Optimal Edge Detectors Revisited.

William McIlhagga

Bradford School of Optometry and Vision Science, University of Bradford, Bradford, England BD7 1DP.

tel. 00441274235957

fax. 00441274235570

mailto: w.h.mcilhagga@bradford.ac.uk

homepage: http://www.brad.ac.uk/acad/lifesci/optometry/index.php/Staff/DrWilliamMcIlhagga

Abstract.

Canny (1986) suggested that an optimal edge detector should maximize both signal-to-noise ratio and localization, and he derived mathematical expressions for these criteria. Based on these criteria, he claimed that the optimal step edge detector was similar to a derivative of a gaussian. However, Canny's work suffers from two problems. First, his derivation of localization criterion is incorrect. Here we provide a more acurate localization criterion and derive the optimal detector from it. Second, and more seriously, the Canny criteria yield an infinitely wide optimal edge detector. The width of the optimal detector can however be limited by considering the effect of the neighbouring edges in the image. If we do so, we find that the optimal step edge detector, according to the Canny criteria, is the derivative of an ISEF filter, proposed by Shen and Castan (1992).

In addition, if we also consider detecting blurred (or non-sharp) gaussian edges of different widths, we find that the optimal blurred-edge detector is the above optimal step edge detector convolved with a gaussian. This implies that edge detection must be performed at multiple scales to cover all the blur widths in the image. We derive a simple scale selection procedure for edge detection, and demonstrate it in one and two dimensions.

Note: This is a slightly improved version of the published article. A number of minor errors and clumsy expressions crept into that paper, as a result of the limited time given to authors by Springer (although they apparently give unlimited time to the reviewers).

1. Introduction

Edges are projections of physical processes, such as changes of reflectance at object boundaries, or a changes of illumination. They are informative cues to the three dimensional structure of the world and, because of this, edge detection is a vital first step in many vision systems. Many edge detectors have been developed (see Peli and Malah (1982), Ziou and Tabbone (1998)), often from informal or ad hoc arguments. Canny (1986), in an influential paper, suggested that edge detectors should optimize two specific performance criteria. First, the edge detector should have a good signal-to-noise ratio, so that edges can be detected even when image quality is poor. Second, edge detectors should accurately localize the edges, to support subsequent visual processes that need a high degree of positional accuracy. Canny suggested that the optimal edge detector maximizes the product of signal-to-noise and localization. He also found it necessary to constrain the smoothness of the edge detector. The resulting constrained optimal filter was similar to a derivative of a gaussian.

Unfortunately, Canny's development of optimal edge detectors contains two significant problems. The first, lesser, problem is that Canny did not provide the correct expression for localization (Tagare and deFigueiredo, 1990, Koplowitz and Greco, 1994). Here we provide a more accurate expression for edge localization. This new measure includes the filter smoothness, which explains why Canny found it necessary to constrain this. The second, bigger, problem is that the Canny criteria imply that the optimal edge detector is infinitely wide. This renders the edge detector useless, because it will pick up an infinite amount of interference from other edges in the image. To solve this problem, we need to include the effects of the other edges on filter performance. This can be done by modelling the other edges in the image as a Brown noise stochastic process.

When we do this, we find that the optimal edge detector is a derivative of an ISEF filter (Shen and Castan, 1992). In addition to providing the correct optimal edge detector for step edges, we also generalize the edge detection task to include detecting blurred edges of any width. When the noise in the image is low, the resultant algorithm is similar to a scheme proposed by Lindeberg (1998), but its optimality properties were not previously known. This optimal edge detection scheme is demonstrated in two dimensions.

2. Canny's Optimal Edge Detector.

We begin by summarising Canny's (1986) approach to edge detection. Consider a linear filter $f(x)$ designed to detect an isolated step edge $g(x)$, located at x=0, in white noise $n(x)$. The filter response $h(x)$ is given by

$$\begin{aligned} h(x) &= \int_{-r}^{+r} g(x-t)\, f(t)\, dt + \int_{-r}^{+r} n(x-t)\, f(t)\, dt \\ &= h_g(x) + h_n(x) \end{aligned} \qquad (1)$$

where $h_g(x)$ and $h_n(x)$ are the filter responses to edge and noise respectively. Edges are marked by peaks in the filter response $h(x)$. The edge is detected by a peak in the filter

reponse, and in the absence of noise, $h_g(0)$ is the sole peak. The filter is zero outside the interval [-*r*, *r*].

The signal to noise ratio of the filter is its response to the edge $h_g(0)$ divided by the r.m.s. response to noise $E[h_n(0)^2]^{1/2}$, which is

$$SNR(f) = \int_{-r}^{+r} g(-x) f(x)\, dx \Big/ n_0 \sqrt{\int_{-r}^{+r} f(x)^2\, dx} \qquad (2)$$

where n_0 is the r.m.s. amplitude of the noise.

In noise, the peak response $h_g(x)$ will occur at some point $x_{\max}$, different from zero. Ideally, however, the peak location should be close to 0. Canny defined the localization of the edge detector to be the reciprocal of the standard deviation of $x_{\max}$ about zero, namely $E[x_{\max}^2]^{-1/2}$. Near the true edge location, the filter response can be approximated by a Taylor expansion:

$$\begin{aligned} h(x) &\approx h(0) + xh'(0) + x^2 h''(0)/2 \\ &= h(0) + xh_g'(0) + x^2 h_g''(0)/2 + xh_n'(0) + x^2 h_n''(0)/2 \end{aligned} \qquad (3)$$

Since $h_g(0)$ is a maximum, the derivative $h_g'(0)$ is zero. Taking the derivative of Equation (3) with respect to *x* and substituting x_{max} gives

$$h'(x_{\max}) = h_n'(0) + x_{\max}(h_g''(0) + h_n''(0)) \qquad (4)$$

Since $h(x_{\max})$ is a maximum, the derivative $h'(x_{\max})$ is zero. Solving Equation (4) for x_{max} gives

$$x_{\max} = \frac{-h_n'(0)}{h_g''(0) + h_n''(0)} \qquad (5)$$

where both $h_n'(0)$ and $h_n''(0)$ are uncorrelated zero-mean gaussian random variables when the noise is gaussian (Koplowitz and Greco, 1994).

Canny assumed $h_n''(0) = 0$[1], so Equation (5) simplifies to $E[x_{\max}^2] = E[h_n'(0)^2] / h_g''(0)^2$, and the localization is then

$$L_C(f) \approx \frac{|h_g''(0)|}{E[h_n'(0)^2]^{1/2}} = \left| \int_{-r}^{+r} g'(-x) f'(x)\, dx \right| \Big/ n_0 \sqrt{\int_{-r}^{+r} f'(x)^2\, dx} \qquad (6)$$

The subscript *C* in $L_C(f)$ indicates that this is Canny's expression for localization.

[1] To be precise, Canny's assumption was that $E[h_n'(x_{\max})^2] \approx E[h_n'(0)]^2$, which is only guaranteed when $h_n''(x) \approx 0$.

Both $SNR(f)$ and $L_C(f)$ are proportional to the ratio of edge amplitude A to noise amplitude n_0. To obtain performance criteria that depend only on the filter, Canny defined Σ and Λ_C as

$$SNR(f) = \frac{A}{n_0}\Sigma(f), \quad \text{and} \quad L_C(f) = \frac{A}{n_0}\Lambda_C(f) \qquad (7)$$

Both Σ and Λ_C can be optimized simultaneously by finding a filter f which optimizes their product,

$$Opt(f) = \Sigma(f)\Lambda_C(f) \qquad (8)$$

Unfortunately, the optimizing filter produces multiple noise peaks in the vicinity of the edge, making it hard to decide which of the peaks is the edge.

To lessen this problem, Canny constrained the average distance between the noise peaks to be greater than some fraction of the filter width. The average distance between noise peaks is proportional to

$$Z(f) = \frac{E[h_n'(0)^2]^{1/2}}{E[h_n''(0)^2]^{1/2}} = \sqrt{\int_{-r}^{+r} f'(x)^2\,dx \Big/ \int_{-r}^{+r} f''(x)^2\,dx} \qquad (9)$$

The quantity $Z(f)$ is a measure of filter smoothness. The filter which maximized the product $Opt(f)$ subject to a constraint on $Z(f)$ was similar to a gaussian derivative (Canny 1986).

3. Two Problems with Canny's Edge Detector.

It has been previously noticed that Canny's localization measure $L_C(f)$ is incorrect (Tagare and deFigueiredo, 1990, Koplowitz and Greco, 1994). In section 3.1 below we derive a more correct localization criterion, which incorporates the smoothness functional $Z(f)$ when the edge amplitude is small. This means it is no longer necessary to impose a constraint on $Z(f)$ to get a good edge detection filter.

It has not however been previously noticed that the optimal filter, according to Canny's criteria, must be infinitely wide. In section 3.2, we show why this occurs. An infinitely wide filter is useless for edge detection in real images. In section 4, we suggest a solution to the infinite width problem, which leads us to a different form of optimal edge detector than that proposed by Canny.

3.1 Canny's Localization is Incorrect.

Tagare and deFigueiredo (1990) and Koplowitz and Greco (1994) have noted that Canny's assumption that $h_n''(0)=0$ is most likely wrong. By defining standardized normal random variables $X = h_n'(0)/E[h_n'(0)^2]^{1/2}$ and $Y = h_n''(0)/E[h_n''(0)^2]^{1/2}$, Equation (5) can be written as

$$x_{\max} = \left(\frac{E[h_n'(0)^2]^{1/2}}{E[h_n''(0)^2]^{1/2}} \right) \frac{X}{h_g''(0)/E[h_n''(0)^2]^{1/2} + Y} = Z(f) \frac{X}{h_g''(0)/E[h_n''(0)^2]^{1/2} + Y} \quad (10)$$

X and Y are uncorrelated if the noise $n(x)$ is white gaussian noise. Note that

$$h_g''(0)/E[h_n''(0)^2]^{1/2} = L_C(f)Z(f) \quad (11)$$

so Equation (10) can be written as

$$x_{\max} = \frac{1}{L_C(f)} \left(\frac{[L_C(f)Z(f)]X}{[L_C(f)Z(f)] + Y} \right) \quad (12)$$

For brevity, we will write Ω for the product $L_C(f)Z(f)$. The edge detector localization $L(f)$ is then

$$L(f) = E[x_{\max}^2]^{-1/2} = L_C(f)E[\left(\frac{\Omega X}{\Omega + Y} \right)^2]^{-1/2} \quad (13)$$

This is Canny's localization $L_C(f)$ divided by the standard deviation of a ratio of normal random variables, $E[(\Omega X/(\Omega + Y))^2]^{1/2}$. Unfortunately this ratio has an undefined first moment and infinite higher moments (Marsaglia 1965, Hinkley 1969), which makes the localization $L(f)$ zero. Apparently, then, it is impossible to localize a step edge! However, the infinite moments of the ratio $\Omega X/(\Omega + Y)$ are due to a subset of events where the denominator $(\Omega + Y)$ is close to zero (Marsaglia, 2006). These events most likely occur when the edge couldn't be detected at all (because of noise), rather than being detected infinitely far from its true location.

If we avoid these events by conditioning on a nonzero denominator, $|\Omega + Y| > \varepsilon$ for small ε, then the moments of the ratio $\Omega X/(\Omega + Y)$ do exist (Marsaglia, 2006). The first moment is zero. There is however no closed form for the second moment, but it can be estimated by Monte Carlo methods. Figure 1 shows Monte-Carlo estimates of $E[(\Omega X/(\Omega + Y))^2]^{-1/2}$ plotted as a function of Ω. A good approximation to these estimates is

$$E[\left(\frac{\Omega X}{\Omega + Y} \right)^2]^{-1/2} \approx \left(0.5 + \frac{1}{\pi} \arctan\left(\frac{\Omega - 3.2}{0.3} \right) \right)^{1.2} \quad (14)$$

This approximation is also plotted in Figure 1.

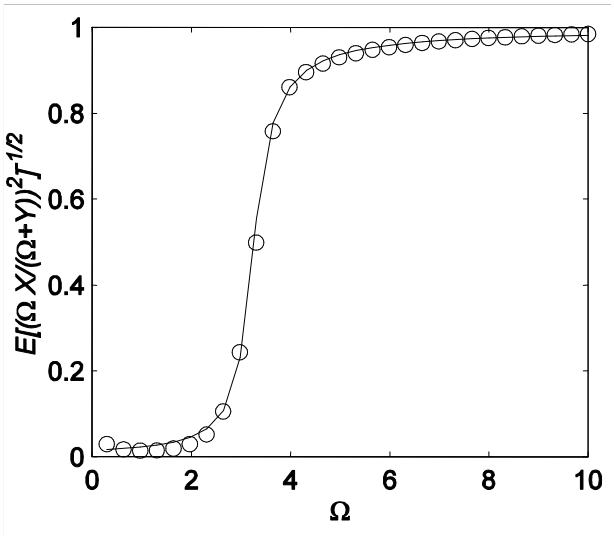

Figure 1. Monte Carlo estimates of $E[(\Omega X/(\Omega+Y))^2]^{-1/2}$ as a function of the parameter Ω. One Monte Carlo run consisted of 10,000 samples of the ratio $\Omega X/(\Omega+Y)$ at a particular value of Ω, from which $E[(\Omega X/(\Omega+Y))^2]^{-1/2}$ is calculated. The circles are medians of 1000 such runs at the same value of Ω. Medians were used to discount extreme values of the ratio $\Omega X/(\Omega+Y)$. The number of samples is sufficient that repeated runs produce almost no change in the plotted estimates. The curve shows the approximation given by Equation (14).

Substituting Equation (14) into Equation (13), and expanding the abbreviation Ω gives

$$L(f) \approx L_C(f)\left(0.5+\frac{1}{\pi}\arctan\left(\frac{L_C(f)Z(f)-3.2}{0.3}\right)\right)^{1.2} \qquad (15)$$

We can, as before, define an optimality measure in terms of $\Sigma(f)$ and $\Lambda_C(f)$ as

$$Opt(f) = \Sigma(f)\Lambda_C(f)\left(0.5+\frac{1}{\pi}\arctan\left(\frac{(A/n_0)\Lambda_C(f)Z(f)-3.2}{0.3}\right)\right)^{1.2} \qquad (16)$$

Unlike Canny's criterion, this still involves the edge-to-noise ratio A/n_0 and the functional $Z(f)$. A set of optimal detectors for different A/n_0 ratios is shown in Figure 2. Note that these are unconstrained optima; it is no longer necessary to impose a constraint on filter smoothness $Z(f)$. These are very similar to detectors derived by Canny for different constraints on $Z(f)$.

The choice of optimal detector depends on the ratio A/n_0, but we do not know what this is in advance. We can derive a compromise filter, which works reasonably well at all ratios, and Canny's choice was a filter rather similar to the one for A/n_0=1.33 in Figure 2. However, we will not commit to a choice here, as there is a more serious problem with Canny's edge detector.

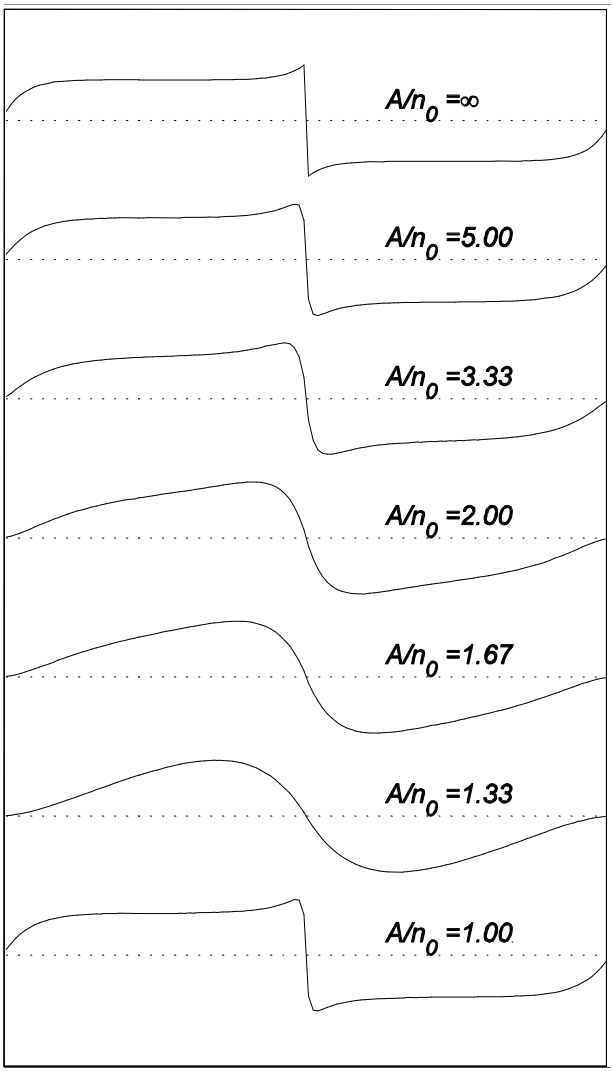

Figure 2. Optimal edge detectors for different edge to noise ratios A/n_0. The topmost filter is the best for very low noise; the influence of the functional *Z*(*f*) here is minor. It is not a matched filter but is smoothed at either end because of the need to minimize the derivative at the boundary. The filters for high noise levels (bottom filter) begin to look like the low noise filters again, because performance at low noise levels is not much dependent on *Z*(*f*) either.

3.2. The Optimal Edge Detector is Infinitely Wide.

The width of the edge detection filter has so far been set at some arbitrary value *r*. It would be useful to have an optimality argument for choosing the best filter width; indeed, without it, it is hard to claim that the filter at any particular width *r* is optimal. Canny (1986) argued that the signal-to-noise ratio $\Sigma(f)$ increases with filter width while the localization $\Lambda_C(f)$ decreases so that the product $\Sigma(f)\Lambda_C(f)$ is constant. Thus all widths are equally good, and some other consideration must be used to select the appropriate width. However, the reduction in filter localization as width increases is a consequence of Canny's constraint on

the filter smoothness $Z(f)$. He forced wider filters to be smoother than narrow filters, and this made localization poorer in the wider filters. When this constraint on filter smoothness is removed – as it can be thanks to the new formulation for localization given by Equation (15) – we find that wide optimal filters are as good as narrow ones for localization.

Consider the filter in Figure 3(a). This can be widened to improve the signal to noise ratio while keeping localization constant. One way of doing this is by first locating nonzero peaks or plateaus in the filter (shown by circles in Figure 3(a)) and then stretching them out (Figure 3(b)). Such peaks must always exist because the filter is odd, and is zero at the endpoints and the centre. The localization $L_C(f)$ of the stretched filter is unchanged because the added points all have a derivative $f'(x)$ of zero, so they do not change the integrals in $L_C(f)$. Equally, the introduction of points where $f'(x)$ =0 does not change the numerator of $Z(f)$ in Equation (9). The second derivative $f''(x)$ is unchanged except at the endpoints of the stretch (where it actually decreases), so the denominator of $Z(f)$ is likewise unaffected. Hence the localization $L(f)$ of the filter is unchanged because $L_C(f)$ and $Z(f)$ are unchanged.

$SNR(f)$ is improved by this stretching process, intuitively because the stretched filter looks more like a step edge. Consider the stretched filter in Figure 3(b). For simplicity, let n_0 = 1 and $g(x)$=1 or -1. Let the maximum value of the filter itself be f_{max}, and let y be the total length of the plateau after stretching. The SNR for the stretched filter is

$$\frac{1}{2}SNR(f) = \frac{\int_0^{+r} f(x)\,dx + yf_{\max}}{\sqrt{\int_{-r}^{+r} f(x)^2\,dx + yf_{\max}^2}} \qquad (17)$$

If stretching the filter increases the *SNR*, then the derivative of $SNR(f)^2$ with respect to y should be positive at y=0. The derivative is

$$\frac{1}{4}\frac{\partial SNR(f)^2}{\partial y}\bigg|_{y=0} = 2f_{\max}\frac{\int f(x)\,dx}{\int f(x)^2\,dx} - \left(f_{\max}\frac{\int f(x)\,dx}{\int f(x)^2\,dx}\right)^2 \qquad (18)$$

and it will be positive when

$$2 > f_{\max}\frac{\int f(x)\,dx}{\int f(x)^2\,dx} \qquad (19)$$

The fraction on the right hand side attains a maximum when $f(x)$ is a constant, say α, and it then has the value $1/\alpha$. However, if $f(x) = \alpha$ then $f_{\max} = \alpha$, so the maximum value of the right hand side of the inequality is 1. We conclude that the derivative of *SNR* with respect to y is positive, and stretching the filter improves *SNR*.

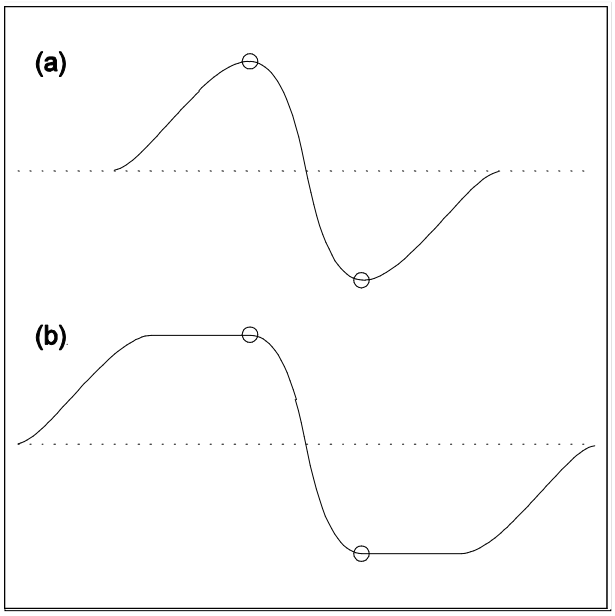

Figure 3. This shows how a filter can be altered to improve *SNR* while keeping localization L_C and smoothness, Z, constant. The upper filter is the original (narrow) filter. The peaks in the filter are identified by circles. These peaks are stretched out to form plateaus in (b). These plateaus do not affect localization and they increase signal-to-noise.

Since the stretched filter has increased $SNR(f)$ and unchanged $L(f)$, it has increased $Opt(f)$, from which it follows that the optimum filter must be infinitely wide[2]. This undesirable outcome is a consequence of the simplified edge model used so far. Canny's edge detection filter is optimized to detect an *isolated* edge in white noise, and under those conditions, the optimum filter is indeed infinitely wide, because there are no other edges that might interfere with detection. In real images, though, an infinitely wide filter will integrate over infinitely many edges, and will obviously be useless at detecting any one edge in particular. The only principled way to solve this infinite-width problem is to change the edge model to account for neighbouring edges. We do this in the next section.

4 Optimal Edge Detection in The Presence of Other Edges.

In real applications we are interested in detecting one edge in the presence of many other edges. A wide edge detector will have better signal-to-noise ratio for a single edge, but will be more likely to overlap other edges than a narrow detector. If a filter overlaps other edges, they will interfere unpredictably with the detection of the edge we are interested in. In this respect the other edges behave like noise, and they can be modelled as a stochastic process. Images typically have a $1/\omega^2$ power spectrum, where ω is the spatial frequency (Burton and Moorhead, 1987, Field 1987) which make them similar to Brown noise, albeit with a more interesting phase spectrum (Tadmor and Tolhurst, 1993). Given that the phase spectrum is unlikely to affect how the other edges interfere with detection of the edge we're interested in,

[2] This is different from the infinite filters of Deriche (1987) or Sarkar and Boyer (1991),which have a finite second moment.

it is sufficient for our purposes to model the other edges as simple Brown noise. This Brown noise will be what limits the width of the edge detector.

4.1 The Optimal Detector in Brown Noise.

Suppose the Brown noise in the image has a power spectrum C^2/ω^2. This is added to the white noise, with power spectrum n_0^2, to give a total noise power spectrum of $C^2/\omega^2 + n_0^2$. With brown noise, the expressions for *SNR*, L_C and Z are easier to handle in the Fourier domain. Corresponding to Equations (2), (6), and (9), these are:

$$\begin{aligned} SNR(F) &= \int G(\omega)F(\omega)\,d\omega \Big/ \sqrt{\int |F(\omega)|^2\,(C^2/\omega^2 + n_0^2)\,d\omega} \\ L_C(F) &= \left|\int \omega^2 G(\omega)F(\omega)\,ds\right| \Big/ \sqrt{\int \omega^2\,|F(\omega)|^2\,(C^2/\omega^2 + n_0^2)\,d\omega} \\ Z(F) &= \sqrt{\int \omega^2\,|F(\omega)|^2\,(C^2/\omega^2 + n_0^2)\,d\omega \Big/ \int \omega^4\,|F(\omega)|^2\,(C^2/\omega^2 + n_0^2)\,d\omega} \end{aligned} \qquad (20)$$

Here, $F(\omega)$ is the Fourier transform of the filter $f(x)$, and $G(\omega) = -iA/\omega$ is the Fourier transform of the step edge $g(x)$ with amplitude *A*. These expressions follow from Parseval's theorem and the derivative theorem (Bracewell, 1986).

Let $W(\omega)$ be the filter with Fourier transform

$$W(\omega) = \frac{i\omega}{\sqrt{C^2 + n_0^2\omega^2}} = i\omega B(\omega) \qquad (21)$$

which is the derivative operator $i\omega$ multiplied by a 1st-order Butterworth filter $B(\omega) = (C^2 + n_0^2\omega^2)^{-1/2}$. This filter whitens the noise, because $|W(\omega)|^2\,(C^2/\omega^2 + n_0^2) = 1$. The edge detector $F(\omega)$ can be factored into a product of the whitening filter $W(\omega)$ and a post-whitening detector $K(\omega)$,

$$F(\omega) = W(\omega)K(\omega). \qquad (22)$$

where $K(\omega)$ is the Fourier transform of some filter $k(x)$. Since $W(0) = 0$, this factorization requires that $F(0) = 0$, which is satisfied since $f(x)$ is an odd filter. By substituting the product $W(\omega)K(\omega)$ for $F(\omega)$, we can rewrite Equations (20) very simply in terms of the post-whitening detector $K(\omega)$ as

$$\begin{aligned} SNR(K) &= \int G_W(\omega)K(\omega)\,d\omega \Big/ \sqrt{\int |K(\omega)|^2\,d\omega} \\ L_C(K) &= \left|\int \omega^2 G_W(\omega)K(\omega)\,d\omega\right| \Big/ \sqrt{\int \omega^2\,|K(\omega)|^2\,d\omega} \\ Z(K) &= \sqrt{\int \omega^2\,|K(\omega)|^2\,d\omega \Big/ \int \omega^4\,|K(\omega)|^2\,d\omega} \end{aligned} \qquad (23)$$

where $G_W(\omega) = W(\omega)G(\omega) = A/\sqrt{C^2 + n_0^2\omega^2}$ is the Fourier transform of a whitened step edge, $g_W(x)$. The optimality criterion for the post-whitening detector $K(\omega)$ is thus

$$Opt(K,A) = \Sigma(K)\Lambda_C(K)\left(0.5 + \frac{1}{\pi}\arctan\left(\frac{A\Lambda_C(K)Z(K) - 3.2}{0.3}\right)\right)^{1.2} \qquad (24)$$

where as before $SNR(K) = A\Sigma(K)$ and $L_C(K) = A\Lambda_C(K)$. Here we explicitly note the functional dependence of the optimality criterion on the edge amplitude *A* by writing it as $Opt(K,A)$. Note that after whitening the noise amplitude is 1, so it can be dropped from Equation (24). However, the noise amplitude implicitly affects the shape and height of $G_W(\omega)$, and thus of $\Sigma(K)$ and $\Lambda_C(K)$.

The optimal detection filter $K(\omega)$ maximizes $Opt(K,A)$. In addition, we have to require the spatial version of the filter $k(x)$ to be non-negative. This is because the whitened edge $g_W(x)$ is entirely positive or entirely negative depending on the sign of the edge amplitude *A*[3]. If $k(x)$ has negative lobes, it would be impossible to determine whether a positive peak in the filter output was due to the overlap between a positive edge $g_W(x)$ and the positive centre of $k(x)$, or to a negative edge $-g_W(x)$ which lines up with a negative part of $k(x)$. This constraint is necessary because without it the optimal detector $k(x)$ does have wide negative regions at some values of edge amplitude *A*.

When the edge amplitude *A* is large, $Opt(K,A)$ simplifies to $\Sigma(K)\Lambda_C(K)$, and is maximized by the matched filter $K(\omega) = G_W(\omega)$ (Canny, 1986). In this case, the edge detector $F(\omega)$ is

$$F(\omega) = W(\omega)G_W(\omega) = \frac{i\omega}{C^2 + n_0^2\omega^2} \qquad (21)$$

which, up to a multiplicative constant, corresponds to the spatial filter

$$f(x) = sign(x)\exp(-(C/n_0)\,|\,x\,|) \qquad (22)$$

The width of the filter varies according to the ratio of white noise to brown noise. In the extreme case of *C*=0, the filter is an infinitely wide step edge, which is consistent with the infinitely wide filter found in section 3.2. The other extreme, as n_0 tends to zero, yields a derivative operator.

The filter in equation (22) is identical to one previously suggested by Shen and Castan (1992). It is the derivative of an infinite symmetric exponential filter (ISEF), so we will refer to it as a DISEF filter. Shen and Castan (1992) derived the DISEF filter using an isolated edge model, like Canny, but used different, albeit related, optimality criteria. They also considered the problem of multiple edges, but they modelled the edges in an image by a random telegraph-signal (RTS), which switches randomly between two values. Under the RTS model, Shen and Castan derived the scale factor in the exponential of equation (22) as $\sqrt{4\lambda^2 + C^2/n_0^2}$, where λ is the switching density, rather than C/n_0. Shen (1995) further considers a sum-of-RTS's edge model, more similar to the Brown noise model used here, but

[3] The whitened edge $g_W(x)$ is, to within a multiplicative constant, $AK_0(|\,x\,|)$, where $K_n(x)$ is the modified Bessel function of the second kind of order *n*. If *A* is positive, $g_W(x) > 0$ for all *x*.

this yields a complicated edge detection scheme. Despite these differences of detail, there is a remarkable convergence of results between the Canny approach used here and Shen and Castan's edge detector.

4.2 A Compromise Edge Detector.

The DISEF filter of Equation (22) is only optimal for large edge amplitudes *A*. When *A* is smaller, the optimal post-whitening detector $K(\omega)$ is not matched to the whitened edge. Figure 4 shows some examples of optimal post-whitening detectors in the spatial domain, $k(x)$, and their corresponding edge detectors $f(x)$, for different values of *A*. These result from numerical optimization of the spatial domain version of equation (24). The dependence of the detectors on the unknown amplitude *A* means it is impossible to find an edge detector which is universally optimal. We can nonetheless propose a compromise filter, which works reasonably well at all amplitudes *A*. A common characteristic of the detectors in Figure 4 is that they are all wider than the matched filter obtained when $A = \infty$. A good compromise edge detector might therefore be obtained by simply widening, or blurring, the matched filter slightly.

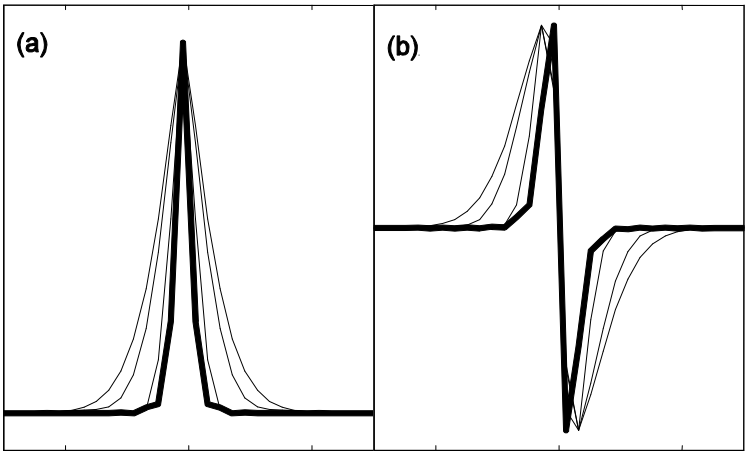

Figure 4: (a) Plot of some optimal post-whitening detectors $k(x)$, for various values of edge amplitude *A*, given *C*=1 and n_0=0.8. The thick lined profile is the matched detector. The wider detectors are for intermediate values of *A*. The plots are 30 pixels wide. (b) Plot of corresponding edge detectors $f(x)$, which are whitened versions of the detectors $k(x)$. The thick-lined edge detector is the DISEF filter (Equation 22) which corresponds to the matched post-whitening detector in (a).

We need to define what we mean by a good compromise. Let K_A be the detector which maximizes $Opt(K, A)$ at some edge amplitude *A*. The matched detector (Equation 22) is K_∞. We wish to find a compromise detector $\hat{K}$ which is good in a minimax sense compared to all the optimal detectors K_A. That is, we wish to find a detector $\hat{K}$ which minimizes the maximum loss

$$\sup_A \left\{ Opt(K_A, A) / Opt(\hat{K}, A) \right\} \qquad (23)$$

In addition, we want to limit the loss for the compromise detector at high edge contrasts, namely $Opt(K_\infty,\infty)/Opt(\hat{K},\infty)$, to a "reasonable" value, so that the compromise detector does not throw away too much performance when the edge has high contrast. Finally, we would like the compromise filter to be mathematically simple. We restrict ourselves to filters $\hat{K}$ produced by convolving the matched filter K_∞ with a simple nonnegative smoothing function; a gaussian smoother was found to produce good results. Thus our compromise detector is given by

$$\begin{aligned}\hat{K}(\omega,\sigma) &= K_\infty(\omega)\exp(-(\omega\sigma)^2) \\ &= \exp(-(\omega\sigma)^2)/\sqrt{C^2+n_0^2\omega^2}\end{aligned} \qquad (24)$$

where σ is the smoothing parameter. The choice of compromise filter thus boils down to the choice of σ.

Figure 5 compares the performance of a compromise filter to the optimal filter performance. Here, the noise parameters are set to C=1 and n_0 =0.05 (for the top set of curves) or n_0 =0.8 (for the lower set of curves). The crosses show the performance of the matched filter, given by $Opt(K_\infty, A)$, for various values of A. The thick solid lines show $Opt(K_A, A)$. Each point in this curve gives the performance of a different optimal filter K_A. The thin line shows the performance of a compromise filter $Opt(\hat{K}, A)$. This filter is given by smoothing the matched filter for that C/n_0 ratio by a small 3-point gaussian equal to [0.2221 0.5557 0.2221] ≅ [$^2/_9$, $^5/_9$, $^2/_9$] . The optimal detector is no more than 85% better than this compromise detector, and the matched filter is no more than 40% better than it, while the compromise filter often greatly exceeds the matched filter's performance.

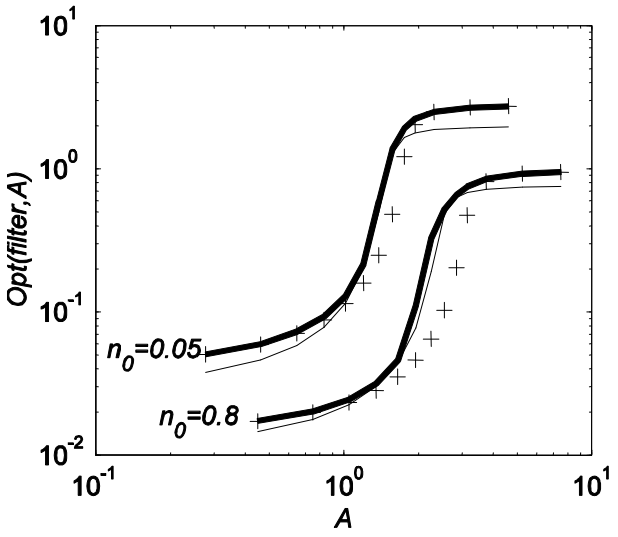

Figure 5. Performance of edge detectors at different noise levels. The brown noise was set to C=1. White noise density n_0 was either 0.05 (top curves) or 0.8 (bottom curves). In each group of curves, the thick solid line shows the performance of the optimal detectors at different values of the ratio A. The crosses show the performance of the matched post-whitening filter K_∞. The matched filter suffers a substantial loss of performance for some values of A. The thin line shows the performance of the best compromise filter, which is the matched filter convolved with a three-point gaussian.

To summarize, the best compromise edge detection algorithm in one dimension is as follows:

Algorithm 1:

1) Estimate C and n_0 from the image, or use sensible presets.
2) Convolve the image with a 3 point gaussian [$^2/_9$, $^5/_9$, $^2/_9$]
3) Convolve the resultant blurred image with the DISEF filter $f(x) = sign(x)\exp(-(C/n_0)|x|)$.
4) Positive-valued peaks or negative-valued troughs in the output of this filter represent edges, if they are strong enough.

This edge detection scheme can be easily extended to two dimensions (although the optimality properties of this extension are unknown). To develop the 2D scheme, we can write the optimal 1D edge detector as a series of filters

$$\begin{aligned} F(\omega) &= K(\omega)W(\omega) \\ &= \{B(\omega)G_3(\omega)\}\{i\omega B(\omega)\} \end{aligned}$$

where in the second line we have factored the whitening filter $W(\omega)$ into the derivative $i\omega$ and a Butterworth smoothing filter $B(\omega)$, and the optimal post-whitening filter $K(\omega)$ is factored into the product of a Butterworth filter (which is matched to the whitened edge) and the three point gaussian $G_3(\omega)$. We can rearrange this to give

$$F(\omega) = i\omega B(\omega)^2 G_3(\omega)$$

where $B(\omega)^2$ is the ISEF filter. To extend the algorithm to two dimensions, we replace $B(\omega)^2$ and $G_3(\omega)$ with their two-dimensional counterparts, and replace the derivative with the directional derivative. This yields the following algorithm:

Algorithm 2:

1) Estimate C and n_0 from the image, or use sensible presets.
2) Convolve the image rows, then the columns, with a 3 point gaussian [$^2/_9$, $^5/_9$, $^2/_9$]
3) Convolve the resultant blurred image with the circularly symmetric ISEF filter $\exp(-(C/n_0)\sqrt{x^2+y^2})$.
4) Compute the directional derivatives in all directions. These can easily be computed from the direction derivatives along the rows and columns.
5) Peaks in the output of this filter across space and derivative direction represent edges, if they are strong enough. It is usually necessary to use a hysteresis algorithm (Canny, 1986) to sort these peaks into coherent edges.

The performance of this algorithm is shown in Figure 7, middle panel. One can see that in the absence of significant noise, it appears to perform well. However, shadow edges are poorly represented.

5. Detecting Blurred Edges.

Many edges are not step edges. Generally, softer or blurred edges are caused by defocus, self-shadowing of a curved surface, or by shadow penumbrae (see e.g. Elder and Zucker 1998, Elder 1999). These various kinds of blurred edge can be modelled as a step edge blurred with a gaussian function,

$$g(x,\sigma) = g(x) * gauss(x,\sigma) \qquad (25)$$

where $gauss(x,\sigma) = \exp(-x^2/(2\sigma^2))/\sqrt{2\pi\sigma^2}$ is a unit gaussian distribution with width σ. A step edge is the limiting case of zero blur, i.e. $g(x,0)$.

5.1 The Optimal Detector for Blurred Edges

Introducing blurred edges changes the expressions for $SNR(K)$ and $L_C(K)$ in Equation (23) to the following:

$$\begin{aligned} SNR(K) &= \int \{G_W(\omega)Gauss(\omega,\sigma)\}K(\omega)\, d\omega \Big/ \sqrt{\int |K(\omega)|^2\, d\omega} \\ L_C(K) &= \left|\int \omega^2 \{G_W(\omega)Gauss(\omega,\sigma)\}K(\omega)\, d\omega\right| \Big/ \sqrt{\int \omega^2 |K(\omega)|^2\, d\omega} \end{aligned} \qquad (26)$$

Here $Gauss(\omega,\sigma)$ is the Fourier transform of $gauss(x,\sigma)$ in Equation (25), and $\{G_W(\omega)Gauss(\omega,\sigma)\}$ is the Fourier transform of the whitened gaussian edge. The product of $SNR(K)$ and $L_C(K)$ is, as before, maximized by the matched filter $K(\omega,\sigma) = G_W(\omega)Gauss(\omega,\sigma)$. Hence the optimal edge detector for a gaussian edge of width σ, $F(\omega,\sigma)$, is given by

$$F(\omega,\sigma) = W(\omega)G_W(\omega) = \frac{i\omega}{C^2 + n_0^2\omega^2} Gauss(\omega,\sigma) \qquad (27)$$

which is just a blurred DISEF filter. The optimal spatial filter is then

$$f(x,\sigma) = sign(x)\exp(-(C/n_0)|x|) * gauss(x,\sigma) \qquad (28)$$

where * denotes convolution.

Each edge detector $f(x,\sigma)$ is optimized for a particular edge blur, but will respond to edges with other blurs as well. How can we select the appropriate blur detector for the edge, given that we don't know in advance what the blur is? When we restrict ourselves to matched detectors, this problem is easily solved. Let $k(x,\sigma)$ be the spatial version of $K(\omega,\sigma)$ above, and let $h(x,\sigma)$ be the detector output obtained by convolving $k(x,\sigma)$ with the whitened signal. The detector output $h(x,\sigma)$ forms a scale space representation of the input signal (Witkin, 1983). If we normalize the detector $k(x,\sigma)$ so that $\int k(x,\sigma)^2\, dx = 1$, then its expected response at any point is identical to the signal to noise ratio, $SNR(k)$. Hence a peak in the scale space $h(x,\sigma)$ is a local maximum of $SNR(k)$. This will also be a local maximum of $L_C(k)$, since *SNR* and L_C are maximized by the same matched filter (Canny, 1986). The peak in the scale space $h(x,\sigma)$ is therefore a local maximum of the product $SNR(k)L_C(k)$ and identifies the locally optimal detector.

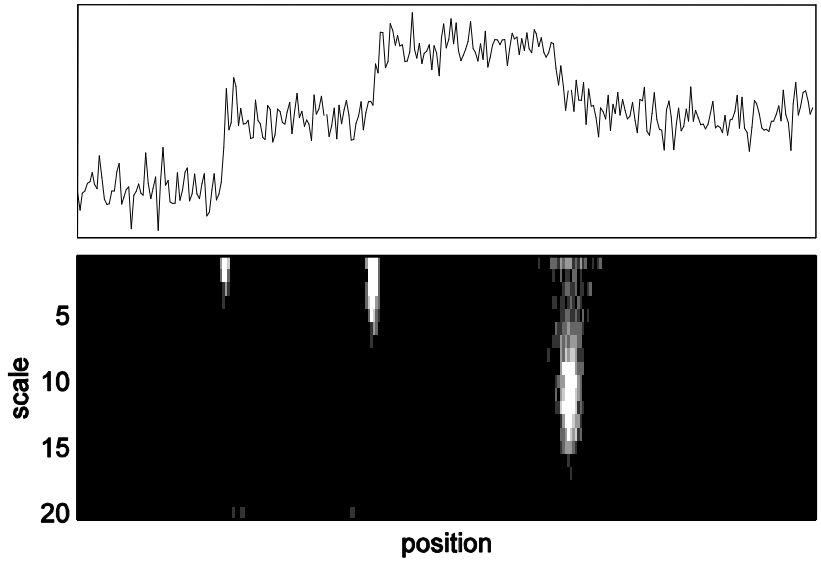

Figure 6. The top image shows an example of a signal with three gaussian edges, with widths of 1, 3 and 8 from left to right. One thousand such images were generated, with different noise samples, and the blurred-edge detection algorithm run on them. Each image was then convolved with a DISEF filter and gaussian blurs of different sizes to yield a scale space map $h(x,\sigma)$. Strong peaks in the scale space indicate edges. The bottom image shows all the edges found in 1000 runs of the algorithm, given C=0.5 and n_0=0.2. Most of the time, the edge is located within a few pixels of its true location, but identification of the blur is less accurate. In particular, there is a tendency to detect the edge at a finer scale than it really is. If we look at the example image, however, we can see that sometimes the added noise does make the edge (particularly the far right one) appear sharper than it is. Incidentally, in equivalent conditions human observers also seem to have this problem (May & Georgeson, 2007).

The performance of this model in visualized in Figure 6. The top panel shows an example noisy edge. The bottom panel shows all the local maxima (after thresholding to remove noise) accumulated over 1000 runs of the optimal scale-space edge detector described above, each with a different noise sample but the same set of edges. The position of the edges is fairly well estimated, but the blur less so.

The matched scale-space detector described above takes on an interesting form when the white noise n_0 is zero. In this case, the whitening operator is a simple derivative, and the whitened edge $g_W(x)$ is an impulse function. The matched post-whitening detector is $k(x,\sigma) = g_W(x) * gauss(x,\sigma)$, which is simply $gauss(x,\sigma)$. The norm of this is $\int gauss(x,\sigma)^2\ dx = 1/(2\sqrt{\pi\sigma})$, hence $k(x,\sigma) = \sqrt{2}\pi^{1/4}\sigma^{1/2} gauss(x,\sigma)$ is a normalized post-whitening filter. Since whitening is simply a derivative operation, this means that the blurred edge detector is, to within a constant scaling, just

$$f(x,\sigma) = \sigma^{1/2} \frac{d}{dx} gauss(x,\sigma) \qquad (30)$$

This is identical to Lindeberg's (1998) edge detection scheme (also suggested for human edge perception by Georgeson et al. 2007), so these proposals can both be understood as optimal blurred edge detectors for minimal amounts of white noise.

5.2 Two Dimensions

As with step edges, the blurred edge detector can be extended to a 2D algorithm, in a similar way. The optimal one-dimensional detector $F(\omega,\sigma)$ for an edge of blur σ is given by

$$\begin{aligned} F(\omega,\sigma) &= K(\omega,\sigma)W(\omega) \\ &= \{\lambda Gauss(\omega,\sigma)B(\omega)\}\{i\omega B(\omega)\} \\ &= i\omega Gauss(\omega,\sigma)B(\omega)^2 \end{aligned}$$

where $\lambda = \left(\int |K(\omega,\sigma)|^2 \, d\omega\right)^{-1/2}$ is the normalization factor for the filter. For simplicity, the three-point gaussian used to make a compromise filter has been left out here. To extend this to two dimensions, we simply replace $B(\omega)^2$ and $Gauss(\omega,\sigma)$ with their two-dimensional counterparts, and replace the derivative with the directional derivative. The appropriate normalization factor is trickier. At first sight it would seem reasonable to use $\lambda = \left(\int |K(\omega_1,\omega_2,\sigma)|^2 \, d\omega_1 d\omega_2\right)^{-1/2}$, where ω_1 and ω_2 are the row and column frequencies. However, this does not work well at all. The normalization is only appropriate along the direction of the derivative, since that is the only direction in which the post-whitening detector K can be expected to match the edge. Because of this, the appropriate normalization is one dimensional,

$$\lambda_{2D} = \left(\int \left|\int K(\omega_1,\omega_2,\sigma)\, d\omega_1\right|^2 d\omega_2\right)^{-1/2}$$

That is, we integrate along one dimension parallel to the edge, then square and integrate along the other dimension perpendicular to the edge[4]. This leads to Algorithm 3, for 2D blurred edge detection.

Algorithm 3:

1) Estimate C and n_0 from the image, or use sensible presets.
2) Convolve the image with the circularly symmetric ISEF filter $\exp(-(C/n_0)\sqrt{x^2+y^2})$.
3) For each scale σ,
    a. convolve the image with a two-dimensional gaussian $gauss(x,y,\sigma)$. The convolution is then scaled by the normalization factor λ given above.
    b. Compute the directional derivatives in all directions θ. These can easily be computed from the direction derivatives along the rows and columns.

[4] I'm not 100% sure this is the correct normalization, as it would depend on the model for the image in the vicinity of an edge.

4) Local peaks in the output of this filter across space (x,y), derivative direction θ, and scale σ, represent edges, if they are strong enough. It is necessary to use a hysteresis algorithm (Canny, 1986) to sort these peaks into coherent edges.

The performance of this algorithm on a relatively noise free image is shown in Figure 7, right hand panel. Compared to the middle panel (Algorithm 2), which only detects step edges, we can see that Algorithm 3 correctly picks up the shadow edges. However, its performance in some other parts of the image is not as clean as the step edge algorithm.

One useful feature of this edge detection model is its ability to cope with noise when the parameters of the whitening filter, C and n_0, are estimated from the image. This is shown in Figure 8. As the white noise increases, the smoothing provided by the Butterworth filter also increases.

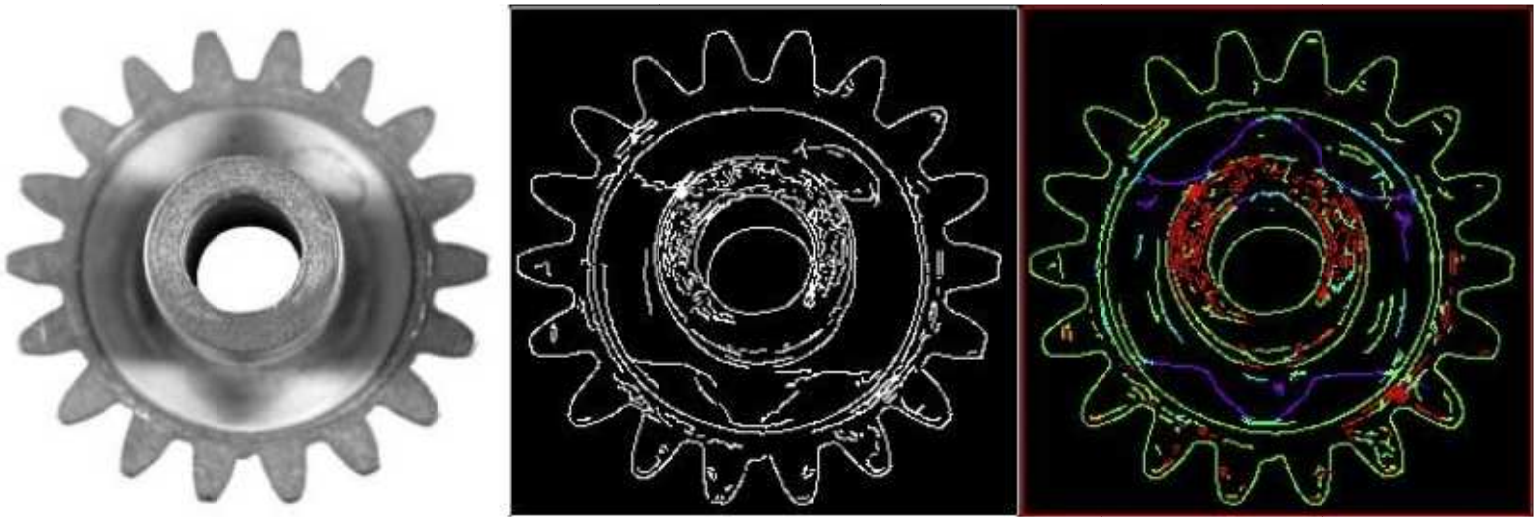

Figure 7: The optimal 2D edge detector in action. The grey scale image on the left is analyzed by the 2D step edge detector (Algorithm 2) in the middle panel. This algorithm is good at detecting many edges, but fails when given a blurred edge, such as the shadows. The right hand panel shows the edges found by Algorithm 3. The edge scale is colour coded red, green, cyan, purple from finest to coarsest scales. This algorithm correctly identifies the shadow edges as well as some blurred grooves. In both panels, the image noise parameters C and n_0 were estimated from the image.

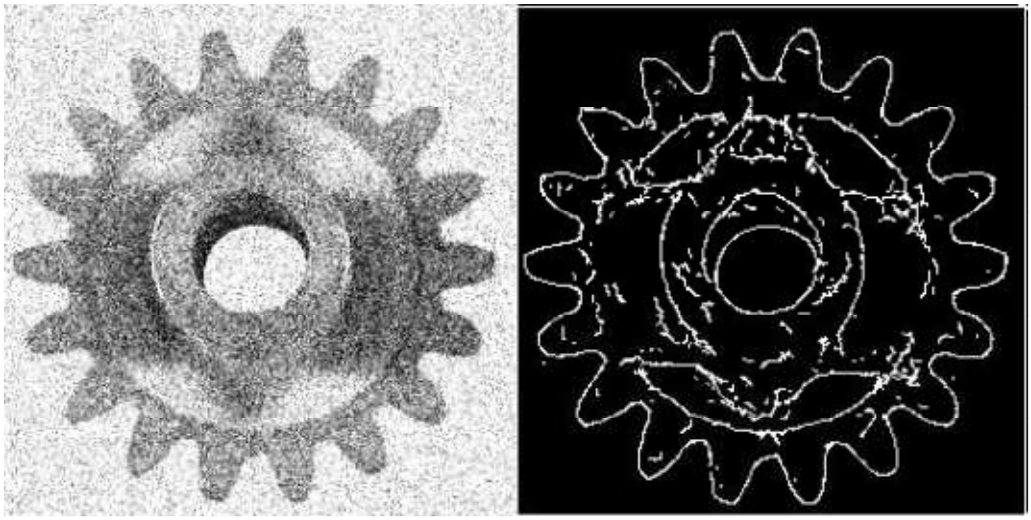

Figure 8. The grey scale image in Figure 7, with added noise. The result of running Algorithm 3 on it is shown on the right panel. Visually, the noise has eliminated much of the detail in the image, but most of what remains is detected by the algorithm, and most of the noise can be rejected.

6. Conclusion.

Canny's (1986) paper was a significant contribution to the methodology of edge detection. Prior to it, the actual performance criteria for edge detectors were rarely stated explicitly. After it, it is more or less impossible to propose an edge detector without reference to the Canny criteria. After such a major methodological advance, the issue of what an optimal edge detector actually looks like is perhaps less vital, although since we must use them, it is important to get it right. When we do so, we find that the optimal step edge detector is not similar to a derivative of a gaussian filter, but is instead the derivative of an exponential filter (DISEF) proposed by Shen and Castan (1992). In addition, once we have solved the optimal detector for step edges, it is relatively easy to extend to the task of detecting edges of different blurs. One remaining problem with the optimal detector proposed here is localization: the very complex localization criterion means that no one filter can be optimal at all edge contrasts.

The ISEF filter is optimal when the noise has a power spectrum of the form $C^2 + n_0^2$. However, this sometimes is not the true form of the noise. For example, if the imaging device has poor optical quality (such as a cheap webcam, a CCTV camera, or the human eye), the optical blur will change the slope of the Brown noise. In this case, the whitening filter will change, and one must replace the ISEF filter with something else; namely, a filter with power spectrum $K(\omega) = |W(\omega)|^2$, where the whitening filter $W(\omega)$ changes with the image statistics. Thus the algorithms in this paper can be altered to adapt to image statistics. A procedure like this is potentially behind adaptation effects in human vision (e.g. Wainwright, 1999, Webster & Georgeson, 2002).

Finally, the optimality of the edge detectors here was only shown for the 1D case. While the 2D algorithms perform well, it is still an open problem what criteria should be used to develop optimal 2D detectors.

References.

Bracewell, R.N. (1986) The Fourier Transform and Its Applications, *McGraw-Hill, New York.*

Burton GJ, Moorhead IR (1987) Color and spatial structure in natural scenes. *Applied Optics*, 26 157-170

Canny, J (1986) A Computational Approach to Edge Detection, *IEEE Trans. Patt. Anal. Image Proc*. 8(6) 679-698.

Deriche, R. (1987) Using Canny´s Criteria to Derive a Recursively Implemented Optimal Edge Detector, *International Journal of Comp. Vision*, 1(2), 167-187

Elder, J & Zucker, S. (1998) Local Scale Control for Edge Detection and Blur Estimation, *IEEE Trans. Patt. Anal. Machine Intell*. 20(7) 699-716

Elder, J (1999) Are Edges Incomplete?, *International Journal of Comp. Vision*, 34, 97-122.

Field, D (1987) Relations between the statistics of natural images and the repsonse properties of cortical cells, *J. Opt. Soc. Am*. A 4 2379-2394

Georgeson, M. A., May, K. A., Freeman, T. C. A., & Hesse, G. S. (2007). From filters to features: Scale–space analysis of edge and blur coding in human vision. *Journal of Vision*, 7(13):7, 1–21, http://journalofvision.org/7/13/7, doi:10.1167/7.13.7.

Hinkley D.V (1969). On the Ratio of Two Correlated Normal Random Variables. *Biometrika*, 56, 635–639.

Koplowitz, J & Greco, V (1994) On the Edge Location Error for Local maximum and Zero-Crossing Edge Detectors, *IEEE Trans. Patt. Anal. Image Proc*. 16(12) 1207-1212.

Lindeberg, T (1998) Edge Detection and Ridge Detection with Automatic Scale Selection, *International Journal of Comp. Vision*, 30(2), 117--154.

Marsaglia G (1965) Ratios of Normal Variables and Ratios of Sums of Uniform Variables. *Journal of the American Statistical Association*, 60, 193–204.

Marsaglia, G (2006) Ratios of Normal Variables, *Journal of Statistical Software*, 16(4) 1-10.

May, K. and Georgeson, M (2007) Blurred edges look faint, and faint edges look sharp, *Vision Research* 47 1705–1720

Peli, T. & Malah, D. (1982) A Study of Edge Detection Algorithms, *Computer Graphics and Image Processing*, 20, 1-21.

Petrou, M. & Kittler, J. (1991) Optimal Edge Detectors for Ramp Edges, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 483-491

Sarkar, S. & Boyer, K.L. (1991) On Optimal Infinite Impulse Response Edge Detection Filters, *IEEE Trans. Pattern Analysis Machine Intelligence*, 13, 1154-1171

Shen, J & Castan, S. (1992) An optimal linear operator for step edge detection, *CVGIP: Graphical Models and Image Processing*, 54, 112-133

Shen J (1995) Multi-Edge Detection by Isotropical 2-D ISEF Cascade, *Pattern Recognition*, 28, 1871-1885.

Tadmor, Y & Tolhurst, DJ (1993), Both the phase and the amplitude spectrum may determine the appearance of natural images, *Vision Research*, 33(1):141-145

Tagare, H & deFigueiredo, R (1990) On the Localization Performance Measure and Edge Detection, *IEEE Trans. Patt. Anal. Image Proc*. 12(12) 1186-1190.

Wainwright, M. J. (1999) Visual adaptation as optimal information transmission, *Vision Research*. 39, 3960–3974

Webster, M.A. Georgeson, M. & Webster, S.M. (2002) Neural adjustments to image blur, *Nature Neuroscience*, 5(9) 839-840

Witkin, A. P. (1983) Scale–space filtering, *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, 1019–1022.

Ziou, D. & Tabbone, S. (1998) Edge Detection Techniques - An Overview, *International Journal of Pattern Recognition and Image Analysis*, 8, 537-559.