COMPUTER GRAPHICS AND IMAGE PROCESSING 12, 286–308 (1980)

Finding Structure in Co-Occurrence Matrices for Texture Analysis *

STEVEN W. ZUCKER AND DEMETRI TERZOPOULOS

Computer Vision and Graphics Laboratory, Department of Electrical Engineering, McGill University, Montreal, Quebec, Canada

Co-occurrence matrices are a popular representation for the texture in images. They contain a count of the number of times that a given feature (e.g., a given gray level) occurs in a particular spatial relation to another given feature. However, because of the large number of spatial relations that are possible within an image, heuristic or interactive techniques have usually been employed to select the relation to use for each problem. In this paper we present a statistical approach to finding those spatial (or other) relations that best capture the structure of textures when the co-occurrence matrix representation is used. These matrices should thus be well suited for discriminations that are structurally based.

1. INTRODUCTION

Texture can be viewed as a global pattern arising from the repetition, either deterministically or randomly, of local subpatterns [11]. The structure resulting from this repetition is often important in discriminating between different textures. In this paper, we present a statistical approach to selecting a description from within a well-known class of representations that best reflects such structure. The class of representations is the co-occurrence matrix introduced by Haralick [6], which has been used for remote sensing, biomedical, and many other application areas (see, e.g., the recent survey [5]).

The co-occurrence matrix is essentially a two-dimensional histogram of the number of times that pairs of intensity values (or, more generally, arbitrary local features) occur in a given spatial relationship. Thus, it forms a summary of the subpatterns that could be formed by intensity pairs and the frequency with which they occur.

The success of this kind of representation is tied directly to the fidelity with which it captures the structure of the underlying texture. To understand the importance of selecting the spatial relationship properly, consider a textural pattern with a dominant banded (i.e., almost constant) structure in the horizontal direction, but with a random structure in the vertical direction. If the co-occurrence matrix were built using a spatial relationship between vertical

* This research was supported by MRC Grant MA-6154.

0146-664X/80/030286-23$02.00/0

pairs of pixels (at some separation), then it would appear as if the underlying texture were perfectly random. On the other hand, if a horizontal spatial relationship were used, then it would appear as if the texture were highly structured. For a co-occurrence matrix computed as an average over several directions, the description would be somewhere between these two extremes.

In the above example, only one variable (orientation) was considered in defining the spatial relationship on which to base the co-occurrence description, and this variable was shown to be critical in capturing the structure of the texture. In more realistic situations, such dependencies are rarely that obvious. Furthermore, similar difficulties arise in attempting to choose a spatial relation on which to base first-order gray-level statistics [10] or generalized co-occurrence matrices [3]. In fact, for generalized co-occurrence matrices, the situation is further complicated by the wide choice of local features that can be used.

The specific problem to be addressed in this paper is the selection of spatial relationships for defining co-occurrence matrices such that these matrices maximally reflect the structure of the underlying texture. The model adopted for structure is statistical and is based on a (chi-square) measure of independence of the rows and columns of the co-occurrence matrix. Thus it provides quantitative support for the design of co-occurrence-based classification schemes, and supplements the interactive and heuristic (in particular, the exhaustive [10]) tools currently available.

The measure of independence is formulated by interpreting the co-occurrence matrix as a contingency table [7]. These notions are formally defined in the next two sections, and are followed by example applications to several Brodatz [2] textures and to the LANDSAT-1 images considered in [10]. The dependency measure clearly indicates the size of the structural units in the Brodatz textures. It further provides quantitative corroboration for Weszka, Dyer, and Rosenfeld's empirical observations about the merits of certain spatial relationships for the LANDSAT images. Finally, it leads to the specification of co-occurrence matrices upon which successful classifiers can be designed.

2. CO-OCCURRENCE MATRICES FOR TEXTURE CLASSIFICATION

In this section, we formally define co-occurrence matrices and list several feature functions that can be computed over them. Such features usually provide the basis for classification in practical problems.

Let f be a rectangular, discrete picture containing a finite number of gray levels. f is defined over the domain

$$\mathfrak{D} = \{(i, j) \mid i \in [0, n_i),\ j \in [0, n_j),\ i, j \in I\}$$

by the relation

$$f = \{((i, j), k) \mid (i, j) \in \mathfrak{D},\ k = f(i, j),\ k \in [0, n_g),\ k \in I\},$$

where I is the set of integers, n_i and n_j are the horizontal and vertical dimensions of f, and n_g is the number of gray levels in f.

The unnormalized co-occurrence matrix, Ψ, is a square matrix of dimension n_g and is a function of both the image f and a displacement vector $\bar{d} = [\Delta i, \Delta j]$

in the (i, j) plane; i.e.,

$$\Psi(f, \bar{d}) = \{\psi_{ij}(f, \bar{d})\}.$$

Its entries, ψ_{ij}, are the unnormalized frequencies of co-occurring gray levels in f which are separated by the spatial relation $\bar{d}$:

$$\psi_{ij}(f, \bar{d}) = \#\{((k_1, l_1), (k_2, l_2)) \mid (k_1, l_1), (k_2, l_2) \in \mathfrak{D}, \\ f(k_1, l_1) = i, f(k_2, l_2) = j, [k_2, l_2] - [k_1, l_1] = \bar{d}\},$$

where # denotes the number of elements in the set.

Symmetric co-occurrence matrices are generated by pooling frequencies of gray-level occurrences that are separated by both $\bar{d}$ and $-\bar{d}$. That is, the symmetric co-occurrence matrix, Φ, is defined by the relation

$$\Phi(f, \bar{d}) = \Psi(f, \bar{d}) + \Psi(f, -\bar{d}).$$

For the remainder of this paper, we shall deal with symmetric co-occurrence matrices exclusively.

It is often convenient to normalize co-occurrence matrices so that they approximate discrete joint probability densities of co-occurring gray levels. The appropriate normalization is accomplished by dividing each entry in the co-occurrence matrix by the total number of paired occurrences

$$\Phi_N = \frac{1}{N}\Phi, \qquad \text{where} \qquad N = \sum_i \sum_j \phi_{ij}.$$

Texture classification is usually accomplished by using certain characteristic features of co-occurrence matrices. That is, the values of a number of feature functions can be used to summarize the content of the matrices. Fourteen such functions were introduced by Haralick *et al.* [6], four of which appear to be the most widely used in practice. These functions are listed below since they are used for the classification experiments described in Section 5.

(1) Angular second moment: $\text{ASM} = \sum_i \sum_j \phi_{ij}^2$;

(2) Contrast: $\text{CON} = \sum_i \sum_j (i - j)^2 \phi_{ij}$;

(3) Correlation: $\text{COR} = \sum_i \sum_j (ij\phi_{ij} - \mu_x\mu_y)/(\sigma_x\sigma_y)$;

(4) Entropy: $\text{ENT} = -\sum_i \sum_j \phi_{ij} \log \phi_{ij}$.

ϕ_{ij} denotes the (i, j)th element of a normalized co-occurrence matrix Φ_N, μ_x, and σ_x are the mean and standard deviation of the marginal probability vector, obtained by summing over the rows of Φ_N, and μ_y and σ_y are the corresponding statistics for the column sums.

3. CONTINGENCY TABLES AND χ^2 SIGNIFICANCE TESTS

A new point of view toward co-occurrence matrices can be developed by interpreting intensity pairs (or local feature values) in an image as samples ob-

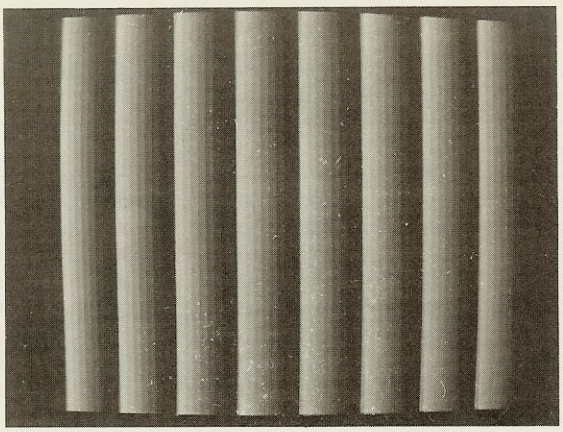

FIG. 1. A banded texture (gray levels 1 to 8 repeated every eight columns).

tained from a (two-dimensional) random process. The rows and columns of a co-occurrence matrix separate the samples into various classes based on observed intensities. The matrix thus tabulates the frequencies of samples belonging to each class.

Such tables have been used for some time in statistics, and are called "contingency" tables. They typically take the following form:

A \ B	1 $\cdots$ j $\cdots$ n	Row totals
1	$x_{11} \cdots x_{1j} \cdots x_{1n}$	r_1
$\vdots$	$\vdots \quad \vdots \quad \vdots$	$\vdots$
i	$x_{i1} \cdots x_{ij} \cdots x_{in}$	r_i
$\vdots$	$\vdots \quad \vdots \quad \vdots$	$\vdots$
m	$x_{m1} \cdots x_{mj} \cdots x_{mn}$	r_m
Column totals	$c_1 \cdots c_j \cdots c_n$	N

In the above table, x_{ij} is the number of times that variable A has been observed to fall into class i while B has fallen into class j. Furthermore, the various row and column totals are indicated.

If we interpret variables A and B as image pixels at either end of $\bar{d}$, and the classes into which they can fall as the gray levels $[0, n_g)$, the correspondence between co-occurrence matrices and contingency tables becomes clear. An unnormalized co-occurrence matrix, Ψ, is a square contingency table with $m = n = n_g$.

The importance of adopting this interpretation of co-occurrence matrices is that it allows the formulation of a precise statistical measure for the amount of textural structure that is contained in any particular matrix. For motivation, we present the following simple example. Consider a highly structured, one-dimensional, banded texture (Fig. 1) that repeats every eight columns. If we restrict the displacement vector $\bar{d}$ to lie in the horizontal direction, different

TABLE 1

Co-Occurrence Matrices, over Two Spatial Relationships, for the Banded Texture of Fig. 1[a]

Spatial relation: $\bar{d} = [3, 0]$							
0.0000	0.0000	0.0000	0.0574	0.0000	0.0656	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0656	0.0000	0.0574	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0656	0.0000	0.0574
0.0574	0.0000	0.0000	0.0000	0.0000	0.0000	0.0656	0.0000
0.0000	0.0656	0.0000	0.0000	0.0000	0.0000	0.0000	0.0656
0.0656	0.0000	0.0656	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0574	0.0000	0.0656	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0574	0.0000	0.0656	0.0000	0.0000	0.0000
Spatial relation: $\bar{d} = [8, 0]$							
0.1250	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.1250	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.1250	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.1250	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.1250	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.1250	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1250	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1250

[a] Note that the matrix becomes diagonal for distances that are integer multiples of eight pixels.

values of $|\bar{d}|$ would lead to co-occurrence matrices with very different entries. In particular, a choice of $|\bar{d}| = 8 \cdot n$, where n is an integer, would lead to a diagonal Φ (Table 1), conveying the banded structure; any other choice would miss the important regularity, thus falsely indicating a more random underlying texture. This example shows that different choices for $\bar{d}$ can lead to very different descriptions of the same textural pattern.

Our notion of structure, conveyed by co-occurrence matrices, is related to the strength of the statement that can be made about variable B given observations about A (and vice versa). If the texture is highly structured and the co-occurrence matrix reflects this structure, then observations about A should bias the probabilities of observing various classes for B. On the other hand, if the structure is not being captured, then observations about A will not influence the probabilities for B. In other words, A and B will be independent. The amount of structure conveyed by a co-occurrence matrix clearly depends on the choice of variables A and B; that is, it is a function of $\bar{d}$.

A quantitative measure of this structure can be obtained by hypothesizing that the variables A and B are independent, and then using a chi-square goodness of fit test to determine the degree to which this hypothesis can be rejected by the observed data (i.e., the image). For textures with a lot of structure, it should be rejected overwhelmingly. Operationally, the hypothesis translates into a statement about row/column independence in contingency tables.

In general, the chi-square test is employed to determine whether observed

frequencies of occurrences in a set of randomly drawn samples appear to have been drawn from an assumed distribution. The test involves calculation of the quantity

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - e_i)^2}{e_i}, \tag{1}$$

where the x_i and e_i represent observed and expected frequencies in the ith class, respectively. As the number of samples drawn approaches infinity the above distribution approaches that of a chi-square function with $k - 1$ degrees of freedom.

The chi-square test can still be applied in situations where the expected frequencies depend on unknown parameters, if maximum likelihood estimators are used to estimate these parameters ([7, Sect. 9.6]; but see also [1, Sect. 3.1.1]). Furthermore, it becomes necessary to subtract one degree of freedom for each parameter estimated.

To formulate the independence hypothesis consider an arbitrary contingency table. Let p_{ij} be the probability corresponding to the cell in the ith row and jth column, and let $p_{i.}$ be the probability corresponding to the ith row and $p_{.j}$ the probability corresponding to the jth column. The hypothesis that the two variables, A and B, are independent may now be written as

$$H_0\colon\ p_{ij} = p_{i.}p_{.j}, \qquad i = 1, \ldots, m,\ \ j = 1, \ldots, n.$$

If N denotes the total number of samples, i.e.,

$$N = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij},$$

then the measure of compatibility between observed and expected frequencies is (from (1))

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(x_{ij} - Np_{ij})^2}{Np_{ij}}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(x_{ij} - Np_{i.}p_{.j})^2}{Np_{i.}p_{.j}} \tag{2}$$

under hypothesis H_0. Since $p_{i.}$ and $p_{.j}$ are unknown, it is necessary to compute their maximum likelihood estimates. First, however, note that

$$\sum_{i=1}^{m} p_{i.} = 1 \qquad \text{and} \qquad \sum_{j=1}^{n} p_{.j} = 1. \tag{3}$$

Thus $(m - 1) + (n - 1) = m + n - 2$ parameters must be estimated. Therefore, the number of degrees of freedom for testing H_0 is

$$\begin{aligned} \nu &= (\text{number of cells}) - 1 - (\text{number of estimated parameters}) \\ &= (m - 1)(n - 1). \end{aligned} \tag{3.1}$$

To determine the maximum likelihood estimates of $p_{i\cdot}$ and $p_{\cdot j}$, note that the data samples are discrete and independent. Then the likelihood of the sample (i.e., the probability of obtaining the sample in the order of its occurrence) is

$$L = \prod_{i=1}^{m} \prod_{j=1}^{n} p_{ij}^{x_{ij}}$$

which, under H_0, becomes

$$\begin{aligned} L &= \prod_{i=1}^{m} \prod_{j=1}^{n} (p_{i\cdot} p_{\cdot j})^{x_{ij}} \\ &= \prod_{i=1}^{m} \prod_{j=1}^{n} p_{i\cdot}^{x_{ij}} \prod_{i=1}^{m} \prod_{j=1}^{n} p_{\cdot j}^{x_{ij}} \\ &= \prod_{i=1}^{m} p_{i\cdot}^{\sum_{j=1}^{n} x_{ij}} \prod_{j=1}^{n} p_{\cdot j}^{\sum_{i=1}^{m} x_{ij}} \\ &= \prod_{i=1}^{m} p_{i\cdot}^{r_i} \prod_{j=1}^{n} p_{\cdot j}^{c_j}, \end{aligned}$$

where

$$r_i = \sum_{j=1}^{n} x_{ij}$$

and

$$c_j = \sum_{i=1}^{m} x_{ij}$$

are the sums of the frequencies in the ith row and jth column, respectively. It is convenient to express one of the $p_{\cdot j}$, say $p_{\cdot n}$, in terms of the others by using relation (3).

Hence

$$L = (1 - \sum_{j=1}^{n-1} p_{\cdot j})^{c_n} \prod_{i=1}^{m} p_{i\cdot}^{r_i} \prod_{j=1}^{n-1} p_{\cdot j}^{c_j}.$$

Taking logarithms, we have

$$\log L = c_n \log (1 - \sum_{j=1}^{n-1} p_{\cdot j}) + \sum_{i=1}^{m} r_i \log p_{i\cdot} + \sum_{j=1}^{n-1} c_j \log p_{\cdot j}.$$

A maximum likelihood estimate of $p_{\cdot j}$ may be obtained by differentiating with respect to $p_{\cdot j}$ and setting the derivative to zero:

$$\frac{\partial \log L}{\partial p_{\cdot j}} = -c_n/(1 - \sum_{j=1}^{n-1} p_{\cdot j}) + \frac{c_j}{p_{\cdot j}} = 0.$$

Now

$$1 - \sum_{j=1}^{n-1} p_{\cdot j} = p_{\cdot n}.$$

Thus

$$p_{\cdot j} = \frac{p_{\cdot n}}{c_n} c_j = \lambda c_j,$$

where λ is independent of j. Since this must be true for all $j = 1, \ldots, n$,

$$1 = \sum_{j=1}^{n} p_{\cdot j} = \lambda \sum_{j=1}^{n} c_j = \lambda N.$$

Therefore, $\lambda = 1/N$, and the maximum likelihood estimate of $p_{\cdot j}$ is

$$\hat{p}_{\cdot j} = \frac{c_j}{N}.$$

Similarly, the maximum likelihood estimate of $p_{i\cdot}$ can be shown to be

$$\hat{p}_{i\cdot} = \frac{r_i}{N}.$$

Replacing $p_{i\cdot}$ and $p_{\cdot j}$ in the expression for χ^2 (Eq. (2)) by their maximum likelihood estimates, we obtain

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(x_{ij} - (r_i c_j/N))^2}{r_i c_j/N}. \tag{4}$$

If N is sufficiently large and H_0 is true, then (4) will possess a chi-square distribution with $(m - 1)(n - 1)$ degrees of freedom.

A more computationally efficient expression for (4) may be obtained after some algebra:

$$\chi^2 = N\left(\sum_{i=1}^{m} \sum_{j=1}^{n} \frac{x_{ij}^2}{r_i c_j} - 1\right). \tag{5}$$

Computing (4) or (5) over a square contingency table yields a direct measure of the significance of H_0 with respect to that table. The level of significance is a function of the number of degrees of freedom and may be found by inspection of χ^2 tables [7] or by numerical integration of the χ^2 distribution. If χ^2 (adjusted by the proper number of degrees of freedom) exceeds a critical value, χ_0^2 (usually 0.05), then H_0 is rejected; otherwise H_0 is accepted. Or, for comparison purposes, χ^2 provides a continuous measure of structure within contingency tables.

4. AN ALGORITHM FOR SELECTING CO-OCCURRENCE MATRICES FOR TEXTURE CLASSIFICATION

Since we are interested in co-occurrence matrices that reflect the greatest amount of structure in the underlying texture, it is straightforward to devise algorithms to select the best matrix (or matrices) from a set of candidate matrices (these candidates may be obtained, for example, by using different spatial rela-

tions $\bar{d}$). One merely determines the goodness of fit of H_0 (by applying (4) or (5)) and selects the matrix (or matrices) yielding the highest value of χ^2.

As long as the co-occurrence matrices are complete (i.e., contain no zero entries) the theory presented in the previous section is applicable as stated. However, in the analysis of real texture pictures, one or more allowable gray levels may never occur. This causes entire rows and columns (due to symmetry) in co-occurrence matrices to have zero entries. When such matrices are interpreted as contingency tables, each degenerate row will result in one of the r_i having a value of zero, while each degenerate column will result in a zero c_j. Thus, zero maximum likelihood estimates would be obtained for one or more of the $p_{i\cdot}$ or $p_{\cdot j}$ parameters.

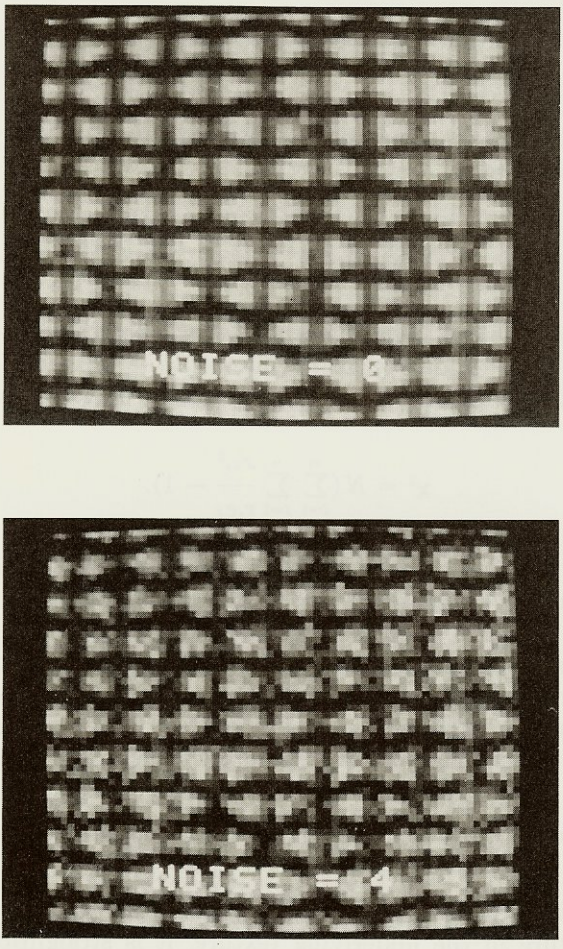

FIG. 2a. A texture that has been subjected to several magnitudes of additive noise in order to vary the structure.

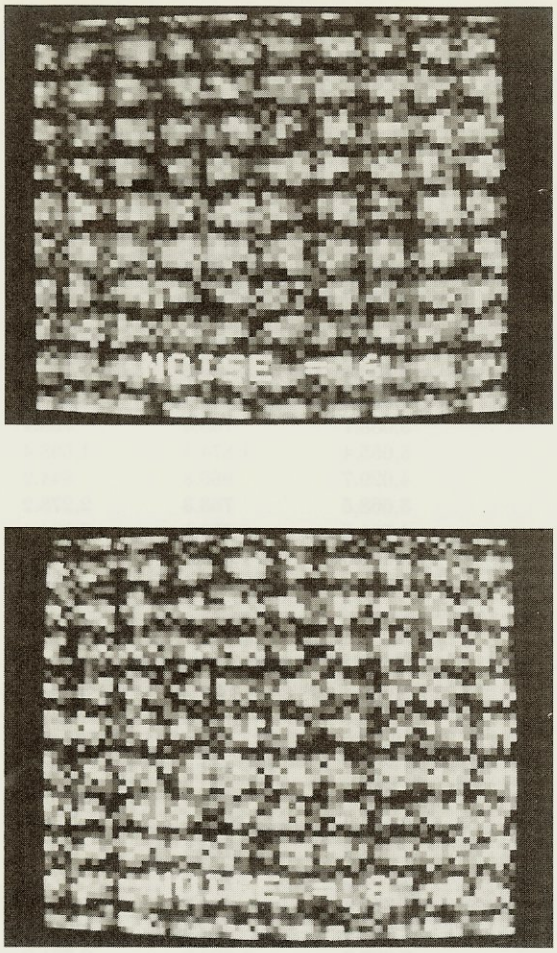

FIG. 2a—*Continued.*

The presence of nonestimable parameters requires that special action be taken in the application of (4) or (5), as well as adjustments in the number of degrees of freedom. Extensive discussion of the problems caused by incomplete contingency tables may be found in [1].

For our purposes, since each nonestimable parameter contains no information about a particular category (i.e., intensity), incomplete contingency tables may be collapsed into smaller, complete tables by eliminating all degenerate rows and columns. Equation (3.1) should then be used to determine the proper number of degrees of freedom for the complete table. That is, if ν is the result of (3.1) for the original incomplete table, $\mathfrak{Z}_e$ the number of cells with zero estimates, and $\mathfrak{Z}_p$ the number of nonestimable parameters, the adjusted number of degrees of freedom is given by

$$\nu' = \nu - \mathfrak{Z}_e + \mathfrak{Z}_p.$$

In actual practice, the proper number of degrees of freedom for any co-occur-

TABLE 2

χ^2 Values for Co-Occurrence Matrices Computed from the Images in Fig. 2a

Distance	Angles			
	0°	45°	90°	135°
Noise = 0				
2	4,781.5	1,100.4	1,261.2	1,200.4
3	5,046.8	2,659.0	2,199.7	2,601.1
4	4,082.2	1,097.8	738.2	1,114.0
5	3,537.9	923.5	1,853.3	896.9
6	5,466.1	6,112.5	5,749.3	6,794.2
7	6,048.3	2,267.1	1,826.3	2,323.1
8	3,688.7	738.4	738.1	830.7
9	3,055.4	1,874.4	1,698.4	1,873.6
10	4,029.7	986.8	844.2	925.9
11	3,668.5	763.3	2,278.2	627.8
12	4,761.5	2,533.1	7.934.4	2,739.0
13	10,492.9	1,487.2	2,217.7	1,472.0
14	7,745.0	564.9	779.6	597.4
Noise = 4				
2	1,636.4	500.7	501.7	559.3
3	1,605.7	1,159.0	621.2	1,066.0
4	1,476.8	580.8	361.1	639.6
5	1,543.1	483.4	820.0	451.4
6	1,983.1	2,455.8	2,082.4	2,675.5
7	2,136.2	993.0	902.0	1,073.4
8	1,568.3	440.6	368.9	345.4
9	1,255.2	963.8	571.9	819.9
10	1,372.8	494.7	372.5	428.2
11	1,202.4	404.5	910.8	376.9
12	2,037.8	1,090.9	2,781.8	1.246.3
13	4,142.1	707.7	1,068.9	741.8
14	2,910.2	403.4	364.3	380.3
Noise = 6				
2	972.9	407.8	366.9	377.7
3	934.6	599.6	555.3	700.9
4	1,002.0	394.6	418.7	387.6
5	843.6	519.7	635.2	454.2
6	1,073.6	1,065.7	1,013.1	1,209.3
7	1,141.2	669.1	572.2	633.1
8	920.6	292.8	346.2	315.3
9	778.2	582.0	447.3	484.0
10	791.2	326.1	343.9	377.9
11	853.5	439.6	552.8	418.4
12	1,119.9	715.4	1,233.5	749.2
13	1,352.0	492.4	585.8	530.7
14	1,303.5	316.6	344.6	313.3

TABLE 2—*Continued*

Distance	Angles			
	0°	45°	90°	135°
Noise = 8				
2	714.8	446.0	407.6	433.7
3	608.4	528.4	438.8	485.9
4	601.3	386.5	448.4	438.1
5	601.3	445.1	527.2	481.5
6	629.1	687.7	582.3	682.8
7	651.8	611.1	438.5	463.2
8	525.8	378.4	353.2	420.3
9	500.1	530.7	380.5	426.2
10	624.0	379.5	371.5	358.8
11	533.9	398.2	522.5	361.6
12	586.3	508.8	623.0	429.6
13	677.5	451.1	576.1	422.9
14	730.5	383.2	402.3	314.3

rence matrix is not required by our selection algorithm, because absolute levels of significance need not be computed. Rather, only relative comparisons of the magnitudes of χ^2 values (computed over candidate matrices) need be made; the degrees of freedom are constant and hence cancel out.

5. EXPERIMENTS

Two different kinds of experiments were performed. The first kind was designed to show that the χ^2 measure did indeed vary with the amount of structure in the texture pattern, while the second kind involved the classification of textures.

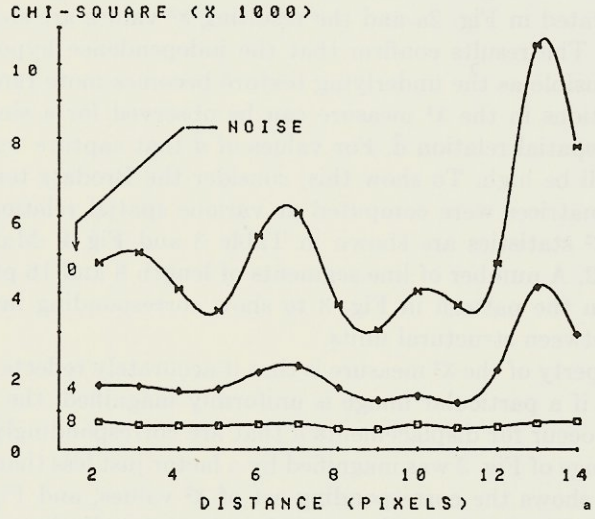

FIG. 2b. Graph of χ^2 vs distance (at 0°) for several magnitudes of additive noise. The curves were generated from selected entries in Table 2.

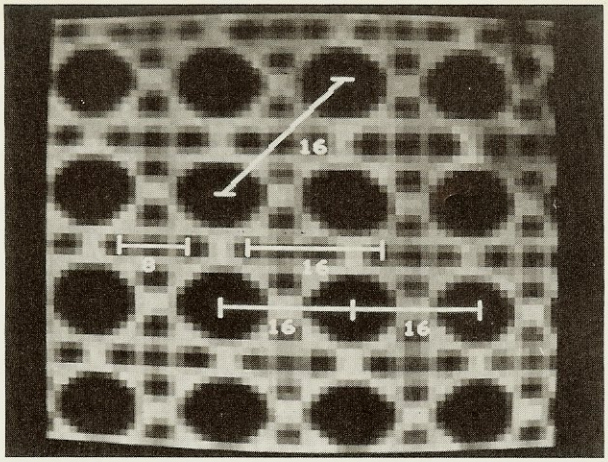

FIG. 3. A texture pattern on which the displacements (in pixels) between several major structural units are superimposed.

To show the relation between χ^2 values computed over co-occurrence matrices and the degree of structure present in the underlying texture, we used a number of Brodatz patterns. These texture samples were digitized into 256×256 images with 256 gray levels. The image intensities were then requantized into 16 equiprobable intensities in order to compensate for varying brightnesses and contrasts among the originals. The amount of structure in each image was varied by the pointwise addition of (uniform, zero mean) pseudorandom noise over different ranges (i.e., uniformly distributed in $[-N, N]$). χ^2 values for co-occurrence matrices constructed over several spatial relationships (various distances over the four angles, 0, 45, 90, and 135°) were computed. As expected, the χ^2 statistics decreased with increasing noise at all values of $\bar{d}$ for all images tested. A typical texture is illustrated in Fig. 2a and the resulting χ^2 values are shown in Table 2 and in Fig. 2b. The results confirm that the independence hypothesis, H_0, becomes more plausible as the underlying texture becomes more random.

Similar variations in the χ^2 measure can be observed for a single image as a function of the spatial relation $\bar{d}$. For values of $\bar{d}$ that capture texture structure very well, χ^2 will be high. To show this, consider the Brodatz texture in Fig. 3. Co-occurrence matrices were computed at various spatial relationships and the corresponding χ^2 statistics are shown in Table 3 and Fig 4. Maxima occur for $|\bar{d}| = 16$ and 32. A number of line segments of length 8 and 16 pixels have been superimposed on the pattern in Fig. 3 to show corresponding image scales and the distances between structural units.

A further property of the χ^2 measure is that it accurately reflects image magnification. That is, if a particular image is uniformly magnified, the maxima of the χ^2 measure will occur for displacements $\bar{d}$ that are correspondingly enlarged. For example, the image of Fig. 3 was magnified by a factor just less than 2, as shown in Fig. 5. Table 4 shows the corresponding set of χ^2 values, and Fig. 6 is a graph of some of them. Maximal χ^2 values clearly occur at a displacement $|\bar{d}| = 24$, almost twice the displacement obtained for the original image. In both images,

TABLE 3

χ^2 Values for Co-Occurrence Matrices Computed from the Texture of Fig. 3

Distance	Angles			
	0°	45°	90°	135°
2	22,458.4	8,320.1	14,298.8	8,731.5
4	7,295.5	7,983.4	3,788.8	7,159.5
6	8,583.4	5,165.5	4,920.6	4,959.8
8	9,718.9	4,649.2	4,325.6	5,217.7
10	7,030.4	4,986.1	3,949.4	4,615.4
12	6,022.6	2,478.0	4,160.5	2,857.6
14	22,041.3	13,154.4	21,233.0	9,864.8
16	53,866.1	22,583.3	40,782.3	34,186.0
18	12,413.4	2,069.3	7,000.1	4,332.1
20	3,796.6	2,858.9	2,329.0	3,066.1
22	4,820.5	2,674.4	2,809.5	2,903.3
24	4,963.5	2,694.2	2,558.3	3,400.9
26	3,413.6	1,608.0	2,028.5	2,215.3
28	5,338.6	2,658.9	6,942.3	1,097.8
30	20,597.7	14,588.6	23,054.7	8,588.1
32	25,696.2	5,582.3	13,537.1	12,547.6
34	6,747.6	778.5	2,489.9	2,041.3
36	1,950.7	1,401.9	1,320.9	1,299.0
38	2,724.4	2,078.3	1,950.6	2,146.6
40	2,478.9	1,454.2	1,244.9	2,236.6

the displacement between the major structural units in the texture are accurately reflected by the χ^2 measure.

Next, we present the results of several texture classification experiments. A

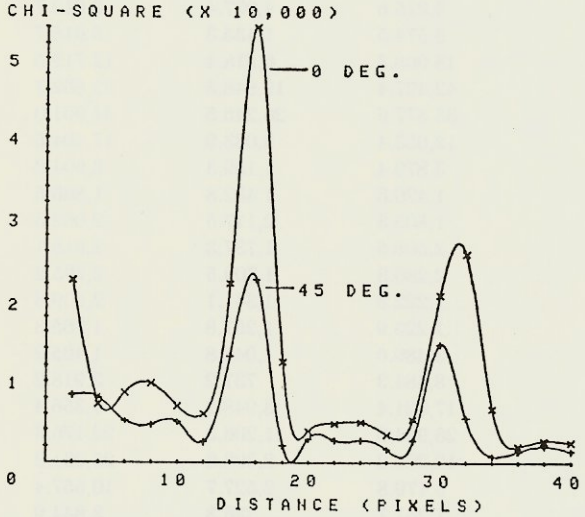

FIG. 4. Graph of χ^2 vs distance at two orientations. The curves were generated from selected entries in Table 3. Note how the displacements between structural units in Fig. 3 are reflected by the location of the maxima in the curves.

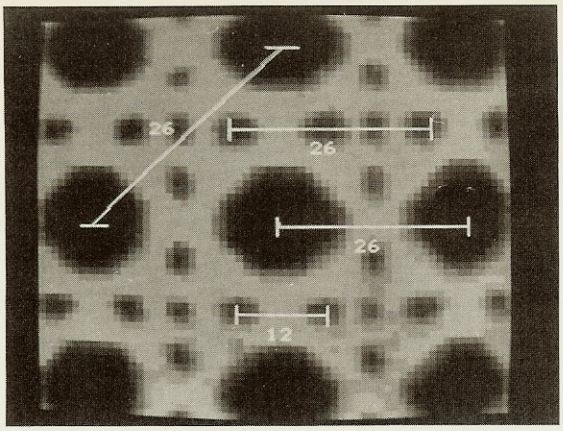

FIG. 5. The texture of Fig. 3 uniformly magnified. Displacements between major structural units are shown.

TABLE 4

χ^2 Values for Co-Occurrence Matrices Computed from the Texture of Fig. 5

Distance	Angles			
	0°	45°	90°	135°
8	3,452.6	5,724.2	3,987.1	5,379.1
10	4,177.5	3,223.3	5,131.1	3,205.7
12	4,889.5	4,073.3	5,333.7	3,959.5
14	3,884.0	2,798.1	4,709.8	2,486.6
16	2,328.3	2,442.9	3,923.9	2,394.3
18	3,375.6	4,037.4	3,012.7	3,271.7
20	8,574.5	1,435.3	5,015.7	1,382.5
22	18,968.3	6,518.4	12,719.5	7,777.2
24	42,427.4	19,848.5	32,862.7	21,826.0
26	33,877.6	26,246.5	44,901.0	16,567.0
28	12,053.4	8,033.9	17,494.5	4,676.1
30	3,879.4	1,159.3	6,604.2	714.6
32	1,420.6	1,432.8	1,886.5	1,192.8
34	1,805.3	2,117.5	2,062.5	1,407.9
36	2,508.5	1,737.3	2,815.1	1,291.4
38	2,280.8	1,709.5	2,795.2	1,156.9
40	1,222.8	1,341.1	2,509.8	948.9
42	1,223.9	1,257.8	1,785.3	1,145.7
44	3,489.6	1,046.8	1,405.2	449.9
46	8,484.3	737.2	3,918.2	1,730.2
48	17,021.4	3,948.2	9,356.4	6,447.5
50	26,934.0	11,206.3	22,176.3	10,927.8
52	12,280.2	8,308.6	25,231.2	6,049.5
54	5,179.8	2,527.7	10,557.4	2,089.6
56	1,261.7	236.3	3,844.9	756.4
58	620.4	662.4	924.3	1,294.4
60	1,064.8	893.5	1,164.0	1,519.2

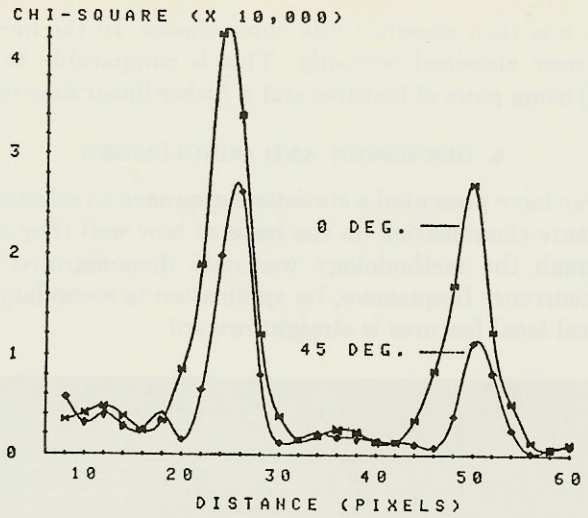

FIG. 6. Graph of selected χ^2 values from Table 4. Compared to Fig. 4, the maxima occur at larger distances due to the magnification of the structural units.

minimum Mahalanobis distance, linear discriminant classifier was used, the details of which are given in the Appendix.

The first experiment involved five Brodatz textures which are shown in Fig. 7. Twenty-five samples were taken from each texture (class) by extracting 64 × 64 windows over the equiprobability quantized images. Co-occurrence matrices were computed over several distances and orientations of 0, 45, 90, and 135°. χ^2 values were computed for each matrix and the distance yielding the maximum χ^2 measure, averaged over the four orientations, was chosen. The ASM, CON, and COR feature functions were computed over the four chosen matrices (at the above distance). The feature vectors (length 6) were generated by computing the mean and range of the values of each feature function. The classifier was trained on the 100 feature vectors (25 in each class). One hundred new windows were then selected and classified. The classification was 100% correct.

The second classification experiments involved the set of LANDSAT-1, Eastern Kentucky terrain images used by Weszka *et al.* [10] in their main study. The data set consists of three groups consisting of 60 images each. Each image is 64 × 64 pixels and has been histogram flattened to cover 64 gray levels (see Fig. 8). Co-occurrence matrices were, once again, computed for each window over several spatial relationships. For all of the 180 windows, $\bar{d} = [1, 0]$ and $\bar{d} = [0, 1]$ yielded matrices with the largest χ^2 values. Hence, these spatial relationships should be preferred in a classification of the above texture sample. Table 5 shows typical χ^2 results. It is interesting to note that these findings support Rosenfeld's observation that displacements of 1 pixel in the horizontal paired with size 1 in the vertical direction yielded the best classification results.

To further corroborate this observation, we trained our classifier on the basis of matrices selected by maximal χ^2 values. Feature vectors (length 3) were constructed for each of the 180 matrices using the ASM, CON, and ENT features.

The training set was then classified into three classes. In the final analysis, 85% of the samples were classified correctly. This is comparable to the best result obtained in [10] using pairs of features and a Fisher linear discriminant classifier.

6. DISCUSSION AND CONCLUSIONS

In this paper we have presented a statistical approach to selecting co-occurrence matrices for texture classification on the basis of how well they captured texture structure. Although the methodology was only demonstrated for matrices of intensity pair occurrence frequencies, its application to secondary images derived from more general local features is straightforward.

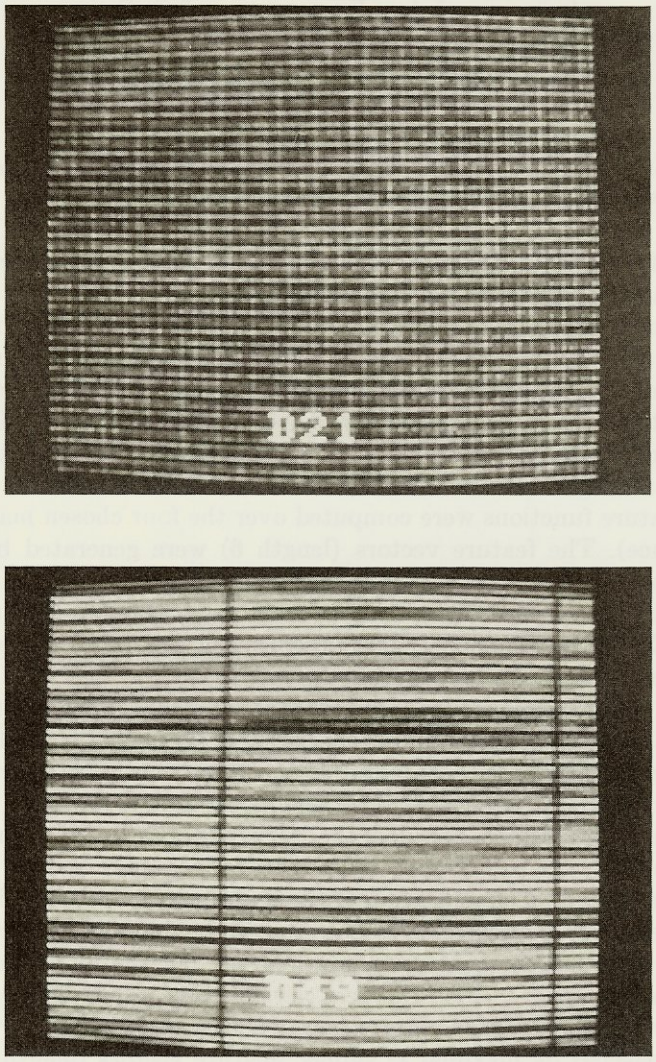

FIG. 7. The five Brodatz textures used in the first classification experiment.

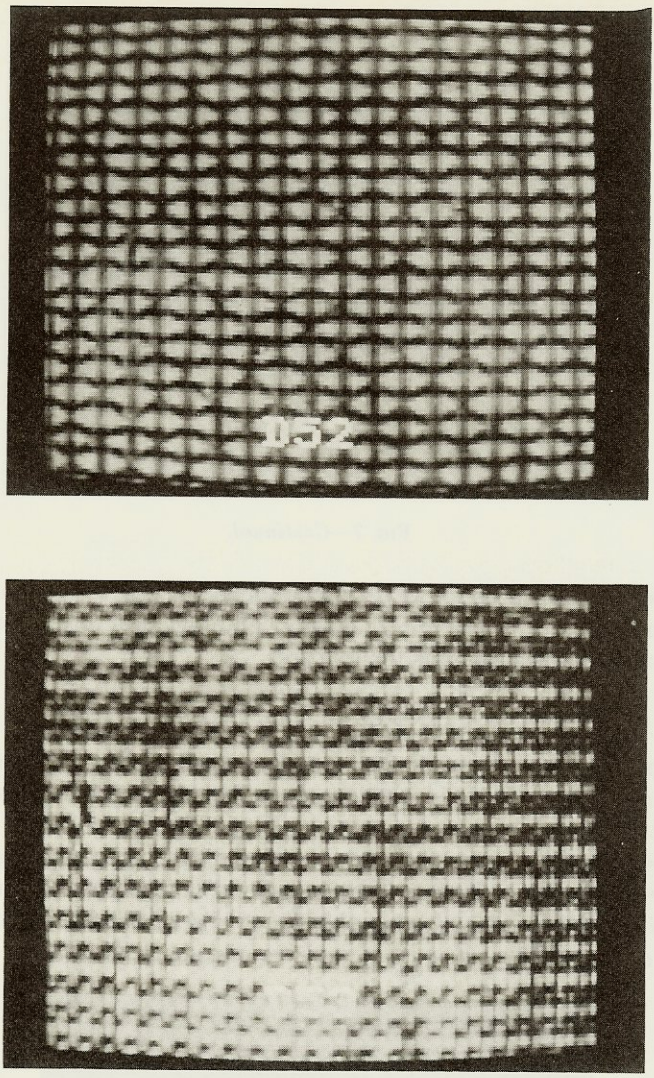

FIG. 7—*Continued.*

The measure of texture structure captured by co-occurrence matrices, a χ^2 statistic, was used to select such matrices for classification. This implies that the feature vectors for different samples in a particular classification experiment may be computed from matrices derived from different spatial relationships. An important variation on the technique presented here is to actually use the chosen $\bar{d}$ values themselves as features for classification. Such a feature should separate magnified images of identical textures very well, and should also reflect size variations between texture primitives.

The χ^2 measure of texture structure is only one possibility for quantifying the association between variables in contingency tables. Goodman and Kruskal [8] have characterized several other measures of association which should also be

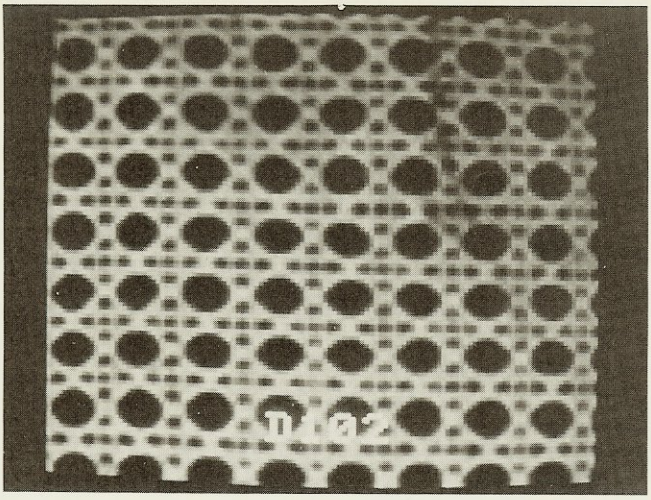

FIG. 7—*Continued.*

examined, both as features for texture classification and as measures of texture structure. Such measures may provide the beginnings of a formal bridge between statistical and structural models for texture.

APPENDIX: A LINEAR DISCRIMINANT CLASSIFIER

In this appendix we describe the classifier used in our texture experiments. It is a minimum Mahalanobis distance classifier that computes a set of linear discriminant functions.

Let ω_i, $i = 1, 2, \ldots, c$, be the set of classes and $\bar{X}$ be a sample (i.e., feature vector). The decision rule divides feature space into c decision regions $R_1, \ldots, R_c$. Let $g_i(\bar{X})$, $i = 1, 2, \ldots, c$, denote the set of discriminant functions. If $g_i(\bar{X}) > g_j(\bar{X})$ for all $i \neq j$ then $\bar{X}$ is in R_i and the decision rule assigns $\bar{X}$ to class ω_i.

FIG. 8. LANDSAT terrain images used in the second classification experiment. A typical image is shown from each of the three groups.

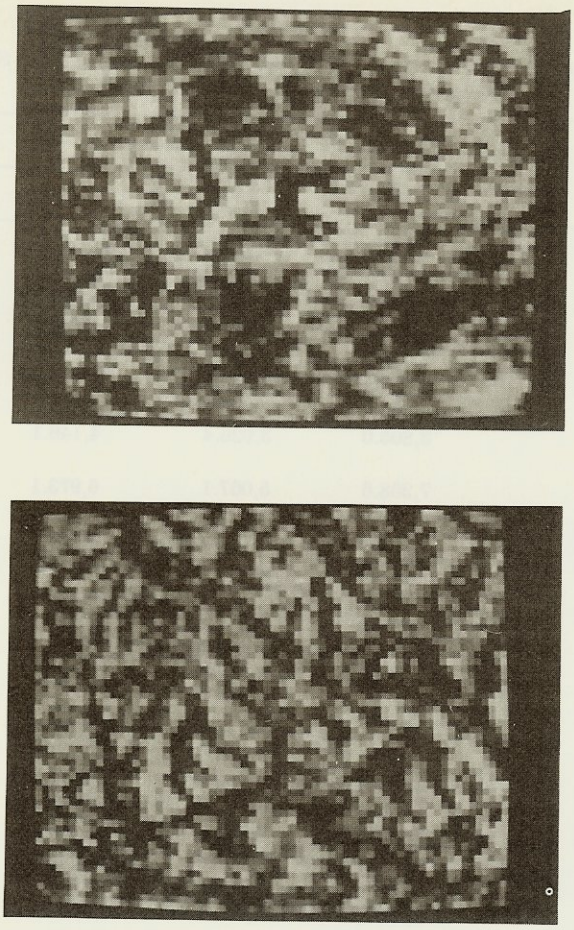

FIG. 8—*Continued.*

It can be shown [4] that minimum error rate classification can be obtained with the discriminant functions

$$g_i(\bar{X}) = \log p(\bar{X} \mid \omega_i) + \log P(\omega_i),$$

where $p(\bar{X} \mid \omega_i)$ is the likelihood of ω_i with respect to $\bar{X}$, and $P(\omega_i)$ is the a priori probability of samples falling into ω_i. To evaluate the functions, we assume that $p(\bar{X} \mid \omega_i)$ are multivariate normal densities. Furthermore, we assume that the covariance matrices are identical for all the classes (i.e., $\Sigma_i = \Sigma$). Thus, geometrically, the clusters for all classes are hyperellipses of equal size and shape, with the cluster of the ith class centered about mean vector $\overline{\mu_i}$. Under these conditions, the discriminant functions are linear and the resulting decision boundaries are hyperplanes.

Suppose that there are c classes and that the classifier is to be trained on a pool of N samples, with n_k, $k = 1, 2, \ldots, c$, samples belonging to each class. In addition, suppose that X_i is the ith feature of $\bar{X}$ and that $\bar{X}$ has length d. For each

TABLE 5

χ^2 Values for Co-Occurrence Matrices, Computed over Several Spatial Relationships for the LANDSAT Images[a]

Distance	Angles			
	0°	45°	90°	135°
1	8,795.4	6,737.6	11,994.6	6,998.7
2	5,026.1	4,641.0	5,577.7	4,267.7
3	4,246.3	3,897.1	4,904.9	4,315.7
4	4,103.0	4,009.3	4,497.7	4,177.0
5	4,023.4	3,894.6	4,438.9	4,089.9
6	4,079.3	4,066.3	4,172.4	4,242.1
7	4,003.6	3,879.2	4,106.7	4,006.4
8	3,805.0	3,926.4	4,146.1	4,201.5
1	7,268.6	5,067.1	6,973.1	4,949.7
2	4,660.9	4,131.5	3,985.6	4,002.5
3	4,112.1	3,744.2	4,257.4	3,974.1
4	4,056.8	4,035.2	4,356.5	4,037.5
5	3,970.7	3,950.1	4,143.4	4,039.1
6	4,152.3	3,968.3	4,488.4	4,013.9
7	4,041.1	4,043.4	4,018.1	3,928.8
8	3,958.3	4,043.3	3,892.7	4,121.6
1	13,628.7	9,618.6	17,446.3	11,182.7
2	7,282.4	5,560.0	7,814.8	6,275.2
3	5,858.3	4,450.6	5,486.4	4,861.8
4	4,988.2	4,329.8	4,616.3	4,390.2
5	4,473.5	4,309.3	4,305.2	4,265.8
6	4,419.1	4,185.8	4,073.6	4,229.1
7	3,904.4	4,293.5	4,185.6	4,085.2
8	3,863.1	4,300.1	4,458.9	4,186.0

[a] The tables were generated from the images in Fig. 8, in the order of their appearance.

class, $k = 1, 2, \ldots, c$, we compute the means

$$\mu_j^k = \frac{1}{n_k} \sum_{i=1}^{n_k} X_{ij}^k, \qquad j = 1, 2, \ldots, d,$$

and the matrix of cross products of deviations from the means

$$D^k = \{d_{jl}^k\} = \sum_{i=1}^{n} (X_{ij}^k - \mu_j^k)(X_{il}^k - \mu_l^k), \qquad j, l = 1, 2, \ldots, d.$$

The pooled covariance matrix is then

$$\Sigma = \{\sigma_{ij}\} = \sum_{k=1}^{c} D^k / ((\sum_{k=1}^{c} n_k) - c).$$

Let σ_{ij}^{-1} be the (i, j)th element of Σ^{-1}. The coefficients of the linear discriminant

functions are then given by

$$W_i^k = \sum_{j=1}^{d} \sigma_{ij}^{-1} \mu_j^k, \qquad i = 1, 2, \ldots, d,$$

and the constant term by

$$W_0^k = -\tfrac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \sigma_{ij}^{-1} \mu_i^k \mu_j^k.$$

Thus, the discriminant functions are

$$g_k(\bar{X}) = W_0^k + \sum_{i=1}^{d} W_i^k X_i^k.$$

The decision rule assigns $\bar{X}$ to the class yielding the largest discriminant function. The confidence in the classification is given by

$$P_L = 1 / \sum_{k=1}^{c} e^{(g_k - g_L)},$$

where L is the class whose discriminant function gives the largest value, g_L, for $\bar{X}$.

One may obtain a general measure of the usefulness of a set of discriminant functions obtained from a particular training set by computing a generalized Mahalanobis D^2 statistic. Let m_i, $i = 1, 2, \ldots, d$, be the common means for all c groups.

$$m_i = \sum_{k=1}^{c} n_k \mu_i^k / \sum_{k=1}^{c} n_k.$$

The statistic is given by

$$D^2 = \sum_{i=1}^{d} \sum_{j=1}^{d} \sigma_{ij}^{-1} \sum_{k=1}^{c} n_k (\mu_i^k - m_i)(\mu_j^k - m_j).$$

D^2 can be used as a chi square (under the assumption of normality), with $m(k-1)$ degrees of freedom, to test the hypothesis that the d feature mean values are the same in all c groups (i.e., the groups are nonseparable).

REFERENCES

1. Y. M. M. Bishop, S. E. Feinberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, Mass., 1976.
2. P. Brodatz, *Textures: A Photographic Album*, Dover, New York, 1966.
3. L. S. Davis, S. Johns, and J. K. Aggarwal, Texture analysis using generalized co-occurrence matrices, in *Proc. IEEE Conf. on Pattern Recognition and Image Processing, Chicago 1978.*
4. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
5. R. M. Haralick, Statistical and structural approaches to texture, *Proc. IEEE* **67**, 1979, 786–804.
6. R. M. Haralick, K. Shanmugam, and I. Dinstein, Textural features for image classification, *IEEE Trans. Systems Man Cybernet.* **SMC-3**, 1973, 610–621.
7. P. G. Hoel, *Introduction to Mathematical Statistics*, 4th ed., Wiley, 1971.
8. L. A. Goodman and W. H. Kruskal, Measures of association for cross classifications, I–IV,

J. Amer. Statist. Assoc. **49**, 1954, 732–764; **54**, 1959, 123–163; **58**, 1963, 310–364; **67**, 1972, 415–421.

9. R. L. Plackett, *The Analysis of Categorical Data*, Griffin, London, 1974.
10. J. Weszka, C. R. Dyer, and A. Rosenfeld, A comparative study of texture measures for terrain classification, *IEEE Trans. Systems Man Cybernet.* **SMC-6**, 1976, 269–285.
11. S. W. Zucker, Toward a Model of Texture, *Computer Graphics Image Processing* **5**, 1976, 190–202.