On the optimal edge detector

M. Petrou
Informatics Department, Rutherford Appleton Laboratory,
Chilton, Didcot, Oxon. OX11 OQX

J. Kittler
Electronic & Electrical Engineering Department
University of Surrey, Guildford GU2 5XH

Advancing further the theory of the optimal edge detector, as has been developed by Canny and Spacek, we derive the 'definitive' optimal edge operator. We show that the cubic spline approximation is, in practice, as good as the optimal edge detector.

1. INTRODUCTION

Canny [1] was the first to set the foundations of the theory of an optimal edge detector: Good signal to noise ratio, good locality and maximum suppression of false responses. He derived quantitative measures for these three qualities and combined the first two of them to form a measure of performance for an edge detector. He then maximized that measure under the additional constraint of maximum suppression of false responses. The equations he arrived at were long and complicated. Eventually, he proposed the derivative of a Gaussian as the best approximation to the optimal operator derived by the method described above. This operator is simple in form and its performance measure is 80% that of the optimal operator.

Spacek [4] picked up the threads from where Canny had left them and formed a performance measure combining all three quantitative measures Canny had derived. In doing so, he simplified the differential equation, the solution of which gives the form of the optimal filter. As a result, his optimal filter is different from Canny's.

Apart from the different form of the total performance measure used, the works of Canny and Spacek appear at first sight to differ in some other respect too. Spacek, right from the beginning, set the boundary conditions which his filter must satisfy: It must be antisymmetric ($g(0)=0$) and smoothly going to zero at its *finite* limits at $\pm w (g(\pm w) = 0, g'(\pm w) = 0)$, with given maximum amplitude ($g(x_m) = k$ where $g'(x_m) = 0$). Canny, on the other hand, had developed his theory assuming filters of infinite extent. In practice, however, he put finite limits to his integrals and imposed exactly the same boundary conditions as Spacek, although at the end he approximated his filter with one which does not go smoothly to 0 at $\pm w$.

The final form of the filter equation Spacek derived depended on six parameters, the numerical values of which had to be chosen so that the performance was optimal. In order to simplify the process, Spacek fixed two of those parameters and determined the remaining four from the boundary conditions.

In the work presented here, we extend Spacek's work by maximizing the performance measure with respect to the extra two parameters. We thus derive the "definitive optimal" filter and critically compare it with the non-optimal ones. In section 2 we present the optimization process, in section 3 we discuss implementation details and in section 4 we present our conclusions.

2. THE MAXIMIZATION PROCESS

Spacek, using only the boundary conditions, and before any maximization process, derived a first approximation to his filter in the form of a cubic spline.

$$cs(x) = x^3 + 2x^2 + x$$

We shall refer to this filter later on and use it as a reference point for checking the performance of the various optimal filters.

By assuming an ideal step edge corrupted by white Gaussian noise, one can derive the quantitative measures of the characteristics of a filter as follows:

Measure for the signal to noise ratio:

$$S = \frac{\int_{-w}^{0} g(x)\, dx}{\left[2\int_{-w}^{0} g^2(x)\, dx\right]^{1/2}} \qquad (1)$$

AVC 1988 doi:10.5244/C.2.30

Good locality measure inversely proportional to the standard deviation of the distribution of points where the edge is supposed to be:

$$L = \frac{| g'(0) |}{[2 \int_{-w}^{0} g'^2(x) dx]^{1/2}} \qquad (2)$$

Measure for the maximum suppression of false edges proportional to the distance between the neighboring maxima of the response of the filter to white noise:

$$C = \left[\frac{\int_{-w}^{0} g'^2(x)\, dx}{\int_{-w}^{0} g''^2(x)\, dx} \right]^{\frac{1}{2}} \qquad (3)$$

We define the total performance measure as:

$$P = (2\, S\, L\, C)^2 = \frac{\left[\int_{-w}^{0} g(x)\, dx \right]^{\frac{1}{2}} g'^2\,(0)}{\int_{-w}^{0} g^2(x)\, dx \int_{-w}^{0} g''^2(x)\, dx} \qquad (4)$$

Our purpose is to choose g(x) so that this quantity is maximum. To do this is enough to extremize any one of the integrals appearing in the above quantity, assuming that the remaining integrals are constant [2]. We choose to minimize $\int g^2(x)dx$ assuming that

$$\int_{-w}^{0} g(x)\, dx = c_1 \;\; and \;\; \int_{-w}^{0} g''^2(x)\, dx = c_2 \qquad (5)$$

where c_1 and c_2 are some constants. Using the method of Lagrange multipliers we define the function $Z(g, g', g'')$ as:

$$Z\,(g, g', g'') = g^2(x) + \mu_1 g''^2(x) + \mu_2 g\,(x) \qquad (6)$$

where μ_1 and μ_2 are the Lagrange multipliers. This function must satisfy Euler's equation:

$$Z_g - \frac{dZ_{g'}}{dx} + \frac{d^2 Z_{g''}}{dx^2} = 0 \qquad (7)$$

By substitution, we derive the differential equation for the optimal filter :

$$2\, g\,(x) + \mu_2 + 2\, \mu_1\, g''''(x) = 0 \qquad (8)$$

So far, we have followed exactly Spacek's steps in the derivation of this equation. From now on our process differs slightly so that all possible solutions can be found in the most general form. We assume that the Lagrange multipliers which appear in this equation are complex. The general solution of the above differential equation is then:

$$\begin{aligned} g\,(x) = {} & A_1 exp(\beta x)[\cos(\alpha x)+i\sin(\alpha x)] + \\ & A_2 exp(-\alpha x)[\cos(\beta x)+i\sin(\beta x)] + \\ & A_3 exp(-\beta x)[\cos(\alpha x)-i\sin(\alpha x)] + \\ & A_4 exp(\alpha x)[\cos(\beta x)-i\sin(\beta x)] + A_5 \end{aligned} \qquad (9)$$

where α, β are real and A_1, A_2, A_3, A_4, A_5 are complex.

This is a complex filter which depends on twelve real parameters. We shall choose some of these parameters so that the imaginary part of the filter vanishes identically. If we use superscripts R and I to indicate the real and imaginary parts of a quantity, we obtain:

$$\begin{aligned} g^{R}(x) = {} & exp(\beta x)[A_1{}^{R}\cos(\alpha x)-A_1{}^{I}\sin(\alpha x)] + \\ & exp(-\alpha x)[A_2{}^{R}\cos(\beta x)-A_2{}^{I}\sin(\beta x)] + \\ & exp(-\beta x)[A_3{}^{R}\cos(\alpha x)+A_3{}^{I}\sin(\alpha x)] + \\ & exp(\alpha x)[A_4{}^{R}\cos(\beta x)+A_4{}^{I}\sin(\beta x)] + A_5{}^{R} \end{aligned} \qquad (10)$$

$$\begin{aligned} g^{I}(x) = {} & exp(\beta x)[A_1{}^{I}\cos(\alpha x)+A_1{}^{R}\sin(\alpha x)] + \\ & exp(-\alpha x)[A_2{}^{I}\cos(\beta x)+A_2{}^{R}\sin(\beta x)] + \\ & exp(-\beta x)[A_3{}^{I}\cos(\alpha x)-A_3{}^{R}\sin(\alpha x)] + \\ & exp(\alpha x)[A_4{}^{I}\cos(\beta x)-A_4{}^{R}\sin(\beta x)] + A_5{}^{I} \end{aligned} \qquad (11)$$

The trivial solution of the equation $g^{I}(x) = 0$ is the vanishing of all its coefficients. This leads to a difference of boxes operator, which, however, does not satisfy the boundary conditions we imposed at $\pm w$. So, this solution is not acceptable. There are only two other possible solutions:

i) $\alpha = 0$ (or $\beta = 0$) which leads to:

$$A_2{}^{I} + A_4{}^{I} = 0\,,\; A_2{}^{R} - A_4{}^{R} = 0\,,$$
$$A_1{}^{I} = 0\,,\; A_3{}^{I} = 0\,,\; A_5{}^{I} = 0$$

Using these in (10) we obtain the solution which Spacek derived for $\mu_1 < 0$:

$$g\,(x) = K_1 exp(Ax) + K_2 exp(-Ax) + K_3 cos(Ax) + K_4 sin(Ax) + K_5 \qquad (12)$$

ii) $\alpha = \beta$ (or $\alpha = -\beta$) which leads to:

$$A_1{}^{I} = -A_4{}^{I}\,,\; A_1{}^{R} = A_4{}^{R}\,,\; A_2{}^{I} = -A_3{}^{I}\,,$$
$$A_2{}^{I} = A_3{}^{I}\,,\; A_5{}^{I} = 0$$

Using these conditions in (10) we obtain the solution which Spacek derived for $\mu_1 > 0$:

$$g(x) = exp(Ax)\,[K_1 sin(Ax) + K_2 cos(Ax)] + \quad (13)$$
$$exp(-Ax)\,[K_3 sin(Ax) + K_4 cos(Ax)] + K_5$$

The parameters K_1, K_2, K_3, K_4, K_5 and A are *real* quantities to be determined by the boundary conditions and the constraints (5).These are six equations, enough to determine the six unknowns. However, the constants c_1 and c_2 which appear in equations (5) are unknown themselves.

Spacek fixed A and K_5 to be 1/w and k respectively. He then determined the remaining constants from the boundary conditions. In our approach, we use the boundary conditions to express four of the parameters in terms of the other two, then we substitute $g(x;K_3,K_4)$ in the expression for P and choose the values for K_3 and K_4 so that P takes its maximum value. One glance at the formulae involved is enough to convince anybody that this task is almost humanly impossible! However, it is easy with the use of algebraic computing. The system we used was **REDUCE.** The boundary conditions at 0 and $\pm w$ were solved in terms of K_1, K_2 and K_5 as functions of A, K_3 and K_4. Subsequently the system

$$g(x_m, A; K_3, K_4) = k$$
$$g'(x_m, A; K_3, K_4) = 0 \quad (14)$$

was considered. By solving it, one can formally obtain $A(K_3,K_4)$. Then the values of K_3 and K_4 can be obtained as solutions of the system:

$$\frac{\partial P}{\partial K_3} = 0, \quad \frac{\partial P}{\partial K_4} = 0 \quad (15)$$

Since (14) cannot be solved analytically, one must also compute numerically the quantities $\frac{\partial A}{\partial K_3}, \frac{\partial A}{\partial K_4}$ which appear in (15).

The above formal way of calculating K_3 and K_4, however, is not appropriate in this case: $P(K_3,K_4)$ is not a smooth function of its arguments with one well defined maximum. Instead, it has lots of local maxima and if one guesses an initial solution of (15) and tries to iterate to the true solution, most likely one will arrive at a local maximum without any means of telling whether this is also a global maximum.

To overcome this problem, we calculated $P(K_3,K_4)$ at the knots of a dense grid covering the K_3, K_4 space and by inspection we chose those values of K_3, K_4 for which P was maximum. We followed this process for both solutions (12) and (13). The optimal filter is given by (13) with parameter values:

$A = 1.44609$
$K_1 = -5.22049$
$K_2 = -0.86171$
$K_3 = 0.6$
$K_4 = -1.0$
$K_5 = 1.86171$

These values were calculated for w = 1.0, k = 1.0 and they have to be appropriately rescaled for different values of the half-width and the amplitude of the filter.

Table 1 gives the values of the partial and the total performance measures for the optimal and suboptimal operators (given by equations (13) and (12) respectively) as well as the Spacek and the cubic spline operators.

Table 1. The performance measures for the various filters.

	Optimal	Suboptimal	Spacek	Cubic spline
S	0.6059	0.6052	0.6040	0.6038
L	1.8781	1.8856	1.9337	1.9365
C	0.1879	0.1873	0.1828	0.1826
P	0.1829	0.1828	0.1824	0.1823

We notice that Spacek's operator is better than the spline operator by 0.0001 while the optimal operator is better than Spacek's by 0.0005! At first sight this seems a significant improvement (fivefold!) over Spacek's result. In the next section we are going to examine how important this improvement is in practice.

3. IMPLEMENTATION AND RESULTS

The derived filter is antisymmetric and thus directional. So it cannot be generalized to two dimensions. Spacek proposed a clever way of overcoming this problem: He integrated the filter to obtain a symmetric smoothing operator which can trivially be generalized to two dimensions. After the image has been smoothed by this filter, the edges can be found as the maxima of a simple difference operator which approximates the first derivative of the intensity. Following the same method we derived the optimal smoothing filter:

$$h(x) = w\, exp(Ax/w)[L_1\, sin(Ax/w) + L_2\, cos(Ax/w)] +$$
$$w\, exp(-Ax/w)[L_3\, sin(Ax/w) + L_4\, cos(Ax/w)] +$$
$$L_5 x + L_6 \qquad for\ -w \le x \le 0 \quad (16)$$
$$h(x) = h(-x) \qquad for\ x \ge 0$$

where

$$L_1 = -3.70549$$
$$L_2 = 2.65552$$
$$L_3 = -0.97478$$
$$L_4 = 0.24369$$
$$L_5 = 3.28037$$
$$L_6 = -1.89921$$

The constant of integration has been chosen so that $h(\pm w) = 0$ and the coefficients have been scaled so that h(0)=1.

All the development so far has been based on continuous functions. The derived filter now has to be discretized by sampling. To make sure that the filter does not lose its optimal qualities by sampling, we first calculate its Fourier coefficients, assuming that it is a periodic function of period 2w. We obtain:

$$h(x) = 0.405 + 0.495\cos(\pi x/w) + 0.0875\cos(2\pi x/w) + 0.00457\cos(3\pi x/w) + 0.00553\cos(4\pi x/w) + 0.000567\cos(5\pi x/w) + 0.00109\cos(6\pi x/w) + 0.000146\cos(7\pi x/w) + \ldots \quad (17)$$

Since h(x) is support limited, its spectrum is expected, in general, to contain all the harmonic frequencies. We notice, however, that from the 5th harmonic onwards the Fourier coefficients become insignificant. We can then assume that our filter is band limited with a cutoff frequency $\Omega_f = \frac{5\pi}{w}$. In fact, if we keep only the first six terms in the Fourier expansion we can reconstruct the filter so that its value at 0 is 0.9982 instead of 1 and at w is -0.0021 instead of 0.

According to the sampling theorem then, the filter will retain all its properties if it is sampled with frequency $\frac{2\pi}{T} \geq 2\Omega_f$ where T is the frequency of the sampling function. Given that we sample the filter at every pixel this inequality is satisfied if w (the half width of the filter) is at least 5 pixels wide. So, these filters are rather large, at least 10 by 10 pixels.

We applied the various filters to a real remotely sensed image in order to compare their performance. Remotely sensed images are notoriously difficult and noisy. The reason we chose such an image was because of the availability of the 'perfect segmentation' against which we could compare our edge detection. A human photointerpreter had drawn by hand the meaningful edges of the image: field boundaries, drains, rivers, roads, lake outlines, forest outlines, etc.

We defined two measures of performance for the various edge detectors: The fraction of all correctly detected edge pixels i.e. those which had at least one 'correct' edge pixel of the hand segmentation within their 3 x 3 neighborhood (CD), and the fraction of the 'correct' pixels which were not detected at all, i.e. they did not have any edge pixel detected by the edge detector within their 3 x 3 neighborhood (UD). Of course, one cannot expect very high performance values for any filter since the human photointerpreter does not simply use an edge detector, but higher level knowledge as well which enables him to ignore texture and noise edges no matter how strong they are.

Table 2 summarises the results of the application of a Gaussian smoothing filter (i.e. the one proposed by Canny), the Spacek, the optimal and the spline operators, to the same 256 x 256 image. The variance of the Gaussian was chosen so that the value of the mask at w was 0.001. All filters were 11 x 11 pixels in size and the same threshold was used for all of them. We notice from this table that the optimal filter has superior performance over the other filters, as was expected. However, given that the 'correct' edge pixels in the image were about 6000, the better performance does not seem to be statistically significant.

Table 2. Results of the appplications of the filters to a remotely sensed image.

Filter	CD	UD
Gaussian	0.4216	0.3868
Spacek	0.4249	0.4296
optimal	0.4259	0.4314
cubic spline	0.4249	0.4296

One may argue that such an image is rather inappropriate for testing the performance of step edge detectors. Indeed, our image contained a high percentage of delta function edges (lines), on which all the above filters are expected to perform poorly, and the rest of the edges were ramps rather than steps. Of course, that is how real images are, but, to be fair to the filters we gave them another chance by applying them to a synthetic image containing only step edges with additive Gaussian noise. Fig. 1a shows this image, which we call 'the stamp'. The intensity of the image jumps by 20 units in each step and the standard deviation of the noise is 20 units as well, so that our image has '100% noise'. The perforations are chosen to have smaller dimension than the dimension of the filter (they are 8x8, while the filter used was 13 x 13) so that we can check how well they are preserved. Fig. 1b shows the 'perfect edge detection' against which we shall compare the results of our filters.

Figs. 2a,b,c,d show the results of applying the Gaussian, the spline, the Spacek and the optimal filters respectively. The same algorithm and the same threshold was applied to all filters. It is obvious that the performance of the last three filters is indistinguishable. This is reflected in the quantitative performance measures as well, shown in Table 3.

Table 3. Results of the application of the filters to the synthetic image.

Filter	CD	UD
Gaussian	0.8227	0.2112
Spacek	0.8603	0.2367
optimal	0.8638	0.2394
cubic spline	0.8599	0.2370

One thing which is common in the results of the filters on the two images is that the Gaussian filter performs worse with respect to the CD measure, but better with respect to the UD measure, in comparison to the other filters, i.e. the Gaussian filter leaves undetected far fewer pixels. The reason is that the Gaussian filter is 'slimmer' than the others in the sense that the weights it attaches to the peripheral pixels are smaller than the weights given to the same pixels by the other filters. As a result the Gaussian filter smooths less than the other filters, and several true but weak edge pixels are retained alongside lots of noise pixels. This interpretation is shown schematically in Fig. 3 with the help of a Venn diagram.

4. CONCLUSIONS

We applied an elaborate and careful optimisation process extending the work of Canny and Spacek and derived the optimum edge detector. Although the peformance criterion shows that this operator must be better than the others, the improvement is only marginal. If one compares the results of the simple cubic spline operator with those of Spacek's and the optimal one, one sees very slight difference in the performance. If we consider only the signal to noise ratio, we conclude that if the optimal filter can cope with an image which has been corrupted with white Gaussian noise of variance σ_0, the cubic spline filter will perform equally well on an image which has been corrupted with Gaussian noise of variance $1.007\sigma_0$! One must remember that the cubic spline filter was not determined by any optimization process, but only by the boundary conditions and the general shape requirements we imposed on our filter. It seems then, that the performance of a filter is largely determined by its gross shape and the boundary conditions it satisfies and to a lesser extent by its detailed shape. We conclude, therefore,that the cubic spline operator is almost as good as one can hope to obtain.

The future of edge detection must lie in the algorithm development rather than the mere convolution of an image with a suitably chosen filter. Examples of such algorithms are those developed by Canny [1] (Canny algorithm as opposed to just Canny filter), Fleck [3] and Wilson & Span [5].

Acknowledgments

We are grateful to the NERC Unit of Thematic Information Systems at Reading University, and in particular to David Mason and Alan Cross for kindly providing the remotely sensed images and their hand segmentations.

We also wish to thank A. Conway for carefully typesetting this paper.

This work was supported by an Alvey grant (project number MMI26).

REFERENCES

1. Canny, J. F. 'Finding edges and lines in images', MIT AI Lab.Tech. Report 720 (1983).
2. Courant, R. & Hilbert, D. 'Methods of mathematical physics'. Vol. 1, (1953) Wiley Interscience, New York, USA.
3. Fleck, M. M. 'Representing space for practical reasoning', in the'Proceedings of the third Alvey Vision Conference', (1987) p. 275.
4. Spacek, L. A. 'Edge detection and motion detection'. Image and Vision Computing, **4**, (1986) p. 43.
5. Wilson, R. & Spann, M. 'Image segmentation and uncertainty', Research Studies Press ltd., (1987) John Wiley & Sons Inc.

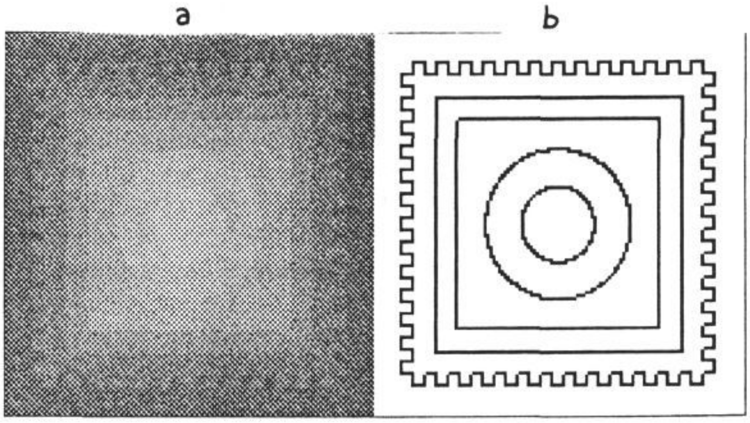

Fig. 1

a) the "stamp" with 100% noise.

b) The "perfect edges" for the "stamp".

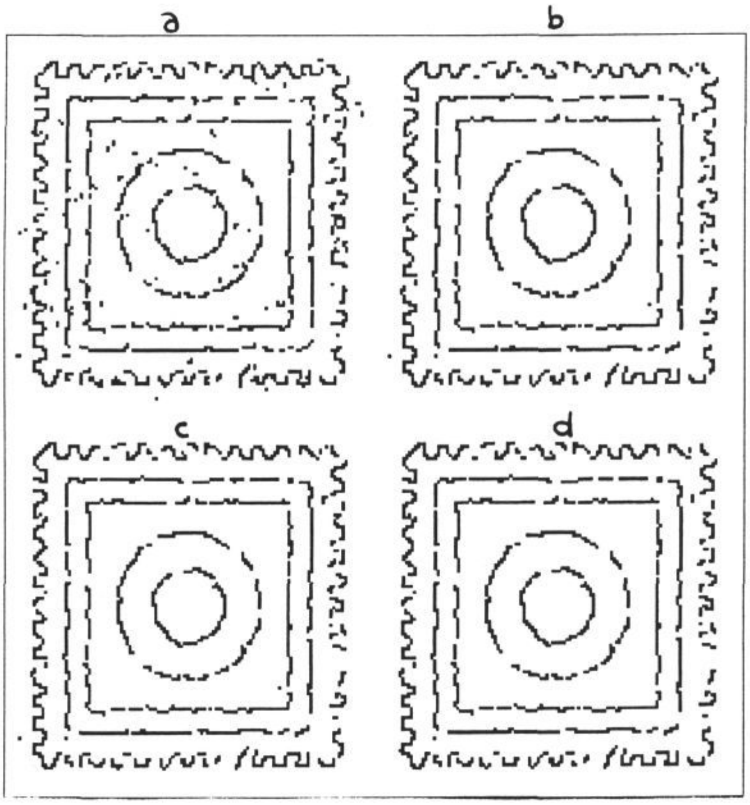

Fig. 2

a) Output of the Gaussian filter

b) Output of the "spline" filter

c) Output of the Spacek filter

d) Output of the optimal filter.

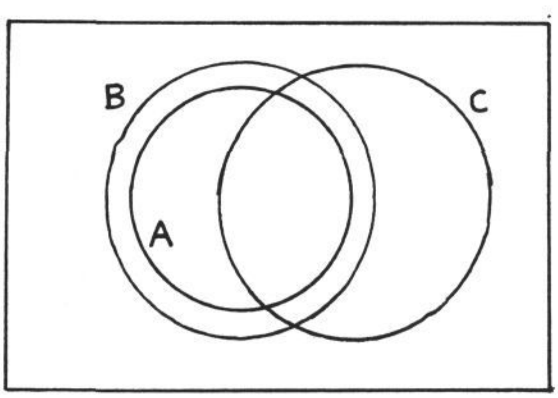

Fig. 3

A: set of edge pixels detected by the optimal filter

B: set of edge pixels detected by the Gaussian filter ($A \subset B$)

C: set of correct edge pixels.