Scale-Space Properties of Quadratic Feature Detectors

Paul Kube and Pietro Perona

Abstract—Feature detectors using a quadratic nonlinearity in the filtering stage are known to have some advantages over linear detectors; here, we consider their scale-space properties. In particular, we investigate whether, like linear detectors, quadratic feature detectors permit a scale selection scheme with the "causality property," which guarantees that features are never created as scale is coarsened. We concentrate on the design most common in practice, i.e., one dimensional detectors with two constituent filters, with scale selection implemented as convolution with a scaling function. We consider two special cases of interest: constituent filter pairs related by the Hilbert transform, and by the first spatial derivative. We show that, under reasonable assumptions, Hilbert-pair quadratic detectors cannot have the causality property. In the case of derivative-pair detectors, we describe a family of scaling functions related to fractional derivatives of the Gaussian that are necessary and sufficient for causality. In addition, we report experiments that show the effects of these properties in practice. Thus we show that at least one class of quadratic feature detectors has the same desirable scaling property as the more familiar detectors based on linear filtering.

Index Terms—Feature detection, edge detection, scale space, nonlinear filtering, energy filters, quadratic filters, causality.

1 Introduction

Images contain information at multiple scales of resolution, and detecting image features across a range of scales is an important step in many visual tasks. Given an image, one may obtain coarse scale versions of it by "blurring" or "lowpass-filtering" it in order to perform this analysis. This idea, often called "multiscale image analysis" or "scale-space analysis," dates at least from Rosenfeld and Thurston [1] and Marr [2]. Whenever the same feature is present at multiple scales, it is convenient to *detect* it at coarse scale and *localize* it by propagating the results to fine scales [3]. For this reason it is thought to be important that features detected at a given resolution were not created gratuitously at that scale by the blurring process, but rather are "grounded" in image detail at a finer resolution. A multiscale feature detection method that never introduces features as the scale is coarsened has the desirable property of *causality* (the term in this context is due to Koenderink [4]; the property has also been called "monotonicity" [5], "well-behavedness" [6], "nice scaling behavior" [7], and "the evolution property" [8]).

It is known that edge detectors which operate by marking edges at zeros, level-crossings, or extrema, in the output of a linear filter acting on the image have the causality property if scale is selected by convolution of the image with a Gaussian of appropriate variance [5], [7], [9], [10].

- *P. Kube is with the Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA 92093. E-mail: kube@cs.ucsd.edu.*
- *P. Perona is with the California Institute of Technology, MS 136-97, Pasadena, CA 91125, and with the Dipartmento di Elettronica ed Informatica, Università di Padova, Italy. E-mail: perona@caltech.edu.*

Manuscript received Dec. 15, 1994. Recommended for acceptance by J. Daugman. An earlier version of this work appeared in the Proceedings of the Third European Conference on Computer Vision, May 1994.
For information on obtaining reprints of this article, please send e-mail to: transpami@computer.org, and reference IEEECS Log Number P96073.

Scale selection by means of nonlinear diffusion [11] has also been studied (see survey chapters on linear and nonlinear diffusions in [12]). The idea of analyzing images by nonlinear combination of the outputs of quadrature pairs of linear filters dates from the late 1970s and early 1980s [13], [14], [15]. More recently, it has been observed that for feature detection quadratic filters have advantages over linear filters, particularly because they are able to detect and localize features with complex structure [16], [17], [18], [19], [20], [21], [22], [23]. However, the question whether these quadratic or "energy" detectors permit a causal scale selection technique has remained open.

We address this question restricting ourselves to one-dimensional quadratic feature detectors, with scale selected by convolution of the image with a "scaling" function. We concentrate on the most common design, detectors with two constituent filters, and we consider two cases of interest: constituent filters related by the Hilbert transform, and constituent filters related by the first spatial derivative. We show that, in the case of Hilbert-pair filters, there exists no scaling function giving the causality property. In the case of derivative-pair filters, we describe a family of scaling functions related to fractional derivatives of the Gaussian that are necessary and sufficient for causality. Thus we show that at least one class of quadratic feature detectors has the same desirable scaling property as the more familiar detectors based on linear differential filtering.

Definitions, notation and assumptions are given in Section 2. Section 3 states the theorems about the scale-space properties of Hilbert-pair and derivative-pair quadratic feature detectors; the proofs of the theorems are in the Appendix. These results are supplemented with experimental observations reported in Section 4. Section 5 discusses the results and directions for future work.

2 Scale, Causality, and Feature Detection

Let $h(x)$ be a real-valued signal. In this paper we assume $x \in \mathbb{R}$, giving direct application to time signals, images with unidirectionally oriented edges, and images with features that are one-dimensional on a scale comparable to the detector filter size. Let $\sigma > 0$ be a real valued scale parameter, with increasing σ corresponding to coarser scales. We are interested in finding the location of features in the signal h at a scale σ; or, more to the point, in how this feature detection process behaves as σ is varied.

A linear feature detector marks a feature in $h(x)$ at position $x = x_0$ and scale $\sigma = \sigma_0$ if x_0 is a local extremum in x of the linearly filtered signal

$$L(x,\sigma) = (f * g^{\sigma} * h)(x) \tag{1}$$

with $\sigma = \sigma_0$. Here $f(x)$ is the impulse response of a linear shift invariant filter, $g^{\sigma}(x)$ is a scaling function of the form $g(x/\sigma)/\sigma$, and "$*$" denotes convolution. Scale selection is a matter of filtering the signal h by convolution with the scaling function for some value of σ; since convolution is associative, it is equivalent to view this as using a detector filter with impulse response $f * g^{\sigma}$, which depends on the scale parameter.

A quadratic (or "energy") feature detector marks a feature in the signal h at position $x = x_0$ and scale $\sigma = \sigma_0$ if x_0 is a local maximum in x of the nonlinearly filtered signal

$$E(x,\sigma) = \sum_{f \in \mathcal{F}} \left[\left(f * g^{\sigma} * h \right)(x) \right]^2 \tag{2}$$

with $\sigma = \sigma_0$. Here, $\mathcal{F}$ is the set of impulse responses of the *constituent filters* of the detector. The process of scale selection is analogous to the case of the linear detector (1), but the quadratic nonlinearity gives a detector with interestingly different properties.

It is known [5], [7], that linear detectors have the causality property if (and only if) the scaling function is the Gaussian $g^{\sigma}(x) \propto \exp(-x^2/2\sigma^2)/\sigma$. Our question is whether, like linear detectors, there are quadratic feature detectors that have the causality property; that is, whether there exist choices of constituent filters $\mathcal{F}$ and scaling function g^{σ} which guarantee that, for every signal h, new maxima are never introduced in E as σ increases.

We make these assumptions:

1) The constituent filter impulse responses $f \in \mathcal{F}$, the parameterized scaling function g^{σ}, and the signal h are functions $\mathbb{R} \mapsto \mathbb{R}$ which are sufficiently well behaved that E as defined in (2) is smooth and bounded.
2) The scaling function $g^{\sigma}(x)$ is of the form $g(x/\sigma)/\sigma$, for some even function $g(\cdot)$: The scaling function has no preferred direction or scale.
3) $g(\cdot)$ is *not* such that $g(x/\sigma)/\sigma \propto g(x)$ for all σ: scaling must do more than multiply by a scalar.
4) $g(x)$ has a differentiable Fourier transform $G(u)$ (and so $g(x/\sigma)/\sigma$ has Fourier transform $G(\sigma u)$).
5) Each $f \in \mathcal{F}$ has a Fourier transform $F(u)$ nonzero for u in some open set of $\mathbb{R}$.
6) The signal h has a Fourier transform $H(u)$.

These assumptions are in the spirit of those found in the study of scaling functions for linear differential feature detectors, e.g., [5] and [7]; differences are minor and technical.

We will concentrate on the case of quadratic feature detectors with a set of two constituent filters $\mathcal{F} = \{f^a, f^b\}$, such that f^b is the Hilbert transform of f^a multiplied by a nonzero constant, or such that f^b is the first derivative of f^a multiplied by a nonzero constant. These "Hilbert-pair" and "derivative-pair" quadratic detector designs comprise all existing practice of which we are aware.

2.1 Conditions for Causality Failure

If a quadratic feature detector with constitutent filters $\mathcal{F}$ and scaling function g^{σ} fails to have the causality property, then there is some signal h, location x_0, and scale σ_0 such that a local maximum of $E(x, \sigma)$ is created with increasing σ at $x = x_0$, $\sigma = \sigma_0$. Since E is smooth, (x_0, σ_0) is a degenerate critical point of E with respect to x, and generically the set of critical points $E_x(x, \sigma) = 0$ in a neighborhood of (x_0, σ_0) form an upward opening parabola (a "noncausal" fold catastrophe) as shown in Fig. 1 [6], [24], [25]. $E(x, \sigma)$ qualitatively changes shape as a function of x in the neighborhood of the generic noncausal degenerate critical point, as shown in Fig. 2. For $\sigma < \sigma_0$, there is no maximum of E in a neighborhood of x_0; at $\sigma = \sigma_0$, E is locally cubic in the neighborhood of x_0; with $\sigma > \sigma_0$, there is a local maximum of E near x_0.

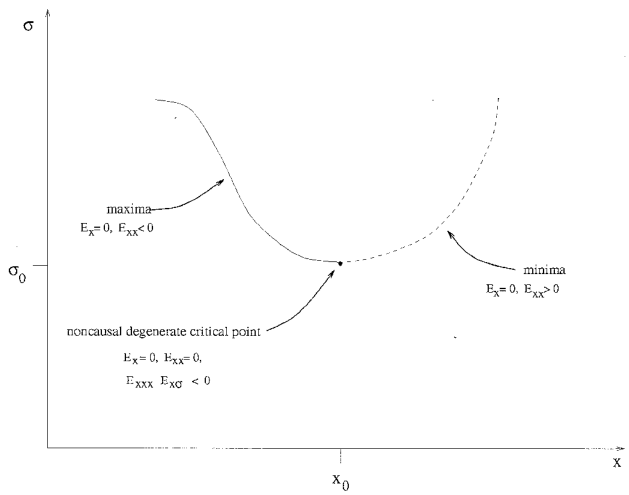


Fig. 1. The generic scale-space signature of a causality failure, i.e., a creation of a new maximum, in $E(x, \sigma)$. See text for discussion.

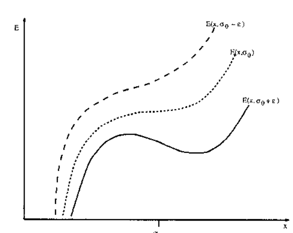


Fig. 2. Typical behavior of $E(x,\sigma)$ as a function of x near a noncausal degenerate critical point (x_0,σ_0). ϵ is small and positive.

It is straightforward to express the conditions of such a causality failure in terms of partial derivatives of E at the degenerate critical point (cf. [7]): A noncausal fold catastrophe occurs in $E(x, \sigma)$ at a point $x = x_0$, $\sigma = \sigma_0$ if and only if

$$\begin{aligned} E_x(x_0, \sigma_0) &= 0, \\ E_{xx}(x_0, \sigma_0) &= 0, \\ E_{x\sigma}(x_0, \sigma_0)\, E_{xxx}(x_0, \sigma_0) &< 0. \end{aligned} \tag{3}$$

Thus, showing the existence of a signal h that satisfies the conditions (3) for E as defined in (2) is sufficient to show that the correpsonding multiscale quadratic feature detector does not have the causality property. If any signal that produces a causality failure can be transformed into one that produces a generic causality failure, the converse is also true, and the nonexistence of h to satisfy (3) implies that the detector has the causality property. Alternatively, following Hummel [8], one could try to prove causality by showing that for any h, $E_x(x, \sigma)$ is the solution to a partial differential equation that satisfies a minimum and a maximum principle. (This minimum-principle approach is, of course, not promising for proving that a detector does *not* have the causality property.) Here, we will take the first approach, and investigate whether or not there exists a signal h to satisfy the conditions (3) for particular quadratic detector designs, i.e., for particular choices of constituent filters $\mathcal{F}$ and scaling function g^{σ} satisfying Assumptions 1 through 6.

2.2 Toward Investigating the Causality Property

We will consider whether generic causality failures can occur in $E(x, \sigma)$ at $x = 0$; since E is shift invariant with respect to the signal h, this is without loss of generality. Here we develop a useful notation for the problem. Essentially, we will seek an expression for the conditions (3) which reduces the question of the causality property to the question of solving a constrained system of linear equations.

Define

$$\Delta^{f}_{n,m} \equiv \frac{\partial^{n+m}}{\partial x^{n} \partial \sigma^{m}} \left(f * g^{\sigma} * h \right)(0).$$

Partial derivatives of E in σ and x involved in the conditions (3) for $x_0 = 0$ consist of terms of this form, viz.:

$$\begin{aligned} E_x(0,\sigma) &= 2\sum_{f\in\mathcal{F}} \Delta^{f}_{0,0}\Delta^{f}_{1,0} \\ E_{xx}(0,\sigma) &= 2\sum_{f\in\mathcal{F}} \left[\left(\Delta^{f}_{1,0}\right)^2 + \Delta^{f}_{0,0}\Delta^{f}_{2,0} \right] \\ E_{xxx}(0,\sigma) &= 2\sum_{f\in\mathcal{F}} \left[3\Delta^{f}_{1,0}\Delta^{f}_{2,0} + \Delta^{f}_{0,0}\Delta^{f}_{3,0} \right] \\ E_{x\sigma}(0,\sigma) &= 2\sum_{f\in\mathcal{F}} \left[\Delta^{f}_{0,1}\Delta^{f}_{1,0} + \Delta^{f}_{0,0}\Delta^{f}_{1,1} \right] \end{aligned}$$

Recalling from Assumption 2 that $g^{\sigma}(x)$ has the form $g(x/\sigma)/\sigma$, and writing F, G, and H for the Fourier transforms of f, g, and h, respectively, we have from the convolution and scaling theorems [26]

$$\left(f * g^{\sigma} * h \right)(x) = \int_{-\infty}^{\infty} e^{2i\pi ux} F(u) G(\sigma u) H(u)\, du \tag{4}$$

and therefore

$$\Delta^{f}_{n,m} = \int_{-\infty}^{\infty} (2i\pi)^{n} u^{n+m} F(u) G^{(m)}(\sigma u) H(u)\, du \tag{5}$$

where $G^{(m)}$ is the mth derivative of G. Note that $G^{(m)}(u)$ is real, and even or odd in u depending on whether m is even or odd.

We represent the signal $h(x)$ as the sum of K sinusoidal components; since $h(x)$ is real, we can then write, for some complex coefficients $h_1, \ldots, h_K$

$$H(u) = \sum_{k=1}^{K} \left[\delta(u - u_k) h_k + \delta(u + u_k) h_k^{*} \right]$$

where all the frequencies u_k are positive (the "DC term" contributes nothing of interest), and "*" denotes complex conjugation. Thus

$$\begin{aligned} \Delta^{f}_{n,m} = \sum_{k=1}^{K} (2i\pi)^{n} \Big[& u_k^{n+m} F(u_k) G^{(m)}(\sigma u_k) h_k \\ & + (-1)^{n} u_k^{n+m} F^{*}(u_k) G^{(m)}(\sigma u_k) h_k^{*} \Big]. \end{aligned}$$

We note that this discretization of the spectrum of h is without loss of generality for our purposes, since K can be arbitrarily large, and a generic causality failure is robust to sufficiently small perturbations of the integrals (5). Our approach here can be compared with the "spatial" discretization of the signal assumed in [7]; we remark that in the context of digital signal processing applications, discretization in both domains is routine.

For further simplification, we can restrict our attention to the case of constituent filters whose impulse response functions are even- or odd-symmetric in x. (That this restriction to even-odd detectors is without loss of generality follows from Lemma 1 in the Appendix.) Letting the constituent filters $\mathcal{F} = \{f^{e}, f^{o}\}$, with $f^{e}(x)$ even symmetric and $f^{o}(x)$ odd, the Fourier transforms $F^{e}(u)$ and $F^{o}(u)$ of $f^{e}(x)$ and $f^{o}(x)$ are then even real and odd imaginary, respectively. Writing h_k^{e} for the even (real) part and h_k^{o} be the imaginary (odd) part of each signal coefficient h_k, so that $h_k = h_k^{e} + h_k^{o}$, $h_k^{*} = h_k^{e} - h_k^{o}$, we then have, for n even,

$$\begin{aligned} \Delta^{f^e}_{n,m} &= \sum_{k} (2i\pi)^{n} \left[2u_k^{n+m} F^{e}(u_k) G^{(m)}(\sigma u_k) h_k^{e} \right] \\ \Delta^{f^o}_{n,m} &= \sum_{k} -(2i\pi)^{n} \left[2u_k^{n+m} \frac{1}{i} F^{o}(u_k) G^{(m)}(\sigma u_k) \frac{1}{i} h_k^{o} \right] \end{aligned}$$

and, for n odd,

$$\begin{aligned} \Delta^{f^e}_{n,m} &= \sum_{k} (2i\pi)^{n} i \left[2u_k^{n+m} F^{e}(u_k) G^{(m)}(\sigma u_k) \frac{1}{i} h_k^{o} \right] \\ \Delta^{f^o}_{n,m} &= \sum_{k} -(2i\pi)^{n} i \left[2u_k^{n+m} \frac{1}{i} F^{o}(u_k) G^{(m)}(\sigma u_k) h_k^{e} \right] \end{aligned}$$

Now, we introduce the real column K-vectors $\mathbf{f}^{o,m}$, $\mathbf{f}^{e,m}$, the systems $\mathbf{h}^{e}$, $\mathbf{h}^{o}$, with the kth component of each specified as

$$\mathbf{f}_k^{e,m} = F^e(u_k)G^{(m)}(\sigma u_k)$$
$$\mathbf{f}_k^{o,m} = \frac{1}{i}F^o(u_k)G^{(m)}(\sigma u_k)$$
$$\mathbf{h}_k^e = 2h_k^e$$
$$\mathbf{h}_k^o = \frac{2}{i}h_k^o$$

Further, let $\mathbf{U}$ be the $K \times K$ diagonal matrix of positive frequencies $diag(u_1, \ldots, u_K)$. Then we have

$$\Delta_{n,m}^{f^e} = (2i\pi)^n\left[\mathbf{h}^{eT}\mathbf{U}^{n+m}\mathbf{f}^{e,m}\right]$$
$$\Delta_{n,m}^{f^o} = -(2i\pi)^n\left[\mathbf{h}^{oT}\mathbf{U}^{n+m}\mathbf{f}^{o,m}\right]$$
$$\Delta_{n,m}^{f^e} = (2i\pi)^n i\left[\mathbf{h}^{oT}\mathbf{U}^{n+m}\mathbf{f}^{e,m}\right]$$
$$\Delta_{n,m}^{f^o} = -(2i\pi)^n i\left[\mathbf{h}^{eT}\mathbf{U}^{n+m}\mathbf{f}^{o,m}\right]$$

where "^{T}" denotes matrix transpose. In fact, to study the conditions for generic causality failure (3) we will need only to consider partial derivatives with respect to σ of order 0 and 1. So, to simplify the notation further, define

$$\mathbf{f}^e \equiv \mathbf{f}^{e,0}$$
$$\mathbf{f}^o \equiv \mathbf{f}^{o,0}$$
$$\mathbf{f}^{e\prime} \equiv \mathbf{f}^{e,1}$$
$$\mathbf{f}^{o\prime} \equiv \mathbf{f}^{o,1}$$

and, finally, introduce the symbols

$$v_1 \equiv \mathbf{h}^{eT}\mathbf{f}^e = \Delta_{0,0}^{f^e}$$
$$v_2 \equiv \mathbf{h}^{eT}\mathbf{U}\mathbf{f}^o = \Delta_{1,0}^{f^o} / 2\pi$$
$$v_3 \equiv \mathbf{h}^{eT}\mathbf{U}^2\mathbf{f}^e = -\Delta_{2,0}^{f^e} / 4\pi^2$$
$$v_4 \equiv \mathbf{h}^{eT}\mathbf{U}^3\mathbf{f}^o = -\Delta_{3,0}^{f^o} / 8\pi^3$$
$$v_1' \equiv \mathbf{h}^{eT}\mathbf{U}\mathbf{f}^{e\prime} = \Delta_{0,1}^{f^e}$$
$$v_2' \equiv \mathbf{h}^{eT}\mathbf{U}^2\mathbf{f}^{o\prime} = \Delta_{1,1}^{f^o} / 2\pi$$
$$w_1 \equiv \mathbf{h}^{oT}\mathbf{f}^o = -\Delta_{0,0}^{f^o}$$
$$w_2 \equiv \mathbf{h}^{oT}\mathbf{U}\mathbf{f}^e = -\Delta_{1,0}^{f^e} / 2\pi$$
$$w_3 \equiv \mathbf{h}^{oT}\mathbf{U}^2\mathbf{f}^o = \Delta_{2,0}^{f^o} / 4\pi^2$$
$$w_4 \equiv \mathbf{h}^{oT}\mathbf{U}^3\mathbf{f}^e = \Delta_{3,0}^{f^e} / 8\pi^3$$
$$w_1' \equiv \mathbf{h}^{oT}\mathbf{U}\mathbf{f}^{o\prime} = -\Delta_{0,1}^{f^o}$$
$$w_2' \equiv \mathbf{h}^{oT}\mathbf{U}^2\mathbf{f}^{e\prime} = -\Delta_{1,1}^{f^e} / 2\pi$$

With this notation, we can state the issue as follows: An even-odd quadratic feature detector with constituent filters $\mathcal{F} = \{f^e, f^o\}$ and scaling function g^σ has the (generic) causality property if and only if for no diagonal positive frequency matrix $\mathbf{U}$ do there exist vectors $\mathbf{h}^e$, $\mathbf{h}^o$ to solve the systems

$$\begin{pmatrix} \mathbf{f}^{eT} \\ (\mathbf{U}\mathbf{f}^o)^T \\ (\mathbf{U}^2\mathbf{f}^e)^T \\ (\mathbf{U}^3\mathbf{f}^o)^T \\ (\mathbf{U}\mathbf{f}^{e\prime})^T \\ (\mathbf{U}^2\mathbf{f}^{o\prime})^T \end{pmatrix} \mathbf{h}^e = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_1' \\ v_2' \end{pmatrix} \tag{6}$$

$$\begin{pmatrix} \mathbf{f}^{oT} \\ (\mathbf{U}\mathbf{f}^e)^T \\ (\mathbf{U}^2\mathbf{f}^o)^T \\ (\mathbf{U}^3\mathbf{f}^e)^T \\ (\mathbf{U}\mathbf{f}^{o\prime})^T \\ (\mathbf{U}^2\mathbf{f}^{e\prime})^T \end{pmatrix} \mathbf{h}^o = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_1' \\ w_2' \end{pmatrix} \tag{7}$$

subject to the constraints (3), namely

$$E_x \propto v_2\, w_1 - v_1\, w_2 = 0 \tag{8}$$

$$E_{xx} \propto v_2^2 - v_1 v_3 + w_2^2 - w_1 w_3 = 0 \tag{9}$$

$$E_{xxx}\, E_{x\sigma} \propto (3\, v_3\, w_2 + v_1\, w_4 - 3\, v_2\, w_3 - v_4\, w_1) \times (-\, v_1'\, w_2 - v_1\, w_2' + v_2\, w_1' + v_2'\, w_1) < 0 \tag{10}$$

Therefore, the issue is whether these constrained systems have solutions. We discuss answers to this question in the following sections.

3 Scale-Space Theorems

We are considering quadratic feature detectors with two constituent filters. If one of the constituent filters is the Hilbert transform of the other (perhaps multiplied by a nonzero constant), we call it a Hilbert-pair detector. This is the most common type of quadratic detector in practice; motivation for this design has come from work in psychophysical modeling [27] as well as consideration of its computational properties [16], [28]. Because of the quadrature phase relationship between the constituent filters, these are sometimes called "energy" feature detectors.

Suppose a one-dimensional function $f(x)$ has Fourier transform $F(u)$; then the Hilbert transform of f has Fourier transform $i\ \mathrm{sgn}(u)F(u)$. This gives a particularly simple form for the systems (6) and (7), since when the constituent filters f^e and f^o are Hilbert transforms of each other we have (up to sign) $\mathbf{f}^e = \mathbf{f}^o = \mathbf{f}$. We find that there always exist $\mathbf{h}^e$, $\mathbf{h}^o$ to solve these constrained systems, and so there exists a signal $h(x)$ that leads the detector to a causality failure:

Theorem 1. *No Hilbert-pair quadratic feature detector has the causality property.*

The proof, in the Appendix, is constructive, in that for appropriate σ_0, it gives linear systems which can be solved to compute a signal $h(x)$ which produces a causality failure at scale $\sigma = \sigma_0$ and $x = 0$. In the case of Gaussian scaling and differential constituent filters, this can be done at *any* scale σ. An example of constructing a causality failure in such a quadratic detector is shown in Fig. 3; compare with Fig. 2.

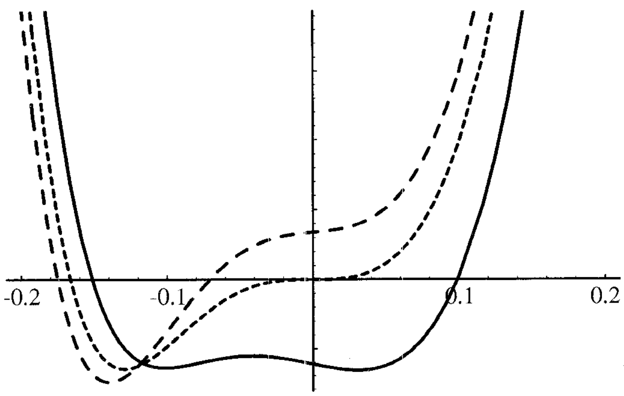


Fig. 3. An example of a causality failure in Hilbert-pair quadratic detector. A new maximum appears with increasing σ in $E(x, \sigma)$ at $x = 0$, $\sigma = 0.1$. The constituent filters are a first-derivative operator and its Hilbert transform; scaling is by convolution with a Gaussian. The signal $h(x)$ that produced this causality failure was constructed by solving the constrained system described in Section 2, with frequencies $u_1 = 1$, $u_2 = 2$, $u_3 = 3$, $u_4 = 4$. The dashed, dotted, and solid curves graph $E(x, \sigma)$ as a function of x for $\sigma = 0.04$, 0.1, 0.16 respectively. The new maximum can be observed in the $E(x, 0.16)$ curve to the left of $x = 0$.

If one of the constituent filters in a quadratic detector is the first derivative of the other (perhaps multiplied by a nonzero constant), we call it a derivative-pair detector. Some properties of derivative-pair detectors have been been discussed by Kube [21]. As with Hilbert-pair detectors, we obtain simplifications in the form of the systems (6), (7), since if, for example, $f^o(x) = df^e(x)/dx$, then $\mathbf{f}^o = 2\pi \mathbf{U}\mathbf{f}^e$.

Here however the result is more favorable, and we find that there exists a family of scaling functions that give derivative-pair detectors the causality property. This family includes the familiar Gaussian and its even derivatives, but is somewhat more general. We define it as follows:

If a function $g(x)$ has a Fourier transform $G(u)$, then the tth derivative of g (with respect to x) has Fourier transform $(2\pi i u)^t G(u)$. If t is not an integer, then this is the Fourier transform of a fractional derivative of g. In general, a fractional derivative of an even, real function will not be real or even. For example, if g^σ is a Gaussian

$$g(x) = \frac{r}{s}\exp\left(\frac{-x^2}{s^2}\right)$$

for some nonzero r and positive s, then the tth derivative of g^σ has Fourier transform

$$(2\pi i u)^t \sqrt{2\pi} r \, \exp\left(-s^2\pi^2u^2\right)$$

which is not real and even, and so cannot be the transform of a scaling function, unless t is an even integer. The related function

$$r|u|^t \sqrt{2\pi} \, \exp\left(-s^2\pi^2u^2\right)$$

is, however, real and even for all real t. Expressed as a function of σu,

$$G(\sigma u) = |\sigma u|^t \sqrt{2\pi} r \, \exp\left(-s^2\pi^2\sigma^2u^2\right), \tag{11}$$

this is the Fourier transform of an admissible scaling function. Lacking a better name, we will call such a function a *modified fractional derivative of the Gaussian*. This defines a family of scaling functions; it includes, for example, the normal density function (when $r = 1$, $s = \sqrt{2}$, $t = 0$), and the second derivative of the normal density function (when $r = -4\pi/\sigma^2$, $s = \sqrt{2}$, $t = 2$), etc. We will find that scaling functions from this family are involved in the only known examples of quadratic feature detectors with the causality property:

THEOREM 2. *A derivative-pair quadratic feature detector has the causality property if and only if its scaling function is an modified fractional derivative of the Gaussian.*

The proof is in the Appendix. We note that there exist function pairs that are related both by the Hilbert transform *and* the first derivative, e.g., $\mathcal{F} = \{\cos(x), -\sin(x)\}$. However, these functions are pure sinusoids, and the apparent contradiction between Theorems 1 and 2 is ruled out by the assumptions given in Section 2, in particular Assumption 5.

Experiments showing the implications of these theorems in feature detection on real images are discussed in the next section.

4 Results of Experiments

Section 3 stated theorems to the effect that causality failures can occur in quadratic feature detector schemes which use Hilbert-pair filters and Gaussian scaling, but cannot generically occur if Gaussian scaling is used with derivative-pair filters. However, these results leave open the questions whether causality failures occur often in practice using Hilbert-pair detectors on real images, and whether nongeneric causality failures occur in practice with derivative-pair detectors. We have attempted to address these questions with experiments we report here.

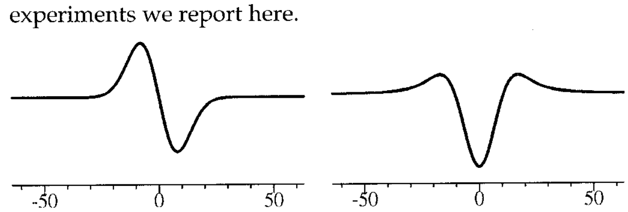


Fig. 4. Typical scaled constituent filter impulse responses $f^o * g^s$, $f^e * g^s$ for the Hilbert-pair quadratic feature detector discussed in the text. Left, the first derivative of a Gaussian with standard deviation eight pixels; right, its Hilbert transform.

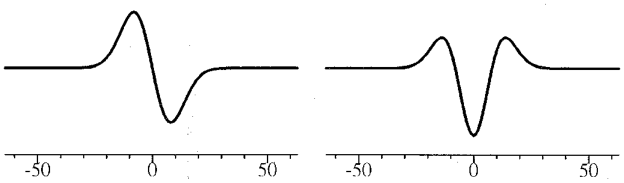


Fig. 5. Typical scaled constituent filter impulse responses $f^o * g^s$, $f^e * g^s$ for the derivative-pair quadratic feature detector discussed in the text. Left, the first derivative of a Gaussian with standard deviation eight pixels; right, its first derivative.

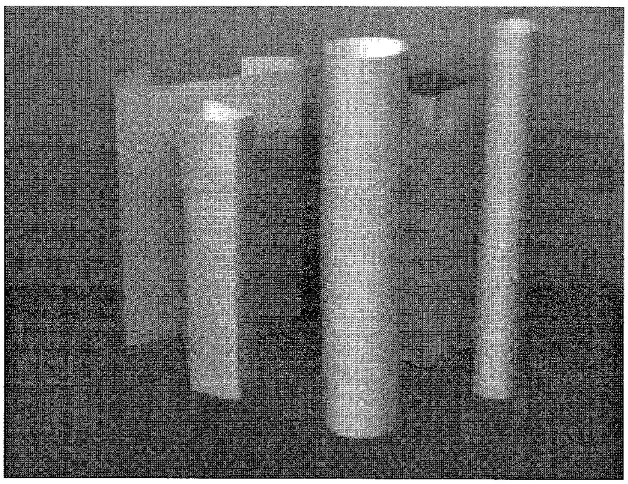


Fig. 6. Typical arrangement of objects used in creating the images used in experiments. A row of pixels from an image of such objects was used as the one-dimensional signal $h(x)$ for feature detection.

Six scenes were created by placing matte-surfaced right rectangular prisms and cylinders randomly on a table; Fig. 6 shows a typical arrangement of objects. Each scene was imaged with camera geometry and illumination such that edges in the images were predominantly vertical. A typical row of 512 pixels was taken from each such image; this one-dimensional signal was subjected to quadratic feature detection as defined in Section 2, using Gaussian scaling with σ in the range 0.5 to 64 pixels. A region of each image from which the row of pixels was taken and the graph of the pixel intensities in that row are shown in Figs. 7-12.

In one set of experiments, the quadratic detector had as constituent filters the first-derivative operator and its Hilbert transform; in the other set of experiments, the quadratic detector had as constituent filters the first-derivative operator and its first derivative (i.e., the second derivative operator). Impulse responses of the constituent filters at a scale of eight pixels are shown in Figs. 4 and 5, respectively. In each experiment, local maxima were detected with no thresholding at scales in the range 0.5 to 64 pixels, in 0.5 pixel steps. The resulting scale space representations of the multiscale features found are also shown in Figs. 7 through 11. Apparent causality failures are indicated with arrows.

It is interesting that while the shapes of the constituent filter impulse responses are quite similar, the performance of these two types of quadratic feature detector are qualitatively different. For each image, the Hilbert-pair quadratic detector exhibited causality failure. Some features which were introduced with increasing scale were ephemeral, but others were quite robust and persisted over a range of scales. The derivative-pair quadratic detector generated no observed noncausal features. These experimental results are, of course, consistent with the theoretical results stated in Section 3.

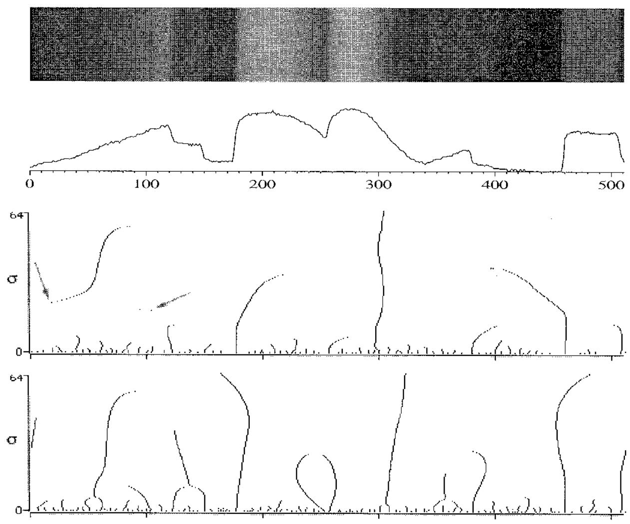


Fig. 7. From top to bottom: An image with one-dimensional edges; the graph of image intensity for the top row of pixels in the image; feature scale-space generated from that 1D signal with a Hilbert-pair quadratic feature detector; feature scale-space generated from that 1D signal with a derivative-pair quadratic feature detector. Gaussian scaling and periodic convolution are used in each case. The Hilbert-pair detector exhibits causality failures (indicated by arrows); the derivative-pair filter does not. See Figs. 8, 9, 10, and 11 for other examples. See text for discussion.

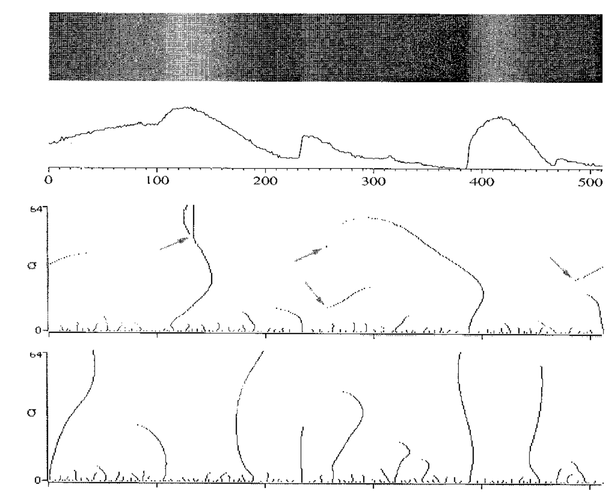


Fig. 8.

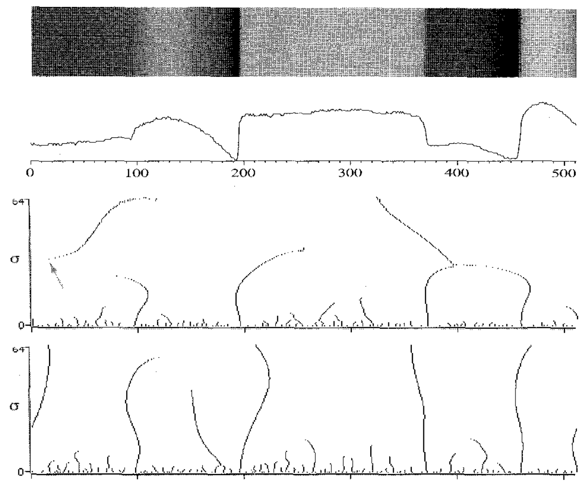


Fig. 9.

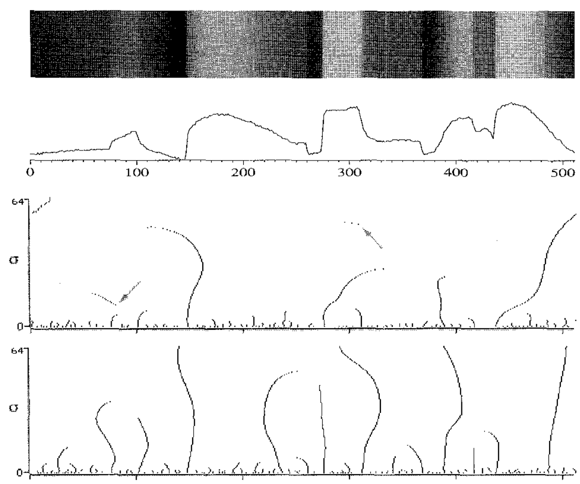


Fig. 10.

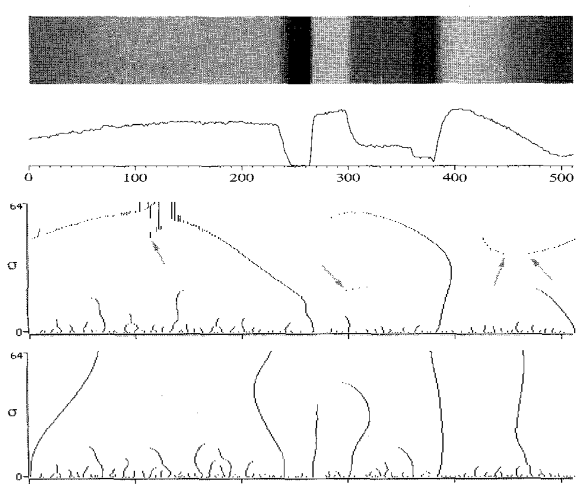


Fig. 11.

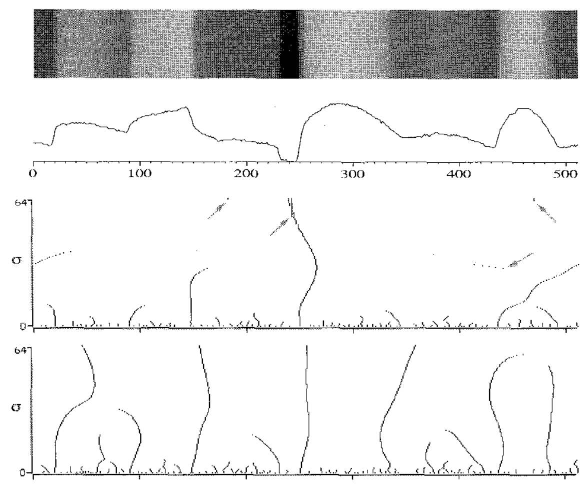


Fig. 12.

5 Discussion

The scale-space properties of quadratic feature detectors being of potential interest, we have investigated whether any quadratic detectors have the causality property. We considered one-dimensional derivative-pair and Hilbert-pair detectors, and found that the former, but not the latter, can have the property. Further, we showed that there is a family of scaling functions, related to fractional derivatives of the Gaussian, which are necessary and sufficient for causality with derivative-pair detectors. We proved theorems to this effect, and showed results on real images that show that the theoretical results hold in practice.

Ronse [28] has studied Hilbert-pair detectors and concluded, on the basis of the relation between maxima in E and phases of the Fourier components of the signal h, that feature points are stable under convolution of the image with a zero-phase scaling function such as the Gaussian. We emphasize that this is not true for all features in every signal h; as we have shown, Hilbert-pair detectors are incapable of having the causality property, for any scaling function. However, it is true for feature locations relative to which the phases of all Fourier components of h are identical (this includes features at $x = 0$ for some $h(x)$ which are even or odd symmetric in x). Though it does not address the performance of Hilbert-pair detectors on more general kinds of features, this is a remarkable property, and one that is not shared by linear or derivative-pair quadratic detectors.

We have restricted our scope in various ways, and corresponding generalizations of our results are possible. For example, we have concentrated on one-dimensional signals. A treatment of higher dimensions could be pursued with the same kind of techniques; if with increasing spatial dimension there remains only one isotropic scale parameter σ, generic nondegenerate critical points will always have essentially a one-dimensional structure [24], [25]. (The Hilbert transform and the first derivative do not extend uniquely to higher dimensions, so a family of filters parameterized by rotation would have to be considered.)

In addition, we have only considered the properties of quadratic detectors with two constituent filters, and only special cases of these. The cases treated encompass all examples of quadratic detectors in the literature, but others may be useful to investigate. Some extensions of the results here are immediate; for example, any two-filter design which leads to independent rows in the systems (6) and (7) will permit a proof of noncausality along the lines given for the Hilbert-pair case. Other generalizations may be difficult; the question of the causality property for quadratic detectors is equivalent to deciding whether a certain system of quadratic inequalities has a feasible solution, and in its general form this appears to be a hard problem [29].

One can also question how important the causality property really is in feature detector design. It may not be essential in practice to have all features at every scale continuously traceable to features at scale zero. In our experiments, some of the features introduced with increasing scale in Hilbert-pair detectors were ephemeral, persisting only over a narrow range of scales. A multiscale image processing system which computes image representations at only discrete scales may not observe such short-lived events at all. These ephemeral features often have low contrast, and are spatially unstable; postprocessing may be able to eliminate many of them. Further, some persistent noncausal features may have physical significance, even though they do not appear at some range of finer scales. In general, scale-space properties may have to be balanced against other properties of a feature detector, such as reliable detection of desired features, number of local maxima in the filter response associated to each feature, and so on.

Nevertheless, we have shown that the type of quadratic detector that has heretofore received most of the attention in the literature, the Hilbert-pair detector, does not have the causality property; and these results should direct increased interest toward the derivative-pair type, which does.

Appendix
Proofs of Theorems

Here we prove Theorems 1 and 2 stated in Section 3, using notation developed in Section 2, and relying on Assumptions 1 through 6 stated there.

We will find the following lemmas useful:

LEMMA 1. *Let g^σ be any scaling function, and let θ be any Fourier-transformable odd-symmetric linear shift-invariant operator (e.g., the Hilbert transform or first derivative operator). Then*

1) *If for every even function f^e, the even-odd detector with constituent filters $\{f^e, \theta f^e\}$ and scaling function g^σ has the causality property, then for every function f, the detector with constituent filters $\{f, \theta f\}$ and scaling function g^σ has the causality property.*
2) *If for every even function f^e, the even-odd detector with constituent filters $\{f^e, \theta f^e\}$ and scaling function g^σ fails to have the causality property, then for every function f, the detector with constituent filters $\{f, \theta f\}$ and scaling function g^σ fails to have the causality property.*

PROOF. Let $\Theta(u)$ be the Fourier representation of the operator θ. Then, following (4), we can write (2) as

$$E(x,\sigma) = \left[\left(f * g^\sigma * h\right)(x)\right]^2 + \left[\left(\theta f * g^\sigma * h\right)(x)\right]^2$$
$$= \left[\int_{-\infty}^{\infty} e^{2i\pi ux} F(u)H(u)G(\sigma u)\, du\right]^2 + \left[\int_{-\infty}^{\infty} e^{2i\pi ux}\Theta(u)F(u)H(u)G(\sigma u)\, du\right]^2$$

and clearly $E(x, \sigma)$ remains unchanged, and so exhibits the same scale-space properties, if $F(u)H(u)$ is replaced by an identical (generalized) function of u. Taking the two parts of the lemma in turn:

1) We show the contrapositive. With scaling function g^σ, suppose some filter pair $\{f, \theta f\}$ exhibits a failure of the causality property on a signal h. Let $G(\sigma u)$, $F(u)$, $\Theta(u)F(u)$, $H(u)$ be the corresponding Fourier transforms. Select an even f^e such that its Fourier transform $F^e(u)$ is nowhere zero, and consider the signal h' with Fourier transform
$$H'(u) = \frac{F(u)H(u)}{F^e(u)}.$$
Substituting, we see that with the same scaling function, the even-odd pair $\{f^e, \theta f^e\}$ must exhibit a causality failure on h'.
2) For any f, we show there exists a signal h that will exhibit a failure of the causality property with filters $\{f, \theta f\}$. Let $F(u)$ be the Fourier transform of f. Select even f^e such that its Fourier transform $F^e(u)$ is zero exactly where $F(u)$ is zero. Now, since by hypothesis $\{f^e, \theta f^e\}$ with scaling function g^σ fails to have the causality property, there exists a signal h' on which this even-odd pair exhibits the failure. Let $H'(u)$ be the Fourier transform of such a signal, and consider the signal h with Fourier transform
$$H(u) = \begin{cases} \dfrac{F^e(u)H'(u)}{F(u)} & \text{for } u \text{ such that } F(u) \neq 0 \\ 0 & \text{elsewhere.} \end{cases}$$
Again, substituting, we see that with the same scaling function, $\{f, \theta f\}$ must exhibit a causality failure on h. □

LEMMA 2. *Let $F(u)$ and $G(u)$ be the Fourier transforms of some constituent filter transfer function and scaling function, respectively. Then for any integers m, n, k such that $n \le m$, $k > m - n$, there exists a matrix of positive frequencies $\mathbf{U} = diag(u_1, \ldots, u_k)$ such that the matrix*

$$\begin{pmatrix} \left(\mathbf{U}^n \mathbf{f}\right)^T \\ \left(\mathbf{U}^{n+1}\mathbf{f}\right)^T \\ \cdots \\ \left(\mathbf{U}^{m-1}\mathbf{f}\right)^T \\ \left(\mathbf{U}^m \mathbf{f}\right)^T \end{pmatrix} \quad (12)$$

has full rank, where $\mathbf{f}_i = F(u_i)G(\sigma u_i)$ *for any* σ *in some open set of* $\mathbb{R}$.

PROOF. The matrix (12) can be factored

$$\begin{pmatrix} (\mathbf{U}^n\mathbf{f})^T \\ (\mathbf{U}^{n+1}\mathbf{f})^T \\ \cdots \\ (\mathbf{U}^{m-1}\mathbf{f})^T \\ (\mathbf{U}^m\mathbf{f})^T \end{pmatrix} = \begin{pmatrix} u_1^n & \cdots & u_k^n \\ u_1^{n-1} & \cdots & u_k^{n-1} \\ \vdots & \ddots & \vdots \\ u_1^{m-1} & \cdots & u_k^{m-1} \\ u_1^m & \cdots & u_k^m \end{pmatrix} diag(\mathbf{f}_i, \ldots, \mathbf{f}_k)$$

$$= \mathbf{M}\ diag(\mathbf{f}_i, \ldots, \mathbf{f}_k)$$

Since the matrix $\mathbf{M}$ is Vandermonde, $\mathbf{M}$ will be full rank if the frequencies u_i are distinct. The matrix $diag(\mathbf{f}_1, \ldots \mathbf{f}_k)$ will be nonsingular just in case no element of $\mathbf{f}$ is 0; but there always exist k distinct frequencies to make this true for σ in some open set, since by Assumptions 4 and 5 $G(u)$ is differentiable and not strictly zero, and $F(u)$ is nonzero in some open set. Thus there exist k distinct frequencies to make the product of these matrices, viz. (12) full rank. □

THEOREM 1. *No Hilbert-pair quadratic feature detector has the causality property.*

PROOF. By Lemma 1, it suffices to show that, under the assumptions stated in Section 2, there exist a diagonal matrix of positive frequencies $\mathbf{U}$ and vectors $\mathbf{h}^e$, $\mathbf{h}^o$ to solve the systems (6) and (7) subject to the constraints (8) through (10) for some σ. If f^o is the Hilbert transform of f^e, then, in this notation, $\mathbf{f}^e = \mathbf{f}^o = \mathbf{f}$, and the systems (6) and (7) become

$$\begin{pmatrix} \mathbf{f}^T \\ (\mathbf{Uf})^T \\ (\mathbf{U}^2\mathbf{f})^T \\ (\mathbf{U}^3\mathbf{f})^T \\ (\mathbf{Uf}')^T \\ (\mathbf{U}^2\mathbf{f}')^T \end{pmatrix} \mathbf{h}^e = \Psi\mathbf{h}^e = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_1' \\ v_2' \end{pmatrix} \tag{13}$$

$$\begin{pmatrix} \mathbf{f}^T \\ (\mathbf{Uf})^T \\ (\mathbf{U}^2\mathbf{f})^T \\ (\mathbf{U}^3\mathbf{f})^T \\ (\mathbf{Uf}')^T \\ (\mathbf{U}^2\mathbf{f}')^T \end{pmatrix} \mathbf{h}^o = \Psi\mathbf{h}^o = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_1' \\ w_2' \end{pmatrix} \tag{14}$$

We will prove the theorem for this case; it should be clear that multiplying the filters by nonzero scalars only trivially complicates the proof.

The proof proceeds by cases to consider possible dependency relationships among the rows of the matrix Ψ. For each case, we show that there exist $\mathbf{h}^e$, $\mathbf{h}^o$ to solve the systems (13) and (14) subject to the constraints (8) through (10).

Begin by selecting σ and $\mathbf{U} = diag(u_1, \ldots, u_K)$, $K > 5$, such that the frequencies u_k are distinct, and such that $\mathbf{f}_k = G(u_k\sigma)F(u_k) \neq 0$ for $k = 1, \ldots, K$. (Since $F(u)$ is nonzero for u in some open subset, and $G(u)$ is differentiable and not strictly zero, this is possible for any K.) Now, the first four rows of the matrix Ψ are independent, from Lemma 2. The cases we consider exhaust the possible dependence relationships between the vectors $\mathbf{Uf}'$, $\mathbf{U}^2\mathbf{f}'$ and the first four rows of Ψ.

CASE 1. $\mathbf{Uf}'$, $\mathbf{U}^2\mathbf{f}'$ are independent of $\{\mathbf{f}, \mathbf{Uf}, \mathbf{U}^1\mathbf{f}, \mathbf{U}^3\mathbf{f}\}$. Then the matrix Ψ is nonsingular and there exist $\mathbf{h}^e$, $\mathbf{h}^o$ to solve the systems for any arbitrary values of the constants v_i, w_i. We can select, for example,

$$v_1 = w_1 = v_2 = w_2 = v_3 = w_3 = 1 \tag{15}$$

$$v_4 = 1 \qquad w_4 = 1 \tag{16}$$

$$v_1' = 0 \qquad v_2' = 0 \tag{17}$$

$$w_1' = -1 \qquad w_2' = -1 \tag{18}$$

which satisfy the constraints (8) through (10).

CASE 2. $\mathbf{Uf}'$ is independent of $\{\mathbf{f}, \mathbf{Uf}, \mathbf{U}^1\mathbf{f}, \mathbf{U}^3\mathbf{f}\}$, but $\mathbf{U}^2\mathbf{f}'$ is dependent, say

$$(\mathbf{U}^2\mathbf{f}')^T = (\alpha\ \beta\ \gamma\ \delta) \begin{pmatrix} \mathbf{f}^T \\ (\mathbf{Uf})^T \\ (\mathbf{U}^2\mathbf{f})^T \\ (\mathbf{U}^3\mathbf{f})^T \end{pmatrix} \tag{19}$$

for some real numbers α, β, γ, δ. Then there exist $\mathbf{h}^e$, $\mathbf{h}^o$ to solve the systems for arbitrary values of all the constants v_i, w_i except v_2', w_2', for which we must have

$$v_2' = \alpha v_1 + \beta v_2 + \gamma v_3 + \delta v_4$$

$$w_2' = \alpha w_1 + \beta w_2 + \gamma w_3 + \delta w_4.$$

We can select, for example,

$$v_1 = w_1 = v_2 = w_2 = v_3 = w_3 = 1$$

$$v_4 = 0 \qquad w_4 = 1.$$

This satisfies (8) and (9) and makes $E_{xxx} > 0$. Then we select w'_1, v'_1 such that

$$w'_1 - v'_1 - \delta < 0$$

this makes $E_{x\sigma} < 0$, and so satisfies (10).

CASE 3. $\mathbf{Uf}'$ is dependent on $\{\mathbf{f}, \mathbf{Uf}, \mathbf{U}^1\mathbf{f}, \mathbf{U}^3\mathbf{f}\}$, say

$$(\mathbf{Uf}')^T = (\alpha\ \beta\ \gamma\ \delta) \begin{pmatrix} \mathbf{f}^T \\ (\mathbf{Uf})^T \\ (\mathbf{U}^2\mathbf{f})^T \\ (\mathbf{U}^3\mathbf{f})^T \end{pmatrix}$$

But then

$$\left(\mathbf{U}^2\mathbf{f}'\right)^T = (\alpha\ \beta\ \gamma\ \delta)\begin{pmatrix}(\mathbf{Uf})^T\\ \left(\mathbf{U}^2\mathbf{f}\right)^T\\ \left(\mathbf{U}^3\mathbf{f}\right)^T\\ \left(\mathbf{U}^4\mathbf{f}\right)^T\end{pmatrix}$$

and so

$$v'_1 = \alpha v_1 + \beta v_2 + \gamma v_3 + \delta v_4$$
$$w'_1 = \alpha w_1 + \beta w_2 + \gamma w_3 + \delta w_4$$
$$v'_2 = \alpha v_2 + \beta v_3 + \gamma v_4 + \delta v_5$$
$$w'_2 = \alpha w_2 + \beta w_3 + \gamma w_4 + \delta w_5$$

for some real numbers α, β, γ, δ, where v_5, w_5 are such that

$$\begin{pmatrix}\mathbf{f}^T\\ (\mathbf{Uf})^T\\ \left(\mathbf{U}^2\mathbf{f}\right)^T\\ \left(\mathbf{U}^3\mathbf{f}\right)^T\\ \left(\mathbf{U}^4\mathbf{f}\right)^T\end{pmatrix}\mathbf{h}^e = \begin{pmatrix}v_1\\ v_2\\ v_3\\ v_4\\ v_5\end{pmatrix} \qquad (20)$$

$$\begin{pmatrix}\mathbf{f}^T\\ (\mathbf{Uf})^T\\ \left(\mathbf{U}^2\mathbf{f}\right)^T\\ \left(\mathbf{U}^3\mathbf{f}\right)^T\\ \left(\mathbf{U}^4\mathbf{f}\right)^T\end{pmatrix}\mathbf{h}^o = \begin{pmatrix}w_1\\ w_2\\ w_3\\ w_4\\ w_5\end{pmatrix} \qquad (21)$$

Note that, from Lemma 2, systems (20), (21) are full rank, and so can be solved for any values of the v_i, w_i. We consider subcases.

CASE IIIA. β, γ, δ not all 0. We are free to select $v_1 = w_1 = 1$, $v_2 = w_2 = 2$, $v_3 = 8$, $w_3 = w_4 = w_5 = 0$; this satisfies constraints (8), (9). Then (10) is satisfied if

$$8(48 - v_4)\beta + (48 - v_4)(-16 + v_4)\gamma$$
$$- (48 - v_4)(2v_4 - v_5)\ \delta < 0$$

But this can be done by suitable choice of v_4, v_5. This can be seen as follows. Suppose $\delta \neq 0$. Then choose $v_4 = 16$, and note that $256\beta - 32\,(64 - v_5)\delta$ can be made negative by choice of v_5. On the other hand, suppose $\delta = 0$; then $8(48 - v_4)\ \beta + (48 - v_4)(-16 + v_4)\ \gamma$ can be made negative by choice of v_4, since the ratio of coefficients, viz., $v_4 - 16$, can be made to take any desired value by choice of v_4, and not both γ, β are 0.

CASE IIIB. $\beta = \gamma = \delta = 0$. This implies

$$\mathbf{Uf}' = \alpha\mathbf{f}$$

If this fails to hold for some choice of $\mathbf{U}$ and σ, then one of the other Cases 1, 2, 3a, or 3b holds, and the causality property fails to hold. But therefore, from the definition of $\mathbf{f}$ and $\mathbf{f}'$, to avoid failure of the causality property, G must be a solution to the differential equation

$$uF(u)\ G'(\sigma u) = \alpha F(u) G(\sigma u), \qquad (22)$$

wherever $F(u)\ G(\sigma u) \neq 0$. Now solutions to (22) have the form

$$G(\sigma u) = r\ (\sigma u)^s \qquad (23)$$

so $G(\sigma u) \propto G(u)$, and G is not the Fourier transform of an admissible scaling function, since it violates Assumption 3. Therefore, there is no scaling function that gives a Hilbert-pair detector with the causality property. □

THEOREM 2. *A derivative-pair quadratic feature detector has the causality property if and only if its scaling function is a modified fractional derivative of the Gaussian.*

PROOF. We will show that, under the assumptions stated in Section 2, there exist a diagonal matrix of positive frequencies $\mathbf{U}$ and vectors $\mathbf{h}^e$, $\mathbf{h}^o$ to solve the systems (6), (7) subject to the constraints (8) through (10) for some σ, unless the scaling function is a modified fractional derivative of the Gaussian (defined in Section 3); in which case there are no solutions. (Again, as for Theorem 1, the restriction to even-odd filter pairs suffices, by Lemma 1.) If $f^e(x) = \frac{\xi}{2\pi}\frac{d}{dx}f^o(x)$ for some nonzero scalar ξ, then, in this notation, we can put $\mathbf{f}^o = \mathbf{f}$ and $\mathbf{f}^e = \xi\mathbf{f}$, and the systems (6), (7) become

$$\begin{pmatrix}\xi(\mathbf{Uf})^T\\ (\mathbf{Uf})^T\\ \xi\left(\mathbf{U}^3\mathbf{f}\right)^T\\ \left(\mathbf{U}^3\mathbf{f}\right)^T\\ \xi\left(\mathbf{U}^2\mathbf{f}'\right)^T\\ \left(\mathbf{U}^2\mathbf{f}'\right)^T\end{pmatrix}\mathbf{h}^e = \begin{pmatrix}v_1\\ v_2\\ v_3\\ v_4\\ v'_1\\ v'_2\end{pmatrix} \qquad (24)$$

$$\begin{pmatrix}\mathbf{f}^T\\ \xi\left(\mathbf{U}^2\mathbf{f}\right)^T\\ \left(\mathbf{U}^2\mathbf{f}\right)^T\\ \xi\left(\mathbf{U}^4\mathbf{f}\right)^T\\ (\mathbf{Uf}')^T\\ \xi\left(\mathbf{U}^3\mathbf{f}'\right)^T\end{pmatrix}\mathbf{h}^o = \begin{pmatrix}w_1\\ w_2\\ w_3\\ w_4\\ w'_1\\ w'_2\end{pmatrix} \qquad (25)$$

(Here we do not suppress the multiplicative constant ξ, since its role is not as obvious as in the case of Theorem 1.) These are in turn equivalent to the systems

$$\begin{pmatrix}(\mathbf{Uf})^T\\ \left(\mathbf{U}^3\mathbf{f}\right)^T\\ \left(\mathbf{U}^2\mathbf{f}'\right)^T\end{pmatrix}\mathbf{h}^e = \begin{pmatrix}v_2\\ v_4\\ v'_2\end{pmatrix} \qquad (26)$$

$$\begin{pmatrix} \mathbf{f}^T \\ \left(\mathbf{U}^2\mathbf{f}\right)^T \\ \left(\mathbf{U}\mathbf{f}'\right)^T \end{pmatrix} \mathbf{h}^o = \begin{pmatrix} w_1 \\ w_3 \\ w_1' \end{pmatrix} \tag{27}$$

together with the identities

$$v_1 = \xi v_2,\ v_3 = \xi v_4,\ v_1' = \xi v_2',\ w_2 = \xi w_3.$$

We will find it useful also to introduce the symbol $w_5 = w_4/\xi$. Substituting these identities in (8) through (10), we find that a generic causality failure requires the existence of $\mathbf{h}^e$, $\mathbf{h}^o$ to solve the systems (26) and (27) subject to the constraints

$$E_x \propto v_2 w_1 - \xi^2 v_2 w_3 = 0 \tag{28}$$

$$E_{xx} \propto v_2^2 - \xi^2 v_2 v_4 + \xi^2 w_3^2 - w_1 w_3 = 0 \tag{29}$$

$$\begin{aligned} E_{xxx}E_{x\sigma} \propto &\left(3\xi^2 v_4 w_3 + \xi v_2 w_4 - 3 v_2 w_3 - v_4 w_1\right) \\ &\times\left(-\xi^2 v_2' w_3 - \xi v_2 w_2' + v_2 w_1' + v_2' w_1\right) < 0 \end{aligned} \tag{30}$$

We proceed by cases to show these conditions can be satisfied, if and only if the detector does not use modified fractional derivative Gaussian scaling.

Now, (28) is satisfied just in case $v_2 = 0$ or $w_1 = \xi^2 w_3$. If $v_2 = 0$, (29) is satisfied just in case $w_3 = 0$, or $w_1 = \xi^2 w_3$. If $w_1 = \xi^2 w_3$, (29) is satisfied just in case $v_2 = 0$ or $v_2 = \xi^2 v_4$. So to satisfy (28) and (29), at least one of the following must hold:

- $v_2 = 0$ and $w_3 = 0$
- $v_2 = 0$ and $w_1 = \xi^2 w_3$
- $w_1 = \xi^2 w_3$ and $v_2 = \xi^2 v_4$

We will consider each of these in the cases that follow.

CASE 1. $\mathbf{U}\mathbf{f}'$ and $\mathbf{U}^3\mathbf{f}'$ are both independent of $\{\mathbf{f}, \mathbf{U}^2\mathbf{f}, \mathbf{U}^4\mathbf{f}\}$ (and so $\mathbf{U}^2\mathbf{f}'$ is independent of $\{\mathbf{U}\mathbf{f}, \mathbf{U}^3\mathbf{f}\}$). A causality failure can be constructed, because the systems (26) and (27) can be made full rank by Lemma 2, and thus there exist $\mathbf{h}^e$, $\mathbf{h}^o$ to solve them for any choice of the v_i, w_i. Selecting, for example, $v_2 = w_3 = 0$, $w_1 = v_4 = v'_2 = 1$ satisfies the constraints $E_x = E_{xx} = 0$, $E_{xxx} < 0$, $E_{x\sigma} > 0$.

Therefore, to avoid noncausality, the systems (26) and (27) must be defective.

CASE 2. $\mathbf{U}\mathbf{f}'$ is dependent on $\{\mathbf{f}, \mathbf{U}^2\mathbf{f}, \mathbf{U}^4\mathbf{f}\}$, say

$$\left(\mathbf{U}\mathbf{f}'\right)^T = \left(\alpha\ \beta\ \gamma\right)\begin{pmatrix} \mathbf{f}^T \\ \left(\mathbf{U}^2\mathbf{f}\right)^T \\ \left(\mathbf{U}^4\mathbf{f}\right)^T \end{pmatrix}$$

Then

$$\left(\mathbf{U}^2\mathbf{f}'\right)^T = \left(\alpha\ \beta\ \gamma\right)\begin{pmatrix} \left(\mathbf{U}\mathbf{f}\right)^T \\ \left(\mathbf{U}^3\mathbf{f}\right)^T \\ \left(\mathbf{U}^5\mathbf{f}\right)^T \end{pmatrix}$$

$$\xi\left(\mathbf{U}^3\mathbf{f}'\right)^T = \xi\left(\alpha\ \beta\ \gamma\right)\begin{pmatrix} \left(\mathbf{U}^2\mathbf{f}\right)^T \\ \left(\mathbf{U}^4\mathbf{f}\right)^T \\ \left(\mathbf{U}^6\mathbf{f}\right)^T \end{pmatrix}$$

we consider the systems

$$\begin{pmatrix} \left(\mathbf{U}\mathbf{f}\right)^T \\ \left(\mathbf{U}^3\mathbf{f}\right)^T \\ \left(\mathbf{U}^5\mathbf{f}\right)^T \end{pmatrix} h^e = \begin{pmatrix} v_2 \\ v_4 \\ v_5 \end{pmatrix}$$

$$\begin{pmatrix} \mathbf{f}^T \\ \left(\mathbf{U}^2\mathbf{f}\right)^T \\ \left(\mathbf{U}^4\mathbf{f}\right)^T \\ \left(\mathbf{U}^6\mathbf{f}\right)^T \end{pmatrix} h^e = \begin{pmatrix} w_1 \\ w_3 \\ w_5 \\ w_6 \end{pmatrix}$$

where we have introduced the symbol w_6. By Lemma 2, these systems can be solved for arbitrary values of the v_i and w_i thereby determining (28) through (30). Treating subcases, corresponding to the possible ways of satisfying the constraints (28) and (29):

CASE 2A. $v_2 = 0$ and $w_3 = 0$. Substituting into (30), we get

$$E_{xxx}E_{x\sigma} \propto -\gamma v_4 v_5 w_1^2 - \beta v_4^2 w_1^2$$

which can be made negative by suitable choice of v_4, v_5, w_1, if $\gamma \neq 0$. If $\gamma = 0$, this can be made negative if and only if $\beta > 0$. Therefore, no noncausality can arise in this case, if and only if $\gamma = 0$, and $\beta \leq 0$.

CASE 2B. $v_2 = 0$ and $w_1 = \xi^2 w_3$. We have

$$E_{x\sigma} = 0$$

so no noncausality is possible.

CASE 2C. $w_1 = \xi^2 w_3$ and $v_2 = \xi^2 v_4$. We have

$$\begin{aligned} E_{xxx}E_{x\sigma} \propto &-\gamma\xi^2 v_4^2\left(w_3 - w_5\xi^2\right)\left(-w_5 + w_6\xi^2\right) \\ &-\beta\xi^4 v_4^2\left(-w_3 + w_5\xi^2\right)^2. \end{aligned}$$

This can be made negative by suitable choice of v_4, w_3, w_5, w_6, if $\gamma \neq 0$, because the ratio of the coefficients of β, γ is

$$\frac{w_5\xi^2 - w_3}{w_6\xi^2 - w_5}$$

which can be made to take any value. However, if $\gamma = 0$, as in Case 2a, since the coefficient of β cannot be made negative, no noncausality is possible as long as $\beta \leq 0$.

CASE 3. $\mathbf{U}^2\mathbf{f}'$ is dependent on $\{\mathbf{U}\mathbf{f}, \mathbf{U}^3\mathbf{f}\}$, say

$$\left(\mathbf{U}^2\mathbf{f}'\right)^T = \left(\alpha\ \beta\right)\begin{pmatrix} \left(\mathbf{U}\mathbf{f}\right)^T \\ \left(\mathbf{U}^3\mathbf{f}\right)^T \end{pmatrix}.$$

Then

$$(\mathbf{Uf'})^T = (\alpha\ \beta)\begin{pmatrix}\mathbf{f}^T \\ (\mathbf{U}^2\mathbf{f})^T\end{pmatrix}$$

and

$$(\mathbf{U}^3\mathbf{f'})^T = (\alpha\ \beta)\begin{pmatrix}(\mathbf{U}^2\mathbf{f})^T \\ (\mathbf{U}^4\mathbf{f})^T\end{pmatrix},$$

which is just a special case of Case 2.

CASE 4. $\mathbf{U}^3\mathbf{f'}$ is dependent on $\{\mathbf{f}, \mathbf{U}^2\mathbf{f}, \mathbf{U}^4\mathbf{f}\}$, say

$$(\mathbf{U}^3\mathbf{f'})^T = (\gamma\ \alpha\ \beta)\begin{pmatrix}\mathbf{f}^T \\ (\mathbf{U}^2\mathbf{f})^T \\ (\mathbf{U}^4\mathbf{f})^T\end{pmatrix}.$$

Then

$$(\mathbf{U}^2\mathbf{f'})^T = (\gamma\ \alpha\ \beta)\begin{pmatrix}(\mathbf{U}^{-1}\mathbf{f})^T \\ (\mathbf{Uf})^T \\ (\mathbf{U}^3\mathbf{f})^T\end{pmatrix}$$

and

$$(\mathbf{Uf'})^T = (\gamma\ \alpha\ \beta)\begin{pmatrix}(\mathbf{U}^{-2}\mathbf{f})^T \\ \mathbf{f}^T \\ (\mathbf{U}^2\mathbf{f})^T\end{pmatrix}.$$

We proceed by introducing the symbols v_0, w_0 and considering the systems

$$\begin{pmatrix}(\mathbf{U}^{-1}\mathbf{f})^T \\ (\mathbf{Uf})^T \\ (\mathbf{U}^3\mathbf{f})^T\end{pmatrix}\mathbf{h}^e = \begin{pmatrix}v_0 \\ v_2 \\ v_4\end{pmatrix}$$

$$\begin{pmatrix}(\mathbf{U}^{-2}\mathbf{f})^T \\ \mathbf{f}^T \\ (\mathbf{U}^2\mathbf{f})^T \\ (\mathbf{U}^4\mathbf{f})^T\end{pmatrix}\mathbf{h}^o = \begin{pmatrix}w_0 \\ w_1 \\ w_3 \\ w_5\end{pmatrix}$$

which, again by Lemma 2, can be solved for arbitrary choice of values for v_0, v_2, v_4, w_0, w_1, w_3, and w_5. And, again, treating subcases corresponding to the possible ways of satisfying the constraints (28) and (29):

CASE 4A. $v_2 = 0$ and $w_3 = 0$. Substituting, we find

$$E_{xxx}E_{x\sigma} \propto -\gamma v_0 v_4 w_1^2 - \beta v_4^2 w_1^2$$

and this can be made negative if and only if $\gamma \neq 0$ or $\beta > 0$; this is the same situation as Case 2a.

CASE 4B. $v_2 = 0$ and $w_1 = \xi^2 w_3$. Then $E_{x\sigma} = 0$ and so noncausality is impossible.

CASE 4C. $w_1 = \xi^2 w_3$ and $v_2 = \xi^2 v_4$. We get

$$E_{xxx}E_{x\sigma} \propto -\beta v_4^2 \xi^4 \left(-w_3 + w_5\xi^2\right)^2$$

which can be made negative if and only if $\beta > 0$; this is the same situation as Case 2c.

Thus from consideration of the possible cases, we find that the causality property must hold if and only if

$$\mathbf{Uf'} = \alpha\mathbf{f} + \beta\mathbf{U}^2\mathbf{f}$$

for some α, and $\beta \leq 0$. And so, from the definitions of $\mathbf{f}$ and $\mathbf{f'}$, for the detector to have the causality property it is necessary and sufficient that G be a solution to the differential equation

$$uF(u)G'(\sigma u) = \alpha F(u)G(\sigma u) + \beta u^2 F(u)G(\sigma u) \qquad (31)$$

for some α, and $\beta \leq 0$, whenever $F(u)G(\sigma u) \neq 0$. Recalling from Assumption 2 that we require G to be of the form $G(\sigma u)$ and even, (31) has such a solution just in case

$$G(\sigma u) = r\,|\,\sigma u\,|^t \exp(-su^2\sigma^2) \qquad (32)$$

for some r, s, t, $s > 0$. That is, the scaling function must be a modified fractional derivative of the Gaussian. □

ACKNOWLEDGMENTS

This research was supported in part by the U.S. National Science Foundation (RIA-IRI 9211651, NYI-IRI 9306155, Caltech ERC for Neuromorphic Systems Engineering), and by the U.S. Office of Naval Research (N00014-93-1-0990). The authors are grateful to John Canny, Bill Helton, Jean-Michel Morel, and Christian Ronse for helpful discussions and for reading drafts of the paper. Comments and suggestions by PAMI Associate Editor John Daugman were very useful and much appreciated.

REFERENCES

[1] A. Rosenfeld and M. Thurston, "Edge and Curve Detection for Visual Scene Analysis," *IEEE Trans. Computers*, vol. 20, pp. 562-569, May 1971.

[2] D. Marr and E. Hildreth, "Theory of Edge Detection," *Proc. Royal Soc. London*, vol. B207, pp. 187-217, 1980.

[3] A.P. Witkin, "Scale-Space Filtering," *Int'l Joint Conf. Artificial Intelligence*, 1983, pp. 1,019-1,021, Karlsruhe.

[4] J.J. Koenderink, "The Structure of Images," *Biol. Cybern.*, vol. 50, pp. 363-370, 1984.

[5] J. Babaud, A. Witkin, M. Baudin, and R. Duda, "Uniqueness of the Gaussian Kernel for Scale-Space Filtering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 26-33, Jan. 1986.

[6] J.J. Clark, "Singularity Theory and Phantom Edges in Scale Space," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, no. 5, pp. 720-727, 1988.

[7] A. Yuille and T. Poggio, "Scaling Theorems for Zero Crossings," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 15-26, Jan. 1986.

[8] R.A. Hummel and A.R. Moniot, "Reconstructions From Zero Crossings in Scale-Space," *IEEE Trans. Acoust., Speech, Signal Processing*, Dec. 1989.

[9] L. Wu and Z. Xie, "Scaling Theorems for Zero-Crossings," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, pp. 46-54, Jan. 1991.

[10] V. Anh, J.Y. Shi, and H.T. Tsui, "Scaling Theorems for Zero Crossings of Bandlimited Signals," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 3, pp. 309-320, Mar. 1996.

[11] P. Perona and J. Malik, "Scale-Space and Edge Detection Using Anisotropic Diffusion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629639, July 1990.

[12] *Geometry-Driven Diffusion in Computer Vision*, B.M. ter Haar Romeny, ed. Kluwer, 1994.

[13] G. H. Granlund, "In Search of a General Picture Processing Operator," *Computer Graphics and Image Processing*, vol. 8, pp. 155-173, 1978.

[14] E. Adelson and J. Bergen, "Spatiotemporal Energy Models for the Perception of Motion," *J. Optical Soc. Am.*, vol. 2, no. 2, pp. 284-299, 1985.

[15] J. Daugman, "Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two Dimensional Visual Cortical Filters," *J. Optical Soc. Am.*, vol. 2A, no. 7, pp. 1,160-1,169, 1985.

[16] M.C. Morrone and R.A. Owens, "Feature Detection From Local Energy," *Pattern Recognition Letters*, vol. 6, pp. 303-313, 1987.

[17] R. Owens, S. Venkatesh, and J. Ross, "Edge Detection is a Projection," *Pattern Recognition Letters*, vol. 9, pp. 233-244, 1989.

[18] W.T. Freeman and E.H. Adelson, "The Design and Use of Steerable Filters," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, pp. 891-906, 1991.

[19] P. Perona and J. Malik, "Detecting and Localizing Edges Composed of Steps, Peaks and Roofs," *Proc. Third Int'l Conf. Computer Vision*, Osaka, Japan, pp. 52-57, 1990.

[20] P. Perona and J. Malik, "Boundary Detection Using Quadratic Filters: Performance Criteria and Experimental Assessment," *Proc. SPIE Conf. on Applications of Artificial Intelligence*, SPIE vol. 1,708, Orlando, Apr. 1992.

[21] P. Kube, "Properties of Energy Edge Detectors," *Proc. IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition*, 1992.

[22] P. Perona, "Deformable Kernels for Early Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 488-499, May 1995.

[23] G.H. Granlund and H. Knuttson, *Signal Processing for Computer Vision*. Kluwer, 1995.

[24] T. Poston and I. Stewart, *Catastrophe Theory and its Applications*. London: Pitman, 1978.

[25] V.I. Arnold, *Singularity Theory*. Cambridge Univ. Press, 1981.

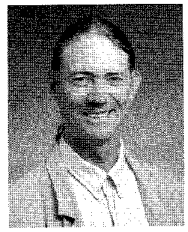
[26] R.N. Bracewell, *The Fourier Transform and Its Applications*. McGraw Hill, 1978.

[27] M.C. Morrone and D.C. Burr, "Feature Detection in Human Vision: A Phase Dependent Energy Model," *Proc. Royal Soc. of London*, vol. B235, pp. 221-245, 1988.

[28] C. Ronse, "On Idempotence and Related Requirements in Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, pp. 484-491, 1993.

[29] A.I. Barvinok, "Feasibility Testing for Systems of Real Quadratic Equations," *STOC*, 1992.

[30] P. Kube and P. Perona, "Scale-Space Properties of Quadratic Edge Detectors," *Proc. Third European Conf. Computer Vision*, J.-O. Eklundh, ed., *LNCS-Series*, vol. 800-801, pp. I-115-122. Stockholm: Springer-Verlag, May 1994.



Paul Kube received the SB degree from MIT and the PhD from the University of California at Berkeley. Since 1988 he has been on the faculty of the Department of Computer Science and Engineering at the University of California, San Diego, where he is also a member of the Institute for Neural Computation, the Institute for Biomedical Engineering, and the Interdisciplinary Group in Cognitive Science. Dr. Kube's research interests include machine vision, pattern recognition, and artificial intelligence.

Pietro Perona is an assistant professor of electrical engineering with the California Institute of Technology. He was graduated with a degree in electrical engineering from the University of Padova, Italy, in 1985 and received a PhD in electrical engineering and computer science from the University of California at Berkeley in 1990.

In 1990-91 he was a postdoctoral fellow with the Laboratory for Information and Decision Systems of MIT. He is also an adjunct associate professor with the University of Padova, Italy. In 1994 Dr. Perona received an NSF National Young Investigator award. Dr. Perona's research interests are in computer vision, visual psychophysics, computational modeling of human vision, and prosciutto making. His work includes research on partial differential equations for visual processing, modeling of human texture segmentation and classification, visual psychophysics on texture popout, shape perception and motion perception, 3D motion analysis for visual navigation, and object recognition.