Optimal, Efficient, Recursive Edge Detection Filters

S. Sarkar and K. L. Boyer
Signal Analysis and Machine Perception Laboratory
Department of Electrical Engineering
The Ohio State University, Columbus, OH 43210

Abstract

We outline the design of an optimal, efficient, infinite impulse response edge detection filter. We compute the optimal filter based on Canny's SNR - Localization product criterion and a constraint on the spurious noise response. However, we incorporate an appropriate spatial width measure for infinite impulse responses directly *in the expression for spurious response. The three criteria are then maximized using the variational method and non-linear constrained optimization. We tabulate the optimal filter parameters for various performance criteria and present a complete methodology for implementation using approximating recursive digital filtering. The approximating filter is separable into two linear filters operating in two orthogonal directions. The implementation is very simple and computationally efficient, having constant time execution with respect to scale, facilitating real time hardware implementation.*

I. Introduction

The value of edge detection as a starting point for many computer vision and image analysis paradigms is well known. Of the existing approaches [1, 2, 3], detecting edges by linear filtering is one of the most common. Canny [4] approached it by formulating three criteria desired in any edge detection filter: good detection, good localization, and low spurious response. He maximized the product of the first two criteria while keeping the spurious response criterion constant. Using the variational approach, he derived a set of finite extent step edge detection filters corresponding to various values of the spurious response criterion (or multiple response criterion, as we call it). Canny approximates the finite extent filter derived for step edges by the first derivative of a Gaussian. As Deriche [5] points out, no analysis was done in this approximating stage. The derived filter was of finite extent, but was approximated by an infinite extent filter.

Our approach is more direct; we formulate the three criteria as appropriate for a filter of infinite impulse response and, using the calculus of variations, optimize the composite criteria. Although the filter we derive is also well approximated by first derivative of a Gaussian, we are able to achieve a superior recursively implemented approximation directly. *We do not recommend using the first derivative of a Gaussian.*

We proceed directly to infinite impulse response (IIR) filters. An IIR filter can be approximated by a recursive digital filter which is spatially scaled by a change of coefficients, not the filter order. Hence, these filters exhibit constant run time with respect to scale. For FIR filters, scaling induces a change in the convolution mask producing a corresponding change in the number of computations required at the time of execution. IIR filters are storage efficient and so are very amenable to sequential hardware implementation. The filter we present is separable, i.e. it can be decomposed into separate row and column convolutions, so we can process all the columns and the rows simultaneously, making the edge detection process parallelizable.

*This research funded by the Department of Psychiatry at The Ohio State University, by the NASA Center for the Commercial Development of Space, and by an equipment grant from the National Science Foundation.

II. Optimal Edge Detection Criteria

We derive the operator for a 1D edge and then extend to the 2D case. Let the impulse response of the desired filter be $f(x)$, and denote the edge itself by $G(x)$. The edge is assumed to be centered at $x = 0$. Noise is assumed to be additive, white, and Gaussian with variance σ^2. The three optimality criteria are as follows, the first two being those proposed by Canny [4].

Good Detection: In white Gaussian noise the probability of false detection decreases with increasing signal to noise ratio. The expression for signal to noise ratio (SNR) at the edge location is:

$$SNR = \frac{\left| \int_{-\infty}^{\infty} G(-x) f(x) dx \right|}{\sigma \sqrt{\int_{-\infty}^{\infty} f^2(x) dx}} \tag{1}$$

Good Localization: This criterion is expressed as the inverse of the variance of the position of the maximum of the filter response about the correct location. A Taylor series expansion of the filter output yields the mean-squared value of the deviation, $E[x_o^2]$:

$$E[x_o^2] = \frac{\sigma^2 \int_{-\infty}^{\infty} f'^2(x) dx}{\left[\int_{-\infty}^{\infty} G'(x) f'(x) dx \right]^2}$$

Localization is defined as the reciprocal of $E[x_o^2]$:

$$Localization = \frac{\left| \int_{-\infty}^{\infty} G'(-x) f'(x) dx \right|}{\sigma \sqrt{\int_{-\infty}^{\infty} f'^2(x) dx}} \tag{2}$$

Low Multiple Response: This criterion constrains the number of spurious maxima within the operator's spatial spread. The expected distance between maxima of the output noise is set to some constant, k, times the operator width, W. An expression for the mean distance between maxima of the filter output in response to noise is found [4], using a result due to Rice [6], to be of the form:

$$x_{max} = 2\pi \left(\frac{\int_{-W}^{W} f'^2(x) dx}{\int_{-W}^{W} f''^2(x) dx} \right)^{1/2} = kW \tag{3}$$

It is desirable that k is as large as possible so that the distance between spurious maxima is as great as possible.

The multiple response criterion must be modified to apply to filters of infinite extent. Canny and Deriche [5] each set $x_{max} = kW$, where W is the width of the operator and k is a constant. In Canny's case, W was well defined, so he sought to maximize x_{max}. But for an operator of infinite extent, we need an alternate expression for the width of the operator. It is easy to argue that the optimal filter for a symmetric edge will be symmetric while the operator for an anti-symmetric edge will be anti-symmetric [4]. Given that $G(x)$ is either purely symmetric or purely anti-symmetric, the filter $f(x)$ will be purely symmetric or purely anti-symmetric, respectively, about $x = 0$. Thus, in either event, $f^2(x)$ will be symmetric about $x = 0$. We can therefore define the width of the filter as the normalized root mean square deviation of the function $f^2(x)$ about the origin:

$$\bar{W} = \left(\frac{\int_{-\infty}^{\infty} x^2 f^2(x) dx}{\int_{-\infty}^{\infty} f^2(x) dx} \right)^{1/2} \tag{4}$$

Since this expression, the RMS spatial width, leads to the right scaling properties for the multiple response criterion, its use is justified. Note that this is analogous to the expression for the RMS bandwidth of a filter. Torre and Poggio used the same expression for spatial width [7] (their Equation 3.10). We substitute this expression for W in the righthand side of Equation (3) and change the limits on the integrals to $\pm\infty$ and our modified multiple response criterion (MRC) becomes:

$$MRC = 2\pi \left(\frac{\int_{-\infty}^{\infty} f'^2(x) dx}{\int_{-\infty}^{\infty} f''^2(x) dx} \right)^{1/2} \left(\frac{\int_{-\infty}^{\infty} f^2(x) dx}{\int_{-\infty}^{\infty} x^2 f^2(x) dx} \right)^{1/2} = k \tag{5}$$

As this criterion is increased (numerically), the resulting filter will exhibit fewer spurious responses. Note that the multiple response criterion is invariant with respect to spatial scale, i.e., $MRC[f(x/a)] = MRC[f(x)]$ producing a multiple response criterion which is a function of the filter's inherent *shape* and not its scale.

III. Design of the Optimal IIR Step Edge Detector

A. Optimality criterion for step edges

Although these criteria can be used to derive detectors for various types of edges, we restrict the remainder of the discussion to step edges. For a step edge $G(x) = Au(x)$, where $u(x)$ is the unit step and A is the amplitude of the edge. Following Canny, we modify the signal to noise ratio and the localization so that they are invariant to signal amplitude and noise power by defining Λ and Σ in the following manner: $\Sigma = \frac{\sigma}{A} SNR \quad \Lambda = \frac{\sigma}{A} Localization$.

Substituting for $G(x)$ in the expressions for SNR and *Localization*, we have the product $\Sigma\Lambda$, Equation (6), which is invariant with respect to spatial scale. We then maximize this product, subject to the multiple response constraint given by Equation (5).

$$\Sigma(f)\Lambda(f') = \frac{\left| \int_{-\infty}^{0} f(x) dx \right|}{\sqrt{\int_{-\infty}^{\infty} f^2(x) dx}} \frac{|f'(0)|}{\sqrt{\int_{-\infty}^{\infty} f'^2(x) dx}} \tag{6}$$

B. The Variational Approach

We now seek a function, $f(x)$, which maximizes $\Sigma\Lambda$, subject to multiple response criterion (MRC). The calculus of variations [8] offers perhaps the most elegant solution. To maximize $\Sigma\Lambda$ we appropriately extremize its integrals. Following Canny, we transform the maximization of (6) to constrained minimization involving only the integral functionals. One integral is chosen to be extremized while the rest are held constant. Since $\int_{-\infty}^{0} f''^2(x) dx$ is in the denominator of the multiple response criterion, Equation (5), we choose to minimize it holding the rest constant with the boundary conditions: $f(0) = 0$, $f(-\infty) = 0$, $f'(0) = c5$, $f'(-\infty) = 0$. The solution is then a function of those undetermined constants which are in turn chosen to maximize our optimality criterion. By reciprocity, we could have chosen to extremize any of the other integrals while keeping the rest constant, and the solution would be the same. Since the optimality criterion is now a function of the reduced set of constants, it is specified by them.

Using the fact that the integrals over $\pm\infty$ will be twice the value of the integrals over $-\infty$ to 0 and the isoperimetric constraint [8], we form a composite functional $\Psi(x, f, f', f'')$ using Lagrange multipliers:

$$\Psi(x, f, f', f'') = f''^2(x) + \lambda_1 f^2(x) + \lambda_2 f'^2(x) + \lambda_3 x^2 f^2(x) + \lambda_4 f(x) \tag{7}$$

We want to minimize the integral of this functional, $\Psi(x, f, f', f'')$, denoted by $J[f]$, over all admissible functions $f(x)$.

$$J[f] = \int_{-\infty}^{0} \Psi(x, f, f', f'') dx$$

A necessary condition for a minimum is given by the Euler equation corresponding to the functional Ψ [8]:

$$f''''(x) - \lambda_2 f''(x) + (\lambda_1 + \lambda_3 x^2) f(x) = -\frac{\lambda_4}{2} \tag{8}$$

Any solution of this equation is a function of the unknown $\lambda's$. Therefore, before actually solving for f we have to identify the constraints on the $\lambda's$. $\lambda_3 > 0$ and $\lambda_4 < 0$ because their associated functions are to be minimized and maximized, respectively. Other constraints can be found by considering the second variation, $\delta^2 J$, of Ψ, which leads to:

$$\int_{-\infty}^{0} \left[g''^2 + \lambda_2 g'^2 + (\lambda_1 + \lambda_3 x^2) g^2 \right] dx \geq 0 \tag{9}$$

Equation (9) must hold for all admissible functions $g(x)$. Note that (9) is always true if λ_2 and λ_1 are positive because λ_3 is always positive (is associated with the integral which is to be minimized) and the rest of the terms are squares. For the case when either λ_1 or λ_2 is negative, the following conditions must hold for the variation to be always positive.

$$\left\{ \begin{array}{ll} \lambda_2^2 \leq 4\lambda_1 & \text{for } \lambda_1 \geq 0 \text{ and } \lambda_2 \leq 0 \\ \lambda_1^2 \leq \lambda_2 \lambda_3 & \text{for } \lambda_2 \geq 0 \text{ and } \lambda_1 \leq 0 \end{array} \right\} \tag{10}$$

Notice that both λ_1 and λ_2 being negative does not assure that the second variation, $\delta^2 J$, is always positive. We constrain our solution using (10). Again, when both λ_1 and λ_2 are positive, the second variation is positive; in that case we impose no further constraints.

C. Constrained Optimization

The solution, f, to the differential equation (8) is a function of five parameters, $\lambda_1, \lambda_2, \lambda_3, \lambda_4$, and $f'(0)$. $f'(0)$ has to be considered because of the unknown boundary condition, $f'(0) = c5$, which can be set to some *a priori* value because optimality is unaffected by filter amplitude. However, the constrained optimization (discussed later) goes faster if we let the slope of the filter at $x = 0$ be unknown.

Thus, our composite criterion, consisting of the product, $\Sigma\Lambda$, subject to MRC, was a function of $f(x)$ but is now a function of five variables. Values for the variables must be found which maximize the compos-

ite criterion. Since direct evaluation of the integrals appears difficult, we find a numerical solution to Equation (8) and then do a nonlinear constrained optimization of the composite criterion as a function of the five variables. This constrained optimization problem can be transformed into an unconstrained problem by the penalty method [9], which introduces penalties for violating each constraint. This forms a new function of the $\lambda's$ and $f'(0)$, $p(f)$, to be optimized:

$$p(f) = \Sigma\Lambda - \sum_{i=1}^{N} \mu_i P_i$$

P_i, the penalty for the i^{th} constraint, is nonzero if the constraint is violated. For greater μ_i, the corresponding constraint is satisfied more strictly. The complete paradigm is illustrated in Fig. 1. At each cycle we find the $\lambda's$ for which $p(f)$ is a maximum. For the next cycle we increase the $\mu's$ and use the $\lambda's$ of the previous step as the initial conditions, continuing until the constraints are satisfied within acceptable error.

D. Optimal Operators

Table 1 tabulates the optimum values of the λ parameters, $\Sigma\Lambda$, and MRC. Using Table 1 one can pick the $\lambda's$ corresponding to one's edge detection filter requirements in terms of $\Sigma\Lambda$ and MRC, and solve Equation (8) for the optimal filter. Notice that the filter corresponding to MRC of 2.685, for which we present results below, has the highest value for the product of $\Sigma\Lambda$ and MRC.

Plots of the optimal filter responses for various MRC are shown in Fig. 2. Notice the optimal filter decays more slowly for lower MRC than for higher MRC. As the MRCincreases, the operator gets tighter in spatial spread; the MRC term forces the solution to decay faster. As MRC approaches zero the filter tends to a signum function, or an infinite difference of boxes, which is to be expected because that maximizes the $\Sigma\Lambda$ product and produces the matched filter (use Schwarz's inequality).

IV. Recursive implementation

One of the most important criteria for an edge detector, from a practical viewpoint, is computational efficiency. We use the fact that IIR digital filters have infinite response to design a recursive approximation to our optimal detector because an IIR implementation requires a fixed (and reasonable) number of multiplications independent of scale. The filter with MRC, the mean distance between spurious maxima in units of rms filter width, equal to 2.685 is chosen to be approximated.

A. 1D Approximation

Our filter is obviously non-causal. We achieve full non-causal filtering by using two causal filters as defined below. The total filter response is a sum of two halves, $f_-(x)$ and $f_+(x)$. The one dimensional convolution of a signal, $s(x)$, by $f(x)$ can be realized by summing the response of a causal IIR filter, having impulse response $f_+(x)$, to the signal, $s(x)$, and its time (or space, really) reversed counterpart, $s(-x)$. The sampled version, $f(n)$, is to be approximated by an IIR filter. $f_+(x)$ is well approximated by (see Fig 3), in the minimum squared error sense:

$$h(x) = Ae^{-\alpha x}(cos(\phi) - cos(\beta\alpha x + \phi)) \text{ for } \alpha \geq 0 \qquad (11)$$

Although there is no theoretical justification for this expression, it forms a nice fit and involves the fewest number of terms we could reasonably expect. Introducing more exponentially damped sinusoids would improve the fit, but the increased complexity is unwarranted. This approximation has a squared error of 0.051, while the first derivative of a Gaussian has an error of 0.4. Roughly speaking, this approximation has an order of magnitude less residual error and is more easily implemented than the Gaussian derivative. In fact, the Gaussian derivative *itself* would then have to be approximated.

B. Design Procedure

We implement a discrete version of the filter $h(x)$, denoted by $h(n)$:

$$h(n) = Ae^{-\alpha n\tau}(cos(\phi) - cos(\beta\alpha n\tau + \phi)) \text{ for } n \geq 0 \qquad (12)$$

where τ is the sampling interval. The unit pulse response, $h(n)$, can be realized by an IIR digital filter having two zeros and three poles with the recursive form as:

$$y(n) = b_1y(n-1)+b_2y(n-2)+b_3y(n-3)+a_1x(n-1)+a_2x(n-2) \text{ for } n \geq 0 \qquad (13)$$

where $y(n)$ is the output sequence in response to input sequence $x(n)$ and

$$b_1 = e^{-\alpha\tau}(1 + 2cos(\beta\alpha\tau)) \; ; \quad b_2 = -b_1e^{-\alpha\tau} \; ; \quad b_3 = e^{-3\alpha\tau}$$

$$a_1 = h(1) = Ae^{-\alpha\tau}(cos(\phi) - cos(\beta\alpha\tau + \phi))$$

$$a_2 = h(2) - b_1a_1 = Ae^{-2\alpha\tau}(cos(\phi) - cos(2\beta\alpha\tau + \tau)) - b_1a_1$$

Finally, non-causal filtering of the input sequence, $\{x(n)\}$, is achieved as follows. Denote the time-reversed input sequence as $\{x_r(n)\}$, the output in response to the input sequence as $\{y_+(n)\}$, and the response to the time-reversed part as $\{y_-(n)\}$. The final output sequence $\{y(n)\}$ is given by $y(n) = y_+(n) - y_{-r}(n)$ for every n, with $\{y_{-r}(n)\}$ the time-reversed version of $\{y_-(n)\}$ and the subtraction due to the odd-symmetry of the filter.

C. Extension to 2D

The 1D filter can be extended to 2-D by applying the optimal filter perpendicular to the edge and a low-pass projection function, which will average (and so lower the variance of) the noise, parallel to the edge. This is easy if we choose the projection function to be the integral of the edge detection filter. Observing that the slope of a surface in any direction can be determined from its slope in two orthogonal directions, we apply a 2-D separable filter for the x and y directions, $f_x(r,c)$ and $f_y(r,c)$ respectively, which are separable and can be expressed in the form $g(r)h(c)$ and $g(c)h(r)$, respectively. $g(x)$, the lowpass projection function, is the integral of $h(x)$ and has a shape very similar to a Gaussian (see Fig. 4). The sampled version $g(n)$ is given by:

$$g(|n|) = e^{-\alpha|n|\tau}\left[\frac{cos(\phi)}{\alpha} - \frac{\alpha cos(\alpha\beta|n|\tau + \phi) - \alpha\beta sin(\alpha\beta|n|\tau + \phi)}{\alpha^2 + \alpha^2\beta^2}\right]$$

As before the non-causal filter can be realized as the sum of two identical causal filters operating in opposite directions. The filter, $g(n)$, with z-transform $G(z)$, can be realized by a recursive filter having three poles and two zeros:

$$G(z) = \frac{a_0 + a_1z + a_2z^2}{1 - b_1z - b_2z^2 - b_3z^3}$$

where

$$b_1 = e^{-\alpha\tau}(1 + 2cos(\beta\alpha\tau)) \; ; \quad b_2 = -b_1e^{-\alpha\tau} \; ; \quad b_3 = e^{-3\alpha\tau}$$

$$a_0 = g(0) = \frac{cos(\phi)}{\alpha} - \left(\frac{\alpha cos(\phi) - \alpha\beta sin(\phi)}{\alpha^2 + \alpha^2\beta^2}\right)$$

$$a_1 = g(\tau) - b_1a_0 \; ; \quad a_2 = g(2\tau) - b_1g(\tau) - b_2a_0$$

With $\{x(n)\}$ and $\{x_r(n)\}$ the input sequence and its time-reversed counterpart, respectively, let the outputs be $\{y_+(n)\}$ and $\{y_-(n)\}$, respectively. Then the final output is $y(n) = y_+(n)+y_{-r}(n)-x(n)$ for every n, where $\{y_{-r}(n)\}$ is the time reversed version of $\{y_-(n)\}$. The input value is subtracted because otherwise the value of the effective impulse response is twice its value at $n = 0$.

To summarize, our overall edge detection filter operates as follows. We first apply the non-causal projection filter, $G(z)$, to the rows of the

image and the non-causal edge detection filter, $H(z)$, to the columns. This estimates a smoothed image gradient in the row direction. Similarly, we apply $G(z)$ to the columns and $H(z)$ to the rows, to get an estimate of the smoothed gradient in the column direction. The number of multiplications per pixel is 40, and is independent of the size (scale) of the filter used.

From the two gradient directions we can estimate the maximum gradient direction, which is the direction perpendicular to the edge. We then use non maximum suppression on the gradient magnitude in the maximum gradient direction. Hysteresis thresholding based on a probability of error criterion can then be performed (see [10] for more details) to clean up the resulting edges.

V. Results

Because of space restrictions the number of illustrations has been severely cut. For a more detailed performance analysis the reader is referred to [11].

A. Operator Performance

To demonstrate quantifiable results, we created synthetic images with various signal to (zero-mean, white, Gaussian) noise ratios. Figure 5 shows a sequence of noisy images and the corresponding edge detection filter output. No thresholding has been done; we have plotted *all* the local maxima. Note how well the operator performs in the presence of very adverse noise. Using lower values of the scale parameter, α, flags fewer spurious edges, but the localization of the edges detected is impaired by the increased width of the operator. Higher values of α will produce more edges, but the true edges will be located more precisely.

The performance of the operator on a Magnetic Resonance Image (MRI) of a coronal slice of a human brain is shown in Fig. 6 with $\alpha = 0.5525$. The next test image is shown in Fig. 8. This is a very difficult image from an edge detection standpoint because it exhibits several different edge signatures. The edges due to the hair are certainly not step like, while the stripes on the coat represent line edges. The background exhibits many ridge and ramp edges. The result with $\alpha = 0.55248$ is shown in Fig. 8(b). Most edges are detected and located with good precision, even when not step like. Post processing in the form of hysteresis thresholding based on the operator response estimate reduces the number of spurious responses drastically as shown in Fig. 8(c).

B. Performance Comparison with Other Operators

We now compare the performance of our operator with that of Deriche and with the first derivative of a Gaussian. The first derivative of a Gaussian is implemented as a truncated Gaussian convolution followed by central differencing. The algorithm for the Gaussian convolution used is due to Sotak and Boyer [12] in their implementation of the Laplacian-of-Gaussian operator. It should be pointed out that the Gaussian derivative as implemented is necessarily an approximation, suggesting that we may find some difference between expected and observed performance.

Our operator has $\Sigma\Lambda = 1.21$ and $MRC = 2.8$. To construct the most meaningful comparison, we choose the Deriche operator having the same $\Sigma\Lambda$ product. The Deriche operator is given by

$$d(x) = -e^{-\alpha|x|}\sin(\omega x)$$

Letting $\alpha = m\omega$ we have

$$\Sigma\Lambda = \frac{2m}{\sqrt{m^2+1}} \text{ where } m \geq 0 \tag{14}$$

$$MRC = \sqrt{\frac{2m^2(m^2+1)^2}{(5m^2+1)(6m^4+3m^2+1)}} \text{where } m \geq 0 \tag{15}$$

The shape of the Deriche operator is characterized by m alone. The maximum value of MRC for the Deriche operator is about 2.6. Maximum $\Sigma\Lambda$ achievable is 2.0, as $m \longrightarrow \infty$, when $MRC = 1.623$. The value of m required for our purposes is 0.763, and the corresponding MRC is 2.48. The scaling constants are chosen so the operators have the same rms width. Equating the widths of the two filters, we find that α of our operator and ω of the Deriche operator are related by $\frac{\alpha}{\omega} = 1.03336$.

The filter performance measures $\Sigma\Lambda$ and MRC can be plotted to provide insight into their behavior. For this we refer to Fig. 7. The solid line corresponds to our operator, the dashed line corresponds to the Deriche operator, and the dotted line is the Canny operator using the r.m.s. width of the filter to calculate the MRC criterion. The "dot and dash"ed line corresponds to Canny's optimal filters using the finite extent of the filter as its width in the expression of the MRC[1]. The curve for the Deriche operator terminates at $\Sigma\Lambda = 2.0$, the maximum $\Sigma\Lambda$ possible for that filter. Our filter characteristic lies above that of the Deriche operator and Canny's operator for all values of the multiple response criterion. This means that ours has highest $\Sigma\Lambda$ product, and hence is the best of them. The dot in Fig. 7 corresponds to the first derivative of a Gaussian as an infinite length filter with $\frac{1}{\alpha\sigma} = 0.9255092$. The similarity of our optimal operator to the Gaussian derivative is also reflected in the performance of the two operators (see Figs. 9c and 9b).

Although there is not a marked difference between our operator and the first derivative of a Gaussian, the same cannot be said about the Deriche operator, which in some instances exhibits two responses to a single edge. Note the double response to the lake edge in Fig. 9(d). The same can be said about other prominent edges; the Deriche operator marks as edge points some which the other operators do not[2].

It is evident from the edge images that our operator outperforms the other two. Theoretically, the first derivative of a Gaussian should have the fewest number of spurious responses because it has $MRC = 3.245$, whereas our operator has $MRC = 2.8$ and the Deriche operator has $MRC = 2.4$[3]. Of course, since all of our implementations (except the Deriche operator) are approximations to the theoretical filters, there are some performance penalties. Nonetheless, our filter is both superior theoretically and more easily approximated, than the Gaussian derivative.

VI. Conclusion

We derived an optimal infinite impulse response filter for detecting step edges. The filter was approximated by an IIR digital filter, the implementation of which is simple, computationally efficient (constant time), and amenable to hardware realization. We presented results with extremely noisy images, indicating the excellent performance of our operator.

[1] We can not really compare the Canny operator with the other filters because it is a finite filter while the other two are infinite length filters. We include it for completeness.

[2] We should point out that the first derivative of a Gaussian performs better at the image boundaries because the algorithm of Sotak and Boyer incorporates d.c. padding [12] to prevent border erosion. Thus, the poor performance of both our operator and that of Deriche along the image border is an artifact of the finite size of the image, and not a drawback of the operators *per se*.

[3] We compare our operator with $MRC = 2.8$ to the Gaussian derivative on the basis of equivalent rms width. We should note, however, that our operator for $MRC = 3.245$, the same as the Gaussian derivative, has $\Sigma\Lambda = 1.024$, while the Gaussian derivative has $\Sigma\Lambda = 0.92$.

VII. References

[1] L. S. Davis, "A survey of edge detection techniques," *Computer Graphics and Image Processing*, vol. 4, no. 3, pp. 248–270, September 1975.

[2] M. Brady, "Computational approaches to image understanding," *Comput. Surveys*, vol. 14, no. 1, pp. 3–71, March 1982.

[3] V. S. Nalwa and T. O. Binford, "On detecting edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 699–714, November 1986.

[4] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–714, November 1986.

[5] R. Deriche, "Using Canny's criteria to derive a recursively implemented optimal edge detector," *International Journal of Computer Vision*, pp. 167–187, 1987.

[6] S.O.Rice, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, no. 24, pp. 46–156, 1945.

[7] V. Torre and T. A. Poggio, "On edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 2, pp. 147–163, March 1986.

[8] R.Courant and D.Hilbert, *Methods of Mathematical Physics*. New-York: Wiley Interscience, 1953.

[9] D. G. Luenberger, *Introduction to Linear and Non-Linear Programming*. Reading, MA: Addison-Wesley, 1973.

[10] J. Canny, "Finding edges and lines in images," Technical Report AI-TR-720, MIT Artificial Intelligence Laboratory, June 1983.

[11] S. Sarkar, "Optimal, efficient detection and low level perceptual organization of image edge features," M.S. Thesis, The Ohio State University, 1990.

[12] G. E. Sotak and K. L. Boyer, "The Laplacian-of-Gaussian kernel: A formal analysis and design procedure for fast, accurate convolution and full-frame output," *Computer Vision, Graphics, and Image Processing*, vol. 48, no. 2, pp. 147–189, 1989.

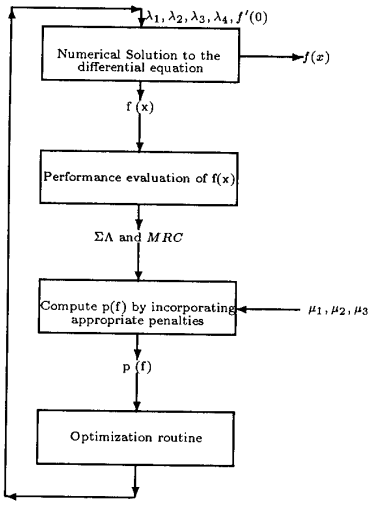

Figure 1: Flowchart followed in deriving the operator.

Table 1: Parameters and filter performance measures for various values of Signal to Noise Ratio / Localization product and Multiple Response Criterion.

	Optimal filter parameters							
n	MRC	$\Sigma\Lambda$	$(\Sigma\Lambda)MRC$	λ_1	λ_2	λ_3	λ_4	$f'(0)$
1	0.2	6.162	1.232	8.6858	11.279	0.0743	-11.459	6.3056
2	0.5	3.992	1.996	2.6726	5.2331	0.0942	-10.464	7.4101
3	1.0	2.676	2.676	2.3779	4.3177	0.2717	-9.5082	9.143
4	1.2	2.442	2.930	2.2315	4.1450	0.4354	-9.4538	9.175
5	1.5	2.159	3.239	3.0621	5.0021	1.7579	-10.6563	8.202
6	1.7	1.976	3.359	2.6585	4.0986	2.3317	-10.4989	8.3580
7	1.9	1.789	3.399	2.4594	3.4480	3.0531	-10.0365	8.4486
8	2.3	1.483	3.403	1.8121	1.0544	3.4169	-10.1624	8.8656
9	2.5	1.367	3.418	1.4842	0.5160	5.8983	-9.6812	8.9019
10	2.685	1.348	3.620	2.0399	-1.1554	4.3156	-9.1171	9.4152
11	2.8	1.224	3.427	1.4598	-1.1813	7.8178	-9.12822	9.4285
12	3.0	1.157	3.471	1.7821	-2.3817	4.8560	-8.7322	9.8911
13	3.245	1.024	3.322	2.6481	-3.2314	2.6771	-7.9314	12.1645

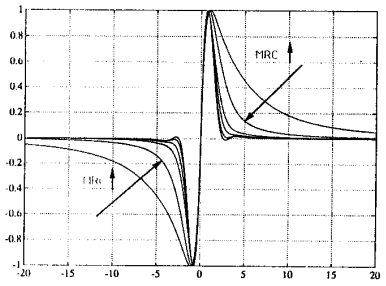

Figure 2: Optimal operators. The arrow indicates increasing MRC.

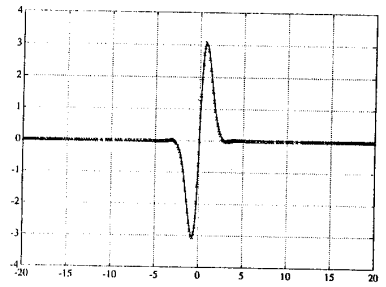

Figure 3: Plot of the optimal filter (solid line) and the approximating exponentially damped raised sinusoid (x's).

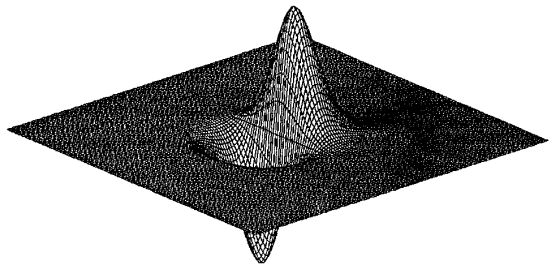

Figure 4: Perspective plot of the optimal 2-D edge operator.

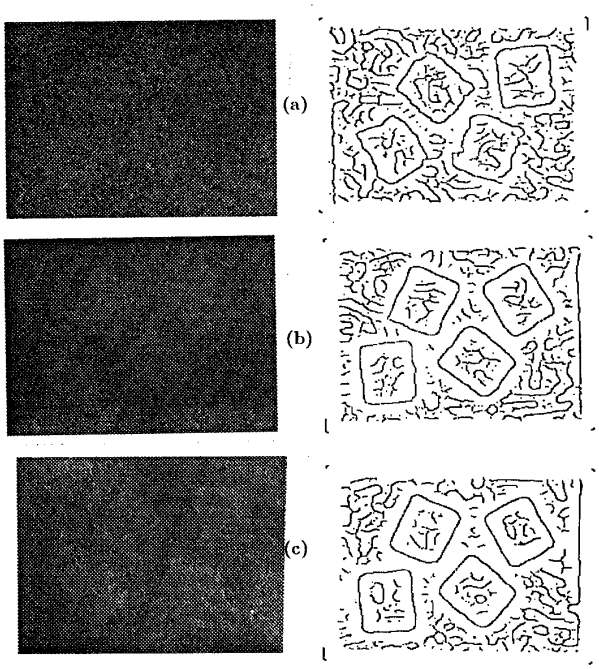

Figure 5: Performance of the operator derived in this paper in the presence of various level of noise (a) SNR=0.35 with $\alpha = 0.13822$ (b) SNR=0.707 with $\alpha = 0.230$ (c) SNR=1.0 with $\alpha = 0.230$.

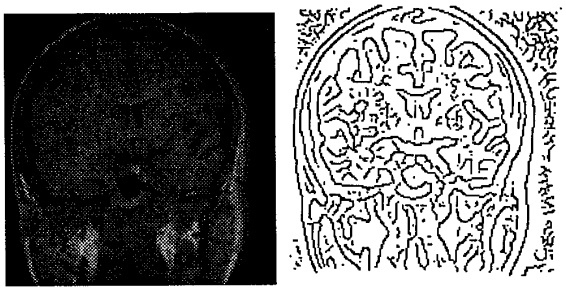

Figure 6: Magnetic Resonance Image of the brain; coronal slice.

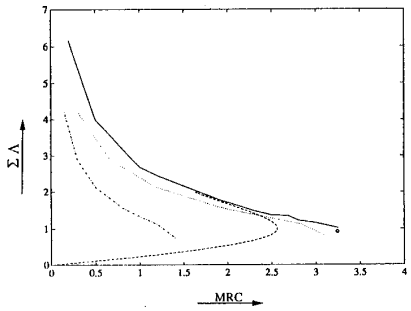

Figure 7: Plot of $\Sigma\Lambda$ vs. MRC for (1) Deriche operator (dashed line), Canny operator (2) (dotted line) using the r.m.s. extent to evaluate the MRC (3) ("dot and dash"ed line) using the extent of the finite filter to evaluate the MRC (4) The operator derived in this paper (solid line). The dot corresponds to the first derivative of Gaussian.

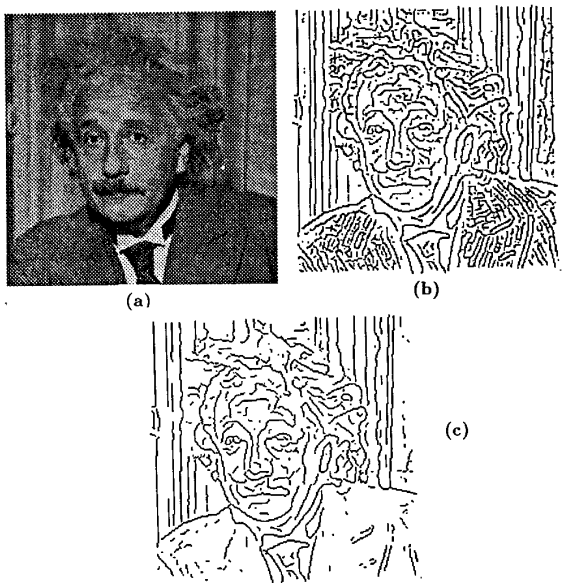

Figure 8: (a) Picture of Albert Einstein. (b) Edges detected with the optimal operator detected in this paper, $\alpha = 0.55248$. (c) Edges after hysteresis thresholding.

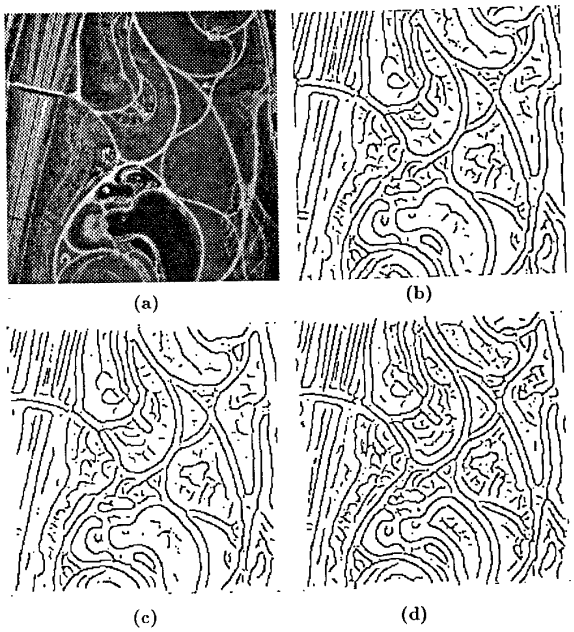

Figure 9: (a) Aerial image used. (b) Edges detected with the first derivative of a Gaussian, σ=3.13. (c) Edges detected with the operator proposed in this paper, with parameters $\alpha = 0.3453058, \beta = 1.201178310$ and $\phi = 0.770567179$.(d) Edges detected by Deriche operator using the following parameters, $m = 0.763, \omega = 0.334162$.