JSC-14859

JUN 19 1979

Lockheed Electronics Company, Inc.

A SUBSIDIARY OF LOCKHEED CORPORATION

1830 NASA Road 1, Houston, Texas 77058
Tel. 713-333-5411

NASA CR-160261

(NASA-CR-160261) LOCAL NEIGHBORHOOD TRANSITION PROBABILITY ESTIMATION AND ITS USE IN CONTEXTUAL CLASSIFICATION (Lockheed Electronics Co.) 33 p HC A03/MF A01 CSCL 12A G3/65

N79-27916

Unclas 27975

TECHNICAL MEMORANDUM

LOCAL NEIGHBORHOOD TRANSITION PROBABILITY ESTIMATION AND ITS USE IN CONTEXTUAL CLASSIFICATION

By

C. B. Chittineni

Approved By: T. C. Minter

T. C. Minter, Supervisor
Techniques Development Section

May 1979

LEC-13344

CONTENTS

FIGURES

Figure		Page

1. INTRODUCTION

In the literature there is considerable interest in the incorporation of contextual information into classification, especially in the development of methods for character recognition. Generally, one of two basic approaches has been followed, the table look-up method or the Markov approach. The table look-up method is based on the assumption that every word in the text is selected from a known finite table. A word of text is classified by comparing it with every word in the table having the same length 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 leads to the estimation of the probabilities of all possible pairs, triples, etc., of characters; i.e., transition probabilities from sample text.

Abend (ref. 1) derived optimal procedures when a Markov dependence exists between the states of nature, and Raviv (ref. 2) gives the results of applying such procedures for the recognition of English text. Chow (ref. 3), using a nearest-neighbor dependence method, obtained the structure and parameters of a recognition network for patterns represented by binary matrices.

Use of contextual analysis in speech recognition is considered by Alter (ref. 4). Welch and Salter (ref. 5) present an algorithm for the incorporation of contextual information in the classification of picture elements (pixels) in an image. Chittineni (ref. 6) discusses the use of context with the linear classifiers.

All of these approaches assume the availability of, or estimation from a sample, the transition probabilities. In the application of the above techniques for classification of imagery data such as that obtained in remote sensing, it is difficult to estimate the transition probabilities; and very often they vary from one image to the other.

This paper presents a simple model for the transition probabilities in terms of a single parameter Θ and presents methods for the incorporation of contextual information into classification in terms of Θ. Techniques for locally estimating Θ, based on classifier decisions and using a maximum likelihood method, are developed. The paper is organized as follows.

Section 2 presents a model for the transition probabilities. Section 3 discusses the incorporation of context into classification. Section 4 develops techniques for locally estimating the parameter of transition probability model using the maximum likelihood method. Section 5 presents conclusions. Appendix A develops some results for estimating the parameter of transition probabilities under the assumption of different transition probability models for horizontal and vertical neighbors. Appendix B presents a multitemporal interpretation of the techniques developed in the paper for remote sensing applications by minimizing the registration errors and incorporating context into classification.

2. A MODEL FOR REPRESENTING TRANSITION PROBABILITIES

Let i and j be the neighboring picture elements (pixels) with pattern vectors X_i and X_j and class labels ω_i and ω_j respectively. Let ω_i and ω_j take values r and s. Let $P(\omega = r)$ be the *a priori* probability of class r. If it is assumed that the labels of pixels i and j are independent, then

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

Similarly, if it is assumed that the labels of the pixels i and j are completely dependent,

$$\left.\begin{aligned} P(\omega_i = r|\omega_j = s) &= 0 \\ \text{and} \qquad & \\ P(\omega_i = r|\omega_j = r) &= 1 \end{aligned}\right\} \tag{2}$$

Because generally a dependence exists between neighboring pixels, this dependency is modeled through a parameter Θ, which lies between 0 and 1 as

$$\left.\begin{aligned} &P(\omega_i = r|\omega_j = s) = (1 - \Theta)P(\omega_i = r) \\ \text{and} \\ &P(\omega_i = r|\omega_j = r) = (1 - \Theta)P(\omega_i = r) + \Theta \end{aligned}\right\} \tag{3}$$

where

$$0 \leq \Theta \leq 1 \tag{4}$$

From equations (1), (2), and (3), it is easily seen that, $\Theta = 1$ denotes complete dependence and that $\Theta = 0$ denotes independence.

The following shows that this definition satisfies the postulates of probability. Let there be M classes. Consider that

$$\begin{aligned} \sum_{r=1}^{M} P(\omega_i = r|\omega_j = s) &= P(\omega_i = s|\omega_j = s) + \sum_{\substack{r=1 \\ \neq s}}^{M} P(\omega_i = r|\omega_j = s) \\ &= \left[(1 - \Theta)P(\omega_i = s) + \Theta\right] + \sum_{\substack{r=1 \\ \neq s}}^{M} (1 - \Theta)P(\omega_i = r) \\ &= 1 - \Theta + \Theta = 1 \end{aligned}$$

thus satisfying the probability rule. However, it is to be noted that the dependencies between the neighboring pixels can be modeled through some other parameter; for example, by replacing $\Theta = \frac{\alpha}{1 + \alpha}$, then the dependencies depend on α, $0 \leq \alpha \leq \infty$, by replacing $\Theta = \frac{e^{-\beta}}{1 - e^{-\beta}}$, then the dependencies depend on β, $-\infty \leq \beta \leq \infty$. This paper assumes that the spatial dependencies are modeled according to equations (3) and (4).

3. CONTEXTUAL CLASSIFIERS

Using the transition probabilities model of the previous section, this section develops methods for incorporating contextual information into the classifier decision process.

3.1 SPATIALLY UNIFORM CONTEXT

It is assumed that Θ holds hood for transition probabilities representation in the neighborhood under consideration. Consider a neighborhood of pixels shown in the following figure.

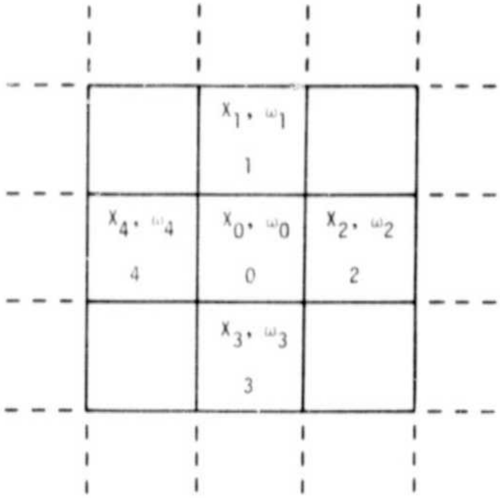

Figure 1.— Four neighbors of pixel 0.

The pattern vectors and class labels of these pixels are denoted by X_i, ω_i, $i = 0, 1, 2, \cdots, 4$. Suppose that pixel 0 is under consideration. Pixel 0 is classified into class i_0 on the basis of *a posteriori* probabilities $p(\omega_0 = i_0 | X_0, X_1, \cdots, X_4)$. Let $f = p(X_0, X_1, \cdots, X_4)$. Consider the following:

$$
\begin{aligned}
p(\omega_0 = i_0 | X_0, X_1, \cdots, X_4) &= \frac{p(\omega_0 = i_0, X_0, X_1, \cdots, X_4)}{p(X_0, X_1, \cdots, X_4)} \\
&= \frac{p(\omega_0 = i_0, X_0, X_1, \cdots, X_4)}{f} \\
&= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_4=1}^{M} \frac{p(\omega_0 = i_0, X_0, \omega_1 = i_1, X_1, \cdots, W_4 = i_4, X_4)}{f} \\
&= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_4=1}^{M} \frac{p(X_0, X_1, \cdots, X_4 | \omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4) P(\omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4)}{f}
\end{aligned} \tag{5}
$$

Making an assumption that the probability density function of a pattern given its label is independent of other labels and patterns, one can write the following:

$$
\begin{aligned}
p(X_0, X_1, \cdots, X_4 | \omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4) &= p(X_0 | \omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4, X_1, \cdots, X_4) P(X_1, \cdots, X_4 | \omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4) \\
&= p(X_0 | \omega_0 = i_0) P(X_1, \cdots, X_4 | \omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4) \\
&= \prod_{j=0}^{4} p(X_j | \omega_j = i_j)
\end{aligned}
\tag{6}
$$

Now consider

$$
\begin{aligned}
P(\omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4) &= P(\omega_0 = i_0) P(\omega_1 = i_1, \cdots, \omega_4 = i_4 | \omega_0 = i_0) \\
&= P(\omega_0 = i_0) P(\omega_1 = i_1 | \omega_0 = i_0, \omega_2 = i_2, \cdots, \omega_4 = i_4) P(\omega_2 = i_2, \cdots, \omega_4 = i_4 | \omega_0 = i_0) \\
&= P(\omega_0 = i_0) P(\omega_1 = i_1 | \omega_0 = i_0) P(\omega_2 = i_2 | \omega_0 = i_0, \omega_3 = i_3, \omega_4 = i_4) P(\omega_3 = i_3, \omega_4 = i_4 | \omega_0 = i_0) \\
&= P(\omega_0 = i_0) \prod_{j=1}^{4} P(\omega_j = i_j | \omega_0 = i_0)
\end{aligned}
\tag{7}
$$

Note that in equation (7), it was assumed that the labels of the pixels are independent of the labels of the nonneighboring pixels. From equations (5), (6), and (7), the following is obtained.

$$
\begin{aligned}
p(\omega_0 = i_0 | X_0, X_1, \cdots, X_4) &= \frac{P(\omega_0 = i_0) p(X_0 | \omega_0 = i_0)}{p(X_0, X_1, \cdots, X_4)} \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_4=1}^{M} \prod_{j=1}^{4} p(X_j | \omega_j = i_j) P(\omega_j = i_j | \omega_0 = i_0) \\
&= \frac{P(\omega_0 = i_0) p(X_0 | \omega_0 = i_0)}{p(X_0, X_1, \cdots, X_4)} \prod_{j=1}^{4} \sum_{i_j=1}^{M} \left[p(X_j | \omega_j = i_j) P(\omega_j | \omega_0 = i_0) \right]
\end{aligned}
\tag{8}
$$

However, the denominator of equation (8) can be written as

$$
\begin{aligned}
p(X_0, X_1, \cdots, X_4) &= p(X_1, \cdots, X_4) p(X_0 | X_1, \cdots, X_4) \\
&= p(X_1, \cdots, X_4) \sum_{i_0=1}^{M} p(X_0, \omega_0 = i_0 | X_1, \cdots, X_4) \\
&= p(X_1, \cdots, X_4) \sum_{i_0=1}^{M} p(X_0 | \omega_0 = i_0) p(\omega_0 = i_0 | X_1, \cdots, X_4)
\end{aligned}
$$

Now consider

$$P(\omega_0 = i_0|X_1, \cdots, X_4) = \sum_{i_1=1}^{M} \cdots \sum_{i_4=1}^{M} p(\omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4|X_1, \cdots, X_4)$$

$$= \frac{1}{P(X_1, \cdots, X_4)} \sum_{i_1=1}^{M} \cdots \sum_{i_4=1}^{M} p(X_1, \cdots, X_4|\omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4)P(\omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4)$$

$$= \frac{1}{P(X_1, \cdots, X_4)} \sum_{i_1=1}^{M} \cdots \sum_{i_4=1}^{M} P(\omega_0 = i_0) \left[\prod_{j=1}^{4} p(X_j|\omega_j = i_j)P(\omega_j = i_j|\omega_0 = i_0)\right]$$

$$= \frac{1}{P(X_1, \cdots, X_4)} P(\omega_0 = i_0) \prod_{j=1}^{4} \sum_{i_j=1}^{M} p(X_j|\omega_j = i_j)P(\omega_j = i_j|\omega_0 = i_0)$$

Hence, the following is obtained:

$$p(X_0, X_1, \cdots, X_4) = \sum_{i_0=1}^{M} p(X_0|\omega_0 = i_0)P(\omega_0 = i_0) \prod_{j=1}^{M} \sum_{i_j=1}^{M} p(X_j|\omega_j = i_j)P(\omega_j = i_j|\omega_0 = i_0) \tag{9}$$

Consider

$$\sum_{i_j=1}^{M} p(X_j|\omega_j = i_j)P(\omega_j = i_j|\omega_0 = i_0) = \sum_{\substack{i_j=1 \\ \neq i_0}}^{M} p(X_j|\omega_j = i_j)P(\omega_j = i_j|\omega_0 = i_0) + p(X_j|\omega_j = i_0)P(\omega_j = i_0|\omega_0 = i_0)$$

$$= (1 - \Theta) \sum_{i_j=1}^{M} p(X_j|\omega_j = i_j)P(\omega_j = i_j) + \Theta p(X_j|\omega_j = i_0) \tag{10}$$

Using equations (9) and (10) in equation (8), one gets

$$p(\omega_0 = i_0|X_0, X_1, \cdots, X_4) = \frac{P(\omega_0 = i_0)p(X_0|\omega_0 = i_0) \prod_{j=1}^{4} \left[(1 - \Theta) \sum_{i_j=1}^{M} p(X_j|\omega_j = i_j)P(\omega_j = i_j) + \Theta p(X_j|\omega_j = i_0)\right]}{\sum_{i_0=1}^{M} P(\omega_0 = i_0)p(X_0|\omega_0 = i_0) \prod_{j=1}^{M} \left[(1 - \Theta) \sum_{i_j=1}^{M} p(X_j|\omega_j = i_j)P(\omega_j = i_j) + \Theta p(X_j|\omega_j = i_0)\right]}$$

$$= \frac{p(\omega_0 = i_0|X_0)\left\{\prod_{j=1}^{4} \left[(1 - \Theta) + \Theta \frac{p(\omega_j = i_0|X_j)}{P(\omega_j = i_0)}\right]\right\}}{\sum_{i_0=1}^{M} p(\omega_0 = i_0|X_0)\left\{\prod_{j=1}^{4} \left[(1 - \Theta) + \Theta \frac{p(\omega_j = i_0|X_j)}{P(\omega_j = i_0)}\right]\right\}} \tag{11}$$

Equation (11) can be used to update the posteriori probabilities of the classes of pixel 0 with the incorporation of contextual information. For classification of pixel 0, the decision rule becomes the following: Decide $X_0 \varepsilon \omega_0 = i_0$, which maximizes g_{i_0}, where

$$g_{i_0} = p(\omega_0 = i_0 | X_0) \prod_{j=1}^{4} \left[(1 - \Theta) + \Theta \frac{p(\omega_j = i_0 | X_j)}{P(\omega_j = i_0)} \right]$$

3.2 SEQUENTIAL OR MARKOVIAN DEPENDENCE

This section considers the sequential or Markovian dependence between neighboring pixels with the transition probabilities as described in section 2 in terms of parameter Θ. This sequential model can be used to classify a pixel using contextual information as follows. Consider a 3×3 neighborhood of pixel 0, shown in figure 2.

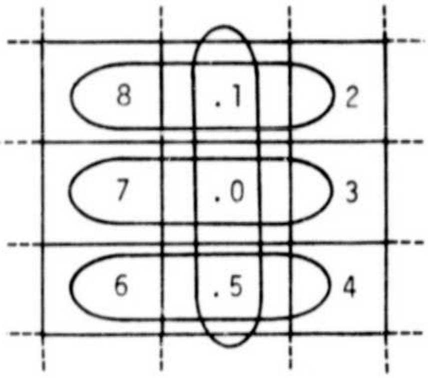

Figure 2.— Illustration of 3×3 neighborhood.

Consider pixel 1. Its posteriori probabilities are updated using the information from patterns of pixels 8 and 2 and similarly for pixels 7, 0, 3 and 6, 5, 4. Finally, the posteriori probabilities of the labels of the pattern of pixel 0 are obtained using the ones of pixels 1 and 5. Now consider the sequential neighborhood shown in figure 3.

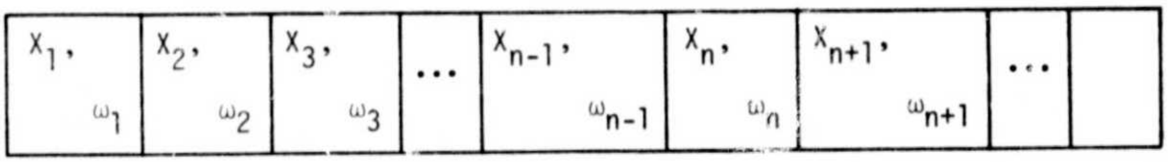

Figure 3.— Illustration of a sequential neighborhood.

Assume that the posteriori probabilities $p(\omega_{n-1} = i|X_1, \cdots, X_{n-1})$, $i = 1, 2, \cdots, M$ are known. Then the problem is to update the posteriori probabilities of X_n using the information from X_n, X_{n+1}, and $X_1, \cdots, X_{n-1}$. The following assumptions are made. Given the identification of the label of the n-1*th* pixel, the label of the n*th* pixel does not depend on the patterns of pixels 1, 2, $\cdots$, n - 1. That is,

$$p(\omega_n = k|\omega_{n-1} = j, X_1, \cdots, X_{n-1}) = P(\omega_n = k|\omega_{n-1} = j) \tag{12}$$

It is also assumed that given the identification of the label of a pattern, its density does not depend on any other information. That is,

$$p(X_n|\omega_n = k, \text{any other } X \text{ or } \omega) = p(X_n|\omega_n = k) \tag{13}$$

With these assumptions, the contextual relations using sequential or Markovian dependence are developed. Now consider

$$\begin{aligned}
p(\omega_n = k|X_1, \cdots, X_{n-1}) &= \sum_{i=1}^{M} p(\omega_n = k, \omega_{n-1} = i|X_1, \cdots, X_{n-1}) \\
&= \sum_{i=1}^{M} p(\omega_n = k|\omega_{n-1} = i, X_1, \cdots, X_{n-1})p(\omega_{n-1} = i|X_1, \cdots, X_{n-1}) \\
&= \sum_{i=1}^{M} P(\omega_n = k|\omega_{n-1} = i)p(\omega_{n-1} = i|X_1, \cdots, X_{n-1}) \\
&= \sum_{i=1}^{M} \frac{P(\omega_n = k)}{P(\omega_{n-1} = i)} P(\omega_{n-1} = i|\omega_n = k)p(\omega_{n-1} = i|X_1, \cdots, X_{n-1}) \\
&= \sum_{\substack{i=1 \\ \neq k}}^{M} \frac{P(\omega_n = k)}{P(\omega_{n-1} = i)} P(\omega_{n-1} = i|\omega_n = k)p(\omega_{n-1} = i|X_1, \cdots, X_{n-1}) \\
&\quad + \frac{P(\omega_n = k)}{P(\omega_{n-1} = k)} P(\omega_{n-1} = k|\omega_n = k)p(\omega_{n-1} = k|X_1, \cdots, X_{n-1}) \\
&= \sum_{\substack{i=1 \\ \neq k}}^{M} \frac{P(\omega_n = k)}{P(\omega_{n-1} = i)} (1 - \Theta)P(\omega_{n-1} = i)p(\omega_{n-1} = i|X_1, \cdots, X_{n-1}) \\
&\quad + \frac{P(\omega_n = k)}{P(\omega_{n-1} = k)} \left[(1 - \Theta)P(\omega_{n-1} = k) + \Theta\right]p(\omega_{n-1} = k|X_1, \cdots, X_{n-1}) \\
&= (1 - \Theta)P(\omega_n = k) \sum_{i=1}^{M} p(\omega_{n-1} = i|X_1, \cdots, X_{n-1}) + \Theta \frac{P(\omega_n = k)}{P(\omega_{n-1} = k)} p(\omega_{n-1} = k|X_1, \cdots, X_{n-1}) \\
&= (1 - \Theta)P(\omega_n = k) + \Theta \frac{P(\omega_n = k)}{P(\omega_{n-1} = k)} p(\omega_{n-1} = k|X_1, \cdots, X_{n-1})
\end{aligned} \tag{14}$$

Using similar arguments, one can write the following:

$$
\begin{aligned}
p(X_n | X_1, \cdots, X_{n-1}) &= \sum_{i=1}^{M} p(X_n, \omega_n = i | X_1, \cdots, X_{n-1}) \\
&= \sum_{i=1}^{M} p(X_n | \omega_n = i, X_1, \cdots, X_{n-1}) p(\omega_n = i | X_1, \cdots, X_{n-1}) \\
&= \sum_{i=1}^{M} p(X_n | \omega_n = i) p(\omega_n = i | X_1, \cdots, X_{n-1})
\end{aligned}
\tag{15}
$$

Now the posteriori probabilities of the label of pattern X_n are updated using the information from pattern X_n and patterns $X_1, X_2, \cdots, X_{n-1}$ as

$$
\begin{aligned}
p(\omega_n = k | X_1, \cdots, X_{n-1}, X_n) &= \frac{P(\omega_n = k, X_1, \cdots, X_{n-1}, X_n)}{P(X_1, \cdots, X_n)} \\
&= \frac{p(X_n | \omega_n = k, X_1, \cdots, X_{n-1}) p(\omega_n = k | X_1, \cdots, X_{n-1}) p(X_1, \cdots, X_{n-1})}{p(X_n | X_1, \cdots, X_{n-1}) p(X_1, \cdots, X_{n-1})} \\
&= \frac{p(X_n | \omega_n = k) p(\omega_n = k | X_1, \cdots, X_{n-1})}{\sum_{i=1}^{M} p(X_n | \omega_n = i) p(\omega_n = i | X_1, \cdots, X_{n-1})} \\
&= \frac{\frac{1}{P(\omega_n = k)} p(\omega_n = k | X_n) p(\omega_n = k | X_1, \cdots, X_{n-1})}{\sum_{i=1}^{M} \frac{1}{P(\omega_n = i)} p(\omega_n = i | X_n) p(\omega_n = i | X_1, \cdots, X_{n-1})} \\
&= \frac{(1 - \Theta) p(\omega_n = k | X_n) + \Theta \left[p(\omega_n = k | X_n) p(\omega_{n-1} = k | X_1, \cdots, X_{n-1}) / P(\omega_{n-1} = k) \right]}{\sum_{i=1}^{M} \left\{ (1 - \Theta) p(\omega_n = i | X_n) + \Theta \left[p(\omega_n = i | X_n) p(\omega_{n-1} = i | X_1, \cdots, X_{n-1}) / P(\omega_{n-1} = i) \right] \right\}}
\end{aligned}
\tag{16}
$$

The information in patterns $X_1, \cdots, X_n$ in obtaining the label of pattern X_{n+1} can be written as follows:

$$
\begin{aligned}
p(\omega_{n+1} = j|X_1, \cdots, X_n) &= \sum_{i=1}^{M} p(\omega_{n+1} = j, \omega_n = i|X_1, \cdots, X_n) \\
&= \sum_{\substack{i=1 \\ \neq j}}^{M} P(\omega_{n+1} = j|\omega_n = i)p(\omega_n = i|X_1, \cdots, X_n) + P(\omega_{n+1} = j|\omega_n = j)p(\omega_n = j|X_1, \cdots, X_n) \\
&= (1 - \Theta)P(\omega_{n+1} = j) \sum_{i=1}^{M} p(\omega_n = i|X_1, \cdots, X_n) + \Theta p(\omega_n = j|X_1, \cdots, X_n) \\
&= (1 - \Theta)P(\omega_{n+1} = j) + \Theta p(\omega_n = j|X_1, \cdots, X_n)
\end{aligned}
\tag{17}
$$

Consider

$$
\begin{aligned}
p(X_{n+1}|X_1, \cdots, X_n) &= \sum_{j=1}^{M} p(X_{n+1}, \omega_{n+1} = j|X_1, \cdots, X_n) \\
&= \sum_{j=1}^{M} p(X_{n+1}|\omega_{n+1} = j)p(\omega_{n+1} = j|X_1, \cdots, X_n)
\end{aligned}
\tag{18}
$$

Using the patterns $X_1, X_2, \cdots, X_n, X_{n+1}$, one has

$$
\begin{aligned}
p(\omega_n = k|X_1, \cdots, X_n, X_{n+1}) &= \sum_{j=1}^{M} p(\omega_n = k, \omega_{n+1} = j|X_1, \cdots, X_n, X_{n+1}) \\
&= \frac{\sum_{j=1}^{M} p(X_{n+1}|\omega_{n+1} = j, \omega_n = k, X_1, \cdots, X_n)p(\omega_{n+1} = j|\omega_n = k, X_1, \cdots, X_n)p(\omega_n = k, X_1, \cdots, X_n)}{p(X_{n+1}, X_n, \cdots, X_1)} \\
&= \frac{\sum_{j=1}^{M} p(X_{n+1}|\omega_{n+1} = j)P(\omega_{n+1} = j|\omega_n = k)p(\omega_n = k|X_1, \cdots, X_n)}{p(X_{n+1}|X_1, \cdots, X_n)}
\end{aligned}
\tag{19}
$$

The numerator of equation (19) can be written as follows:

$$
\begin{aligned}
&\sum_{\substack{j=1 \\ \neq k}}^{M} p(X_{n+1}|\omega_{n+1} = j)P(\omega_{n+1} = j|\omega_n = k)p(\omega_n = k|X_1, \cdots, X_n) + p(X_{n+1}|\omega_{n+1} = k)P(\omega_{n+1} = k|\omega_n = k)p(\omega_n = k|X_1, \cdots, X_n) \\
&= (1 - \Theta)p(\omega_n = k|X_1, \cdots, X_n) \sum_{j=1}^{M} p(X_{n+1}|\omega_{n+1} = j)P(\omega_{n+1} = j) + \Theta p(X_{n+1}|\omega_{n+1} = k)p(\omega_n = k|X_1, \cdots, X_n)
\end{aligned}
\tag{20}
$$

Now from equations (17), (18), (19), and (20), one obtains

$$
\begin{aligned}
p(\omega_n = k|X_1, \cdots, X_n, X_{n+1}) &= \frac{(1-\theta)p(\omega_n = k|X_1, \cdots, X_n)\sum_{j=1}^{M} p(X_{n+1}|\omega_{n+1} = j)P(\omega_{n+1} = j) + \theta p(X_{n+1}|\omega_{n+1} = k)p(\omega_n = k|X_1, \cdots, X_n)}{\sum_{j=1}^{M} p(X_{n+1}|\omega_{n+1} = j)[(1-\theta)P(\omega_{n+1} = j) + \theta p(\omega_n = j|X_1, \cdots, X_n)]} \\
&= \frac{\left[(1-\theta)p(\omega_n = k|X_1, \cdots, X_n) + \theta \frac{p(\omega_{n+1} = k|X_{n+1})}{P(\omega_{n+1} = k)} p(\omega_n = k|X_1, \cdots, X_n)\right]}{\left[(1-\theta) + \theta \sum_{j=1}^{M} \frac{p(\omega_{n+1} = j|X_{n+1})}{P(\omega_{n+1} = j)} p(\omega_n = j|X_1, \cdots, X_n)\right]}
\end{aligned} \tag{21}
$$

From equations (16) and (21), one obtains the desired result.

3.3 UNSUPERVISED MAXIMUM LIKELIHOOD PARAMETER ESTIMATION

One of the methods of unsupervised learning or clustering is to assume the component densities of the mixture density as normal with unknown means and covariance matrices and to draw samples independently from the mixture density and estimate the parameters of the mixture density using maximum likelihood technique. Let $\mathcal{X} = \{X_1, \cdots, X_n\}$ be a set of n unlabeled samples that are drawn independently from the mixture density:

$$p(X|\alpha) = \sum_{j=1}^{M} p(X|\omega = j, \alpha_j)P(\omega = j)$$

where α is a vector of parameters of the mixture density and α_j is a vector of parameters of the *jth* component density. The likelihood of the observed samples is, by definition, the joint density,

$$p(\mathcal{X}|\alpha) = \prod_{k=1}^{n} p(X_k|\alpha)$$

If $p(X|\omega = i)$ are assumed to be multivariate normal with the means μ_i and covariance matrices Σ_i, the equations for the local maximum likelihood estimates $\hat{\mu}_i$, $\hat{\Sigma}_i$, and $\hat{P}(\omega = i)$ under the constraints of $0 \leq \hat{P}(\omega = i) \leq 1$ and $\sum_{i=1}^{M} \hat{P}(\omega = i) = 1$ are given by the following (ref. 7).

$$\hat{P}(\omega = i) = \frac{1}{n}\sum_{k=1}^{n} \hat{P}(\omega = i|X_k, \hat{\alpha}_i)$$

$$\hat{\mu}_i = \frac{\sum_{k=1}^{n} \hat{P}(\omega = i|X_k, \hat{\alpha}_i)X_k}{\sum_{k=1}^{n} \hat{P}(\omega = i|X_k, \hat{\alpha}_i)}$$

$$\hat{\Sigma}_i = \frac{\sum_{j=1}^{n} \hat{P}(\omega = i|X_k, \hat{\alpha}_i)\left[(X_k - \hat{\mu}_i)(X_k - \hat{\mu}_i)^T\right]}{\sum_{k=1}^{n} \hat{P}(\omega = i|X_k, \hat{\alpha}_i)}$$

where

$$\hat{P}(\omega = i|X_k, \hat{\alpha}_i) = \frac{p(X_k|\omega = i, \hat{\alpha}_i)\hat{P}(\omega = i)}{\sum_{j=1}^{M} p(X_k|\omega = j, \hat{\alpha}_j)\hat{P}(\omega = j)}$$

$$= \frac{|\hat{\Sigma}_i|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(X_k - \mu_i)^T \hat{\Sigma}_i^{-1}(X_k - \hat{\mu}_i)\right]\hat{P}(\omega = i)}{\sum_{j=1}^{M} |\hat{\Sigma}_j|^{-\frac{1}{2}} \exp\left[\frac{1}{2}(X_k - \hat{\mu}_j)^T \hat{\Sigma}_j^{-1}(X_k - \hat{\mu}_j)\right]\hat{P}(\omega = j)}$$

In the application of this technique to the clustering of images in the spectral domain, the parameters are updated after a split-and-merge sequence. Updating the parameters involves the computation of the posteriors. The contextual algorithms presented in sections 3.1 and 3.2 can be used for updating the posteriors, with the estimates of transition probabilities from the local neighborhood using the techniques developed in sections 4.1 and 4.2.

4. LOCAL NEIGHBORHOOD MAXIMUM LIKELIHOOD ESTIMATION OF PARAMETER Θ

This section derives an expression for the likelihood function of patterns X_0, X_1, X_2, X_3, X_4 given the parameter Θ and obtains Θ by maximizing this function.

4.1 NEIGHBORHOOD OF FOUR

Consider a local neighborhood of pixel 0, illustrated in figure 1. The likelihood function of X_0, X_1, $\cdots$, X_4 given Θ can be written as follows:

$$\begin{aligned} p(X_0, X_1, \cdots, X_4|\Theta) &= \sum_{i_0=1}^{M}\sum_{i_1=1}^{M}\cdots\sum_{i_4=1}^{M} p(X_0, \omega_0 = i_0, X_1, \omega_1 = i_1, \cdots, X_4, \omega_4 = i_4|\Theta) \\ &= \sum_{i_0=1}^{M}\sum_{i_1=1}^{M}\cdots\sum_{i_4=1}^{M} p(X_0, X_1, \cdots, X_4|\omega_0 = i_0, \cdots, \omega_4 = i_4, \Theta)P(\omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4|\Theta) \end{aligned} \tag{22}$$

Consider

$$\begin{aligned} p(X_0, X_1, \cdots, X_4|\omega_0 = i_0, \cdots, \omega_4 = i_4, \Theta) &= p(X_0|X_1, \cdots, X_4, \omega_0 = i_0, \cdots, \omega_4 = i_4, \Theta)p(X_1, \cdots, X_4|\omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4, \Theta) \\ &= p(X_0|\omega_0 = i_0)p(X_1|X_2, \cdots, X_4, \omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4, \Theta)p(X_2, \cdots, X_4|\omega_0 = i_0, \cdots, \omega_4 = i_4, \Theta) \\ &= \prod_{j=0}^{4} p(X_j|\omega_j = i_j) \end{aligned} \tag{23}$$

To obtain equation (23), it was assumed that the probability density of a pattern, given its label, is independent of any other information. Consider

$$\begin{aligned} P(\omega_0 = i_0, \omega_1 = i_1, \cdots, \omega_4 = i_4|\Theta) &= P(\omega_0 = i_0|\Theta)P(\omega_1 = i_1, \cdots, \omega_4 = i_4|\omega_0 = i_0, \Theta) \\ &= P(\omega_0 = i_0)P(\omega_1 = i_1|\omega_0 = i_0, \omega_2 = i_2, \cdots, \omega_4 = i_4, \Theta)P(\omega_2 = i_2, \cdots, \omega_4 = i_4|\omega_0 = i_0, \Theta) \\ &= P(\omega_0 = i_0)\prod_{j=1}^{4} P(\omega_j = i_j|\omega_0 = i_0, \Theta) \end{aligned} \tag{24}$$

The derivation of equation (24) is based on the assumption that the label of a pattern is indpendent of the labels of the nonneighboring patterns. From equations (22), (23), and (24), the following is obtained:

$$\begin{aligned} p(X_0, X_1, \cdots, X_4|\Theta) &= \sum_{i_0=1}^{M}\sum_{i_1=1}^{M}\cdots\sum_{i_4=1}^{M}\left[\prod_{j=0}^{4} p(X_j|\omega_j = i_j)\right]P(\omega_0 = i_0)\left[\prod_{j=1}^{4} P(\omega_j = i_j|\omega_0 = i_0, \Theta)\right] \\ &= \sum_{i_0=1}^{M} P(\omega_0 = i_0)p(X_0|\omega_0 = i_0)\sum_{i_1=1}^{M}\cdots\sum_{i_4=1}^{M}\left[\prod_{j=1}^{4} p(X_j|\omega = i_j)P(\omega_j = i_j|\omega_0 = i_0, \Theta)\right] \end{aligned} \tag{25}$$

Interchanging the product and the summations results in the following expression:

$$p(X_0, X_1, \cdots, X_4|\Theta) = \sum_{i_0=1}^{M} P(\omega_0 = i_0)p(X_0|\omega_0 = i_0) \prod_{j=1}^{4} \left[\sum_{i_j=1}^{M} p(X_j|\omega_j = i_j)P(\omega_j = i_j|\omega_0 = i_0, \Theta) \right] \tag{26}$$

Since $\prod_{j=0}^{4} p(X_j)$ is independent of Θ, dividing both sides of the above equation by it and noting that the *a priori* probabilities are independent of pattern location, one obtains

$$\frac{p(X_0, X_1, \cdots, X_4|\Theta)}{\prod_{j=0}^{4} p(X_j)} = \sum_{i_0=1}^{M} p(\omega_0 = i_0|X_0) \left\{ \prod_{j=1}^{4} \left[\sum_{i_j=1}^{M} \frac{p(\omega_j = i_j|X_j)}{P(\omega_j = i_j)} P(\omega_j = i_j|\omega_0 = i_0, \Theta) \right] \right\}$$

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

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

$$= \sum_{i_0=1}^{M} p(\omega = i_0|X_0) \left\{ \prod_{j=1}^{4} \left[(1 - \Theta) + \Theta \frac{p(\omega = i_0|X_j)}{P(\omega = i_0)} \right] \right\} \tag{27}$$

Let

$$L(\Theta) = \frac{p(X_0, X_1, \cdots, X_4|\Theta)}{\prod_{j=0}^{4} p(X_j)} \tag{28}$$

Expanding equation (27) yields

$$L(\Theta) = (1 - \Theta)^4 + \Theta(1 - \Theta)^3 A + \Theta^2(1 - \Theta)^2 B + \Theta^3(1 - \Theta)C + \Theta^4 D \tag{29}$$

where

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

$$B = \sum_{i_0=1}^{M} \frac{p(\omega = i_0|X_0)}{p^2(\omega = i_0)} [p(\omega = i_0|X_1)p(\omega = i_0|X_2 + p(\omega = i_0|X_1)p(\omega = i_0|X_3) + p(\omega = i_0|X_1)p(\omega = i_0|X_4)$$

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

$$C = \sum_{i_0=1}^{4} \frac{p(\omega = i_0|X_0)}{p^3(\omega = i_0)} [p(\omega = i_0|X_1)p(\omega = i_0|X_2)p(\omega = i_0|X_3) + p(\omega = i_0|X_1)p(\omega = i_0|X_2)p(\omega = i_0|X_4)$$

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

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

To maximize $L(\Theta)$, the derivative of $L(\Theta)$ is taken with respect to Θ, and the resulting expression is equated to zero. This results in

$$\frac{\partial L(\Theta)}{\partial \Theta} = a\Theta^3 + b\Theta^2 + c\Theta + d = 0 \tag{30}$$

where

$$a = 4 - 4A + 4B - 4C + 4D, \qquad b = -12 + 9A - 6B + 3C$$

$$c = 12 - 6A + 2B, \qquad d = -4 + A$$

Equation (30) is a cubic equation with real coefficients; hence, it will have either three real roots or one real root and two complex roots. With a change of variable (ref. 8),

$$Z = a\Theta + b \tag{31}$$

one obtains from equation (30)

$$Z^3 + 3HZ + G = 0 \tag{32}$$

where

$$\left.\begin{aligned} H &= ac - b^2 \\ G &= a^2d - 3abc + 2c^3 \end{aligned}\right\} \tag{33}$$

Let

$$
\left.
\begin{aligned}
p &= \frac{1}{2}\left[-G + \sqrt{G^2 + 4H^3}\right] \\
&\text{and} \\
q &= \frac{1}{2}\left[-G - \sqrt{G^2 + 4H^3}\right]
\end{aligned}
\right\}
\tag{34}
$$

The roots of the cubic equation $y^3 - 1 = 0$ are 1, $-\frac{1}{2} + \frac{1}{2}\sqrt{-3}$ and $-\frac{1}{2} - \frac{1}{2}\sqrt{-3}$. If either of the imaginary roots is represented by u, the other is u^2. That is,

$$
y^3 - 1 = (y - 1)(y - u)(y - u^2) \tag{35}
$$

Then the roots of equation (32) can be shown to be

$$
\left.
\begin{aligned}
z_1 &= \sqrt[3]{p} + \frac{-H}{\sqrt[3]{p}} \\
z_2 &= \omega \sqrt[3]{p} + \frac{-H}{\sqrt[3]{\ }} \\
z_3 &= \omega^2 \sqrt[3]{\ } + \omega \frac{-H}{\sqrt[3]{\ }}
\end{aligned}
\right\}
\tag{36}
$$

The roots of equation (30), Θ_i, i = 1, 2, 3 can thus be obtained from equations (31) and (36).

The $L(\Theta)$ is a continuous function in Θ. The Θ in the range $0 \leq \Theta \leq 1$, which maximizes $L(\Theta)$, can be found using figure 4 (ref. 9).

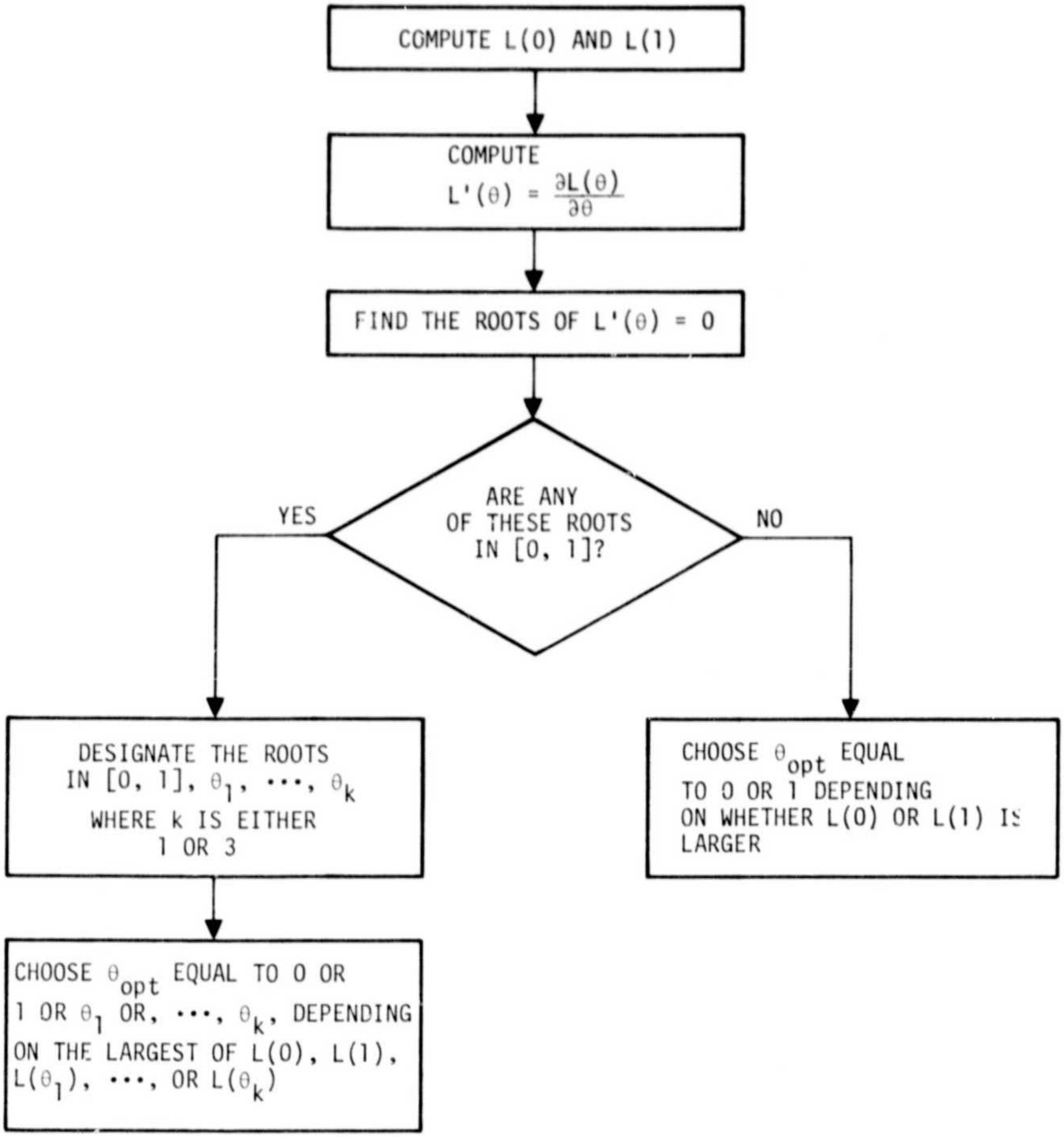

Figure 4.– Procedure for finding Θ_{opt} in the range $0 \leq \Theta \leq 1$, which gives the global maximum for $L(\Theta)$.

The Θ_{opt} of this section can be used with the spatially uniform context algorithm presented in section 3.1.

4.2 NEIGHBORHOOD OF TWO

Consider a local neighborhood of pixel 0, illustrated in figure 5.

X_1	X_0	X_2
1	0	2

Figure 5.— Two neighbors of pixel 0.

Let

$$L(\Theta) = \frac{p(X_0, X_1, X_2|\Theta)}{\prod_{j=0}^{2} p(X_j)} \tag{37}$$

Using similar arguments as in section 4.1, one obtains an expression for $L(\Theta)$ as follows:

$$L(\Theta) = (1 - \Theta)^2 + \Theta(1 - \Theta)A + \Theta^2 B \tag{38}$$

where

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

and

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

Taking the derivative of $L(\Theta)$ with respect to Θ and equating the resulting expression to 0 yields

$$\frac{\partial L(\Theta)}{\partial \Theta} = 2(\Theta - 1) + (1 - 2\Theta)A + 2\Theta B = 0 \tag{39}$$

The root Θ_1 of equation (39) is given by

$$\Theta_1 = \frac{1}{2}\left[\frac{2 - A}{1 - A + B}\right] \tag{40}$$

Because $L(\Theta)$ is a continuous function of Θ, the optimal value of Θ, Θ_{opt} in the range $0 \leq \Theta_{opt} \leq 1$, which gives maximum value for $L(\Theta)$, can be found using a procedure similar to that given in figure 4. The Θ_{opt} can then be used with the sequential contextual algorithm presented in section 3.2.

5. CONCLUSIONS

This paper considers the problem of incorporating contextual or spatial information into classification. The dependencies between neighboring patterns are modeled through a single parameter Θ, which describes the transition probabilities of the classes of the neighboring patterns.

Expressions are derived for updating the posteriori probabilities of the classes of the pattern under consideration using contextual information both for a spatially uniform contextual model and for sequential or Markovian dependencies between neighboring patterns. A likelihood function for the patterns in the neighborhood of the pattern under consideration, given the parameter Θ, is derived, and the optimal value of Θ can be obtained by maximizing the likelihood function.

The techniques presented in this paper can be used for the incorporation of contextual information both for supervised and unsupervised classifications. Incorporation of context in unsupervised learning or clustering by using maximum likelihood estimates for the parameters of a mixture density with the component Gaussian densities is briefly described.

Instead of using one parameter Θ in the neighborhood of the pattern under consideration, as shown in appendix A, transition probability models with different parameters can be used. The techinques, as shown in appendix B, can be extended for multitemporal or time-varying situations such as those encountered in remote sensing.

The procedures developed for a local neighborhood estimate of Θ can be used under some other modeling of transition probabilities as long as the transition probability modeling satisfies the probability postulates. For example, Θ can be replaced with $\frac{\alpha}{1 + \alpha}$, where α lies between 0 and ∞ or with $\frac{e^{-\beta}}{1 + e^{-\beta}}$, where β can be between $-\infty$ and ∞.

APPENDIX A

ESTIMATION OF TRANSITION PROBABILITIES WITH DIFFERENT PARAMETERS IN THE LOCAL NEIGHBORHOOD

This appendix develops some results for estimating the parameters of transition probabilities under different models for horizontal and vertical neighbors. Let θ_H and θ_V be the parameters of transition probability model for horizontal and vertical neighbors respectively. For the local neighborhood illustrated in figure 1, consider the following equation from section 4.1:

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

$$= \sum_{i_0=1}^{M} p(\omega = i_0|X_0)\left\{\prod_{\substack{j=1\\ \neq 2\\ \neq 4}}^{4}\left[(1 - \theta_V) + \theta_V \frac{p(\omega = i_0|X_j)}{p(\omega = i_0)}\right]\right\}\left\{\prod_{\substack{j=1\\ \neq 1\\ \neq 3}}^{4}\left[(1 - \theta_H) + \theta_H \frac{p(\omega = i_0|X_j)}{p(\omega = i_0)}\right]\right\}$$

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

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

$$+ \theta_V^2(1 - \theta_H)^2\beta_V + \theta_V^2(1 - \theta_H)\theta_H\alpha\beta_{HV} + \theta_V^2\theta_H^2\beta_{VH} \qquad \text{(A-1)}$$

where

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

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

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

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

$$\alpha_{VH} = \sum_{i_0=1}^{M} \frac{p(\omega = i_0|X_0)}{p^2(\omega = i_0)}[p(\omega = i_0|X_1) + p(\omega = i_0|X_3)][p(\omega = i_0|X_2) + p(\omega = i_0|X_4)]$$

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

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

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

To determine θ_V and θ_H that maximize equation (A-1), one takes partial derivations of equation (A-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 (A-1) with respect to θ_V, equating the resulting expression to zero, and solving for θ_V, one obtains

$$\theta_V = (1/2)\frac{a_{N2}\theta_H^2 + a_{N1}\theta_H + a_{N0}}{a_{D2}\theta_H^2 + a_{D1}\theta_H + a_{D0}} \qquad \text{(A-2)}$$

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 (A-1) with respect to θ_H, equating the resulting expression to zero, and solving for θ_H, one obtains

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

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 for θ_V from equation (A-2) into (A-3) results in a fifth-order algebraic equation whose roots can be obtained by numerical methods (refs. 10 and 11). Let the resulting roots be $\theta_{Hr}(i)$, $i = 1, \cdots, 5$. From equation (A-2), one then obtains the corresponding values for $\theta_{Vr}(i)$, $i = 1, \cdots, 5$.

Let

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

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

$$\overline{\delta}_1 = \begin{bmatrix} 0 \\ \\ 1 \end{bmatrix}, \ \overline{\delta}_2 = \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix}, \ \overline{\delta}_3 = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}, \ \overline{\delta}_4 = \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix} \tag{A-5}$$

Now the optimum value of $\overline{\theta}$, $\overline{\theta}_{opt}$ for $0 \leq \theta_H \leq 1$ and $0 \leq \theta_V \leq 1$, which maximizes equation (A-1), can be obtained from the flow diagram given in A-1. The above analysis can be generalized for obtaining the parameters of transition probabilities which have different parameters for more than two directions.

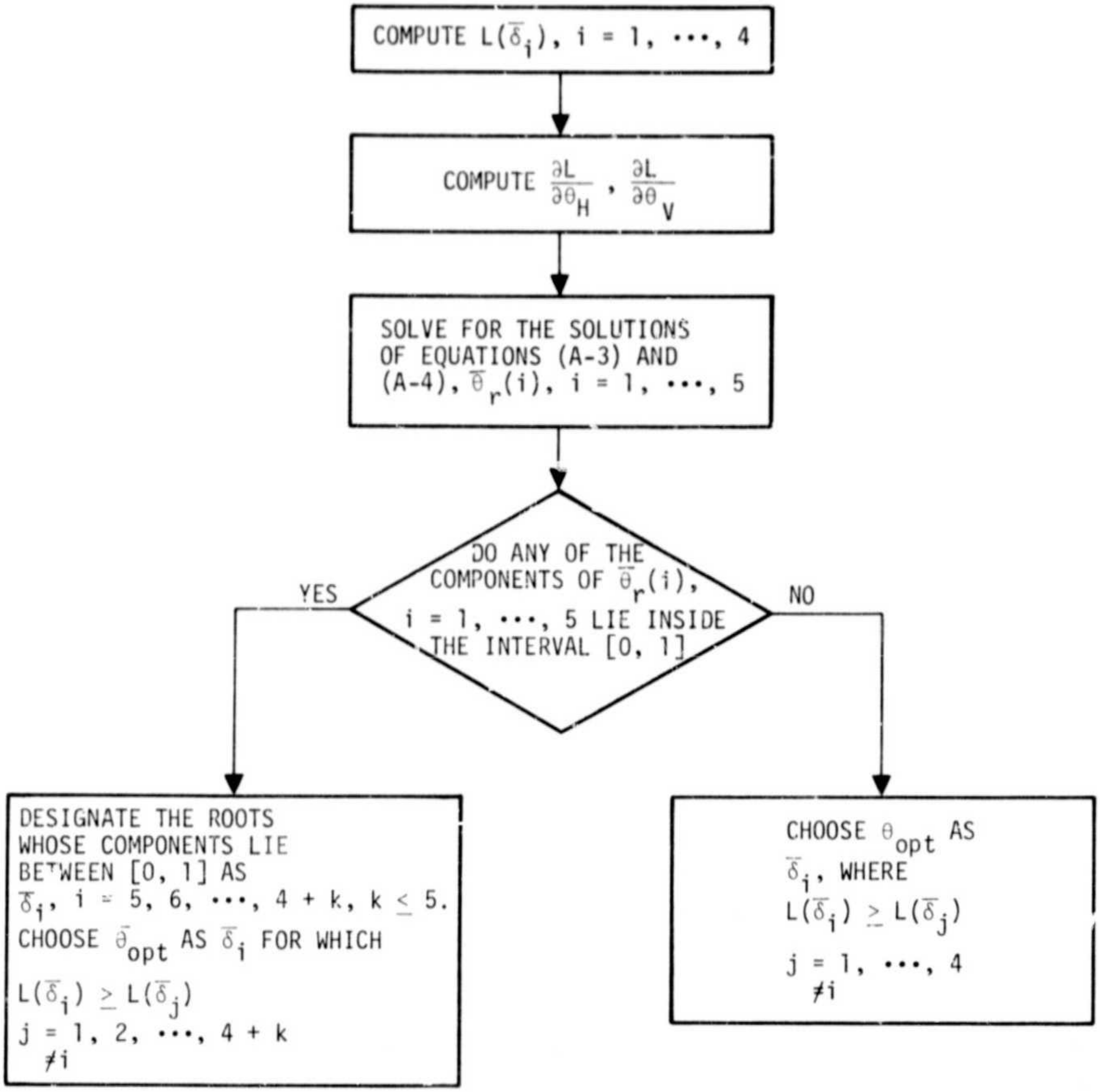

Figure A-1.— Procedure for finding $\overline{\theta}_{opt}$, which gives a global maximum of $L(\theta)$.

APPENDIX B

MULTITEMPORAL INTERPRETATION OF CONTEXT

This appendix gives a multitemporal interpretation of the theory developed in the paper for applications in remote sensing. 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. These passes are registered, and classification is performed on the registered data. It is assumed that there are r acquisitions. For the pixel under consideration, in each acquisition a data vector X_i, $i = 1, 2, \cdots, r$ is acquired. Suppose that the acquisitions $2, \cdots, r$ are registered with respect to acquisition 1. In registration, errors are encountered. Let the classifier be trained on the data from these individual acquisitions, obtaining the probability density functions $p(X|\omega = i)$, $i = 1, 2, \cdots, M$. The following paragraphs discuss the application of the theory developed in the paper in obtaining the label of the pixel under consideration using data X_i, $i = 1, 2, \cdots, r$ and by minimizing the effect of registration errors and incorporating the context. The pixel is classified using the decision rule: Classify it class $\omega = j$ if

$$p(\omega = j|X_i, \cdots, X_r) \geq p(\omega = i|X_1, \cdots, X_r) \qquad \text{(B-1)}$$

$$i = 1, 2, \cdots, M$$

$$\neq j$$

The registration errors are assumed to be modeled through the model for transition probabilities given in equations (3) and (4). Since $p(X_1, \cdots, X_r)$ is independent of i, equation (B-1) is equivalent to classifying the pixel as $\omega = j$ if

$$P(\omega = j)p(X_1, \cdots, X_r|\omega = j) \geq P(\omega = i)p(X_1, \cdots, X_r|\omega = i) \qquad \text{(B-2)}$$

$$i = 1, 2, \cdots, M$$

$$\neq j$$

From the theory developed in the paper, one has

$$
\begin{aligned}
p(X_1, \cdots, X_r|\omega = j) &= \sum_{i_1=1}^{M} \cdots \sum_{i_r=1}^{M} p(X_1, \omega_1 = i_1, \cdots X_r, \omega_r = i_r|\omega = j) \\
&= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \cdots \sum_{i_r=1}^{M} p(X_1, \cdots, X_r|\omega_1 = i_1, \cdots, \omega_r = i_r, \omega = j)P(\omega_1 = i_1, \omega_2 = i_2, \cdots, \omega_r = i_r|\omega = j) \\
&= \sum_{i_1=1}^{M} \cdots \sum_{i_r=1}^{M} \left[\prod_{j=1}^{r} p(X_j|\omega_j = i_j)\right] p(\omega_1 = i_1|\omega = j) \prod_{j=2}^{r} P(\omega_j = i_j|\omega_1 = i_1) \\
&= \sum_{i_1=1}^{M} p(X_j|\omega_1 = i_1)P(\omega_1 = i_1) \prod_{j=2}^{r} \left[\sum_{i_j=1}^{M} p(X_j|\omega_j = i_j)P(\omega_j = i_j|\omega_1 = i_1)\right] \\
&= \sum_{i_1=1}^{M} p(X_j|\omega_1 = i_1)P(\omega_1 = i_1) \prod_{j=2}^{r} \left[(1 - \theta) \sum_{i_j=1}^{M} p(X_j = i_j|\omega_j = i_j) + \theta P(X_j|\omega_j = i_1)\right]
\end{aligned} \tag{B-3}
$$

The following assumptions are made in the derivation of equation (B-3). The density function of X_j given its label identification is independent of any other information. The $\omega_1 = i_1$, the class of X_1, does not depend on the label of combined data, $X_1, \cdots, X_r$. Because the acquisitions 2, $\cdots$, r are assumed to be registered with respect to the first acquisition, the transition probabilities are assumed to obey

$$P(\omega_j = i_j|\omega_1 = i_1, \text{any other } \omega) = P(\omega_j = i_j|\omega_1 = i_1)$$

Using arguments similar to the ones in sections 4.1 and 4.2, one can write the likelihood function of $X_1, X_2, \cdots, X_r$ given θ as

$$
\begin{aligned}
\frac{p(X_1, X_2, \cdots, X_r|\theta)}{\prod_{j=1}^{r} p(X_j)} &= \frac{1}{\prod_{j=1}^{r} p(X_j)} \left[\sum_{i_1=1}^{M} \cdots \sum_{i_r=1}^{M} p(X_1, \omega_1 = i_1, \cdots, X_r, \omega_r = i_r|\theta)\right] \\
&= \frac{1}{\prod_{j=1}^{r} p(X_j)} \left[\sum_{i_1=1}^{M} \cdots \sum_{i_r=1}^{M} \left\{\prod_{j=1}^{r} p(X_j|\omega_j = i_j)\right\} P(\omega_2 = i_2, \cdots, \omega_r = i_r|\omega_1 = i_1, \theta)P(\omega_1 = i_1|\theta)\right] \\
&= \sum_{i_1=1}^{M} p(\omega_1 = i_1|X_1) \left\{\prod_{j=2}^{r} \left[(1 - \theta) + \frac{p(\omega_j = i_1|X_j)}{P(\omega_j = i_1)} \theta\right]\right\}
\end{aligned} \tag{B-4}
$$

The θ can be obtained by maximizing equation (B-4); it is used in equation (B-3) in obtaining the label $\omega = j$ of $X_1, \cdots, X_r$. It is to be noted that this multitemporal interpretation can easily be coupled with the contextual classification techniques developed in the paper.

APPENDIX C

REFERENCES

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

2. 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.

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

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

5. 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.

6. Chittineni, C. B.: Research Plan for Developing and Evaluating Classifiers. LEC-13300 (JSC-14849), Apr. 1979.

7. Duda, R. O.; and Hart, P. E.: Pattern Classification and Scene Analysis. John Wiley, New York, 1973.

8. Burnside, W. S.; and Panton, A. W.: The Theory of Equations. S. Chand and Co., New Delhi, 1967.

9. Cooper, L.; and Steinberg, D.: Introduction to Methods of Optimization. W. B. Saunders Company, Philadelphia, 1970, pp. 87-133.

10. Kelly, L. G.: Handbook of Numerical Methods and Its Applications. Addison-Wesley Publishing Company, Inc., 1967, pp. 356-359.

11. Bingham, J. A. C.: An Improvement to Iterative Methods of Polynomial Factorization. Commun. ACM, vol. 10, Jan. 1967, pp. 57-60.