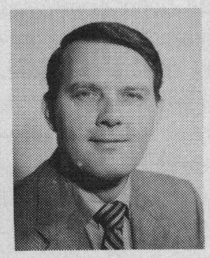
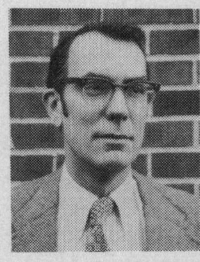
[8] W. W. Plummer, "Asynchronous arbiters," *IEEE Trans. Comput.*, vol. C-21, pp. 37–42, Jan. 1972.
[9] S. H. Unger, "Asynchronous sequential switching circuits with unrestricted input changes," *IEEE Trans. Comput.*, vol. C-20, pp. 1437–1444, Dec. 1971.
[10] D. B. Armstrong, A. D. Friedman, and P. R. Menon, "Design of asynchronous circuits assuming unbounded gate delays," *IEEE Trans. Comput.*, vol. C-18, pp. 1110–1210, Dec. 1969.
[11] "Decisions, Decisions," *Sci. Amer.*, p. 43, Apr. 1973.
[12] T. J. Chaney and C. E. Molnar, "Anomalous behavior of synchronizer and arbiter circuits," *IEEE Trans. Comput.* (Corresp.), vol. C-22, pp. 421–422, Apr. 1973.
[13] M. Hurtado, "Structure and performance of asymptomically bistable dynamic systems," D.Sc. dissertation, Washington Univ., St. Louis, Mo., 1974, to be published.



George R. Couranz (S'66–M'70) received the B.S.E.E. degree from Washington University, St. Louis, Mo., in 1960. From 1960 to 1966 he served as an officer in the United States Air Force, during which time he received the M.S.E.E. degree from the Air Force Institute of Technology, Wright-Patterson Air Force Base, Dayton, Ohio. In 1966 he joined the Computer Systems Laboratory of Washington University where he worked on parallel asynchronous processing techniques. He received the D.Sc. from Washington University in 1970.

He is presently with Raytheon Company, Sudbury, Mass., involved in systems design and associative processing.



Donald F. Wann (M'54) was born in St. Louis, Mo., on March 28, 1932. He received the B.E. degree from Yale University, New Haven, Conn., in 1953, and the M.S. and D.Sc. degrees in electrical engineering from Washington University, St. Louis, Mo., in 1957 and 1961, respectively.

He joined the School of Engineering and Applied Science at Washington University as an Assistant Professor in 1961. He served as Associate Chairman of the Department of Electrical Engineering from 1965 to 1970, where he is currently Professor. He is the Associate Director of the Computer Systems Laboratory, and his research work is in the area of image processing and special-purpose computer design for solving problems in biomedicine.

Dr. Wann is a member of Eta Kappa Nu, Sigma Xi, and Tau Beta Pi, the American Association for the Advancement of Science, the Association for Computing Machinery, and the Pattern Recognition Society.

On the Quantitative Evaluation of Edge Detection Schemes and Their Comparison with Human Performance

JERRY R. FRAM AND EDWARD S. DEUTSCH, MEMBER, IEEE

***Abstract*—A technique for the quantitative evaluation of edge detection schemes is presented. It is used to assess the performance of three such schemes using a specially-generated set of images containing noise. The ability of human subjects to distinguish the edges in the presence of noise is also measured and compared with that of the edge detection schemes. The edge detection schemes are used on a high-resolution satellite photograph with varying degrees of noise added in order to relate the quantitative comparison to real-life imagery.**

***Index Terms*—Edge detection schemes, image processing, quantitative evaluation.**

Manuscript received March 15, 1973; revised May 15, 1974. This work was supported in part by the National Aeronautics and Space Administration under Grant NGR21-002-351.
J. R. Fram is with Finn and Fram, Inc., Sun Valley, Calif.
E. S. Deutsch is with P. S. Ross & Partners, Toronto, Ont., Canada.

I. INTRODUCTION

ONE of the fundamental problems in the area of image processing is the location of boundaries of objects within an image. While a considerable number of edge detection schemes have been devised towards this end no comprehensive attempt has been made to compare the various schemes available. Such a comparison is necessary in order to establish the performance limitations of each scheme under the various conditions encountered. While the work reported here is only a first step in this important direction, it forms a groundwork for the long needed overall evaluation.

Characteristics of edge detection schemes, which should be investigated for comparative purposes include the following; edge orientation biases, edge detection in the presence of noise, range in scale of edge detectability,

ability to detect blurred edges, ability to detect curved edges, ability to extract an edge in the presence of other edges, and computer speed and storage requirements. The performance of edge detectors in the presence of noise is investigated here; this characteristic was selected because it is important in a wide range of applications and can be isolated relatively easily from other characteristics. It is recognized that in order to thoroughly compare edge detection schemes all the other enumerated characteristics must be investigated.

The approach followed here has one principle feature—the comparative results must be quantitative. We thus attempt to find reproducible numbers corresponding to parameters which accurately reflect the edge detectors' performance in the presence of noise. This is most readily accomplished with artificial pictures. However, the true test of an edge detector is in relation to its performance with real pictures; the numbers derived from artificial images must therefore be shown to have relevance there.

The edge detection schemes due to Hueckel [1]–[3], Macleod [4], [5] and Rosenfeld [6]–[8] are investigated here, all of which are briefly summarized. Human responses to essentially the same visual information are evaluated thereafter, and compared with the computer performances. Finally, in order to investigate the relevance of the parameters chosen to compare the various edge detection schemes, they are tried on the real-life imagery of a satellite photograph to which varying degrees of noise were added.

The only other attempt to quantify edge detection performance, known to the investigators, was that of Herskovits [9]. These quantification methods are easily interpretable but are hard to apply to most edge detection schemes. In contrast, while the exact meaning of the parameters to be calculated here is admittedly open to discussion, they can be estimated easily for a wide range of edge detection schemes. It will be seen that our parameters, in some sense, reflect the quality of edge detection, but a comparison with the more straightforward approach of Herskovits is useful. Such a comparison is discussed in another work by the authors [10].

II. THE STRUCTURE OF THE TEST IMAGES

A set of images was sought for which both the noise and the edge signal could be characterized numerically and in which these were to be the only variables allowed to change. The edge signal of a picture is parameterized by establishing two regions with different mean grey levels so that the regions could be said to be separated by an edge. The edge signal's strength could thus be parameterized as the difference in mean grey level of the two regions. The noise in these regions could be given by the variance in their grey level. It was decided to keep this variance largely independent of position so that the noise could be said to characterize each image as a whole. Approximately ten pictures were generated for each such contrast.

These test images consisted of 36×36 matrices each of which was divided into three zones as shown in Fig. 1. Picture points in zones 1 and 2 were assigned grey levels which were selected randomly from Gaussian distributions of means g_1 and g_2, respectively, and $\sigma = 24$. These distributions were truncated at 0 and 63, respectively, the minimum and maximum grey levels used. Elements in zone 3 were assigned values in essentially the same manner, however, the mean of the grey level distribution for each column was obtained by interpolating between zones 1 and 2.

Specifically, the truncation points of the Gaussian distributions were set at -0.5 and 63.5; this resulted in reducing the variance of the distributions pushing their means towards 32. Table I summarizes the structure of the test images and gives the corrections made due to the effects of truncation. Parameters g_1 and g_2 and *nominal noise* refer to the averages and standard deviations of the Gaussian distributions, and *nominal contrast* is just the difference between g_1 and g_2. *Actual contrast* and *actual noise* are the noise and contrast values, respectively, after the truncation corrections were made. Typical examples of images are shown in Fig. 2(a).

III. ON THE QUANTIFICATION OF EDGE DETECTOR PERFORMANCE

A. A Standard Form of Edge Detector Output

The output of edge detectors may come in a variety of forms, each with its individual meaning. As a result, the parameterization of edge detector performance must be done in different ways, each tailored to the individual program. This inherent feature is a possible source of bias whose effect cannot be readily calculated. In an effort to confine this possible bias to a particular stage of the quantification of edge detector performance and, hopefully, also to limit it, the quantification of edge detector performance was divided into two steps:

1) The edge detector output was put into a standard form by a procedure which took into account its various characteristics.

2) From the standard form, two parameters reflective of the performance of the edge detector were calculated. This calculation was done in the identical manner for all edge detectors.

The standard form chosen was a binary plane in which a 1 denoted that the corresponding point was considered an edge point and a 0 meant otherwise. The size of this plane was the output domain of the edge detector. The three edge detectors evaluated here all produced quantities which could be considered edge weights for every point of the output domain. These quantities were rounded off to the nearest integer for convenience of display and storage. They were then thresholded to produce approximately the number of points, n_{fill}, expected for a well found edge. This number was chosen for each edge detector on the basis of an inspection of a small sample of outputs from the test images so as to more or less optimize the results.

Given n_{fill}, the method of deciding the above threshold for each picture was as follows. For each edge detector output, one can define a monotonically decreasing function $n(t)$ equal to the number of points whose edge weights are greater than or equal to t. The threshold used then is defined by

$$t_{\text{thresh}} = \text{lub}\,\{t \mid n(t) > n_{\text{fill}}\}.$$

The number of points of edge weight near t_{thresh} was always much less than n_{fill}.[1] For this reason, it was thought that rounding off the edge weights did not greatly influence the results.

One might object that the above procedure introduces prior knowledge of the edge into the results via the number n_{fill}. While this is clearly the case, it is hard to see how this procedure could favor one edge detector over another. The informational content of the edge detectors has been altered by transforming them into the standard form, but it is believed that this has been done in a relatively unbiased manner.

B. The Edge Detection Performance Parameters and a Model to Conceptualize Them

For every picture of the sample set, two parameters are calculated from the standardized form of output discussed in the previous section. These may be conceptualized in terms of the following model. It is assumed that each 1 within the binary plane results from either one or possibly both of the two sources, noise or signal. It is further assumed that the 1's derived from the noise are randomly distributed about the whole picture with constant probability, whereas the 1's resulting from signal are restricted to a small subset of points within the plane, which contains the edge. Thus had the positions of the 1's generated by signal and the 1's generated by noise been known, one could form the standard output binary plane by simply ORing these two sets of 1's together.

The first parameter P_1 may be defined on the basis of the above model as the maximum likelihood estimate (MLE) of the ratio of the total number of signal 1's divided by the sum of the number of noise 1's plus the number of signal 1's. The number of noise 1's is normalized to correspond to a standard number of columns in the binary plane making this ratio independent of the size of the output domain of the edge detector. The second parameter, P_2, may be visualized as follows. A row of the edge region is defined as "covered" if it contains at least one 1.

[1] The edge detector described in Section IV-C below produced upon thinning some trivial exceptions to this statement in which $n(0) \gg n_{\text{fill}}$ and $n(1) < n_{\text{fill}}$. However, in these cases, it was manifestly clear that the edge was well found and the resulting parameters reflected this.

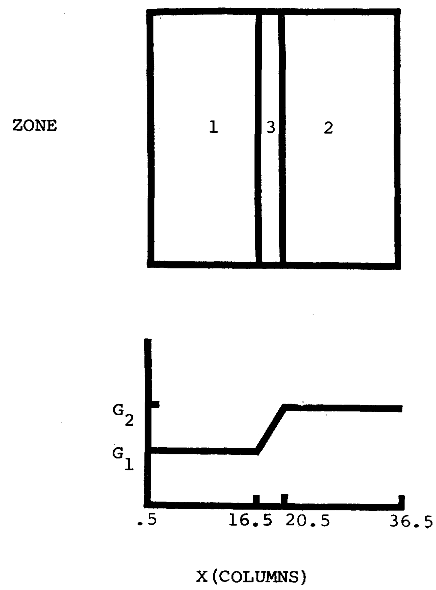


Fig. 1. The layout of a test image. The square indicates the positioning of the three zones of the test image, and the graph gives their mean grey levels. The X positions given are half integral because they indicate boundaries between columns. The mean grey level of each column of zone 3 is obtained by interpolation between zones 1 and 2.

TABLE I
PARAMETERS OF THE SET OF TEST IMAGES

g_1	g_2	NOMINAL CONTRAST	NOMINAL NOISE	ACTUAL CONTRAST	ACTUAL NOISE	NUMBER OF PICTURES
30	33	3	24	1.4	16.4	10
29	35	6	24	2.8	16.3	10
27	36	9	24	4.2	16.3	10
26	38	12	24	5.6	16.2	10
24	39	15	24	6.9	16.2	10
23	41	18	24	8.3	16.1	10
21	42	21	24	9.6	16.0	10
20	44	24	24	11.0	15.9	10
18	45	27	24	12.3	15.8	10
17	47	20	24	13.6	15.7	7

(Note that the edges of the sample are all vertical.) For this purpose all rows of the edge which are "covered" by noise 1's regardless of whether they are also covered by signal 1's are disregarded; P_2 is then set equal to the fraction of remaining edge rows "covered" by signal 1's. It thus provides a measure of the distribution of the signal over the length of the edge. An MLE of this fraction is made.

The explicit definitions of P_1 and P_2 now follow. Let the standard binary plane for edge detector output contain w_1 columns and w_2 rows. Let the "edge region" contain w_1^e columns and w_2^e rows. Here $w_2^e = w_2$, so that the superscript can be dropped from this quantity. Let the number of 1's in the edge region be n^e and the number of 1's outside the edge region be n^0. Note that while the n^e internal 1's are derivable from either signal only, noise only, or both signal and noise, the n^0 external 1's can only come from the noise in accordance with the model above. Parameter P_1 is then defined by

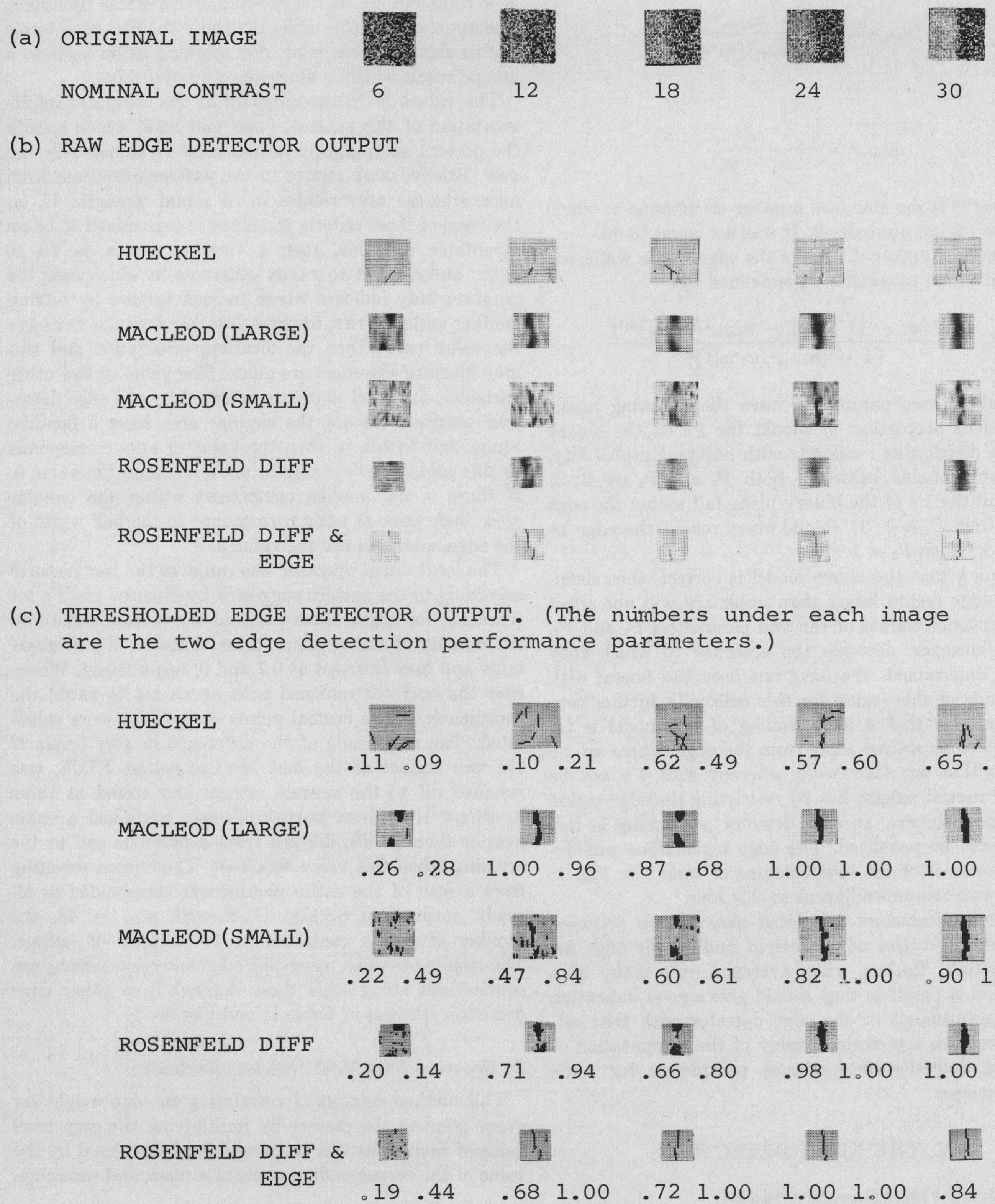


Fig. 2. A pictorial display of the steps taken in computing the edge detection performance parameters.

$$P_1 = \frac{n_{\text{sig}}^e}{n_{\text{sig}}^e + (n_{\text{noise}}^e + n^0)(w_1^{\text{stan}}/w_1)}$$

where

$$n_{\text{sig}}^e = \frac{n^e - n_{\text{noise}}^e}{1 - n_{\text{noise}}^e/w_1^e w_2}$$

and

$$n_{\text{noise}}^e = n^0 \frac{w_1^e}{w_1 - w_1^e}.$$

Where w_1^{stan} is the standard number of columns to which the noise 1's are normalized. It was set equal to 30.

If n^r is the number of rows of the edge region which are "covered" then parameter P_2 is defined by

$$P_2 = \frac{n^r/w_2 - \{1 - [1 - n_{\text{noise}}^e/w_1^e w_2]^{w_1^e}\}}{[1 - (n_{\text{noise}}^e/w_1^e w_2)]^{w_1^e}}.$$

The above two parameters have the following model independent properties: 1) should the 1's on the binary plane be distributed randomly with constant probability, the most probable values of both P_1 and P_2 are 0; 2) should all the 1's of the binary plane fall within the edge region, then $P_1 = 1$; 3) should every row of the edge be "covered," then $P_2 = 1$.

Assuming that the above model is correct, then defining the edge region larger than necessary will not affect the expectation values of the two parameters P_1 and P_2. It will, however, decrease the accuracy to which they may be determined. We have not been too liberal with the bounds of this region for this reason. A further consideration was that a likely failing of the model is for signal 1's to be shifted away from the edge. There are no means within our framework whereby such 1's can be assigned partial weight, but by restricting the edge region to a reasonable size, an edge detector performing in this manner can be penalized. The edge region thus was defined to consist of the four columns of zone 3 in Fig. 1, plus the two columns adjacent to this zone.

The two parameters calculated may not be expected to reflect the degree of success in finding the edge for every picture. Rather, when averaged over many of a given class of pictures, they should give a good indication of the performance of the edge detector with that set. Fig. 2 provides a pictorial display of the computation of the edge detection performance parameters for some typical images.

IV. THE EDGE DETECTORS

A. *The Local Visual Operator (Hueckel)*

This operator uses as input the grey level function derived from a circular area within the image. This input is fitted to that member within a set of ideal edge lines, whose Gaussian error of approximation to the input is minimal. The minimum Gaussian error and the best edge-line are determined by expanding the input in terms of a complete set of Fourier-like orthonormal functions. The approximation is made that only the first nine terms of this expansion are kept; this approximation leads to a unique result which is determined analytically.

The values of two parameters in the computer implementation of this scheme, CONF and DIFF, which specify the pattern acceptance criterion must be supplied by the user. Briefly, CONF relates to the pattern's freedom from noise whereas DIFF relates to its signal strength. If, on the basis of these criteria the input is determined to be an acceptable edge-line, then a variable, SUCCS, is set to TRUE. SUCCS is set to FALSE otherwise in which case the program may indicate where to look further by setting another variable, TRY, to TRUE. If either SUCCS or TRY have the value TRUE, then the location, orientation and two grey levels of the edge[2] are given. The value of two other variables, LINS and BLUR are returned by the edge detection program. Should the circular area have a line-like component within it, then the value of LINS corresponds to this component's strength. BLUR has then the value 0. If there is no line-like component within the circular area, then value of BLUR corresponds to the half width of the edge, and LINS has the value 0.

The local visual operator was run over the test pictures according to the pattern suggested by Hueckel in [3], for searching for edges. Each point of the picture came into the domain of the operator three times on the average. CONF and DIFF were set at 0.7 and 0, respectively. Whenever the operator returned with SUCCS set to TRUE, the coordinates of the central points of the edge were calculated. The magnitude of the difference in grey levels of the two regions of the best ideal edge-line, EDJS, was rounded off to the nearest integer and stored at these locations. If a given location already contained a value greater than EDJS, derived from a previous call to the operator, then this value was kept. The output resulting from a scan of the entire picture was thresholded as already outlined in Section III-A with $n_{\text{fill}} = 48$, the number of points contained in $1\frac{1}{2}$ columns of output. The results obtained using this edge detection scheme are summarized along with those derived from other edge detection schemes in Table II and Fig. 3.

B. *Gaussian Edge Mask Detector (Macleod)*

This method consists of calculating an edge weight for every point of the picture by multiplying the grey level value of each point in a surrounding neighborhood by the value of the corresponding point of a mask and summing.

[2] The operator has three modes; edge-lines, edges, and lines. Only the second mode is discussed here.

TABLE II
FIGURES OF PERFORMANCE FOR THE DETECTOR SCHEMES

EDGE DETECTOR	PARAMETER NUMBER	NOMINAL CONTRAST 3	6	9	12	15	18	21	24	27	30
HUECKEL	1	.00±.03	.13±.07	.03±.06	.21±.06	.23±.11	.26±.09	.35±.09	.66±.07	.58±.07	.67±.07
	2	-.03±.04	.10±.03	.05±.09	.25±.07	.20±.10	.27±.09	.34±.07	.63±.06	.62±.08	.68±.04
MACLEOD (LARGE)	1	.06±.99	.37±.40	.54±.08	.80±.06	.88±.03	.90±.02	.95±.02	.99±.00	.96±.02	.97±.02
	2	.07±.18	.27±.18	.50±.10	.73±.07	.81±.03	.82±.04	.88±.03	.96±.03	.91±.04	.96±.02
MACLEOD (SMALL)	1	.01±.04	.16±.04	.19±.03	.37±.03	.52±.04	.58±.04	.73±.04	.86±.03	.87±.03	.92±.01
	2	-.20±.11	.26±.12	.31±.07	.54±.06	.74±.06	.82±.09	.90±.03	.98±.02	.98±.01	.99±.01
ROSENFELD DIFF ONLY	1	.06±.06	.23±.05	.39±.06	.63±.05	.74±.03	.76±.03	.80±.04	.97±.01	.92±.02	.97±.02
	2	.07±.18	.27±.17	.32±.45	.74±.06	.80±.04	.87±.03	.86±.04	.96±.02	.95±.03	.99±.01
ROSENFELD DIFF & EDGE	1	.03±.06	.16±.04	.24±.06	.47±.04	.55±.05	.66±.04	.68±.04	.74±.06	.82±.05	.81±.03
	2	.07±.13	.36±.07	.47±.11	.83±.05	.90±.06	.95±.03	.98±.00	.98±.02	.99±.01	1.00±.00

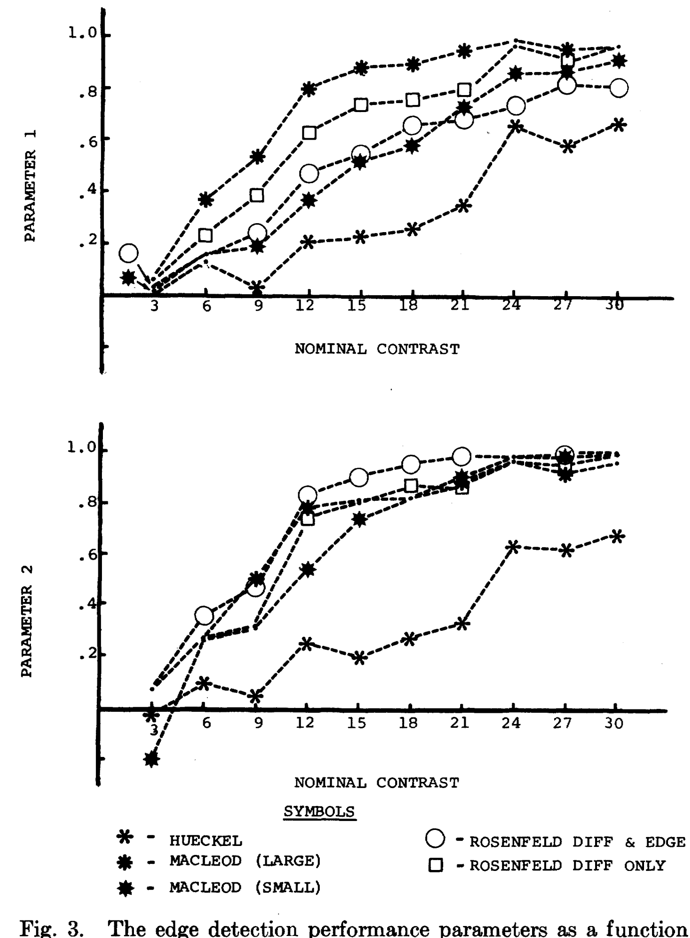


Fig. 3. The edge detection performance parameters as a function of nominal contrast. (The dotted lines are solely for the purpose of guiding the eye.)

The mask is given by

$$w(x,y) = \exp\left[-\left(\frac{y}{t}\right)^2\right]\left\{\exp\left[-\left(\frac{x-p}{p}\right)^2\right] - \exp\left[-\left(\frac{x+p}{p}\right)^2\right]\right\}$$

where x is the component of the distance of the neighboring point from the original point in a direction perpendicular to the direction of the edge in question, and y is the parallel component. Values of p and t are supplied by the user.

This scheme has the desirable feature that, if the mask is centered on an edge of the scale and orientation of interest, then points most likely to indicate such an edge will be weighted most heavily. Points on the edge itself and points likely to lie on neighboring edges will contribute little to the resulting edge weight.

Two masks of the form $w(x,y)$ were tested. The value of x for each point of the two masks was given by the horizontal displacement of the point from the center of the mask, and the value of y was given by its vertical displacement. Thus, in both cases, vertical edges would receive maximum weight. The points of the first mask were assigned values given by $w(x,y)$ with $p = t = 4$ inside a 7×7 square and 0 outside this square. The points of the second mask were similarly assigned values with $p = 4.7$ and $t = 4$ inside a 13×13 square and 0 outside. The absolute values of the corresponding sums were then thresholded by the procedure outlined in Section III-A with n_{fill} set equal to the number of points in four columns of its output domain. The results obtained using both masks are summarized in Table II under Macleod "small" and Macleod "large," respectively.

C. Local Difference Calculations (Rosenfeld)

The implementation of the scheme tested here, that of Hayes [8], consisted of two programs, run sequentially, called DIFF[3] and EDGE. DIFF, calculated edge weights for each point of the picture and EDGE suppressed the nonmaxima of these weights thus essentially thinning the edges.

Briefly, DIFF first calculated the average grey level of

[3] Not to be confused with the *parameter* of the same name used by Hueckel.

a square neighborhood of each point; the square sizes available are $2^k \times 2^k$ where k can be any integer from 1–6. The program then takes the absolute value of the difference of the averages of each adjacent nonoverlapping pair of these neighborhoods whose relative orientation is perpendicular to the edge direction of interest, and assigns this value to a point midway between these neighborhoods.

The edge weights for four orientations may be computed: horizontal, vertical, and the two diagonals. Every point of the picture is then assigned an edge weight with the exception of a frame about the edges. The program, EDGE, suppresses every such edge weight for which there exists a larger edge weight within a distance of 2^{k-1} perpendicular to the edge direction in question.

In testing this scheme, 8×8 square neighborhoods, ($k = 3$), were used, and only the vertical orientation was investigated. The output of DIFF was evaluated using a threshold selected by setting n_{fill} equal to the number of points in four columns of output (see Section III-A). The output of EDGE was also investigated setting n_{fill} to the number of points in 4/3 columns. The results are summarized in Table II and Fig. 2.

V. COMPARATIVE ASSESSMENT OF EDGE DETECTION SCHEMES

A. Computation Considerations

The comparison of the speed of the three schemes tested may not accurately be done without reference to specific hardware and software implementation. However, a rough indication is desirable; the method suggested by Hueckel seems appropriate.

Definition: Cover a picture dense enough with operator applications so that the instance of the pattern may be recognized anywhere in the picture. Divide the total number of necessary operations by the number of points of the picture. The result has the dimension "arithmetic operations per picture point" and will be called the "expense" or the "speed" of the operator [3].

We shall use the term "expense" as this seems far more appropriate to the quantity described. Hueckel provides further clarification.

The definition is not precise. In particular, care should be taken as to an honest choice of application density. Each of the following instructions shall count as one arithmetic operation: +, −, *, /, "add one," increment by the pointer and the like. Faster instructions such as one-word memory transfers or sign-bit complementations do not count. A square root counts as 10 operations [3].

In the case of the schemes of Macleod and Rosenfeld, the "expense" as defined above is a function of picture size and it approaches a minimum as the picture size increases. The extension of the above definition is taken to be the "expense" in this limit. Another feature of these two schemes is that they do not cover a complete range in edge orientations as does the operator of Hueckel. Their performance is best for edge of an optimum orientation, and they cannot be expected to function at all on edges perpendicular to this. In most applications, one would thus require passes with enough optimum orientations to give uniform response, but the number of such optimum orientations required may not be deduced from the observations of this work. We shall assume that two optimum orientations are sufficient.

The "expense" of Hueckel's operator as calculated in [3] is 23 operations per picture point.

In the limit of large pictures the scheme of Macleod requires for horizontal and vertical optimum orientations $\frac{1}{2}(3n^2 + 2n - 9)$ operations per point where n is the number of points on the side of the mask. Assuming these two orientations are sufficient and, for each point, one keeps the greater of the two edge weights thus generated, then "expense" is $\frac{1}{2}(3n^2 + 2n - 11)$ operations per point or 75 for the 7×7 mask and 266 for the 13×13 mask.

The scheme of Rosenfeld, employing DIFF only and keeping the maximum edge weights for all orientations, has an "expense" of $2(k + n_0) - 1$ where k is the power of 2 of the square size as in Section IV-C and n_0 is the number of orientations. If EDGE is used as well, then $2(k + n_0) + kn_0$ operations per point are required. Assuming that two orientations are sufficient, then the "expense" of DIFF alone with $k = 3$ is 9 operations per point and the "expense" of DIFF and EDGE together is 16 operations per point.

Economies may be introduced into each of these methods by various means. To mention a few: 1) The use of convolution with fast Fourier transforms can save with Macleod's scheme; 2) parallel processing could reduce the computation time of all three schemes but most effectively and readily that of Rosenfeld where the figures for "expense" then apply to the total number of operations required; 3) the entire picture does not necessarily have to be examined in order to find its edges. In particular, the additional information which Hueckel's operator supplies could facilitate economies here; 4) optical preprocessing could be beneficially employed by all schemes, but most of all by that of Macleod where the "expense" would then be reduced to one operation per picture point for each orientation. The expense of Rosenfeld's scheme would be reduced by $2 \cdot k$.

A. Edge Detection Performance

The results of evaluating the performance of the edge detectors described in the previous section are given in the graphs of Fig. 3 and Table II. Fig. 2(b) presents a pictorial display of the individual steps made in computing the two parameters using some typical images.

It is readily seen from Fig. 3 that of the two parameters evaluated, parameter 1, which indicates the ratio of the edge detector's signal to its noise consistently shows more pronounced differences in the performance of the various edge detecting programs tested than parameter 2, which gives a measure of the fraction of the edge covered by the signal. Thus, using parameter 1, the performance ratings of the programs tested may be consistently ordered over the nominal contrasts ranging from 6 to 18 with Macleod's scheme employing the larger mask scoring highest; following this scheme are Rosenfeld's using DIFF only, Rosenfeld's scheme using DIFF and EDGE, Macleod's scheme using the smaller mask, and finally Hueckel's operator. Above a nominal contrast 18 this order holds with the exception that Macleod's scheme employing the smaller mask performs slightly better in this parameter than Rosenfeld's DIFF and EDGE. At nominal contrast 3, the ratings indicated by both parameters are consistent with randomness for all programs tested, a feature which is elucidated further in Section VI-C.

It is not unreasonable that the parameter 1 rating of Rosenfeld's edge detection scheme is reduced when nonmaxima suppression is appended. Nonmaxima suppression neither adds more information nor selectively reduces the noise present in the output; its sole function is to compact the information present. In fact, since valid edge information tends to cluster, the process of nonmaxima suppression can be expected to discard a higher fraction of the points containing valid information than those containing spurious information. Nevertheless, the parameter 1 ratings of these two versions are close enough to indicate that major concern is not warranted. Both the parameter 1 and the parameter 2 ratings of these two versions of edge detectors could probably be brought closer together by decreasing t_{thresh} for the version without nonmaxima suppression.

A clear result indicated in Fig. 3 is that Hueckel's operator as implemented here does not perform as well on the sample images as the other two schemes. However, the performance based on the results obtained with images of nominal contrast 24 and above may be adequate for many applications. We should also bear in mind that even at a nominal contrast of 30, the amount of noise in relation to the signal is quite considerable. An increase in application density of this operator would undoubtedly improve its rating, but this would also increase the "expense" defined above in Section V-A.

VI. PSYCHOLOGICAL TESTS

A. Experimental Arrangement

It was thought useful to compare the ability of humans to detect edges within images of the type described in Section II. Accordingly, images of this type were put on 16 mm microfilm using the high resolution cathode ray tube picture output system of the University of Maryland Computer Science Center. 56 such pictures each were generated at the nominal contrasts 3, 6, 9, 12, and 15. At each contrast, the pictures were rotated (by computer), in equal numbers, by the angles 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°, yielding four different possible orientations for the edges themselves. The set was displayed on a microfilm reader to five observers, one observer at a time. The observers were located about 24 feet from the pictures in a darkened room; the displayed size of the pictures was 3 inches square, and they were shown at a rate of approximately one every 6 s.

Each group of 48 pictures (the pictures were serialized) contained equal numbers at each contrast although not necessarily equal numbers of each orientation. Subject to this restriction, the order of presentation of the pictures was determined randomly by computer.

The subjects were asked to identify, for each picture, which of the four possible orientations corresponded most with the edge's direction. They were told the answers to the first 48 pictures immediately after seeing each one and were then shown the remaining pictures in groups of 48. They were encouraged to rest after viewing each group but all elected to skip some rest periods. The order in which the groups of 48 were displayed after the first group was different for each observer in order to determine whether or not this order affected the observer's performance; it appeared not to affect it. The first group of 48 pictures were not used in the evaluation.

B. Results of the Psychological Tests

To analyze the results of the psychological test, the common, approximate, assumption was made that for each picture each observer either saw the edge clearly and identified it correctly or did not see it at all, in which case he would guess its orientation at random. Under this assumption, the maximum likelihood estimate of the fraction, f, of edges actually seen can be easily shown to be

$$f = \frac{n_i/n - r}{1 - r}$$

where n is the number of pictures displayed, n_i is the number of correct answers, and r is the probability of getting a correct answer purely by chance ($\frac{1}{4}$).

Table III summarizes the results.

C. An Ideal Observer

The noise introduced into the sample of pictures has the effect of distorting the edges; given a constant amount of noise, it is clear that those pictures of lower contrast will have edges which are more distorted. The question thus arises in the case of the edges of the lower two contrasts where a sizeable portion of the edges are missed by the observers, whether the edges are really there at all and are just hard to see or whether they are obliterated by the noise to the extent that they cannot be said to be present.

In order to obtain a measure of the above effect, a program was devised to determine the orientation an ideal observer would designate for each edge. The operation of the program is quite straightforward; for each orientation the difference in the sum of grey levels for the two sides is evaluated. The orientation which yields the difference of maximum magnitude is then selected. This program was used to compute the best orientation for all pictures of the sample used for the psychological test.

TABLE III
THE RESULTS OF THE PSYCHOLOGICAL TEST

NOMINAL CONTRAST	3	6	9	12	15	18
MEAN OF	.12	.44	.77	.84	.99	.97
STANDARD DEVIATION OF MEAN OF f	.02	.05	.04	.03	.01	.02

All the "best" orientations obtained in this way agreed with those programmed into the pictures of the highest four contrasts. However, at nominal contrast 6, one such "best" orientation differed from the programmed orientation and at nominal contrast 3, this was true of 25 of the 56 pictures. One may conclude that the edges were relatively well defined from nominal contrast 6 up but were undefined at nominal contrast 3. The poor results of the humans and of the edge detection programs at this contrast should thus be regarded as a reflection on the pictures themselves rather than of the edge detectors.

The results of the psychological test were determined using both the above "best" orientations and the programmed orientations. The value of f obtained using the "best" orientations for nominal contrast 3 was about 0.01 greater than that obtained using the programmed orientations, an entirely consistent result. The "best" orientation result is the one which is given in Table III.

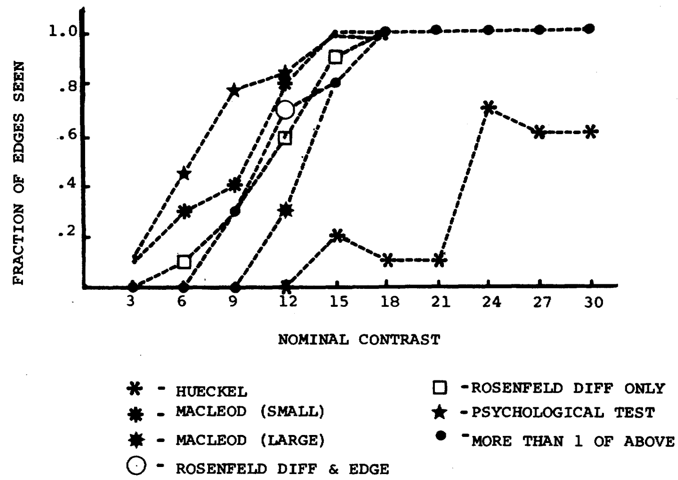


Fig. 4. The results of the psychological test compared with the edge detector programs.

VII. ON THE COMPARISON OF THE RESULTS OF THE PSYCHOLOGICAL TEST AND THE EDGE DETECTOR MEASUREMENTS

The psychological test results cannot be compared directly with the results of the edge detector evaluations because the quantities determined in either case are of a different nature. While the psychological measurement relies on the idea that each observer either sees or does not see an edge in every image, the edge detection performance measurements are quantified with two relatively continuous variables for each picture. The approach chosen for effecting the comparison of these two measurements was to formulate a criterion for reducing the two-parameter information of the edge detector measurements to the binary decision of the psychological measurement.

The criterion used is based on the notion that in order for an edge to be considered found, the 1's in the standard form described in Section III-A should be dense enough in the edge region so that these would not be confused with random fluctuations. Furthermore, the 1's in the edge region should be distributed over a sizeable fraction of the edge—enough of a fraction so that the entire length of the edge might be deduced. These above characteristics are represented, respectively, by the two parameters calculated. In the absence of a rigorous theoretical understanding of the problem, it was decided to pick thresholds for both parameters on the basis of the pictorial displays and consider an edge "seen" if the two parameters of the edge detector output exceed their respective thresholds. The thresholds chosen were 0.4 and 0.6, respectively, corresponding to an average density of 1's inside the edge region 4.3 times that of the average density outside this region and an expectation that more than 60 percent of the edge is "covered." The results thus derived are presented in Fig. 4 along with the results of the psychological test. These results are tabulated in Table IV along with the results using three other sets of thresholds in order to permit the reader to assess the effects of small variations of these thresholds.

It is clear that the criterion used has a degree of arbitrariness. Indeed, the decision as to whether or not an edge has been found, no matter how much information is supplied about the edge detector output, is arbitrary when isolated from specific applications and information available from other sources. Nevertheless, there are some cases where the edge detector output so clearly indicates an edge that it presents sufficient noise-free information for virtually any application. On the other hand, there are other cases where this output appears to give no valid edge information whatsoever. These two extreme cases invariably correspond to the extreme values of the two edge detection performance parameters. It is not unreasonable, therefore, to suggest that thresholding as described above could roughly convert the edge detection performance measurements to the form of the psychological test results.

Fig. 4 indicates that human subjects are superior to the programs tested at the three lowest contrasts. In the case of the best performing program, that of Macleod employing the larger mask, assuming that the likelihood function for its performance rating is a binomial distribution, then this superiority is determined with a certainty of over three standard deviations. This error is not quoted in Table IV due to the uncertainty in the thresholds them-

TABLE IV
EFFECT OF DIFFERENT THRESHOLDS

EDGE DETECTOR	THRESHOLD OF PARAMETER 1	THRESHOLD OF PARAMETER 2	NOMINAL CONTRAST 3	6	9	12	15	18	21	24	27	30
HUECKEL	.4	.6	.0	.0	.0	.0	.2	.1	.1	.7	.6	.6
	.4	.4	.0	.1	.0	.3	.3	.3	.4	.8	.7	.9
	.5	.5	.0	.1	.0	.0	.2	.1	.2	.7	.6	.9
	.6	.6	.0	.0	.0	.0	.1	.0	.1	.6	.4	.6
MACLEOD (LARGE)	.4	.6	.1	.3	.4	.8	1.0	1.0	1.0	1.0	1.0	1.0
	.4	.4	.1	.4	.6	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	.5	.5	.1	.3	.5	.8	1.0	1.0	1.0	1.0	1.0	1.0
	.6	.6	.1	.2	.2	.7	1.0	1.0	1.0	1.0	1.0	1.0
MACLEOD (SMALL)	.4	.6	.0	.0	.0	.3	.8	1.0	1.0	1.0	1.0	1.0
	.4	.4	.0	.0	.0	.4	.9	1.0	1.0	1.0	1.0	1.0
	.5	.5	.0	.0	.0	.1	.6	.7	1.0	1.0	1.0	1.0
	.6	.6	.0	.0	.0	.0	.2	.4	.8	1.0	1.0	1.0
ROSENFELD DIFF ONLY	.4	.6	.0	.1	.3	.6	.9	1.0	1.0	1.0	1.0	1.0
	.4	.4	.0	.1	.4	.8	1.0	1.0	1.0	1.0	1.0	1.0
	.5	.5	.0	.1	.3	.8	.9	1.0	1.0	1.0	1.0	1.0
	.6	.6	.0	.0	.2	.6	.8	.9	1.0	1.0	1.0	1.0
ROSENFELD DIFF AND EDGE	.4	.6	.0	.0	.3	.7	.8	1.0	1.0	1.0	1.0	1.0
	.4	.4	.0	.0	.3	.7	.8	1.0	1.0	1.0	1.0	1.0
	.5	.5	.0	.0	.2	.5	.8	.8	.9	1.0	1.0	1.0
	.6	.6	.0	.0	.0	.1	.5	.7	.7	.8	1.0	1.0

selves. If the two thresholds were lowered to about 0.3 and 0.35, respectively, then Macleod's scheme with the larger mask would yield results comparable to those of the human subjects. Fig. 5 provides the reader with pictorial displays of the additional edge detector outputs of this program which would be considered successful were the thresholds lowered to these values. It is the opinion of the authors that for the most part, these outputs should not be considered successful. They are presented so that the reader may decide for himself. The reader should bear in mind, however, the fuzzy nature of the boundary between successful and unsuccessful outputs. While he may decide that some of the outputs of Fig. 5 are marginally acceptable it is likely that he would find, in the same way, that some of the outputs which exceeded the threshold requirements and are therefore not shown were unacceptable.

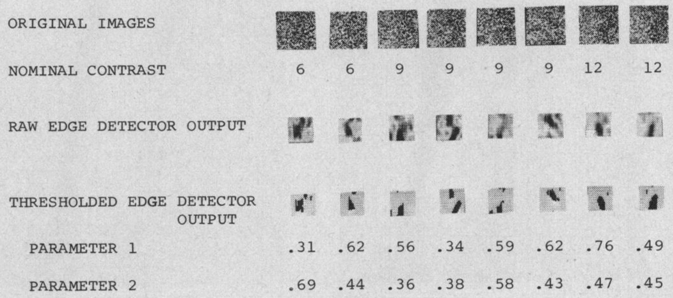


Fig. 5. Outputs of Macleod's edge detector (large) which would have to be regarded as successful in order to rate its performance equal to that of humans in Section VI-3.

VIII. PERFORMANCE OF THE EDGE DETECTOR PROGRAMS ON A REAL PICTURE

As a qualitative check on the results of the previous sections, the edge detection programs evaluated in this work were used on satellite imagery. For this purpose, an ERTS photograph was scanned to produce a 4096×4096 matrix of 64 grey levels. A 216×216 square was selected from this matrix from which 4 test pictures were made. The first consisted of the original scene and the remaining 3 were obtained by adding noise by changing in a random way about 20, 40, and 60 percent, respectively, of the points in the original scene. The 4 test pictures are shown in Fig. 6(a). Addition of the noise was accomplished in the following manner. Let p be the expectation of the fraction of points to be modified, 0.2, 0.4, or 0.6 here. For every point of the picture, a pseudo-random number between 0 and 1 was generated. If this number was greater than p, the point was left unaltered. If it was less than or equal to p, another pseudo-random number was generated, multiplied by 64 and rounded down to the nearest integer. The grey level value of the point was set equal to the result.

The resulting pictures differ from the artificial test images used in this work in that the noise was added in a different manner and it was the noise that was varied rather than the contrast. It is felt, however, that the relationship between the amounts of noise and contrast is really the property of general interest not the magnitude or specific nature of either one. Varying the noise rather than the contrast and using a different kind of noise, then,

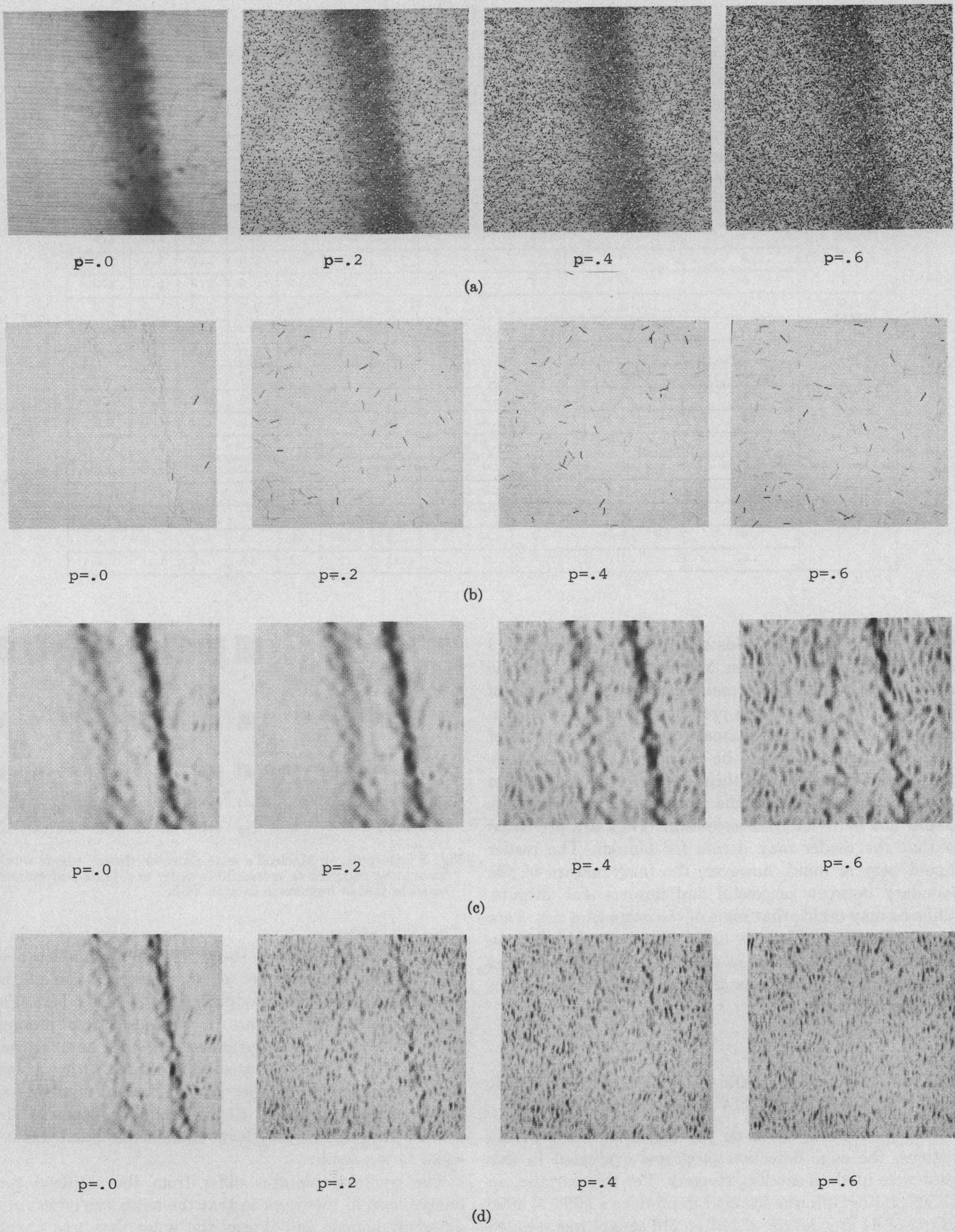
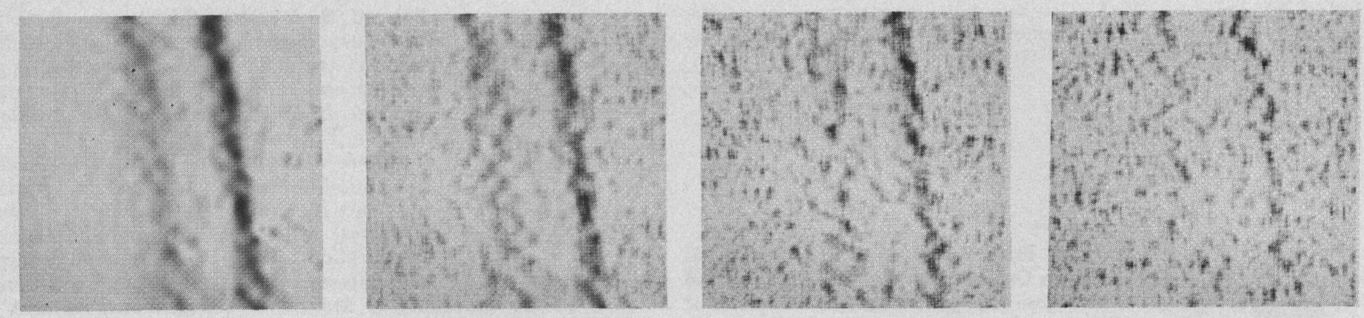
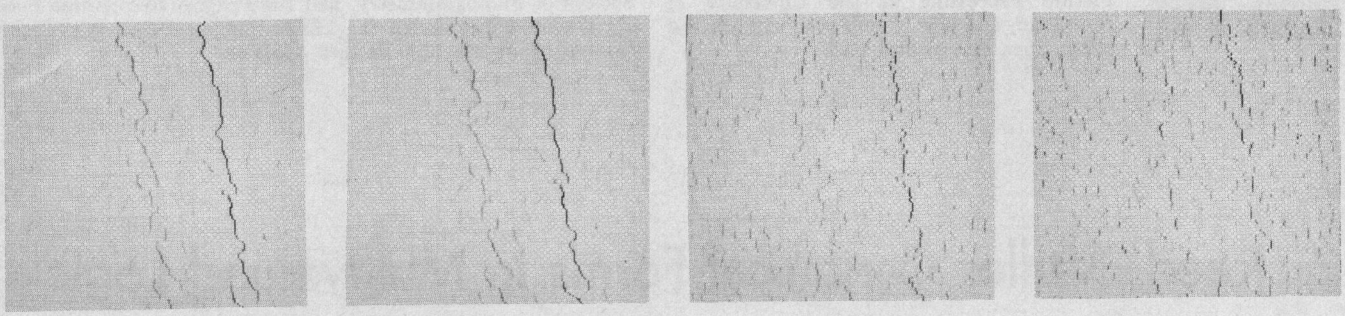


Fig. 6. (a) A satellite photograph with noise artificially added given by p. (See Section VII.) (b) The local visual operator (Hueckel) output from images of Fig. 6(a). (c) The Gaussian mask edge detector (Macleod, larger mask) output from images of Fig. 6(a). (d) The Gaussian mask edge detector (Macleod, smaller mask) output from images of Fig. 6(a). (e) Rosenfeld's size 3 edge detector without nonmaxima suppression output from images of Fig. 6(a) (DIFF). (f) Rosenfeld's size 3 edge detector with nonmaxima suppression output from images of Fig. 6(a) (DIFF and EDGE).



p=.0 p=.2 p=.4 p=.6

(e)



p=.0 p=.2 p=.4 p=.6

(f)

Fig. 6. *Continued.*

tests in a small way the generality of the results of this work as well as their qualitative correctness.

The particular image used for this test was chosen because its dominant feature was a vertical dark stripe yielding two vertical edges. As mentioned in Section III, the tested versions of both Rosenfeld's and Macleod's schemes are biased against horizontal edges to a much greater extent than would normally be the case. Consequently, the absence of all but vertical edges made this image a convenient choice. Also, the simplicity of the image selected made it easy to assess the performance or the edge detectors in the presence of noise without introducing some of the other factors mentioned in Section I.

The outputs of the five edge detector programs operating on these test pictures are shown in Fig. 6(b)–(f). These figures display the edge weights described in Section III multiplied by a constant selected to spread their values up to 63. This procedure for displaying the results worked well for all cases except Hueckel's operator on the noise-free picture where a few very high edge weights had the effect of obscuring the otherwise adequate performance.

The reader may verify the close qualitative agreement between Fig. 6(b)–(f) and the numerical results obtained from the sample of artificial pictures. He may also verify the results of the psychological test by comparing his own ability to distinguish the edges in the noisiest picture to (admittedly, his perception of) the performance of the edge detection programs on this picture.

IX. CONCLUSIONS

A method to quantitatively compare edge detection schemes using artificial pictures was presented. It was used to evaluate the ability of three such schemes to function in the presence of noise. A test on a real picture was run of the edge detection schemes evaluated in this work, and it yielded results of qualitative agreement with those of the new method. A psychological test was also run to determine how well human visual performance compared with that of the edge detection schemes.

REFERENCES

[1] M. H. Hueckel, "An operator which locates edges in digitized pictures," *J. Ass. Comput. Mach.*, vol. 1, pp. 113–125, 1971.

[2] ——, "An operator which locates edges in digitized pictures," Artif. Intel. Proj., Stanford Univ., Stanford, Calif., AIM-105, Oct. 1969.

[3] ——, "A local visual operator which recognizes edges and lines," *J. Ass. Comput. Mach.*, vol. 20, pp. 634–647, Oct. 1973.

[4] I. D. G. Macleod, "On finding structure in pictures," in *Picture Language Machines*, S. Kaneff, Ed. New York: Academic, 1970, p. 231.

[5] ——, "Comments on "techniques for edge detection"," *Proc. IEEE* (Lett.), vol. 60, p. 344, Mar. 1972.

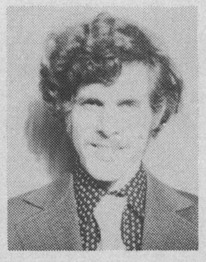
[6] A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. Comput.*, vol. C-20, pp. 562–569, May 1971.

[7] 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.

[8] K. C. Hayes, Jr., and A. Rosenfeld, "Efficient edge detectors and applications," Comput. Sci. Center, Univ. Maryland, College Park, Tech. Rep. TR-207, Nov. 1972.

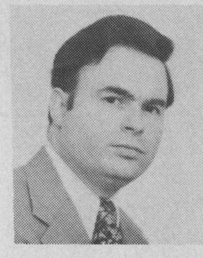
[9] A. Herskovits and T. O. Binford, "On boundary detection," MIT Project MAC, Artificial Intelligence Memo 183, July 1970.

[10] J. R. Fram and E. S. Deutsch, "A quantitative study of the orientation bias of some edge detector schemes," Univ. Maryland, College Park, Tech. Rep. TR 285, Jan. 1974.



Jerry R. Fram was born in Berkeley, California on October 31, 1941. He received the B.A. degree in physics from the University of California, Berkeley, in 1964 and the Ph.D. degree in physics from the University of Maryland, College Park, in 1973.

He taught physics in Central America as a Peace Corps Volunteer. He has worked in Picture Processing at the University of Maryland, and now is employed at Finn and Fram, Inc., Sun Valley, Calif.



Edward S. Deutsch (M'72) was born in Chesham, England, on April 23, 1943. He received the B.Sc. degree in electrical engineering and the Ph.D. degree in computer sciences from the University of London, London, England.

He has worked for the British Post Office's Communications Department on optical character recognition systems, and for Thorn Electrical Industries Ltd., and the Instron Corporation (U.K.) on electrical control problems. He was on the Professorial Staff of the Computer Science Center at the University of Maryland, College Park, his area of interest being the processing of two-dimensional digital information. He was also a private consultant. He is now with P. S. Ross & Partners, Management Consultants, Toronto, Ont., Canada.

Dr. Deutsch is a member of Sigma Xi, the Institution of Electrical Engineers, the Association for Computing Machinery, the American Society of Photogrammetry, and the Pattern Recognition Society. He is also a member of the IEEE Computer Society's Technical Committee on Machine Pattern Analysis.

Reed–Muller Canonical Forms in Multivalued Logic

K. L. KODANDAPANI AND RANGASWAMY V. SETLUR

***Abstract*—Canonical forms for m-valued functions referred to as m-Reed–Muller canonical (m-RMC) forms that are a generalization of RMC forms of two-valued functions are proposed. m-RMC forms are based on the operations $\oplus_m$ (addition mod m) and $\cdot_m$ (multiplication mod m) and do not, as in the cases of the generalizations proposed in the literature, require an m-valued function for m not a power of a prime, to be expressed by a canonical form for M-valued functions, where $M > m$ is a power of a prime. Methods of obtaining the m-RMC forms from the truth vector or the sum of products representation of an m-valued function are discussed. Using a generalization of the Boolean difference to m-valued logic, series expansions for m-valued functions are derived.**

***Index Terms*—m-Reed–Muller canonical (m-RMC) forms of multivalued functions, multivalued Boolean difference, series expansions.**

I. INTRODUCTION

RECENTLY there has been an increasing interest in multivalued logic design. Most of the work in multivalued logic is based on Post algebra [1]. This algebra is based on the connectives $\vee$, $\cdot$, and $\sim_m$ which are defined as follows.

If X and Y are m-valued variables,

$$X \vee Y = \max (X,Y)$$
$$X \cdot Y = \min (X,Y),$$

and $\sim_m X$ is the permutation

$$\begin{pmatrix} 0 & 1 & \cdots & m-1 \\ 1 & 2 & \cdots & 0 \end{pmatrix}$$

of the values of X. The modular algebra [2] based on the connectives $\oplus_m$ (addition modulo m) and $\cdot_m$ (multiplication modulo m) has been shown to be functionally complete, if m is a prime number. It has been observed in [3] that the modular algebra is simple to manipulate in circuit synthesis and minimization.

In any logic system, the study of canonical forms is important. This is partly due to the fact that in many cases, the canonical representation of a function will be the starting point for obtaining a minimal or a reduced expression for a function. The canonical forms based on the connectives $\vee$, $\cdot$, and $\sim_m$ were first studied by Post [1]. Several authors [4]–[6] have also studied canonical forms for multivalued functions based on $\vee$, $\cdot$, and unary connectives other than $\sim_m$.

There have been a few attempts to obtain generalization of Reed–Muller canonical (RMC) forms in two-valued logic to multivalued logic. Cohn [7] studied the canonical forms based on $\oplus_m$ and $\cdot_m$ for m-valued functions when m is a prime number. Pradhan and Patel [8] have proposed an m-valued algebra based on the operations of the field $GF(m)$ for m, a power of a prime. They have shown that any m-valued function for m, a power of a prime, can be expressed in a form similar to the RMC forms of two-valued functions. It is of interest to investi-

Manuscript received October 8, 1973; revised September 29, 1974.
K. L. Kodandapani was with the School of Automation, Indian Institute of Science, Bangalore, India. He is now with the Department of Electrical Engineering and Computer Science Program, Indian Institute of Technology, Kanpur, India.
R. V. Setlur is with the School of Automation, Indian Institute of Science, Bangalore, India.