A Context Classifier

ROBERT M. HARALICK, FELLOW, IEEE, AND HYONAM JOO, STUDENT MEMBER, IEEE

Abstract—**All other things being equal, context classifiers have a higher identification accuracy than pixel independent classifiers. This paper discusses a classifier whose context for each pixel is the best row monotonically increasing path including the given pixel. Use of such a context results in a two pass algorithm in which both the top down and bottom up pass require only two rows of data and whose computational complexity per pixel is constant independent of the size of image or the length of the best context path including the given pixel.**

I. Introduction

A CONTEXT CLASSIFIER is characterized by the fact that it classifies an unknown pixel using the entire context of the image or a substantially sized context neighboring the pixel. Basically, the effect of context is that a pixel's most probable interpretation when viewed in isolation changes when viewed in some context. One might expect the classification accuracy to be higher if an unknown pixel is classified using context rather than when it is classified using only the measurement made on that pixel without context. This is true in most cases. For example, a single pixel is not likely to be classified as water if it is surrounded by pixels classified as ground in a remotely sensed data set. The classification result of the conventional context free classifier leaves many isolated pixels and many small groups of pixels not connected with the blob to which they belong. Thus, in the last few years there has been a trend to increase the use of context in classification.

The use of context in pixel labeling can be found in many papers. The dominant contextual technique has been one of cooperative processing of neighboring pixels by a relaxation technique. Toussaint [7] presented a tutorial survey of techniques for using contextual information in pattern recognition emphasizing the problem of text recognition, and Haralick [4] gave a survey of decision making in context. Tilton [6] and Swain [5] use a p-context array which contains spatial information of $(p\text{-}1)$ pixels surrounding and neighboring the current pixel in the context pixel classification process. They derived the optimal decision rule using the context array and focus their attention on finding an unbiased estimate of the context function which is a statistical characterization of the context to be used in the decision rule. Yu and Fu [9] also noted that the spectral information of the surrounding pixels is correlated with the center pixel being considered. They investigated the spatial correlation between pixels and developed a spatial stochastic recursive contextual classification method. Wharton [8] presented a contextual analysis procedure based on the local frequency distribution of scene components and showed a two-stage contextual classifier.

In this paper, we present a theory and an algorithm of a new context classifier that assigns each pixel the highest probability label given some substantially sized context involving the pixel. The algorithm takes the form of a recursive neighborhood operator and turns out to be a two-pass algorithm first applied in a top down scan of the image and then in a bottom up scan of the image.

Section II describes the new context classifier algorithm. Section III gives the results of the new classifier when applied on the simulated images and the real images. Conclusions are given in Section IV.

II. Context Classifier

The most desirable kind of labeling process would give each pixel the highest probability label given the entire context of the image. The next most desirable kind of labeling process, and the one developed in this paper, gives each pixel the highest probability label given some substantially sized context neighboring the pixel. We show that such an algorithm takes the form of a recursive neighborhood operator first applied in a top down scan of the image and then in a bottom up scan of the image. The algorithm itself is related to a forward dynamic programming algorithm put in a two-dimensional mesh setting.

To explain the meaning of what the algorithm produces, select any pixel in the image. Now consider all the row monotonically increasing paths that begin at any border pixel of the image above the selected pixel, go through the selected pixel, and end at some border pixel of the image below the selected pixel. Each such path represents a context for the pixel. Corresponding to each path and the observed pixel data on the path, there is an associated highest probability label for the given pixel. Among all the paths there is some best path whose associated highest probability label is higher than the highest probability label of every other path. In two scans of the image, the context algorithm described in this paper is able to assign to each pixel of the image the highest probability label coming from its best path. Then, for the context used by the algorithm, the percentage of misclassified pixels is minimized.

The theory for the algorithm requires two distinct ideas. The first idea produces a decomposition for the problem.

Manuscript received March 6, 1986; revised June 18, 1986.

R. M. Haralick was with Machine Vision International, Ann Arbor, MI 48104. He is now with the Department of Electrical Engineering, University of Washington, Seattle, WA 98195.

H. Joo is with Machine Vision International, Ann Arbor, MI 48104.

IEEE Log Number 8610371.

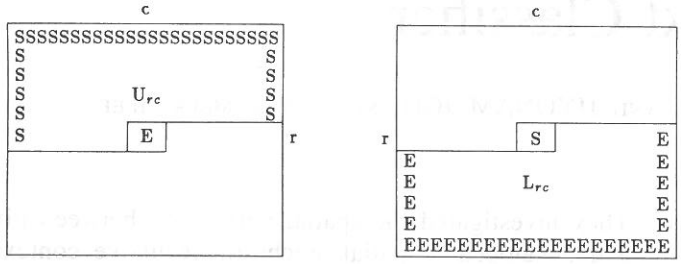

Fig. 1. The set U_{rc} and L_{rc}: U_{rc} is the set of all row monotonically increasing paths beginning on a border of the image above or to the left of the pixel (r, c) and terminating at the pixel (r, c) and not containing any pixel on row r beyond column c. L_{rc} is the set of all row monotonically increasing paths begining at the pixel (r, c) and terminating on the border of the image below or to the right of the pixel (r, c). S and E designate the possible starting and ending pixels of a path; respectively.

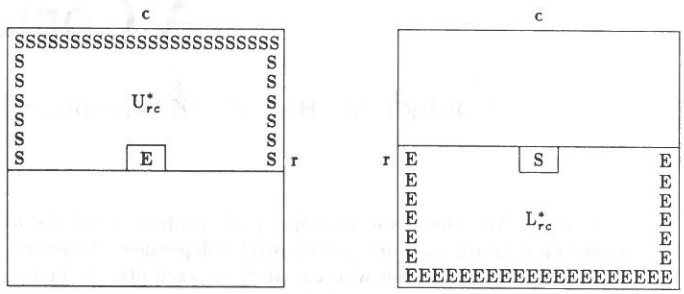

Fig. 2. The set U^*_{rc} and L^*_{rc}: U^*_{rc} is the set of all row monotonically increasing paths beginning on a border of the image at the same row or above pixel (r, c) and terminating at pixel (r, c). L^*_{rc} is the set of all row monotonically increasing paths beginning at the pixel (r, c) and terminating at some border pixel at the same row or below pixel (r, c). S and E designate the possible starting and ending pixels of a path, respectively.

Finding the highest probability label given the best path passing through the pixel can be accomplished by finding two probabilities, the probability for each possible label given the best path beginning above the pixel and terminating at the pixel and the probability for each possible label given the best path beginning at the pixel and terminating below the pixel. Finding these probabilities is what the algorithm accomplishes in the top down scan and the bottom up scan. The decomposition tells how to combine these probabilities to determine the highest probability label given the context of the best path through the pixel. The second idea produces a recursive decomposition which tells how to determine the conditional probability for each label given the data on the pixel's best upper (or lower) path. The decomposition bears a definite similarity to the one used in forward dynamic programming and as well bears some similarity to the iteration technique employed in some relaxation methods.

A. Notational Conventions

We now present basic concepts and notation conventions in order that we have a precise framework within which we can describe the algorithm.

1) A *path* means any connected sequence of pixels, each pixel neighboring its successor, in which the path does not intersect itself. A *row monotonically increasing* path is a path in which each successor pixel is on the same row or one row below its predecessor.

2) The set U_{rc} designates the set of all row monotonically increasing paths that begin at some border pixel of the image above or to the left of pixel (r, c) and terminate at pixel (r, c). The set U_{rc} is not permitted to contain any pixels on row r beyond column c.

3) The set L_{rc} designates the set of all row monotonically increasing paths that begin at (r, c) and terminate at some border pixel below or to the right of pixel (r, c). The set L_{rc} is not permitted to contain any pixels on row r before column c. These are illustrated in Fig. 1.

4) The set Z_{rc} designates the set of all row monotonically increasing paths beginning from a border of the image passing through pixel (r, c) and continuing to another border pixel of the image. Thus, Z_{rc} is just the join of all paths in U_{rc} with the paths in L_{rc}.

5) The set U^*_{rc} designates the set of all row monotonically increasing paths that begin at some border pixel of the image above or to the left of pixel (r, c) and terminate at pixel (r, c) coming to (r, c) from the left, from the right, or any neighboring pixel above (r, c). The difference between the set U^*_{rc} and the set U_{rc} is that the set U_{rc} does not include paths which contain any pixels on the same row r beyond column c while the set U^*_{rc} does.

6) The set L^*_{rc} designates the set of all row monotonically increasing paths that begin at pixel (r, c) and terminate at some border pixel at the same row or below pixel (r, c). While the set L_{rc} does not include paths which contain any pixels on the same row r before column c, the set L^*_{rc} does include those paths. This is illustrated in Fig. 2.

Let X_{ij} designate the measurements of the pixel (i, j). Let Q be any path. $P(X_{ij};\ (i, j) \in Q)$ designates the joint probability of all the measurements taken from the pixels on the path Q. For pixel (r, c), let e_{rc} be the correct but unknown label at pixel (r, c). Then $P(e_{rc}|\, X_{ij};\ (i, j) \in Q)$ designates the conditional probability that the pixel (r, c) takes the label e_{rc} given that the joint measurements made for the pixels on the path Q are $\{X_{ij};\ (i, j) \in Q\}$.

We define $f_{Z_{rc}}(e_{rc})$ to be the probability that pixel (r, c) takes label e_{rc} given that the joint measurements on the best row monotonically increasing path Q is $\{X_{ij};\ (i, j) \in Q\}$. Thus

$$f_{Z_{rc}}(e_{rc}) = \max_{Q \in Z_{rc}} P(e_{rc}|\, X_{ij};\ (i, j) \in Q). \tag{1}$$

Just as $f_{Z_{rc}}(e_{rc})$ designates the probability of label e_{rc} given the joint measurements $\{X_{ij};\ (i, j) \in Q\}$ arising from the best row monotonically increasing path Q in Z_{rc}, let $g_{U_{rc}}(e_{rc})$ designate the probability of label e_{rc} given joint measurements $\{X_{ij};\ (i, j) \in Q\}$ arising from the best row monotonically increasing path in U_{rc} and let $g_{L_{rc}}(e_{rc})$ designate the probability of label e_{rc} given joint measurements $\{X_{ij};\ (i, j) \in Q\}$ arising from the best row monotonically increasing path in L_{rc}.

B. The Top-Bottom Decomposition

In this section we demonstrate the conditions under which $f_{Z_{rc}}(e_{rc})$ has the decomposition

$$f_{Z_{rc}}(e_{rc}) = \frac{1}{\Delta'}\left[\frac{1}{P(X_{rc}|e_{rc})}\, g_{U_{rc}}(e_{rc})\, g_{L_{rc}}(e_{rc})\right] \quad (2)$$

where Δ' is a normalizing constant that will be discussed later.

This decomposition requires some assumptions on the conditional probability of the labels given the measurements and the joint probability of the labels. By Bayes' formula

$$\begin{aligned} P(e_{ij};\,(i,j)\in Q|X_{ij};\,(i,j)\in Q) &= P(X_{ij};\,(i,j)\in Q|e_{ij};\,(i,j)\in Q) \\ &\quad \times \frac{P(e_{ij};\,(i,j)\in Q)}{P(X_{ij};\,(i,j)\in Q)}. \end{aligned}$$

Assuming that the measurements at each pixel depend only on the true label at that pixel and the measurement noise for one pixel does not influence the measurement noise for another pixel, we have

$$P(X_{ij};\,(i,j)\in Q|e_{ij};\,(i,j)\in Q) = \prod_{(i,j)\in Q} P(X_{ij}|e_{ij}). \quad (3)$$

This assumption is really an idealization. It corresponds to a reality in which independent randomness is superimposed on the ideal signal. The degree to which texture dominates is the degree to which this idealization does not hold.

The probability $P(e_{ij};\,(i,j)\in Q)$ is the joint prior probability of having the true labels for each pixel (i,j) on the path Q be e_{ij}. This probability encodes all the information we have about context. If for example, we had independence

$$P(e_{ij};\,(i,j)\in Q) = \prod_{(i,j)\in Q} P(e_{ij})$$

we would discover that the highest probability assignment we could make using the context is precisely the highest probability assignment we could make using only the local pixel information.

The simplest assumption of higher order independence is a Markov-like assumption in which the joint prior probability becomes a function expressible as the product of functions whose arguments are the label pairs for successive pixels in the path Q. Letting $R(Q)$ designate the set of all pairs of succesive pixels in the path Q, we have

$$P(e_{ij};\,(i,j)\in Q) = \prod_{((i,j),\,(k,l))\in R(Q)} A(e_{ij}, e_{kl}). \quad (4)$$

Using the two assumptions described above, we will decompose $f_{Z_{rc}}(e_{rc})$ into two components as suggested by (2).

$$\begin{aligned} f_{Z_{rc}}(e_{rc}) &= \max_{Q\in Z_{rc}} P(e_{rc}|X_{ij};\,(i,j)\in Q) \\ &= \max_{Q\in Z_{rc}} \sum_{\substack{e_{ij} \\ (i,j)\in Q^-}} P(e_{ij}|X_{ij};\,(i,j)\in Q) \end{aligned} \quad (5)$$

where Q^- designates the set of all pixels in path Q except the pixel (r,c) and the summation over all e_{ij} where $(i,j) \in Q^-$ designates an iterated summation, one sum for each pixel $(i,j)\in Q^-$, the sum taken over all possible values for the label the pixel (i,j) can take.

Substituting (3) and (4) into (5), we have

$$f_{Z_{rc}}(e_{rc}) = \max_{Q\in Z_{rc}} \frac{1}{\Delta}\left[\sum_{\substack{e_{ij} \\ (i,j)\in Q^-}} \prod_{(i,j)\in Q} P(X_{ij}|e_{ij}) \prod_{((i,j),\,(k,l))\in R(Q)} A(e_{ij}, e_{kl})\right] \quad (6)$$

where

$$\begin{aligned} \Delta &= P(X_{ij};\,(i,j)\in Q) \\ &= \sum_{\substack{e_{ij} \\ (i,j)\in Q}} P(e_{ij}, X_{ij};\,(i,j)\in Q) \\ &= \sum_{\substack{e_{ij} \\ (i,j)\in Q}} \prod_{(i,j)\in Q} P(X_{ij}|e_{ij}) \prod_{((i,j),\,(k,l))\in R(Q)} A(e_{ij}, e_{kl}). \end{aligned}$$

It is obvious that Δ is a normalizing constant that makes the sum of the terms inside the bracket of (6) divided by Δ over all e_{rc} to be one.

Since Z_{rc} can be decomposed as the join of U_{rc} and L_{rc}, there results

$$\begin{aligned} f_{Z_{rc}}(e_{rc}) = \max_{Q_1\in U_{rc},\, Q_2\in L_{rc}} \frac{1}{\Delta}\Bigg[& P(X_{rc}|e_{rc}) \sum_{\substack{e_{ij} \\ (i,j)\in Q_1^-}} \sum_{\substack{e_{ij} \\ (i,j)\in Q_2^-}} \prod_{(i,j)\in Q_1} \\ & \cdot P(X_{ij}|e_{ij}) \prod_{(i,j)\in Q_2} P(X_{ij}|e_{ij}) \prod_{((i,j),\,(k,l))\in R(Q_1)} \\ & \cdot A(e_{ij}, e_{kl}) \prod_{((i,j),\,(k,l))\in R(Q_2)} A(e_{ij}, e_{kl})\Bigg]. \end{aligned}$$

Rearranging the above equation, we can group all expressions involving Q_1 together and all expressions involving Q_2 together, and we obtain

$$\begin{aligned} f_{Z_{rc}}(e_{rc}) &= \frac{1}{\Delta'}\Bigg[\frac{1}{P(X_{rc}|e_{rc})} \times \max_{Q_1\in U_{rc}} \Bigg\{\frac{1}{\Delta_1} \sum_{\substack{e_{ij} \\ (i,j)\in Q_1^-}} \prod_{(i,j)\in Q_1} \\ &\quad \cdot P(X_{ij}|e_{ij}) \prod_{((i,j),\,(k,j))\in R(Q_1)} A(e_{ij}, e_{kl})\Bigg\} \\ &\quad \times \max_{Q_2\in L_{rc}} \Bigg\{\frac{1}{\Delta_2} \sum_{\substack{e_{ij} \\ (i,j)\in Q_2^-}} \prod_{(i,j)\in Q_2} P(X_{ij}|e_{ij}) \\ &\quad \cdot \prod_{((i,j),\,(k,j))\in R(Q_2)} A(e_{ij}, e_{kl})\Bigg\}\Bigg] \\ &= \frac{1}{\Delta'}\left[\frac{1}{P(X_{rc}|e_{rc})}\, g_{U_{rc}}(e_{rc})\, g_{L_{rc}}(e_{rc})\right] \end{aligned}$$

where

$$\Delta = \Delta' \Delta_1 \Delta_2$$

$$\Delta_m = \sum_{\substack{e_{ij} \\ (i,j)\in Q_m}} \prod_{(i,j)\in Q_m} P(X_{ij}|e_{ij}) \prod_{((i,j),(k,l))\in R(Q_m)} A(e_{ij}, e_{kl}),$$

for $m = 1, 2$.

Thus if

$$g_{U_{rc}}(e_{rc}) = \max_{Q_1 \in U_{rc}} \left\{ \frac{1}{\Delta_1} \sum_{\substack{e_{ij} \\ (i,j)\in Q_1^-}} \prod_{(i,j)\in Q_1} P(X_{ij}|e_{ij}) \prod_{((i,j),(k,j))\in R(Q_1)} A(e_{ij}, e_{kl}) \right\}$$

and if

$$g_{L_{rc}}(e_{rc}) = \max_{Q_2 \in L_{rc}} \left\{ \frac{1}{\Delta_2} \sum_{\substack{e_{ij} \\ (i,j)\in Q_2^-}} \prod_{(i,j)\in Q_2} P(X_{ij}|e_{ij}) \prod_{((i,j),(k,j))\in R(Q_2)} A(e_{ij}, e_{kl}) \right\}$$

we have the decomposition suggested in (2).

C. The Recursive Decomposition

In this section we give a recursive decomposition of $g_{U_{rc}}$ in terms of the neighboring $g_{U_{rc-1}}$ and $h_{U^*_{r-1c-1}}$, $h_{U^*_{r-1c}}$ and $h_{U^*_{r-1c+1}}$ where the h_{U^*} function will be defined in the next paragraph. The recursive decomposition of $g_{L_{rc}}$ is similar. By definition

$$\begin{aligned} g_{U_{rc}}(e_{rc}) &= \max_{Q \in U_{rc}} P(e_{rc}|X_{ij};\ (i,j) \in Q) \\ &= \max_{Q \in U_{rc}} \sum_{\substack{e_{ij} \\ (i,j)\in Q^-}} P(e_{ij};\ (i,j) \in Q | X_{ij};\ (i,j) \in Q) \\ &= \max_{Q \in U_{rc}} \frac{1}{\Delta_1} \sum_{\substack{e_{ij} \\ (i,j)\in Q^-}} \prod_{(i,j)\in Q} P(X_{ij}|e_{ij}) \prod_{((i,j),(k,l))\in R(Q)} \\ &\qquad \cdot A(e_{ij}, e_{kl}). \end{aligned} \tag{7}$$

Just as $g_{U_{rc}}$ designates the probability of label e_{rc} given measurements $\{X_{ij};\ (i,j) \in Q\}$ arising from the best row monotonically increasing path in U_{rc}, let $h_{U^*_{rc}}$ designate the probability of label e_{rc} given measurements $\{X_{ij};\ (i,j) \in Q\}$ arising from the best row monotonically increasing path in U^*_{rc}.

Then, by definition

$$\begin{aligned} h_{U^*_{rc}}(e_{rc}) &= \max_{Q \in U^*_{rc}} P(e_{rc}|X_{ij};\ (i,j) \in Q) \\ &= \max_{Q \in U^*_{rc}} \sum_{\substack{e_{ij} \\ (i,j)\in Q^-}} P(e_{ij};\ (i,j) \in Q | X_{ij};\ (i,j) \in Q) \\ &= \max_{Q \in U^*_{rc}} \frac{1}{\Delta_1^*} \sum_{\substack{e_{ij} \\ (i,j)\in Q^-}} \prod_{(i,j)\in Q} P(X_{ij}|e_{ij}) \prod_{((i,j),(k,l))\in R(Q)} \\ &\qquad \cdot A(e_{ij}, e_{kl}). \end{aligned} \tag{8}$$

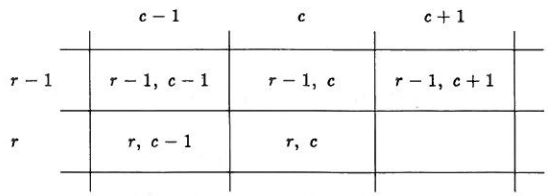

Fig. 3. The decomposition for $g_{U_{rc}}$: The best path to (r, c) must have come from $(r, c-1)$, $(r-1, c-1)$, $(r-1, c)$, or $(r-1, c+1)$ where (r, c) means the location of the pixel in the image at row r and column c.

where Δ_1^* is computed similarly as Δ_1.

To do the decomposition for $g_{U_{rc}}$ we need to recognize that whatever the best path is, the best path to (r, c) must have come from $(r, c-1)$, $(r-1, c-1)$, $(r-1, c)$, or $(r-1, c+1)$. Because the best path cannot cross itself, if the best path came from $(r, c-1)$, then the path must be in U_{rc-1}. However, there is no damage of the path crossing itself if the best path comes from $(r-1, c-1)$. Hence, such a path must be in U^*_{r-1c-1}. Likewise, a best path coming from $(r-1, c)$ must be in U^*_{r-1c} and a best path coming from $(r-1, c+1)$ must be in U^*_{r-1c+1}. This is illustrated in Fig. 3.

Using these ideas, we can rewrite (7) as

$$\begin{aligned} g_{U_{rc}}(e_{rc}) = \max_{Q_i} \Bigg[& \max_{Q_1 \in U_{rc-1}} \sum_{e_{rc-1}} W_1(Q_1)\, A(e_{rc-1}, e_{rc}), \\ & \max_{Q_2 \in U^*_{r-1c-1}} \sum_{e_{r-1c-1}} W_2(Q_2)\, A(e_{r-1c-1}, e_{rc}), \\ & \max_{Q_3 \in U^*_{r-1c}} \sum_{e_{r-1c}} W_3(Q_3)\, A(e_{r-1c}, e_{rc}), \\ & \max_{Q_4 \in U^*_{r-1c+1}} \sum_{e_{r-1c+1}} W_4(Q_4)\, A(e_{r-1c+1}, e_{rc}) \Bigg], \end{aligned} \tag{9}$$

where, for $m = 1, \cdots, 4$

$$W_m(Q_m) = \frac{P(X_{rc}|e_{rc})}{\Delta_m} \sum_{\substack{e_{ij} \\ (i,j)\in Q_m^-}} \prod_{(i,j)\in Q_m} P(X_{ij}|e_{ij}) \prod_{((i,j),(k,l))\in R(Q_m)} A(e_{ij}, e_{kl}).$$

Examining the first term in the maximization of (9) and comparing it to $g_{U_{rc-1}}(e_{rc-1})$ as defined by (7), we discover that the expressions are almost identical. The only difference is that the expression for $g_{U_{rc-1}}(e_{rc-1})$ involves a summation iterated over all $(i,j) \in Q^-$ while the first expression in the maximization involves a summation iterated over all $(i,j) \in Q$. Let M be the number of possible labels for a pixel at any position (i,j). Then at (r, c), we have M best paths one for each possible label and a corresponding maximum conditional probability. Let Q^* be the best path for label e^*_{rc} where $g_{U_{rc}}(e^*_{rc})$ is maximum. For this path and this label e^*_{rc}, define a conditional probability for label e^*_{rc} by

$$g_{U_{rc}}(e_{rc}|e_{rc}^*) = P(e_{rc}|X_{ij};\ (i,j) \in Q^*).$$

Similarly, define the conditional probability $h_{U_{rc}^*}(e_{rc}|e_{rc}^*)$ of label e_{rc} at (r, c) given the best path Q^* by

$$h_{U_{rc}^*}(e_{rc}|e_{rc}^*) = P(e_{rc}|X_{ij};\ (i,j) \in Q^*).$$

Now, if the best path for label e_{rc} at (r, c) comes from $(r, c-1)$, it should come from one of those M best paths computed at pixel $(r, c-1)$. Therefore, the first term of (9) can be rewritten as

$$\max_{Q_1 \in U_{rc-1}} W_1(Q_1)A(e_{rc-1}, e_{rc})$$

$$= \max_{e_{rc-1}^*} \left[\frac{P(X_{rc}|e_{rc})}{\Delta_1} \times \sum_{e_{rc-1}} g_{U_{rc-1}}(e_{rc-1}|e_{rc-1}^*)A(e_{rc-1}, e_{rc}) \right].$$

A similar argument applies for the second, third, and fourth terms of the maximization. After comparing them with $h_{U_{r-1c-1}^*}$, $h_{U_{r-1c}^*}$, and $h_{U_{r-1c+1}^*}$ and using the same reasoning, this results in

$$g_{U_{rc}}(e_{rc}) = \max$$

$$\left[\max_{e_{rc-1}^*} \left\{ \frac{P(X_{rc}|e_{rc})}{\Delta_1} \sum_{e_{rc-1}} g_{U_{rc-1}}(e_{rc-1}|e_{rc-1}^*)A(e_{rc-1}, e_{rc}) \right\} \right.$$

$$\max_{e_{r-1c-1}^*} \left\{ \frac{P(X_{rc}|e_{rc})}{\Delta_2} \sum_{e_{r-1c-1}} h_{U_{r-1c-1}^*}(e_{r-1c-1}|e_{r-1c-1}^*) \cdot A(e_{r-1c-1}, e_{rc}) \right\}$$

$$\max_{e_{r-1c}^*} \left\{ \frac{P(X_{rc}|e_{rc})}{\Delta_3} \sum_{e_{r-1c}} h_{U_{r-1c}^*}(e_{r-1c}|e_{r-1c}^*)A(e_{r-1c}, e_{rc}) \right\}$$

$$\left. \max_{e_{r-1c+1}^*} \left\{ \frac{P(X_{rc}|e_{rc})}{\Delta_4} \sum_{e_{r-1c+1}} h_{U_{r-1c+1}^*}(e_{r-1c+1}|e_{r-1c+1}^*) \cdot A(e_{r-1c+1}, e_{rc}) \right\} \right]. \quad (10)$$

Equation (10) says that for each label value e_{rc}, the conditional probability of e_{rc} given $\{X_{ij};\ (i,j) \in Q$, the path giving the highest conditional probability$\}$ can be obtained on the basis of the just previously computed $g_{U_{rc-1}}$ and the previously computed $h_{U_{r-1c-1}^*}$, $h_{U_{r-1c}^*}$, $h_{U_{r-1c+1}^*}$ coming from the row above the current row r (see Fig. 4).

So providing we can demonstrate a way to compute $h_{U_{rc}^*}$, (10) specifies a recursive neighborhood operator that scans the image in a top-down left-right scan to produce $g_{U_{rc}}(e_{rc})$ for each pixel (r, c) and for each label e_{rc}.

Fortunately, we are able to provide an algorithm for computing $h_{U_{rc}^*}$. Its development proceeds along similar lines to that of $g_{U_{rc}}$. For a path from U_{rc}^* to such (r, c), it must first have gone through $(r, c-1)$, $(r, c+1)$, $(r-1, c-1)$, $(r-1, c)$, or $(r-1, c+1)$. Furthermore, if it went through $(r, c-1)$, since the path cannot cross itself, it must be a path in U_{rc-1}. Thus we have

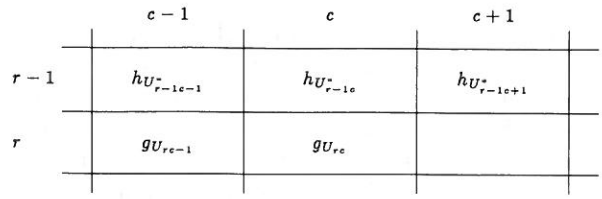

Fig. 4. The computation of $g_{U_{rc}}$: Using h_{U^*} values computed in the row above the current row, $g_{U_{rc}}$ can be computed recursively from $h_{U^*_{r-1c-1}}$, $h_{U^*_{r-1c}}$, $h_{U^*_{r-1c+1}}$, and the just previously computed $g_{U_{rc-1}}$ in the left to right scan of the image.

$$h_{U_{rc}^*}(e_{rc}) = \max \left[g_{U_{rc}}(e_{rc}), \max_{Q_2 \in U_{rc+1}^*} \sum_{e_{rc+1}} W_2(Q_2)A(e_{rc+1}, e_{rc}) \right]. \quad (11)$$

In a similar manner as in the development of (10), we can rewrite (11) as

$$h_{U_{rc}^*}(e_{rc}) = \max \left[g_{U_{rc}}(e_{rc}), \max_{e_{rc+1}^*} \left\{ \frac{P(X_{rc}|e_{rc})}{\Delta_2} \sum_{e_{rc+1}} h_{U_{rc+1}^*}(e_{rc+1}|e_{rc+1}^*)A(e_{rc+1}, e_{rc}) \right\} \right]. \quad (12)$$

Equation (12) states that $h_{U_{rc}^*}$ can be recursively computed from $g_{U_{rc}}$ and the previous $h_{U_{rc+1}^*}$ in a right-left scan of a row done after $g_{U_{rc}}$ has been computed. To start the recursive calculation (12) we take $h_{U_{rc}^*}(e_{rc}) = g_{U_{rc}}(e_{rc})$ for that column position c that is the rightmost column position.

In summary, (10) and (12) give the following algorithm for the computation of $g_{U_{rc}}(e_{rc})$. From (10) we perform a top-down left-right scan of the image recursively computing $g_{U_{rc}}$ from the previous $g_{U_{rc-1}}$ and the $h_{U_{r-1c-1}^*}$, $h_{U_{r-1c}^*}$, and $h_{U_{r-1c+1}^*}$, which had been computed on the previous row. Following the computation of $g_{U_{rc}}$ for all pixels on row r, we perform a right-left scan of row r using (12) to compute $h_{U_{rc}^*}$. An obvious mirror image derivation applies to $g_{L_{rc}}$. To compute it, we perform a bottom-up right-left scan of the image recursively computing $g_{L_{rc}}$ from the previous $g_{L_{rc+1}}$ and the $h_{L_{r+1c-1}^*}$, $h_{L_{r+1c}^*}$, and $h_{L_{r+1c+1}^*}$, which had been computed on the previous row from the bottom-up scan. Following the computation of $g_{L_{rc}}$ for all pixels on row r, we perform a left-right scan of row r and compute $h_{L_{rc}^*}$.

As soon as $g_{L_{rc}}(e_{rc})$ has been computed, it can be combined with $g_{U_{rc}}(e_{rc})$ as given in (2) to compute $f_{Z_{rc}}(e_{rc})$. Then the label e_{rc}, which maximizes $f_{Z_{rc}}(e_{rc})$, can be determined and pixel (r, c) labeled with this label.

III. Experimental Results and Discussion

To show the increase in classification accuracy of the new context classifier, several experiments were performed on the images. First, two types of simulated images are used to examine the accuracy improvement of the new classifier as compared to the pixel independent Bayes

quadratic classifier, given that the class conditional mean vectors, covariance matrices, and class prior probabilities were known. Then it is applied to real images to investigate its performance in a more realistic case. We also present one experiment on a thermal image to compare the ability to detect an object in the image and two experiments on Landsat data to compare the classification accuracy.

In the derivation of the new context classifier algorithm, we used a Markov-like assumption and expressed $P(e_{ij};\ (i, j) \in Q)$ as the product of functions whose arguments are the label pairs for successive pixels in the path Q, (4). To compute the function $A(e_{ij}, e_{kl})$ for the testing purpose in this paper, we assumed a Gaussian stationary two-dimensional process. Here, stationary means that the correlation between neighboring pixels is position independent in the image. The function A is estimated for four directional pairs of pixels: horizontal, vertical, and two diagonals. We approximate A by the observed frequency distribution of each pair of labels in all four directions from the ground truth data.

The first simulated image is generated from a real Landsat image following the method used by Swain *et al.* [5]. From the sample image and the ground truth data, we first estimate the mean vectors μ_i and the covariance matrices Σ_i for each class i. Measurement vectors for each pixel on the simulated image are produced by a pseudo-random multi-dimensional Gaussian random number generator having the required means and covariances. Then a simulated image with the following characteristics is created.

1) Each pixel in the simulated image represents the same class as in the ground truth data.
2) Each class has a multi-dimensional Gaussian normal distribution having the mean and covariance matrix estimated from the sample image.
3) All pixels are class-conditionally independent of adjacent pixels.

The real image consists of a subimage collected by the Landsat multispectral scanner (MSS), which has known ground truth. This is a 151 × 151 subframe of a scence of Roanoke, VA, taken on April 13, 1976 (shown later in Fig. 13) that contains five ground cover classes

class 1: urban or built-up land
class 2: agricultural land
class 3: rangeland
class 4: forest land
class 5: water.

It has four spectral bands. The ground truth data and the first band of the simulated image are shown in Figs. 5 and 6, respectively. Due to the small sample size, ground cover classes 3 and 5 are not used in the simulated image.

The pixel-independent Bayes classifier and the context classifier are applied on the simulated image and the results are shown in Figs. 7 and 8. Contingency tables of the classified images are given in Table I. The overall classification accuracy is measured as the ratio of the

Fig. 5. Ground truth data of the simulated image 1.

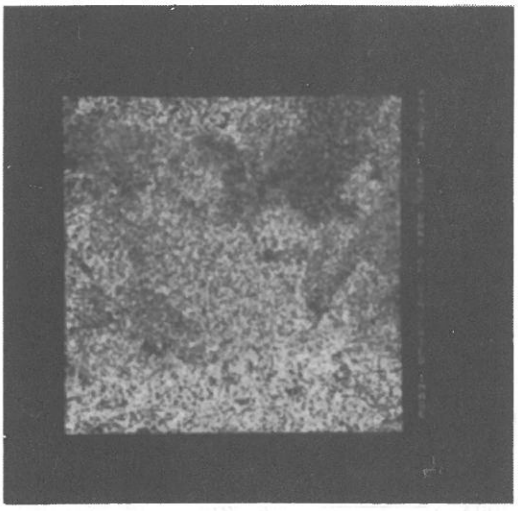

Fig. 6. First band of the simulated image 1.

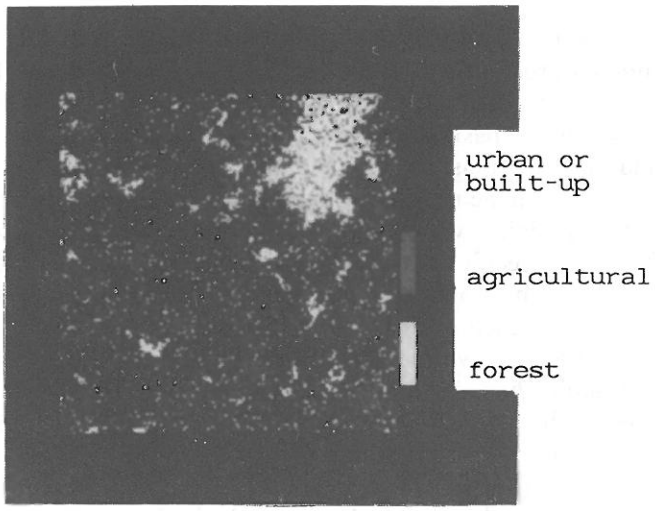

Fig. 7. Classified result of the simulated image 1-classified by the Bayes classifier.

Fig. 8. Classified result of the simulated image 1—classifed by the context classifier.

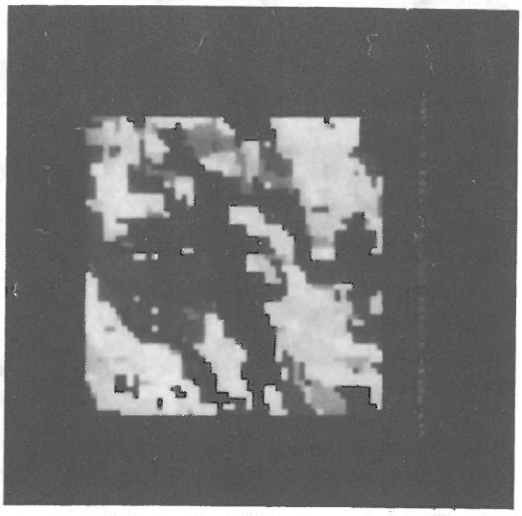

Fig. 9. Markov pseudo-random image as a ground truth data for the simulated image 2: Transition probability $p = 0.7$ and SNR = 16.

TABLE I
CONTINGENCY TABLES FOR CLASSIFICATION RESULTS OF SIMULATED IMAGE 1
(Column = assigned class, row = true class, URB = urban or built-up land, AGR = agricultural land, FST = forest land. Scale factor for the number of pixels = 10.)

(a) *Bayes classifier result*

class	URB	AGR	FST	total	Acc(%)*
URB	1315	76	46	1437	91.5
AGR	462	88	29	579	15.2
FST	90	9	167	266	62.8
total	1867	173	242	2282	68.8**

(b) *Context classifier result*

class	URB	AGR	FST	total	Acc(%)*
URB	1421	10	6	1437	98.9
AGR	400	163	16	579	28.2
FST	34	3	229	266	86.1
total	1855	176	251	2282	79.4**

* *percent of correct classification*

** *overall classification accuracy : ratio of the number of correctly classified pixels to the total number of classified pixels*

number of correctly classified pixels to the total number of classified pixels and Table I shows that the context classifier gained 10.6-percent increase in overall classification accuracy over the pixel-independent Bayes classifier.

The second type of simulated image is generated by extending the method suggested by Devijver [1] to the two-dimensional space. We use a simple two-dimensional Markov source model, which is a six-state model with transition probabilities parameterized by p, $0 \leq p \leq 1$, defined by

$$P(e_{ij} | e_{i-1j}, e_{ij-1}) = P(e_{ij} | e_{i-1j}) P(e_{ij} | e_{ij-1})$$

where

$$P(e_{ij} | e_{i-1j}) = \begin{cases} p, & \text{if } e_{ij} = e_{i-1j} \\ \dfrac{1-p}{5}, & \text{otherwise} \end{cases}$$

$$P(e_{ij} | e_{ij-1}) = \begin{cases} p, & \text{if } e_{ij} = e_{ij-1} \\ \dfrac{1-p}{5}, & \text{otherwise} \end{cases}$$

A pseudo-random Markov image which is used as a ground truth image for the simulated image is generated as follows:

1) The top left-most pixel in the image is assigned randomly with equal prior probability for each possible class.
2) The selection of class assignments for successive pixel positions is performed in a top to bottom and left to right scan of the image.
3) The class of each pixel at (i, j) is determined by the transition probabilities given the labels of the neighboring north $(i-1, j)$ and west $(i, j-1)$ pixels. The unit interval is partitioned proportional to the transition probability $P(e_{ij} | e_{i-1,j}, e_{i,j-1})$ for each class e_{ij}. A pseudo-random number between 0 and 1 is generated and the class is chosen according to the subinterval in which the pseudo-random number falls.

One of the simulated images with $p = 0.7$ generated by such a method is shown in Fig. 9.

The class conditional distributions of the measurements are chosen to be two-dimensional normal with covariance matrices being each the identity matrix. The mean vectors of the classes are located at the vertices of a regular hexagon inscribed within a circle of radius R. Ignoring context, the signal to noise ratio (SNR) is measured by the

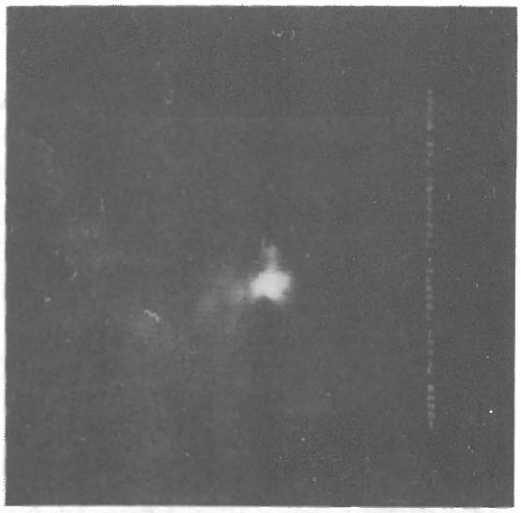

Fig. 10. Test image 1: 8–12-μm thermal image of size 200 × 200 taken at Grafenwoehr, Germany. This image contains one object (white) at the center.

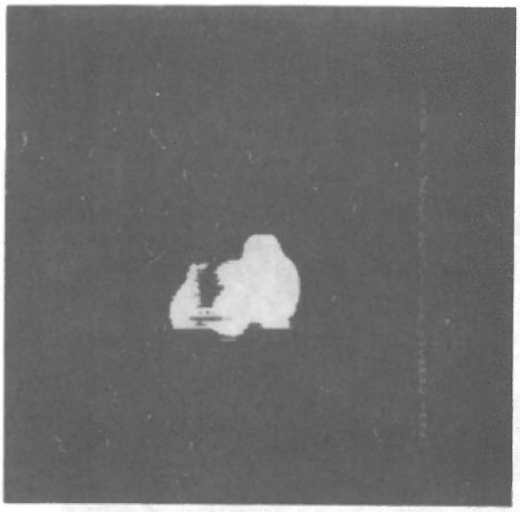

Fig. 11. Classified result of test image 1—classified by the Bayes classifier.

TABLE II
OVERALL CLASSIFICATION ACCURACY OF THE BAYES CLASSIFIER AND THE CONTEXT CLASSIFIER WHEN APPLIED TO THE SECOND TYPE OF SIMULATED IMAGES WITH DIFFERENT SNR AND TRANSITION PROBABILITY PARAMETER p.

	p = 0.2 SNR = 16	p = 0.4 SNR = 16	p = 0.4 SNR = 9	p = 0.7 SNR = 16
Bayes classifier	96.12	95.88	86.96	96.28
Context classifier	96.40	96.88	89.16	98.24

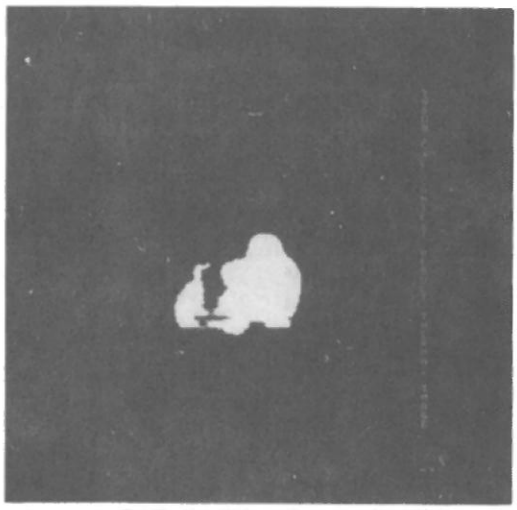

Fig. 12. Classified result of test image 1—classified by the context classifier.

ratio of the traces of the between class and within class covariance matrices. In our configuration SNR = R^2.

Four 50 × 50 images are generated and classified by the pixel independent Bayes quadratic classifier and the context classifier. The performance of the two classifiers are shown in Table II in terms of overall classification accuracy. We can clearly see the improved accuracy in the result of the context classifier.

Now we extend the test to two real data sets. The first is an 8–12-μm thermal image of size 200 × 200 taken at Grafenwoehr, Germany which contains one object in the center (see Fig. 10). For the classification of the image, the two texture features of entropy and inverse difference moment are measured from the image [3]. The original measurement in the thermal image and the two texture features compose the measurement vector. Training samples are selected randomly from the image and mean vectors and covariance matrices are estimated assuming that the underlying class conditional probability is multidimensional normal. Figs. 11 and 12 show the result image classified by the pixel-independent Bayes classifier and the context classifier, respectively. We can see the effect of the context classifier by noting the difference in the boundary lines of the classified object between Figs. 11 and 12. Compared to the smooth boundary lines in the result of the context classifier, the pixel-independent Bayes classifier leaves uneven boundary lines.

To give the numerical comparison of classification accuracy, two digital remote sensing data sets that have known ground truth data are classified by both the pixel-independent Bayes classifier and the context classifier. The first image is the one used to generate the first simulated image. It has four spectral bands and the first band of the image is shown in Fig. 13. The objective of the analysis was to discriminate three ground cover classes 1, 2, and 4. Classified results of both classifiers are shown in Figs. 14 and 15 and the contingency tables are given in Table III. The overall classification accuracy in Table III is again measured as the ratio of the number of correctly classified pixels to the total number of classified pixels. It shows that the context classifier gained 3.5-percent increase in overall classification accuracy over the pixel-independent Bayes classifier. Examining the result of the context classifier (Fig. 15), it can be seen that within each class it assigns the ambiguous noisy pixels to their most likely class, leaving almost homogeneous regions for each class. Also, it smoothed noisy boundaries between each class.

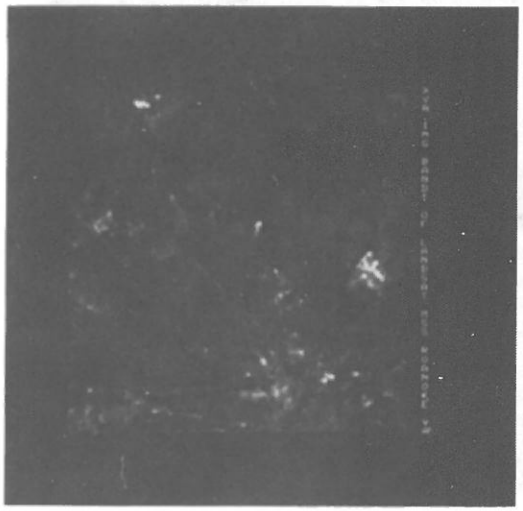

Fig. 13. Test image 2: 151 × 151 subframe of MSS scene of Roanoke, VA, taken on April 13, 1976 (first band). Longitude from 79°52′ to 80°00′ W; latitude from 37°15′ to 37°23′ N.

Fig. 14. Classified result of test image 2—classified by the Bayes classifier.

Fig. 15. Classified result of test image 2—classified by the context classifier.

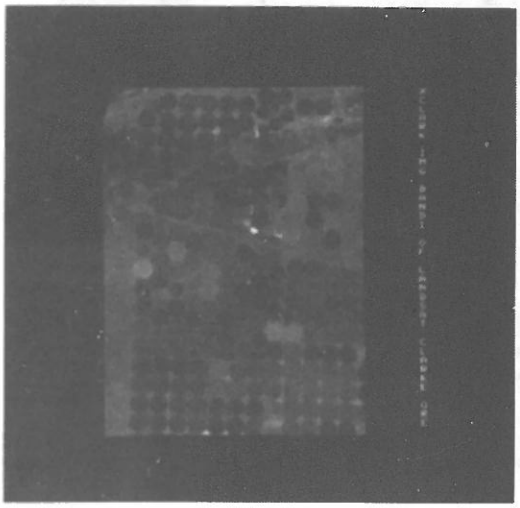

Fig. 16. Test image 3: First band of 200 × 150 subimage of LANDSAT image showing the Clarke, OR, 7.5--min U.S. Geological Survey quadrangle map taken in June 1979.

TABLE III
CONTINGENCY TABLES FOR CLASSIFICATION RESULTS OF TEST IMAGE 2
(Column = assigned class, row = true class, URB = urban or built-up land, AGR = agricultural land, FST = forest land. Scale factor for the number of pixels = 10.)

(a) *Bayes classifier result*

class	URB	AGR	FST	total	Acc(%)*
URB	1233	172	31	1436	85.9
AGR	371	168	39	578	29.1
FST	79	7	166	252	65.9
total	1683	347	236	2266	69.2**

(b) *Context classifier result*

class	URB	AGR	FST	total	Acc(%)*
URB	1222	192	23	1437	85.0
AGR	295	248	35	578	42.9
FST	67	7	178	252	70.6
total	1584	447	236	2267	72.7**

* *percent of correct classification*

** *overall classification accuracy : ratio of the number of correctly classified pixels to the total number of classified pixels*

The second real data set is a four band image of size 200 × 175 taken July 1979 over Clarke, Oregon. The area corresponds to the Clarke, OR, 7.5-min U.S. Geological Survey quadrangle map. Figs. 16 and 17 show the first band of the image and the ground truth data. This image contains eight ground cover classes.

class 1: wheat	class 5: beans
class 2: alfalfa	class 6: apples
class 3: potatoes	class 7: pasture (irrigated)
class 4: corn	class 8: rangeland

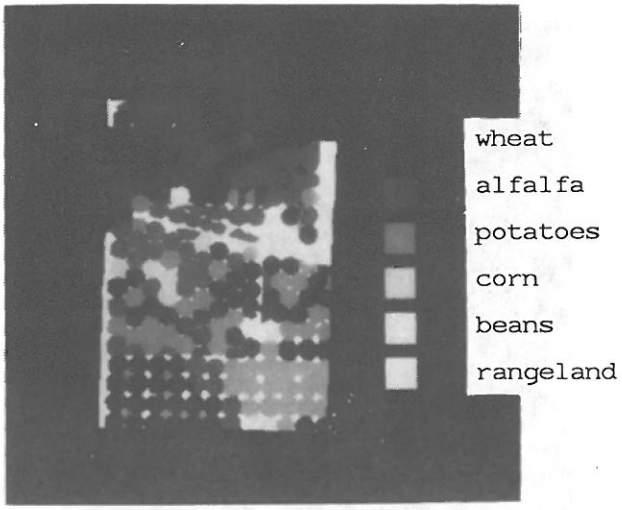

Fig. 17. Ground truth data for test image 3.

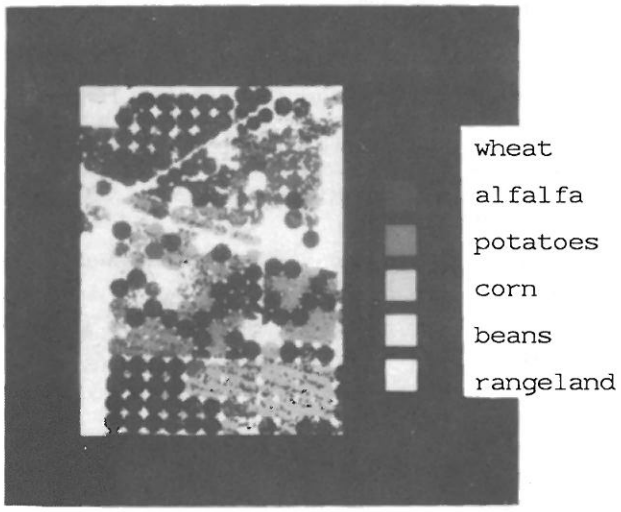

Fig. 18. Classified result of test image 3—classified by the Bayes classifier.

One forth of the image is randomly selected as a training set to estimate the class-conditional covariance matrices and mean vectors. Because of the small sample size, class 6 and 7 are excluded from the ground truth data, and the classifiers are applied to separate classes 1 through 5 and class 8. Figs. 18 and 19 show the classified results of the pixel-independent Bayes classifier and the context classifier. Their contingency tables are given in Table IV. Jenson [2] also tested this image and reported the overall classification accuracy of 64 percent for AMOEBA and 65 percent for ISOCLS. Since we used more extensive ground truth data than did Jenson *et al.* in their experiment, a direct comparision cannot be made. However, we find that the new classifier (73.2-percent overall classification accuracy) gained a 4.1-percent increase in the overall classification accuracy over the pixel-independent Bayes classifier. We performed more tests on Landsat images and consistently found that the new context classifier gained 3–8-percent increase in overall classification accuracy compared to the pixel independent Bayes classifier under Gaussian distribution assumption.

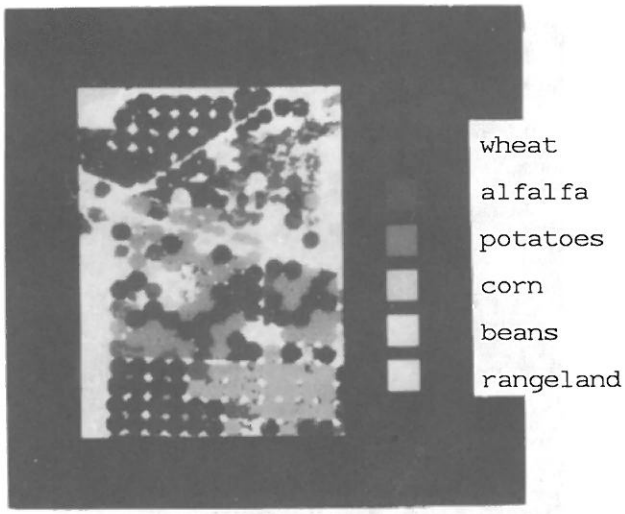

Fig. 19. Classified result of test image 3—classified by the context classifier.

TABLE IV
CONTINGENCY TABLES FOR CLASSIFICATION RESULTS OF TEST IMAGE 3
(Column = assigned class, row = true class.)

(a) *Bayes classifier result*

class	WHT	ALF	POT	CRN	BNS	RNG	total	Acc(%)*
WHT	10234	427	561	20	4	626	11872	86.2
ALF	1252	2405	2073	18	11	913	6672	36.0
POT	745	470	4872	5	1	431	6524	74.7
CRN	83	40	85	98	44	432	782	44.8
BNS	17	77	24	10	16	337	481	3.3
RNG	154	261	208	18	10	3274	3925	83.4
total	12485	3680	7823	169	86	6013	30256	69.1**

(b) *Context classifier result*

class	WHT	ALF	POT	CRN	BNS	RNG	total	Acc(%)*
WHT	10620	317	333	78	24	500	11872	89.5
ALF	901	2286	2190	49	616	630	6672	34.3
POT	306	324	5547	37	43	267	6524	85.0
CRN	89	14	23	250	154	252	782	32.0
BNS	16	15	7	68	275	100	481	57.2
RNG	176	156	240	25	152	3176	3925	80.9
total	12108	3112	8340	507	1264	4925	30256	73.2**

WHT : *wheat* CRN : *corn*
ALF : *alfalfa* BNS : *beans*
POT : *potatoes* RNG : *rangeland*

* *percent of correct classification*

** *overall classification accuracy : ratio of the number of correctly classified pixels to the total number of classified pixels*

The increase in classification accuracy and efficiencies can be compared to the other context classifiers proposed by several authors. Yu and Fu [9] used a stationary stochastic process on a two-dimensional plane as a model to extract the spatial correlation parameters, which is the context information used in their recursive context classifier. When tested on a real image, their results gained about a 7-percent increase in classification accuracy over the conventional Bayes classifier. Tilton *et al.* [6] and Swain *et al.* [5] derived an unbiased estimate of the context distribution $G(\theta^p)$ they used in the context classifier. The function $G(\theta^p)$ is the relative frequency with which

θ^p occurs in the array θ, where θ^p is the true label of the p-context array θ and the p-context array is a local context, a set of pixels neighboring the pixel in consideration. The optimal estimate of $G(\theta^p)$, designated by $T_{\theta p}(X)$, is a function that minimizes the mean squared error $E[T_{\theta p}(X) - G(\theta^p)]^2$. They reported 2–6-percent improvement in classification accuracy over the noncontext maximum likelihood classifier.

From the above descriptions, the computation cost of those two contextual classifiers increases exponentially as the size of the context increases. In comparison, the new context classifier gives about the same improvement in classification accuracy but requires only a fixed amount of computation per pixel. Further improvements can also be made in the context classifier described in this paper by obtaining a proper model for computation of the function $A(e_{ij}, e_{kl})$.

IV. Conclusions

A context classifier was designed so that it gives each pixel the highest probability label given some substantially sized context including the pixel. Applied on a simulated image, the context classifier shows better classification compared to the pixel-independent Bayes classifier. On real images, it was observed that the context classifier gained 4–8-percent increase in overall classification accuracy over the pixel-independent Bayes classifier under Gaussian distribution assumption.

Using assumptions (3) and (4) we were able to derive a two-pass algorithm that can be expressed as in (10) and (12). Even though we used a simple frequency-measuring method to estimate the joint probability of label pairs for successive pixels in the path, we obtained a higher classification accuracy for the context experiments. We also compared the improvement in overall classification accuracy to the other context classifiers developed so far and found comparable results.

References

[1] P. A. Devijver, "A comparative study of decision making algorithms in hidden Markov chains," Philips Res. Lab., Brussels, Belgium, Rep. R. 481, Oct. 1984.

[2] S. K. Jenson, T. R. Loveland, and J. Bryant, "Evaluation of AMOEBA: A spectral-spacial classification method," *J. Appl. Photographic Eng.*, vol. 8, no. 3, pp. 159–162, June 1982.

[3] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[4] R. M. Haralick, "Decision making in context," *IEEE Trans. Pattern Analysis Machine Intell.*, vol. PAMI-5, no. 4, pp. 417–428, July 1983.

[5] P. H. Swain, S. B. Vardeman, and J. C. Tilton, "Contextual classification of multispectral image data," *Pattern Recognition*, vol. 13, no. 6, pp. 429–441, 1981.

[6] J. C. Tilton, S. B. Vardeman, and P. H. Swain, "Estimation of context for statistical classification of multispectral image data," *IEEE Trans. Geosci. Remote Sensing*, vol. GE-20, no. 4, pp. 445–452, Oct. 1982.

[7] G. T. Toussaint, "The use of context in pattern recognition," *Pattern Recognition*, vol. 10, pp. 189–204, 1978.

[8] S. W. Wharton, "A contextual classification method for recognizing land use patterns in high resolution remotely sensed data," *Pattern Recognition*, vol. 15, no. 15, pp. 317–324, 1982.

[9] T. S. Yu and K. S. Fu, "Recursive contextual classification using a spatial stochastic model," *Pattern Recognition*, vol. 16, no. 1, pp. 89–108, 1983.

*

Robert M. Haralick (S'67–M'69–SM'76–F'84) was born in Brooklyn, NY, on September 30, 1943. He received the B.A. degree in mathematics from the University of Kansas in 1964, the B.S. degree in electrical engineering in 1966, the M.S. degree in electrical engineering in 1967, and the Ph.D. degree from the University of Kansas in 1969.

He has worked with Autonetics and IBM. In 1965, he worked for the Center for Research, University of Kansas, as a Research Engineer, and in 1969 he joined the faculty of the Electrical Engineering Department there where he last served as a Professor from 1975 to 1978. In 1979, he joined the faculty of the Electrical Engineering Department at Virginia Polytechnic Institute and State University where he was a Professor and Director of the Spatial Data Analysis Laboratory. From 1984 to 1986 he served as Vice President of Research at Machine Vision International, Ann Arbor, MI. He now holds a Boeing chaired professorship in the Department of Electrical Engineering at the University of Washington, Seattle. He has done research in pattern recognition, multi-image processing, remote sensing, texture analysis, data compression, clustering, artificial intelligence, and general systems theory, and has published over 200 papers. He is responsible for the development of General Image Processing System, (GIPSY), a multi-image processing package that runs on a minicomputer system.

Dr. Haralick is a member of the Association for Computer Machinery and the Pattern Recognition Society.

*

Hyonam Joo (S'86) was born in Seoul, Korea, on August 2, 1953. He received the B.S. degree in electrical engineering from Seoul National University, Seoul, Korea, in 1976 and the M.S. degree in electrical engineering from Virginia Polytechnic Institute and State University, Blacksburg, in 1985. He is currently working toward the Ph.D. degree in electrical engineering from Virginia Polytechnic Institute and State University.

He worked with the Agency for Defense Development, Korea, as a Research Engineer from 1976 to 1982. In addition to his graduate work, he is currently a Reseach Associate at Machine Vision International, Ann Arbor, MI.