Uniqueness of the Gaussian Kernel for Scale-Space Filtering

Article *in* IEEE Transactions on Pattern Analysis and Machine Intelligence · February 1986

DOI: 10.1109/TPAMI.1986.4767749 · Source: IEEE Xplore

4 authors, including:

Michel Baudin

Grokcity Inc.

Uniqueness of the Gaussian Kernel for Scale-Space Filtering

JEAN BABAUD, ANDREW P. WITKIN, MICHEL BAUDIN, MEMBER, IEEE, AND RICHARD O. DUDA, FELLOW, IEEE

***Abstract*—Scale-space filtering constructs hierarchic symbolic signal descriptions by transforming the signal into a continuum of versions of the original signal convolved with a kernal containing a scale or bandwidth parameter. It is shown that the Gaussian probability density function is the only kernel in a broad class for which first-order maxima and minima, respectively, increase and decrease when the bandwidth of the filter is increased. The consequences of this result are explored when the signal—or its image by a linear differential operator—is analyzed in terms of zero-crossing contours of the transform in scale-space.**

***Index Terms*—Difference of Gaussians, Gaussian filters, multiresolution descriptions, scale-space filtering, signal description, waveform description**

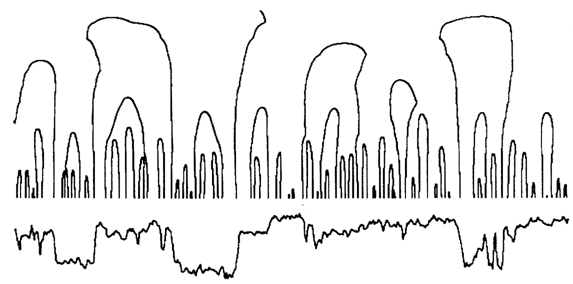


Fig. 1. A signal, and contours of $\phi_{xx} = 0$ in its scale-space image. The x-axis is horizontal; the coarsest scale is on top. The contours describe the behavior of inflections in the Gaussian-smoothed signal, as the standard deviation of the Gaussian is varied.

I. Introduction

It is often useful to describe a signal qualitatively by its extrema and those of its first few derivatives. However, the description obtained depends not only on the signal but on the scale of measurement, i.e., the size of the operator used to detect the features. At too fine a scale, one is swamped with extraneous detail. At too coarse a scale, important features may be missed, with those that survive severely distorted by the effects of excessive smoothing. Frequently there is no good basis for choosing the scale of measurement *a priori*. In fact, it may often be desirable to describe the same signal at more than one scale in the course of interpreting it.

Recognition of the problem of scale has spawned considerable interest in multiresolution descriptions, obtained by passing the signal through filters at several fixed sizes [4], [7]. However, these multiscale techniques share a fundamental shortcoming: they provide no means of relating the descriptions at different scales to one another, or of deciding which to use when. Marr and Hildreth [4] speculated that zero-crossings that spatially coincide over several channels are "physically significant," but this idea has not born fruit.

Scale-space filtering [10] is a qualitative signal description method that deals gracefully with the problem of scale by treating the size of the smoothing kernel as a continuous parameter. As the scale parameter is varied, the extrema or zero-crossings in the smoothed signal and its derivatives in general move continuously, although new extrema appear from time to time at singular points. The behavior of extrema over scale is characterized in terms of the *scale-space image*, the surface $\phi(x, y)$ swept out on the xy-half-plane when a signal $f(x)$ is convolved with a smoothing kernel $g(x, y)$, where y is a positive scale or "bandwidth" parameter.[1] If the events to be detected in f are the zero-crossings of $L(f)$, where L is a linear differential operator, then the behavior of zero-crossings over scale is characterized by the zero-crossing contours of $L(\phi)$. (See Fig. 1.) A unified and organized description is produced from the scale-space image by treating each contour as a single physical event, observed through a continuum of filters, rather than as a set of unrelated events. Each extremum is characterized by the coarsest scale at which it appears and by its location on the signal axis at the finest scale of observation (thus correcting the distorting effects of smoothing.) Finally, these events are used to construct a tree describing the successive partitioning of the signal into finer subintervals as new zero-crossings appear at finer scales.

These simplifications rest on a property of the scale-space image that is evident in Fig. 1: moving from coarse to fine scale, new zero-crossings appear, but existing ones never disappear. In other words, the scale-space contours are closed above but never below. Consequently, extrema observed at any scale may be localized by their projections at the finest available scale; and the partitioning of the sig-

Manuscript received October 15, 1984; revised August 15, 1985. Recommended for acceptance by W.E.L. Grimson.

J. Babaud, A. P. Witkin, and M. Baudin are with Schlumberger Computer Aided Systems, Palo Alto, CA 94304.

R. O. Duda is with Syntelligence, Sunnyvale, CA.

IEEE Log Number 8406215.

[1] By a *scale* parameter, we mean one for which larger values denote greater degrees of smoothing, e.g., the standard deviation of a Gaussian kernel. A *bandwidth* parameter is just the reciprocal of the scale parameter. In this paper, we use y to denote the latter.

nal by extrema moving from coarse to fine forms a strict hierarchy.

In this paper, we prove that over a broad class of kernels and signals this desirable property holds uniquely for the Gaussian. Roughly speaking, we show that the gaussian is the only kernel for which local maxima of ϕ always increase and local minima always decrease as the bandwidth of the filter is increased; with other kernels, there are "nonpathological" signals that violate this natural requirement. This property means that, for any well-behaved signal f, the peaks and valleys of $\phi(x, y)$ become monotonically more pronounced as the bandwidth y increases, and that zero-crossing contours observed at low bandwidth cannot in general vanish with increasing bandwidth.

In proving this result, we must be careful in stating the conditions on the families of signals f and kernels g that we admit. We must also pay particular attention to the case of zero-curvature extrema; such extrema can arise from signals that we would not like to call pathological. These might give rise to closed zero-crossing contours, although we have not so far discovered cases in which they do.

Thus, we begin by giving precise statements of the admissible families of signals and kernels, limiting our attention to first-order extrema, where the curvature is nonzero. We then develop general expressions for the partial derivatives of ϕ, which are used to express the monotonicity property. The proof of sufficiency is straightforward. The proof of necessity is obtained by constructing a special family of signals f that cause the monotonicity condition to be violated for all but Gaussian kernels. Finally, we note that the monotonicity condition is not satisfied at all extrema, and discuss the implications of these results.

II. The Mathematical Results

A. Signals and Kernels

The calculations that follow assume that we are dealing with signals and kernels we can meaningfully convolve, and that any linear differential operator can be applied to the result of this convolution. A general framework in which this is true is obtained by restricting our attention to infinitely differentiable kernels, vanishing at infinity faster than any inverse of polynomial, and to signals that are continuous linear functionals on the space of these kernels.

Exact definitions are found in [8] where such kernels are called "rapidly decreasing functions," and such signals "tempered Schwartz distributions." Unless otherwise specified, all limits are taken in the sense of weak convergence. Convolution, integration, and differentiation are meant for or with respect to Schwartz distributions.

This allows consideration of the most general kind of signals and kernels for which convolution is well defined. For example, every measurable function whose value is bounded by a polynomial is a tempered distribution. Provided it does not rise too fast at infinity, the signal can be as irregular as the user wants; its image by any differential operator is defined using distribution calculus.

We also need to define a first-order extremum of a twice-differentiable function f: $R \to R$ as a point $z \in R$ such that $\dot{f}(x) = 0$ and $\ddot{f}(x) \neq 0$. (By a first-order extremum of f we mean an extremum with nonzero curvature, i.e., a zero-crossing of the first derivative but not of the second.) It is very common in applications to overlook the possibility of extrema of higher orders. However, a function as simple as x^4 has a minimum at $x = 0$, for which the first nonzero derivative is of order four. The function

$$f(x) = e^{(1/x^2)}$$

is defined for $x \neq 0$, and extends by continuity to $x = 0$, where it has a strict minimum at which all its derivatives vanish. As we shall see, when there are higher-order extrema, even a Guassian kernel cannot guarantee good behavior.

B. The Uniqueness Theorem

We can now state the central result of this paper.

Proposition: Let $g(\cdot, y)$, for all $y > 0$ be a family of infinitely differentiable, rapidly decreasing kernels $R \to R$, satisfying the following conditions:

1) y is a bandwidth parameter for g. In other words, there exists a kernel h such that, for all $x \in R$ and $y > 0$,

$$g(x, y) = yh(xy).$$

2) g is symmetrical in x: for all $x \in R$

$$g(-x, y) = g(x, y).$$

3) For all $y > 0$, $g(\cdot, y)$ is normalized:

$$\int_{-\infty}^{\infty} g(u, y)\, du = \int_{-\infty}^{\infty} h(v)\, dv = 1.$$

4) There exists an integer p such that

$$h^{(2p)}(0) \neq 0.$$

Then, given any admissible signal f, the necessary and sufficient conditions for all first-order maxima (respectively minima) of the convolution

$$\phi(x, y) = f(x) * g(x, y) = \int_{-\infty}^{\infty} f(u)\, g(x - u, y)\, du$$

to increase (respectively decrease) monotonically as y increases is that $g(\cdot, y)$ be the Gaussian

$$g(x, y) = \frac{1}{\sqrt{2\pi}}\, y\, e^{-1/2(xy)^2}.$$

That is, the Gaussian guarantees that wherever

$$\phi_x = \frac{\partial \phi}{\partial x} = 0$$

and

$$\phi_{xx} = \frac{\partial^2 \phi}{\partial x^2} \neq 0$$

the *monotonicity condition*

$$\phi_y \phi_{xx} < 0$$

will be satisfied.

C. Comments

We have placed several restrictions on the family of kernels that we will consider. These limitations deserve some discussion.

First, we require that g be infinitely differentiable and rapidly decreasing so that we can admit distributions (such as the Dirac delta) as legitimate signals. This requirement excludes certain popular smoothing functions, such as the exponential

$$g(x, y) = \frac{y}{2} e^{-|xy|}$$

for which

$$h(x) = \tfrac{1}{2} e^{-|x|},$$

and the "moving-window" average

$$g(x, y) = \begin{cases} y/2 & |xy| < 1 \\ 0 & |xy| \geq 1. \end{cases}$$

for which

$$h(x) = \tfrac{1}{2} 1_{[-1,1]}(x).$$

However, such functions can be approximated arbitrarily closely in the sense of weak convergence by rapidly decreasing, infinitely differentiable functions. Alternatively, we could have replaced the Dirac distribution used in proving necessity by any of a number of well known continuous approximations at the cost of making the proof more cumbersome.

Condition 1) amounts to a definition of what we mean by a scale parameter. It ensures that two kernels of the family corresponding to two different values of y are not identical. The use of the word "scale" can be justified by noting that if $U(x)$ is obtained by convolving an arbitrary signal $f(x)$ with $h(x)$, then $U(xy)$ is the convolution of $f(xy)$ with $yh(xy)$. Thus $U(xy)$ is the image of $f(xy)$ by the scaled kernel.

Intuitively, condition 1) says that the scale parameter influences the kernel by stretching it along the x-axis while keeping its area invariant. Note that if $H(\omega)$ is the Fourier transform of the kernel h, then the Fourier transform of g is $H(\omega/y)$. Thus, y directly scales the frequency variable, and it is natural to call y the *bandwidth parameter*.

The symmetry condition 2) gives an equal weight to the "past" and the "future" of f, and thereby excludes causal filters from consideration. This constraint is appropriate for the purpose of signal description because the entire signal may be presumed to be available. We note in passing that the symmetry assumption implies that h is even, and hence that its first derivative $\dot{h}$ is odd.

Condition 3), that the kernel has unit area, gives the filter the desirable property that if the signal f is a constant, say, $f = a$, then $\phi = a$.

Condition 4) is needed in the proof for technical reasons. The requirement that not all the derivatives of h vanish at 0 is a mild one.

The essence of the proposition is that, as y increases, all the first-order maxima of $\phi(\cdot, y)$ increase and all its first-order minima decrease if and only if the kernel h is Gaussian. An example is given below showing that the Gaussian kernel does not satisfy a similar montonicity condition for higher order extrema, i.e., points $x \in R$ at which the first nonzero partial derivative of $\phi(x, y)$ in x is of order $2n$ with $n > 1$.

The convolution of any distribution f with $g(\cdot, y)$ inherits the regularity properties of $g(\cdot, y)$. Therefore, the infinite differentiability of h guarantees that all partial derivatives of $\phi(x, y)$ are continuous functions on the xy-half-plane defined by $y > 0$ and $x \in R$.

D. The Proof

1) Sufficiency of the Gaussian: In the next section we derive general expressions for the partial derivatives ϕ_y and ϕ_{xx} for any kernel. It is easy to show that the gaussian satisfies the monotonicity condition merely by substituting a Gaussian function into those expressions and noting that the standardized Gaussian

$$h(x) = \frac{1}{\sqrt{2\pi}} e^{(1/2)x^2}$$

satisfies the differential equation

$$\ddot{h}(x) = -[h(x) + x\dot{h}(x)].$$

However, a more illuminating proof can be obtained through observing that the Gaussian is the Green's function for the one-dimensional diffusion equation. In particular, let ϕ be viewed as a function of space x and time t, and let $f(x)$ be its initial value. Then, if ϕ satisfies

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{\partial \phi}{\partial t}$$

it is well known that ϕ is given by the convolution of f with the Gaussian kernel

$$g(x, 1/\sqrt{2t}) = \frac{1}{\sqrt{4\pi t}} e^{-(1/4t)x^2}.$$

Conversely, the convolution of $f(x)$ with this kernel must satisfy the diffusion equation. Thus, if we change variables and let $y = 1/\sqrt{2t}$, it follows that

$$\frac{\partial \phi}{\partial y} = \frac{\partial \phi}{\partial t}\frac{\partial t}{\partial y} = -\frac{1}{y^3}\frac{\partial \phi}{\partial t} = -\frac{1}{y^3}\frac{\partial^2 \phi}{\partial x^2},$$

which leads to

$$\phi_y \phi_{xx} = \frac{-1}{y^3}[\phi_{xx}]^2.$$

Thus, the condition $\phi_y \phi_{xx} < 0$ is satisfied not only at every first-order extremum of ϕ in x, but everywhere ϕ_{xx}

$\neq 0$. The sufficiency part of the proof is therefore trivial, and the remarkable part of the proposition is the necessity assertion. Its proof is more difficult, and requires the explicit form of the partial derivatives of ϕ.

2) The Partial Derivatives: The partial derivatives ϕ_{xx} and ϕ_y are easily computed.

$$\phi(x, y) = \int_{-\infty}^{\infty} f(u)\, g(x-u, y)\, du = y \int_{-\infty}^{\infty} f(u)\, h(y(x-u))\, du.$$

Thus

$$\phi_x = y^2 \int_{-\infty}^{\infty} f(u)\, \dot{h}(y(x-u))\, du,$$

$$\phi_{xx} = \int_{-\infty}^{\infty} f(u)\, \ddot{h}(y(x-u))\, du,$$

and

$$\phi_y = \int_{-\infty}^{\infty} f(u)\, [h(y(x-u)) + y(x-u)\, \dot{h}(y(x-u))]\, du.$$

These results can be placed in a more convenient form through the change of variables $v = y(x-u)$, which leads to

$$\phi(x, y) = \int_{-\infty}^{\infty} f\left(x - \frac{v}{y}\right) h(v)\, dv, \tag{1}$$

$$\phi_x = y \int_{-\infty}^{\infty} f\left(x - \frac{v}{y}\right) \dot{h}(v)\, dv, \tag{2}$$

$$\phi_{xx} = y^2 \int_{-\infty}^{\infty} f\left(x - \frac{v}{y}\right) \ddot{h}(v)\, dv, \tag{3}$$

and

$$\phi_y = \frac{1}{y} \int_{-\infty}^{\infty} f\left(x - \frac{v}{y}\right) [h(v) + v\dot{h}(v)]\, dv. \tag{4}$$

3) Necessity of the Gaussian: There are many kernel functions besides the Gaussian that satisfy conditions 1)–3). However, we shall now show that only the Gaussian satisfies the monotonicity condition for all signals f.

The following lemma will enable us to restrict our attention to kernels such that $h(0) \neq 0$ or $\dot{h} \neq 0$

Lemma: A kernel h satisfying conditions 1)–4) of the proposition and such that $h(0) = 0$ and $\ddot{h}(0) = 0$ cannot satisfy the monotonicity condition.

By condition 4), the derivatives of h at 0 cannot be all zero. The symmetry condition 2) then implies that the first nonzero derivative of h at 0 is of order $2p$ for a positive integer p, so that, in the neighborhood of 0,

$$h(x) = ax^{2p} + o(x^{2p}).$$

Choosing $f(x) = \frac{1}{2}[\delta(x-\epsilon) + \delta(x+\epsilon)]$ with $\epsilon > 0$, we have

$$\phi(x, y) = \frac{a}{2} y^{2p+1}[(x-\epsilon)^{2p} + (x+\epsilon)^{2p}] + o(y^{2p+1}x^{2p}).$$

f is symmetric, and therefore

$$\phi_x(0, y) = 0.$$

Furthermore,

$$\phi_{xx}(0, y) = 2p(2p-1)\, ay^{2p+1}\epsilon^{2p-2} + o(y^{2p+1}x^{2p-2})$$

and

$$\phi_y(0, y) = (2p+1)\, ay^{2p}\epsilon^{2p} + o(y^{2p}x^{2p}).$$

The monotonicity condition is violated, the lemma is proved, and we may assume henceforth that $h(0)$ and $\ddot{h}(0)$ are not both zero.

The remainder of the proof is obtained by constructing a signal $f(x)$ that causes all kernels meeting conditions 1)–4) to have at least one first-order extremum of ϕ in x at which the monotonicity condition is violated. There are many such signals, but the following is sufficient:

$$f(x) = \beta\delta(x) + \tfrac{1}{2}[\delta(x-1) + \delta(x+1)].$$

Here δ is the Dirac distribution, and β is a parameter at our disposal. From (2)–(4) it follows that

$$\phi_x(x, y) = y^2\{\beta\dot{h}(xy) + \tfrac{1}{2}[\dot{h}((x-1)y) + \dot{h}((x+1)y)]\},$$

$$\phi_{xx}(x, y) = y^3\{\beta\ddot{h}(xy) + \tfrac{1}{2}[\ddot{h}((x-1)y) + \ddot{h}((x+1)y)]\},$$

and

$$\phi_y(x, y) = \beta[h(xy) + xy\dot{h}(xy)] + \tfrac{1}{2}[h((x-1)y) + (x-1)y\, \dot{h}(x-1)y)] + \tfrac{1}{2}[h((x+1)y) + (x+1)y\, \dot{h}(x+1)y)].$$

Now both f and h are symmetric, so we can expect ϕ to have an extremum at $x = 0$. Since h and $\ddot{h}$ are even and $\dot{h}$ is odd, when $x = 0$ the above equations simplify to

$$\phi_x(0, y) = y^2[\beta\dot{h}(0) + \tfrac{1}{2}\dot{h}(-y) + \tfrac{1}{2}\dot{h}(y)] = 0,$$

$$\phi_{xx}(0, y) = y^3[\beta\ddot{h}(0) + \ddot{h}(y)],$$

and

$$\phi_y(0, y) = \beta h(0) + h(y) + y\dot{h}(y).$$

Thus

$$\phi_y(0, y)\, \phi_{xx}(0, y) = y^3[\beta\ddot{h}(0) + \ddot{h}(y)]\, [\beta h(0) + h(y) + y\dot{h}(y)].$$

Now suppose that we fix y and examine the behavior of this expression as a function of β. In general, it is a quadratic function of β with two real roots, one at

$$\beta_1 = -\frac{\ddot{h}(y)}{\ddot{h}(0)},$$

and the other at

$$\beta_2 = -\frac{h(y) + y\dot{h}(y)}{h(0)}.$$

Unless either a) there are no roots, or b) the two roots coincide, for any value of y there will be values for β for which the *monotonicity condition* is violated. The first possibility is excluded by condition 4) and the lemma. Since $\dot{h}(0) = 0$, we must have $h(0) \neq 0$. Thus, it is necessary that $\beta_1 = \beta_2$ for all values of y, i.e., that

$$\frac{\ddot{h}(y)}{h(y) + y\dot{h}(y)} = \frac{\ddot{h}(0)}{h(0)}$$

for all y.

Since $h(0)\,\ddot{h}(0)$ is the coefficient in β^2 of $\phi_y \phi_{xx}$, the *monotonicity condition* implies that $h(0)$ and $\ddot{h}(0)$ have opposite signs. Letting

$$a^2 = -\frac{\ddot{h}(0)}{h(0)},$$

we have

$$\frac{d}{dy}\left[\frac{dh}{dy}\right] = \ddot{h}(y) = -a^2[h(y) + y\dot{h}(y)]$$

$$= -a^2 \frac{d}{dy}[yh(y)].$$

Using the fact that $\dot{h}(0) = 0$ and integrating twice, we obtain

$$\frac{dh}{dy} = -a^2 y h(y)$$

$$\int_{h(0)}^{h(y)} \frac{dh}{h} = \ln\frac{h(y)}{h(0)} = -a^2 \int_0^y y\,dy = -\frac{1}{2}a^2y^2$$

or

$$h(y) = h(0)\, e^{-1/2a^2y^2}$$

This shows that the only kernel satisfying conditions 1)–4) and the *monotonicity condition* is Gaussian.

E. The Domain of Validity of the Proposition

1) Behavior of the Smoothed Signal at Higher-Order Extrema: To include higher order extrema, the *monotonicity condition* would have to be restated to say that wherever the first nonzero partial derivative of ϕ in x is of order $2p$, the first nonzero partial derivative of ϕ in y is of order $2q + 1$, and

$$\frac{\partial^{2p}\phi}{\partial x^{2p}} \frac{\partial^{2q+1}\phi}{\partial y^{2q+1}} < 0.$$

The proposition states that only the kernel to satisfy this with $p = 1$ and $q = 0$ is the Gaussian. A special case of the function f used to prove the proposition is used here to show that it cannot be generalized.

If h is a Gaussian kernel, f is as in (5), and β and y are such that $\beta\ddot{h}(0) + \ddot{h}(y) = 0$ then ϕ has an extremum of order $p > 1$ at $x = 0$.

$$\beta = \frac{\ddot{h}(y)}{\ddot{h}(0)} = (y^2 - 1)\, e^{-y^2/2}.$$

Then, at $x = 0$,

$$\frac{\partial\phi}{\partial x} = \frac{\partial^2\phi}{\partial x^2} = \frac{\partial^3\phi}{\partial x^3} = 0$$

and

$$\frac{\partial^4\phi}{\partial x^4} = y^7(y^2 - 3)\, h(y).$$

Thus, whenever $y^2 \neq 3$, this defines an extremum of order 2 for ϕ. If we substitute $t = (1/2y^2)$, and write the diffusion equation for ϕ at $x = 0$, we get

$$\frac{\partial\phi}{\partial t} = \frac{\partial^2\phi}{\partial x^2} = 0$$

and

$$\frac{\partial^2\phi}{\partial t^2} = \frac{\partial}{\partial t}\left[\frac{\partial^2\phi}{\partial x^2}\right] = \frac{\partial^2}{\partial x^2}\left[\frac{\partial\phi}{\partial t}\right] = \frac{\partial^4\phi}{\partial x^4}.$$

At $x = 0$, ϕ has a first order extremum in t, and therefore in y, of the same nature as its second order extremum in x. This is incompatible with the generalized monotonicity condition.

2) Relation with the Diffusion Equation: Our sufficiency proof is based on the fact that $\phi(x, \sqrt{1/2t})$ is the solution of the diffusion equation in (x, t) with initial condition $f(x)$ and a unit diffusion coefficient. We saw that ϕ then satisfies a much stronger *monotonicity condition* than the one stated in the proposition.

If we consider the temperature distribution in an infinite rod with a heat diffusion coefficient $k(x)$ that is not constant, then the temperature profile of the rod at time t is a solution of

$$k\frac{\partial^2\phi}{\partial x^2} + \frac{\partial k}{\partial x}\frac{\partial\phi}{\partial x} = \frac{\partial\phi}{\partial t}$$

As a consequence,

$$\frac{\partial\phi}{\partial x} = 0 \quad \text{and} \quad \frac{\partial^2\phi}{\partial x^2} \neq 0$$

implies

$$\frac{\partial^2\phi}{\partial x^2}\frac{\partial\phi}{\partial t} > 0.$$

This is exactly equivalent to our *monotonicity condition*, and no stronger statement can be made. However, in the general case of a variable diffusion coefficient, the solution of the equation is not the convolution of the initial condition with a kernel.

3) The Boundary Conditions: When dealing with lowpass filters of variable bandwidths, one would like to think that the original signal is recovered when the bandwidth tends to infinity, and that the average of the signal is obtained in the zero-bandwidth case. As is shown in the following paragraphs, the first of these limit properties, when $y \to +\infty$, is satisfied by all signals without any further restriction, but the desired behavior when $y \to 0$ requires the additional assumption that the signal does have a constant average.

The Infinite-Bandwidth Case: The kernels defined by conditions 1)–4) of the proposition behave like acceptable low-pass filters in the infinite bandwidth case, because properties 1) and 2) ensure that

$$\lim_{y \to +\infty} \phi(\cdot, y) = f$$

and therefore $\phi(x, \infty)$ to be $f(x)$ itself. This can be verified by considering the Fourier transforms.

As we mentioned earlier, for all $y > 0$, the Fourier transform of $g(x, y) = yh(xy)$ is $H(\omega/y)$, where H is the Fourier transform of h. For all ω, $H(\omega/y) \to H(0)$ when $y \to +\infty$. If $h \geq 0$ and verifies 2) h is a probability density function, and it is known that pointwise convergence in the Fourier transform space implies weak convergence in the original space. Therefore, in that case, $g(x, y)$ converges weakly to a Dirac impulse at x. The restriction that $h \geq 0$ is lifted by decomposing h in the form: $h = h^+ - h^-$, where h^+ and h^- are both positive, continuous, and rapidly decreasing.

The Zero-Bandwidth Case: We want the smoothed signal to tend to a constant when $y \to 0$, and this imposes a further restriction on the type of signals we are considering, namely that f has a constant average $\bar{f}$. For all $x \in R$,

$$\lim_{T \to +\infty} \frac{1}{2T} \int_{-T}^{+T} f(x - a)\, da = \lim_{T \to +\infty} f * \frac{1}{2T} 1_{[-T,T]}(x) = \bar{f} \in R. \tag{7}$$

This is a first-order stationarity condition. If even a linear drift is added to a signal satisfying (7), then an average $\bar{f}(x)$ will exist for all $x \in R$, but it will not be constant. If such a drift is present in an observed signal, it must be subtracted from it before scale-space filtering can be used meaningfully on the residual.

A general kernel h is continuous and rapidly decreasing, and can therefore be uniformly approximated by step functions. If it also meets the symmetry condition 2) then it is clear that it can be developed in the form

$$h = \sum_{i=1}^{+\infty} \alpha_i \frac{1}{2t_i} 1_{[-t_i, t_i]}$$

where the series converges uniformly. If h also meets the normalization condition 3), then we must have

$$\sum_{i=1}^{+\infty} \alpha_i = 1.$$

Therefore

$$\lim_{T \to +\infty} f * g(x, 1/T) = \sum_{i=1}^{+\infty} \alpha_i \lim_{T \to +\infty} f * \frac{1}{2Tt_i} 1_{[-Tt_i, Tt_i]}(x) = \bar{f}$$

The change of variable $y = 1/T$ then yields

$$\lim_{y \to 0} \phi(x, y) = \bar{f}$$

for all signals f satisfying (7) and all kernels meeting conditions 1)–4). Therefore, provided $\bar{f} \neq 0$, the condition expressed in (7) is sufficient to ensure that the zero-crossing contours of ϕ never reach the line $y = 0$.

F. Existence of the Contours and Singularities

The implicit function theorem ensures that the equation $\phi(x, y) = 0$ uniquely defines an infinitely differentiable function $\alpha : V \to R$ in a neighborhood V of x_0 such that $\phi(x, \alpha(x)) = 0$ if and only if there exists y_0 such that

$$\phi(x_0, y_0) = 0$$

and

$$\left.\frac{\partial \phi}{\partial y}\right|_{(x_0, y_0)} \neq 0.$$

By symmetry in x and y, this implies that a unique zero-crossing contour goes through a point (x, y) such that $\phi(x, y) = 0$ if and only if grad $\phi(x, y) \neq 0$.

We know from the proposition that the Gaussian is the only kernel to ensure that $\phi_x = 0$ and $\phi_{xx} \neq 0$ implies $\phi_y \phi_{xx} < 0$. Therefore, with the Gaussian kernel, there is a unique zero-crossing contour through every (x, y) such that $\phi(x, y) = 0$ and $\phi_x(x, y) \neq 0$ or $\phi_{xx}(x, y) \neq 0$.

Points (x, y) at which $\phi(x, y) = \phi_x(x, y) = \phi_y(x, y) = \phi_{xx}(x, y) = 0$ are singular, in the sense that they can belong to either no zero-crossing contour or to more than one. They can be classified as follows:

a) $\phi(x, y) = 0$ and the first nonzero partial derivative of ϕ in x at (x, y) is of even order. The arguments of Section II-E-1 show that (x, y) is an extremum of ϕ both in x and y, and that the *monotonicity condition* is violated at (x, y). If it is also a minimum of ϕ in (x, y), then (x, y) is an isolated zero, although we have discovered no such cases so far.

b) $\phi(x, y) = 0$ and the first nonzero partial derivative of ϕ in x at (x, y) is of odd order. (x, y) is not an extremum, and may belong to several zero-crossing contours of ϕ. Such cases are easily constructed (see Fig. 2) but are seldom observed in irregular numerical data.

c) $\phi(x, y) = 0$ and all derivatives of ϕ in x are zero at (x, y). The following argument shows that this happens only in in the degenerate case $f = 0$.

For all $n \in N$, differentiating the Gaussian kernel gives

$$\frac{\partial^n \phi}{\partial x^n} = A_n y^n \int_{-\infty}^{+\infty} f\left(x - \sqrt{2}\, \frac{u}{y}\right) H_n(u)\, e^{-u^2} du = 0$$

where A_n is a constant and H_n is the Hermite polynomical of degree n. The above equation means that $f(x - \sqrt{2}u/y)$, as a function of u, is orthogonal to all Hermite polynomials. This can only be true if $f = 0$.

However, in numerical computation, f and ϕ are known only in the form of tables of values with a finite precision.

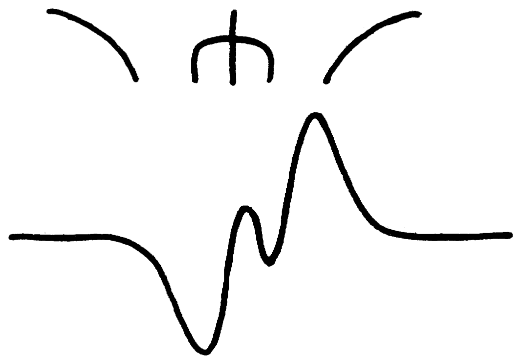


Fig. 2. The function $f(x) = -xe^{x^2} + xe^{x^2 4}$ illustrates a singularity at which zero-crossing contours in the scale-space image intersect. Adapted from an example by D. Vanderschel (personal communciation).

An extremum is identified by comparison of a table entry with the prior and next entries, not by calculation of the first nonzero derivative. An extremum that is smooth enough generates three equal entries in a table of values, and is indistinguishable from local constancy. This causes singular points of type c) to be observed in practice for nonzero f.

One remarkable feature of the zero-crossing contours obtained by Gaussian scale-space filtering is that singularities are due not to a lack of regularity, but to an excess of smoothness in the signal.

III. Application to Scale-Space Filtering

A. Contouring Zero Crossings

We now want to apply the proposition to the problem of tracking significant "events" in the xy-half-plane. By proving the above proposition, we have established that the Gaussian kernel is the only one to meet the *monotonicity condition* on first-order extrema. We will now explore the consequences of this on the shape of the zero-crossing contours of a transformed signal.

The events we want to track are defined by $L[\phi(x, y)] = 0$, where L is a linear differential operator, such as the Laplacian. By the properties of convolutions,

$$L[\phi(x, y)] = L[f(x)] * g((x, y) = f(x) * L[g(x, y)].$$

In other words, we can first apply the operator to the signal and then smooth the result, or apply the operator to the smoothing kernel and then convolve the result with the signal. The theory of distributions ensures that this is always meaningful if the signals are tempered distributions and the kernels are rapidly decreasing functions. If we think of $L[f(x)]$ as being the signal, then the events that we want to track are merely the zeros of ϕ itself. Thus, with no loss in generality we can restrict our attention to the behavior of the zero-crossing contours defined by

$$\phi(x, y) = 0.$$

B. The Problem of Closing Contours

Let us assume that we are doing Gaussian scale-space filtering and that there are no singular points in the xy-half-plane.

Coarse-to-fine tracking of zero-crossing events is only meaningful if the contours originate in the signal f. A closed zero-crossing contour in the xy-half-plane cannot conveniently be related to any feature of f, and therefore fails to provide useful signal description information.

A limited development of ϕ in the neighborhood of a y-extremum (x_m, y_m) of a zero-crossing contour yields

$$\phi(x, y) = \phi(x_m, y_m)\,(y - y_m) + \tfrac{1}{2}\phi_{xx}(x_m, y_m)\,(x - x_m)^2 + 0(y - y_m, x - x_m).$$

At (x_m, y_m), the tangent to the contour is the line $y = y_m$, and therefore $\phi_x(x_m, y_m) = 0$. $\phi_{xx}(x_m, y_m) \neq 0$, because (x_m, y_m) would otherwise be a singular point. The proposition then implies that $\phi_y(x_m, y_m)\ \phi_{xx}(x_m, y_m) < 0$. The cross term in ϕ_{xy} is dominated by the linear term in ϕ_y and is neglected.

It follows that the contour $\phi = 0$ is locally parabolic, and concave up (i.e., toward increasing bandwidth) because $\phi_y \phi_{xx} < 0$. If a contour were to close, then it would contain at least one y extremum in the vicinity of which the contour would be concave down. The proposition therefore guarantees that there will be no closed zero-crossing contours in the absence of singular points.

The *monotonicity condition* of the proposition is actually stronger than necessary to prevent closed contours, since it also forbids all y-extrema in the neighborhood of which a contour is concave down. A contour could reverse its path so as to generate such an extremum without necessarily being closed.

IV. Observations

The Gaussian function plays a fundamental role in probability theory, communication theory, and in the physics of diffusion processes. In probability theory, it is connected to additive random processes through the Central Limit Theorem. Among probability density functions having a given variance, it is well known to be the one with maximum entropy. As a smoothing kernel, it has the plesant property that it has the same form as its Fourier transform. As the Green's function for the diffusion equation, it provides "monotonic" smoothing of the initial function. Thus, in retrospect, it is not surprising that it should be the unique smoothing kernel with that property.

In the more specialized area of waveform analysis, we showed that the Gaussian is unique in guaranteeing that extrema may be added but not deleted in moving from coarse to fine scale. The partitioning of the signal into intervals may therefore be described by a coarse-to-fine tree. This hierarchic structure has proved useful in providing simplified "sketches" of signals [11], in histogram analysis [2], and in curve description [1].

A topic of interest is the extension of this useful tree structure to the description of images. There is ample precedent for the use of Gaussians and their derivatives in edge detection: Marr [5] argued that the smoothing/resolution tradeoff of a Gaussian made it particularly impor-

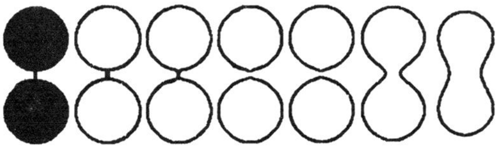


Fig. 3. The splitting and merging of zero-crossing contours over scale is illustrated by the "dumbell" shape at left. Black and white represent 1 and −1, respectively. The sequence of contour maps show zeros in the convolution of the shape with a Gaussian. The standard deviation of the Gaussian doubles with each map, from left to right. At both the finest and coarsest scales, a single contour divides the plane into two regions. On an intermediate range the contour splits into two, dividing the plane into three regions.

tant for early vision processing, while Rodieck and Stone [6] and Wilson and Bergen [9] have found evidence that this kind of detector is used by the eye.

Recently, [12] extended the present result to two dimensions, showing that zero-crossing contours (and other level curves) in Gaussian-convolved functions cannot vanish with decreasing scale. (Yuille & Poggio [12] addressed the related question of reconstructing a signal from its zero crossing over scale and Hummel *et al.* [3] addressed the more general problem of deblurring Gaussian blur.) In one dimension, the monotonicity condition guarantees that we can build an interval tree. Unfortunately, in two dimensions the situation is more complicated. Zero-crossing contours, while they may not *vanish*, are free to split and merge at increasingly fine scale. The regions bounded by zero-crossing contours therefore split and merge as well and their behavior over scale may not be described by a simple tree. The resulting difficulties are illustrated in Fig. 3, which shows a simple "dumbell" shape and the zeros of its Gaussian convolution at several scales. At both fine and coarse scales, a single contour divides the plane into two regions. On an intermediate range the contour splits into two, dividing the plane into three regions.

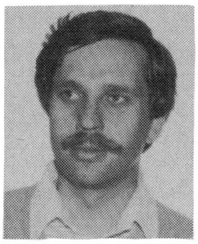
References

[1] H. Asada and M. Brady, "The curvature primal sketch," Massachusetts Inst. Technol., Cambridge, AI Memo 758, Feb. 1984.
[2] M. J. Carlotto, "Histogram analysis using a scale-space approach," in *Proc. CVPR*, 1985, pp. 334–340.
[3] R. A. Hummel, B. Kimia, and S. W. Zucker, "Deblurring Gaussian blur," McGill Univ., Montreal, P. Q., Canada, Tech. Rep. 83-13R, 1984.
[4] D. Marr and E. C. Hildredth, "Theory of edge detection," *Proc. Roy. Soc. London.*, vol. B 207, pp. 187-217, 1980.
[5] D. Marr, *Vision*. San Francisco, CA: W. H. Freeman, 1982.
[6] R. W. Rodieck and J. Stone, "Analysis of receptive fields of cat retinal ganglion cells," *J. Neurophysiol*, vol. 28, pp. 833–849, 1965.
[7] A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. Comput.*, vol. C-20, pp. 562–569, May 1971.
[8] W. Rudin, *Functional Analysis*, New York: McGraw-Hill, 1974, pp. 135–226.
[9] H. R. Wilson and J. R. Bergen, "A four mechanism model for spatial vision," *Vision Res.*, vol. 19, pp. 19–32, 1979.
[10] A. P. Witkin, "Scale space filtering," in *Proc. Int. Joint Conf. Artificial Intell.*, Karlsruhe, 1983.
[11] "Scale space filering: A new approach to multi-scale description," in *Image Understanding 1984*, S. Ullman and W. Richards, Eds. Norwood, NJ: Ablex, 1984.
[12] A. L. Yuille and T. Poggio, "Fingerprints theorems," *Proc. AAAI*, pp. 362–365, 1984.

Jean Babaud is a graduate of the Ecole Polytechnique, Paris, France, and also received the B.S. from the University of Toulouse, Toulouse, France.

He was a Meteorological Engineer from 1945 to 1947, when he joined the Charles Baudouin Corporation as a Development Engineer. He became the Technical Director of ACB Corp. in 1950 and its President in 1962. He later joined Schlumberger Instrumentation, where he became General Manager in 1965 and President in 1967. He was Director of Programs and Products for Schlumberger companies involved in electronics and instrumentation in Europe from 1969 to 1971, and then became Vice-President of Schlumberger Ltd. in charge of electronics and instrumentation programs. He has published papers in the field of measurement and holds several patents.

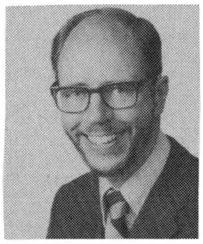
Andrew P. Witkin, photograph and biography not available at the time of publication.



Michel Baudin (M'82) received the B. S. degree in mathematics from the University of Paris-South, Paris, France, in 1975, and the M.S. degree in engineering from the Ecole des Mines de Paris in 1977.

He was engaged in research on the theory of multidimensional random point processes and its applications between 1977 and 1980, at the Earthquake Research Institute of the University of Tokyo, the Ecole des Mines de Paris, and the Institute of Statistical Mathematics of Tokyo. He joined Schlumberger-Fairchild in 1981 and took a leading role in the development of the computer integrated manufacturing system, INCYTE, now in use in most Fairchild factories. He is co-inventor of INCYTE and has published eight papers on subjects ranging from the theory of random point processes to the economics of computer integrated manufacturing.

Mr. Baudin is a member of the American Society for Quality Control and the Statistical Society of Japan.



Richard O. Duda (S'57–M'58–SM'65–F'80) was born in Evanston, IL, on April 27, 1936. He recieved the B.S. and M.S. degrees in engineering from the University of California, Los Angeles, in 1968 and 1969, respectively, and the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1962.

He joined the staff of Stanford Research Institute, Menlo Park, CA, in that year, and participated in research on pattern recognition, machine, vision, and expert system. At SRI he coauthored the book *Pattern Classification and Scene Analysis* with Dr. P. E. Hart, and was one of the developers of the PROSPECTOR system for mineral exploration. In 1980 he joined the newly established Fairchild Laboratory for Artificial Intelligence Research, where his research included work on knowledge-based signal interpretation. In 1983 he became one of the founders of Syntelligence, Sunnyvale, CA, where he is currently a Senior Scientist.

Dr. Duda is a member of Tau Beta Pi, Sigma Xi, Phi Beta Kappa, the Association for Computing Machinery, and the American Association for the Advancement of Science. He is a past Associate Editor of the IEEE TRANSACTIONS ON COMPUTERS.