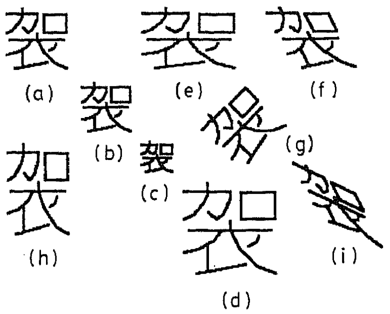

Fig. 13. Examples of pattern transformation for the " 袈 " pattern.

where

$$u_{ij} = z_1 \cdot x_{ij} + z_2 \cdot y_{ij} + x_0$$

$$v_{ij} = z_3 \cdot x_{ij} + z_4 \cdot y_{ij} + y_0.$$

This function is a two-dimensional affine transformation and it may be almost enough for computer graphics of character patterns. Fig. 12 shows some examples of this transformation. In the figure, the (a) pattern has a 60×60 pixels of initial region and is supposed to be original. These patterns are generated by operating the following parameters, respectively:

$$\text{(a)}: \begin{pmatrix} 1.0 & 0 \\ 0 & 1.0 \end{pmatrix} \quad \text{(b)}: \begin{pmatrix} 0.8 & 0 \\ 0 & 0.8 \end{pmatrix} \quad \text{(c)}: \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$$

$$\text{(d)}: \begin{pmatrix} 1.5 & 0 \\ 0 & 1.5 \end{pmatrix} \quad \text{(e)}: \begin{pmatrix} 1.0 & 0 \\ 0 & 1.5 \end{pmatrix} \quad \text{(f)}: \begin{pmatrix} 1.0 & 0 \\ 0.5 & 1.0 \end{pmatrix}$$

$$\text{(g)}: \begin{pmatrix} \cos 45 & -\sin 45 \\ \sin 45 & \cos 45 \end{pmatrix} \quad \text{(h)}: \begin{pmatrix} 1.5 & 0 \\ 0 & 1.0 \end{pmatrix} \quad \text{(i)}: \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}.$$

In the figure, patterns from (a) to (e) and (h) can also be generated by changing their initial regions, but the rest cannot be generated without the transformation. While pattern sizes from (a) to (d) are changed from 30×30 to 90×90 pixels, they each keep their form. If we consider the linewidth according to their sizes, we can get more natural patterns. In these transformations, the parameters except $z_i (i = 1 - 4)$ are kept constant. If division parameters a_1, a_2, a_3 and block parameters t, d are varied, pattern shapes also change.

Our generation system has two functions, that is, generation and registration. Even if a pattern is not encoded in the dictionary, we can generate the pattern directly by inputting code strings of its subpatterns and register it in the dictionary. Some applications such as graphical pattern generation do not need the registration of all the character patterns.

V. Concluding Remarks

We reported a coding and generating method of structural characters, especially Chinese and Korean characters. In the method, we defined three types of blocks, i.e., fundamental blocks, composite blocks, and a special block called the null block. If fundamental blocks which correspond to strokes are well defined, this algorithm can be applied to other structural characters. This pattern encoding method is not so far from linguistic concepts of Chinese or Korean characters, and a person who does not know this algorithm will be able to encode a pattern or modify code strings with a little training. Also, the generation program itself is very compact because of the recursiveness of the algorithm. Then it can be executed on most of the low cost microcomputers, and can be applied to printing, education, and computer graphics.

References

[1] M. Nisimaki *et al.*, Special issue on optical character recognition, *J. Inst. Electron. Commun. Eng. Japan*, vol. 61, no. 2, pp. 111–135, Feb. 1978.
[2] W. Stallings, "Chinese character recognition," in *Syntactic Pattern Recognition, Application.* New York: Academic, 1977, ch. 5, pp. 95–123.
[3] T. Agui *et al.*, "A method of recognition and representation of Korean characters by tree grammar," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, pp. 245–251, July 1979.
[4] S. Mori *et al.*, "Research on machine recognition of handprinted characters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, pp. 386–405, July 1984.
[5] K. S. Fu, *Syntactic Method in Pattern Recognition.* New York: Academic, 1974.
[6] ——, "Introduction to syntactic pattern recognition," in *Syntactic Pattern Recognition, Application.* New York: Academic, 1977, ch. 1, pp. 1–28.
[7] B. Rankin, "A linguistic study of formation of Chinese characters," PhD. dissertation, Univ. Pennsylvania, Philadelphia, 1965.
[8] B. Rankin and S. Siegal, "A grammar for component combination in Chinese characters," NBS Tech. Note 296, 1966.
[9] W. Stallings, "The morphology of Chinese characters: A survey of models and applications," *Comput. Humanities*, vol. 9, pp. 13–24, 1975.
[10] S. K. Chang, "An interactive system for Chinese character generation and retrieval," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-3, pp. 257–265, May 1973.
[11] E. M. Huang and C. Y. Suen, "Computational analysis of the structural compositions of a frequently used Chinese character set," in *Proc. Int. Conf. Text Processing Large Character Set (ICTP'83)*, in Japan, Oct. 1983, pp. 292–297.
[12] O. Fujimura and R. Kagaya, "Structural patterns of Chinese characters," *Annu. Bull. Res. Inst. Logopedics Phoniatrics*, Univ. Tokyo, vol. 3, pp. 131–148, 1969.
[13] H. Nagahashi *et al.*, "A pattern generation method of Chinese characters," *Trans. Inst. Electron. Commun. Eng., Japan*, vol. E65, pp. 607–613, 11, Nov. 1982.
[14] T. Agui and H. Nagahashi, "A coding method of Chinese characters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, 4, pp. 333–341, Oct. 1979.
[15] H. Nagahashi *et al.*, "Pattern generation of Chinese characters in terms of graphics," in *Proc. Int. Conf. Text Processing Large Character Set (ICTP'83) in Japan*, Oct. 1983, pp. 281–286.
[16] K. Ogawa *et al.*, "On a size and type transformation of Kanji pattern," *Trans. Inst. Electron. Commun. Eng.*, Japan, vol. J65-D, pp. 234–241, Feb. 1982.

Sum and Difference Histograms for Texture Classification

MICHAEL UNSER

***Abstract*—The sum and difference of two random variables with same variances are decorrelated and define the principal axes of their associated joint probability function. Therefore, sum and difference histograms are introduced as an alternative to the usual co-occurrence matrices used for texture analysis. Two maximum likelihood texture classifiers are presented depending on the type of object used for texture characterization (sum and difference histograms or some associ-**

Manuscript received June 10, 1983; revised May 27, 1985. Recommended for acceptance by S. W. Zucker.

The author was with the Signal Processing Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland. He is now with the Biomedical Engineering and Instrumentation Branch, National Institutes of Health, Bethesda, MD 20892.

IEEE Log Number 8405366.

ated global measures). Experimental results indicate that sum and difference histograms used conjointly are nearly as powerful as co-occurrence matrices for texture discrimination. The advantage of the proposed texture analysis method over the conventional spatial gray level dependence method is the decrease in computation time and memory storage.

***Index Terms*—Classification, co-occurrence matrices, image processing, texture.**

I. Introduction

Texture is the term used to characterize the surface of a given object or phenomenon and is undoubtedly one of the main features used in image processing and pattern recognition. Texture is essentially a neighborhood property. Recent work by Haralick provides a comprehensive survey of most existing structural and statistical approaches to texture [1].

The Spatial Grey Level Dependence Method (SGLDM) [2] is certainly one of the most powerful statistical texture analysis algorithms. It is based on the estimation of the joint probability functions of two picture elements in some given relative position (co-occurrence or spatial grey level dependence matrices). The justification for the use of this type of characterization is given by experiments on human texture perception indicating that second-order probabilities of the form measured by the co-occurrence matrices play an important role in human texture discrimination [3]–[6]. The spatial gray level dependence matrices are mostly used as intermediate matrices and dimensionality reduction is performed in computing features of the type described in [2]. Nevertheless, it has been shown recently [7], [8] that this last step can be avoided and that the co-occurrence matrices can be used directly as the input features of an efficient texture classifier.

During the past years, several new texture models have been introduced and have been claimed to be superior to co-occurrence matrices for texture analysis. The first is Laws' model based on texture energy measures, which has been described in [9]. The second is the Gaussian Markov random field model, which has been studied by R. Kashyap and others [10]. Although these approaches have been found to be superior to the SGLDM, in some restricted cases, no quick conclusions should be made: the two forementioned methods extract statistical information that can be shown to be completely contained in the spatial autocorrelation or covariance function (second moments) of the underlying texture fields. As an immediate consequence, these approaches should be less performant than the SGLDM (which includes the feature "correlation") when all possible pairwise configuration in the domain of interest are considered. However, in such a case, the dimensionality of the feature vector, when using SGLDM, would be significantly more important than for the other methods.

The main drawback to using the SGLDM is the large memory requirement for storing the co-occurrence matrices. One sometimes has the paradoxical situation in which the objects (SGLD matrices) used for texture characterization are more voluminous than the original images from which they are derived. It is also clear that because of their large dimensionality, the co-occurrence matrices are very sensitive to the size of the texture samples on which they are estimated.

The purpose of this paper is to present an alternative to usual SGLDM which is nearly as accurate for classification and which reduces the memory requirement. A co-occurrence matrix is replaced by estimates of the first order probability functions along its principal axes, namely the sum and difference histograms. Different texture features are introduced. In Section IV, two maximum likelihood or Bayesian texture classifiers are presented. The first classifier uses the different histograms directly as features and leads to an extremely fast implementation. The second classifier is more conventional and uses as features a set of global measures computed from the sum and difference histograms. Finally, experiments indicate that this texture analysis method performs nearly as well as the SGLDM.

II. Second-Order Texture Description

A discrete texture image defined on a $K \times L$ rectangular grid is denoted by $\{y_{k,l}\}$, $(k = 1, \cdots, K;\ l = 1, \cdots, L)$ and is considered to be a realization of a bidimensional stationary and ergodic process. Let $G = \{1, 2, \cdots, N_g\}$ be the set of the N_g quantized grey levels. It is common to characterize the spatial organization of the picture elements by a set of second-order statistics. Consider two picture elements in a relative position fixed by (d_1, d_2):

$$\begin{cases} y_1 = y_{k,l} \\ y_2 = y_{k+d_1, l+d_2}. \end{cases} \tag{1}$$

The relative displacement (d_1, d_2) may be equivalently characterized by a distance d in radial units and an angle θ with respect to the horizontal axis. The discrete joint probability function (PF) of these two random variables is $P(y_1, y_2)$. The probability of observing the gray levels i and j at a fixed relative position specified by (d_1, d_2) is

$$\text{Prob}\,\{y_{k,l} = i,\ y_{k+d_1,l+d_2} = j\} = P(i, j; d_1, d_2) = P(i, j) \tag{2}$$

and does not depend on the absolute indexes (k, l).

A. Spatial Gray Level Dependence Matrix

Let D be a subset of indexes specifying a texture region to be analyzed. The spatial gray level co-occurrence or dependence matrix with parameters (d_1, d_2) is defined as [2]

$$c(i, j; d_1, d_2) = c(i, j) = \text{Card}\,\{(k, l) \in D,\ y_{k,l} = i \ \text{and}\ y_{k+d_1,l+d_2} = j\}. \tag{3}$$

The total number of counts is

$$N = \text{Card}\,\{D\} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} c(i, j) \tag{4}$$

where Card { } refers to the number of elements of a set. The normalized co-occurrence matrix is an estimate of the joint PF defined by (2); one has

$$\hat{P}(i, j) = c(i, j)/N \simeq P(i, j). \tag{5}$$

B. Principal Axis

The covariance matrix associated with the pair of random variables y_1 and y_2 is given by

$$C_y = \sigma_y^2 \cdot \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{6}$$

where

$$\sigma_y^2 \cdot \rho = E\{(y_1 - \mu) \cdot (y_2 - \mu)\} \ \text{and}\ \mu = E\{y_1\} = E\{y_2\}. \tag{7}$$

Because of stationarity, one has that

$$E\{(y_1 - \mu)^2\} = E\{(y_2 - \mu)^2\} = \sigma_y^2. \tag{8}$$

The eigenvalues and eigenvectors of the matrix C_y are solution of the matrix equation

$$C_y \cdot u = \lambda \cdot u. \tag{9}$$

These quantities are given by

$$\begin{cases} \lambda_1 = \sigma_y^2 \cdot (1 + \rho) \\ \lambda_2 = \sigma_y^2 \cdot (1 - \rho) \end{cases} \tag{10}$$

and

$$u_1 = \left[\frac{1}{\sqrt{2}}\ \frac{1}{\sqrt{2}}\right]^T \ \text{and}\ u_2 = \left[\frac{1}{\sqrt{2}}\ \frac{-1}{\sqrt{2}}\right]^T. \tag{11}$$

The eigenvectors are the axes of inertia of the joint probability function $P(y_1, y_2)$ and the eigenvalues are the variances along those axes. The vectors u_1 and u_2 are the eigenvectors of any 2×2 covariance matrix where $\mathrm{Var}\{y_1\} = \mathrm{Var}\{y_2\}$. Therefore, the sum and difference define the principal axes of any second-order PF of a stationary process. The linear transform

$$\begin{cases} z_1 = (y_1 + y_2)/\sqrt{2} \\ z_2 = (y_1 - y_2)/\sqrt{2} \end{cases} \tag{12}$$

will produce two random variables z_1 and z_2 which are decorrelated. One has the following properties:

$$\mathrm{Var}\{z_1\} = \lambda_1 = \sigma_y^2 \cdot (1 + \rho) \tag{13}$$

$$\mathrm{Var}\{z_2\} = \lambda_2 = \sigma_y^2 \cdot (1 - \rho) \tag{14}$$

$$\mathrm{Covar}\{z_1 \cdot z_2\} = E\{(z_1 - E\{z_1\}) \cdot (z_2 - E\{z_2\})\} = 0. \tag{15}$$

C. Approximation of Second-Order Probabilities

When the uncorrelated random variables z_1 and z_2 are also independent, the joint probability function can be computed from

$$P(y_1, y_2) = P(z_1, z_2) = P_s(z_1) \cdot P_d(z_2). \tag{16}$$

This expression is always true for Gaussian random variables. For arbitrarily distributed random variables, uncorrelatedness is a necessary but not sufficient condition for independence. The last equality will therefore not always be satisfied. Nevertheless, the product of the first order PF's along the principal axis can still be used as a close approximation of the joint PF.

$$\hat{P}_y(i, j) = c_0 \cdot P_s(i + j) \cdot P_d(i - j) \simeq P_y(i, j) \tag{17}$$

where c_0 is a normalization constant chosen in order to guarantee that

$$\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \hat{P}_y(i, j) = 1 \tag{18}$$

when the support of (i, j) is finite. This condition reflects the constraint that the pairs (i, j) are only defined in the square domain shown in Fig. 1. It takes into account the fact that the product $P_s(i + j) \cdot P_d(i - j)$ is not necessarily zero outside this region.

The degree of deviation of the true probability distribution $P_y(i, j)$ from its approximation $\hat{P}_y(i, j)$, $(i, j = 1, \cdots, N_g)$, can quantitatively be measured by

$$\begin{aligned} I(P, \hat{P}) &= \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_y(i, j) \cdot \log \{P_y(i, j)/\hat{P}(i, j)\} \geq 0 \qquad (19) \\ &= \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_y(i, j) \cdot \log \{P_y(i, j)/(P_s(i + j) \cdot \\ &\quad \cdot P_d(i - j))\} - \log \{c_0\} \geq 0 \\ &= H_s + H_d - H_y - \log \{c_0\} \geq 0 \qquad (20) \end{aligned}$$

where H_y, H_s, and H_d are the entropy measures defined by

$$H_y = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_y(i, j) \cdot \log \{P_y(i, j)\} \tag{21}$$

$$H_s = -\sum_{k=2}^{2N_g} P_s(k) \cdot \log \{P_s(k)\}$$

and

$$H_d = -\sum_{l=-N_g+1}^{N_g-1} P_d(l) \cdot \log \{P_d(l)\}. \tag{22}$$

The relative entropy $I(P, \hat{P})$ is a measure of interdependence [11, p. 59] and is equal to zero if and only if $P(i, j) = \hat{P}(i, j)$, for $ij = 1, \cdots, N_g$, that is, when the random variables z_1 and z_2 are independent. The similarity measure between the sets of probabilities $\{P(i, j)\}$ and $\{\hat{P}(i, j)\}$ is sometimes called the directed divergence

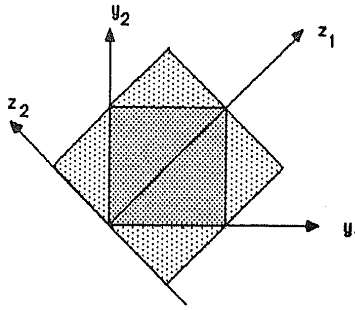

Fig. 1. Geometric representation of the sum and difference transformation in the space of y_1 and y_2.

or the Kullback information and has some interesting properties [12], [13]. The practical interest of (20) is that it is possible to test the hypothesis of independence in simply comparing the sum and difference entropies to the entropy of the associated co-occurrence matrix. The closer $I(P, \hat{P})$ is to zero, the better is the approximation defined by (17).

III. Texture Features

It was shown that the sum and difference define the principal axes of the second-order PF of a stationary process. Therefore, it is suggested that the usual co-occurrence matrices used for texture description be replaced by their associated sum and difference histograms which can be estimated directly from the image. Difference statistics on their own have already been used for texture analysis and have been compared to other methods [14], [15]. In this study, the complementarity of sum and difference statistics is emphasized. Their use is justified by the approximation of a joint PF by a combination of first order statistics of uncorrelated transformed variables.

In the following text, the nonnormalized sum and difference, associated with the relative displacement (d_1, d_2), are defined as

$$\begin{cases} s_{k,l} = y_{k,l} + y_{k+d_1,l+d_2} \\ d_{k,l} = y_{k,l} - y_{k+d_1,l+d_2}. \end{cases} \tag{23}$$

Thus, the dynamic range of the sum and difference is generally twice the range of the original image.

A. Sum and Difference Histograms

The sum and difference histograms with parameters (d_1, d_2) over the domain D are defined in a manner very similar to the spatial grey level co-occurrence or dependence matrix definition:

$$h_s(i; d_1, d_2) = h_s(i) = \mathrm{Card}\{(k, l) \in D, s_{k,l} = i\} \tag{24}$$

$$h_d(j; d_1, d_2) = h_d(j) = \mathrm{Card}\{(k, l) \in D, d_{k,l} = j\}. \tag{25}$$

As before, the total number of counts is

$$N = \mathrm{Card}\{D\} = \sum_i h_s(i) = \sum_j h_d(j). \tag{26}$$

The normalized sum and differences histograms

$$\hat{P}_s(i) = h_s(i)/N; \qquad (i = 2, \cdots, 2N_g) \tag{27}$$

$$\hat{P}_d(j) = h_d(j)/N; \qquad (j = -N_g + 1, \cdots, N_g - 1) \tag{28}$$

are estimates of the sum and difference PF's defined by

$$P_s(i) = \mathrm{prob}\{s_{k,l} = i\}; \qquad (i = 2, \cdots, 2N_g) \tag{29}$$

$$P_d(j) = \mathrm{prob}\{d_{k,l} = j\}; \qquad (j = N_g + 1, \cdots, N_g - 1). \tag{30}$$

B. Global Histogram Features

In many applications, it may be necessary to reduce the dimensionality of the set of characteristics used for texture description. Statistical information can be extracted from the histograms by

TABLE I
EXAMPLES OF GLOBAL HISTOGRAM FEATURES

FEATURE	FORMULA
mean	$\mu = \sum_i i \cdot \hat{P}(i)$
variance	$\sigma^2 = \sum_i (i-\mu)^2 \cdot \hat{P}(i)$
q^{th} moment about the mean	$m_q = \sum_i (i-\mu)^q \cdot \hat{P}(i)$
entropy	$H = \sum_i -\hat{P}(i) \cdot \log(\hat{P}(i))$

TABLE II
EXAMPLES OF THE MOST COMMONLY USED TEXTURE FEATURES COMPUTED FROM THE CO-OCCURRENCE MATRICES AND THEIR EQUIVALENT FORM COMPUTED FROM THE SUM AND DIFFERENCE HISTOGRAMS

TEXTURE FEATURE	CO-OCCURRENCE MATRIX	SUM AND DIFFERENCE HISTOGRAMS
mean	$f_1 = \sum_i \sum_j i \cdot \hat{P}(i,j)$	$= \frac{1}{2} \sum_i i \cdot \hat{P}_s(i) = \mu$
variance	$f_2 = \sum_i \sum_j (i-\mu)^2 \cdot \hat{P}(i,j)$	$= \frac{1}{2}\left(\sum_i (i-2\mu)^2 \cdot \hat{P}_s(i) + \sum_j j^2 \cdot \hat{P}_d(j) \right)$
energy	$f_3 = \sum_i \sum_j \hat{P}(i,j)^2$	$\simeq \sum_i \hat{P}_s(i)^2 \cdot \sum_j \hat{P}_d(j)^2$
correlation	$f_4 = \sum_i \sum_j (i-\mu) \cdot (j-\mu) \cdot \hat{P}(i,j)$	$= \frac{1}{2}\left(\sum_i (i-2\mu)^2 \cdot \hat{P}_s(i) - \sum_j j^2 \cdot \hat{P}_d(j) \right)$
entropy	$f_5 = \sum_i \sum_j -\hat{P}(i,j) \cdot \log(\hat{P}(i,j))$	$\simeq -\sum_i \hat{P}_s(i) \cdot \log(\hat{P}_s(i)) - \sum_j \hat{P}_d(j) \cdot \log(\hat{P}_d(j))$
contrast	$f_6 = \sum_i \sum_j (i-j)^2 \cdot \hat{P}(i,j)$	$= \sum_j j^2 \cdot \hat{P}_d(j)$
homogeneity	$f_7 = \sum_i \sum_j \frac{1}{1+(i-j)^2} \cdot \hat{P}(i,j)$	$= \sum_j \frac{1}{1+j^2} \cdot \hat{P}_d(j)$
cluster shade	$f_8 = \sum_i \sum_j (i+j-2\mu)^3 \cdot \hat{P}(i,j)$	$= \sum_i (i-2\mu)^3 \cdot \hat{P}_s(i)$
cluster prominence	$f_9 = \sum_i \sum_j (i+j-2\mu)^4 \cdot \hat{P}(i,j)$	$= \sum_i (i-2\mu)^4 \cdot \hat{P}_s(i)$

computing quantities such as the mean, the variance and the entropy. Table I gives a list of features that can be computed from both sum and difference histograms.

Haralick [2] has proposed a variety of measures that can be employed to extract useful textural inforamation from the spatial co-occurrence matrices. It is suggested here to evaluate these features directly from the sum and difference in order to simplify the computation. Consider the set of 14 features introduced in [2]; nine of them can be computed exactly from the sum and difference histograms while the remaining five can be approximated very closely assuming mutual independence between s and d (for example: the energy and the entropy). Table II shows the equivalence between the most widely used features computed from the co-occurrence matrices and those computed from their associated sum and difference histograms. Using this approach the gain in computation time (a double summation is replaced by a simple summation) and memory requirement can be important depending on intensity resolution (a factor $N_g/4$). Consider the case of an image quantified with $N_g = 32$ gray levels. The use of the proposed method will enable a computation eight times faster than the previously mentioned features and will need memory eight times smaller than the usual evaluation based on the co-occurrence matrices.

IV. CLASSIFICATION

In this section two maximum likelihood or Bayesian classifiers will be presented. They differ in the type of features that are used and in the assumptions that are made concerning the parametric family of class conditional probability density functions of the feature vector associated with different types of textures.

Consider the problem of classifying a given texture sample, with associated N-dimensional feature vector $\boldsymbol{x}$ according to K texture classes ω_i. $(i = 1, \cdots, K)$. For simplicity, it is assumed that the prior probabilities of $\boldsymbol{x}$ belonging to the different classes are all the same. A class ω_i is characterized by the conditional probability density function $f(\boldsymbol{x}|\omega_i)$. The maximum likelihood or Bayesian decision rule that minimizes the total probability of false classification is given by [16]

$$f(\boldsymbol{x}|\omega_j) = \text{Max}\,\{f(\boldsymbol{x}|\omega_i\}, \qquad (i = 1, \cdots, K) \rightarrow \text{choose class } \omega_j. \tag{31}$$

The optimum maximum likelihood criterion requires knowledge of the class conditional probability density functions $f(\boldsymbol{x}|\omega_i)$. For practical reasons, a parametric representation of these functions has to be chosen. Two different cases are investigated next.

A. Sum and Difference Histograms

It is possible to characterize a particular texture sample by a collection of sum and difference histograms that have been estimated for different relative displacements d_1 and d_2. In the following development, the case of a unique histogram is considered. Subsequent results can easily be adapted for the case where more than one histogram is used, assuming mutual independence.

Let $\{h(i)\}$, $(i = 1, \cdots, N)$ represent a particular histogram that has been chosen for texture representation. The different counts $h(i)$ can be presented as an N-dimensional feature vector

$$\boldsymbol{x} = [h(1)\ h(2) \cdots h(N)]^T = [x_1 x_2 \cdots x_N]^T. \tag{32}$$

It is assumed that the components of this feature vector are distributed according to a multinomial distribution

$$f(\boldsymbol{x}_j) = \left(\sum_{i=1}^{N} x_i\right)! \prod_{i=1}^{N} P_j(i)^{x_i}/x_i! \tag{33}$$

where $P_j(i)$ is the probability of having the particular sum or difference equal to i on the texture of type j. This assumption is equivalent to considering that the counting process can be seen as the result of N independent generalized Bernoulli trials [17, pp. 74–76] where the mutually exclusive events are the observation of a sum or difference of a given value. Because of the stationarity and ergodicity of the underlying random process, this assumption becomes more and more realistic as the size of the estimation domain increases.

It can be shown that the quantity $\hat{P}(i) = x_i/N$, computed on texture j, is the maximum likelihood estimate of $P_j(i)$. Therefore, the parameters of distributions $f(x|\omega_i)$, $(i = 1, \cdots, K)$, can be estimated in a learning phase on texture samples of sufficient size. Taking the logarithm of (33) and neglecting the terms that are common to the different PDF's, the maximum likelihood decision rule (31) can be shown to be equivalent to

$$U_i(x) = \min\{U_j(x)\}, \qquad (j = 1, \cdots, K) \rightarrow \text{choose class } \omega_i \tag{34}$$

where $U_j(x)$ is given by

$$U_j(x) = -\sum_{i=1}^{N} x_i \cdot \log\{P_j(i)\}. \tag{35}$$

One has to remember that $x_i = h(i)$ represents the number of occurrences of a sum or difference equal to i over the domain of interest D. Therefore, using (24) or (25), it is possible to replace the summation in the feature space by a summation over the spatial domain.

$$U_j(x) = \sum\sum_{(k,l)\in D} -\log\{P_j(y_{k,l} \pm y_{k+d_1,l+d_2})\}. \tag{36}$$

Applying this equation, the whole process can be accomplished by a simple table lookup procedure. Optimum classification is performed with only K additions per pixel without an explicit evaluation of the difference or sum histograms. It is straightforward to extend the method when using more than one histogram and assuming mutual independence. The resulting decision function is obtained by simple summation of the individual contributions. The block diagram of a system performing the classification of $(M \times N)$ texture samples is depicted in Fig. 2. This texture classifier offers multiple advantages. First of all, it is computationally very attractive. The decision functions $U_j(x)$ can be computed in approximately the same time as would be required to estimate the histograms explicitly. Secondly, there is absolutely no loss of information of the kind that would result from some heuristic dimensionality reduction.

B. Global Features

A texture sample can be characterized by a set of global features as described in Section III-B, computed from sum and difference histograms for different relative positions d_1 and d_2. As before, these values are ordered in an N-dimensional feature vector $x = [x_1 \cdots x_N]^T$. It is convenient to assume that the feature vector is distributed according to a multivariate normal distribution

$$f(x|\omega_j) = (2\pi)^{N/2}\,|R_j|^{-1/2} \cdot \exp\{-\tfrac{1}{2}(x - m_j)^T \cdot R_j^{-1} \cdot (x - m_j)\} \tag{37}$$

where m_j is the mean vector and R_j the covariance matrix associated to the class ω_j:

$$m_j = E\{x|\omega_j\} \quad \text{and} \quad R_j = E\{(x - m_j)(x - m_j)^T|\omega_j\}. \tag{38}$$

It is possible to compute these parameters in a learning procedure on a sufficiently large number of texture samples. Let $\{x_i\}$, $(i = 1, \cdots, M)$ be the M feature-vectors associated with a set of reference texture samples of type j. The maximum likelihood estimates of the distribution parameters are given by [16]

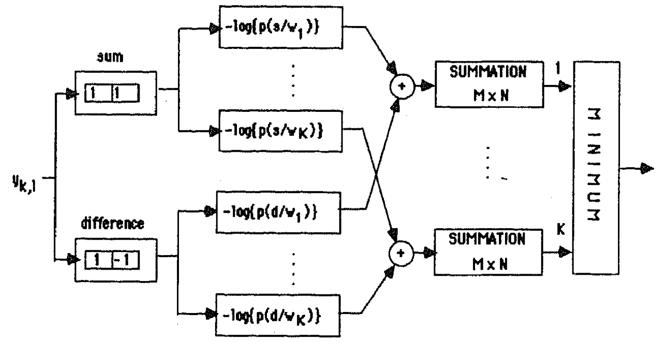

Fig. 2. Block diagram of a texture classifier using sum and difference histograms as features.

$$\hat{m}_j = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{39}$$

$$\hat{R}_j = \frac{1}{M}\sum_{i=1}^{M} (x_i - \hat{m}_j)(x_i - \hat{m}_j)^T. \tag{40}$$

Justification for the use of a Gaussian model is given by the particular family of texture features that have been chosen. Most of them can be expressed as a spatial average over the domain D. For this reason, one can expect their individual first-order statistics to be almost Gaussian. The multivariate Gaussian probability density function is also the maximum entropy distribution that satisfies the constraints given by (38). In this situation, the maximum likelihood decision rule (31) can be shown to be equivalent to

$$D_i(x) = \min\{D_j(x)\}, \qquad (j = 1, \cdots, K) \rightarrow \text{choose class } \omega_i \tag{41}$$

where $D_j(x)$ is given by

$$D_j(x) = (x - m_j)^T \cdot R_j^{-1} \cdot (x - m_j) + \log\{|R_j|\}. \tag{42}$$

V. Experimental Results

This section presents an experimental evaluation of the proposed approach and a comparison with methods based on the computation of spatial gray level dependence matrices. For this purpose, we have selected 12 Brodatz textures [18].The texture samples were digitized with a 256 × 256 spatial resolution and 256 gray levels. Spatial transducer nonuniformities were compensated by a local normalization procedure. A histogram equalization, producing an output image with 32 equiprobable gray levels, was then performed, in order to obtain texturres with approximately the same first order statistics. The advantage of such a treatment is twofold: first of all, it reduces the sensitivity to lighting condition and to gain adjustment of the transducer, and secondly, it guarantees that first order statistics are not taken into account for discrimination. The preprocessed images are shown in Fig. 3.

A. Sum and Difference Approximation

These textures were used to compute four co-occurrence matrices, for $d = 1$ and $\theta = 0, 45, 90, 135°$, as well as their associated sum and difference histograms. The maximum likelihood estimates of the corresponding sets of probabilities were obtained by normalization, applying (5) and (27), (28), respectively. The validity of the approximation of a second order distribution by a product of sum and difference histograms has been tested in evaluating the interdependence measure defined by (20). The detailed results of computation for all textures in Fig. 3 are given in Table III, for $d = 1$ and $\theta = 0°$. Equations (21) and (22) were evaluated using the binary logarithm. The second order entropy H_y has to be compared with the maximum value $2 \log_2\{32\} = 10$, corresponding to the equiprobability of all possible pairwise configurations. A consequence of preprocessing, which imposed flat histograms for all textures, is that this limiting case also represents the approximated distribution which would be obtained in assuming mutual pixel in-

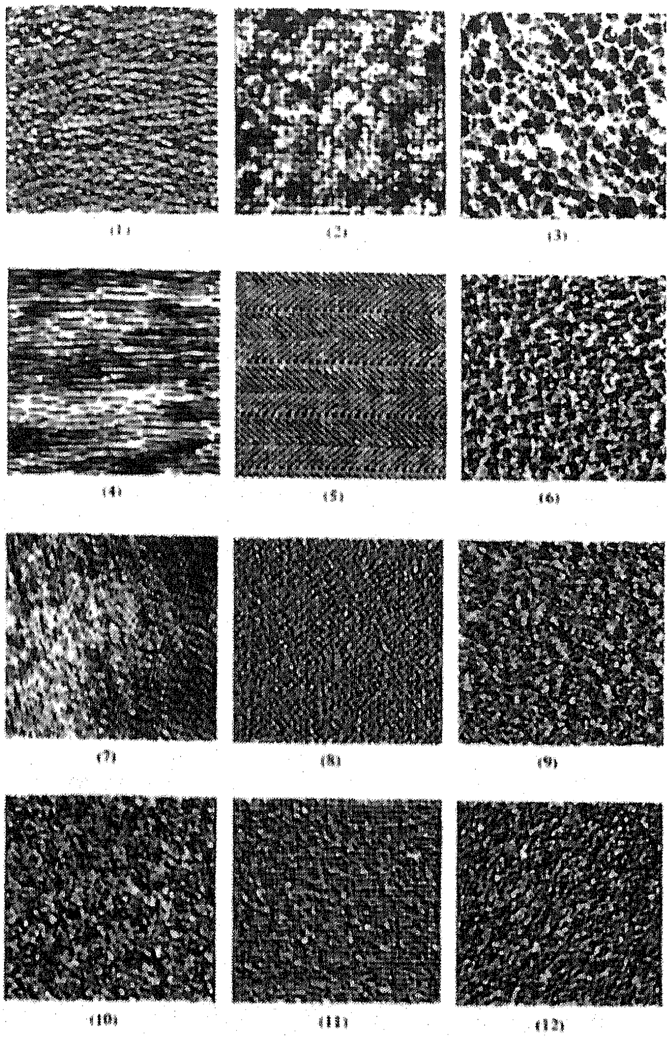

Fig. 3. Preprocessed Brodatz textures used for classification (256 × 256 pixels with 32 equiprobable grey levels).

TABLE III
DETAILED COMPUTATION OF THE INTERDEPENDENCE MEASURE (d = 1 AND θ = 0) FOR THE 12 TEXTURES IN FIG. 3

Texture	H_y	H_s	H_d	$\log_2\{c_0\}$	$I(P_y, \hat{P}_y)$
(1)	9.308	5.946	4.730	1.163	0.205
(2)	9.005	5.897	4.729	1.236	0.386
(3)	8.712	5.910	4.444	1.315	0.327
(4)	7.512	5.664	3.548	1.253	0.448
(5)	9.570	5.943	4.989	1.203	0.160
(6)	9.480	5.951	4.862	1.182	0.151
(7)	8.825	5.806	5.363	1.941	0.402
(8)	9.842	5.895	5.364	1.243	0.173
(9)	9.794	5.920	5.262	1.231	0.157
(10)	9.571	5.954	4.959	1.201	0.142
(11)	9.697	5.930	5.122	1.212	0.142
(12)	9.638	5.943	5.077	1.213	0.169

dependence. Histogram equalization has also the effect to enlarge both sum and difference histograms. This explains the relatively high values of $\log_2 \{c_0\}$ indicating important side effects, due to the fact that the product $P_s(i + j) \cdot P_d(i - j)$ is nonzero outside the domain defined by $ij = 1, \cdots, N_g$. Nevertheless, the quality of the approximation is acceptable and generally much better than the one that would be obtained by assuming pixel independence. The measures of interdependence in the four principal directions are given in Table IV. These values are roughly the same for all directions and are of the same order of magnitude for all textures that have been considered.

TABLE IV
INTERDEPENDENCE MEASURES IN THE FOUR PRINCIPAL DIRECTIONS WITH d = 1 COMPUTED FOR THE 12 TEXTURES IN FIG. 3

Texture	$\theta=0°$	$\theta=45°$	$\theta=95°$	$\theta=135°$
(1)	0.205	0.247	0.250	0.248
(2)	0.386	0.392	0.380	0.394
(3)	0.327	0.339	0.327	0.334
(4)	0.448	0.469	0.468	0.467
(5)	0.160	0.202	0.170	0.198
(6)	0.151	0.167	0.160	0.173
(7)	0.402	0.412	0.454	0.404
(8)	0.173	0.175	0.147	0.182
(9)	0.157	0.176	0.165	0.177
(10)	0.142	0.167	0.143	0.156
(11)	0.142	0.169	0.137	0.158
(12)	0.169	0.175	0.165	0.200

TABLE V
CLASSIFICATION OF THE 12 TEXTURES IN FIG. 3 USING SUM AND DIFFERENCE HISTOGRAMS IN THE FOUR PRINCIPAL DIRECTIONS WITH d = 1: P_{c1}: PROB. OF CORRECT CLASSIFICATION FOR THE TRAINING SET, N_{c1}: NUMBER OF CORRECTLY CLASSIFIED TEXTURE SAMPLES BELONGING TO THE TRAINING SET, P_{c2}: PROB. OF CORRECT CLASSIFICATION FOR THE UNKNOWN SET, N_{c2}: NUMBER OF CORRECTLY CLASSIFIED TEXTURE SAMPLES BELONGING TO THE UNKNOWN SET

Features	N	Size	P_{c1}	N_{c1}	P_{c2}	N_{c2}
sum and dif histograms	8x64	16x16	91.86%	1411 out of 1536	88.22%	1355 out of 1536
sum and dif histograms	8x64	32x32	98.94%	380 out of 384	96.88%	372 out of 384
sum and dif histograms	8x64	64x64	100%	96 out of 96	97.92%	94 out of 96
sum histograms	4x64	16x16	59.44%	913 out of 1536	55.40%	851 out of 1536
sum histograms	4x64	32x32	82.03%	315 out of 384	76.04%	292 out of 384
sum histograms	4x64	64x64	91.66%	88 out of 96	84.38%	81 out of 96
difference histograms	4x64	16x16	89.84%	1380 out of 1536	87.17%	1339 out of 1536
difference histograms	4x64	32x32	98.70%	379 out of 384	94.27%	362 out of 384
difference histograms	4x64	64x64	100%	96 out of 96	96.88%	93 out of 96
co-occurrence matrices	$4x32^2$	16x16	96.42%	1481 out of 1536	86.72%	1332 out of 1536
co-occurrence matrices	$4x32^2$	32x32	98.96%	380 out of 384	93.23%	358 out of 384
co-occurrence matrices	$4x32^2$	64x64	100%	96 out of 96	95.83%	92 out of 96

B. Classification Based on Sum and Difference Histograms

Sum and difference histograms have been applied to the classification of the 12 textures displayed in Fig. 3. The upper half of each was used as training set to estimate eight sum and difference probabilities along the directions of 0, 45, 90, and 135 degrees and d = 1. Classification was then performed for nonoverlapping subimages of dimension 16 × 16, 32 × 32, and 64 × 64, belonging to both training and unknown sets (upper and lower parts of the images, respectively). The classifier that was used is an extended version of the one shown in Fig. 2. The contributions of the probabilities corresponding to the different orientations have been added at the inputs of the summation blocks. Table V gives the results of classification in terms of the computed probability of correct classification for different window sizes. Results of classification using the sum (respectively difference) histograms on their own have been reported for comparison. The classification rates are always better for the training set than for the unknown set. This indicates a strong interdependence between the performances of the method and the quality of the estimation of the distribution parameters. As it can be expected, classification becomes more accurate as the window size increases. Classification is performed almost perfectly for window sizes larger than 32 × 32. In this experiment, difference histograms appear to be significantly more powerful than sum histograms for texture discrimination. Besides, these quantities perform a little better when used jointly than when used separately.

For comparison, the results of the classification of the same textures, using as features the four co-occurrence matrices in the directions of 0, 45, 90, and 135 degrees, have also been included.

TABLE VI
CLASSIFICATION OF THE 12 TEXTURES IN FIG. 3 FOR VARIOUS FEATURE SETS USING GLOBAL MEASURES BASED ON PAIRS OF PIXELS IN THE FOUR PRINCIPAL DIRECTIONS WITH $d = 1$: P_c: PROB. OF CORRECT CLASSIFICATION, N_c; NUMBER OF CORRECTLY CLASSIFIED TEXTURE SAMPLES (USING THE "LEAVING ONE OUT" METHOD)

Features	N	Size	P_c	N_c
Co-occurrence matrices				
set 1 ($d=1, \theta=0,45,90,135°$)	16	16x16	88.82 %	10243 out of 11532
set 1 ($d=1, \theta=0,45,90,135°$)	16	32x32	97.74 %	2639 out of 2700
set 1 ($d=1, \theta=0,45,90,135°$)	16	64x64	100. %	588 out of 588
set 2 ($d=1, \theta=0,45,90,135°$)	24	16x16	91.19 %	10516 out of 11532
set 2 ($d=1, \theta=0,45,90,135°$)	24	32x32	99.63 %	2690 out of 2700
set 2 ($d=1, \theta=0,45,90,135°$)	24	64x64	100. %	588 out of 588
Sum and dif. histograms				
set 3 ($d=1, \theta=0,45,90,135°$)	16	16x16	88.06 %	10155 out of 11532
set 3 ($d=1, \theta=0,45,90,135°$)	16	32x32	97.19 %	2624 out of 2700
set 3 ($d=1, \theta=0,45,90,135°$)	16	64x64	100. %	588 out of 588
set 4 ($d=1, \theta=0,45,90,135°$)	20	16x16	91.06 %	10501 out of 11532
set 4 ($d=1, \theta=0,45,90,135°$)	20	32x32	99.33 %	2682 out of 2700
set 4 ($d=1, \theta=0,45,90,135°$)	20	64x64	100. %	588 out of 588

These results were obtained by a similar classification algorithm, for additional details refer to [8]. The performances of the co-occurrence matrices are slightly better, considering the training set only. Surprisingly, there is a decrease in performances for the classification of the unknown set, when compared with sum and difference histograms. This could be a consequence of the large dimensionality of the feature vector. The classification rates obtained with sum and difference histograms are very competitive—these features being also more robust. For approximately equivalent results in texture discrimination, sum and difference histograms may be preferred to the usual co-occurrence matrices for their relative compactness as texture descriptors.

C. Classification Based on Global Features

For this classification experiment, the images were divided in square regions with 50 percent overlap. The experimental data set for each class consists of 961 texture samples of dimension 16×16, 225 samples of dimension 32×32, and 49 samples of dimension 64×64. Four symmetric co-occurrence matrices ($d = 1$ and $\theta = 0, 45, 90, 135$ degrees), as well as their corresponding sum and difference histograms, have been computed on every texture sample and have been used to evaluate four different sets of features:

- *Feature Set 1*—Each co-occurrence matrix is characterized by the four most commonly used Haralick features [2]: "energy," "entropy," "correlation," and "inertia."
- *Feature Set 2*—Each co-occurrence matrix is characterized by the "energy," "entropy," "inertia," "cluster shade," "cluster prominence," and "homogeneity" features, which were proposed more recently by Conners *et al.* [19], [20].
- *Feature Set 3*—Each co-occurrence matrix is characterized by the approximation of the four most commonly used Haralick features, using sum and difference histograms: "correlation," "inertia," and approximated "energy" and "entropy."
- *Feature Set 4*—Each co-occurrence matrix is characterized by the "correlation," "inertia," "cluster shade," "cluster prominence," and "homogeneity" features. This feature set was obtained in selecting the quantities in Sets 1 and 2 that can be directly computed from the sum and difference histograms.

The method described in section IV-B was used to classify the texture samples based on their measured feature values. For each pattern that has been tested, the training was performed on the remaining samples ("leaving one out" method), using the maximum likelihood estimates of the distribution parameters defined by (39) and (40). The results of classification are given in Table VI. It appears that the use of the sum and difference approximation (Set 3 instead of Set 1) does not change the performances in texture classification very much. The best results are obtained with feature Sets 3 and 4. The former is slightly more performant than the latter, which was obtained in choosing the co-occurrence features that can be computed from the sum and difference histograms. This set might be of practical interest because it can be evaluated directly, by spatial averaging, at the output of the sum and difference filters. This technique enables a significant savings in computation and memory storage, as it is not necessary to estimate explicitly the auxilary sum and difference histograms—and even more so the co-occurrence matrices.

From these experiments, it can be concluded that global texture features derived from the sum and difference histograms provide almost equivalent results in texture classification than those obtained from the usual co-occurrence matrices.

VI. CONCLUSION

A simplification of the very popular Spatial Grey Level Dependence Method for texture analysis has been proposed. The usual co-occurrence matrices are replaced by their associated sum and difference histograms. This is justified by the fact that the sum and difference define the principal axes of second-order probability functions of a stationary random process and are therefore decorrelated. A set of global histogram measures was introduced. It was also shown that the usual texture features (or some close approximations) associated to the co-occurrence matrices could be computed directly from the sum and difference histograms. The advantage of this procedure is the reduction in computation time and memory storage.

Two maximum likelihood texture classifiers have been introduced. The first one considers the sum and difference histograms as the component of a feature vector. The assumption of a class conditional multinomial distribution leads to an extremely fast implementation which avoids an explicit evaluation of the feature vector. The second classifier is based on global measurements extracted from the histograms. Experimental results indicate that sum and difference histograms are almost as powerful as co-occurrence matrices for texture discrimination.

REFERENCES

[1] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, pp. 786–804, May 1979.

[2] R. M. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, pp. 610–621, Nov. 1973.

[3] B. Julesz, "Visual pattern discrimination," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 84–92, Feb. 1962.
[4] B. Julesz *et al.*, "Inability of humans to discriminate between visual textures that agree in second order statistics—Revisited," *Perception*, vol. 2, pp. 391–405, 1973.
[5] A. Gagalowicz, "Visual discrimination of stochastic texture fields based upon second order statistics," in *Proc. 5th Int. Joint Conf. Pattern Recognition*, Miami Beach, FL, Dec. 1–4, 1980, pp. 786–789.
[6] ——, "A new method for texture field synthesis. Some applications to the study of human vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-3, pp. 520–533, Sept. 1981.
[7] A. L. Vickers and J. W. Modestino, "A maximum likelihood approach to texture classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, pp. 61–68, Jan. 1982.
[8] M. Unser, "A fast texture classifier based on cross entropy minimisation," Proc. EUSIPCO-83, in *Signal Processing II: Theories and Applications*, H. W. Schüssler, Ed. Amsterdam, The Netherlands: Elsevier, 1983, pp. 261–264.
[9] M. Pietikaïnen, A. Rosenfeld, and L. S. Davis, "Experiments with texture classification using averages of local matches," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-13, pp. 421–426, 1983.
[10] R. L. Kashyap, R. Chellappa, and A. Khotanzad, "Texture classification using features derived from random field models," *Pattern Recognition Lett.*, vol. 1, pp. 43–50, 1982.
[11] S. Watanabe, *Knowing and Guessing*. New York: Wiley, 1969.
[12] S. Kullback, *Information Theory and Statistics*. New York: Dover, 1957.
[13] R. W. Johnson, "Axiomatic characterization of the directed divergences and their linear combinations," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 709–716, Nov. 1979.
[14] J. S. Weszka, C. Y. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 269–285, Apr. 1976.
[15] R. W. Conners and C. A. Harlow, "A theoretical comparison of texture algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, pp. 204–222, May 1980.
[16] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic, 1972.
[17] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1965.
[18] P. Brodatz, *Textures—A Photographic Album for Artists and Designers*. New York: Dover, 1966.
[19] R. W. Conners, "Toward a set of statistical features which measure visually perceivable qualities of textures," in *Proc. Pattern Recognition Image Processing Conf.*, Aug. 1979, pp. 382–390.
[20] R. W. Conners and C. A. Harlow, "Toward a structural textural analyser based on statistical methods," *Comput. Graphics Image Processing*, vol. 12, pp. 224–256, 1980.