JOURNAL OF COMPLEXITY 4, 106–123 (1988)

A Regularized Solution to Edge Detection

T. POGGIO, H. VOORHEES, AND A. YUILLE

Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 545 Technology Square, Cambridge, Massachusetts 02139

We assume that edge detection is the task of measuring and localizing changes of light intensity in the image. As discussed by V. Torre and T. Poggio (1984), "On Edge Detection," AI Memo 768, MIT AI Lab), edge detection, when defined in this way, is a problem of numerical differentiation, which is ill posed. This paper shows that simple regularization methods lead to filtering the image prior to an appropriate differentiation operation. In particular, we prove (1) that the variational formulation of Tikhonov regularization leads to a convolution filter, (2) that the form of this filter is similar to the Gaussian filter, and (3) that the regularizing parameter λ in the variational principle effectively controls the scale of the filter.

1. INTRODUCTION

Edge detection does not have a precisely defined goal. The word "edge" itself, which refers to physical properties of objects, is somewhat of a misnomer. Several years of experience have shown that the ideal goal of detecting and locating physical edges in the surfaces being imaged is very difficult and still out of reach (for a review see Brady, 1982; Hildreth, 1985). Edge detection has come to be defined as the first step in this goal of detecting physical changes such as object boundaries—the operation of detecting and locating changes in intensity in the image. Other processes which operate on these measurements of intensity changes will then group boundaries and attempt to label and characterize them in terms of the properties of the 3-D surfaces. A solution to the problem of detecting and correctly characterizing physical edges requires, in general, high level knowledge and interaction with several early vision modules; it cannot be fully obtained just in terms of operations on image intensities.

In this narrow sense, we assume that edge detection—this first step in processing the image—is a process that measures, detects, and localizes

0885-064X/88 $3.00

changes of intensity. We assume, in particular, that derivatives must be estimated correctly to label the critical points in the image intensity array, characterize their local properties (are they minima or maxima or saddle points?), and begin to relate intensity changes to the underlying physical process (are they shadow edges or depth discontinuities?). As a consequence, several different types of derivatives of the image—and not only the Laplacian (Torre and Poggio, 1984)—possibly at different scales, may have to be estimated.[1] From this point of view edge detection is a problem of numerical differentiation of images (Torre and Poggio, 1984). The problem is not straightforward, and attempts over many years have proven its difficulties. Considered as a problem of numerical differentiation, edge detection turns out to be an ill-posed problem. As discussed by Poggio and Torre (1984), mathematically ill-posed problems are problems where the *solution* either *does not exist* or is *not unique* or *does not depend continuously on the data.*

Numerical differentiation is a (mildly) ill-posed problem because its solution does not depend continuously on the data. It is therefore natural to try to solve this problem by using regularization techniques developed in recent years for dealing with mathematically ill-posed problems. The problem can be regularized by the use of a wide class of filters (Torre and Poggio, 1984, Sect. 2.4; see also Duda and Hart, 1973). In the following section we consider two specific regularizing operators which are in a sense very natural and very simple.

2. Regularizing Edge Detection

To regularize an ill-posed problem and make it well posed, one has to introduce generic constraints on the problem. In this way, one attempts to force the solution to lie in a subspace of the solution space, where it is well defined. The basic idea of regularization techniques is to restrict the space of acceptable solutions by choosing the function that minimizes an appropriate functional. We consider here *standard regularization theory* to find a well-behaved intensity z (so that it can be later differentiated) from discrete and noisy data y. The problem is thus to regularize the problem of finding z from the data y such that $Az = y$. In the case of edge detection considered as numerical differentiation, we want an approximation z to the intensity data y_i at sample points x_i that is well behaved under differentiation. Thus we consider an operator A which samples the function z on the lattice such that $Az|_{x_i} = z|_{x_i}$ for $i = 1, \ldots, N$. Standard regularization as formulated by Tikhonov transforms the problem into the following variational principle:

[1] A very similar problem arises in the characterization of surface properties—in particular their differential properties—from depth data.

Find z that minimizes

$$|Az - y|^2 + \lambda|Pz|^2, \tag{1}$$

where λ is a regularization parameter and P is a stabilizing operator. Thus, the regularization parameter λ controls the compromise between the degree of regularization of the solution and its closeness to the data.

The discrete problem can be formulated rigorously as an ill-posed problem. Bertero *et al.* (1986) discuss in detail this formulation and its connection with other techniques. The problem is then to find a suitable norm and a suitable stabilizing functional $|Pf|$. It is natural to choose for P the simplest form of Tikhonov's stabilizing functionals (Tikhonov and Arsenin, 1977)

$$|Pz|^2 = \int \sum_{r=0}^{p} c_r(\xi) \left(\frac{d^r z}{d\xi^r}\right)^2 d\xi,$$

where $c_r(\xi)$ are non-negative weighting factors. In 1-D, we may choose $P = d^k/dx^k$ and the normal L^2 norm. For $k = 2$ this choice corresponds to a constraint of smoothness on the approximated intensity profile z, with $|Pf| = d^2f/dx^2$. Its *physical justification* is that the noiseless image has to be smooth in the sense that its derivatives must be bounded because the image is band limited by the optics. Band-limited functions have bounded derivatives because $f' \leq \Omega M$, where $M = \sup F(\omega)$, Ω is the cut-off frequency, and $F(\omega)$ is the Fourier transform of $f(x)$. Notice that for this simple reason it is meaningless to speak of edges as discontinuities in the *image* intensities. Physically, the constraint of smoothness allows us to effectively eliminate the noise that creeps in after or during the sampling and transduction process and makes the operation of differentiation unstable. We stress that this is not the only stabilizing functional possible for this problem, although it is probably the simplest one (Torre and Poggio, 1984). The order of the stabilizer does not have a great influence (Tikhonov and Arsenin, 1977), but see Appendix 1. Our results (later reported by Terzopoulos, 1986) can be very easily generalized to all Tikhonov stabilizers.

2.1 *Standard Regularization Method, Approximating Splines, and the Edge Detection Filter*

Tikhonov's regularization method leads to the following problem: Find the f that minimizes

$$\sum (y_i - f(x_i))^2 + \lambda \int (f''(x))^2\, dx, \tag{2}$$

where λ is the regularization parameter which can be found as described later. This problem was considered originally by Reinsch (1967) in the case of numerical differentiation and by Schoenberg (1964) for the problem of graduation (in the original Reinsch article, however, λ is a Lagrange multiplier!). Both Schoenberg and Reinsch gave the solution in terms of approximating cubic splines. We prove that for most practical purposes, the approximating spline function can be obtained by convolving the data point y_i with the cubic spline *convolution* filter R shown in Fig. 2 (see also Appendixes 1 and 3; Poggio *et al.*, 1985). We then have the following result:

PROPOSITION. *The solution to Eq.* (2)—*the regularized solution to the problem of numerical differentiation—in the case of inexact data, can be obtained by convolving the data with a convolution filter which is* (*a*) *a cubic spline, and* (*b*) *similar to a Gaussian* (*it is in particular a Butterworth-like filter*).

The exact assumptions under which the proposition is valid are discussed in Appendix 1 and in Poggio *et al.* (1985). First, the data must be given on a regular grid (as is the case for an image). Second, the image data must either go to zero at infinity or be periodic. Under these conditions, the filtering operation is space invariant and linear (the Euler–Lagrange equations corresponding to the quadratic variational problem are linear). Thus the approximating spline can be obtained by a convolution operation. Note that the result that the regularizing operator corresponding to a quadratic variational principle is a convolution filter—for data on a regular grid and toroidal boundary conditions—is valid beyond the case of numerical differentiation, provided the boundary conditions are appropriate and the operator A of Eq. (1) is space invariant.

3. Regularization Parameter and Comparison with the Gaussian Filter

Figure 1 shows the filter R obtained by solving the variational principle Eq. (1) in Appendix 1. Its shape and size depends on the regularizing parameter λ. Figure 2a shows the first derivative of the filter for different

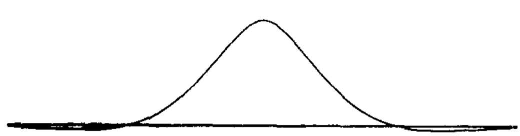

FIG. 1. Filter R derived by regularization principles.

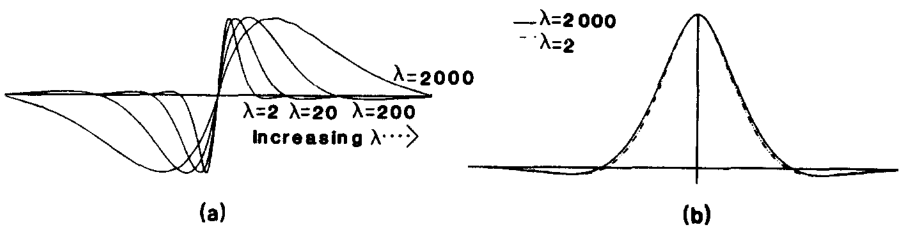

FIG. 2. Filter R and first derivative for different values of regularizing parameter λ. (a) λ affects size of (the first derivative of) R but (b) does not affect shape of R. (Amplitudes are normalized in (a); both amplitudes and widths are scaled linearly and independently in (b) for comparison.)

values of λ. The continuous version of the filter, derived in Appendix 2, is practically indistinguishable from the discrete filter, as shown by numerical comparisons. The smoothing parameter λ controls the effective size of the filter. It does not significantly affect the shape of the filter, but only its size, as shown in Fig. 2b. Appendix 3 makes this more precise. Changing λ amounts to scaling the size of the filter up or down. If λ is small, smoothness is unimportant, and the filter will tend to be an interpolating filter and therefore be similar to a δ function. On the other hand, with a very large λ, the main weight is on smoothness, and the filter will tend to be very large. The continuous form of the filter suggests that the role of λ is similar to the role of σ for the Gaussian (for 1-D $\lambda \simeq \sigma^4$, as shown in Poggio *et al.* (1985); also in Appendix 3).

The regularization filters derived here appear to be quite similar to the Gaussian distribution from the point of view of a numerical implementation on the computer. The same is true in the continuous case for a noncausal Butterworth filters (the order depends on the order of the stabilizer). Graphs of the filter R, and its first and second derivatives are shown with those of the Gaussian in Fig. 3. Marr and Hildreth (1980; Hildreth, 1980) have argued that the Gaussian is an optimal smoothing filter for detection due to its localization properties in both the spatial and frequency domains. The fact that the Gaussian is quite similar to the optimal filter derived here using regularization principles provides further mathematical justification for the use of a Gaussian-like filter in edge detection. Notice that our derivation is more general and simpler than Canny's (who also derived a variational principle), while leading to a filter that is indistinguishable for all practical purposes. Furthermore, the spline filter provides directly the solution to the problem of designing a digital filter (a problem which has to be solved separately in the case of a strictly continuous derivation, such as for the Gaussian filter).

If the boundary conditions are not periodic or natural, then the deriva-

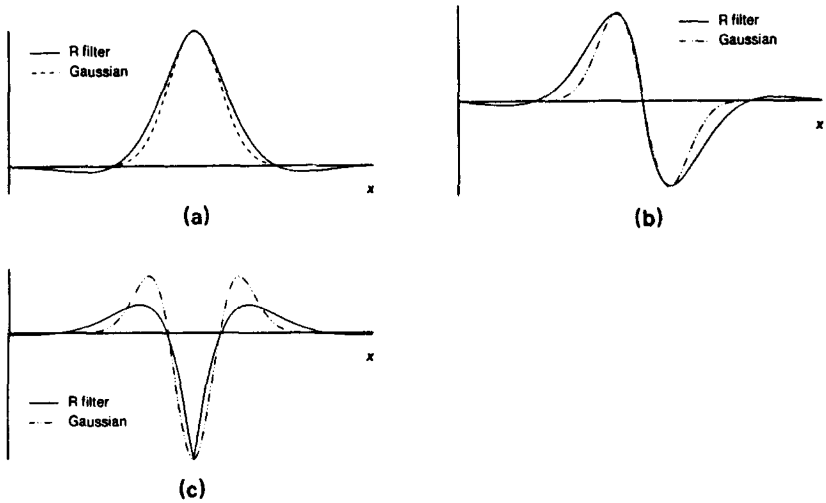

FIG. 3. Comparison of one-dimensional regularizing filter R with Gaussian: (a) zeroth, (b) first, and (c) second derivatives.

tion of Reinsch (see Poggio *et al.*, 1985, Appendix 1.1) provides the correct Green function for those boundary conditions.

4. DISCUSSION

Several questions and extensions suggest themselves in a natural way. Here we list some of them.

4.1. *Finding the Optimal* λ

Regularization theory can give the optimal value of λ if the errors on the smoothing criteria and the error of the approximations are known in advance. If the integral of f''^2 is less than E, and the sum of $(y_i - f(x_i))^2$ is less than ε, then $\lambda = \varepsilon/E$ (see Bertero, 1982; Tikhonov, 1963; Tikhonov and Arsenin, 1977). Normally, however, errors on the data or on the smoothness conditions are not known in advance. Regularization theory provides several methods for finding the optimal smoothing parameter λ under this circumstance, usually assuming Gaussian noise. We want to indicate here two main methods: (1) Tikhonov's method, for convolution-type problems, as is the case here, and (2) the cross-validation method and the generalized cross-validation method (Wahba, 1980). Geiger and Poggio (1987) have implemented a method based on the first approach to estimate the optimal scale of the filter from image data. Their solution also offers a tentative explanation for two perceptual phenomena.

4.2 *Relation with Other Edge Detectors*

Because of the close similarity of our spline filters to Gaussians, the edge detector that we derive in this paper is very similar to edge detectors proposed previously. Marr and Poggio (1977) proposed the difference of two Gaussians as an approximation to the second derivative of a Gaussian. Marr and Hildreth (1980; Hildreth, 1980) have shown that the second derivative of a Gaussian is, indeed, very close to the difference of Gaussians, as simple inspection of the diffusion equation shows. J. Canny's filter (Canny, 1983) is very close to the derivative of a Gaussian, and Haralick's cubic polynomial interpolant (Haralick, 1982) is again similar to Canny's filter. Our derivation justifies the use of a Gaussian or a filter very close to a Gaussian as the best filter for edge detection. Regularization theory yields derivative-of-Gaussian-like filters as the optimal filter in a simpler, more general, and, we believe, more rigorous way, than previous derivations. In particular, our result makes clear that the Butterworth filter regularizes the ill-posed problem of numerical differentiation. The regularizing constraint here is that the norm of the derivatives in the noise-free image is small.

It is interesting that we derive a filter very similar to Canny's, based on simpler and more general principles that are not restricted to the optimal detection of *step* edges. It is also interesting to note that the Laplacian of the regularization filter, like the Laplacian of the Gaussian, can be approximated by a difference of Gaussians (although not as well). While the Laplacian of the Gaussian is best approximated by a space constant ratio (the ratio of scales of the two Gaussians) $\gamma \approx 1.6$ (Marr and Hildreth 1980; Hildreth 1980; see also Grimson and Hildreth, 1985), increasing the ratio to $\gamma \approx 4$ results in a function which better fits the main (excitatory) lobe of the regularizing filter, as shown in Fig. 4.

4.3. *Extension to Two Dimensions*

Our approach can be extended to two dimensions by formulating the regularization principle in two dimensions. Instead of Eq. (2), one would then have the problem of minimizing

$$\sum_{i,j} (y_{ij} - f(x_{ij}))^2 + \lambda \iint (Pf)^2 \, dx \, dy. \tag{3}$$

The main problem is the choice of the operator P. If we consider a Tikhonov stabilizer (see Poggio and Torre, 1984), then a choice for P that is smooth enough to allow the use of second derivatives of the regularized image is

$$P = \nabla^2 \mathbf{\nabla}.$$

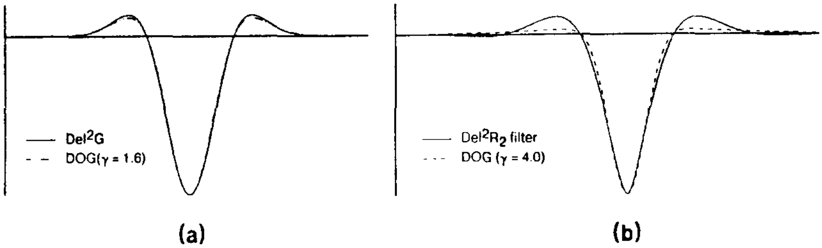

FIG. 4. Difference of two-dimensional Gaussians as approximations to (a) Laplacian of Gaussian, (b) Laplacian of two-dimensional regularizing filter R_2. The ratio between the scales (σ) of the two Gaussians is γ.

This choice is used in Appendix 2 to derive the filter for the two-dimensional continuous case. The filter is shown in Fig. 5. The choice of the derivative to be used on the filter is a *separate*, important issue that we do *not* address in this paper. Torre and Poggio (1984) discuss the properties of several two-dimensional differential operators, including the second directional derivative along the gradient.

If P is chosen to be the quadratic variation or the square Laplacian, the resulting approximations, known as thin plate splines (Wahba, 1980; Terzopoulos, 1984a), are not smooth enough for finding zeros of second derivatives of a function,[2] as implied by Terzopoulos (1984b).

Clearly, formulations of this type are also relevant for the problem of surface interpolation and approximation in the sense of Grimson (1982) and Terzopoulos (1984a). In the case of sparse data, which they considered, the variational principle does not lead to a convolution filter, although it does lead to a standard Green function. On a regular grid it leads to a convolution filter similar to the Gaussian. As a practical implication, evenly spaced surface data (for example, laser range data) may be interpolated or approximated effectively by Gaussian convolution. Hence, tasks which involve differentiating surface data, such as computing lines of curvature (Brady *et al.* 1985), could use the simpler convolution method to smooth the data. Since Reinsch's method (see Poggio *et al.*, 1985, Appendix 1.1) can deal with boundary conditions different from periodic ones, the corresponding Green function may be used to prevent smoothing across depth discontinuities.

The results of applying the Laplacian of the two-dimensional regularization filter and the Laplacian of a Gaussian to an image are shown in Fig. 6. As expected, due to the similarity of the two filters, both edge detection operators yield similar results.

[2] We are indebted to Demetri Terzopoulos for this remark.

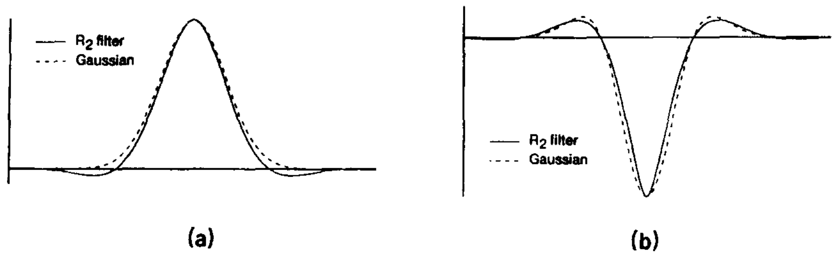

FIG. 5. Cross section of (a) two-dimensional regularization filter R_2 and (b) its Laplacian.

5. CONCLUSION: EDGE DETECTION AS DETECTION OF "INSUFFICIENT SMOOTHNESS"

We have considered part of the problem of edge detection as a problem of numerical differentiation. The discrete intensity data have to be approximated by a smooth function before differentiation. We found that this process is equivalent to convolving the data with the desired derivative of a generalized spline filter.

A complementary point of view is the following: The smooth function z that minimizes Eq. (1) will deviate locally from the ideal smoothness conditions by occasionally giving large absolute values of (Pz). Intuitively, these should identify nearby "edges" in the intensity image (note again that mathematical discontinuities cannot occur in the image because of its band-limited character). Thus, in our 1-D example, large values of f''^2 should point to a nearby edge. This corresponds to identifying edges by appropriate level crossings of the second derivative of a Gaussian filtered image (and successively localizing them by zero crossings). One can carry this argument even further. After identification of locations where the smoothness condition $(Pz)^2$ is "violated" too much (a threshold

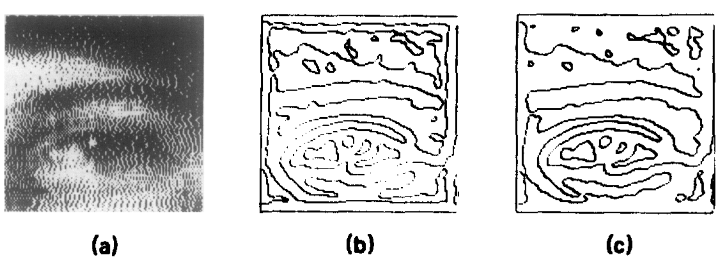

FIG. 6. Comparison of Laplacian of two-dimensional regularization filter and Laplacian of Gaussian as Edge Detectors: (a) image I, (b) zero crossings of $\nabla^2 R_2 * I$, (c) zero crossings of $\nabla^2 G_2 * I$.

must be chosen), "edges" placed there serve as boundaries for a second regularization step (convolution cannot be used now, at least not near the boundaries; instead we must use either the Green function or direct minimization of Eq. (1)).

APPENDIX 1. A DERIVATION OF THE REGULARIZATION FILTER FOR DISCRETE DATA IN ONE DIMENSION

In this derivation we show that the regularization method, described by Eq. (4) of the text, yields a convolution filter which is a cubic spline. Standard results from the calculus of variations guarantee that our solution has continuous second derivatives.

Again, the problem is to minimize

$$\lambda \int (f''(x))^2\, dx + \sum_i (f(x_i) - y_i)^2. \tag{1}$$

We find the minimum by sending $f(x) \to f(x) + \delta f(x)$ and setting the first variation of (1) to zero,

$$\lambda \int f''''(x)\, \delta f(x)\, dx + \sum_i (f(x_i) - y_i)\, \delta f(x_i) = 0. \tag{2}$$

This yields the Euler–Lagrange equation

$$\lambda f''''(x) + f(x) \sum_i \delta(x - x_i) = \sum_i y_i \delta(x - x_i). \tag{3}$$

So far, we have deliberately not specified boundary conditions. For infinite, or toroidal boundary conditions, the function $f(x)$ can be determined in terms of the $\langle y_i \rangle$ by a convolution filter if and only if the data points $\langle x_i \rangle$ are regularly spaced. This is because the system is then translation-invariant. We will show this explicitly and give a method for constructing the convolution filter.

The function in (1) is convex, and hence has a unique minimum, so there is a unique solution for $f(x)$ in (3). Thus, we only need to see whether a convolution can solve (3). We try a solution

$$f(x) = R(x) * y(x) = \int R(x - \zeta) y(\zeta)\, d\zeta, \tag{4}$$

where "$*$" denotes convolution, $R(x)$ is a filter, and $y(x) = \Sigma_i y_i \delta(x - x_i)$. We substitute (4) into (3) and obtain

$$\lambda \sum_i y_i R''''(x - x_i) + \sum_i \sum_j y_j R(x_i - x_j) \delta(x - x_i) - \sum_i y_i \delta(x - x_i) = 0.$$

We compare coefficients of y_i and obtain

$$\lambda R''''(x - x_i) + \sum_j R(x_j - x_i)\delta(x - x_j) - \delta(x - x_i) = 0. \tag{6}$$

If the $\langle x_i \rangle$ are not evenly spaced, then these equations are inconsistent, and no convolution filter exists. If the $\langle x_i \rangle$ are evenly spaced, then the set of equations in (6) reduces to a single equation

$$\lambda R''''(x) + \sum_j R(x_j)\delta(x - x_j) - \delta(x) = 0. \tag{7}$$

The solutions to (7) correspond to cubic splines "stitched" together at the points $\{x_i\}$. Let $R_i(x)$ denote the solution in the range $x_i < x < x_{i+1}$. We write

$$R_i(x) = \alpha_i x^3 + \beta_i x^2 + \gamma_i x + \delta_i. \tag{8}$$

The splines are stitched together so that $R(x)$, $R'(x)$, and $R''(x)$ are continuous at the points $\{x_i\}$. From (7), we see that R''' has a discontinuity of $-(1/\lambda)R(x_i)$ at x_i. It is straightforward to find the relations between the $R_i(x)$ and $R_{i-1}(x)$ in terms of the parameters α_i, β_i, γ_i, and δ_i.

This gives

$$\begin{aligned}
\alpha_{n+1} &= \alpha_n - \frac{1}{6\lambda} R(nh) \\
\beta_{n+1} &= \beta_n - \frac{nh}{2\lambda} R(nh) \\
\gamma_{n+1} &= \gamma_n - \frac{(nh)^2}{2\lambda} R(nh) \\
\delta_{n+1} &= \delta_n + \frac{(nh)^3}{6\lambda} R(nh),
\end{aligned} \tag{9}$$

where h is the spacing between the lattice points, $h = x_{i+1} - x_i$, and

$$R(nh) = \alpha_n(nh)^3 + \beta_n(nh)^2 + \gamma_n(nh) + \delta_n. \tag{10}$$

The continuous 1-D filter is derived in Poggio *et al.* (1985) as

$$R(x, \lambda) = \frac{1}{2\lambda^{1/4}} e^{-|x|/\sqrt{2}\lambda^{1/4}} \cos\left(\frac{|x|}{\sqrt{2}\lambda^{1/4}} - \frac{\pi}{4}\right). \tag{11}$$

Its shape is very similar to the discrete filter.

The choice of the stabilizer. The simplest form of Tikhonov's stabilizing functionals, $\langle f'(x)\rangle_2$ leads to a filter with a discontinuous first derivative at the origin. This means that the filtered signal will not necessarily be smooth enough to have well-defined second derivatives. We have found empirically that the filtered signal is typically very bumpy. For signals of noisy step edges there are often many zero crossings, violating Canny's criterion of single response to a feature. It is of course possible to have other Tikhonov stabilizers giving rise to filters of different degrees of smoothness.

APPENDIX 2: THE REGULARIZATION FILTER FOR CONTINUOUS DATA IN TWO DIMENSIONS

We can extend the results of the previous section to two dimensions. Our generalization of the smoothing function $\int (f''(x))^2\, dx$ is

$$\iint \nabla^2 \mathbf{\nabla} f(\boldsymbol{x}) \nabla^2 \mathbf{\nabla} f(\boldsymbol{x})\, d\boldsymbol{x}. \tag{1}$$

Courant and Hilbert (1953) show that the Euler–Lagrange equation is

$$\nabla^2 \nabla^2 \nabla^2 f = 0. \tag{2}$$

We write the regularization of the two-dimensional smoothing problem in the form

$$\iint (f(\boldsymbol{x}) - y(\boldsymbol{x}))^2\, d\boldsymbol{x} + \lambda \iint (\nabla^2 \mathbf{\nabla} f(\boldsymbol{x}) \nabla^2 \mathbf{\nabla} f(\boldsymbol{x}))\, d\boldsymbol{x}. \tag{3}$$

The Euler–Lagrange equations of the combined system are

$$\lambda \nabla^2 \nabla^2 \nabla^2 f(\boldsymbol{x}) + f(\boldsymbol{x}) = y(\boldsymbol{x}). \tag{4}$$

This equation is translation-invariant and so, for boundary conditions at infinity or periodic, the solution can be written as a convolution

$$f(\boldsymbol{x}) = R_2(\boldsymbol{x}) * y(\boldsymbol{x}), \tag{5}$$

where

$$\lambda \nabla^2 \nabla^2 \nabla^2 R_2(\boldsymbol{x}) + R_2(\boldsymbol{x}) = \delta(\boldsymbol{x}), \tag{6}$$

where $\delta(\boldsymbol{x})$ is the Dirac delta function. Again, observe that for small λ the filter $R_2(\boldsymbol{x})$ tends to the delta function.

To solve (6) we take its Fourier transform, and find

$$FR_2(\boldsymbol{\omega}) = \frac{1}{\lambda\omega^6 + 1}. \tag{7}$$

Equation (7) gives the Fourier transform of the continuous filter: it is of the form of a non-causal Butterworth filter. Equation (7) can be easily generalized to the general form of a Tikhonov stabilizer (instead of Eq. (1)).

From Eq. (6) one obtains

$$R_2(\boldsymbol{x}) = \frac{1}{2\pi} \int \frac{e^{+i\boldsymbol{\omega}\cdot\boldsymbol{x}}}{\lambda\boldsymbol{\omega}^6 + 1}\, d\boldsymbol{\omega}. \tag{8}$$

We express $\boldsymbol{x}$ and $\boldsymbol{\omega}$ in polar coordinates

$$\boldsymbol{x} = (r, \phi), \qquad \boldsymbol{\omega} = (w, \theta). \tag{9}$$

So

$$R_2(\boldsymbol{x}) = \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} \frac{e^{+i\omega \cos(\theta-\phi)}}{\lambda w^6 + 1}\, w\, dw\, d\theta. \tag{10}$$

Now

$$\int_0^{2\pi} e^{iwr \cos(\theta-\phi)}\, d\theta = \pi J_0(wr), \tag{11}$$

where J_0 is the zero order Bessel function. This gives

$$R_2(\boldsymbol{x}) = \frac{1}{2} \int_0^\infty \frac{J_0(wr)}{\lambda w^6 + 1}\, w\, dw, \tag{12}$$

that is, R_2 is the Hankel transform of $1/(\lambda w^6 + 1)$. This integral can be solved numerically, as shown in Poggio *et al.* (1985), Appendix 3.

APPENDIX 3: THE REGULARIZATION FILTER AS AN APPROXIMATION TO THE GAUSSIAN

In this appendix we show that the regularization filter approximates the Gaussian, both in one and two dimensions.

Comparison of one-dimensional filters. We show that the one-dimensional regularization filter is an approximate solution to the diffusion equation and therefore approximates a Gaussian.

The one-dimensional regularization filter is given by

$$R(x, \mu) = \frac{\mu}{2} e^{-\mu x/\sqrt{2}} \cos\left(\frac{\mu x}{\sqrt{2}} - \frac{x}{4}\right). \tag{1}$$

We expand this in a Taylor series in $\mu x/2$ to get

$$R(x, \mu) = \frac{\mu}{2\sqrt{2}}\left[1 - \left(\frac{\mu x}{\sqrt{2}}\right)^2 + O(\mu x)^3\right]. \tag{2}$$

This expression is valid when μx is small, i.e., when x is small compared to $\lambda^{1/4}$. In this case the first two terms which we denote by $\tilde{R}(x, \mu)$ are a good approximation to the function. We calculate

$$\frac{\partial^2 \tilde{R}}{\partial x^2} = \frac{-\mu^3}{2\sqrt{2}} \tag{3}$$

and

$$\frac{\partial \tilde{R}}{\partial \mu} = \frac{1}{2\sqrt{2}} + O(\mu x)^2, \tag{4}$$

which satisfy, to order $(\mu x)^2$, the equation

$$\frac{\partial^2 \tilde{R}}{\partial x^2} = -\mu^3 \frac{\partial \tilde{R}}{\partial \mu}. \tag{5}$$

Thus this function obeys the diffusion equation,

$$\frac{\partial^2 \tilde{R}}{\partial x^2} = \frac{\partial \tilde{R}}{\partial t} \tag{6}$$

with parameter $t = \frac{1}{2}\mu^{-2} = \frac{1}{2}\lambda^{1/2}$, in the region where μx is small. As μ decreases, this region gets larger and the region in which the function approximates a Gaussian increases.

This theoretical analysis supports the numerical results (for discrete data) which show that R can be approximated by a Gaussian. Furthermore, recalling that the standard deviation σ of the Gaussian is given by $\sigma = \sqrt{2t}$, the analysis shows that the standard deviation of the corresponding Gaussian is $\lambda^{1/4}$.

A comparison of R with the Gaussian $G = (e^{-x^2/2\sigma^2})$ can also be done directly in the Fourier domain.

Comparison of two-dimensional filters. We now consider the two-dimensional case. The regularized filter can be written in terms of a Fourier integral

$$R_2(\boldsymbol{x}, \lambda) = \frac{1}{2\pi} \int \frac{e^{-i\boldsymbol{\omega}\cdot\boldsymbol{x}}}{1 + \lambda\boldsymbol{\omega}^6} \, d\boldsymbol{\omega}. \tag{7}$$

We perform the transformation $\boldsymbol{\omega} \to \lambda^{1/6}\boldsymbol{\omega}$ to obtain, with $\underline{\mu} = 1/\lambda^{1/6}$,

$$R_2(\boldsymbol{x}, \underline{\mu}) = \frac{\underline{\mu}}{2\pi} \int \frac{e^{-i\boldsymbol{\omega}\cdot\underline{\mu}\boldsymbol{x}}}{1 + \boldsymbol{\omega}^6} \, d\boldsymbol{\omega}. \tag{8}$$

We expand the exponential in a power series

$$R_2(\boldsymbol{x}, \underline{\mu}) = \frac{\underline{\mu}}{2\pi} \int \frac{1}{1 + \boldsymbol{\omega}^6} \Big(1 + i\boldsymbol{\omega} \cdot \underline{\mu}\boldsymbol{x} + \frac{1}{2} (i\boldsymbol{\omega} \cdot \underline{\mu}\boldsymbol{x})^2 + O(\underline{\mu}^3\boldsymbol{x}^3)\Big) \, d\boldsymbol{\omega}. \tag{9}$$

Thus we have

$$R_2(\boldsymbol{x}, \underline{\mu}) = \frac{\underline{\mu}}{2\pi} \int \frac{1}{1 + \boldsymbol{\omega}^6} \Big(1 + \frac{1}{2} (i\boldsymbol{\omega} \cdot \underline{\mu}\boldsymbol{x})^2 + O(\underline{\mu}^3\boldsymbol{x}^3)\Big) \, d\boldsymbol{\omega}. \tag{10}$$

Note that the linear term drops out due to asymmetry of the integrand. Keeping the first two terms on the right-hand side of (10), and denoting this approximation by $\tilde{R}_2(\boldsymbol{x}, \underline{\mu})$, we calculate

$$\nabla^2 \tilde{R} = \frac{-\underline{\mu}^3}{\pi} \int \frac{\boldsymbol{\omega}^2 \, d\boldsymbol{\omega}}{1 + \boldsymbol{\omega}^6} \tag{11}$$

and

$$\frac{\partial \tilde{R}}{\partial \underline{\mu}} = \frac{1}{2\pi} \int \frac{1}{1 + \boldsymbol{\omega}^6} \, d\boldsymbol{\omega}. \tag{12}$$

Thus as before, the approximation $\tilde{R}_2(\boldsymbol{x}, \lambda)$ satisfies the diffusion equation with t proportional to $\lambda^{1/2}$. The exact function of proportionality can be calculated from (12).

Again, a direct comparison of the Gaussian with our regularizing filter R_2 is done easily in the Fourier plane. Both filters are circularly symmetric and therefore depend only on the radial frequency w.

ACKNOWLEDGMENTS

Several people contributed valuable comments to the memo version of this paper, especially Eric Grimson, Ellen Hildreth, and Demetri Terzopoulos. Thanks go to Elizabeth Willey for more than typing.

Research for this paper was done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory is provided, in part, by the Advanced Research Projects Agency of the Department of Defense. This research has also been partially supported by a gift from the Artificial Intelligence Center of Hughes Aircraft Corporation. H. Voorhees was supported by a Fellowship in Artificial Intelligence from NL Industries.

REFERENCES

BERTERO, M. (1982), Problemi lineari non ben posti e metodi di regolarizzazione, *in* "Problem non ben posti ed inversi," Istituto di Analisi Globale, Firenze.

BERTERO M., TORRE, V., AND POGGIO, T. (1986), "Ill-posed Problems in Early Vision," MIT AI Lab No. 924, November.

BLAKE, A. (1983), The least-disturbance principle and weak constraints, *Patt. Recogn. Lett.* **1,** 393–399.

BRADY, J. M. (1982), Computational approaches to image understanding, *Comput Surv.* **14,** 3–71.

BRADY, J. M., PONCE, J., YUILLE, A. L., AND ASADA, H. (1985), Describing surfaces, *in* "Proc. 2nd International Symposium, Robotics Research" (H. Harafusa, and H. Inoue, Eds.), MIT Press, Cambridge, MA.

CANNY, J. F. (1983), "Finding Edges and Lines in Images," AI-TR-720, MIT AI Lab, June.

COURANT, R., AND HILBERT, D. (1953), "Methods of Mathematical Physics," Vol. 1, Interscience, New York.

DUDA, R. O., AND HART, P. E. (1973), "Pattern Classification and Scene Analysis," Wiley, New York.

GEIGER, D., AND POGGIO, T. (1987), "The Lambda Parameter," AI-Memo 923, MIT AI Lab, January.

GEMAN, S., AND GEMAN, D. (1984), Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intelligence* **6,** 721–741.

GRADSHTEYN, I. S., AND RYZHIK, I. M. (1980), "Table of Integrals, Series and Products," Academic Press, New York.

GREVILLE, T. N. E. (1969), Introduction to spline functions, *in* "Theory and Applications of Spline Functions," pp. 1–36, Academic press, New York.

GRIMSON, W. E. L. (1981), "From Images to Surfaces," MIT Press, Cambridge, MA.

GRIMSON, W. E. L. (1982), A computational theory of visual surface interpolation, *Philos. Trans. R. Soc. London Ser. B* **298,** 395–427, 1982.

GRIMSON, W. E. L., AND HILDRETH, E. (1985), Comments on "Digital step edges from zero crossings of second directional derivatives," *IEEE Trans. PAMI* **7,** 121–127.

HARALICK, R. M. (1980), Edge and region analysis for digital image data, *Comput. Graphics Image Proc.* **12,** 60–73.

HARALICK, R. M. (1981), The digital edge, *in* "Proc. 1981 Conf. on Pattern Recognition and Image Processing," pp. 285–294, Dallas, TX.

HARALICK, R. M. (1982), Zero-crossings of second directional derivative edge operator, *in* "SPIE Proc. on Robot Vision," Arlington, VA.

HILDRETH, E. C. (1980), "Implementation of a Theory of Edge Detection," AI-TR-579, MIT AI Lab.

HILDRETH, E. C. (1985), "Edge Detection," AI-Memo 858, MIT AI Lab.

LUNSCHER, W. H. H. (1983), The asymptotic optimal frequency domain filter for edge detection, *IEEE Trans. PAMI* **6,** 678–680.

MARR, D. C., AND HILDRETH, E. C. (1980), Theory of edge detection, *Proc. R. Soc. London Ser. B* **207,** 187–217.

MARR, D., AND POGGIO, T. (1979), A computational theory of human stereo vision, *Proc. R. Soc. London Ser. B* **204,** 301–328 (also AI Memo 451, MIT AI Lab, 1977).

MARROQUIN, J. (1984), "Surface Reconstruction Preserving Discontinuities," Artificial Intelligence Lab. Memo, No. 792, MIT Press, Cambridge, MA.

MARROQUIN, J. (1985), "Probabilistic Solution of Inverse Problems," PhD thesis, MIT, Cambridge, MA.

MOROZOV, V. A. (1984), "Methods for Solving Incorrectly Posed Problems," Springer-Verlag, New York.

POGGIO, T., AND TORRE, V. (1984), "Ill-posed Problems and Regularization Analysis in Early Vision," AI Memo 773, MIT AI Lab.

POGGIO, T., TORRE, V., AND KOCH, C. (1985), Computational vision and regularization theory, *Nature* **317,** 314–319.

POGGIO, T., VOORHEES, H., AND YUILLE, A. (1985), "A Regularized Solution to Edge Detection," AI Memo 833, MIT AI Lab.

REINSCH, C. H. (1967), Smoothing by spline functions, *Numer. Math.* **10,** 177–183.

ROSENFELD, A., AND KAK, A. C. (1982), "Digital Picture Processing," 2nd ed., Academic Press, New York.

SCHOENBERG, I. J. (1946), Contributions to the problem of approximation of equidistant data by analytic functions, *Quart. Appl. Math.* **4,** 45–99, 112–141.

SCHOENBERG, I. J. (1964), Spline functions and the problem of graduation, *Proc. Natl. Acad. Sci. USA* **52,** 947–950.

SHANMUGAM, K. F., DICKEY, F. M., AND GREEN, J. A. (1979), An optimum frequency domain filter for edge detection in digital pictures, *IEEE Trans. PAMI* **1,** 37–49.

TERZOPOULOS, D. (1984a), "Multiresolution Computation of Visible-Surface Representations," Ph.D. thesis, Dept. of EECS, MIT, Cambridge, MA.

TERZOPOULOS, D. (1984b), Controlled smoothness stabilizers for ill-posed visual problems involving discontinuities, *in* "Proc. Image Understanding Workshop, SAIC."

TERZOPOULOS, D. (1984c), "Multigrid Relaxation Methods and the Analysis of Lightness, Shading, and Flow," Artificial Intelligence Laboratory Memo 803, MIT, Cambridge, MA.

TERZOPOULOS, D. (1986), "Regularization of Inverse Problems Involving Discontinuities," *IEEE Trans. PAMI* **8,** 613–624.

TIKHONOV, A. N. (1963), Solution of incorrectly formulated problems and the regularization method, *Sov. Math. Dokl.* **4,** 1035–1038.

TIKHONOV, A. N., AND ARSENIN, V. Y. (1977), "Solutions of Ill-Posed Problems," Winston and Sons, Washington, DC.

TORRE, V., AND POGGIO, T. (1984, 1986), "On Edge Detection," AI Memo 768, MIT AI Lab, 1984. Also available in a somewhat different version in *IEEE Trans. PAMI* **2,** 147–163.

WAHBA, G. (1980), "Ill-posed Problems: Numerical and Statistical Methods for Mildly, Moderately, and Severely Ill-posed Problems with Noisy Data," Tech. Report 595, University of Wisconsin, Madison.

WITKIN, A. (1983), "Scale-Space Filtering," Proc. IJCAI, 1019-1021, Karsruhe.

YUILLE, A. L., AND POGGIO, T. (1983a), "Scaling Theorems for Zero-crossings," AI Memo 722, MIT AI Lab.

YUILLE, A. L., AND POGGIO, T. (1983b), "Fingerprint Theorems for Zero-crossings," AI Memo 730, MIT AI Lab.