An Edge Detection Performance Measure Incorporating Structural Errors.

J. Besuijen, F. van der Heyden

University of Twente, Department of Electrical Engineering, Control Systems and Computer Engineering
P.O. Box 217, 7500 AE Enschede, The Netherlands

RESUME

Une nouvelle mesure à évaluer les résultats des méthodes de détection des contours des objets est proposé. Au contraire des mesures existants la mesure proposé utilise la structure des contours et des erreurs. Une méthode consistante pour la génération des contours idéal et des images à tester est introducée.

SUMMARY

At the moment edge detection methods do not provide results acceptable for higher level image analysis methods as image segmentation and shape measurement and recognition. An edge detection performance measure tuned to these applications would contribute considerably to the evaluation of the numerous edge detection methods proposed in literature. Up to now edge detection performance measures only use simple detection error statistics and can be used on simple test images only. The new performance measure proposed here is based on the following principle features required of edge and contour images for further image processing: The edges should be complete, and without false edge points, as thin as possible, in the right position, and any clustering of errors in the detector output should be avoided.

Introduction.

The detection of all and only the relevant edges in an image still is a complex task. The use of simple gradients methods directly on an image degraded by noise and/or blur can lead to missed edge pixels and the detection of many false edge points. Suppressing noise by low–pass filtering makes it difficult to localize edges and resolve detailed image structure. More recent zerocrossing techniques give more promising results, but claims for optimality do not always extend to the minimization of errors in the detector output, e.g. position errors and structural artifacts with the Marr–Hildreth method [1,2].

All these different types of errors make it of course very difficult to compare the results of the different edge detector schemes. During the evaluation of our latest contribution to the specter of edge detection methods, it appeared that quantitative edge detection evaluation methods had not developed as fast as image processing in general.

Methods for edge detection evaluation were among others developed by Fram and Deutsch[3], Abdou and Pratt[4], Peli and Malah[5]. None of these methods uses any structural information about the occurrence of errors. Furthermore Fram and Deutsch defined all pixels on the slope of an edge as edge pixels; Abdou and Pratt could evaluate only straight lines in 64x64 test images and Peli and Malah used 32x32 test images with only 16 grey levels.

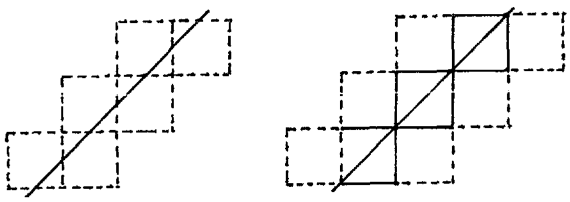
With the state of the art of image processing, a new figure of merit for edge detection methods is required, that can deal with a large specter of test images with 256 grey levels and preferably of a size of 512x512 pixels.

Test image generation.

The main problem in quantitative edge detection evaluation is the definition of the ideal edge in combination with the method of generation of the underlying undegraded test image. The set of test images should include all possible forms and directions of edges and the possibility to evaluate the performance at junctions and intersections of edges, as this structural information provides important clues for higher level image processing.

A modular method of test image generation was chosen, which constructs test images and ideal images from continuous curves. A primitive test image is formed by a continuous curve dividing the image in two regions of different constant grey level. Several primitive test images can be combined by superposition to form more realistic test images. The test images can be degraded by addition of noise, blurring with a point spread function, or superposition with a slowly varying grey level image. All pixels intersected by the continuous curve get a grey level proportional to the areas within the different grey level fields. This way a wide variety of test images can be generated from simple planar functions.

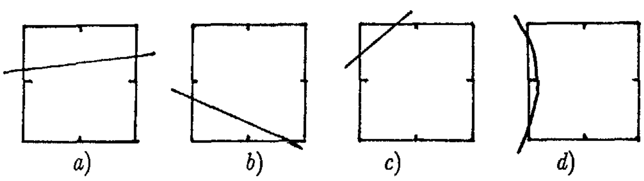
The ideal edge is usually defined as either a 4– or an 8–connected set of pixels. When either of these definitions is used strictly, this could for small offsets of diagonal lines, which leave the generated test image unchanged, result in large differences in the number of pixels detected correctly, as shown in figure 1.



a) *figure 1.* b)

A small offset in the position of the diagonal edge decides whether pixels l or r are selected for the ideal edge.
a) 8–connected. b) 4–connected.

To overcome this problem a maximum and a minimum ideal edge are defined. The maximum ideal edge is the set of all pixels, which are intersected by the continuous curve(s) used to generate the test image. The minimum ideal edge is a subset of the maximum ideal edge, which pixels are at least 8–connected from beginning till end of the curve. This is achieved by defining the minimum ideal edge as the set of all pixels which are intersected by the continuous curve(s), and have the entrance and exit points of the curve in different sides of the pixel, and have at least one half of a pixel–side between the entrance and exit point of the curve, on each side of the curve. This is illustrated in figure 2.



a) b) c) d)

figure 2.

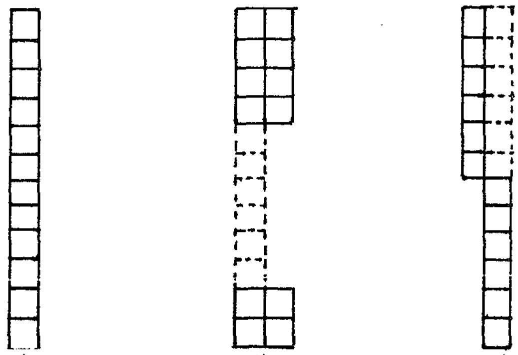
Examples of pixels intersected by the continuous curve. Pixels a) and b) fulfill the conditions stated above to be part of the minimum ideal edge, c) and d) do not.

Detection errors.

These definitions can only be meaningful in combination with the definitions of the errors. The obvious errors are:

1. Missing edge pixels.
2. False alarm pixels.

Most edge detection evaluation methods use these two types of errors to calculate a Figure Of Merit (FOM). It is clear that when an edge is only slightly displaced this is a less severe error, then when parts of the edge are missing all together in the edge detector output. This is why Abdou and Pratt[3] introduced the concept of displaced edge pixels. Unfortunately they were so rigorous as to regard all detected pixels as displaced edge pixels, even false alarm pixels at such a distance from the ideal edge, that no human observer would ever link them together. This results in dependence on the size of the test image and bias towards edge detectors with a blurred edge output, as shown in figure 3.



a) b) c)

figure 3.

a) The ideal edge. The edge detector outputs of b) and c) have the same FOM according to Abdou and Pratt. The outlines of the missing pixels of the ideal edge in b) and c) are dashed.

From the above example it becomes clear that displacement is only relevant in a small zone around the ideal edge, and can only be measured if the structure of the edge and the errors is incorporated in the evaluation method. Therefore three other types of errors are added to the two errors above:

3. Displaced edge pixels.
4. Superfluous edge pixels.
5. Clustering of errors.

To distinguish errors 3 and 4 from 1 and 2, an edge zone of width w is defined around the minimum ideal edge. The different types of errors can now be defined as follows:

1. Missing edge pixels: All pixels of the minimum ideal edge not detected, which can not be bypassed by a chain of pixels within the edge zone.
2. False Alarm pixels: All pixels detected outside the edge zone.
3. Displaced pixels: All bypasses as mentioned in 1.
4. Superfluous edge pixels: All pixels detected within the edge zone but outside the maximum ideal edge, which are not displaced pixels.
5. Clustering error: The average length of a connected chain of errors under 1, 2 or 4.

A displaced pixel has a displacement 0 if it is an element of the maximum ideal edge, otherwise the displacement is equal to the displaced pixel's shortest distance to a pixel of the minimum ideal edge.

The edge zone is currently determined by alternative 4– and 8–connected binary growth of the minimum ideal edge. The distance of the displaced pixels is determined by the number of grow–steps required to create that pixel in the edge zone.

Connectivity and search procedures for bypasses.

It is not difficult to determine the connectivity of certain groups of error pixels, but it can be very time consuming, esp. if the number of false alarms is high. The search procedure for the bypassing of missing pixels of the minimum ideal edge is in fact also a connectivity check: If a string of superfluous pixels is connected to the strings of pixels on both ends of the gap in the detected edge, it forms a bypass. This bypass then in turn has to be relieved of its superfluous pixels.

The bypass procedure does not take much time, because test images in general do not contain many edges and therefore the number of pixels in the edge zone is usually small. Depending on the number of false alarm pixels in the output image typical computing times for a 512x512 image are between 30 seconds and 4 minutes, on a computer with 4.5 Mb memory and some special image processing hardware.

Performance measurement.

The performance measure can then be composed from a weighted sum of the following partial errors, although experiments show that more insight about detector faults can be gained from separate evaluation of the partial errors.

Miss Error:

$$\epsilon_m = \frac{\#\ \text{missing pixels . average length}}{\#\ \text{minimum ideal edge pixels}}$$

False Alarm Error:

$$\epsilon_f = a_{FA} \frac{\#\ \text{false alarm pixels . average length}}{\#\ \text{pixels outside the edge zone}}$$

Blur Error:

$$\epsilon_b = \frac{\#\ \text{superfluous pixels . average length}}{\#\ \text{pixels in the edge zone}}$$

Displacement Error:

$$\epsilon_d = \frac{\#\ \text{displaced pixels . average displacement}}{\#\ \text{minimum ideal edge pixels}}$$

where # stands for "number of".

The factor a_{FA} in the false alarm error was introduced, because a result of an edge detection is already useless if only half of all the pixels outside the edge zone are false alarm errors. It seems desirable to tune both the factor a_{FA} and the width of the edge zone w to the resolution required from the edge detector.

Results.

To illustrate the capabilities of the performance measure two results are shown below. The test image consisted of a circle of radius 50 in an image of 512x512 pixels, with a background level 111 and the grey level inside the circle 143. Noise was added with a signal–to–noise ratio of 1.

Two edge detectors were tested, the Marr–Hildreth detector [1] with $\sigma = 3$, and the Combination of directional derivatives detector (CODD) [2] extended to four directional derivatives by including two diagonal derivatives, also with $\sigma = 3$. The parameters used for the performance measure were: a width w of 2, and false alarm factor a_{FA} of 2. The threshold on the edge detector output was interactively optimized to the best overall performance.

The minimal ideal edge consisted of 284 pixels, the edge zone of 1252 pixels and there were 232604 pixels outside the edge zone; the borders of the image are not used because the convolution is ill–defined there. Figure 4 shows the 128x128 out–takes with the circle of ideal edges, edge zones and the result images.

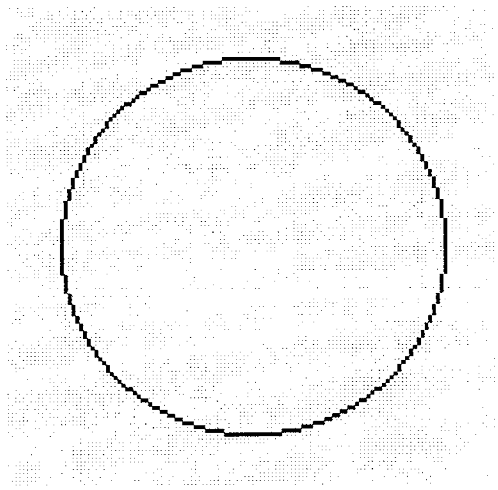


fig. 4a) The minimum ideal edge, which is 8–connected.

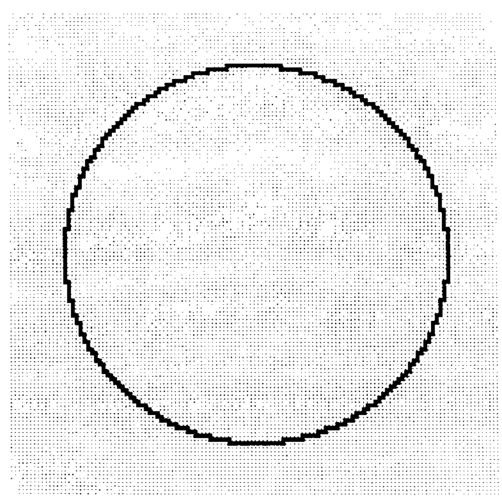


fig. 4b) The maximum ideal edge, which is 4–connected.

The results for the performance measure are given in table 1.

table 1

	Marr–Hildreth	CODD
ϵ_m	0.931	0
ϵ_f	0.276	0.00017
ϵ_b	0.054	0.348
ϵ_d	0.22	0.21

The performance measures clearly quantify the differences in fig 4d and 4e. The Marr–Hildreth operator misses many edge pixels and has many false edge pixels and both highly clustered. The CODD operator has no missing edge pixels, the contour is closed, and nearly no false alarms, but considerably more blur.

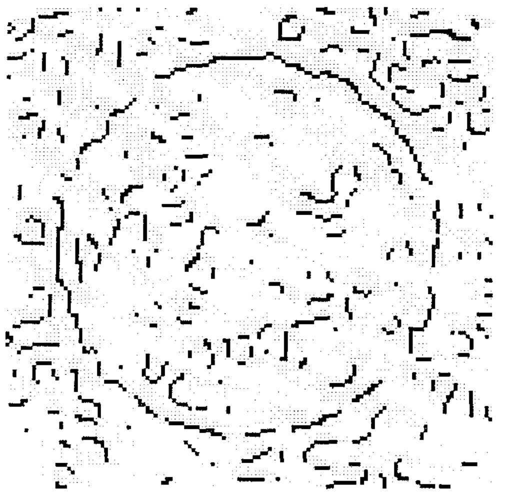


fig. 4d) *The thresholded result for the Marr–Hildreth operator with $\sigma = 3$.*

Conclusion.

The results clearly show that the set of errors introduced here, give a good overall view of the performance of edge detectors even if they differ considerably in their output. To express the performance of an edge detector in one figure, an extensive study in the comparative gravity of the errors is necessary. But even without this figure, the 4 partial performance measures give a greater insight in the particular faults of edge detection methods, and by evaluating the different types of errors a performance measure tuned to a certain application can be developed.

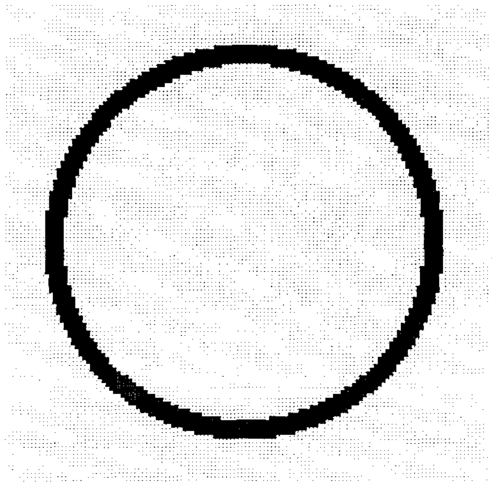


fig. 4c) The edge zone of width $w = 2$.

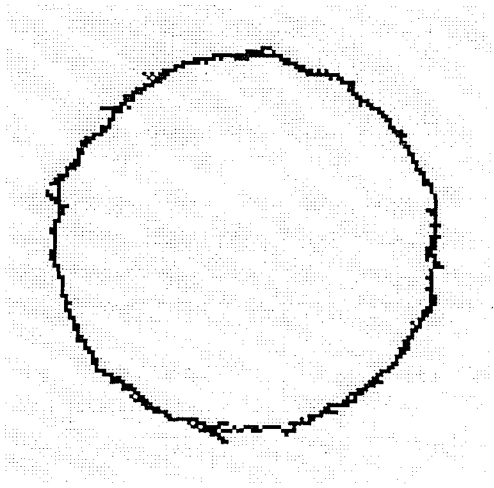


fig. 4e) *The thresholded result for the CODD operator with $\sigma = 3$.*

References.

1. D. Marr and E.C. Hildreth, "Theory of edge detection" Proc. Royal Soc. London B, vol 207, 1980, 187–217.
2. J.Besuijen, "Edge detection by combining directional derivatives", EUSIPCO–86, pp. 887–891, North–Holland, Amsterdam, 1986.
3. J.R. Fram and E.S. Deutsch, "On the quantitative evaluation of edge detection schemes and their comparison with performance", IEEE Trans. Comput. C–24, 1975, 616–627.
4. I.E. Abdou and W.K. Pratt, "Quantitative design and evaluation of enhancement/thresholding edge detectors", Proc. IEEE 67, 1979, 753–763.
5. T. Peli and D. Malah, "A Study of edge detection algorithms", CGIP–20, 1982, 1–21.