Edge Evaluation Using Local Edge Coherence

LES KITCHEN AND AZRIEL ROSENFELD, FELLOW, IEEE

Abstract—A method of evaluating edge detector output is proposed, based on the local good form of the detected edges. It combines two desirable qualities of well-formed edges—good continuation and thinness. It yields results generally similar to those obtained with measures based on discrepancy of the detected edges from their known ideal positions, but it has the advantage of not requiring ideal positions to be known. It can be used as an aid to threshold selection in edge detection (pick the threshold that maximizes the measure), as a basis for comparing the performances of different detectors, and as a measure of the effectiveness of various types of preprocessing operations facilitating edge detection.

I. Introduction

THE CONCEPT of an *edge* is a difficult one to define precisely. The stimulus conditions that cause the perception of an edge by humans are by no means simple to describe. There are many well-known visual paradoxes in which an edge is clearly seen where none physically exists (see, for example, Cornsweet [4], Dember [5], or Gregory [9]). In the analysis of images by computer, exactly what constitutes an edge depends greatly on the objectives of the analysis.

Keeping the above in mind, we can nonetheless regard an *edge* as the boundary between two adjacent regions in an image, each region homogeneous within itself, but differing from the other with respect to some given local property. Thus, an edge should ideally be line-like.

In this paper, we restrict our attention to the simplest case, luminance edges, although the edge evaluation techniques we present below are applicable to color or texture edges as well. Luminance edges in an image have many possible causes in the original scene: discontinuities in surface properties (such as reflectance), in surface orientation, in illumination (shadows, for example), or in depth (causing occlusion of one surface by another). However the interpretation of the cause of an edge will not concern us here.

Luminance edges (henceforth just *edges*) are important features in image analysis, and, accordingly, many schemes have been devised for detecting them. Here we are concerned chiefly with so-called enhancement/thresholding edge detectors. In the enhancement step, an operator which computes local luminance differences is applied to an image. Such an operator will have a high response when positioned on the boundary between two regions but little or no response within each region. (The operators discussed below also compute an estimate of the direction of luminance change.) In the next step the edges in the image are extracted by suitably thresholding the operator output. The final result of processing is a binary picture, pixels deemed to be on an edge (*edge pixels*) having the value one, all others (*nonedge pixels*) having the value zero.

It is of interest to evaluate the quality of the output of an edge detector, both to compare one detector scheme with another, and also to study the behavior of a given detector under different conditions and parameter settings. Several authors have proposed techniques for edge evaluation. In the next section we review their work.

II. Survey of Previous Work

Fram and Deutsch [6], [7] studied the effect of noise on various edge detector schemes. For this purpose they used synthetic images composed of three vertical panels. The outer panels were of two different grey levels; the narrow inner panel interpolated between these grey levels. It was considered that the position of the edge was defined by this central panel and only here should an edge detector respond. Images were generated for a number of different levels of contrast between the two outer panels, and to each image was added identically distributed zero-mean Gaussian noise.

Several different edge enhancement techniques were applied to these test images, and thresholds were chosen so that the number of detected edge points was as close as possible to the number of points expected for a well-found edge, based on inspection of a sample of detector outputs. The thresholded output was evaluated according to two measures. The first, P_1, estimated what fraction of the detected edge pixels were actually edge points. The second, P_2, estimated what fraction of the vertical extent of the edge was covered by detected edge pixels. These estimates are possible because it was known that edge pixels actually due to the edge could be found only in the central panel, and it was assumed that edge pixels due to noise would be uniformly distributed throughout the image.

As would be expected, the experimental results showed that edge detector performance, as measured by P_1 and P_2, improves when contrast is increased relative to noise. They also demonstrated that some edge detection schemes perform consistently better than others.

Manuscript received January 12, 1981. This work was supported by the Defense Advanced Research Projects Agency and the U.S. Army Night Vision Laboratory under Contract DAAG-53-76C-0138 (DARPA order 3206).

The authors are with the Computer Vision Laboratory, Computer Science Center, University of Maryland, College Park, MD 20742.

0018-9472/81/0900-0597$00.75 ©1981 IEEE

While their measures are directly applicable only to vertical edges, Fram and Deutsch also performed experiments with synthetic oblique edges. They did this by the expedient of numerically rotating the enhancement output until it corresponded to a vertical edge. It could then be thresholded and evaluated as if it had been vertical. By this means they examined the sensitivity of the detectors they used to edge orientation.

The approach of Abdou and Pratt [2] is more analytic. (See also Abdou [1].) Using a simple model for the digitization of a straight edge passing through the center of an operator's domain, they geometrically analyzed the sensitivity of a number of edge enhancement operators to the orientation of the edge. They similarly analyzed the fall-off of operator response with displacement from the center of the domain for straight edges with vertical and diagonal orientations.

They described a statistical design procedure for threshold selection in noisy images with vertical and diagonal edges. Using additive Gaussian noise as an example, they derived the conditional probability distributions of operator response for a number of enhancement operators, given the existence or nonexistence of an actual edge. They could thus compute for each operator the probabilities of correct and false detection as a function of threshold and of noise level. By this means they showed the superiority of some detection schemes over others. They also presented a pattern-classification approach to threshold selection using training samples of edge and nonedge neighborhoods and gave experimental results for a number of edge detectors in discriminating edge from nonedge neighborhoods using this approach. These results show a similar ordering of the quality of the various edge detection schemes.

More relevant to the present paper, Abdou and Pratt provided another experimental comparison of the various edge detector schemes using Pratt's figure of merit of edge quality [12]. They used synthetic test images very similar to those of Fram and Deutsch above. The only difference worth remarking on is that Abdou and Pratt vary the relative strength of signal to noise by holding the contrast constant and changing the standard deviation of the added noise. Pratt's figure of merit is based on the displacement of each detected edge pixel from its ideal position (known from the geometry of the synthetic image), with a normalization factor to penalize for too few or too many edge points being detected. Its definition is

$$F = \frac{\sum_{i=1}^{I_A} \frac{1}{1 + \alpha(d(i))^2}}{\max\{I_A, I_I\}}$$

where I_A is the actual number of edge pixels detected; I_I is the ideal number of edge pixels expected (known from the geometry of the synthetic image); $d(i)$ is the miss distance of the ith edge pixel detected; and α is a scaling factor to provide a relative weighting between smeared edges, and thin but offset edges. For these experiments Abdou and Pratt set $\alpha = 1/9$. Like Fram and Deutsch's parameters P_1 and P_2, this figure of merit was implemented for vertical edges, but Abdou and Pratt also present a modification of it for diagonal edges.

Unlike Fram and Deutsch, Abdou and Pratt used the less arbitrary procedure of choosing thresholds so as to maximize the figure of merit. The experimental results showed, as one would expect, that the figure of merit declines with increased noise, and also again showed the superiority of certain edge detection schemes over others.

The work of Bryant and Bouldin [3] is different in several respects. They used real aerial photographs instead of synthetic images. Their threshold selection was based on accepting a fixed upper percentile of the distribution of enhanced edge output. More significantly, they proposed two quite distinct edge evaluation measures. One, called *absolute grading*, is based on the correlation of the edge detector output with an ideal "key" output, this key being determined apparently by hand. Their other technique, called *relative grading*, is rather novel. Omitting the details, it is based on comparing the output of a number of detectors, and rating each detector by how often it agrees with the consensus of the other detectors in deciding whether an edge exists at each pixel. By these means they compared a number of edge detectors and were able to some extent to quantify the improvement in edge output achieved by such post-processing as edge-linking and edge-thinning. They also gave an example of effect of threshold level on the absolute grade of an edge detector.

Relative grading, while an interesting idea, suffers from a number of weaknesses. Its results depend on the details of the consensus determination used and on the particular mix of operators chosen for comparison. Most important, it is completely oblivious to detection errors made by all detectors and may even penalize a good detector that does not make an error made by a majority of bad detectors.

Aside from relative grading, all methods discussed above require prior knowledge of the location of the actual edge, since they are more or less based upon the discrepancy between the detected edge pixels and the ideal position of the edge. This is fine for experiments with controlled synthetic images but raises questions when applied to real images, since the determination of edges in such pictures is very much the subjective decision of a human observer. The techniques are completely inapplicable to images for which the actual edge locations are unknown.

Further, the discrepancy between detected and ideal edge is not the sole determinant of the quality of edge output (see Fig. 1). Here we have two detected edges, both of equal discrepancy from the ideal. However, one of them is clearly preferable, since the detected edge is continuous rather than fragmented. It is clear that some attention should be paid to the good form of the detected edge.

Finally, none of the above edge evaluation measures take any account of the edge direction information produced by most edge enhancement operators. This information is used in many applications and is an important consideration in determining the good form of an edge. Even though a set of edge pixels may lie in the shape of a well-formed

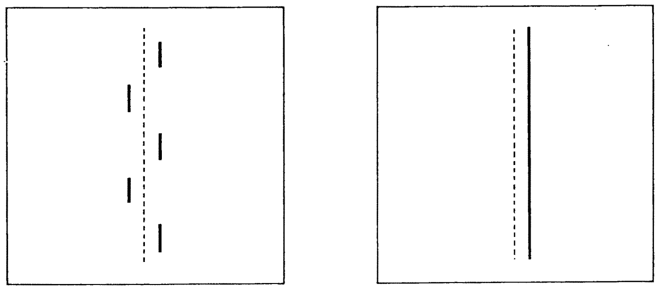

Fig. 1. (a) Disconnected edge. (b) Well-connected edge. Both have equal displacement from ideal edge position. (Ideal edge position shown by dotted line, detected edge by heavy line.)

edge, something is amiss if the estimated edge directions are chaotic. Ideally the luminance gradient direction should be everywhere perpendicular to the edge, and perpendicular in the same sense.

III. Local Edge Coherence

Bearing in mind the deficiencies of the above techniques, we have developed an edge evaluation measure based solely on the criterion of good edge formation, without using any prior knowledge of ideal edge location. This new measure is intended as a supplement to existing measures, not a replacement, since it is clear that a measure which disregards the correct location of an edge cannot be a fully adequate measure (although the results presented below show that it is quite good). For example, an edge detector that systematically mislocates edges will receive, by our scheme, an evaluation measure equal to that of a detector which perfectly locates edges.

However, since the new measure does not require prior knowledge of edge location, it can be used much more freely, in particular on images for which this knowledge is lacking. In addition to the standard uses of comparing edge detector schemes, the new measure can be used for selecting and adjusting edge operators as they are applied to an actual image. For example, an edge detector threshold can be chosen so as to maximize the edge evaluation measure. This will be the threshold which extracts the best-formed edges. (This parallels the work of Weszka and Rosenfeld [13] on threshold evaluation for segmentation of regions. One of their techniques rated a threshold level on the basis of the busyness of the resulting thresholded image.) Notice that this process of edge operator selection and adjustment can be performed completely automatically, without any human intervention. In applications where edge extraction is an important part of the processing, the edge evaluation measure can serve as an indication of image quality.

The approach we have used is based on what we call *local edge coherence*. Essentially we examine every three by three neighborhood of the thresholded edge output, taking into account the direction output as well. If the center of the neighborhood is an edge pixel, then we call the neighborhood an *edge neighborhood* and rate it on the basis of two criteria, *continuation* and *thinness*, which should both be exhibited by a well-formed edge passing through the center of the neighborhood. Both these criteria are based on the working definition of an *edge* given in Section I. It should be locally line-like, with due regard for the consistency of direction of luminance changes. *Continuation* requires, ideally, that adjacent to the central pixel, along the edge (this is perpendicular to the gradient direction of luminance change at the center), that there be two edge pixels with almost identical direction which form the continuation of that edge. *Thinness* requires, ideally, that all the other six pixels of the neighborhood be nonedge pixels. The continuation and thinness ratings of an entire edge output can be measured as the fraction of edge neighborhoods satisfying these respective criteria.

Of course, for most images, very few edge neighborhoods will perfectly satisfy these two criteria, because of digitization problems and even slight noise. We therefore compute instead continuation and thinness scores, ranging from zero to one, with the overall scores being averaged over every edge neighborhood in the output. These scores are designed to take the value one for perfectly formed edge neighborhoods, dropping off only slightly for almost well-formed neighborhoods, but falling eventually to zero for badly formed neighborhoods.

The continuation score is computed as follows. Let $|\alpha - \beta|$ represent the absolute difference between two angles α and β, the difference ranging from zero to π radians. Let

$$a(\alpha, \beta) = \frac{\pi - |\alpha - \beta|}{\pi}.$$

This function ranges from one for identical angles α and β, linearly down to zero for angles that differ by half a revolution, that is, point in opposite directions. It thus measures the extent to which the two angles agree in direction.

Let us number the neighbors of an edge pixel as shown in Fig. 2. Let d stand for the edge gradient direction at the center pixel, and let $d_0, d_1, \cdots, d_7$ stand for the edge gradient directions at each of the eight neighbors, respectively. Let

$$L(k) = \begin{cases} a(d, d_k) \cdot a\left(\frac{\pi k}{4}, d + \frac{\pi}{2}\right), & \text{if neighbor } k \text{ is an edge pixel} \\ 0, & \text{otherwise.} \end{cases}$$

This function measures how well a neighboring pixel continues on the left the edge which passes through the central pixel. It is zero when the neighbor is not an edge pixel, since no continuation exists. When the neighbor is an edge pixel, its rating is composed of two factors: the first, $a(d, d_k)$, measures how well the edge gradient direction at the neighbor agrees with that at the center. The second factor, $a(\pi k/4, d_k + \pi/2)$, measures how close neighbor k is to the expected direction of leftward continuation of the edge, based on the direction at the center. The term $\pi k/4$ is the direction to neighbor k, and the term $d + (\pi/2)$ is at right angles to the gradient direction and therefore lies

3	2	1
4		0
5	6	7

Fig. 2. Numbering system for neighbors.

along the edge. Similarly we define

$$R(k) = \begin{cases} a(d, d_k) \cdot a\left(\frac{\pi k}{4}, d - \frac{\pi}{2}\right), & \text{if neighbor } k \text{ is an edge pixel} \\ 0, & \text{otherwise} \end{cases}$$

which measures how well neighbor k continues the edge toward the right.

Of the three neighboring pixels lying to the left of the central edge gradient direction, the one with the highest value of $L(k)$ is taken as the left continuation. Similarly, of the three pixels on the right, the one with the best value for $R(k)$ is taken as the right continuation of the edge. The average of these two best continuations is taken as the continuation measure C for the entire neighborhood.

The thinness measure T for the neighborhood is computed as that fraction of the remaining six pixels of the neighborhood which are nonedge pixels. This will range from one for a perfectly thin edge, down to zero for a very blurred edge.

Neither of these measures is independently useful for edge evaluation, as will be explained below. However, a linear combination of the two

$$E = \gamma C + (1 - \gamma)T$$

serves quite well for suitable values of γ. This parameter γ can be adjusted to give a relative biasing of the measure E in favor of well-connected edges as against thin edges. The choice of γ will also be discussed below.

While this approach to edge evaluation is a little *ad hoc*, no simpler technique seemed able to capture the notion of a locally well-formed edge. We were first led to investigate the possibility of an edge evaluation measure based on good form by an observation on compatibility coefficients for relaxation labeling [11]. The arrays of compatibility coefficients showed a particular diagonal tendency when derived from images with clear edges which was far less pronounced when derived from noisy or blurred images. We attempted to devise an edge evaluation measure based on characteristics of the compatibility coefficient arrays, and later on characteristics of the edge direction cooccurrence matrices, which are closely related. Preliminary experiments showed that none of these measures were satisfactory, although they suggested that a measure based on good form could ultimately be developed. Several techniques based on local properties of the edge output were investigated, culminating in the method presented here. This measure is intuitively reasonable, and more important, performs quite well, as the experimental results below demonstrate.

One defect of our measure (though shared by all others) is that it can only be applied after thresholding. We endeavored to remedy this by devising methods that treat all pixels as potential edge pixels, but weight their contributions by a function of their edge magnitudes. Unfortunately the enormous number of low-magnitude pixels distorts the measure, unless the weighting function is of such a form as to be tantamount to thresholding.

It should be pointed out that this approach can be easily adapted to measuring the good form of other features, such as lines or corners, which are normally detected by some sort of template matching.

IV. Experiments

We present here some experiments which investigate the behavior of a number of edge detection schemes under various conditions. To permit a comparison, we have tried to make our experimental setup as similar as possible to that of Abdou and Pratt. We have used the same edge detection schemes (although our measure also makes use of edge direction information), the same noise model (additive independent zero-mean Gaussian noise), the same threshold selection criterion (choosing that threshold which maximizes the evaluation measure), and for one series of experiments, essentially the same test image.

A. Test Images and Edge Detectors Tested

Two test images of edges were used. The first, 64 by 64 pixels, consisted of a left panel with grey level 115, a right panel with grey level 140, and a single central column of intermediate grey level 128. This we will call the "vertical edge" image. It is virtually the same as one of the test images used by Abdou and Pratt. In order to present conveniently edges at all orientations, we chose a second test image consisting of concentric light rings (gray level 140) on a dark background (grey level 115). This image was originally generated as a 512 by 512 image, with a central dark circle of radius 64, surrounded by three bright rings of width 32, these being separated by two dark rings of the same width, with a dark surround. The decision as to whether a pixel should be light or dark was based on its Euclidean distance from the center of the image. Then this image was reduced to size 128 by 128, by replacing each 4 by 4 block with a single pixel having the average grey level of the block. The reduction gave a convenient way of approximating, for curved edges, the digitization model used by Abdou and Pratt. We call this the "rings" image. While the edges in this test image are curved, they are locally almost straight, at all possible orientations.

To study the effects of noise, independent zero-mean Gaussian noise was added to each of the test images at seven different signal to noise ratios: 1, 2, 5, 10, 20, 50, and 100. Following Pratt, the signal-to-noise ratio (SNR) is

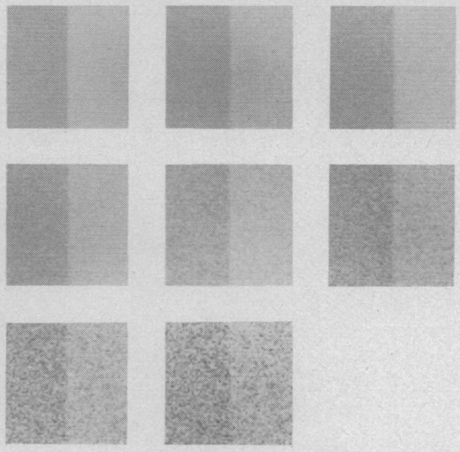

Fig. 3. Vertical edge test image, with various levels of noise. From left to right, top to bottom: no noise, SNR 100, 50, 20, 10, 5, 2, and 1. (Note: images are at twice the scale of those in Fig. 4.)

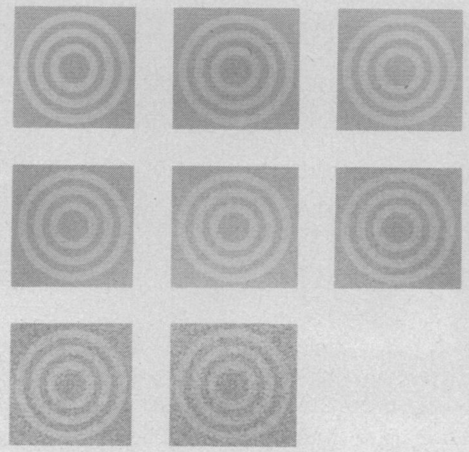

Fig. 4. Rings test image, with various levels of noise. From left to right, top to bottom: no noise, SNR = 100, 50, 20, 10, 5, 2, and 1.

defined to be

$$\mathrm{SNR} = \left(\frac{h}{\sigma}\right)^2$$

where h is the edge contrast (in this case 25), and σ is the standard deviation of the noise, adjusted to give the selected values of SNR. As an extreme case, we used an additional 64 by 64 test image with no well-formed edges, just Gaussian noise with mean 128 and standard deviation 16.

Fig. 3 shows the vertical edge image, noise free as well as with the various levels of added noise. Fig. 4 shows the same for the rings image. At the higher signal-to-noise ratios, the noise is almost imperceptible to the human eye. However, it is quite significant to the edge detectors used, since all of them have only small domains.

Ten different edge enhancement schemes were tested. The first group are the so-called "differential" operators. These measure the horizontal and vertical components of the luminance change by applying a pair of linear masks. The edge gradient direction is computed from these two components using the inverse tangent; and the edge gradient magnitude is computed either as the square root of the sum of squares of the two components, or as a sum (or max) of absolute values, for computational simplicity. Three different pairs of masks were used: those defined by Prewitt, Sobel, and Roberts. Since the edge magnitude can be computed in two ways, this gives six methods altogether. The second group are the "template-matching" operators: three-level, five-level, Kirsch, and compass-gradient. Each of these applies eight masks at every neighborhood. The edge magnitude is taken to be the strongest response out of these eight masks, and the edge gradient direction is given by the preferred orientation of the strongest responding mask. For details on and references to all these operators, see Abdou and Pratt [2].

B. Detailed Evaluation of One Detector

Before presenting an overall comparison of these edge detection schemes, we would like to examine in detail the results of the edge evaluation on a single scheme in order to discuss the properties of the edge evaluation measure itself. For this we have chosen the three-level template matching operator because it performed consistently better than any of the other operators in the comparison experiments described below. Even so, the results of the edge evaluation measure follow much the same pattern for the other operators as well.

Fig. 5 shows the histogram of edge magnitude outputs for the three-level operator applied to the rings image with SNR 50, and Fig. 6 shows the edge magnitude thresholded at nine levels equally spaced through its range. In Fig. 7 are shown plots of edge evaluation against threshold for various values of the weighting factor γ. Fig. 8 shows the same data but plotted instead against the fraction of pixels which are edge pixels at each threshold, scaled logarithmically. This is a better way of presenting the data, since it is the selection of edge pixels that really matters, not the threshold directly.

We see that the thinness measure alone ($\gamma = 0.0$) is of little use for edge evaluation. It reaches its maximum value at high thresholds since it rates a set of isolated edge pixels higher than an even slightly blurred edge. On the other hand the continuation measure performs reasonably well by itself ($\gamma = 1.0$), reaching a maximum value at a threshold which selects quite a good set of edge pixels. (This peak is more pronounced in Fig. 8, since changing the threshold near the maximum produces only a small change in the population of edge pixels. Notice that this threshold lies in the valley of the histogram.) However a close examination shows that these edges are several pixels thick. Better results are achieved with a lower value of γ. For the rest of this paper we will use $\gamma = 0.8$, since this value seems to give the best compromise between continuation and thinness.

Table I shows the maximum values of the edge evaluation measure, and the thresholds at which they occur, for the various values of γ. Fig. 9 shows the thresholded edge magnitude for a range of closely spaced thresholds around that at which the edge evaluation takes its maximum value

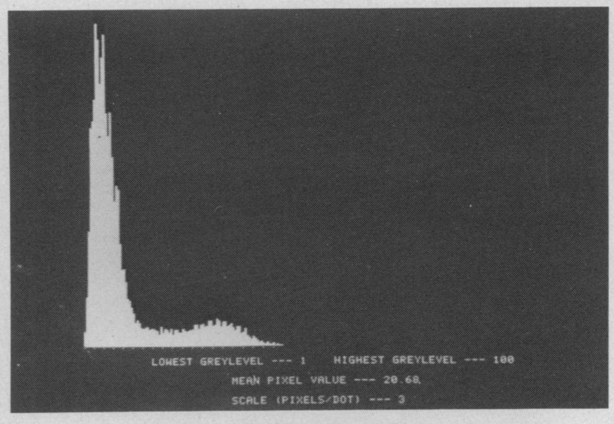

Fig. 5. Histogram of edge magnitude obtained by applying three-level operator to rings image at SNR 50.

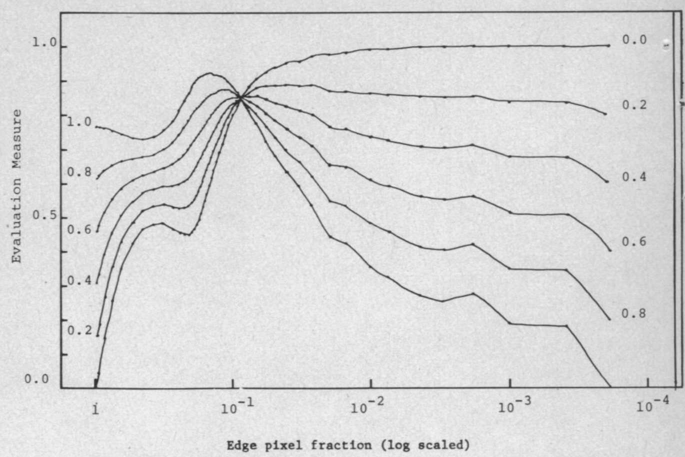

Fig. 8. Using rings test image at SNR 50 and three-level operator: edge evaluation against fraction of edge pixels at each threshold, for various values of parameter γ.

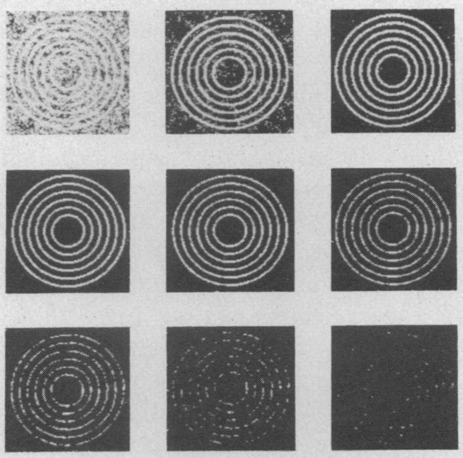

Fig. 6. Edge pixels extracted by thresholding edge magnitude (three-level operator) on rings image at SNR 50. Thresholds, from left to right, top to bottom: 10, 20, 30, 40, 50, 60, 70, 80, and 90 percent of range.

TABLE I
MAXIMUM VALUE OF EDGE EVALUATION MEASURE, AND THRESHOLD AT WHICH THIS OCCURS

$\gamma =$	0.0	0.2	0.4	0.6	0.8	1.0
Maximum value	1.000	0.890	0.856	0.852	0.876	0.922
At threshold (percent of range)	88–98	74	68	60	56	46

For various values of γ. Rings image, SNR = 10, three-level operator.

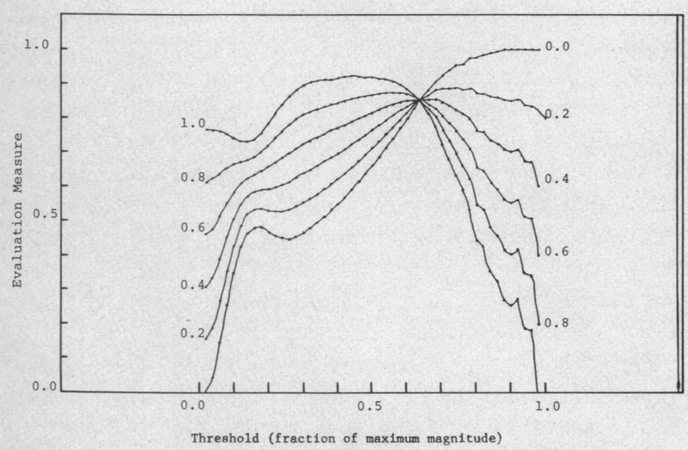

Fig. 7. Using rings test image at SNR 50 and three-level operator: edge evaluation against threshold for various values of parameter γ.

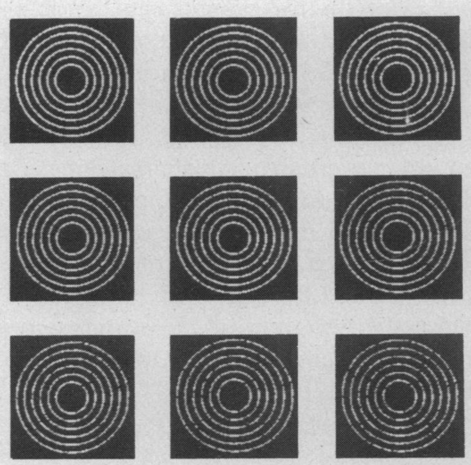

Fig. 9. Edge pixels extracted by thresholding edge magnitude from three-level operator on rings image at SNR 50. Thresholds from left to right, top to bottom: 48, 50, 52, 54, 56, 58, 60, 62, 64 percent of range.

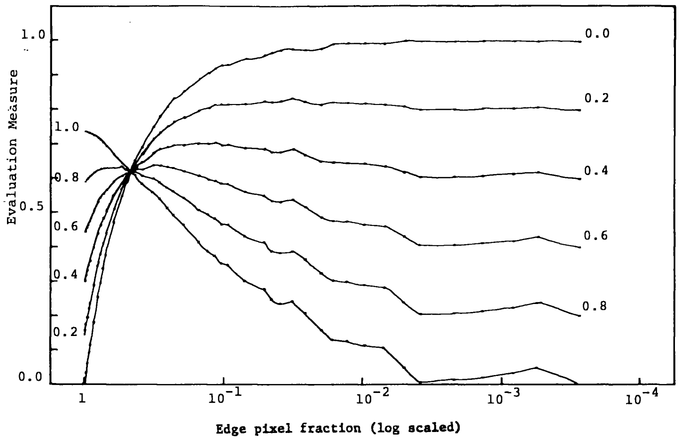

Fig. 10. Using test image of pure noise, and three-level operator: edge evaluation against fraction of edge pixels, for various values of parameter γ.

for $\gamma = 0.8$. Even though we have chosen $\gamma = 0.8$, two remarks should be made. First, values of γ from 0.6 to 0.9 produce similar results. Second, γ can be adjusted depending on the relative seriousness of broken edges as against thickened edges, for a given application. In general γ should be fairly high, since filling breaks in edges is usually a more difficult task than edge thinning.

To show that the peaks in Fig. 8 are actually caused by more or less well-formed edges, we give in Fig. 10 an analogous plot but for the test image of pure noise. The forms of the curves are quite different, without any well-defined peaks for the higher values of γ. However, this graph does reveal a noteworthy property of the edge evaluation measure: even on an image of pure noise, it is possible to choose a threshold which gives a moderately high value of the edge evaluation measure. At first thought this may seem to be a defect. However, on reflection, it is clear that this is an inevitable characteristic of any such measure based on local good form. Because of overlap between the neighborhoods to which the edge operators are applied, an isolated noise point will produce a correlated set of edge pixels. For example, the three-level operator will produce a tiny ring. Even though this ring is highly curved, it is coherent and will receive a moderate edge evaluation score. This evaluation score for isolated noise spots can be computed analytically as an intrinsic property of each edge detection scheme. For an image of pure noise, as used for Fig. 10, the evaluation is somewhat lower than one would expect, apparently because of interference between adjacent noise pixels. In summary, even in a noisy image there will be a certain occurrence of well-formed edges, either by accidental alignment or as an artifact of the edge detection scheme used. It is not the fault of the edge evaluation measure that it reflects this unavoidable property of the images and detection schemes used.

Fig. 11 illustrates the effects of various levels of noise on the edge evaluation measure. For clarity only a subrange of the data is plotted. Outside this subrange the plots for the different noise levels tend to converge. The results show a consistent pattern: the peaks decrease in step with the signal-to-noise ratio. Below SNR = 10 there are no clear

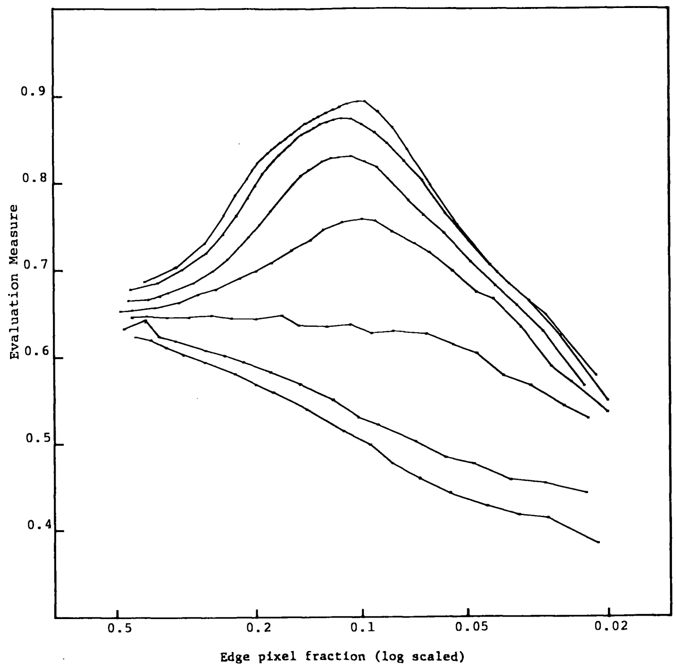

Fig. 11. Using rings test image and three-level operator: edge evaluation ($\gamma = 0.8$) against fraction of edge pixels at each threshold, for various values of SNR (top to bottom curve: 100, 50, 20, 10, 5, 2, 1).

peaks, but the shapes of the curves show that the presence of edges still has some effect on the edge evaluation. (Although we have not pursued the matter, this suggests that an edge evaluation measure might be based on the value E of the measure for the given image relative to the value measured for the same detector on an image of pure noise. But such a relative measure would be useful only for cases of high noise, when E has no clear peaks, and would not be a good means of comparing the outputs of different edge operators.) Away from the peaks, the evaluations for the different noise levels tend to become similar, while retaining the same ordering. This shows that a poor threshold leads to a bad selection of edges, no matter how noisy the original image.

All the above results are pretty much what one would expect intuitively from a measure of edge quality. They thus serve to confirm the validity of the edge evaluation measure. While the figures show the results for the rings image, the results for the vertical image are similar, and if anything, more distinct, since a vertical edge can be more cleanly digitized and has not even the slightest curvature.

C. Comparison of Detectors

Having established that the measure E behaves well, we now present a comparison among the ten edge enhancement operators mentioned above. Every operator was applied to the test image at the seven different signal-to-noise ratios, and at each noise level the threshold was adjusted to maximize E. Figs. 12 and 13 show these maximum values for the differential and template-matching operators, respectively, using the rings image. As expected, these results show that the three by three operators are far better than the two by two operators at detecting edges in the presence

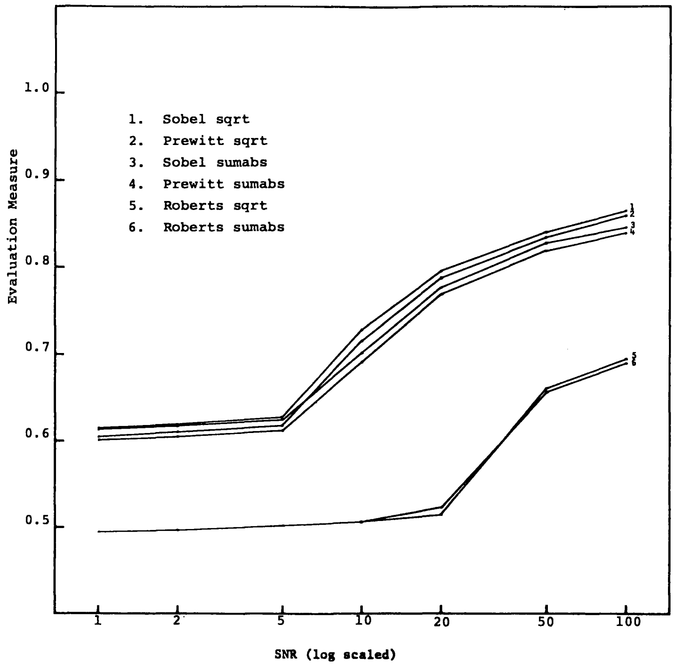

Fig. 12. Using rings test image: maximum edge evaluation ($\gamma = 0.8$) against SNR, for differential operators.

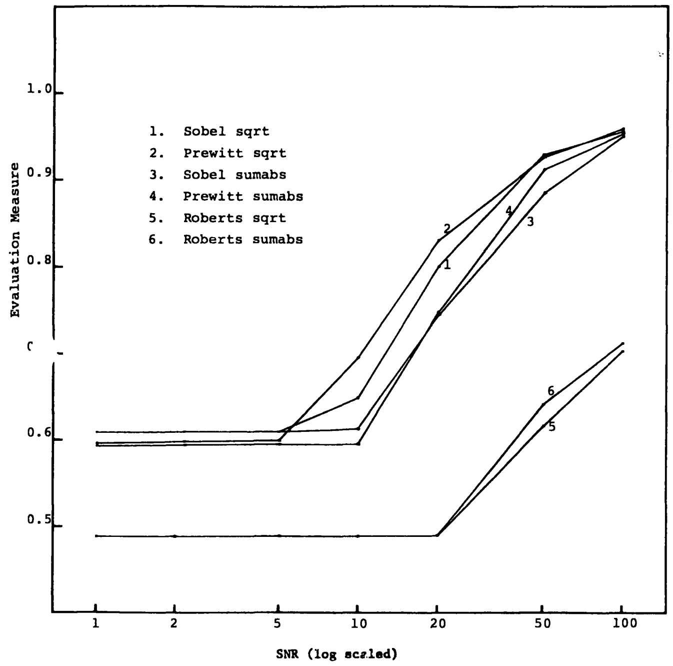

Fig. 14. Using vertical edge test image: maximum edge evaluation ($\gamma = 0.8$) against SNR, for differential operators.

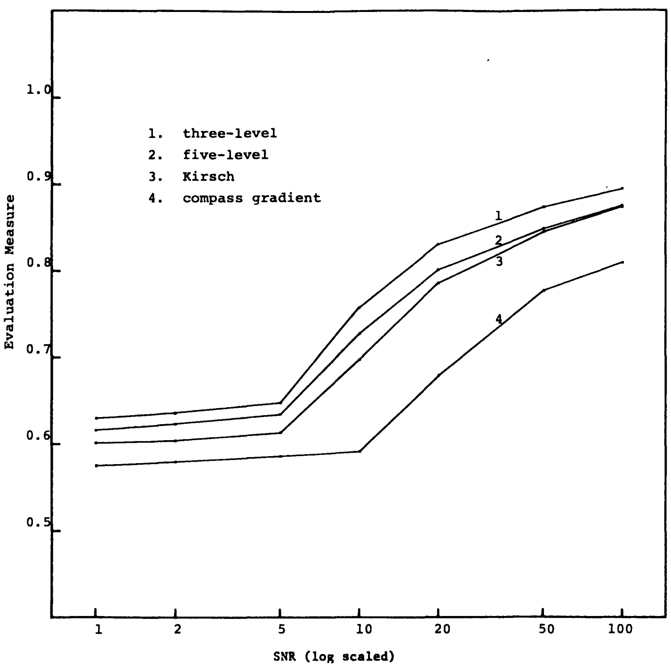

Fig. 13. Using rings test image: maximum edge evaluation ($\gamma = 0.8$) against SNR, for template matching operators.

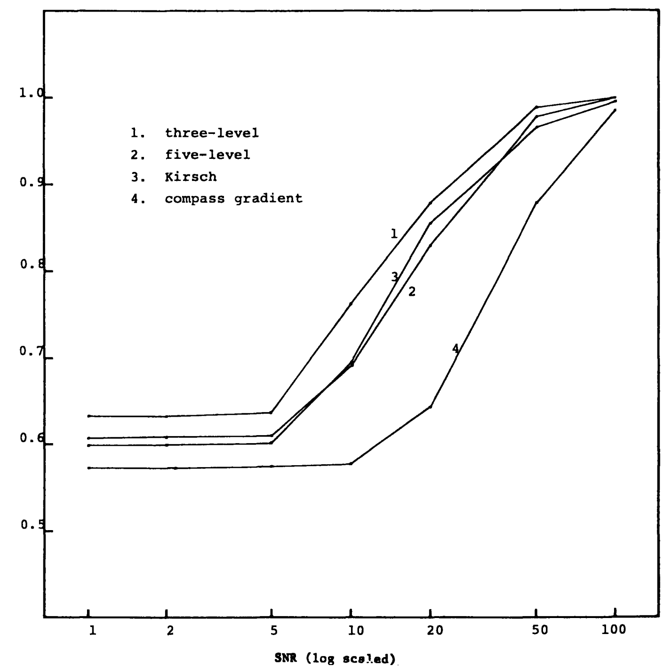

Fig. 15. Using vertical edge test image: maximum edge evaluation ($\gamma = 0.8$) against SNR, for template matching operators.

of noise. Among the three by three operators, the three-level operator is clearly the best, and the compass gradient the worst. The other four operators produce results of about the same quality. The same ordering is preserved if we subtract out the intrinsic response for each operator on pure noise, although the separations are not so great.

Analogous results for the vertical edge image are shown in Figs. 14 and 15. They are not directly comparable, especially at the lower SNR's, because the rings image has a greater density of edges. However some general remarks can be made. First, as explained earlier, the vertical edge gives a higher evaluation. Second, the evaluations of the four three by three differential operators are more spread out. This can be attributed to relative orientation biases in the four operators which are brought out by the vertical edge, but which are cancelled out over the full range of edge orientations in the rings image.

Overall, this comparison is in accord with the findings of Abdou and Pratt. Our results differ from theirs only when the difference between operators is small by both measures.

They also find the three by three operators consistently better than the two by two. However, at the highest signal-to-noise ratios, the performance of the two by two operators, according to their figure of merit, approaches that of the three by three, while our measure still reveals a considerable difference. This shows that while the two by two operators can properly locate edges at low noise levels, they poorly estimate the edge direction.

By both their measure and ours, the compass gradient is the worst of the three by three operators, but their figure of merit, while rating the three-level operator fairly highly, does not show it as clearly superior in all cases. These small discrepancies are not at all surprising, since the two edge evaluation schemes, after all, measure quite different characteristics of edges. The general agreement between the two schemes is encouraging. It serves both to confirm, in large part, the edge operator ratings of Abdou and Pratt, but from a different perspective; and also to strengthen our confidence in the usefulness of the measure E.

D. Effects of Other Operations

We also conducted some experiments to measure the effect on E of a number of commonly used operations: mean filtering, median filtering, nonmaximum suppression of edge output, and an edge improvement technique devised by Peleg [10]. Details of these experiments, omitted here, can be found in [8], but in summary, these operations generally affect E consistently with subjective judgments of edge quality.

V. Conclusion

We have presented a method for evaluating the quality of edge detector output based solely on the local good form of the detected edges. It combines two desiderata of a well-formed edge—good continuation and thinness. This measure behaves as one would like under the effects of change of threshold, noise, and other operations. The comparison experiments show that the results obtained with this measure are similar to those obtained with a measure based on the discrepancy of the detected edge from a known actual edge position. The small differences between the two methods reveal some properties of the operators not brought out by the other approach.

Like other evaluation measures, ours can be used to compare the effectiveness of different edge detection schemes and edge improvement schemes on synthetic images. However, since our measure does not require knowledge of the true location of edges, it has much wider application. It can be used to adjust parameters, such as thresholds, for optimum detection of edges in real images for which edge location is unknown. The evaluation of the detected edges can also serve as an indication of the quality of the original image. Further, the approach of using local coherence can be extended to the evaluation of other local feature detectors.

Acknowledgment

The help of D. Lloyd Chesley in preparing this paper is gratefully acknowledged.

References

[1] I. E. Abdou, "Quantitative methods of edge detection," Image Processing Institute, Univ. Southern California, Los Angeles, CA, USCIPI Rep. 630, 1978.

[2] I. E. Abdou and W. K. Pratt, "Quantitative design and evaluation of enhancement/thresholding edge detectors," *Proc. IEEE*, vol. 67, no. 5, pp. 753–763, May 1979.

[3] D. J. Bryant and D. W. Bouldin, "Evaluation of edge operators using relative and absolute grading," *Proc. IEEE Comp. Soc. Conf. on Pattern Recognition and Image Processing*, Chicago, IL, pp. 138–145, Aug. 1979.

[4] T. N. Cornsweet, *Visual Perception*. New York: Academic, 1970, pp. 268–284.

[5] W. N. Dember, *Psychology of Perception*. New York: Holt, Rinehart and Winston, 1966, pp. 196–206.

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

[7] ____, "On the quantitative evaluation of edge detection schemes and their comparison with human performance," *IEEE Trans. Comp.*, vol. C-24, no. 6, pp. 616–628, June 1975.

[8] L. Kitchen and A. Rosenfeld, "Edge evaluation using local edge coherence," Univ. Maryland, College Park, MD, Computer Sci. Tech. Rep. 981, Dec. 1980.

[9] R. L. Gregory, *Concepts and Mechanisms of Perception*, New York: Charles Scribner's Sons, 1974, pp. xxx–xxxx.

[10] S. Peleg, "Straight edge enhancement and mapping," Univ. Maryland, College Park, MD, Computer Sci. Tech. Rep. 694, Sept. 1978.

[11] S. Peleg and A. Rosenfeld, "Determining compatibility coefficients for relaxation processes," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, no. 7, pp. 548–555, July 1978.

[12] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1978, pp. 495–501.

[13] J. S. Weszka and A. Rosenfeld, "Threshold evaluation techniques," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, no. 8, pp. 622–629, Aug. 1978.