

- [18] D. E. Morrison and R. E. Henkel, Eds., *The Significance Test Controversy*. Chicago: Aldine, 1970.
- [19] D. W. Peterson and F. C. Schewpe, "Code for a general purpose system identifier and evaluator (GPSIE)," *IEEE Trans. Automat. Contr.*, vol. AC-19, p. 852, 1974.
- [20] E. H. Phelps Brown, "The underdevelopment of economics," *Economic J.*, vol. 82, pp. 1-10, Mar. 1972.
- [21] P. A. Samuelson, "Interactions between the multiplier analysis and the principle of acceleration," *Rev. Economics and Statistics*, vol. 21, pp. 75-78, May 1939.
- [22] T. P. Schultz, "An economic model of family planning and fertility," *J. Political Economy*, vol. 77, pp. 153-180, Apr. 1969.
- [23] F. C. Schewpe, *Uncertain Dynamic Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1973.
- [24] P. M. Senge, "Statistical estimation of feedback models," *Simulation*, vol. 28, no. 6, pp. 177-184, June 1977.
- [25] ———, "The system dynamics national model investment formulation: A comparison to the neoclassical model" [Ph.D. dissertation System Dynamics Group, Alfred P. Sloan School of Management], Cambridge, MA: Massachusetts Institute of Technology, 1978.
- [26] A. Sinai, "Introduction to applied economic and financial market forecasting" [Lecture Notes, Part I: Forecasting with Econometric Models: Method, Outputs, and Performance]. Lexington, MA: Data Resources, 1975.
- [27] ———, "Econometric modelbuilding with applications to industry and finance" [Lecture Notes, Part IB: Econometric Models: Construction, Validation, and Forecasting], Lexington, MA: Data Resources, 1976.
- [28] H. Theil, *Principles of Econometrics*. New York: Wiley, 1971.
- [29] G. D. N. Worswick, "Is progress in economic science possible?" *Economic J.*, vol. 82, pp. 73-86, Mar. 1972.
- [30] A. Zellner, "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," *J. American Statistical Association*, vol. 57, pp. 348-368, June 1962.

Textural Features Corresponding to Visual Perception

HIDEYUKI TAMURA, MEMBER, IEEE, SHUNJI MORI, AND TAKASHI YAMAWAKI

Abstract—Textural features corresponding to human visual perception are very useful for optimum feature selection and texture analyzer design. We approximated in computational form six basic textural features, namely, coarseness, contrast, directionality, line-likeness, regularity, and roughness. In comparison with psychological measurements for human subjects, the computational measures gave good correspondences in rank correlation of 16 typical texture patterns. Similarity measurements using these features were attempted. The discrepancies between human vision and computerized techniques that we encountered in this study indicate fundamental problems in digital analysis of textures. Some of them could be overcome by analyzing their causes and using more sophisticated techniques.

I. INTRODUCTION

IN THE STUDY of digital picture recognition, edge detection and region analysis have been considered as basic approaches to feature extraction and analysis of an image. Some success has been attained by appropriate

Manuscript received September 22, 1977; revised February 13, 1978. This work was supported by the Pattern Information Processing System Project of the Agency of Industrial Science and Technology, Ministry of International Trade and Industry, Japan.

H. Tamura was with the Division of Electrical Engineering, National Research Council of Canada, Ottawa, ON K1A 0R8, Canada, on leave from the Pattern Processing Section, Information Sciences Division, Electrotechnical Laboratory, 2-6-1 Nagata-cho, Chiyoda-ku, Tokyo 100, Japan.

S. Mori is with the Information Sciences Division, Electrotechnical Laboratory, Tokyo 100, Japan.

T. Yamawaki was with the Department of Electronics and Communication, Waseda University, Tokyo 160, Japan. He is now with the Nippon Electric Co., Ltd., Tokyo 108, Japan.

combinations of local operators if the tone in a region is almost uniform. Currently, much work is directed towards the textured regions which have internal structure such as stripes or grids, i.e., regions which are too complicated to be processed by local operators alone.

Various kinds of textural features for image classification have been proposed [1]–[6]. It should be noted that all these texture measures are defined for a segmented region. In order to determine the boundaries of textured regions, in other words to find texture edges, further segmentation techniques are required. Unfortunately, since the current texture edge detectors [7]–[11] work successfully only between regions with different average gray-levels, it seems that these textural features are not directly usable for the region segmentation problem. However, indirectly, it is possible to use these measurements as the "gray-levels" in the segmentation process as done by Carlton and Mitchell [12]. Thus we believe it to be significant even to develop the segmentation of images and to refine measures which reflect specified textural properties as faithfully as possible.

In either classification or segmentation, it is important to select the set of properties to be used for identifying two or more textured regions. Textural properties are defined for a region or subimage, not for a point. Pairs of regions are often used in psychological experiments and in fundamental studies of computational measurements. In this situation, if two adjacent regions differ only in scale, contrast, orientation, or shape of repetitive elements, we may be able to perceive two different textures and detect the texture boun-

dary between these two regions. However, in segmentation of a large image containing many textured regions, the same discriminating criterion is not necessarily used for each pair of individual regions. For example, let us consider two directional regions, one of which is a rotated version of the other and in which all other properties are the same. When this pair of regions is juxtaposed, a texture edge detector is required to segment them. On the other hand, if these are crop fields in a remotely sensed image, they should belong to the same category on the land-use map even if they are adjacent. Similar comments can be made for pairs of regions in which contrast is different, in which regularity of repetition varies, and so on.

It depends on a given situation and purpose whether we should treat such regions as the same category or not. In most problems of digital picture analysis, these conditions can be specified by a human being, i.e., an analyst for that problem. Thus it is preferable to decompose textural properties into parameters such as coarseness, contrast, directionality, etc. Especially in the man-aided approach, it is more convenient to have a set of features which are appropriate to human perception of visual textures.

From this point of view, the texture measures so far proposed are not suitable. Above all, Haralick gave an easily computable method [1] which was best evaluated in two comparative studies [13], [14]. However, all of his 14 features are not obvious visually. Sometimes even random selection of features may give satisfactory accuracy in a classification problem, especially if they are orthogonal by accidents. Here, it is important to analyze misassignments and to know the behavior of each feature. It may be applied to classifier or analyzer design oriented towards a problem.

Motivated by the above considerations, our challenge is to develop the textural features approximating visual perception. Bajscy described some outdoor scenes as "blob-like," "homogeneous," or "directional" and tried to combine them with features in Fourier domain [4]. In her study, however, all descriptions were derived only from her subjective view and did not correspond to the Fourier features quantitatively. We proceed to generalize her descriptions by using more representative textural patterns which are perceptually evaluated by many human subjects.

With respect to perception of textures, considerable work has been performed in the psychological fields. Above all, Julesz verified that the discrimination of textures depends mostly on the difference in second-order statistics [15]–[18]. However, as he mentioned himself [16], the work is directed towards the extreme behavior of the perceptual mechanism for studying the extent to which one can just perceive differences in artificially produced patterns when all familiar cues are removed. This standpoint is quite far from ours. For our purpose, we consider more general and familiar features of naturally occurring textures in which no illusion can influence the human subjects.

In this paper, we describe the psychological experiments on basic textural properties as mentioned above and discuss the results to assess the corresponding computational measures derived from the psychological specifications.

II. BASIC TEXTURAL PROPERTIES

A. Structure of Texture Patterns

A strict definition for visual texture is difficult. However, from an analytical point of view, we may regard texture as what constitutes a macroscopic region. Its structure is simply attributed to the repetitive patterns in which elements or primitives are arranged according to a "placement rule" [19]. Hence it can be written as

$$f = R(e)$$

where R is denoting a placement rule (or relation) and e is denoting an element. We can get texture f if R is a function which satisfies notions mentioned above, and we can express "microtexture" if e is recursively replaced by f . On the more sophisticated model, see Zucker [20].

Texture is often considered on two levels, statistical and structural. On the statistical level, texture is observed from macroscopic view as a functional combination of R and e . Thus the difference between textured regions is measured by using the statistics of local properties. On the other hand, the structural level requires more information about each of R and e . Texture synthesis [21] at the structural level can be performed to produce various kinds of different patterns by varying relatively simple versions of R and/or e . Other types of textures, which are generated on a random field such as time-series model [22], [23] rather than by simple repetition, can be considered to be on the statistical level.

Considering a computational analysis of a given texture pattern, it is fairly difficult to parse its structure generally because both factors R and e influence the texture pattern. It is relatively easy to describe the structure of a region when R and e are well defined or when either R or e is known. This case is called the structural analysis [24], [25]. For the naturally occurring textures, since R and e generally have a considerable variation, it seems impossible to approximate them separately. In this case, we are obliged to take a statistical approach where a set of statistics is extracted from local measurements.

It may be true that second-order statistics produce the difference between textures, but we do not have a universal method of determining which of R and e cause the total difference and what kinds of statistics absorb the variation of each. In effect, an infinite set of texture patterns can be easily generated in addition to natural textures, but the methods that have been developed can discriminate only a part of them. For textures which do not differ so much in their elements, the method may work successfully by using the statistical difference of local properties. However, it is easy to give examples which the method cannot overcome, where the behavior of R and e collectively cancel the difference in their statistics. This is the difficulty in texture analysis.

In practical applications, statistics with limited potential may be powerful enough for a specified problem. This is in fact shown in the two comparative studies mentioned previously. The study by Conners and Harlow [14] theoretically concluded that Haralick's "gray-level co-occurrence matrices" [1] had the best "innate discriminative ability."

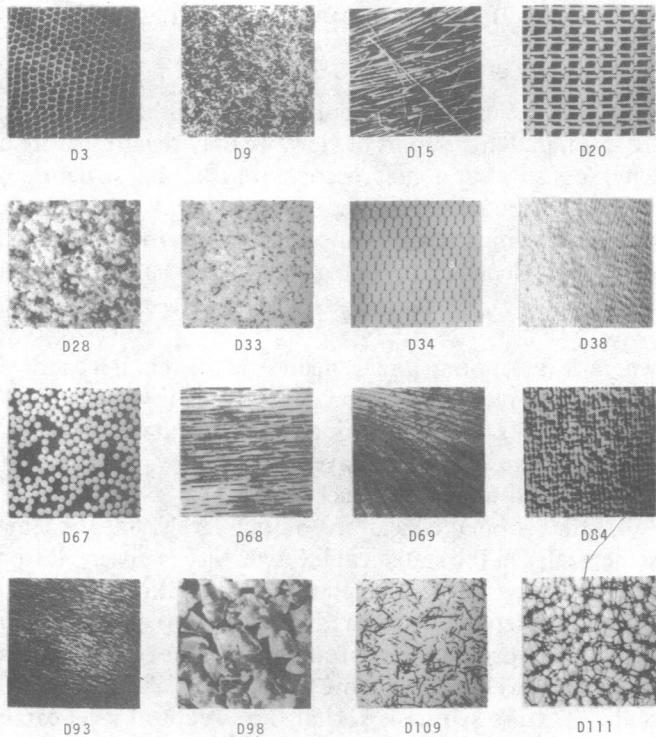


Fig. 1. Natural texture patterns [26].

Nevertheless, it is experimentally shown in the study by Weszka *et al.* [13] that their "gray-level difference statistics," which can be derived from Haralick's matrices, are computationally easier and perform as well as the more complicated "co-occurrence statistics" for their problem of terrain classification. This means that some practical problems can be solved sufficiently by using simpler statistics.

It is significant to investigate the useful methods in a practical situation including limited texture patterns. However, in order to carry the texture analysis techniques to the more generalized stage, we should develop textural features that can be effective on as many typical texture patterns as possible. Here we should remember that all the features are not at the same level. Some of them may be used after others in multistage or hierarchical decision process. Human texture-discrimination seems to employ only simple feature extractors whose outputs are combined in pairs [17]. Hence our initial efforts should be directed towards finding features, each of which reflects one of the simple textural properties.

B. Specifications of Basic Textural Features

Before determining what should be selected as basic parameters for textural properties and how they should be specified, we looked for the descriptions for textural properties in the literature. Then we tried to find general descriptions common over all texture patterns in Brodatz's photographic album [26], which had been extensively used in several fundamental studies. The pictures in Fig. 1 are a part of [26].

Also, these descriptions were given to other investigators working in the pattern recognition field. Some of the given descriptions were peculiar to the objects, e.g., "net-like" for

D3 and D34, "ripple-like" for D38, etc., (original descriptions in Japanese). It may not be impossible to translate each description into computational form, and we cannot predict how such a feature will behave for other patterns than the one from which it was derived. In the extreme case, we may need a specific description for each type of texture. However, these sets of features would be impotent to a newly occurring pattern.

Here we want to have a set of features by which all input patterns are measured and which give well-distributed results. For this purpose, it is required to have both extremes in the concept of each feature, e.g., coarse versus fine for coarseness. We selected the following six specifications which seemed to be common to all visual textures.

1) *Coarseness—Coarse versus Fine*: The coarseness is the most fundamental textural feature and has been much investigated since early studies [27], [28]. Sometimes in the narrow sense, the texture means the coarseness. When two patterns differ only in scale, the magnified one is coarser. For patterns with different structures, the bigger its element size and/or the less its elements are repeated, the coarser it is felt to be. Instead of fine, in some cases, busy or close is used.

2) *Contrast—High Contrast versus Low Contrast*: The simplest method of varying picture contrast is stretching or shrinking of its gray scale; i.e., given a picture $f(x, y)$, we can vary its contrast by multiplying it by a positive constant, obtaining $cf(x, y)$ [29]. This model is designed for changing a picture quality, not picture structure. The contrast knob on a television set is a practical realization of the above model. Thus, based on this idea, when two patterns differ only in gray-level distribution, the difference between their contrast can be measured. However, more factors are supposed to influence the contrast difference between two texture patterns with different structures. We assume the following four factors for contrast:

- i) dynamic range of gray-levels,
- ii) polarization of the distribution of black and white on the gray-level histogram or ratio of black and white areas,
- iii) sharpness of edges,
- iv) period of repeating patterns.

In the narrow sense, contrast stands for picture quality. Factors i) and ii) are used for television images and photo printing. Factor iii) can be seen by comparing two images with the same gray-level distribution. The image with sharp edges has higher contrast. When we see two checkerboard patterns which differ only in scale, we can perceive the difference of contrast caused by Factor iv).

3) *Directionality—Directional versus Nondirectional*: This is a global property over the given region. Directionality involves both element shape and placement rule. Bajcsy [3], [4] divided directionality into two groups, monodirectional and bidirectional, but we measure simply the total degree of directionality. Here the orientation of the texture pattern does not matter; i.e., two patterns which differ only in orientation should have the same degree of directionality.

4) *Line-Likeness—Line-Like versus Blob-Like*: This concept is concerned only with the shape of a texture element.

Other descriptions of the element may be possible in a limited situation, but we did not find any other common descriptions. We expect that this feature supplements the three major ones just mentioned, especially when patterns cannot be discriminated by directionality. For example, we assume that all of the D28, D67, D109, and D111 are nondirectional and that D109 and D111 are considerably more line-like than D28 and D67.

5) *Regularity—Regular versus Irregular*: This is a property for variations of a placement rule. However, it can be supposed that variation of elements, especially in case of natural textures, reduces the regularity on the whole. Additionally, a fine texture tends to be perceived as regular.

6) *Roughness—Rough versus Smooth*: This description was originally meant for tactile textures, not for visual textures. However, when we observe natural textures such as in Fig. 1, we will be able to compare them in terms of rough or smooth. Is this subjective judgement due to the total energy of changes in gray-levels or due to our imaginary tactile sense when we touch it physically?

We are interested in what other features are correlated with roughness as judged by human subjects and the extent to which these features can be approximated mathematically. That is the reason why we add this feature to the following experiments.

III. PSYCHOLOGICAL EXPERIMENTS

A. Experimental Conditions

The purpose of our experiments was to construct psychometric prototypes with which the computational measures could be compared. Thus in the psychological experiments, we did not use the real pictures from Brodatz's album but rather the digital versions.

This data set of visual textures consists of 16 representative patterns shown in Fig. 1. (We believe that this selection has sufficient universality.) A part of each of the original picture was photographed on a 35 mm negative film and scanned by a drum scanner. The data were digitized into 256×256 pixels and quantized into 64 gray-levels, but the quantization was adjusted so that the resultant gray-levels for all pictures cover as wide a range as possible in the 64 levels. We did not perform any other operation on the gray-scale (e.g., histogram flattening or range-width regulation). The digital pictures were displayed on a CRT monitor and photographed, and the slides were shown to the human subjects.

The human subjects consisted of 28 men and 20 women. Although a few people out of them were working in the picture processing field, nobody was concerned with this study. In advance of the experiments, a brief explanation of the basic concept of texture and the six specifications mentioned above was given to them (in Japanese, except for the names of features such as "coarseness-coarse versus fine" which were in English).

The method of judgement adopted in our experiments was the pair comparison, well known in the psychometrics field [30]. All possible pairs of slides from 16 patterns were shown to the human subjects. They were obliged to make

TABLE I
ORDERING BY AUTHORS OF 16 TYPICAL TEXTURES: $\{f'\}$

| f'_{crs} | f'_{con} | f'_{dir} | f'_{lin} | f'_{reg} | f'_{rgh} |
|------------|------------|-------------------|-------------|-------------|------------|
| (coarse) | (high) | (directional) | (line-like) | (regular) | (rough) |
| D98 | D20 | D68 | D34 | D34 | D98 |
| D20 | D67 | D15 | D68 | D20 | D28 |
| D28 | D111 | D93 | D15 | D84 | D84 |
| D111 | D15 | D69 | D69 | D38 | D15 |
| D69 | D84 | D38 | D109 | D9 | D20 |
| D68 | D98 | D84 | D111 | D111 | D3 |
| D67 | D109 | D20 | D3 | D3 | D9 |
| D109 | D28 | D34 | D38 | D68 | D111 |
| D34 | D3 | D3 | D20 | D93 | D109 |
| D3 | D9 | D98 | D93 | D33 | D67 |
| D33 | D68 | D109 | D84 | D15 | D33 |
| D84 | D69 | D111 | D98 | D67 | D93 |
| D15 | D34 | D9 | D33 | D28 | D34 |
| D9 | D93 | D33 | D9 | D109 | D68 |
| D38 | D33 | D28 | D28 | D69 | D69 |
| D93 | D38 | D67 | D67 | D98 | D38 |
| (fine) | (low) | (non-directional) | (blob-line) | (irregular) | (smooth) |

decisions according to the six specifications, i.e., to choose the coarser, the higher in contrast, the more directional, the more line-like, the more regular, and the rougher pattern of each pair.

We processed these results by applying "the law of comparative judgement" for constructing the psychometrics. This law, developed by Thurstone [31], is used to produce a one-dimensional scale for each stimulus (or feature) from the "total matrix of superiorities" collected through pair comparisons.

B. Considerations of the Results

The experimental results are shown in Fig. 2 as scales on one-dimensional distributions. In the following text and tables, we denote the results of these psychological experiments by f with subscripts abbreviating the feature name, i.e., f_{crs} , f_{con} , f_{dir} , f_{lin} , f_{reg} , and f_{rgh} , and we let $\{f\}$ be the set of them. For comparative purposes, we show in Table I the orderings by the authors for the same 16 objective patterns where the symbol f' is used instead of f . The authors were afraid that they were prone to make biased orderings due to their *a priori* ideas on computational definitions. Consequently, we used many human subjects.

We compare these measurements with the authors' ordering $\{f'\}$ to estimate the dispersion of individual subject's measurements. In order to examine the independence of each feature, we used rank correlation, since all measurements were based on comparative judgement. The well-known Spearman's coefficient of rank correlation is given as¹

$$r_s = 1 - \frac{6}{n^3 - n} \cdot \sum_{i=1}^n d_i$$

where d_i is the difference between the ranks assigned to the i th object in two measurements. This coefficient gives a value between +1 and -1 as does the popular correlation coefficient. In Table II we show the resultant rank correlations between all two measurements out of either $\{f\}$ or $\{f'\}$.

¹ Strictly speaking, it should be noticed that the formula above becomes more complicated if more than one measurement has the same rank. In our preliminary report of this study [32], we used Kendall's coefficient r_k as well as r_s . It can be shown that the expected value of r_k is smaller than that of r_s . The fact r_k provided the uniformly smaller values than r_s in all our measurements. Thus in this paper, we use only r_s to discuss the degree of correspondences between two measurements.

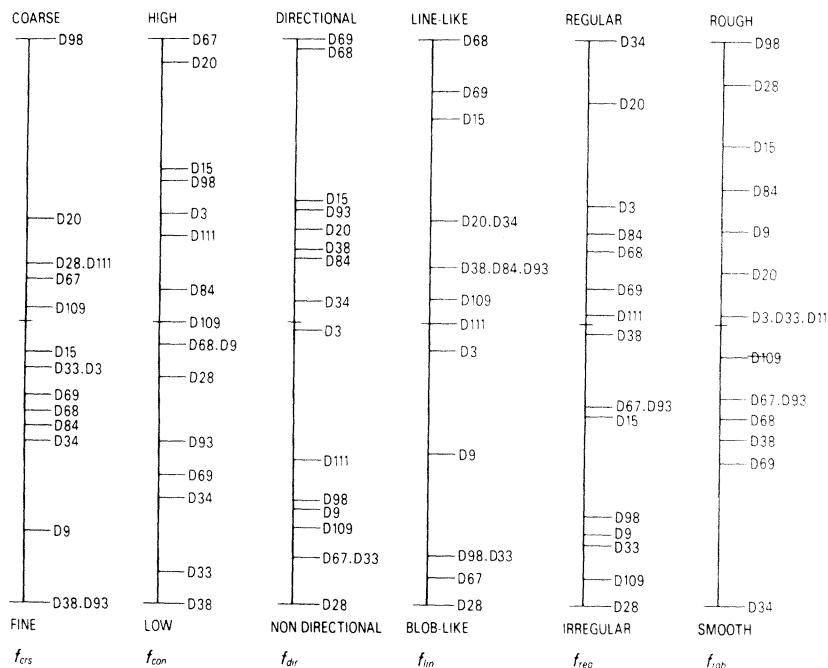


Fig. 2. Psychological measurements of six basic textural features: $\{f\}$.

TABLE II
COEFFICIENTS OF RANK CORRELATION BETWEEN
PSYCHOLOGICAL MEASUREMENTS $\{f\}$ AND $\{f'\}$

| r_s | f_{rgh} | f_{reg} | f_{lin} | f_{dir} | f_{con} |
|---------------------|------------|------------|------------|------------|-----------|
| f_{crs} | 0.535 | -0.210 | -0.372 | -0.383 | 0.658 |
| f_{con} | 0.504 | 0.102 | -0.136 | -0.074 | |
| f_{dir} | -0.361 | 0.594 | 0.923 | | |
| f_{lin} | -0.458 | 0.596 | | | |
| f_{reg} | -0.421 | | | | -0.141 |
| | | | | 0.179 | -0.409 |
| | | | 0.676 | 0.182 | -0.315 |
| | -0.235 | -0.159 | -0.062 | 0.662 | |
| | 0.506 | -0.276 | -0.012 | -0.329 | 0.324 |
| f'_{con} | f'_{dir} | f'_{lin} | f'_{reg} | f'_{rgh} | r_s |
| r_s | | | | | |
| $f_{crs}: f'_{crs}$ | 0.824 | | | | |
| $f_{con}: f'_{con}$ | 0.920 | | | | |
| $f_{dir}: f'_{dir}$ | 0.954 | | | | |
| $f_{lin}: f'_{lin}$ | 0.823 | | | | |
| $f_{reg}: f'_{reg}$ | 0.652 | | | | |
| $f_{rgh}: f'_{rgh}$ | 0.945 | | | | |

Top: Between two measurements from $\{f\}$.

Middle: Between two measurements from $\{f'\}$.

Bottom: Between corresponding measurements in $\{f\}$ and $\{f'\}$.

and between $\{f\}$ and $\{f'\}$ for each feature. In the following, we denote these correlations as $f_{\text{crs}}: f'_{\text{crs}} = 0.824$.

Judging from these results, our observations for the psychological experiments are the following.

a) There exists considerable correlation between f_{crs} , f_{con} , and f_{rgh} . The high score in $f_{\text{crs}}:f_{\text{con}}$ can be understood to be caused by contrast factor iv).

b) Judging from the correlation coefficients, we seem to

visually perceive high roughness in a texture when it is coarse, high in contrast, nondirectional, blob-like, and/or irregular. Above all coarseness and contrast greatly influence roughness, which can be realized intuitively. However, it should be noted that there exists the possibly intrinsic ambiguity between rough and coarse for the untrained subjects. In fact, $f'_{\text{crs}}; f'_{\text{rgb}}$ is not so high as $f_{\text{crs}}; f_{\text{rgb}}$. If we had used smoothness for the feature name instead of roughness, such an ambiguity might have been avoided.

c) The correlation between f_{dir} and f_{lin} is extremely high. Most of the directional textures consist of line-like elements and considerable correlation could be expected. However, as mentioned in Section II-B, we specified line-likeness as a property of a texture element such that D34, D109, and D111 could have high values compared to D38 or D84. In f_{lin} , this was not to be attained, probably because line-likeness was interpreted to be a global property by our human subjects.

d) In the bottom of Table II, we find a remarkable discrepancy between f_{reg} and f'_{reg} . The biggest difference is caused by D69. Although we specified regularity as a property of a placement rule, the experimental results gave relatively high positions to D67 and D69, for which we can describe the texture element easily, whereas the so-called scramble textures such as D9, D28, and D33, for which it is hard to determine the element, were located in the bottom of f_{reg} . It can be supposed that our human subject interpreted regularity as variance of texture elements.

From the observations above, we realized the difficulties in describing and explaining the visual specifications. Unfortunately, it is nearly impossible to repeat psychological experiments under the same condition when changing only a part of the experiment. There existed in our prototypes $\{f\}$ several deficiencies caused by imperfect understanding of the definitions. However, judging from the bottom of Table II, the correspondences between $\{f\}$ and $\{f'\}$ except for

regularity are satisfactory. We cannot expect quite the same measurement for all human beings in visual examination. Individual differences must be expected. In this sense $\{f\}$: $\{f'\}$ could give some criterion of evaluating the differences. It seems too much to expect for the computational techniques to give better correspondence than those which exist between individuals. Hence we attempt to develop computational measures corresponding to $\{f\}$ which give r_s over 0.82.

IV. COMPUTATIONAL DEFINITIONS OF BASIC TEXTURAL FEATURES

This section describes a set of computational measures corresponding to the specifications given in Section II. In the preliminary report [32], we presented the first versions of four texture measures. Since then we have tried to improve them and develop measures for regularity and roughness which were not contained in [32].

The set of measures given in this paper is the best set that we have obtained so far by the criterion of getting the highest r_s between the computational and psychological measurements $\{f\}$. We have always been willing to examine all techniques proposed so far, and to modify them if necessary.

In the following, we describe our present definitions of each feature and also consider some improvements. As in f and f' , we denote each definition or computational result by F with subscripts.

A. Coarseness

Rosenfeld and his colleagues have proposed and examined a variety of coarseness measures since the early days [27]. Considering several problems in coarseness such as the undesirable dependence on contrast (in the narrow sense) or the presence of microtexture, they developed a new method by using various sized operators [28]. The essence of this method is to pick a large size as best when coarse texture is present even though microtexture is also present but to pick a small size when only fine texture is present. This procedure can be summarized in the following steps.

Step 1: Take averages at every point over neighborhoods whose sizes are powers of two, e.g., $1 \times 1, 2 \times 2, \dots, 32 \times 32$. The average over the neighborhood of size $2^k \times 2^k$ at the point (x, y) is

$$A_k(x, y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} f(i, j)/2^{2k}$$

where $f(i, j)$ is the gray-level at (x, y) .

Step 2: For each point, at each point, take differences between pairs of averages corresponding to pairs of non-overlapping neighborhoods just on opposite sides of the point in both horizontal and vertical orientations. For examples, the difference in the horizontal case is

$$E_{k,h}(x, y) = |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)|.$$

Step 3: At each point, pick the best size which gives the highest output value:

$$S_{best}(x, y) = 2^k$$

where k maximizes E in either direction, i.e.,

$$E_k = E_{max} = \max (E_1, E_2, \dots, E_L).$$

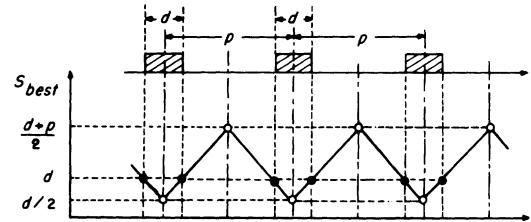


Fig. 3. Output of best size detector for a repetitive pattern (one-dimensional case).

Step 4: Finally, take the average of S_{best} over the picture to be a coarseness measure F_{crs} :

$$F_{crs} = \frac{1}{m \times n} \sum_i^m \sum_j^n S_{best}(i, j)$$

where m and n are the effective width and height of the picture, respectively. In effect, it should be noticed that boundary strips of a picture within the width of the largest operator size 2^L cannot be processed properly.

Now we illustrate the behavior of this method. For simplicity, let us consider one-dimensional repetitive spikes. Fig. 3 shows the output or the best size at each point. If we denote the width of a spike and its period by d and p , respectively, then the expected value for the best size is $(2d + p)/4$. It can be extended to the two-dimensional case. Thus it should be realized that this method is not a size detector of texture elements, but that the final output F_{crs} is influenced by both d and p . The element size d can be obtained only for a checkerboard pattern. This effect is not contrary to our specification of coarseness, and we adopted the Rosenfeld method almost as it was given.

In the stage of determining the best size, if the maximum of E 's appears in more than two operator sizes, the largest size should be taken. This can be readily understood by considering the two-valued checkerboard. At a point on an edge of a square, the value of E stays constant as the averaging size grows (E is the gray-level difference of alternate squares). Just after the averaging size exceeds the size of a square element d , there occurs a falloff in the output of E . Consequently, the best size is determined to be d . This is shown in Fig. 4(a). For natural textures as shown in Fig. 1, there are considerable variations in gray-levels, element size, and placement rule. In such a situation, differences between adjacent large neighborhoods tend to be canceled to some extent by averaging in each, even though macrotexture still produces clear differences in a small-size operator. As shown in Fig. 4(b), there exists a case where the maximum decision may produce a drastic change in the best size even though the difference between the maximum and the second maximum values of E_k is extremely small. We should pay attention to this effect. In texture edge detection [8], it is not too serious which size is taken to be the best size, since the final output at each point is the value of E_{max} , not the size k .

Due to the considerations above, we modified the procedure of determining the best size as follows. At each point, if there exist some k such that $k > k_{max}$ which gives E_{max} and $E_k \geq tE_{max}$, we take the largest k for S_{best} where t is a certain constant less than 1; otherwise we retain the original

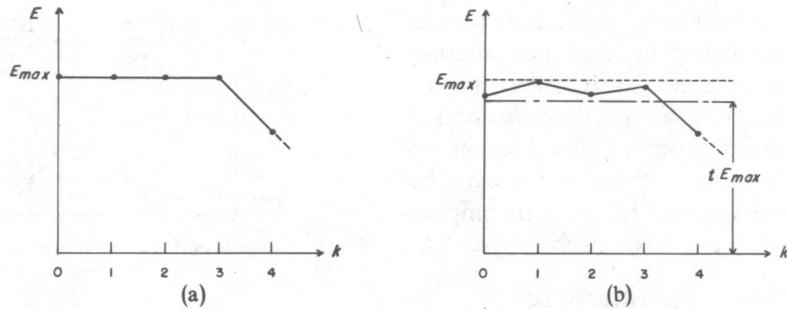


Fig. 4. Behaviors of the best size detector. (a) Ideal case. (b) Noisy case.

procedure. Each value of F_{crs} may be sensitive to the variation of t to some extent, but in practice, total correspondence between f_{crs} : F_{crs} was very stable around $t = 0.9$, as given later.

B. Contrast

The following proposed by Haralick *et al.* [1] was called contrast:

$$f_2 = \sum_n n^2 \left(\sum_{|i-j|=n} p(i, j) \right)$$

where $p(i, j)$ is the (i, j) th entry of the $n \times n$ gray-level spatial dependence matrix. This is the moment of inertia about the main diagonal of the matrix. Thus it can also be obtained from the histogram of the frequencies of gray-level differences as noticed by Weszka *et al.* [13]. What should be noted here is that such a matrix or histogram is derived from the picture on which a gray-scale "histogram flattening" transformation was performed. This transformation is said to be used for the purpose of removing the effects of unequal overall brightness and contrast in the original images [1], [13]. If f_2 is calculated after such a transformation, why should it be called contrast? There is evidently a confusion in terms that seems to be very crucial to our problem as seen in the psychological experiments.

In the sense specified in Section II-B, the contrast that is removed by the histogram flattening is given by Factors i) and ii), i.e., the narrow sense of contrast. On the other hand, the inertia moment of gray-level difference f_2 can reflect Factor iii). Originally, it was defined to be a measure of the amount of local variation present in an image [1]. However, it should be noted that the measure f_2 is a function not only of edge magnitude but also of the number of edges. The more edges an image has, the larger is the value of f_2 , which is contrary to Factor iv). In fact, this feature f_2 has been used as a measure of the fineness [27]. In this sense, f_2 is not appropriate to the concept of contrast that we aim at. As a matter of fact, it is shown later that f_2 does not correlate with the human measurement f_{con} .

With respect to contrast, we attempted to approximate the four factors step by step. The simplest expression for Factor i) is the range of gray-levels. Next, it can be easily seen that the variance σ^2 or standard deviation σ about the mean of the gray-levels probability distribution is more preferable as contrast. As is well known, σ or σ^2 can measure the dispersion in the distribution. In this sense, it reflects

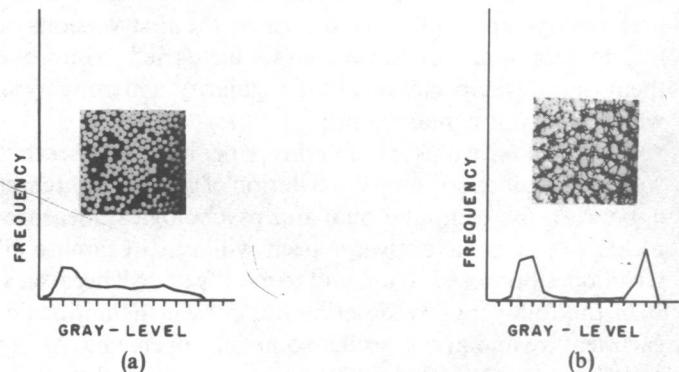


Fig. 5. Examples of gray-level distributions. (a) D111. (b) D67.

Factor ii) to some extent. However, the resultant value is undesirable for a distribution in which a single peak is highly biased to black or white as shown in Fig. 5(a).

For Factor ii), we want a measure of polarization. The kurtosis α_4 is well known for this purpose and can be defined as

$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$

where μ_4 is the fourth moment about the mean and σ^2 is the variance. This measure is normalized with respect to the range so that it can have the minimum value of one in case of twin peaks. This is just Factor ii) and is not influenced by Factor i).

Consequently, we combine σ and α_4 for the second-step measure of contrast as follows:

$$F_{con} = \sigma / (\alpha_4)^n$$

where n is a positive number. In our experiments, we varied n from 8, 4, 2, 1, 1/2, 1/4, to 1/8. Experimentally, $n = 1/4$ yielded the best r_s between f_{con} and F_{con} . This improvement could reduce the values of F_{con} for distributions with biased peaks while almost preserving those for polarized distributions as shown in Fig. 5(b).

Up to this stage, we approximated only Factors i) and ii); i.e., we described the contrast in the narrow sense based on the gray-level distribution. The next step is to incorporate Factors iii) and iv); i.e., contribution to contrast caused by pictorial structure. Fortunately or unfortunately, our improved measure defined above provided quite successful results for 16 texture patterns in Fig. 1. Although it is not perfect, the improvement of a measure by analyzing its principal deficiencies in the previous stages is not appro-

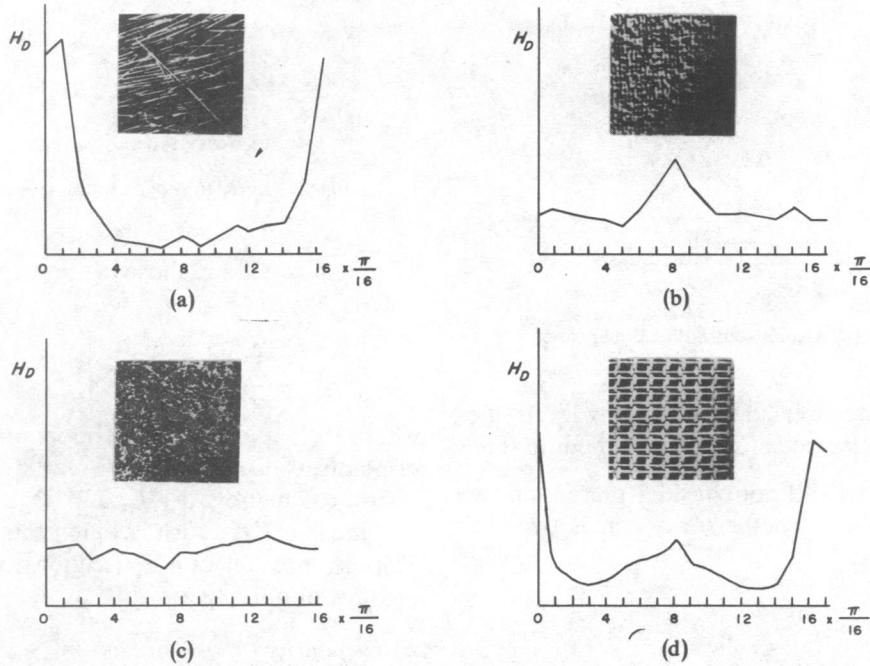


Fig. 6. Examples of local direction histogram H_D . (a) D15. (b) D84. (c) D9. (d) D20.

priate. Hence we leave approximations of Factors iii) and iv) to further studies.

C. Directionality

It is well known that the directionality in an original picture can be preserved in its Fourier power spectrum. We readily find the directionality in the histogram of the Fourier power along lines through the origin, as shown by Bajcsy [3]. However, other textural features in the Fourier domain have not been thoroughly investigated. Currently, it is said that the Fourier features do not behave as well as features in spatial domain, since an aperture effect may degrade the Fourier features [13], [14]. Suppression of aperture effects has been attempted [33], but this method loses the directionality in an original picture. Consequently, it seems uneconomical to compute the time-consuming Fourier transform only for directionality. Therefore, it is desirable to use a faster procedure in spatial domain.

Instead of an histogram of Fourier power, we use an histogram of local edge probabilities against their directional angle. It was shown that this histogram represents sufficiently global features of the input picture such as long lines and simple curves [34]. This method utilizes the fact that gradient is a vector, so it has both magnitude and direction. In the discrete case, the magnitude $|\Delta G|$ and the local edge direction θ are approximated as follows:

$$|\Delta G| = (|\Delta_H| + |\Delta_V|)/2$$

$$\theta = \tan^{-1} (\Delta_V/\Delta_H) + \frac{\pi}{2}$$

where Δ_H and Δ_V are the horizontal and vertical differences measured by following 3×3 operators, respectively,

$$\begin{array}{ccc} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{array} \quad \begin{array}{ccc} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{array}$$

The resultant θ is a real number ($0 \leq \theta < \pi$) measured counterclockwise so that the horizontal direction is zero.

The desired histogram H_D can be obtained by quantizing θ and counting the points with the magnitude $|\Delta G|$ over the threshold t ; i.e.,

$$H_D(k) = N_\theta(k) / \sum_{i=0}^{n-1} N_\theta(i), \quad k = 0, 1, \dots, n-1$$

where $N_\theta(k)$ is the number of points at which $(2k-1)\pi/2n \leq \theta < (2k+1)\pi/2n$ and $|\Delta G| \geq t$. Thresholding $|\Delta G|$ by t is aimed at preventing counting of unreliable directions which cannot be regarded as edge points. In our experiments, we used $n = 16$ and $t = 12$. The shape of each histogram H_D was not sensitive to the value of t .

Some examples of H_D are shown in Fig. 6. Fig. 6(a) tells that D15 is highly directional and slightly inclined to the left; its peak point is located at angle $\pi/16$. On the other hand, from Fig. 6(b), we can see that D84 has a relatively weak directionality, because its peak is not so sharp. The histogram of D9 shown in Fig. 6(c) is nearly flat; consequently, D9 has no directionality. It can be said that almost everybody accepts these descriptions concerning D15, D84, and D9. Thus, H_D can reflect the directionality very faithfully.

A way of measuring the directionality quantitatively from H_D is to compute the sharpness of the peaks. The approach which we adopted is to sum the second moments around each peak from valley to valley, if multiple peaks are determined to exist. This measure can be defined as follows:

$$F_{\text{dir}} = 1 - r \cdot n_p \cdot \sum_p^{n_p} \sum_{\phi \in w_p} (\phi - \phi_p)^2 \cdot H_D(\phi)$$

where

n_p number of peaks,

ϕ_p pth peak position of H_D ,

w_p range of pth peak between valleys,

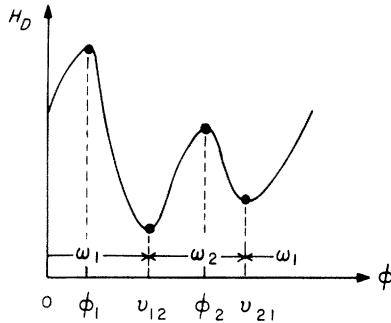


Fig. 7. Definition of symbols for bidirectional case.

r normalizing factor related to quantizing levels of ϕ ,
 ϕ quantized direction code (cyclically in modulo 180°).

In our experiments we did not consider more than two peaks; i.e., we determined whether or not n_p is two. The procedure is as follows:

$$n_p = 2,$$

if

$$\begin{aligned} H_D(v_{12})/H_D(\phi_2) &< 0.5, \\ H_D(v_{21})/H_D(\phi_2) &< 0.5, \end{aligned}$$

and

$$H_D(\phi_2)/H_D(\phi_1) > 0.2,$$

$$n_p = 1, \quad \text{otherwise.}$$

Where v_{12} and v_{21} are the positions of valleys from the first peak ϕ_1 to the second peak ϕ_2 , and vice versa, respectively (see Fig. 7).

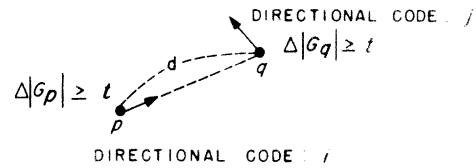
This procedure was designed specifically so that only D20 might be judged to be bidirectional and that $n_p = 1$ might be given to nondirectional patterns such as D9. For general use, the procedure should be much more sophisticated.

D. Line-likeness

It seems necessary to give a more detailed description of line-likeness. We mean by the word line-likeness an element of texture that is composed of lines. For this purpose, when the direction and the neighboring (edge's) direction for a given edge are nearly equal, we regard such a group of edge points as a line.

To be more specific, we construct a direction co-occurrence matrix whose element $P_{Dd}(i, j)$ is defined as the relative frequency with which two neighboring cells separated by a distance d along the edge direction occur on the image, one with the direction code i and the other with the direction code j . This is illustrated in Fig. 8. From this co-occurrence matrix, one can expect that several important textural features are derivable for measuring properties about element shape. As in H_D , it is preferable to ignore directions of trivial edges by using threshold t .

As a measure of line-likeness, we defined the following measure so that co-occurrences in the same direction are

Fig. 8. Entry (i, j) of direction co-occurrence matrix P_{Dd} .

weighted by +1 and those in the perpendicular direction by -1:

$$F_{\text{lin}} = \sum_i^n \sum_j^n P_{Dd}(i, j) \cos \left[(i - j) \frac{2\pi}{n} \right] / \sum_i^n \sum_j^n P_{Dd}(i, j)$$

where P_{Dd} is the $n \times n$ local direction co-occurrence matrix of points at distance.

For calculation of H_D and P_{Dd} we utilized common intermediate steps such as differentiating and thresholding. The distance function d_4 (see [35]) was set to four for all pictures in our experiments.

E. Regularity

It may seem easy to describe the regularity of repetitive patterns in mathematical form. In fact, we can easily synthesize a pattern from a mathematical expression. However, it is not trivial to analyze the regularity of naturally occurring textures. For highly regular patterns some techniques may indicate that they are regular. However, for natural textures which are difficult to describe mathematically, it is fairly difficult to measure a degree of irregularity without any information such as element size or shape.

Here we assume that if any feature of a texture varies over the whole image the image is irregular. Hence we take partitioned subimages and consider the variation of each feature in each subimage. We have defined four independent features up to here. We take the sum of the variation for each of these four features as a measure of regularity; i.e.,

$$F_{\text{reg}} = 1 - r(\sigma_{\text{crs}} + \sigma_{\text{con}} + \sigma_{\text{dir}} + \sigma_{\text{lin}})$$

where r is a normalizing factor and each σ_{xxx} means the standard deviation of F_{xxx} .

F. Roughness

We do not have any good ideas for describing the tactile sense of roughness. According to the results of our psychological experiments on vision, we emphasize the effects of coarseness and contrast, and approximate a measure of roughness by using these computational measures; i.e.,

$$F_{\text{rhs}} = F_{\text{crs}} + F_{\text{con}}$$

Our intention lies in examining to what an extent such a simple approximation corresponds to human visual perception.

V. COMPARISON BETWEEN PSYCHOLOGICAL AND COMPUTATIONAL MEASUREMENTS

The above mentioned features $\{F\}$ were computed for the 16 texture patterns in Fig. 1 which were also used for the psychological experiments. The results are shown in Fig. 9.

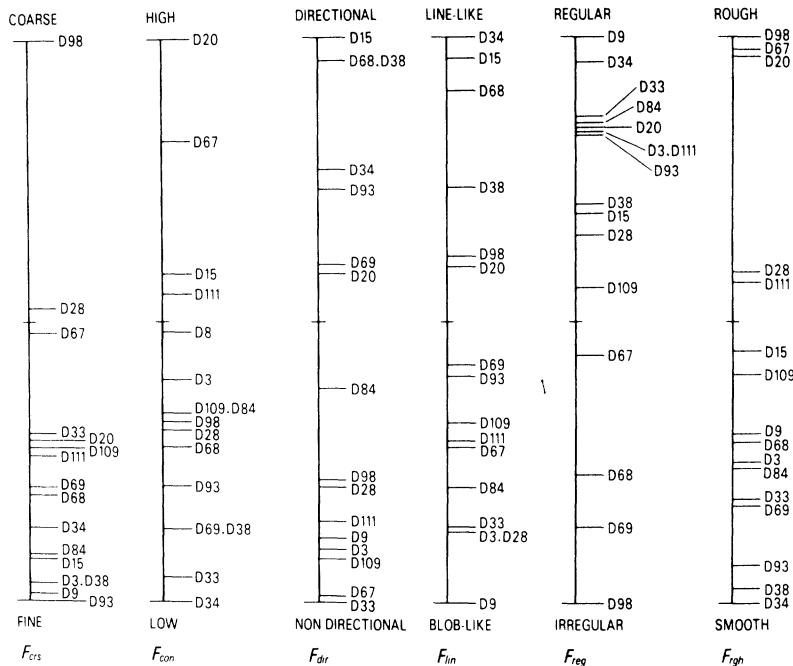
Fig. 9. Computational measurements of six basic textural features: $\{F\}$.

TABLE III
COEFFICIENTS OF RANK CORRELATION BETWEEN
COMPUTATIONAL MEASUREMENTS $\{F\}$

| r_s | F_{crys} | F_{reg} | F_{lin} | F_{dir} | F_{con} |
|------------|------------|-----------|-----------|-----------|-----------|
| F_{crys} | 0.650 | -0.424 | -0.063 | -0.426 | 0.124 |
| F_{con} | 0.756 | -0.024 | -0.233 | -0.250 | |
| F_{dir} | -0.341 | -0.158 | 0.782 | | |
| F_{lin} | -0.180 | -0.313 | | | |
| F_{reg} | -0.376 | | | | |

For comparison, the scales are adjusted to the psychological scales $\{f\}$ in Fig. 2.

The computational measurements were all performed on subimages with 128×128 pixels. The values of F_{con} , F_{dir} , and F_{lin} are the average of measurements over these four nonoverlapped regions in the original picture with 256×256 pixels. With respect to F_{crys} , S_{best} in the intermediate step can be computed effectively only for the inner 64×64 in a 128×128 subimage in case of using the largest 32×32 operator. In this situation, five more subimages can be taken in addition to the above four regions so that all inner effective areas cannot overlap each other. Thus the value of F_{crys} is the average of these total nine samples. F_{reg} is computed from standard deviations of above four features over four or nine samples. Finally, F_{rgh} is from the averaged values of F_{crys} and F_{con} .

The rank correlations were calculated among $\{F\}$ and between $\{f\}$ and $\{F\}$. These are listed in Tables III and IV. Considering these results, we analyze the causes of discrepancies in each feature.

A. Coarseness

The rank correlations with f_{crys} by varying the threshold t are listed in Table V. We can see a very stable behavior

TABLE IV
CORRESPONDENCES BETWEEN PSYCHOLOGICAL AND
COMPUTATIONAL MEASUREMENTS

| r_s |
|-----------------------|
| $f_{crys} : F_{crys}$ |
| $f_{con} : F_{con}$ |
| $f_{dir} : F_{dir}$ |
| $f_{lin} : F_{lin}$ |
| $f_{reg} : F_{reg}$ |
| $f_{rgh} : F_{rgh}$ |

TABLE V
VARIATIONS OF CORRESPONDENCE IN COARSENESS
BY CHANGING THRESHOLD t

| t | $f_{crys} : F_{crys}$ |
|------|-----------------------|
| 0.83 | 0.794 |
| 0.85 | 0.811 |
| 0.86 | 0.831 |
| 0.87 | 0.811 |
| 0.89 | 0.811 |
| 0.91 | 0.811 |
| 1.00 | 0.794 |

around $t = 0.9$. The results for $t = 0.86$ are shown in Fig. 9. The correspondence $f_{crys} : F_{crys} = 0.831$ is satisfactory. We can make the following observations.

a) The main discrepancies between f_{crys} and F_{crys} occur for D3, D20, D34, and D111. These patterns can be interpreted to consist of basic elements which are placed very close to each other and surrounded by lines. It seems that our human subjects perceived them in this manner. On the other hand, the computer must regard the lines in these net-like pictures as the texture elements according to our definition of F_{crys} . If

$p > 2d$, the gap between the repetitive elements ($p - d$) is larger than the element size d . However, if $p < 2d$, the gap width is regarded as the element. If p is close to $2d$, no problem is encountered. However, if d is much smaller than p , as in D3, the expected value $(2d + p)/4$ is close to $p/2$, whereas the width between the lines ($p - d$), which the human being seems to measure, becomes close to p . Furthermore, human beings can detect pentagons or hexagons from D111, which the computer cannot do. All these effects depend on the human ability of perceiving simple shapes. Unfortunately, it is quite difficult to embed such an ability into a computational procedure.

b) On the other hand, this method is powerful for so-called scramble textures in which the basic element is vague. The sizes of basic element in D98, D28, D33, D84, and D9 are not clear, but satisfactory results were obtained about the order of coarseness.

B. Contrast

We examined f_2 as well as F_{con} . As anticipated, f_2 is heavily affected by fineness and has no correlation with F_{con} ; in fact $F_{\text{con}}:f_2 = 0.044$, $F_{\text{crs}}:f_2 = -0.456$.

On the other hand, although our measure F_{con} was designed only for Factors i) and ii), it yielded quite a good score $F_{\text{con}}:F_{\text{con}} = 0.904$. It proves that the narrow sense of contrast is a strong dominant factor even for texture patterns with different structures. Texture is said to be essentially a tone-invariant and spatial property. In this sense, our current measure F_{con} , not containing Factors iii) and iv), is not a textural property. The contrast in our feature set is expected to reflect these two factors in the future.

C. Directionality

To compare the effectiveness of our method, we computed directionality F_{fd} using power histograms from Fourier power spectra. As results, we obtained $f_{\text{dir}}:F_{\text{dir}} = 0.823$ and $f_{\text{dir}}:F_{fd} = 0.776$. This reveals that our local direction histogram H_D is equivalent or better than the Fourier techniques. Directionality is well reflected in Fig. 6. Nevertheless, the correlation of 0.823 is not so excellent as expected. This discrepancy may be caused by the following two reasons.

a) Although the gradual change of direction in D69 shows high directionality perceptually this change is averaged in H_D , resulting in a weak peak. As an extreme case, let us consider the pattern which consists of half circles centered at the same point. Human beings may perceive it as directional to some extent, but we cannot detect the directionality from our histogram H_D . This is one of the limits of using only local operations.

b) The degree of perceptual coarseness or contrast can be measured relatively, i.e., by comparing two or more pictures, while the presence of directionality may be estimated from one. Suppose we are required to classify each texture individually into two groups: directional and nondirectional. We may readily assign [D68, D15, D93, D69, D38, D84, D20, D34] to the first group and the rest to the other group. From this point of view, both f_{dir} and F_{dir} can classify them exactly, but within each group, the ranks are highly

interchanged. It is presumably because the subjects were forced to select the most directional picture in each pair, even if the pictures were very similar.

D. Line-likeness

The correspondence $f_{\text{lin}}:F_{\text{lin}} = 0.713$ is worse than the previous three results. However, taking into consideration the fact that the human subjects had difficulty with line-likeness, it is a desirable result that D34 and D15 took the high ranks in F_{lin} . In fact, the correlation between F_{dir} and F_{lin} is smaller than that between f_{dir} and f_{lin} (i.e., $F_{\text{dir}}:F_{\text{lin}} = 0.782$, $f_{\text{dir}}:f_{\text{lin}} = 0.923$). However, we can observe the following deficiencies in the current F_{lin} :

a) We used the fixed distance four for all pictures. If by accident line crossings exist at this distance, our F_{lin} fails to detect the continuity of directions. This is the case in D3. To avoid this effect, we should use a distance which is adequate to each picture by using the result of F_{crs} ; otherwise we should follow the local direction step by step, but this method would be very time consuming.

b) In our definition of F_{lin} an edge was counted as a line, so that D67 and D98 had the undesirable larger values in F_{lin} than in f_{lin} . To avoid this effect we must use not an edge-detector but a line-detector. It is not easy to overcome these deficiencies. Concerning line-detection, several attempts have so far been performed [36], [37]. However, properly chosen sizes and thresholds are essential in these methods, and sometimes *a priori* knowledge is required. In our situation, we have no information about the presence of lines and their widths, if any. It is rather difficult to design a line-detector which is efficient on any unknown texture.

E. Regularity

We find considerable lack of correspondence between f_{reg} and F_{reg} . As was considered in Section III-B, f_{reg} was not a proper measurement for our specification. The authors' ordering f'_{reg} gave better correspondence with F_{reg} ; i.e., $f_{\text{reg}}:F_{\text{reg}} = 0.246$, $f'_{\text{reg}}:F_{\text{reg}} = 0.779$. Hence, it may be reasonable to regard f_{reg} as an unreliable result in $\{f\}$ while the approximation of regularity by considering variations of the features over subimages can work well.

Improved scores can be expected if the subimages have sizes suitable to their coarseness and to give the proper weight to each contributing feature. Moreover, in order to describe the regularity more precisely, it is indispensable to analyze the complexity of the texture elements and their contribution to the global regularity.

F. Roughness

The correlation between f_{rgh} and F_{rgh} is 0.652. It seems too crude to approximate the roughness simply by combining F_{crs} and F_{con} . We tried to combine F_{rgh} as $F_{\text{crs}} + F_{\text{con}} - F_{\text{dir}}$, but the result was even worse. The principal discrepancies can be observed as follows.

a) D67 is so high in coarseness and contrast that it is ranked as second on F_{rgh} . However, D67 was not perceived so rough by the human subjects.

b) On the other hand, D84 was perceived to be very rough, but it is relatively low in both F_{crs} and F_{con} .

Looking at D67 and D84, we would certainly perceive their basic elements as smooth and rough, respectively, even if their coarseness and contrast were to be changed to some extent. It seems that human beings emphasize the perceptual roughness of texture elements. In this case, one should try to analyze and describe texture elements precisely.

The attained results are shown in Table IV. For the first three features, our improvements achieved successful correspondences with psychological measurements. It follows that the remaining three could be sufficiently improved by due considerations of the causes of discrepancies. Here, what should be noted first is that coarseness plays an important role in all other features. Therefore, it is expected that a better performing coarseness detector will be developed and other textural features can be extracted by using its output. Secondly, development of a method that can describe texture elements faithfully is the remaining significant problem.

VI. TEST OF SIMILARITY MEASUREMENTS

We saw that our set of textural features succeeded in corresponding to human visual perception in each feature. Another interesting problem is to describe similarity of texture pattern by using a certain combination of these features. Here, the problem which we specified to the human subjects and the computer is to find the nearest pattern from 15 other ones for each picture in Fig. 1.

Nine human subjects (not including the authors) had to select the closest pattern for each picture, but they did not have to give reciprocal answers; e.g., if D93 was selected as the closest with respect to D69, any pattern may be selected as the closest to D93 (except of course D93 itself).

The nearest categories were computed for each of the nine samples in each category by using the first four features which are independently defined, i.e., F_{crs} , F_{con} , F_{dir} , F_{lin} . We examined two distance criteria for deciding which is nearest. The first one finds for each sample the category having the maximum likelihood except its own category. In other words, this decision is the case of the minimum Mahalanobis distance to the means of other categories. The other one is the popular Euclidean distance, where all four measurement scales are normalized so that all their means and variances can have the same value.

The results are shown in Table VI. The answers for some categories vary widely. It may be impossible to request only one answer for each category from both human beings and the computer. We can hardly see the correspondence between human and computer results except in the case of D15. It is not trivial to find a good measure for such a correspondence. There exists various answers even among the human beings. Assuming that all these answers are correct, we count the number of computer answers which coincide with any of the human ones. We denote this number by N_{nc} . These values for two criteria are also shown in Table VI. These results, it must be said, are poor. It follows that the simple combination of four features, i.e., F_{crs} , F_{con} , F_{dir} , and F_{lin} , cannot describe similarity well.

TABLE VI
NEAREST CATEGORIES JUDGED BY HUMANS AND COMPUTER

| | Human subjects | Computer | |
|-----------------|-------------------------------------|-----------------------------|----------------------------|
| | | Maximum likelihood | Euclidean distance |
| D3 | D34(5)*, D111(3) | D84(7) D28, D67 | D9(8), D84 |
| D9 | D93(3), D109(2) D33(2), D38, D84 | D3(8), D111 | D3(9) |
| D15 | D68(9) | D68(9)** | D68(9) |
| D20 | D84(6), D111 D34(2) | D93(9) | D67(7), D15(4) |
| D28 | D98(6), D33 D67, D111 | D109(5), D68(4) | D109(9) |
| D33 | D109(4), D38 D28(3), D84 | D98(9) | D109(6), D33(3) |
| D34 | D3(5), D20(2) D111(2) | D68(9) | D38(7), D68(2) |
| D38 | D93(8), D68 | D69(9) | D68(6), D93(3) |
| D67 | D28(5), D3 D84(2), D33 | D109(5), D20(2), D111(2) | D111(9) |
| D68 | D15(5), D38(4) | D98(7), D15(2) | D38(7), D34(2) |
| D69 | D93(6), D38(3) | D98(5), D38(2), D68(2) | D84(4), D38(3), D93(2) |
| D84 | D20(3), D38(3) D15(2), D28 | D68(4), D28 D3(3), D109 | D3(7), D69, D93 |
| D93 | D38(9) | D38(5), D84(2) D109(2) | D69(9) |
| D98 | D28(9) | D109(9) | D28(9) |
| D109 | D9(7), D33, D111 | D28(9) | D111(5), D3 D28(2), D84 |
| D111 | D3(6), D28 D34, D109 | D109(9) | D109(8), D3 |
| N_{nc} | | 28 (19.4%) | 59 (41.0%) |

* Numerals in bracket after category name mean the number of subjects or samples which answered to that category.

** Underlined parts coincide with one of the answers by human subjects.

TABLE VII
NUMBER OF SAMPLES COINCIDING WITH ONE OF HUMAN ANSWERS BY USING MINIMUM EUCLIDEAN DISTANCE UNDER SPECIFIED CONDITIONS

| | F_{crs} | F_{con} | F_{dir} | F_{lin} | Over 4 features | $F_{\text{crs}} \& F_{\text{dir}}$ |
|-----------------|------------------|------------------|------------------|------------------|-----------------|------------------------------------|
| N_{nc} | 49 (34.0%) | 21 (14.6%) | 51 (35.4%) | 32 (22.2%) | 46 (31.9%) | 51 (35.4%) |

It is worth analyzing the results by humans. We asked the human subjects the reasons of their decisions. Most of their explanations indicated that similarity measurement is highly dominated by a few strong cues in a restricted situation, while we used four features equivalently in the above experiment. Especially, the "coarseness" and "directionality" were said to be very significant cues. For example, the human subjects answered the similarity in directionality for the reason why they did not select not D68 but rather D93 for D69, although each of D68 and D69 is a picture of wood grain.

Predicting that the strongest cue dominates the total similarity, we computed the categories having the minimum Euclidean distance for each feature separately in all four features, and only in F_{crs} and F_{dir} . Only the values for these results are shown in Table VII. Although the usefulness of F_{crs} and F_{dir} is proven, we could not approximate all the similarity cases. It may be partly because we do not have a

sufficient number of features which describe the structure of textures. However, more complicated logic seems to be required for the similarity measurement.

VII. CONCLUSION

In this study, we have attempted to develop textural features which correspond to human visual perception. Six basic textural properties (i.e., coarseness, contrast, directionality, line-likeness, regularity, and roughness) have been measured by human subjects. Computational measures have been developed and improved in order that they may correspond to the psychological measurements.

With respect to coarseness, contrast, and directionality, we have attained very successful results. These three features are so significant in global descriptions of textures that they can be expected to be useful separately in cases where the texture differs only in one of them or in combinations for image classification and segmentation problems. Especially, coarseness is a highly essential factor in texture. In order to improve the other features, its results should be utilized.

With respect to line-likeness, regularity, and roughness, we have obtained considerable correspondences between computational and psychological measurements, but more effort is required to describe precisely the texture elements on which these three features depend. For this purpose, we should have many kinds of techniques which can detect differences in a variety of second-order statistics. The meanings of the technique should be clear, because it is nearly impossible to use all differences in general second-order statistics for unknown textures.

On the other hand, simple combinations of our features have not simulated the human similarity measurement well. It is supposedly not because the ability of each feature is insufficient but because more complex mechanisms exist in the human usage of multiple cues. At any rate further studies are necessary in the psychometrical field to investigate the mechanisms. Also, it is important to give well defined specifications to human subjects and analyze the responses for each specification. Indeed, the concepts of uniformity, complexity, homogeneity, etc., are very interesting. We believe that we should continue to develop new computational versions of basic textural features and to improve them.

ACKNOWLEDGMENT

The authors wish to express their thanks to Prof. M. Nagao of Kyoto University for the digitization of the pictures, Dr. K. Mori and his colleagues of the Toshiba Research and Development Center for displaying the pictures, and Dr. T. Kasvand of the National Research Council of Canada for making useful suggestions on the improvement of some computational definitions and his great help in preparing this paper while one of the authors was there as a Visiting Researcher.

REFERENCES

- [1] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, pp. 610-621, Nov. 1973.
- [2] M. M. Galloway, "Texture classification using gray-level run lengths," *Computer Graphics and Image Processing*, vol. 4, pp. 172-179, June 1975.
- [3] R. Bajcsy, "Computer description of textured surfaces," in *Proc. 3rd Int. Joint Conf. Artificial Intelligence*, Aug. 1973, pp. 572-579.
- [4] —, "Computer identification of textured visual scenes," A.I. Lab., Stanford Univ., Palo Alto, CA, Memo. AIM-180, Oct. 1972.
- [5] R. N. Sutton and E. L. Hall, "Texture measurement for automatic classification of pulmonary disease," *IEEE Trans. Comput.*, vol. C-21, pp. 667-676, July 1972.
- [6] O. R. Mitchell, C. R. Myer, and W. Boyne, "A max-min measure for image texture analysis," *IEEE Trans. Comput.*, vol. C-26, pp. 408-414, Apr. 1977.
- [7] A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. Comput.*, vol. C-20, pp. 562-569, May 1971.
- [8] A. Rosenfeld, M. Thurston, and Y. H. Lee, "Edge and curve detection: Further experiments," *IEEE Trans. Comput.*, vol. C-21, pp. 677-715, July 1972.
- [9] W. B. Thompson, "Textural boundary analysis," *IEEE Trans. Comput.*, vol. C-26, pp. 272-276, Mar. 1977.
- [10] L. S. Davis, A. Rosenfeld, and J. S. Weszka, "Region extraction by averaging and thresholding," *IEEE Trans., Syst., Man, Cybern.*, vol. SMC-5, pp. 383-388, May 1975.
- [11] F. Tomita and S. Tsuji, "Extraction of multiple regions by smoothing in selected neighborhoods," *IEEE Trans., Syst., Man, Cybern.*, vol. SMC-7, pp. 107-109, Feb. 1977.
- [12] S. G. Carlton and O. R. Mitchell, "Image segmentation using texture and gray level," in *Proc. IEEE Conf. on Pattern Recognition and Image Processing*, June 1977, pp. 387-391.
- [13] J. S. Weszka, C. R. 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.
- [14] R. W. Connors and C. A. Harlow, "A theoretical comparison of four texture algorithms," Image Analysis Lab., Univ., Missouri-Columbia, Columbia, Tech. Rep. IAL-4-76, Aug. 1976.
- [15] B. Julesz, "Visual pattern discrimination," *IRE Trans., Info. Theory*, vol. IT-8, pp. 84092, Feb. 1962.
- [16] —, "Texture and visual perception," *Scientific American*, vol. 212, pp. 38-54, Feb. 1965.
- [17] —, "Experiments in the visual perception of texture," *ibid*, vol. 232, pp. 34-43, Apr. 1975.
- [18] B. Julesz, E. N. Gilbert, L. A. Shepp, and H. L. Frisch, "Inability of humans to discriminate between visual textures that agree in second-order statistics—revisited," *Perception*, vol. 2, pp. 391-405, 1973.
- [19] A. Rosenfeld, "Visual texture analysis: An overview," Compt. Sci. Center, Univ. Maryland, College Park, Tech. Rep. TR-406, Aug. 1975.
- [20] S. W. Zucker, "Toward a model of texture," *Computer Graphics and Image Processing*, vol. 5, pp. 190-202, June 1976.
- [21] A. Rosenfeld and B. S. Lipkin, "Texture synthesis," in *Picture Processing and Psychopictorics*. New York: Academic, 1970.
- [22] B. H. McCormick and S. N. Jayaramamurthy, "Time series model for texture synthesis," *Int. J. Comp. Inf. Sci.*, vol. 3, pp. 329-343.
- [23] J. T. Tou, D. B. Kao, and Y. S. Chang, "Pictorial texture analysis and synthesis," in *Proc. 3rd Int. Joint Conf. Pattern Recognition*, Nov. 1976, pp. 590-590.
- [24] F. Tomita, M. Yachida, and S. Tsuji, "Detection of homogeneous regions by structural analysis," in *Proc. 3rd Int. Joint Conf. Artificial Intelligence*, Aug. 1973, pp. 564-571.
- [25] S. Tsuji and F. Tomita, "A structural analyzer for a class of textures," *Computer Graphics and Image Processing*, vol. 2, pp. 216-231, Dec. 1973.
- [26] P. Brodatz, *Textures*. New York: Dover 1966.
- [27] A. Rosenfeld and E. B. Troy, "Visual texture analysis," Compt. Sci. Center, Univ. Maryland, College Park, Tech. Rep. TR-116, June 1970.
- [28] K. C. Hayes, Jr., A. N. Shah, and A. Rosenfeld, "Texture coarseness: Further experiments," *IEEE Trans. Syst. Man., Cybern.*, vol. SMC-4, pp. 467-472, Sept. 1974.
- [29] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. New York: Academic, 1976, Chap. 10.
- [30] C. H. Coombs, R. M. Dawes, and A. Tversky, *Mathematical Psychology—An Elementary Introduction*. Englewood Cliffs, N.J.: Prentice-Hall, 1970.
- [31] L. L. Thurstone, "A law of comparative judgement," *Psychol. Rev.*, vol. 34, pp. 273-286, 1927.

- [32] H. Tamura, S. Mori, and T. Yamawaki, "Psychological and computational measurements of basic textural features and their comparison," in *Proc. 3rd Int. Joint Conf. Pattern Recognition*, Nov. 1976, pp. 273-277.
- [33] C. R. Dyer and A. Rosenfeld, "Fourier texture features: Suppression of aperture effects," *IEEE Trans. Syst., Man, Cybern.*, vol. 6, pp. 703-705, Oct. 1976.
- [34] S. Mori, Y. Monden, and T. Mori, "Edge representation in gradient space," *Computer Graphics and Image Processing*, vol. 2, pp. 321-325, Dec. 1973.
- [35] A. Rosenfeld and J. L. Pfaltz, "Distance function on digital pictures," *Pattern Recognition*, vol. 1, pp. 33-61, July 1968.

Quantitative Evaluation of Computer Regenerated Images and Their Use in Storage-Restricted Environments

RAOUF F. H. FARAG, MEMBER, IEEE, AND Y. T. CHIEN, SENIOR MEMBER, IEEE

Abstract—For a given set of images (or pattern samples), a probabilistic grammar may be developed to describe and recognize these images. This grammar, once obtained, may serve as an abstract representation for the images in question. Conversely, images may be regenerated from their grammatical representations. Images regenerated in this manner, however, may contain samples that are not found in the set of original images from which the grammatical representation was developed. A method for evaluating the quality of regenerated images with respect to the set of original images is discussed. Let I be the set of original images and O be the set of images regenerated from the abstract representation of I . Set O consists of two mutually exclusive subsets, A and B . A is the set of images found in I , and B is the set of images not found in I . It is shown that the clustering behavior of set B determines the quality of the regenerated images in the mean-square sense. By defining a scatter matrix for B , several factors that influence the image quality with respect to I are derived. Experimental results on the abstract representation and regeneration of a subset of the Munson data are demonstrated. An interesting application of this regeneration model to pattern recognition in storage-restricted environments is also presented.

I. INTRODUCTION

IN DEALING with a large data base of pictorial information, it may be necessary to consider the generative capacity of probabilistic grammars as a way of storing and manipulating the massive data. This paper discusses this

generative property in the context of regenerating the images from their abstract grammatical representation. In particular, the question of how the quality of regenerated images can be evaluated quantitatively, consistent with the subjective, visual evaluation, will be addressed. The immediate result of regenerating "high quality" images from their abstract grammatical representation is an alternative to the storage of large data bases needed in pattern recognition experiments.

The handling of images has been of considerable interest to researchers in the field of optics and the field of image processing. From the optical point of view the interest is found in noise removal, image enhancement, image restoration, and reconstruction [1]-[4]. In image processing considerable attention is given to image coding for transmission, noise filtering, use of transforms (Fourier, Hadamard, and Walsh) for bandwidth reduction and filtering, etc. [5]-[12]. Levi [1] gives a review of the quantitative image assessment measures. He suggests the use of mean-square error, correlation, signal to noise ratio, and a measurement based on the concept of channel capacity in information theory. In both image processing and optics the problem is a *single* image which has been processed through a distorting system. The interest in these fields is in how to retrieve the input signal in the "best" way possible. In a purely optical system, for instance, the best image would be the least distorted, while in image processing the best image would be the image requiring the least bandwidth while still maintaining a certain "quality" in the image.

The main concern of this paper is different from the above two areas in the sense that we are not handling a single image at a time but rather a collection of images at the input and a

Manuscript received May 31, 1977; revised February 10, 1978. This work was supported in part by the National Science Foundation under Grant GJ36434. An excerpt of this paper was presented at the IEEE International Conference on Cybernetics and Society, Washington, DC, November 1-3, 1976.

R. F. H. Farag is with the Department of Mathematics and Computer Science, St. John's University, Jamaica, NY 11439.

Y. T. Chien is with the Department of Electrical Engineering and Computer Science, University of Connecticut, Storrs, CT 06268.