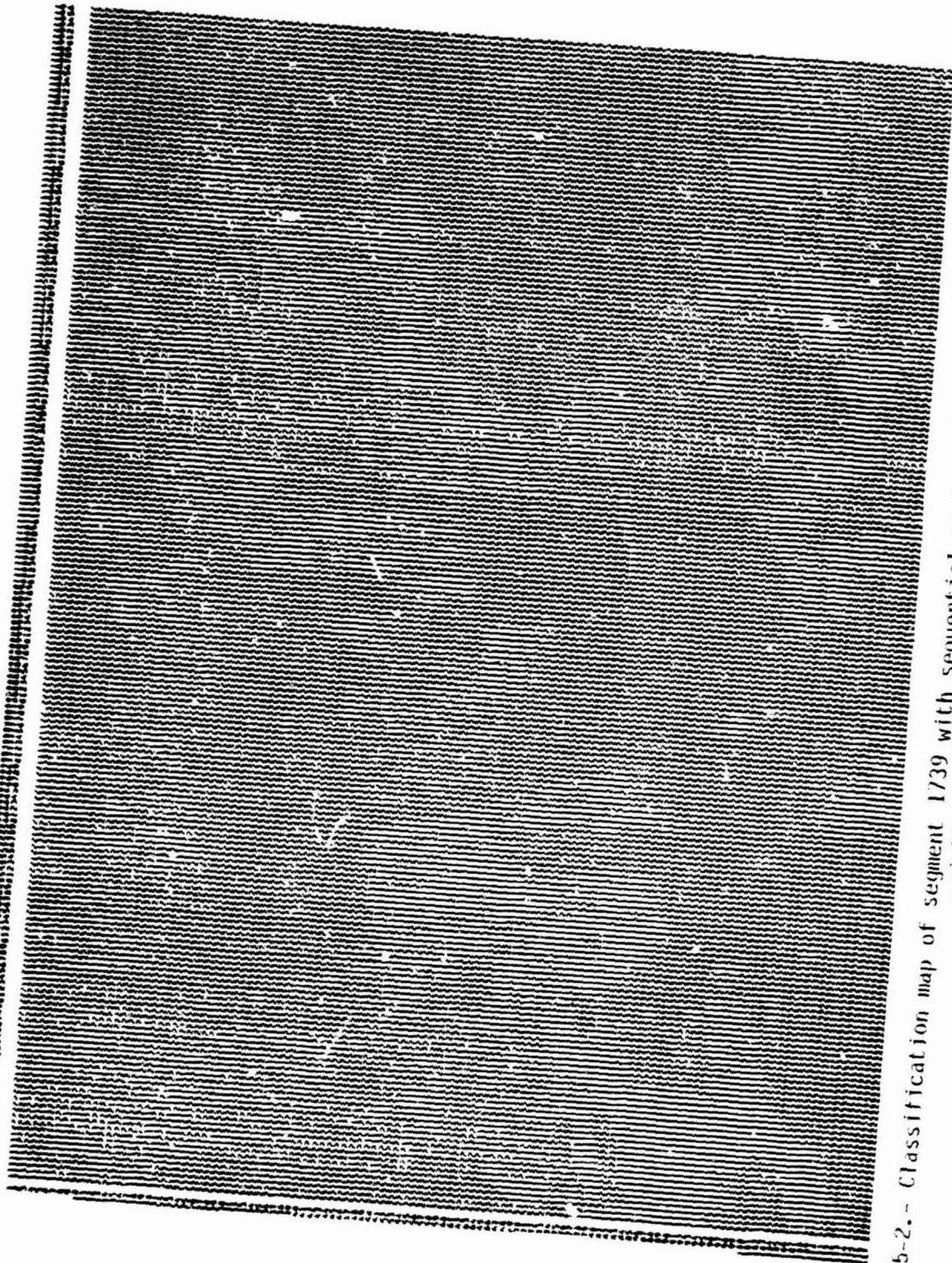
Figure 5-2.- Classification map of segment 1/39 with sequential contextual information for the 3-by-3 neighborhood (one iteration).

Figure 5-3.- Classification of segment 1/39 with sequential contextual information for the 5-by-5 neighborhood (five iterations).

The variance reduction factors obtained without using contextual information and with the use of contextual information from a local neighborhood of size 5 are listed in table 5-4.

TABLE 5-4.- VARIANCE REDUCTION FACTORS WITH AND
WITHOUT CONTEXTUAL INFORMATION

Segment	Location (county, state)	Variance reduction factor	
		Without context	With sequential context, NS = 5
1005	Cheyenne, Colorado	0.5720	0.5430
1060	Sherman, Texas	.6227	.4717
1231	Jackson, Oklahoma	.4407	.4173
1520	Big Stone, Minnesota	.6194	.5216
1604	Renville, North Dakota	.9865	.9741
1675	McPherson, South Dakota	.9985	.9248
1739	Teton, Montana	.9271	.8267

Table 5-4 shows that there is a consistent improvement in the variance reduction factor with the use of contextual information in classification.

6. CONCLUSIONS

In this paper, the problem of incorporating contextual or spatial information
into the classification of imagery data is considered. The contextual infor-
mation is introduced into classification based on the spatial dependencies
between the states of nature of neighboring pixels or based on transition
probabilities. The dependencies between neighboring patterns are modeled with
linear and nonlinear models through a single parameter θ, which describes the
transition probabilities of the classes of the neighboring patterns. An
expression is developed for the likelihood function of the pattern vectors
from a general local neighborhood under the following reasonable assumptions:
(a) The probability density function of a pattern, given its label, is inde-
pendent of other patterns and their labels; and (b) the labels of the pattern
vectors are independent of the labels of their nonneighbors. Specific expres-
sions for the likelihood function are derived for different local neighbor-
hoods and with different transition probabilities models. The parameter θ is
estimated as the one that maximizes the likelihood function.

Expressions are presented for updating the a posteriori probabilities of the
classes of a pixel using information from a single neighbor. It is shown that
these expressions can be used to update the a posteriori probabilities of a
pixel under consideration for spatially uniform context and in a general
sequential neighborhood. The contextual information from two-dimensional
neighborhoods is introduced into the classification of imagery data, also,
through a sequence of sequential neighborhoods.

The techniques presented here are applied to the classification of remotely
sensed MSS imagery data. Computational results for a typical 5-by-5 neighbor-
hood are presented. The classification maps are presented with and without
context, and classification accuracies are given for different sizes of local
neighborhoods.

For a two-class, three-sequential-neighborhood case, expressions are developed for obtaining the transition probabilities without using models. Instead of using one parameter θ in the local neighborhood of the pattern under consideration, as shown in appendix C, transition probabilities models with different parameters in different directions can be used. The techniques, as discussed in appendix D, can be used for multitemporal or time-varying situations such as those encountered in remote sensing.

7. REFERENCES

1. Hanson, A. R.; Riseman, E. M.; and Fisher, E.: Context in Word Recognition. Pattern Recognition, vol. 8, 1976, pp. 35-45.

2. Duda, R. O.; and Hart, P. E.: Experiments in the Recognition of Hand-Printed Text: Part II — Context Analysis. 1968 Fall Joint Computer Conference. AFIFS Conf. Proc., vol. 33, Thompson (Washington, D.C.), 1968, pp. 1139-1149.

3. Abend, K.: Compound Decision Procedures for Unknown Distributions and for Dependent States of Nature. Pattern Recognition. L. Kanal, ed., Thompson (Washington, D.C.), 1968, pp. 207-249.

4. Raviv, J.: Decision Making in Markov Chains Applied to the Problem of Pattern Recognition. IEEE Trans. Information Theory, vol. IT-13, Oct. 1967, pp. 536-551.

5. Alter, R.: Utilization of Contextual Constraints in Automatic Speech Recognition. IEEE Trans. Audio and Electroacoustics, vol. AU-16, Mar. 1968, pp. 6-11.

6. Chow, C. K.: A Recognition Method Using Neighbor Dependence. IRE Trans. Electronic Computers, vol. EC-11, Oct. 1962, pp. 683-690.

7. Kettig, R. L.; and Landgrebe, D. A.: Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects. IEEE Trans. Geoscience Electronics, vol. GE-14, Jan. 1976, pp. 19-26.

8. Haralick, R. M.; Shanmugam, K.; and Dinstein, I.: Textural Features for Image Classification. IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-3, Nov. 1973, pp. 610-621.

9. Swain, P. H.: Bayesian Classification in a Time-Varying Environment. IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-8, Dec. 1978, pp. 879-883.

10. Welch, J. R.; and Salter, K. G.: A Context Algorithm for Pattern Recognition and Image Interpretation. IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-1, Jan. 1971, pp. 24-30.

11. Chittineni, C. B.: Research Plan for Developing and Evaluating Classifiers. Lockheed Electronics Co., Inc., Tech. Memo LEC-13300, JSC-14849, NASA/JSC (Houston), Apr. 1979.

12. Toussaint, G. T.: The Use of Context in Pattern Recognition. Pattern Recognition, vol. 10, 1978, pp. 189-204.

13. Burnside, W. S.; and Panton, A. W.: Theory of Equations. S. Chand and Co. (New Delhi), 1967.

14. Cooper, L.; and Steinberg, D.: Introduction to Methods of Optimization. W. B. Saunders Co. (Philadelphia), 1970, pp. 87-133.

15. Kelley, L. G.: Handbook of Numerical Methods and Its Applications. Addison-Wesley Publishing Co., Inc., 1967, pp. 356-359.

16. Bingham, J. A. C.: An Improvement to Iterative Methods of Polynominal Factorization. Communications of the ACM, vol. 10, Jan. 1967, pp. 57-60.

APPENDIX A

A GENERALIZATION OF SPATIALLY UNIFORM CONTEXT
TO LARGE NEIGHBORHOODS

42

APPENDIX A

A GENERALIZATION OF SPATIALLY UNIFORM CONTEXT
TO LARGE NEIGHBORHOODS

In this appendix, the contextual relationships developed in section 3.2 for spatially uniform context are extended for larger neighborhoods. In particular, the neighborhood shown in figure A-1 is considered. The pixels with the common sides are treated as neighbors, and the diagonal neighbors of the pixels are treated as nonneighbors.
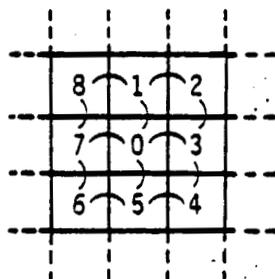
Figure A-1.- Neighboring pixels in a 3-by-3 local neighborhood.

The a posteriori probabilities of the labels of pixel 0, given the information from its local neighborhood, can be written as

$$p(\omega_0 = i_0 | X_0, X_1, \cdots, X_8) = \frac{f_1(i_0)}{f} \qquad (A-1)$$

where

$$f_1(i_0) = \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_8=1}^{M} p(X_0, X_1, \cdots, X_8 | \omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_8 = i_8)$$

$$P(\omega_0 = i_0, \cdots, \omega_8 = i_8) \qquad (A-2)$$

and
$$f = p(X_0, X_1, \cdots, X_8) \qquad (A-3)$$

A-1

43

The notation of equation (A-4) is used in the remainder of this appendix.

$$P(i_0, i_1, \cdots, i_8) = P(\omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_8 = i_8) \tag{A-4}$$

Using equations (13) and (A-4) in equations (A-2) and (A-3), $f_1(i_0)$ and f can be written as

$$f_1(i_0) = \frac{p(\omega_0 = i_0|X_0)}{P(\omega_0 = i_0)} \sum_{i_1=1}^{M} \cdots \sum_{i_8=1}^{M} \left[\prod_{j=1}^{8} \frac{p(\omega_j = i_j|X_j)}{P(\omega_j = i_j)} \right] [P(i_0, i_1, \cdots, i_8)] \tag{A-5}$$

and
$$f = \sum_{i_0=1}^{M} f_1(i_0) \tag{A-6}$$

It was assumed that the labels of the pixels are independent of the labels of their nonneighboring pixels, with the neighboring pixels defined as in figure A-1. Now consider

$$P(i_0, i_1, \cdots, i_8) = P(i_0)P(i_1, \cdots, i_8|i_0)$$

$$= P(i_0)P(i_1|i_0, i_2, \cdots, i_8)P(i_2|i_0, i_3, \cdots, i_8) \cdots$$

$$P(i_7|i_0, i_8)P(i_8|i_0) \tag{A-7}$$

The second term in the right-hand side of equation (A-7) can be written as

$$P(i_1|i_0, i_2, \cdots, i_8) = P(i_1|i_0, i_2, i_8)$$

$$= \frac{P(i_0|i_1, i_2, i_8)P(i_1, i_2|i_8)}{P(i_0|i_2, i_8)P(i_2|i_8)}$$

$$= \frac{P(i_0|i_1)P(i_2|i_1)P(i_1|i_8)}{P(i_0)P(i_2)} \tag{A-8}$$

Similar to equation (A-8), the other terms of equation (A-7) can be shown to be the following.

$$P(i_3|i_0,i_4,i_5,\cdots,i_8) = \frac{P(i_0|i_3)P(i_4|i_3)P(i_3)}{P(i_0)P(i_4)}$$

$$P(i_4|i_0,i_5,\cdots,i_8) = P(i_4|i_5)$$

$$P(i_5|i_0,i_6,\cdots,i_8) = \frac{P(i_0|i_5)P(i_6|i_5)P(i_5)}{P(i_0)P(i_6)} \qquad\qquad (A-9)$$

$$P(i_6|i_0,i_7,i_8) = P(i_6|i_7)$$

$$P(i_7|i_0,i_8) = \frac{P(i_0|i_7)P(i_8|i_7)P(i_7)}{P(i_0)P(i_8)}$$

Using equations (A-8) and (A-9) in equation (A-7) results in

$$P(i_0,i_1,\cdots,i_8) = \frac{\{P(i_0)[P(i_2|i_1)P(i_3|i_2)\cdots P(i_8|i_7)P(i_1|i_8)][P(i_1|i_0)P(i_3|i_0)P(i_5|i_0)P(i_7|i_0)]\}}{P(i_1)P(i_3)P(i_5)P(i_7)}$$

$$(A-10)$$

Expressing the transition probabilities in equation (A-10) in terms of the parameter θ [equation (1) or equation (11)], equations (A-1), (A-5), and (A-6) can be used to incorporate the contextual information from the local neighborhood. As in section 3, θ can be obtained by maximizing the likelihood function of the spectral values of the pixels 0, 1, 2, $\cdots$, 8.

Equation (A-10) also can be used to estimate the transition probabilities in the local neighborhood, if the labels of the pixels are known. For example: In remote sensing, for a selected set of images, the labels of the pixels or ground truth are known. Often it is necessary to estimate the transition probabilities. The following example illustrates, for a few typical neighborhoods, the transition probabilities obtained from the maximization of equation (A-10). The a priori probabilities in the local neighborhood are estimated as an average of the a posteriori probabilities of the classes.

Example: This example illustrates the transition probabilities obtained by computating θ, which maximizes equation (A-10). These are listed for a few typical neighborhoods in table A-1. For neighboring pixels A and B, the notation used for the transition probabilities listed in table A-1 is shown in figure A-2.

TABLE A-1.- MAXIMUM LIKELIHOOD ESTIMATES OF TRANSITION PROBABILITIES
FOR SOME TYPICAL NEIGHBORHOODS

No.	Neighborhood	A priori probabilities	Transition probabilities			
			Linear model		Nonlinear model	
1	2 1 1 / 2 1 1 / 2 1 1	P(ω = 1) = 0.6667 P(ω = 2) = 0.3333	0.8002	0.1998	0.8163	0.1837
			0.4	0.6	0.4737	0.5263
			θ = 0.4		θ = 0.55	
2	2 1 1 / 2 1 1 / 2 2 2	P(ω = 1) = 0.4444 P(ω = 2) = 0.5556	0.6388	0.3612	0.6154	0.3846
			0.2888	0.7112	0.2857	0.7143
			θ = 0.35		θ = 0.5	
3	2 1 1 / 2 2 2 / 2 2 2	P(ω = 1) = 0.2222 P(ω = 2) = 0.7778	0.4166	0.5834	0.3053	0.6947
			0.1666	0.8334	0.1566	0.8434
			θ = 0.25		θ = 0.35	

$p(\omega_A = i | \omega_B = j)$

Figure A-2.- Notation used for the transition probabilities
listed in table A-1.

A-6

APPENDIX B

OPTIMAL TRANSITION PROBABILITIES FOR A TWO-CLASS,
3-BY-3 NEIGHBORHOOD CASE

APPENDIX B

OPTIMAL TRANSITION PROBABILITIES FOR A TWO-CLASS, 3-BY-3 NEIGHBORHOOD CASE

In general, the transition probabilities that maximize the likelihood function can be obtained using optimization methods such as the Davidon-Fletcher-Powell procedure. This requires searching for $M \times M$ parameters, where M is the number of classes. Using the transition probabilities models of section 2, the likelihood function is expressed as a function of a single parameter θ. However, for a two-class, 3-by-3 neighborhood case, expressions for the transition probabilities, which maximize the likelihood function without using models for transition probabilities, are obtained in the following manner.

Let A and B be the neighboring pixels. Let there be two classes. Then we have the following theorems.

Theorem 1: For a two-class case, if the a priori probabilities are position independent in the local neighborhood [i.e., $P(\omega_A = i) = P(\omega_B = i)$], then the transition probabilities are symmetric. That is,

$$P(\omega_A = 1 | \omega_B = 2) = P(\omega_B = 1 | \omega_A = 2) \qquad (B-1)$$

Proof: Consider

$$P(\omega_A = 1 | \omega_B = 2) = 1 - P(\omega_A = 2 | \omega_B = 2)$$

$$= 1 - \frac{P(\omega_A = 2)}{P(\omega_B = 2)} P(\omega_B = 2 | \omega_A = 2)$$

$$= 1 - [1 - P(\omega_B = 1 | \omega_A = 2)]$$

$$= P(\omega_B = 1 | \omega_A = 2)$$

Theorem 2: Let $\theta_1 = P(\omega_A = 1 | \omega_B = 1)$ and $\theta_2 = P(\omega_A = 2 | \omega_B = 2)$. Then, the transition probabilities and the a priori probabilities are related as

$$(1 - \theta_2) = \frac{P_1}{P_2} (1 - \theta_1) \qquad (B-2)$$

B-1

48

where $P_i = P(\omega_A = i) = P(\omega_B = i)$.

Proof: Using the Bayes theorem, we obtain

$$P(\omega_A = 1|\omega_B = 2) = \frac{P(\omega_A = 1)}{P(\omega_B = 2)} P(\omega_B = 2|\omega_A = 1)$$

That is,
$$(1 - \theta_2) = \frac{P_1}{P_2} (1 - \theta_1)$$

Theorems 1 and 2 are used in the following to obtain θ_1 and θ_2. The likelihood function $L(\theta_1,\theta_2)$ for the three-pixel sequential neighborhood of figure 3-4 can be expressed in terms of θ_1 and θ_2 as

$$
\begin{aligned}
L(\theta_1,\theta_2) = {}& \theta_1^2 a_{111} + \theta_1(1 - \theta_1)a_{112} + (1 - \theta_1)(1 - \theta_2)a_{121} \\
& + (1 - \theta_1)\theta_2 a_{122} + \theta_1(1 - \theta_2)a_{211} + (1 - \theta_2)(1 - \theta_1)a_{212} \\
& + \theta_2(1 - \theta_2)a_{221} + \theta_2^2 a_{222}
\end{aligned}
\tag{B-3}
$$

where a_{ijk} are given by

$$a_{ijk} = p(i|X_3) \frac{p(j|X_4)p(k|X_5)}{P(\omega = j)P(\omega = k)} \tag{B-4}$$

and i, j, and k take values 1 or 2. From equations (B-2) and (B-3), the likelihood function can be expressed in terms of parameter θ_1 as

$$L(\theta_1) = b_1 \theta^2 + b_2\theta_1 + b_3 \tag{B-5}$$

where

$$
\left.
\begin{aligned}
b_1 &= (a_{111} - a_{112}) + \frac{P_1}{P_2}(a_{121} - a_{122} - a_{211} + a_{212}) - \left(\frac{P_1}{P_2}\right)^2 a_{221} \\
b_2 &= (a_{112} - a_{122}) + \frac{P_1}{P_2}(-2a_{121} + 2a_{122} + a_{211} - 2a_{212} - a_{221}) + 2\left(\frac{P_1}{P_2}\right)^2 a_{221} \\
b_3 &= a_{122} + \frac{P_1}{P_2}(a_{121} - a_{122} + a_{212} + a_{221}) - \left(\frac{P_1}{P_2}\right)^2 a_{221}
\end{aligned}
\right\}
\tag{B-6}
$$

Let v_1 be the value of θ_1 obtained by differentiating equation (B-5) with respect to θ_1 and equating the resulting expression to zero. That is,

$$v_1 = -\frac{b_2}{2b_1} \qquad (B-7)$$

The parameters θ_1 and θ_2 should lie in the interval 0 to 1. Let v_2 and v_3 be the end points of θ_1. They are given by

$$\left. \begin{array}{l} v_2 = 1 \\[2ex] v_3 = 0, \text{ if } \dfrac{P_1}{P_2} < 1; \text{ otherwise, } v_3 = \left(1 - \dfrac{P_2}{P_1}\right) \end{array} \right\} \qquad (B-8)$$

and

Now, the optimal θ_1 and θ_2 can be obtained as follows. If $0 < v_1 < 1$, choose the optimal value of θ_1, θ_{1opt}, that equals the value v_1, v_2, or v_3 and gives the largest value of $L(\theta_{1opt})$. If v_1 lies outside the interval 0 to 1, choose the value for θ_{1opt} that equals the value v_2 or v_3 and gives the largest value of $L(\theta_{1opt})$; θ_{2opt} is computed from equation (B-2).

Example: For a few typical sequential neighborhoods, this example illustrates the transition probabilities computed using the linear and nonlinear models of section 2, using the procedure of this appendix and using the Davidon-Fletcher-Powell optimization technique. The a posteriori probabilities of the classes are of the 3-by-3 local neighborhood of dot 89 from segment 1739, Teton County, Montana. These are obtained by normalizing the outputs of a linear classifier. The four-sequential neighborhoods are neighborhoods in four directions: 0°, 45°, 90°, and 135°, centering on the central pixel. The a priori probabilities are computed as the average of the a posteriori probabilities in the neighborhood. Class 1 is wheat and class 2 is "other." The a priori probabilities computed for this 3-by-3 neighborhood are

$$\left. \begin{array}{l} P(\omega = 1) = 0.5531 \\[1ex] P(\omega = 2) = 0.4469 \end{array} \right\} \qquad (B-9)$$

The estimated transition probabilities are listed in table B-1.

TABLE B-1.- COMPARISON OF ESTIMATED TRANSITION PROBABILITIES

A posteriori probabilities in the neighborhood	Transition probabilities			
	Linear model	Nonlinear model	Procedure of appendix B	Direct optimization (using Davidon-Fletcher-Powell procedure)
0.5710, 0.8140, 0.3700, 0.4290 0.1860 0.6300 0° direction	0.5531 0.4469 0.5531 0.4469 θ = 0.0	0.1984 0.8016 0.8610 0.1390 θ = -3.95	0.1920 0.8080 1.0 0.0 θ₁ = 0.1920 θ₂ = 0.0	0.1965 0.8035 1.0 0.0
0.8490, 0.8140, 0.1340, 0.1510 0.1860 0.8660 45° direction	0.5531 0.4469 0.5531 0.4469 θ = 0.0	0.4692 0.5308 0.6341 0.3659 θ = -0.4	0.4879 0.5121 0.6338 0.3662 θ₁ = 0.4879 θ₂ = 0.3662	0.4913 0.5087 0.6363 0.3637
0.468, 0.814, 0.717, 0.532 0.186 0.283 90° direction	0.9240 0.0760 0.094 0.906 θ = 0.83	0.8609 0.1391 0.1984 0.8016 θ = 0.8	0.9240 0.076 0.0940 0.9060 θ₁ = 0.924 θ₂ = 0.906	0.9409 0.0591 0.0877 0.9123
0.236, 0.814, 0.818 0.764 0.186 0.182 135° direction	0.5531 0.4469 0.5531 0.4469 θ = 0.0	0.5790 0.4210 0.5269 0.4731 θ = 0.1	0.5123 0.4877 0.6036 0.3964 θ₁ = 0.5123 θ₂ = 0.3964	0.5196 0.4804 0.6079 0.3921

$\theta_1 = 0.1920$, $\theta_2 = 0.0$ (0° direction, Procedure of appendix B)

$\theta_1 = 0.4879$, $\theta_2 = 0.3662$ (45° direction)

$\theta_1 = 0.924$, $\theta_2 = 0.906$ (90° direction)

$\theta_1 = 0.5123$, $\theta_2 = 0.3964$ (135° direction)

Table B-1 shows that the estimated transition probabilities agree well with different procedures. With linear models, the parameter θ tends to be zero for mixed neighborhoods, thus ignoring spatial information from mixed neighborhoods.

APPENDIX C

ESTIMATION OF TRANSITION PROBABILITIES WITH DIFFERENT
PARAMETERS IN THE LOCAL NEIGHBORHOOD

APPENDIX C

ESTIMATION OF TRANSITION PROBABILITIES WITH DIFFERENT
PARAMETERS IN THE LOCAL NEIGHBORHOOD

In this appendix, some results are developed for estimating the transition
probabilities with different parameters in different directions in the local
neighburhood and with interactions in the parameters. The local neighborhood
considered is shown in figure 3-1. The linear model of equation (1) is used
for the transition probabilities. Let θ_H and θ_V be the parameters of the
transition probabilities model for horizontal and vertical neighbors,
respectively. For the local neighborhood illustrated in figure 3-1, consider
the following equation from section 3.2:

$$L(\overline{\theta}) = L(\theta_H, \theta_V) = \frac{p(X_0, X_1, \cdots, X_4)}{\prod_{j=0}^{4} p(X_j)}$$

$$= \sum_{i_0=1}^{M} p(\omega = i_0|X_0) \left\{ \prod_{\substack{j=1 \\ j\neq2 \\ j\neq4}}^{4} \left[(1 - \theta_V) + \theta_V \frac{p(\omega = i_0|X_i)}{p(\omega = i_0)} \right] \right\}$$

$$\left\{ \prod_{\substack{j=1 \\ j\neq1 \\ j\neq3}}^{4} \left[(1 - \theta_H) + \theta_H \frac{p(\omega = i_0|X_j)}{p(\omega = i_0)} \right] \right\}$$

$$= (1 - \theta_V)^2(1 - \theta_H)^2 + (1 - \theta_V)^2(1 - \theta_H)\theta_H\alpha_H + (1 - \theta_V)^2\theta_H^2\beta_H$$

$$+ (1 - \theta_V)\theta_V(1 - \theta_H)^2\alpha_V + (1 - \theta_V)\theta_V(1 - \theta_H)\theta_H\alpha_{VH} + (1 - \theta_V)\theta_V\theta_H^2\alpha\beta_{VH}$$

$$+ \theta_V^2(1 - \theta_H)^2\beta_V + \theta_V^2(1 - \theta_H)\theta_H\alpha\beta_{HV} + \theta_V^2\theta_H^2\beta_{VH} \tag{C-1}$$

where

$$\alpha_H = \sum_{i_0=1}^{M} \frac{p(\omega = i_0|X_0)}{p(\omega = i_0)} [p(\omega = i_0|X_2) + p(\omega = i_0|X_4)]$$

$$\beta_H = \sum_{i_0=1}^{M} \frac{p(\omega = i_0|X_0)}{p^2(\omega = i_0)} [p(\omega = i_0|X_2)p(\omega = i_0|X_4)]$$

$$\alpha_V = \sum_{i_0=1}^{M} \frac{p(\omega = i_0|X_0)}{p(\omega = i_0)} [p(\omega = i_0|X_1) + p(\omega = i_0|X_3)]$$

$$\beta_V = \sum_{i_0=1}^{M} \frac{p(\omega = i_0|X_0)}{p^2(\omega = i_0)} [p(\omega = i_0|X_1)p(\omega = i_0|X_3)]$$

$$\alpha_{VH} = \sum_{i_0=1}^{M} \frac{p(\omega = i_0|X_0)}{p^2(\omega = i_0)} [p(\omega = i_0|X_1) + p(\omega = i_0|X_3)][p(\omega = i_0|X_2)$$

$$+ p(\omega = i_0|X_4)]$$

$$\beta_{VH} = \sum_{i_0=1}^{M} \frac{p(\omega = i_0|X_0)}{p^4(\omega = i_0)} [p(\omega = i_0|X_1)p(\omega = i_0|X_3)][p(\omega = i_0|X_2)p(\omega = i_0|X_4)]$$

$$\alpha\beta_{VH} = \sum_{i_0=1}^{M} \frac{p(\omega = i_0|X_0)}{p^3(\omega = i_0)} [p(\omega = i_0|X_2) + p(\omega = i_0|X_4)][p(\omega = i_0|X_1)p(\omega = i_0|X_3)]$$

$$\alpha\beta_{VH} = \sum_{i_0=1}^{M} \frac{p(\omega = i_0|X_0)}{p^3(\omega = i_0)} [p(\omega = i_0|X_1) + p(\omega = i_0|X_3)][p(\omega = i_0|X_2)p(\omega = i_0|X_4)]$$

To determine θ_V and θ_H that maximize equation (C-1), one takes partial derivatives of equation (C-1) with respect to θ_V and θ_H and solves the resulting equations for θ_V and θ_H. Taking the partial derivative of equation (C-1) with respect to θ_V, equating the resulting expression to zero, and solving for θ_V, one obtains

$$\theta_V = \left(\frac{1}{2}\right) \frac{a_{N2}\theta_H^2 + a_{N1}\theta_H + a_{N0}}{a_{D2}\theta_H^2 + a_{D1}\theta_H + a_{D0}} \tag{C-2}$$

where

$$a_{N2} = 2 - 2\alpha_H + 2\beta_H - \alpha_V + \alpha_{VH} - \alpha\beta_{VH}$$

$$a_{N1} = -4 + 2\alpha_H + 2\alpha_V - \alpha_{VH}$$

$$a_{N0} = 2 - \alpha_V$$

$$a_{D2} = 1 - \alpha_H + \beta_H - \alpha_V + \alpha_{VH} - \alpha\beta_{VH} + \beta_V - \alpha\beta_{HV} + \beta_{VH}$$

$$a_{D1} = -2 + \alpha_H + 2\alpha_V - \alpha_{VH} - 2\beta_V + \alpha\beta_{HV}$$

$$a_{D0} = 1 - \alpha_V + \beta_V$$

Similarly, taking the partial derivative of equation (C-1) with respect to θ_H, equating the resulting expression to zero, and solving for θ_H, one obtains

$$\theta_H = \left(\frac{1}{2}\right) \frac{b_{N2}\theta_V^2 + b_{N1}\theta_V + b_{N0}}{b_{D2}\theta_V^2 + b_{D1}\theta_V + b_{D0}} \tag{C-3}$$

where

$$b_{N2} = 2 - \alpha_H - 2\alpha_V + \alpha_{VH} + 2\beta_V - \alpha\beta_{HV}$$

$$b_{N1} = -4 + 2\alpha_H + 2\alpha_V - \alpha_{VH}$$

$$b_{N0} = 2 - \alpha_H$$

$$b_{D2} = 1 - \alpha_H + \beta_H - \alpha_V + \alpha_{VH} - \alpha\beta_{VH} + \beta_V - \alpha\beta_{HV} + \beta_{VH}$$

$$b_{D1} = -2 + 2\alpha_H - 2\beta_H + \alpha_V - \alpha_{VH} + 2\beta_{VH}$$

$$b_{D0} = 1 - \alpha_H + \beta_H$$

Substituting the expression for θ_V from equation (C-2) into equation (C-3) results in a fifth-order algebraic equation, the roots of which can be obtained by numerical methods (refs. 15, 16). Let the resulting roots be $\theta_{Hr}(i)$; $i = 1, 2, \cdots, 5$. From equation (C-2), corresponding values are obtained for $\theta_{Vr}(i)$; $i = 1, 2, \cdots, 5$.

Let

$$\bar{\theta}_r(i) = \begin{bmatrix} \theta_{Hr}(i) \\ \theta_{Vr}(i) \end{bmatrix} \quad ; \quad i = 1,2,\cdots,5 \tag{C-4}$$

where $\bar{\theta}_r(i)$ is a vector. Let

$$\bar{\delta}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \; \bar{\delta}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \; \bar{\delta}_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \; \bar{\delta}_4 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{C-5}$$

Now, $\bar{\theta}_{opt}$ for $0 \le \theta_H \le 1$ and $0 \le \theta_V \le 1$, which maximizes equation (C-1), can be obtained using a procedure similar to that given in the flow diagram of figure 3-7. The above analysis can be generalized with different parameters for more than two directions and for larger neighborhoods to obtain the transition probabilities.

APPENDIX D

MULTITEMPORAL INTERPRETATION OF CONTEXT

APPENDIX D

MULTITEMPORAL INTERPRETATION OF CONTEXT

In this appendix, a multitemporal interpretation of the theory developed in the paper is given for applications such as those in the machine processing of remotely sensed imagery data. In remote sensing, the sensor system usually makes several passes over the same ground area and acquires a set of data for each pass or acquisition. The data from these passes are registered, and the classification is performed on the registered data. Let there be r acquisitions. For every pixel in acquisition i, a data vector X_i ($i = 1, 2, \cdots, r$) is acquired. Suppose that acquisitions $2, \cdots, r$ are registered with respect to acquisition 1. There will be variations in the data of each pixel from acquisition to acquisition. Also, errors are encountered in registration.

Let the classifier be trained on the data representative of the individual acquisitions, obtaining the probability density functions $p(X|\omega = i)$, $i = 1, 2, \cdots, M$, for each acquisition. This appendix presents the application of the theory of sequential context developed in the paper for the classification of the pixel under consideration. Let X_i be the spectral vector of the pixel under consideration in acquisition i ($i = 1, 2, \cdots, r$) in the classification of the pixel under consideration. This approach takes into account the registration errors and the variations in the data from acquisition to acquisition. The pixel is classified using the decision rule: Classify it to class $\omega = j$, if

$$p(\omega = j|X_1,\cdots,X_r) > p(\omega = i|X_1,\cdots,X_r) \quad ; \quad i = 1,2,\cdots,M \qquad (D-1)$$
$$i \neq j$$

The dependencies from acquisition to acquisition can be modeled through the models of section 2; the transition probabilities can then be estimated using the techniques developed in section 3; the a posteriori probabilities of the classes of the pixel, using data from all the acquisitions, can be computed using the techniques developed in section 4; and the pixel can be classified using equation (D-1).

If there are no errors in registering the data from acquisition to acquisition, the transition probabilities satisfy the following relation.

$$P(\omega_n = k|\omega_{n-1} = i) = 1 \text{ if } i = k$$
$$= 0 \text{ if } i \neq k \qquad (D-2)$$

where ω_n is the class of the pixel under consideration from the nth acquisition. Using sequential context [equation (45)], the a posteriori probabilities of the classes of the pixel under consideration can be written in terms of the pixel spectral vectors from each acquisition as follows.

$$p(\omega_r = i_r|X_1, \cdots, X_r) = \frac{NU(i_r)}{\sum_{i_r=1}^{M} NU(i_r)} \qquad (D-3)$$

where

$$NU(i_r) = \frac{p(\omega_r = i_r|X_r)}{P(\omega_r = i_r)} \left[\sum_{i_{r-1}=1}^{M} P(\omega_r = i_r|\omega_{r-1} = i_{r-1}) \frac{p(\omega_{r-1} = i_{r-1}|X_{r-1})}{P(\omega_{r-1} = i_{r-1})} \right.$$
$$\left. \cdots \sum_{i_1=1}^{M} P(\omega_2 = i_2|\omega_1 = i_1)p(\omega_1 = i_1|X_1) \right] \qquad (D-4)$$

From equations (D-2) through (D-4), equation (D-5) is obtained.

$$p(\omega_r = i_r|X_1, \cdots, X_r)$$

$$= \frac{P(\omega_1 = i_r)p(X_1|\omega_1 = i_r)p(X_2|\omega_2 = i_r)\cdots p(X_r|\omega_r = i_r)}{\sum_{i_r=1}^{M} P(\omega_1 = i_r)p(X_1|\omega_1 = i_r)p(X_2|\omega_2 = i_r)\cdots p(X_r|\omega_r = i_r)} \qquad (D-5)$$

Thus, use of sequential context with assumptions (a) and (b) of section 4 and of equation (D-2) in the classification of a pixel in a multitemporal situation amounts to the class-conditional independence of the pixel spectral vectors of each acquisition. Equation (D-5) can also be written as follows.

$$p(\omega_j = i_j | X_1, \cdots, X_j) = \frac{p(\omega_{j-1} = i_j | X_1, \cdots, X_{j-1}) p(X_j | \omega_j = i_j)}{\sum_{i_j=1}^{M} p(\omega_{j-1} = i_j | X_1, \cdots, X_{j-1}) p(X_j | \omega_j = i_j)} \Bigg\} \quad (D-6)$$

$$\text{for } j = 1, 2, \cdots, r$$

with
$$p(\omega_1 = i_1 | X_1) = \frac{P(\omega_1 = i_1) p(X_1 | \omega_1 = i_1)}{\sum_{i_1=1}^{M} P(\omega_1 = i_1) p(X_1 | \omega_1 = i_1)} \quad (D-7)$$

Equations (D-6) and (D-7) can be interpreted as follows: When the first acquisition is acquired, the a priori knowledge $P(\omega_1 = i_1)$ about the classes of the pixel under consideration is modified into a posteriori probabilities according to equation (D-7). These will become the a priori knowledge for the next acquisition. With the use of the observed spectral vector, the a priori knowledge is modified into a posteriori probabilities according to equation (D-6). When no registration errors are present, equation (D-6) can be used sequentially in a multitemporal situation to incorporate the contextual information in the classification of the pixel under consideration.

However, using the techniques developed in the paper, this multitemporal interpretation can be easily coupled with the spatial information from two-dimensional neighborhoods.