Reprinted by permission from IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS
Vol. SMC-3, No. 6, November 1973, pp. 610-621

Textural Features for Image Classification

ROBERT M. HARALICK, K. SHANMUGAM, AND ITS'HAK DINSTEIN

Abstract—**Texture is one of the important characteristics used in identifying objects or regions of interest in an image, whether the image be a photomicrograph, an aerial photograph, or a satellite image. This paper describes some easily computable textural features based on gray-tone spatial dependancies, and illustrates their application in category-identification tasks of three different kinds of image data: photomicrographs of five kinds of sandstones, 1:20 000 panchromatic aerial photographs of eight land-use categories, and Earth Resources Technology Satellite (ERTS) multispecial imagery containing seven land-use categories. We use two kinds of decision rules: one for which the decision regions are convex polyhedra (a piecewise linear decision rule), and one for which the decision regions are rectangular parallelpipeds (a min–max decision rule). In each experiment the data set was divided into two parts, a training set and a test set. Test set identification accuracy is 89 percent for the photomicrographs, 82 percent for the aerial photographic imagery, and 83 percent for the satellite imagery. These results indicate that the easily computable textural features probably have a general applicability for a wide variety of image-classification applications.**

I. Introduction

With the advent of high-speed general-purpose digital computers it is becoming possible to perform mathematical or algorithmic processes on pictorial data from images of photographic quality. In most of these processes, the pictorial information is represented as a function of two variables (x,y). The image in its digital form is usually stored in the computer as a two-dimensional array. If $L_x = \{1,2,\cdots,N_x\}$ and $L_y = \{1,2,\cdots,N_y\}$ are the X and Y spatial domains, then $L_x \times L_y$ is the set of resolution cells and the digital image I is a function which assigns some gray-tone value $G \in \{1,2,\cdots,N_g\}$ to each and every resolution cell; $I: L_x \times L_y \rightarrow G$. Various two-dimensional analyses are performed on I to achieve specific image-processing tasks such as coding, restoration, enhancement, and classification. In recent years a tremendous amount of computer processing of photographs has occurred, with facilities having been developed to process anything from aerial photographs to photomicrographs [1], [2].

In this paper we are concerned with the task of developing a set of features for classifying or categorizing pictorial data. The classification of pictorial data can be done on a resolution cell basis (such as in identifying the crop category of a resolution cell on satellite imagery) or on a block of contiguous resolution cells (such as in identifying the crop category of an entire agricultural field extending over a large number of resolution cells). The most difficult step in categorizing pictorial information from a large block of resolution cells is that of defining a set of meaningful features to describe the pictorial information from the block of resolution cells. Once these features are defined, image blocks can be categorized using any one of a multitude of pattern-recognition techniques.

In a search for meaningful features for describing pictorial information, it is only natural to look toward the types of features which human beings use in interpreting pictorial information. Spectral, textural, and contextual features are three fundamental pattern elements used in human interpretation of color photographs. Spectral features describe the average tonal variations in various bands of the visible and/or infrared portion of an electromagnetic spectrum, whereas textural features contain information

Manuscript received April 25, 1973; revised June 12, 1973. This work was supported by the NASA Goddard Space Flight Center under Contract NAS5-21822, and the Geographic Sciences Laboratory of the U.S. Army Engineer Topographic Laboratories under Contract DAAK02-70-C-0388.

R. M. Haralick and I. Dinstein are with the School of Electrical Engineering and the Center for Research, Inc., University of Kansas, Lawrence, Kans. 66044.

K. Shanmugam was with the Center for Research, Inc., University of Kansas, Lawrence, Kans. He is now with the Department of Electrical Engineering, Wichita State University, Wichita, Kans. 67218.

about the spatial distribution of tonal variations within a band. Contextual features contain information derived from blocks of pictorial data surrounding the area being analyzed. When small image areas from black and white photographs are independently processed by a machine, then texture and tone are most important.

The concept of tone is based on the varying shades of gray of resolution cells in a photographic image, while texture is concerned with the spatial (statistical) distribution of gray tones. Texture and tone are not independent concepts; rather, they bear an inextricable relationship to one another very much like the relationship between a particle and a wave. Context, texture, and tone are always present in the image, although at times one property can dominate the other.

Texture can be evaluated as being fine, coarse, or smooth; rippled, molled, irregular, or lineated. For example, in the humid tropics, fine texture on radar imagery can be indicative of nonresistant fine-grain sedimentary rocks and unconsolidated sediments, while a coarse texture can be indicative of coarser grained sedimentary rocks. A massive texture with high-contrast components may be indicative of igneous rocks. A hummocky texture can be indicative of eroded igneous rocks.

Texture is an innate property of virtually all surfaces—the grain of wood, the weave of a fabric, the pattern of crops in a field, etc. It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment. Although it is quite easy for human observers to recognize and describe in empirical terms, texture has been extremely refractory to precise definition and to analysis by digital computers. Since the textural properties of images appear to carry useful information for discrimination purposes, it is important to develop features for texture. We present in this paper a computationally quick procedure for extracting textural features of images and discuss the usefulness of these features for discriminating between different kinds of image data.

Early image texture studies have employed autocorrelation functions [3], power spectra [4], restricted first- and second-order Markov meshes [5], and relative frequencies of various gray levels on the unnormalized image [6]. These had some degree of success, but we know little more about texture after finding the results of these experiments than before because they did not try to specifically define, characterize, or model texture. They only used some general mathematical transformation which assigns numbers to the transformed image in a nonspecific way.

Recent attempts to extract textural features have been limited to developing algorithms for extracting specific image properties such as coarseness and presence of edges. Many such algorithms have been developed and tried on special imagery. The subjective parameters (such as the selection of thresholds) associated with the techniques do not enable them to be generalized to imagery other than that processed by the authors. Recently, Rosenfeld and his coinvestigators presented a set of procedures for extracting some textural properties of pictorial data [7]–[9]. In [7] Rosenfeld and Troy described a procedure for obtaining a measure of the texture "coarseness" of images. Their procedures were based on the differences between the gray-tone values of adjacent image elements and on the autocorrelation of the image gray-tone values. In [8] Rosenfeld and Thurston gave a procedure for detecting boundaries separating regions which differ in texture coarseness. In [9] Troy *et al.* described gray-level manipulation procedures which can be used for preprocessing of pictures before applying the algorithms given in [7] and [8]. Procedures for detecting textural properties such as lines and dots have also been suggested by other investigators (see, for example, [10]–[12]). Before applying these procedures to pictures other than the ones processed by the authors of the respective papers, the investigator has to make a choice as to the method appropriate to the picture in question as well as the selection of parameters for the particular method.

We are presenting in this paper a general procedure for extracting textural properties of blocks of image data. These features are calculated in the spatial domain, and the statistical nature of texture is taken into account in our procedure, which is based on the assumption that the texture information in an image I is contained in the overall or "average" spatial relationship which the gray tones in the image have to one another. We compute a set of gray-tone spatial-dependence probability-distribution matrices for a given image block and suggest a set of 14 textural features which can be extracted from each of these matrices. These features contain information about such image textural characteristics as homogeneity, gray-tone linear dependencies (linear structure), contrast, number and nature of boundaries present, and the complexity of the image. It is important to note that the number of operations required to compute any one of these features is proportional to the number of resolution cells in the image block. It is for this reason that we call these features quickly computable.

We also investigate the usefulness of textural features for categorizing or classifying image blocks from three different data sets ranging from high-resolution photomicrographs to low-resolution satellite imagery. The accuracy of classification on multiclass categorization using textural features of these data sets was in the range of 80–90 percent.

II. Textural Features

Our initial perspective of texture and tone is based on the concept that texture and tone bear an inextricable relationship to one another. Tone and texture are always present in an image, although one property can dominate the other at times. The basic intuitively perceived relationships between tone and texture are the following. When a small-area patch of an image has little variation—i.e., little variation of features of discrete gray tone—the dominant property of that area is tone. When a small-area patch has a wide variation of features of discrete gray tone, the dominant property of that area is texture. Crucial to this distinction are the size of the small-area patch, the relative sizes of the discrete features, and the number of distinguishable discrete features. As the number of distinguishable tonal discrete features decreases, the tonal properties will

predominate. In fact, when the small-area patch is only the size of one resolution cell, so that there is only one discrete feature, the only property present is tone. As the number of distinguishable features of discrete gray tone increases within the small-area patch, the texture property will dominate.

One important property of tone–texture is the spatial pattern of the resolution cells composing each discrete tonal feature. When there is no spatial pattern and the gray-tone variation between features is wide, a fine texture results. As the spatial pattern becomes more definite and involves more and more resolution cells, a coarser texture results. An excellent set of photographs of different types of texture may be found in [13].

The preceding description of texture is, of course, a gross simplification and idealization of what actually occurs. Discrete tonal features are really quite fuzzy in that they do not necessarily stand out as entities by themselves. Therefore the texture analysis we suggest is concerned with more general or macroscopic concepts than discrete tonal features.

The procedure we suggest for obtaining the textural features of an image is based on the assumption that the texture information on an image I is contained in the overall or "average" spatial relationship which the gray tones in the image I have to one another. More specifically, we shall assume that this texture information is adequately specified by a set of gray-tone spatial-dependence matrices which are computed for various angular relationships and distances between neighboring resolution cell pairs on the image. All of our textural features are derived from these angular nearest-neighbor gray-tone spatial-dependence matrices.

Gray-Tone Spatial-Dependence Matrices

Suppose an image to be analyzed is rectangular and has N_x resolution cells in the horizontal direction and N_y resolution cells in the vertical direction. Suppose that the gray tone appearing in each resolution cell is quantized[1] to N_g levels. Let $L_x = \{1,2,\cdots,N_x\}$ be the horizontal spatial domain, $L_y = \{1,2,\cdots,N_y\}$ be the vertical spatial domain, and $G = \{1,2,\cdots,N_g\}$ be the set of N_g quantized gray tones. The set $L_y \times L_x$ is the set of resolution cells of the image ordered by their row–column designations. The image I can be represented as a function which assigns some gray tone in G to each resolution cell or pair of coordinates in $L_y \times L_x$; $I: L_y \times L_x \rightarrow G$.

An essential component of our conceptual framework of texture is a measure, or more precisely, four closely related measures from which all of our texture features are derived. These measures are arrays termed angular nearest-neighbor gray-tone spatial-dependence matrices, and to describe these arrays we must emphasize our notion of adjacent or

[1] Variations in lighting, lens, film, developer, and digitizer usually introduce monotonic transformations of the "true" image gray-tone values. Under these conditions, we would want two images of the same scene, one being a gray-tone monotonic transformation of the other to produce the same features. Image normalization by equal-probability quantizing guarantees that images which are monotonic transformations of one another produce the same results [14], [15]. An equal-probability quantizing algorithm is given in Appendix III.

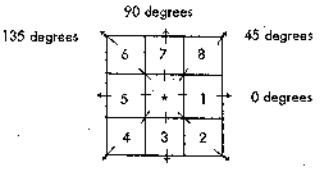


Fig. 1. Resolution cells 1 and 5 are 0° (horizontal) nearest neighbors to resolution cell *; resolution cells 2 and 6 are 135° nearest neighbors; resolution cells 3 and 7 are 90° nearest neighbors; and resolution cells 4 and 8 are 45° nearest neighbors to *. (Note this information is purely spatial, and has nothing to do with gray-tone values.)

nearest-neighbor resolution cells themselves. We consider a resolution cell—excluding those on the periphery of an image, etc.—to have eight nearest-neighbor resolution cells as in Fig. 1.

We assume that the texture-context information in an image I is contained in the overall or "average" spatial relationship which the gray tones in image I have to one another. More specifically, we shall assume that this texture-context information is adequately specified by the matrix of relative frequencies P_{ij} with which two neighboring resolution cells separated by distance d occur on the image, one with gray tone i and the other with gray tone j. Such matrices of gray-tone spatial-dependence frequencies are a function of the angular relationship between the neighboring resolution cells as well as a function of the distance between them. Fig. 2 illustrates the set of all horizontal neighboring resolution cells separated by distance 1. This set, along with the image gray tones, would be used to calculate a distance 1 horizontal gray-tone spatial-dependence matrix. Formally, for angles quantized to 45° intervals the unnormalized frequencies are defined by

$$\begin{aligned}
P(i,j,d,0^\circ) &= \#\{((k,l),(m,n)) \in (L_y \times L_x) \\
&\quad \times (L_y \times L_x) \mid k - m = 0, |l - n| = d, \\
&\quad I(k,l) = i, I(m,n) = j\} \\
P(i,j,d,45^\circ) &= \#\{((k,l),(m,n)) \in (L_y \times L_x) \\
&\quad \times (L_y \times L_x) \mid (k - m = d, l - n = -d) \\
&\quad \text{or } (k - m = -d, l - n = d), \\
&\quad I(k,l) = i, I(m,n) = j\} \\
P(i,j,d,90^\circ) &= \#\{((k,l),(m,n)) \in (L_y \times L_x) \\
&\quad \times (L_y \times L_x) \mid |k - m| = d, \\
&\quad l - n = 0, I(k,l) = i, I(m,n) = j\} \\
P(i,j,d,135^\circ) &= \#\{((k,l),(m,n)) \in (L_y \times L_x) \\
&\quad \times (L_y \times L_x) \mid (k - m = d, l - n = d) \\
&\quad \text{or } (k - m = -d, l - n = -d), \\
&\quad I(k,l) = i, I(m,n) = j\} \qquad (1)
\end{aligned}$$

where # denotes the number of elements in the set.

Note that these matrices are symmetric; $P(i, j; d, a) = P(j, i; d, a)$. The distance metric ρ implicit in the preceding

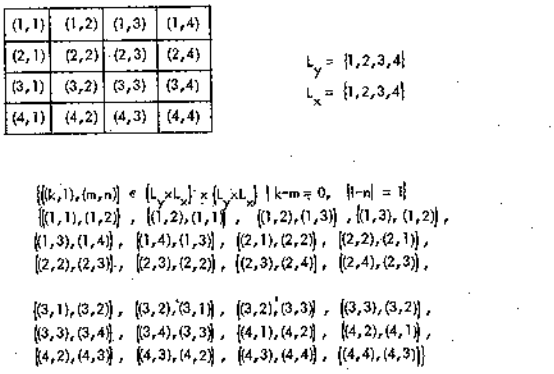


Fig. 2. Set of all distance 1 horizontal neighboring resolution cells on 4 × 4 image.

equations can be explicitly defined by

$$\rho((k,l),(m,n)) = \max\{|k-m|, |l-n|\}.$$

Consider Fig. 3(a), which represents a 4 × 4 image with four gray tones, ranging from 0 to 3. Fig. 3(b) shows the general form of any gray-tone spatial-dependence matrix. For example, the element in the (2,1) position of the distance 1 horizontal P_H matrix is the total number of times two gray tones of value 2 and 1 occurred horizontally adjacent to each other. To determine this number, we count the number of pairs of resolution cells in R_H such that the first resolution cell of the pair has gray tone 2 and the second resolution cell of the pair has gray tone 1. In Fig. 3(c)–3(f) we calculate all four distance 1 gray-tone spatial-dependence matrices.

If needed, the appropriate frequency normalizations for the matrices are easily computed. When the relationship is nearest horizontal neighbor ($d = 1$, $a = 0°$), there will be $2(N_x - 1)$ neighboring resolution cell pairs on each row and there are N_y rows, providing a total of $2N_y(N_x - 1)$ nearest horizontal neighbor pairs (see Fig. 3). When the relationship is nearest right-diagonal neighbor ($d = 1$, $a = 45°$) there will be $2(N_x - 1)$ 45° neighboring resolution cell pairs for each row except the first, for which there are none, and there are N_y rows. This provides a total of $2(N_y - 1)(N_x - 1)$ nearest right-diagonal neighbor pairs. By symmetry there will be $2N_x(N_y - 1)$ nearest vertical neighbor pairs and $2(N_x - 1)(N_y - 1)$ nearest left-diagonal neighbor pairs. After the number of neighboring resolution cell pairs R used in computing a particular gray-tone spatial-dependence matrix is obtained, the matrix is normalized by dividing each entry in the matrix by R.

It is appropriate at this point to comment on the computational aspects of obtaining the gray-tone spatial-dependence matrices. The number of operations required to process an image using our procedure is directly proportional to the number of resolution cells n present in the image. In comparison, the number of operations are of the order of $n \log n$ if one wishes to use Fourier or Hadamard transform to extract texture information. Also, to compute the entries in the gray-tone spatial-dependence matrices, one needs to keep only two lines of image data in core at a time. Thus no severe storage constraints are imposed. For most of the images we have processed, the computations have been carried out on a small digital computer (PDP 15/20 with 12 kwords of core and two DEC tape drives).

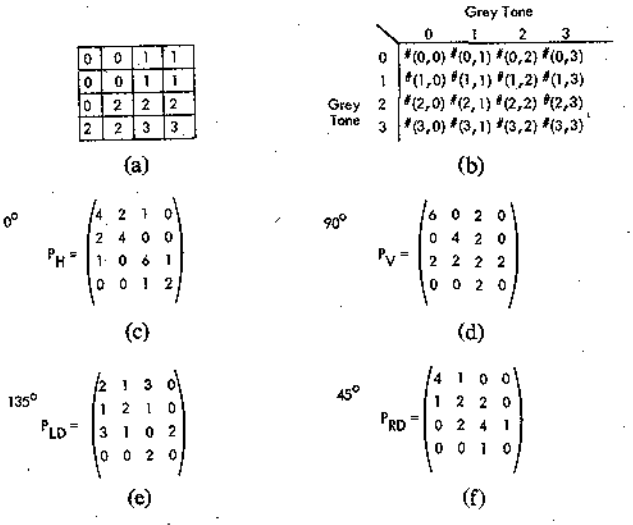


Fig. 3. (a) 4 × 4 image with four gray-tone values 0–3. (b) General form of any gray-tone spatial-dependence matrix for image with gray-tone values 0–3. $\#(i,j)$ stands for number of times gray tones i and j have been neighbors. (c)–(f) Calculation of all four distance 1 gray-tone spatial-dependence matrices.

Textural Features Extracted from Gray-Tone Spatial-Dependence Matrices

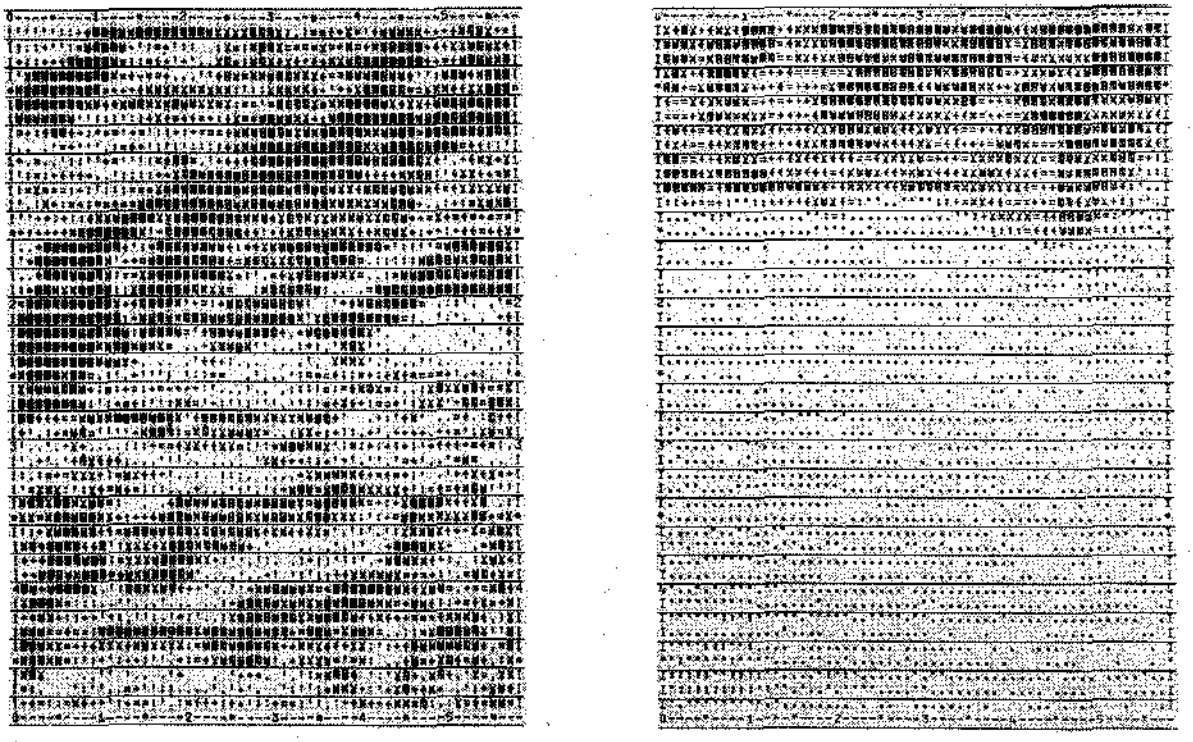
Our initial assumption in characterizing image texture is that all the texture information is contained in the gray-tone spatial-dependence matrices. Hence all the textural features we suggest are extracted from these gray-tone spatial-dependence matrices. The equations which define a set of 14 measures of textural features are given in Appendix I. Some of these measures relate to specific textural characteristics of the image such as homogeneity, contrast, and the presence of organized structure within the image. Other measures characterize the complexity and nature of gray-tone transitions which occur in the image. Even though these features contain information about the textural characteristics of the image, it is hard to identify which specific textural characteristic is represented by each of these features.

For illustrative purposes, we will define 3 of the 14 textural features in this section and explain the significance of these features in terms of the kind of values they take on for two images of distinctly different textural characteristics The features we consider are

$$f_1 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left(\frac{P(i,j)}{R}\right)^2 \tag{2}$$

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{|i-j|=n} \left(\frac{P(i,j)}{R}\right) \right\} \tag{3}$$

$$f_3 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [ijP(i,j)/R] - \mu_x\mu_y}{\sigma_x\sigma_y} \tag{4}$$



Grassland				Water Body		
Angle	ASM	Contrast	Correlation	ASM	Contrast	Correlation
0°	.0128	3.048	.8075	.1016	2.153	.7254
45°	.0080	4.011	.6366	.0771	3.057	.4768
90°	.0077	4.014	.5987	.0762	3.113	.4646
135°	.0064	4.709	.4610	.0741	3.129	.4650
Avg.	.0087	3.945	.6259	.0822	2.863	.5327

(a) (b)

Fig. 4. Textural features for two different land-use category images.

where μ_x, μ_y, σ_x, and σ_y are the means and standard deviations of the marginal distributions associated with $P(i,j)/R$, and R is a normalizing constant.

Fig. 4 shows the digital printout of two 64 × 64 image blocks taken from a satellite picture over the California coastline (NASA ERTS Image no. 1002-18134). The image shown in Fig. 4(a) is a representative sample of grasslands and Fig. 4(b) is a representative sample of water bodies in the area. The values of the features f_1, f_2, and f_3 obtained from gray-tone spatial-dependence matrices for distance $d = 1$, are shown below the images in Fig. 4.

The angular second-moment feature (ASM) f_1 is a measure of homogeneity of the image. In a homogeneous image, such as shown in Fig. 4(b), there are very few dominant gray-tone transitions. Hence the P matrix for this image will have fewer entries of large magnitude. For an image like the one shown in Fig. 2(a), the P matrix will have a large number of small entries, and hence the ASM feature (which is the sum of squares of the entries) in the P matrix will be smaller. A comparison of the ASM values given below the images in Fig. 4 shows the usefulness of the ASM feature as a measure of the homogeneity of the image.

The contrast feature f_2 is a difference moment of the P matrix and is a measure of the contrast or the amount of local variations present in an image. Since there is a large amount of local variation present in the image of Fig. 4(a) compared to the image shown in Fig. 4(b), the contrast feature for the grassland image has consistently higher values compared to the water-body image.

The correlation feature f_3 is a measure of gray-tone linear-dependencies in the image. For both the images shown in Fig. 4, the correlation feature is somewhat higher in the horizontal (0°) direction, along the line of scan. The water-body image consists mostly of a constant gray-tone value for the water plus some additive noise. Since the noise samples are mostly uncorrelated, the correlation features for the water-body image have lower values compared to the grassland image. Also, the grassland image has a considerable amount of linear structure along 45° lines across the image, and hence the value of the correlation feature is

higher along this direction compared to the values for 90° and 135° directions. More examples of the significance of some of the textural features are presented in Appendix II.

The various features which we suggest are all functions of distance and angle. The angular dependencies present a special problem. Suppose image A has features a, b, c, and d for angles 0°, 45°, 90°, and 135°, respectively, and image B is identical to A except that B is rotated 90° with respect to A. Then B will have features c, d, a, and b for angles 0°, 45°, 90°, and 135°, respectively. Since the texture context of A is the same as the texture context of B, any decision rule using the angular features a,b,c,d must produce the same results for c,d,a,b. To guarantee this, we suggest that the angularly dependent features not be used directly. Instead we suggest that two functions of a, b, c, and d, their average and range (which are invariant under rotation), be used as inputs to the classifier.

III. Applications of Textural Features for Image Classification

In this section we present the results of our studies on the usefulness of the textural features for categorizing images. Three data sets were used in our study. These data sets were extracted from photomicrographs of different rocks, from aerial photographs of man-made and natural scenes, and from high-altitude satellite pictures of the earth. A brief description of the data sets and classification algorithms used and the results of classification experiments will be presented. For further details the interested reader is referred to [16]–[18].

Data Set Description

1) *Photomicrographs of Sandstones:* Identification of the types of rocks (sandstones) present in an underground reservoir of crude oil is important in petroleum production studies; hence we undertook a study to identify different types of rocks from the textural characteristics of their photomicrographs. The data set consisted of 243 image blocks of size 64 × 64 and the gray tones of the images were equal-probability quantized into 16 levels. Details of the algorithm used to perform this quantization are given in Appendix III. There were five sandstone categories in the data set, and samples of the photomicrographs are shown in Fig. 5. The textural features for the 243 samples were calculated from distance 1 gray-tone spatial-dependence matrices.

2) *Aerial Photographic Data Set:* This data set consisted of 170 images belonging to eight categories. These 170 images were manually sampled from 9 × 9-in 1:20 000 aerial photography negatives provided by the Army Environmental Topographic Laboratories, Fort Belvoir, Va. [17]. The digitization of the sampled images was done on the Image Discrimination, Enhancement, and Combination System (IDECS)/PDP. The IDECS is an analog–digital near real-time processing system that has been developed at the University of Kansas Center for Research, Inc. A PDP 15/20 computer is interfaced to the IDECS and the system is capable of performing various functions of image processing [25]. The IDECS is capable of processing up to four image inputs. Three flying spot scanners can input transparencies up to a maximum size of 4 × 5 in. A vidicon scanner accepts larger transparencies as well as printed images. Multi-image inputs are electronically registered and combined into a format more suitable for interpretation. Processed images are displayed on a color display unit and/or a black and white monitor. A 24-channel disk is synchronized with the system and used for storage of processed digital images. Either manually or under computer control, the IDECS/PDP is able to perform image enhancement, video level slicing, area integration, category discrimination, display of histograms and scatograms, quantizing, and similar image-processing operations. The algorithm used for quantizing the aerial photographic data set digitized the images into eight levels in such a way that the areas of the different levels (in each image) were equal. Each image was digitized into a 20 × 50 array. The eight categories were residential old (RSOLD), residential new (RESNU), LAKE, SWAMP, MARSH, URBAN, RAIL, and SCRUB or WOOD (SCROD). The first six categories contained 20 samples each, the RAIL and SCROD categories contained 10 and 40 samples, respectively. Samples of the images are shown in Fig. 6.

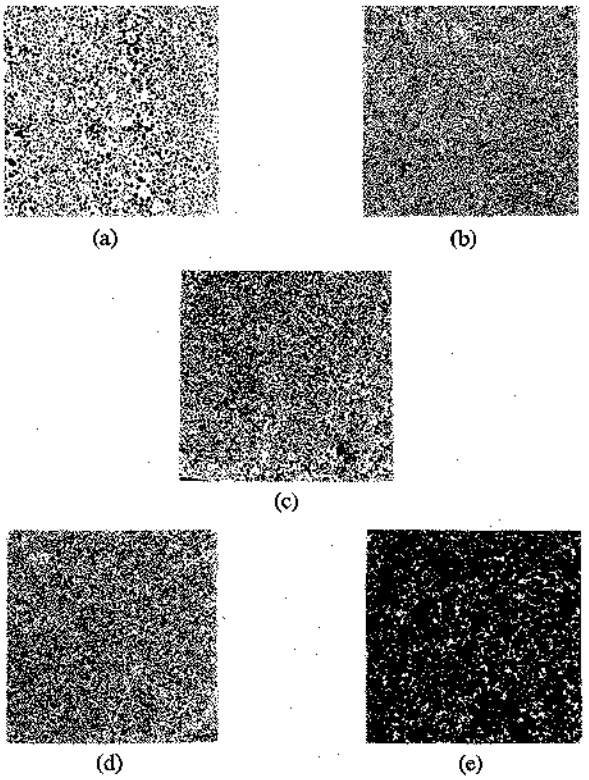


Fig. 5. Samples of photomicrographs of sandstones. (a) Dexter-L. (b) Dexter-H. (c) St. Peter. (d) Upper Muddy. (e) Gaskel.

3) *Data Set Derived from Satellite Imagery:* A variety of satellite imagery is currently used for the remote sensing of Earth resources. The most commonly used satellite

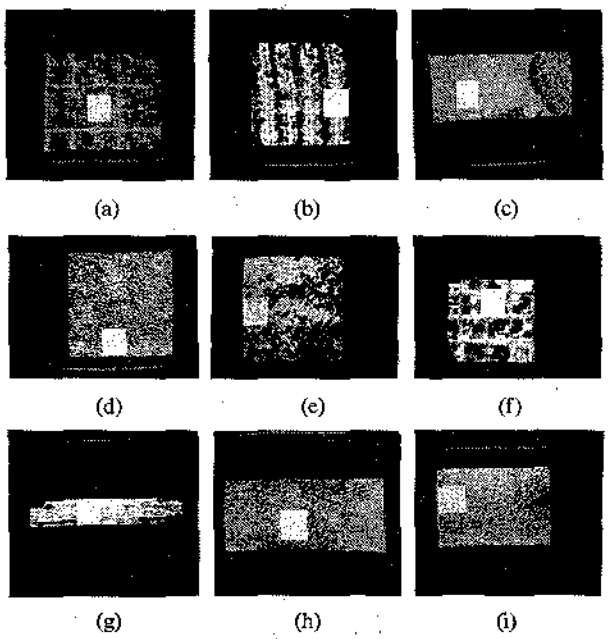


Fig. 6. Samples of aerial photographic data set. (a) RSOLD. (b) RESNU. (c) LAKE. (d) SWAMP. (e) MARSH. (f) URBAN. (g) RAIL. (h) SCRUB. (i) WOOD. (SCRUB and WOOD are classified together as SCROD.)

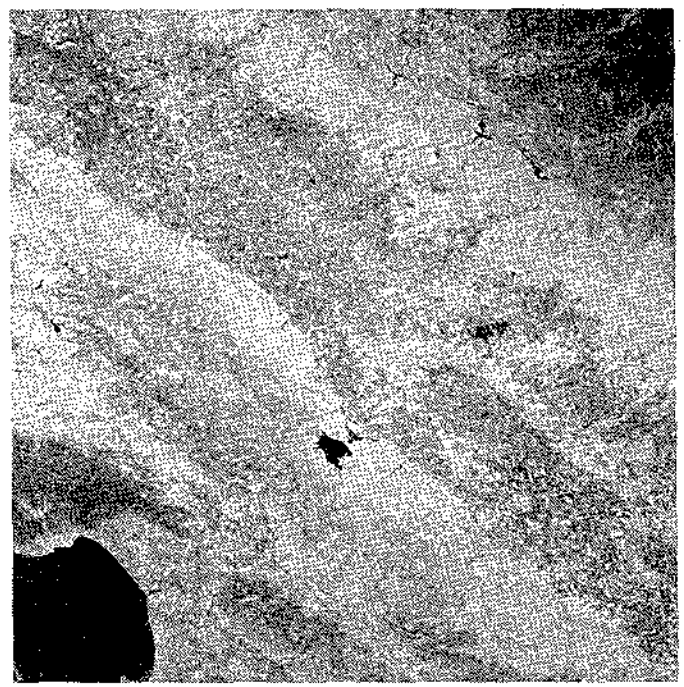


Fig. 7. Satellite picture of Monterey Bay, Calif.

imagery is the multispectral scanner (MSS) imagery, which consists of a set of images of a scene, where each image in the set is created by scanning the radiance of the scene from the satellite in a narrow band of wavelengths. The MSS operates in wavelengths ranging from 0.3 to 16 μm, and up to 24 scanners have been used in the past.

The data set used in our study was derived from a high-altitude four-channel multispectral satellite image taken over the California coastline. The image in one of the four MSS bands is shown in Fig. 7. The ground area covered by the image is about 14 400 mi^2 and the size of the digital image is 2340 $\times$ 3200. Out of this large image, a subset of 624 contiguous image blocks of size 64 $\times$ 64 was taken and used in our study. There were seven land-use categories in the image; these are coastal forest, woodlands, annual grasslands, urban areas, small and large irrigated fields, and water bodies. The land-use category for each 64 $\times$ 64 image block was obtained by interpreting the color composite of the multispectral image set. The textural features were computed from the distance 1 gray-tone spatial-dependence matrix of the second MSS band image in the multispectral image set.

Since the ground truth category information was derived from the color composite image, we used a set of spectral (color) features in addition to the textural features for categorizing these images. The spectral features consisted of the mean and standard deviation of the gray-tone values of the 64 $\times$ 64 image blocks in each of the four spectral bands.

Classification Algorithms

1) *Piecewise Linear Discriminant Function Method:* For categorizing image blocks in data sets 1 and 3, which have a reasonably large number of samples, we used a piecewise linear-discriminant function method. The data sets were partitioned arbitrarily into training and test sets. The classification algorithm was developed using the samples in the training set and it was tested on the samples in the test set.

A variety of algorithms have been discussed in the pattern-recognition literature for obtaining piecewise linear-discriminant functions for pattern classification [19]–[22]. In a widely used algorithm, the pattern space is partitioned into a number of regions using a set of hyperplanes (decision boundaries) whose locations are determined by the sample patterns. Each region is dominated by sample patterns of a particular category. When a new pattern is presented for identification, it is assigned a category depending on the region to which it belongs. If the new pattern X is located in a region dominated by sample patterns of category c_j, then X is classified as coming from category c_j.

For the multicategory problem involving N_R categories, a total of $N_R(N_R - 1)/2$ hyperplanes are used to partition the pattern space. These hyperplanes are defined by a set of weight vectors W_{ij}, $i = 1,2,\cdots,N_R$, $j = 1,2,\cdots,N_R$, $j > i$, which separates the sample patterns belonging to the ith and jth categories. A regression type algorithm given in [19, ch. 4] was used to obtain the weight vectors. After the location of the hyperplanes are determined, the classification of new patterns is done as follows. For each category c_i, the number of hyperplanes V_i which give a positive response when the new pattern X is presented are determined using

$$V_i = \sum_{\substack{j=1 \\ j \neq i}}^{N_R} \frac{|W_{ij}{}^T Z| + W_{ij}{}^T Z}{2|W_{ij}{}^T Z|}, \qquad i = 1,2,\cdots,N_R \quad (5)$$

where Z is the augmented pattern vector obtained by adding a component of value 1 to X, i.e.,

$$Z = \begin{bmatrix} 1 \\ X \end{bmatrix}. \tag{6}$$

X is assigned to category c_j if $V_j = \max\{V_i\}$. If there is a tie between categories c_m and c_n, then X is assigned to c_m if $W_{mn}{}^T Z \geq 0$, or to c_n if $W_{mn}{}^T Z < 0$. Several modifications of the linear-discriminant function method and a multitude of other classification procedures may be found in the references cited.

2) *Min–Max Decision Rule:* For the aerial photograph data set which had a small number of samples per category we used a min–max decision rule for categorizing the images based on their textural features. The decision boundaries were obtained by using all but one of the samples in the data set, and the sample which was left out was then classified. The procedure was repeated for all the samples in the data set to obtain the overall accuracy of the classification scheme. This method of "leaving one sample out" is widely used by investigators in the pattern-recognition field when the data set has a small number of samples. The decision rule is described as follows.

Decision rule: If b_{n_k} and a_{n_k} define the minimum and maximum values of the uniform distribution, $(x_1,x_2,\cdots,x_N)$ is assigned to category k if and only if

$$b_{n_k} \leq x_n \leq a_{n_k}, \qquad n = 1,2,\cdots,N$$

$$\prod_{n=1}^{N} \frac{1}{(a_{n_k} - b_{n_k})} \geq \prod_{n=1}^{N} \frac{1}{(a_{n_j} - b_{n_j})} \tag{7}$$

for all j such that $b_{n_j} \leq x_n \leq a_{n_j}$, $n = 1,2,\cdots,N$. If there exists no k such that $b_{n_k} \leq x_n \leq a_{n_k}$, $n = 1,2,\cdots,N$, then $(x_1,x_2,\cdots,x_n)$ is assigned to category k if and only if

$$\sum_{n=1}^{N} \min \frac{\{|x_n - a_{n_k}|,|x_n - b_{n_k}|\}}{(a_{n_k} - b_{n_k})}$$

$$\geq \sum_{n=1}^{N} \min \frac{\{|x_n - a_{n_j}|,|x_n - b_{n_j}|\}}{(a_{n_j} - b_{n_j})}, \qquad j = 1,2,\cdots,K \tag{8}$$

where K is the number of categories.

Results of Image Classification Studies

1) *Photomicrograph Data Set:* A set of 8 variables, comprising the mean and variance of the textural features f_1, f_2, f_3, and f_9 (for definitions of these features, see Appendix I), computed from distance 1 gray-tone spatial-dependence matrices, was used as an input to the classifier. The data set was arbitrarily divided into a training set of 143 samples and a test set of 100 samples. Piecewise linear-discriminant functions for separating the patterns belonging to the five sandstone categories were derived using the samples in the training set. The contingency table for classifying the samples in the test set is shown in Table I. The overall accuracy of the classification scheme on the test set (obtained by summing the diagonal elements in the contingency table and dividing by the total number of samples in the test set) was 89 percent.

TABLE I
CONTINGENCY TABLE FOR THE CLASSIFICATION OF PHOTOMICROGRAPHS OF SANDSTONES

		ASSIGNED CATEGORY					Total
		Dexter-L	Dexter-H	St. Peter	Upper Muddy	Gaskel	
TRUE CATEGORY	Dexter-L	29	0	1	0	0	30
	Dexter-H	0	15	0	0	0	15
	St. Peter	2	0	22	4	0	28
	Upper Muddy	0	0	4	11	0	15
	Gaskel	0	0	0	0	12	12
	TOTAL	31	15	27	15	12	100

Number of samples in test set = 100; number of samples in training set = 143; overall accuracy of classification of test set = 89 percent.

TABLE II
CONTINGENCY TABLE FOR THE CLASSIFICATION OF THE AERIAL PHOTOGRAPHIC DATA SET

		ASSIGNED CATEGORY								TOTAL
		RSOLD	RESNU	LAKE	SWAMP	MARSH	URBAN	RAIL	SCROD	
TRUE CATEGORY	RSOLD	17	0	0	0	1	1	0	1	20
	RESNU	1	15	0	0	1	2	1	0	20
	LAKE	0	0	19	0	0	0	0	1	20
	SWAMP	0	0	0	19	1	0	0	0	20
	MARSH	0	0	0	0	12	0	0	8	20
	URBAN	2	3	0	0	0	15	0	0	20
	RAIL	0	1	0	0	0	1	5	3	10
	SCROD	0	0	0	0	2	0	0	38	40
	TOTAL	20	19	19	19	17	19	6	51	170

140 out of 170 images or 82.3 percent of the images were correctly classified.

2) *Aerial Photographic Data Set:* A min–max decision rule was used for the classification of 170 images into eight categories. The processing was done as follows. Four gray-tone spatial-dependencies (for four directions) were computed for each image. Eleven textural features (f_1–f_{11} as defined in Appendix I) were computed, yielding a measurement vector of $4 \times 11 = 44$ components for each image. Computation of the mean, range, and deviation of each feature over the four directions reduced the dimensionality to 33. The minimum and maximum statistics were computed for 169 images and the min–max decision rule was used to classify the image which was left out. This procedure was repeated 170 times to classify all the images. Out of these 170 images, 140, or 82.3 percent of the images, were correctly classified. The contingency table is given in Table II.

3) *Satellite Imagery Data Set:* The piecewise linear-discriminant function method was used for identifying the land-use category of each image block in this data set. The input variables to the classifier consisted of the mean variance of the four textural features (f_1, f_2, f_3, and f_9 obtained from the distance 1 gray-tone spatial-dependence matrices) and eight spectral features (comprised of the mean variance of the image gray-tone values) in each of the four spectral bands. The 624 samples in the data set were divided arbitrarily into a training set of 314 samples and a test set of 310 samples. The classifier was trained on the 314 samples in the training set, and each sample in the test set was assigned to one of seven possible land-use categories. The

TABLE III
CONTINGENCY TABLE FOR LAND-USE CLASSIFICATION OF SATELLITE IMAGERY

		ASSIGNED CATEGORY							
		Coastal Forest	Woodlands	Annual Grass-Lands	Urban Area	Large Irrigated Fields	Small Irrigated Fields	Water	Total
TRUE CATEGORY	Coastal Forest	23	1	2	0	0	0	1	27
	Woodlands	0	17	10	0	1	0	0	28
	Annual Grasslands	1	3	109	1	1	0	0	115
	Urban Area	0	3	10	13	0	0	0	26
	Large Irrigated Fields	1	2	6	0	37	2	0	48
	Small Irrigated Fields	0	0	4	0	3	24	0	31
	Water	0	0	0	0	0	0	35	35
	Total	25	26	141	14	42	26	36	310

Number of training samples = 314; number of test samples = 310; accuracy of classification on training set = 84.0 percent; accuracy of classification on test set = 83.5 percent.

contingency table for the classification of the test samples is given in Table III. The overall accuracy of the classifier on the test samples was found to be 83.5 percent.

Classification of multispectral data such as the satellite imagery we have processed is usually carried out using only the spectral features. We attempted a spectral classification on the satellite imagery data and achieved a classification accuracy of 74–77 percent on the test set. This result, compared with the 83.5-percent classification accuracy achieved using a combination of spectral and textural features, shows that a significant improvement in the classification accuracy might result if the textural features are used as additional inputs to the classifier.

IV. Concluding Remarks

We have described a class of quickly computable textural features which seem to have general applicability to many kinds of image data. The textural features are based on statistics which summarize the relative frequency distribution (which describes how often one gray tone will appear in a specified spatial relationship to another gray tone on the image). We have used these features in category-identification tasks of three different kinds of image data. Identification accuracy on independent test sets are 89 percent for the photomicrograph image set (five categories of sandstones), 82 percent for the aerial photographs (eight land-use categories), and 83 percent for the satellite imagery (seven land-use categories).

These initial experimental results are promising. Much work needs to be done, however, on gray-tone normalization on the imagery and the use of features which are invariant under monotonic gray-tone transformations. The reason is that in one important sense, texture is independent of tone. Two people examining photographs of the same texture may actually be seeing two different, though related, kinds of tones in the texture. One photograph may have been developed such that its tones are light and thin, and the other photograph may have been developed such that its tones are dark and heavy. Most people could easily make the observation that the texture on the two images is the same. For a machine to find that the textures are the same, either the images must be probability quantized and the features computed from the probability quantized images (which are invariant under monotonic gray-tone transformations), or the features themselves must be invariant under monotonic gray-tone transformations. Of the textural features described in Appendix I, the angular second-moment, the entropy, the sum entropy, the difference entropy, the information measure of correlation, and the maximal-correlation features have the invariance property. We intend to repeat the experiments reported here using these kinds of features. We expect that these features will provide more generalized results.

Additional investigations are necessary to determine the size of the subimage region and the distances which should be used in computing the gray-tone dependence matrices. Too small a subimage region will not have enough textural information to separate image categories of interest, while a large subimage region may have objects belonging to several different categories. Also, a large subimage region will increase the storage requirements. The distance which must be used in computing the gray-tone spatial-dependencies may be obtained from the autocorrelation function of the image. The distance at which the normalized autocorrelation function of the image becomes too small can serve as an upper bound on the distance which may be used for computing the gray-tone spatial-dependence matrices.

Appendix I
Textural Features

We suggest a set of 28 textural features which can be extracted from each of the gray-tone spatial-dependence matrices. The following equations define these features.

Notation

$p(i,j)$ (i,j)th entry in a normalized gray-tone spatial-dependence matrix, $= P(i,j)/R$.

$p_x(i)$ ith entry in the marginal-probability matrix obtained by summing the rows of $p(i,j)$, $= \sum_{j=1}^{N_g} P(i,j)$.

N_g Number of distinct gray levels in the quantized image.

$\sum_i$ and $\sum_j$ $\sum_{i=1}^{N_g}$ and $\sum_{j=1}^{N_g}$, respectively.

$$p_y(j) = \sum_{i=1}^{N_g} p(i,j).$$

$$p_{x+y}(k) = \sum_{\substack{i=1 \\ i+j=k}}^{N_g} \sum_{j=1}^{N_g} p(i,j), \qquad k = 2,3,\cdots,2N_g.$$

$$p_{x-y}(k) = \sum_{\substack{i=1 \\ |i-j|=k}}^{N_g} \sum_{j=1}^{N_g} p(i,j), \qquad k = 0,1,\cdots,N_g - 1.$$

Textural Features

1) *Angular Second Moment:*

$$f_1 = \sum_i \sum_j \{p(i,j)\}^2.$$

2) *Contrast:*

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{\substack{i=1 \\ |i-j|=n}}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right\}.$$

3) *Correlation:*

$$f_3 = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

where μ_x, μ_y, σ_x, and σ_y are the means and standard deviations of p_x and p_y.

4) *Sum of Squares:* Variance

$$f_4 = \sum_i \sum_j (i - \mu)^2 p(i,j).$$

5) *Inverse Difference Moment:*

$$f_5 = \sum_i \sum_j \frac{1}{1 + (i-j)^2} p(i,j).$$

6) *Sum Average:*

$$f_6 = \sum_{i=2}^{2N_g} i p_{x+y}(i).$$

7) *Sum Variance:*

$$f_7 = \sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i).$$

8) *Sum Entropy:*[2]

$$f_8 = -\sum_{i=2}^{2N_g} p_{x+y}(i) \log \{p_{x+y}(i)\}.$$

9) *Entropy:*

$$f_9 = -\sum_i \sum_j p(i,j) \log (p(i,j)).$$

10) *Difference Variance:*

$$f_{10} = \text{variance of } p_{x-y}.$$

11) *Difference Entropy:*

$$f_{11} = -\sum_{i=0}^{N_g-1} p_{x-y}(i) \log \{p_{x-y}(i)\}.$$

12), 13) *Information Measures of Correlation:*

$$f_{12} = \frac{HXY - HXY1}{\max \{HX, HY\}}$$

$$f_{13} = (1 - \exp [-2.0(HXY2 - HXY)])^{1/2}$$

$$HXY = -\sum_i \sum_j p(i,j) \log (p(i,j))$$

where HX and HY are entropies of p_x and p_y, and

$$HXY1 = -\sum_i \sum_j p(i,j) \log \{p_x(i)p_y(j)\}$$

$$HXY2 = -\sum_i \sum_j p_x(i)p_y(j) \log \{p_x(i)p_y(j)\}.$$

14) *Maximal Correlation Coefficient:*

$$f_{14} = (\text{Second largest eigenvalue of } Q)^{1/2}$$

where

$$Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}.$$

These measures of correlation have some desirable properties which are not brought out in the rectangular correlation measure f_3. For details, see [23], [24].

Referring to (1) in the text, for a chosen distance d we have four angular gray-tone spatial-dependency matrices. Hence we obtain a set of four values for each of the preceding 14 measures. The mean and range of each of these 14 measures, averaged over the four values, comprise the set of 28 features which can be used as inputs to the classifier. In this set of 28 features some of the features are strongly correlated with each other. A feature-selection procedure may be applied to select a subset or linear combinations of the 28 features.

Appendix II
Examples of the Significance of Some Textural Features

The problem "What do the textural features represent?" from a human perception point of view can be a subject for a thorough experiment. There are some intuitive expectations as to properties represented by some features. For example, one might expect the entropy feature to take higher values for more complex images. One might expect to be able to notice some linear dependencies in images with high values of the correlation feature. As a beginning for an investigation of this problem we used an aerial photography data set that consisted of 12 categories, 6 samples in each [17].

We ordered the categories in the following manner. We computed the range and the mode of values that each feature takes for each category. The categories were then ordered for each feature according to the modes, and the ranges were plotted. The prospective of this Appendix is to comment on the ordering of categories by some of the features. Figs. 8–10 show plots of the ranges of three of the features, each accompanied by three pictures. These pictures are of units that have small, medium, and large values of the respective feature. The order of categories of the average angular moment feature (ANGMOM) in Fig. 8 is almost the reverse of the order of categories of the average entropy (ENTROP) in Fig. 9. The three representative pictures show that in both figures the units were ordered according to the complexity of the image. Images with more gray levels have lower average ANGMOM and higher

[2] Since some of the probabilities may be zero, and $\log(0)$ is not defined, it is recommended that the term $\log(p + \varepsilon)$ (ε an arbitrarily small positive constant) be used in place of $\log(p)$ in entropy computations.

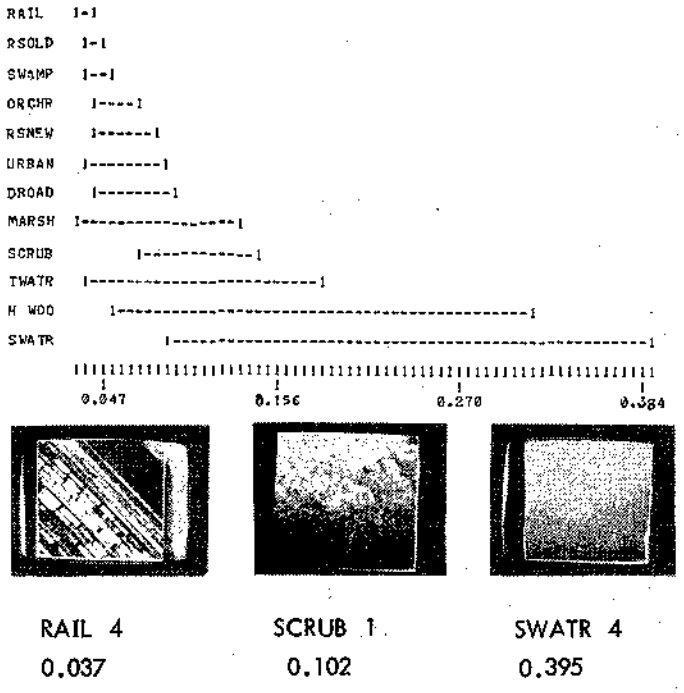


Fig. 8. Range plots and representative pictures for average second angular moment.

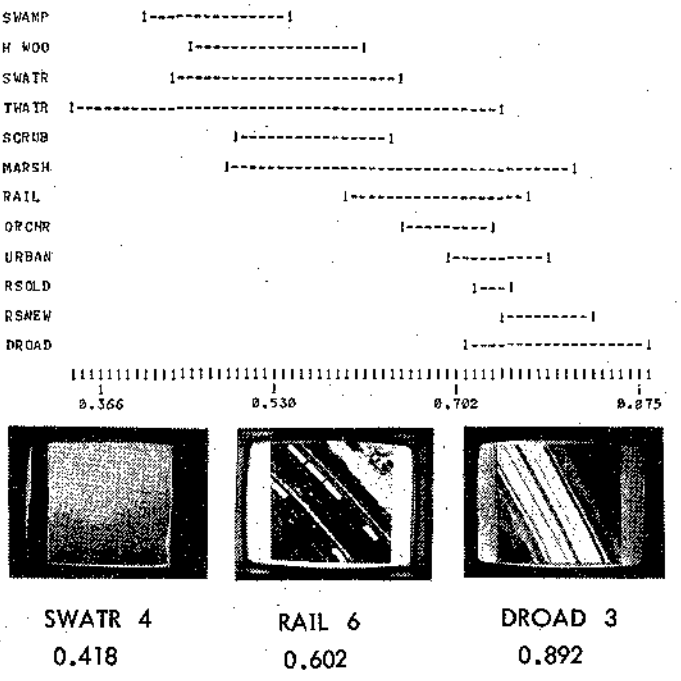


Fig. 10. Range plots and representative pictures for average correlation.

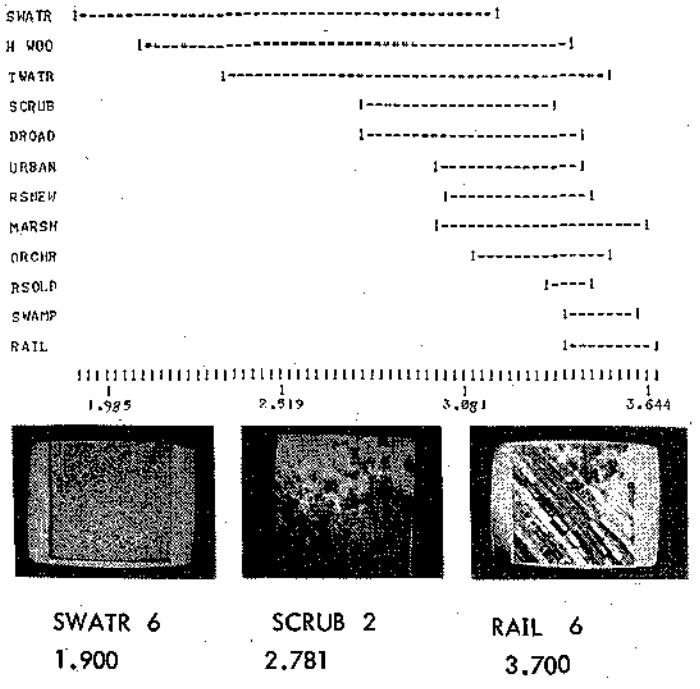


Fig. 9. Range plots and representative pictures for average entropy.

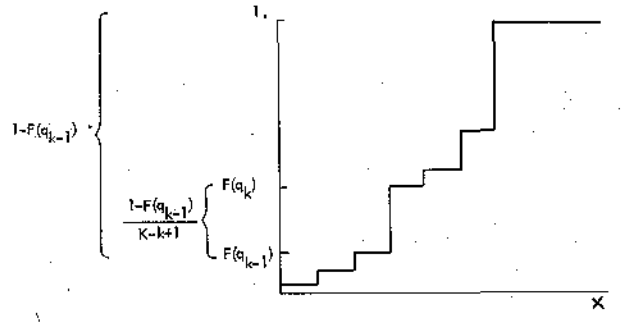


Fig. 11. Quantizing algorithm. At kth iteration, $F(q_k - 1)$ probability has already been allocated to $k - 1$ levels and $1 - F(q_k - 1)$ probability remains to be allocated. If $1 - F(q_k - 1)$ probability is split up equally among remaining $K - k + 1$ quantizing levels to be allocated, each level would get $(1 - F(q_{k-1}))/(K - k + 1)$. Since F is a step function, there is no guarantee that a q_k can be found and that $F(q_k) = F(q_{k-1}) + (1 - F(q_{k-1}))/(K - k + 1)$. Hence we look for a q_k which is closest to satisfying the equality.

average ENTROP. The average ratio (RATIO) in Fig. 10 is lower for natural scenes and higher for man-made scenes. The three representative pictures in Fig. 10 show correspondence between the average RATIO values and the linear dependency of gray levels in the images. These comments reflect our subjective perception. Much more experimentation and analysis should be done in this subject.

Appendix III
Equal-Probability Quantizing

Let X be a random variable with cumulative probability function F_X. Let Q_1, the K-level equal-probability quantizing function for X, be defined by

$Q_1(x) = k$, if and only if

$$\text{lub}\left\{w \mid F_X(w) = \frac{k-1}{K}\right\} \leq x < \text{lub}\left\{w \mid F_X(w) = \frac{k}{K}\right\}$$

where lub denotes least upper bound.

For any strictly monotonic function g, define the random variable Y by $Y = g(X)$. Let Q_2, the K-level equal-prob-

ability quantizing function for Y, be defined by

$Q_2(y) = k$, if and only if

$$\text{lub}\left\{w \mid F_Y(w) = \frac{k-1}{K}\right\} \leq y < \text{lub}\left\{w \mid F_Y(w) = \frac{k}{K}\right\}.$$

The following lemma states that the equal-probability quantization of X produces a random variable which is identical to the equal-probability quantization of Y.

Lemma: $Q_1(X) = Q_2(Y)$.

Proof: $Q_1(X) = k$, if and only if

$$\text{lub}\left\{x \mid F_X(x) = \frac{k-1}{K}\right\} \leq X < \text{lub}\left\{x \mid F_X(x) = \frac{k}{K}\right\}. \quad (9)$$

Since g is strictly monotonic, (9) holds if and only if

$$g\left(\text{lub}\left\{x \mid F_X(x) = \frac{k-1}{K}\right\}\right) \leq g(X) < g\left(\text{lub}\left\{x \mid F_X(x) = \frac{k}{K}\right\}\right). \quad (10)$$

Since g is strictly monotonic, we may reverse the order of g and lub so that (10) holds if and only if

$$\text{lub}\left\{g(x) \mid F_X(x) = \frac{k-1}{K}\right\} \leq g(X) < \text{lub}\left\{g(x) \mid F_X(x) = \frac{k}{K}\right\}. \quad (11)$$

However,

$$F_X(x) = P(X \leq x) = P(g(X) \leq g(x)) = P(Y \leq g(x)) = F_Y(g(x)).$$

Hence, (11) holds if and only if

$$\text{lub}\left\{g(x) \mid F_Y(g(x)) = \frac{k-1}{K}\right\} \leq g(X) < \text{lub}\left\{g(x) \mid F_Y(g(x)) = \frac{k}{K}\right\}. \quad (12)$$

Finally, (12) is equivalent to

$$\text{lub}\left\{y \mid F_Y(y) = \frac{k-1}{K}\right\} \leq Y < \text{lub}\left\{y \mid F_Y(y) = \frac{k}{K}\right\} \quad (13)$$

so that $Q_2(Y) = k$.

Equal Probability Quantizing Algorithm: Let X be a nonnegative random variable with cumulative probability function F_X. Let Q, the K-level equal-probability quantizing function for X, be defined by $Q(x) = k$ if and only if $q_{k-1} \leq x < q_k$. We define $q_0, q_1, q_2, \cdots, q_k$ in an iterative manner. Let $q_0 = 0$. Suppose q_{k-1} has been defined. Then let q_k be the smallest number such that

$$\left|\frac{1 - F_X(q_{k-1})}{K - k + 1} + F_X(q_{k-1}) - F_X(q_k)\right| \leq \left|\frac{1 - F_X(q_{k-1})}{K - k + 1} + F_X(q_{k-1}) - F(q)\right|, \quad \text{for all real } q.$$

Fig. 11 illustrates the equal-probability quantizing algorithm.

REFERENCES

[1] *Special Issue on Digital Picture Processing*, *Proc. IEEE*, vol. 60, pp. 763–922, July 1972.

[2] H. C. Andrews, A. G. Tescher, and R. P. Kruger, "Image processing by digital computer," *IEEE Spectrum*, vol. 9, pp. 20–32, July 1972.

[3] H. Kaizer, "A quantification of textures on aerial photographs," Boston Univ. Res. Lab., Boston, Mass., Tech. Note 121, 1955, AD 69484.

[4] R. Chevallier, A. Fontanel, G. Grau, and M. Guy, "Application du filtrage optique a l'etude des photographies aeriennes," presented at the 11th Congr. Int. Photogrammetric, July 1968.

[5] R. Bixby, G. Elerding, V. Fish, J. Hawkins, and R. Loewe, "Natural image computer," Aeronutronic Division, Philco-Ford Corp., Newport Beach, Calif., Final Tech. Rep., vol. I, Publ. C-4035, May 1967.

[6] E. M. Darling and R. D. Joseph, "Pattern recognition from satellite altitudes," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-4, pp. 38–47, Mar. 1968.

[7] A. Rosenfeld and E. Troy, "Visual texture analysis," Comput. Sci. Cent., Univ. Maryland, College Park, Tech. Rep. 70-116, June 1970.

[8] A. Rosenfeld and M. Thurston, "Visual texture analysis, 2," Comput. Sci. Cent., Univ. Maryland, College Park, Tech. Rep. 70-129, Sept. 1970.

[9] E. B. Troy, E. S. Deutsch, and A. Rosenfeld, "Gray-level manipulation experiments for texture analysis," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, pp. 91–98, Jan. 1973.

[10] *Proc. Symp. Computer Image Processing and Recognition*, Univ. Missouri, Columbia, Aug. 24–26, 1972.

[11] R. M. Pickett, "Visual analysis of texture in the detection and recognition of objects," in *Picture Processing and Psychopictorics*, B. S. Lipkin and A. Rosenfeld, Eds. New York: Academic Press, 1970, pp. 289–308.

[12] A. Rosenfeld and B. S. Lipkin, "Texture synthesis," in *Picture Processing and Psychopictorics*, B. S. Lipkin and A. Rosenfeld, Eds. New York: Academic Press, 1970, pp. 309–345.

[13] P. Brodatz, *Textures*. New York: Dover, 1966.

[14] R. M. Haralick, K. Shanmugam, and I. Dinstein, "On some quickly computable features for texture," in *Proc. Symp. Computer Image Processing and Recognition*, Univ. Missouri, Columbia, vol. 2, pp. 12-2-1–12-2-8, Aug. 1972.

[15] E. E. Hall *et al.*, "A survey of preprocessing and feature extraction techniques for radiographic images," *IEEE Trans. Comput.*, vol. C-20, pp. 1032–1044, Sept. 1971.

[16] R. M. Haralick and D. Anderson, "Texture–tone study with applications to digitized imagery," Univ. Kansas Center for Research, Inc., Lawrence, CRES Tech. Rep. 182–2, Nov. 1971.

[17] R. M. Haralick, I. Dinstein, K. Shanmugam, and D. Goel, "Texture–tone study with applications to digitized imagery," Univ. Kansas Center for Research, Inc., Lawrence, CRES Tech. Rep. 182-4, Dec. 1972.

[18] R. M. Haralick and K. Shanmugam, "Computer classification of reservoir sandstones," *IEEE Trans. Geosci. Electron.*, vol. GE-11, pp. 171–177, Oct. 1973.

[19] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York: Academic Press, 1972.

[20] K. S. Fu and J. M. Mendel, *Adaptive Learning and Pattern Recognition Systems*. New York: Academic Press, 1972.

[21] W. Miesel, *Computer Oriented Approaches to Pattern Recognition*. New York: Academic Press, 1972.

[22] N. J. Nilsson, *Learning Machines*. New York: McGraw-Hill, 1965.

[23] E. H. Linfoot, "An information measure of correlation," *Inform. Contr.*, vol. 1, pp. 85–89, 1957.

[24] C. B. Bell, "Mutual information and maximal correlation measures of dependence," *Ann. Math. Statist.*, vol. 43, pp. 587–595, 1962.

[25] P. N. Anderson *et al.*, "An automated image processing system—IDECS," Ph.D. project, Univ. Kansas, Lawrence, 1972.