Evaluation of Edge Detectors Using Average Risk

L.J. Spreeuwers, F. van der Heijden

University of Twente, Department of Electrical Engineering
7500 AE Enschede, Netherlands

Abstract

A new method for evaluation of edge detectors, based on the average risk of a decision, is discussed. The average risk is a performance measure well-known in Bayesian decision theory. Since edge detection can be regarded as a compound decision making process, the performance of an edge detector is context dependent. Therefore, the application of average risk to edge detection is non-trivial.

This paper describes a method to estimate the probabilities on a number of different types of (context dependent) errors. A weighted sum of these estimated probabilities represents the average risk. The weight coefficients define the cost function. The method is suitable, not only for the comparison of edge operators, but also for the determination of the weaknesses and strengths of a certain edge operator. This is demonstrated with some well-known edge operators.

1. Introduction

Edge detection is an important process in image processing. Many different edge detectors are available, however it is not always clear which edge detector should be used for an application. To select the best edge detector for a certain task, an objective performance measure is required. In statistical detection and classification theory, the average risk of a decision is often used as a performance measure. The lower the average risk, the better the detector will be. The Bayes decision rule results in the lowest possible risk, and thus sets a lower limit for the average risk. However, this decision rule can only be calculated if the conditional probability distribution functions and the prior probabilities of the classes are exactly known. This is generally not the case.

The average risk of an (edge) detector can be calculated using the probabilities on the different types of errors and a cost (or loss) function, that determines how each error is weighted. The probabilities on different classification errors also enable us to find the weaknesses and strengths of edge detectors. Furthermore, they can be used in the design of edge detectors. If we know that a particular edge detector is sensitive for a certain type of error, we may be able to improve this edge detector.

The probability on a certain type of error depends on the characteristics of the scene. Therefore, the performance of an edge detector depends on the type of images it operates on. The application determines the cost function. As an example, the number of boundary junctions in an edge map depends on the scene, while the loss for missing a junction depends on the importance of junctions in further processing.

Except for [3] and [4], on which this work builds, existing edge detector evaluation methods [1-5] use a rather ad hoc combination of measures of different types of errors and generally do not provide measures for the separate errors. Often they can only be used with fixed (very simple) test images. Incidentally, most existing evaluation methods can be generated from the method proposed here by choosing the right cost function. This will be illustrated for the figure of merit (FOM) of Pratt.

The objective of this research has been to design an edge evaluation method, based on the average risk, that can be used to evaluate and design edge detectors for specific types of scenes. Apart from the average risk as a measure of performance, the evaluation method should provide estimations of the probabilities on different types of classification errors.

In this paper a description of the new evaluation method will be given, and the performance of a number of edge detectors will be evaluated.

2. Average risk of an edge detector

We will consider an edge detector as a decision rule that classifies pixels on basis of the grey levels of the pixels in a local neighbourhood, and assume that this decision rule is space invariant. An edge detector classifies pixels in an image as "edge" or "non-edge". Since we want to differentiate between the different types of errors that an edge detector can make, we take into account the following factors:

- Probability of missed and spurious edge elements.

0-8186-2920-7/92 $3.00 © 1992 IEEE

- Clustering of missed and spurious edge elements.
- Displacement errors of edge elements.
- Probability of thickened edge elements.

(This list can be extended if necessary, for instance, to incorporate the behaviour near boundary junctions.)

The proposed evaluation method labels the decisions made by an edge detector. In accordance with the aforementioned factors we define the following labels:

ξ_1: Isolated, missed edge pixel.

ξ_2: Clustered, missed edge pixel (i.e. a pixel belonging to a part of a missed edge segment).

ξ_3: Isolated, spuriously detected edge pixel (false alarm).

ξ_4: Clustered false alarm.

$\xi_{5,n}$: Displacement over n pixel periods.

$\xi_{6,n}$: n periods wide edge pixel (i.e. a pixel belonging to a thickened edge segment the width of which, measured at the pixel position (across the boundary direction), is n pixel periods).

(Incidentally, more than one label can be assigned to one single decision.)

The average risk of a classifier can be calculated using the probabilities on different types of errors, as described by, for instance, Devijver and Kittler [6].

Suppose the loss associated with a label ξ_i is equal to $\lambda(\xi_i|\omega_j)$, where ω_j is the true class (either edge or non-edge) of the corresponding pixel. The probability of a certain type of error equals the joint probability of the true class and a label $p(\xi_i,\omega_j)$. The average risk of the decision rule can be expressed in these probabilities:

$$L = \sum_j \sum_i \lambda(\xi_i|\omega_j)\, p(\xi_i,\omega_j)$$

with the summation of i over all labels, and the summation of j over all true classes.

3. Estimation of error probabilities from images

The probabilities on different types of errors can be estimated from images that are processed by an edge detector, provided that we have the ideal edge map (edge reference map) at our disposal.

We assume that two binary images are available: the edge reference map with the exact positions of the edges, and the detected edge map, i.e. the result of an edge detector. The probabilities on different types of errors can be obtained by assigning the labels to the pixels in the detected edge map, counting the pixels with a certain label and dividing this count by the total number of pixels. Thus, the problem of finding the probabilities on errors is reduced to finding and counting the pixels that are misclassified in a certain way. A scheme of the evaluation process is given in figure 1.

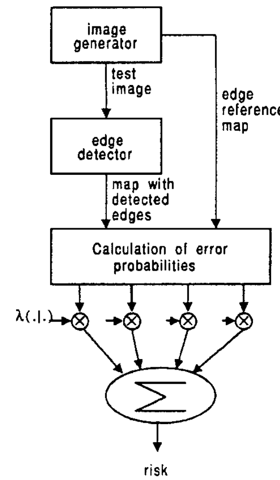

Figure 1.

A method to label the detected edges is presented in the next paragraphs. The labelling can be accomplished using distance transforms and binary logic operations applied to the two maps [7]. It consists of 5 stages:

1. Finding correct edges and false alarms.
2. Finding displaced edges.
3. Determining edge thickness.
4. Finding missed edge pixels.
5. Subdivision of false alarms and missed edge pixels in isolated and clustered false alarms and missed edge pixels.

3.1 Correct edges and false alarms

In the first stage, the correctly classified edges are found by looking in the neighbourhood of the edges in the edge reference map. All detected edges within a certain distance of the edges in the edge reference map are considered to be correct edges. Furthermore, all pixels connected to those pixels in the map with detected edges are considered correct. The false alarms are the remaining pixels in the detected edge map. All correctly classified edges are called the correct detected edge map from now on.

3.2 Displaced edges

Displacement of edges is found by measuring the distance of each edge pixel in the correct detected edge map to the closest edge pixel in the edge reference map. If the correct detected edge map contains edges more than 1 pixel thick the edges must be thinned first.

3.3 Edge thickness

Edge thickness is the width measured in pixel periods of the intersection of the edge segment perpendicular to this segment. However, it is not easy to assign the thickness to edge pixels this way, and thus to estimate the probability on thick edges by counting pixels. A simple approximation of the above can be found by scanning the correct detected edge map horizontally, vertically and diagonally (upper left to bottom right and upper right to bottom left) and determining the widths of the intersection of the edges. The thickness of the edge at the position of a pixel is the minimum of the four intersections. In this way each detected edge pixel can be assigned a thickness.

3.4 Missed edge pixels

The method to find the missed edges is based on filling in the holes in the correct detected edge map using the edge reference map. Once the correct detected edge map is completed, the missed edges can be found as the difference of the completed edge map and the correct detected edge map. The basic idea is to start with the correct detected edge map and add all edges using the edge reference map, to make it look best like the edge reference map.

3.5 Isolated and clustered

Isolated false alarms and missed edge pixels can easily be found, because they have no adjacent false alarms nor missed edge pixels. Clustered false alarms or missed edge pixels are those that are not isolated.

4. Comparison to Pratts figure of merit

The performance measure published by Pratt [1] is the best known and most referenced evaluation method. Pratts figure of merit (fom) is a function of the displacement and thickness errors, and uses a fixed test image. Since the method proposed in this paper generates estimations of the probabilities on displacement errors, it should be possible to obtain Pratts fom by choosing the right cost function.

Pratts fom is given by the formula:

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

where I_I and I_A are the ideal and the actually detected number of edge pixels, and $d(i)$ is the distance from the detected edge pixel to the closest real edge pixel. Pratts fom can be rewritten in terms of the average risk, if the following cost function is chosen:

$$\lambda(displacement = d | edge) = \frac{P(edge)}{1+\alpha d^2}$$

where $P(edge)$ is the prior probability of an edge in the test image. Actually, this is not a cost function, but a credit function, since the "cost" decreases for larger displacements. The costs of all other types of errors are to be chosen 0.

5. Experiments and results

The AVR (average risk) evaluation method has been applied to a number of edge detection operators under various conditions. The selected edge operators are [1]: the Sobel algorithm, the Marr-Hildreth operator, and the Canny operator. In the next paragraph, the selected test image and cost functions are discussed. Subsequently, the AVR of the operators are presented as a function of SNR (signal-to-noise ratio). Finally, it will be demonstrated that the components of the AVR of a certain edge operator can be used to find the weaknesses and strengths of that edge operator.

5.1 Test image and cost functions

The desirable properties of test images (e.g. uniformly distributed edge segments over the image plane) have been discussed in [8]. Test images based on Voronoi tessellations have many of the desired properties. Figure 2 shows the used grey level test image and the corresponding edge reference map. In the experiments, the test image was distorted by additive, Gaussian white noise. The SNR of the image is defined as the ratio between the variance of the difference in grey levels of neighbouring segments and the variance of the noise.

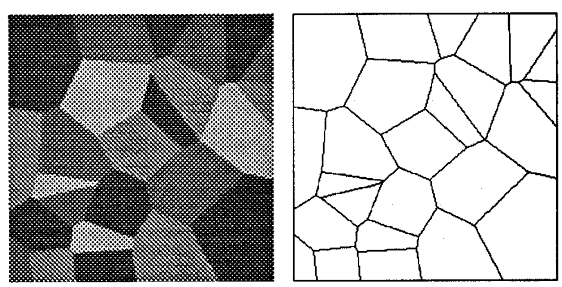

Figure 2. Voronoi test image and edge reference map.

In the experiments the following cost function was chosen:

label	ξ_1	ξ_2	ξ_3	ξ_4	$\xi_{5,n}$	$\xi_{6,n}$
$\lambda(.\mathrm{l}.)$	$\frac{1}{2}$	1	$\frac{1}{2}$	1	$\frac{n}{8}$	$\frac{n-1}{8}$

These costs are reasonably in accordance with general requirements.

5.2 Performance of edge operators

Figure 3 shows the average risk measurements of the tested edge operators as a function of SNR of the test image. Actually, the Marr-Hildreth operator marks all zero crossings of the LOG-filtered test image as edge elements. This produces many spurious edge segments. To circumvent this, in addition, the magnitude of the gradient of the Gaussian-filtered test image is thresholded. The filters in the Canny operator are approximated by Gaussian filters. For both operators two widths of the Gaussian are selected: $\sigma=1$ [pixel period] and $\sigma=2$. In all measurements the thresholds of the operators are selected to give optimal results, i.e. minimal AVR. As expected, the curves are monotonically decreasing. Small values of σ are preferable if SNR is high, and vice versa. On the whole, the Canny operator performs best. For an explanation, see [8].

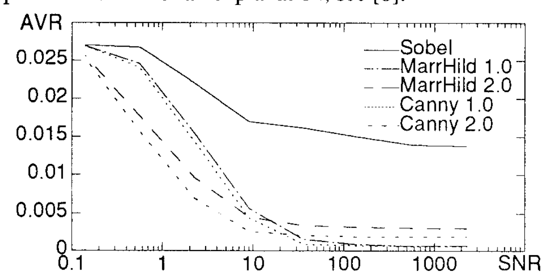

Figure 3. AVR versus SNR for some operators.

The second experiment demonstrates that the AVR measure is appropriate to analyse the behaviour of an edge detector. For that purpose we selected the Canny operator, and a test image with SNR=4. Figure 4 shows AVRs of the operator as a function of the operator width σ. The AVRs differ in the choice of the cost function λ (.l.). The first cost function of the first AVR (solid line), given in the aforementioned table, assigns costs to both displacement errors and detection errors (missed edges, false alarms). This AVR is decomposed into two components: cost for displacements, and costs for detection errors. It is clear that the displacement error increases with the operator width. At the same time, the detection errors decrease. Hence, we have shown that the uncertainty principle, formulated by Canny, holds.

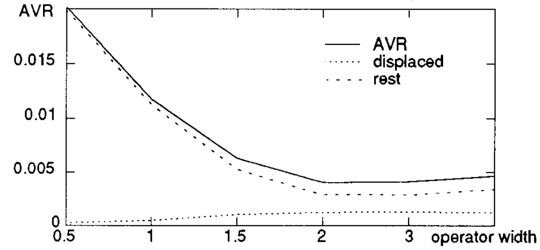

Figure 4. AVRs versus operator width.

6. Conclusions

The presented evaluation method enables us to express the performance of an edge detector in the average risk of a decision it makes. Unlike the expressions other evaluation methods use, the average risk is a well-defined quantity. Furthermore, the presented method provides estimations of the probabilities on different types of errors: currently, isolated and clustered missed edges and false alarms and edge thickness and displacements; this can easily be extended to incorporate other types of errors. The method can work with arbitrary complex test images. A paper on test image generation for more complex scenes is in preparation.

References

[1] Pratt, W.K., Digital Image Processing. Second edition, J.Wiley & Sons, New York 1991.

[2] Fram, J.R., Deutsch, E.S., On the quantitative evaluation of edge detection schemes and their comparison with human performance, IEEE Trans. on Computers vol.c-24, no.6, pp.616-628, 1975

[3] Besuijen, J., Heijden, F. van der, An edge detection performance measure incorporating structural errors, Proc. of the Douzieme colloque gretsi, Juan-les-Pins, June 12-16, 1989

[4] Heijden, F. van der, Quality of edge detectors, Proc. CSN89, Sion, Utrecht, Netherlands, Nov. 1989

[5] Peli, T., Malah, M., A study of edge detection algorithms, Computer graphics and image processing, vol. 20, pp.1-21, 1982

[6] Devijver, P.A., Kittler, J. Pattern recognition: a statistical approach, Prentice Hall, Englewood Cliffs, N.J., 1982

[7] Spreeuwers, L.J., Heijden, F. van der, An Edge Detector Evaluation Method Based on Average Risk. In: Robust Computer Vision (Eds: Forstner/Ruwiedel), Wichmann, Karlsruhe: 1992.

[8] Heijden, F. van der, A Statistical Approach to Edge and Line Detection in Digital Images. Ph.D. Thesis, University of Twente, Enschede 1992.