Local Scale Control for Edge Detection and Blur Estimation

James H. Elder, *Member, IEEE*, and Steven W. Zucker, *Fellow, IEEE*

Abstract—The standard approach to edge detection is based on a model of edges as large step changes in intensity. This approach fails to reliably detect and localize edges in natural images where blur scale and contrast can vary over a broad range. The main problem is that the appropriate spatial scale for local estimation depends upon the local structure of the edge, and thus varies unpredictably over the image. Here we show that knowledge of sensor properties and operator norms can be exploited to define a unique, locally computable *minimum reliable scale* for local estimation at each point in the image. This method for local scale control is applied to the problem of detecting and localizing edges in images with shallow depth of field and shadows. We show that edges spanning a broad range of blur scales and contrasts can be recovered accurately by a single system with no input parameters other than the second moment of the sensor noise. A natural dividend of this approach is a measure of the thickness of contours which can be used to estimate focal and penumbral blur. Local scale control is shown to be important for the estimation of blur in complex images, where the potential for interference between nearby edges of very different blur scale requires that estimates be made at the minimum reliable scale.

Index Terms—Edge detection, localization, scale space, blur estimation, defocus, shadows.

———————————— ✦ ————————————

1 INTRODUCTION

EDGE detectors are typically designed to recover step discontinuities in an image (e.g., [1], [2], [3], [4]), however the boundaries of physical structures in the world generally do not project to the image as step discontinuities. On the left of Fig. 1 is depicted a straight, sharp reflectance edge, slightly off the object plane for the lens system shown, so that the physical edge projects to the image as a blurred luminance transition. In the center of Fig. 1 is shown the shadow of a straight-edged object cast by a spherical light source onto a flat ground surface. Because the light-source is not a point-source, the shadow exhibits a penumbra which causes the shadow edge to appear blurred. On the right is shown a slightly rounded object edge which, when illuminated and viewed from above, also generates a blurred edge in the image.

Since cameras and eyes have finite depth-of-field, light sources are seldom point sources and objects are often smooth, edges in the world will generically project to the image as blurred luminance transitions. This paper generalizes the detection of step discontinuities to encompass this broader, more physically realistic class of edges.

It is important first to distinguish what can and cannot be computed locally at an edge. We have shown [5] that there is in fact a duality between the focal blur and cast shadow scenarios depicted in Fig. 1. Under this duality, the

light source, occluder and ground plane components which constitute the cast shadow model may be exchanged for the aperture, reflectance edge and sensor plane which comprise the geometric optics model of focal blur. Specifically, both situations predict a sigmoidal luminance transition of exactly the same form:

$$I(x) = f(x/r)$$

where

$$f(u) = \frac{1}{\pi}\left(\arccos u - u\sqrt{1 - u^2}\right)$$

A shaded object edge can also be shown to mimic this intensity pattern [5].

The parameter r in this equation determines the degree of blur in the edge. For the case of focal blur, r is determined by the size of the aperture and the relative distances between the lens, image plane and sensor plane. For a cast shadow, the relevant variables are the visual angle of the light source and the distance between the occluder and the

- *J. Elder is with the Centre for Vision Research, Department of Psychology, York University, 4700 Keele St., North York, ON, Canada M3J 1P3. E-mail: jelder@yorku.ca.*
- *S. Zucker is with the Center for Computational Vision and Control, Departments of Computer Science and Electrical Engineering, Yale University, 51 Prospect St., P.O. Box 208285, New Haven, CT 06520–8285. E-mail: zucker–steven@cs.yale.edu.*
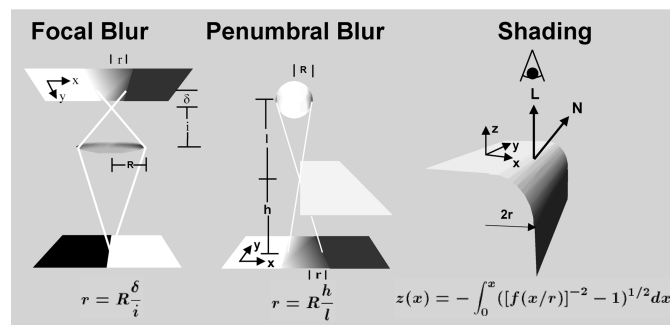
Fig. 1. Edges in the world generically project to the image as spatially blurred. From left to right: Focal blur due to finite depth-of-field; penumbral blur at the edge of a shadow; shading blur at a smoothed object edge.

(a)   (b)   (c)
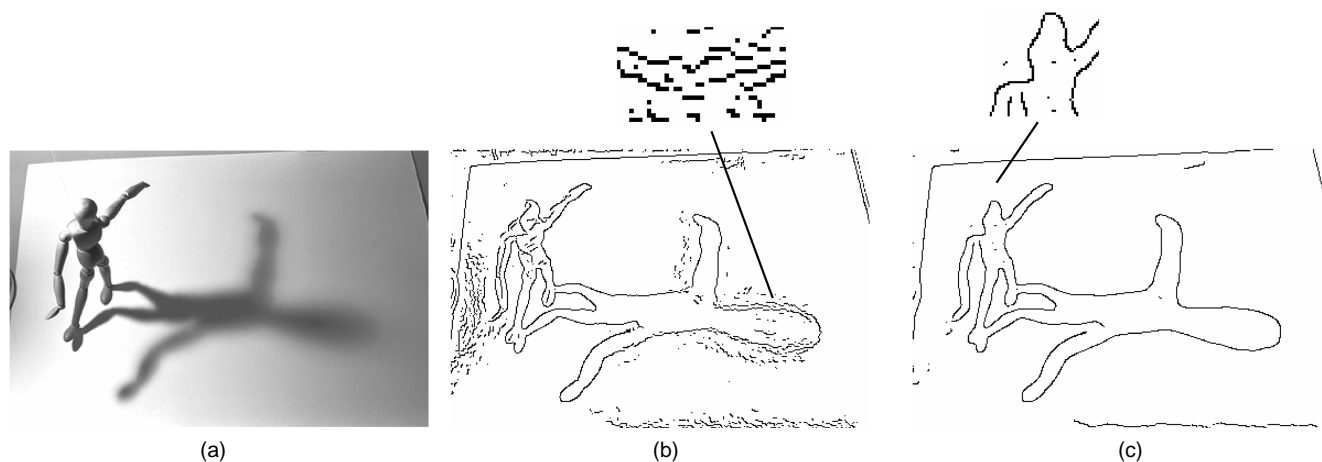
Fig. 2. The problem of local estimation scale. Different structures in a natural image require different spatial scales for local estimation. (a) The original image contains edges over a broad range of contrasts and blur scales. (b) The edges detected with a Canny/Deriche operator tuned to detect structure in the mannequin. (c) The edges detected with a Canny/Deriche operator tuned to detect the smooth contour of the shadow. Parameters are ($\alpha = 1.25$, $\omega = 0.02$) and ($\alpha = 0.5$, $\omega = 0.02$), respectively. See [2] for details of the Deriche detector.

ground surface. For a shaded edge, r is determined by the curvature of the surface. In natural scenes, these variables may assume a wide range of values, producing edges over a broad range of blur scales.

Our conclusions are twofold. First, edges in the world generically project to the image as sigmoidal luminance transitions over a broad range of blur scales. Second, we cannot restrict the goal of the local computation to the detection of a specific *type* of edge (e.g., occlusion edges), since we expect different types of edges to be locally indistinguishable. Thus, the goal of the local computation must be to detect, localize and characterize *all* edges over this broad range of conditions, regardless of the physical structures from which they project.

To illustrate the challenge in achieving this goal, consider the scene shown in Fig. 2. Because the light source is not a point source, the contour of the cast shadow is not uniformly sharp. The apparent blur is of course the penumbra of the shadow: that region of the shadowed surface where the source is only partially eclipsed.

Fig. 2b shows the edge map generated by the Canny/Deriche edge detector [1], [2], tuned to detect the details of the mannequin (the scale parameter and thresholds were adjusted by trial and error to give the best possible result). At this relatively small scale, the contour of the shadow cannot be reliably resolved and the smooth intensity gradients behind the mannequin and in the foreground and background are detected as many short, disjoint curves. Fig. 2c shows the edge map generated by the Canny/Deriche edge detector tuned to detect the contour of the shadow. At this larger scale, the details of the mannequin cannot be recovered, and the contour of the shadow is fragmented at the section of high curvature under one arm.

This example suggests that to process natural images, operators of multiple scales must be employed. This conclusion is further supported by findings that the receptive fields of neurons in the early visual cortex of cat [6] and primate [7] are scattered over several octaves in size. While this conclusion has been reached by many computer vision

researchers (e.g., [8], [1], [9], [10], [11], [12]), the problem has been and continues to be: Once a scale space has been computed, how is it used? Is there any principled way to combine information over scale, or to reason within this scale space, to produce usable assertions about the image?

In this paper, we develop a novel method for local scale adaptatation based upon two goals:

1) Explicit testing of the statistical reliability of local inferences.
2) Minimization of distortion in local estimates due to neighboring image structures.

This method for reliable estimation forms the basis for generalizing edge detection to the detection of natural image edges over a broad range of blur scales and contrasts. Our ultimate objective is the detection of *all* intensity edges in a natural image, regardless of their physical cause (e.g., occlusions, shadows, textures).

2 SCALE SPACE METHODS IN EDGE DETECTION

The issue of scale plays a prominent role in several of the best-known theories of edge detection. Marr and Hildreth [8] employed a Laplacian of Gaussian operator to construct zero-crossing segments at a number of scales and proposed that the presence of a physical edge be asserted if a segment exists at a particular position and orientation over a contiguous range of scale. Canny [1] defined edges at directional maxima of the first derivative of the luminance function and proposed a complex system of rules to combine edges detected at multiple scales. The main problem with these methods is the difficulty in distinguishing whether nearby responses at different scales correspond to a single edge or to multiple edges.

Continuous scale-space methods applied to edge detection have also tended to be complex [13]. In an anisotropic diffusion network [14], the rate of diffusion at each point is determined by a space- and time-varying conduction coefficient which is a decreasing function of the estimated gradient magnitude of the luminance function at the point.

Thus, points of high gradient are smoothed less than points of low gradient. While this is clearly a powerful framework for image enhancement, our goal here is not to sharpen edges, but to *detect* them over a wide range of blur and contrasts. For this purpose, the most troubling property of anisotropic diffusion is its implicit use of a unique threshold on the luminance gradient, below which the gradient decreases with time (smoothing), above which the gradient increases with time (edge enhancement). Unfortunately, there is no principled way of choosing this threshold even for a single image, since important edges generate a broad range of gradients, determined by contrast and degree of blur. Sensor noise, on the other hand, can generate very steep gradients.

Edge focusing methods [11], [15] apply the notion of coarse-to-fine tracking developed for matching problems to the problem of edge detection. The approach is to select the few events (e.g., zero-crossings) that are "really significant" (i.e., survive at a large scale), and then track these events through scale-space as scale is decreased to accurately localize the events in space.

Aside from the computational complexity of this approach, there are two main problems with its application to edge detection. First, for edges, scale has no reliable correspondence to significance. Generally, edges will survive at a higher scale if they are high contrast, in focus, and isolated. However, contrast is a very poor indicator of significance, since objects with similar reflectance functions, when placed in occlusion, can generate very low contrast edges. This is a frequent occurrence, partly because of self-occlusions and the fact that objects with similar reflectance functions are often grouped together in the world (e.g., the pages in front of you now). While one can make the argument that the important objects should be in focus, in practical situations this is often impossible. Finally, the isolation of an edge also does not indicate significance: the proximity of one object to another does not lessen the significance of either object in general. Thus it is inappropriate to use the high end of scale space to select for edge significance.

The second problem lies in the assumption that optimal localization accuracy can be attained at the finest scale. While it is true that at coarse scales the trace of an event tends to wander in space due to interaction with neighboring events, the exact variation in accuracy with scale depends strongly on the separation of neighboring events relative to the level of sensor noise [1]. If events are widely spaced, localization accuracy in fact increases with scale. If sensor noise is high, localization accuracy can become very poor at finer scales.

Jeong and Kim [16] have proposed an adaptive method for estimating a unique scale for edge detection at each point in the image. They pose the problem as the minimization of a functional over the entire image, and use a relaxation method to solve the resulting nonconvex optimization problem. The authors report that their results suffered from the complicated shape of the objective function, and the resulting sensitivity of the selected scale to the initial guess.

More recently, Lindeberg has proposed a method for selecting local scale for edge detection based upon maximiz-ing a heuristic measure of edge strength [17], [18]. The main difficulty with this approach is that the scale thus selected is often too small to provide reliable estimates of the derivatives upon which edge detection is based, leading to dense edge maps containing many artifactual edges. In an attempt to distinguish real edges from artifact, Lindeberg proposes a more global post-detection stage in which a measure of edge significance is defined and integrated along connected chains of pixels. Only edge chains above some (unspecified) threshold value are then considered important. In contrast, our goal in this paper is to develop a completely local method for scale selection which does not require this type of post-processing heuristic to distinguish real edges from artifact.

Our departure from standard approaches to the scale problem in edge detection can thus be summarized by the following observations:

1) There exists no natural scale or gradient threshold which can be defined a priori to distinguish edges from nonedges.
2) Survival at large scales does not distinguish significant from insignificant edges.
3) Localization is not, in general, best at the finest scales.
4) To avoid artifactual edges, selected scales must be large enough to provide reliable derivative estimates.

3 MINIMUM RELIABLE SCALE

The difficulty in reliably recovering structure from images such as Fig. 2 is that the appropriate scale for estimation varies over the image. However, while the scale of the image structure is space-variant, the system used to produce the image from the scene is fixed. This is the typical situation in computer vision: one doesn't know "what's out there," but one *does* know the properties of the sensor, and one can thus compute the statistics of the sensor noise in advance [19]. Given a specific model of the event to be detected (in this case, a luminance edge), and the appropriate operators to be used for this purpose, one can then relate the parameters of the model to a unique minimum scale at which the event can be reliably detected. We call this unique scale the *minimum reliable scale* for the event.

By reliable here we mean that at this and larger scales, the likelihood of error due to sensor noise is below a standard tolerance (e.g., 5 percent for an entire image). This definition does not account for the danger of incorrect assertions due to the influence of scene events nearby in the image, which in any band-limited system must be an increasing function of scale. While attempts have been made by others to explicitly model this phenomenon [1], it is our view that this problem is unlikely to admit such a general solution. For example, while an ensemble of images may yield an estimate of the expected separation between edges, if a sample of the ensemble contains a fine corduroy pattern, this estimate will be of little use. Rather than relying upon such uncertain priors, we argue that the smallest scale which provides a reliable estimate should be used. By selecting the minimum reliable scale at each point in the image, we prevent errors due to sensor noise

while simultaneously minimizing errors due to interference from nearby structure.

By identifying a unique scale for local estimation, we avoid the complexities and ad hoc decisions required to combine responses at multiple scales [8], [1] or to track edges through scale-space [11], [15]. Since the computation is entirely local, we avoid the complexity of the global computation proposed by Jeong and Kim [16]. By adhering to a strict criterion for reliability with respect to sensor noise, we obviate the need for complex postprocessing heuristics to distinguish real from artifactual edges [18].

4 MODELING EDGES, BLUR, AND SENSING

An edge is modeled as a step function $Au(x) + B$ of unknown amplitude A and pedestal offset B, which, for the purposes of this discussion, will be aligned with the y-axis of the image coordinate frame. The focal or penumbral blur of this edge is modeled by the Gaussian blurring kernel[1]

$$g(x, y, \sigma_b) = \frac{1}{2\pi\sigma_b^2} e^{-(x^2+y^2)/2\sigma_b^2}$$

of unknown scale constant σ_b, to generate the error function

$$(A/2)\left(\mathrm{erf}\left(x/\sqrt{2}\,\sigma_b\right) + 1\right) + B$$

Sensor noise $n(x, y)$ is modeled as a stationary, additive, zero-mean white noise process; that is, the noise at a given point in the image is a normally distributed random variable with standard deviation s_n, independent of the signal and the noise at other points in the image. The complete edge model is thus:

$$(A/2)\left(\mathrm{erf}\left(x/\sqrt{2}\,\sigma_b\right) + 1\right) + B + n(x, y) \qquad (1)$$

Estimating the sensor noise statistics for an imaging system is relatively straightforward. For the system used in Fig. 2, a region of a defocused image of a plain flat surface was first selected. This subimage was then high-pass filtered with a unit-power kernel, the two-tap filter $\left(1/\sqrt{2}, -1/\sqrt{2}\right)$. The shading over this subimage varies slowly, and the defocus acts as an additional low-pass filter, so we could be confident that the scene structure contributed negligible energy to the filter output. The following elementary result from the theory of random processes was then exploited:

PROPOSITION 1. *The standard deviation of the result of a linear transformation $\mathcal{L}: \mathfrak{R}^n \rightarrow \mathfrak{R}$ applied to a set of i.i.d. random variables of standard deviation s_n is the product of the L_2 norm of the linear transformation and the standard deviation of the random variables: $s_{\mathcal{L}} = \|\mathcal{L}\|_2 s_n$.*[2]

The proof is straightforward [23].

Thus, the statistics of the unit-power filter output provide an estimate of the statistics of the sensor noise: the standard deviation of the noise for the eight-bit image shown in Fig. 2 is approximately 1.6 quantization levels.

5 RELIABILITY CRITERION

Our edge-detection method depends upon making reliable inferences about the local shape of the intensity function at each point in an image. Reliability is defined in terms of an overall significance level α_I for the entire image and a pointwise significance level α_p. We fix the overall image significance level α_I at 5 percent for the entire image, i.e., we demand that the probability of committing one or more Type I (false positive) errors over all image points be less than 5 percent. Under the assumption of i.i.d. noise, the number of pixels n in the image then determines the pointwise significance level:

$$\alpha_p = 1 - (1 - \alpha_I)^{1/n} \qquad (2)$$

For simplicity, we fix α_p at a constant value determined from the maximum image size used in our experiments: $n = 512 \times 512$ pixels $\rightarrow \alpha_p = 2.0 \times 10^{-7}$. This ensures that the overall significance level α_I is less than 5 percent for all images used in this paper.

6 LOCAL SCALE CONTROL AND GRADIENT DETECTION

A necessary condition for the assertion of an edge at a particular location in the image is a nonzero gradient in the luminance function. The gradient can be estimated using steerable Gaussian first derivative basis filters [24], [25]:

$$g_1^x(x, y, \sigma_1) = \frac{-x}{2\pi\sigma_1^4} e^{-(x^2+y^2)/2\sigma_1^2}$$

$$g_1^y(x, y, \sigma_1) = \frac{-y}{2\pi\sigma_1^4} e^{-(x^2+y^2)/2\sigma_1^2}$$

where σ_1 denotes the scale of the first derivative Gaussian estimator. The response $r_1^\theta(x, y, \sigma_1)$ of a first derivative Gaussian filter $g_1^\theta(x, y, \sigma_1)$ to an image $I(x, y)$ in an arbitrary direction θ can be computed exactly as a weighted sum of the basis filter responses, so that if

$$r_1^x(x, y, \sigma_1) = g_1^x(x, y, \sigma_1) * I(x, y)$$

and

$$r_1^y(x, y, \sigma_1) = g_1^y(x, y, \sigma_1) * I(x, y)$$

then

$$r_1^\theta(x, y, \sigma_1) = \cos(\theta) r_1^x(x, y, \sigma_1) + \sin(\theta) r_1^y(x, y, \sigma_1)$$

At nonstationary points of the luminance function, $r_1^\theta(x, y, \sigma_1)$ has a unique maximum over θ, the gradient magnitude $r_1^{\theta_M}(x, y, \sigma_1)$, attained in the gradient direction $\theta_M(x, y, \sigma_1)$:

---

1. The Gaussian model differs from the geometric model for focal and penumbral blur discussed in Section 1. It has been argued that the Gaussian model better accounts for the various aberrations in practical imaging systems, and it has been used widely in depth-from-defocus work [20], [21], [22].

2. We use the Gaussian function both as a model for the probability distribution of sensor noise and as a smoothing function for local estimation. For clarity, we use the symbol s for the standard deviation of the Gaussian when it is used as a model of noise, and the symbol σ for the scale of the Gaussian when it is used as a smoothing filter.

$$r_1^{\theta_M}(x, y, \sigma_1) = \sqrt{\left(r_1^x(x, y, \sigma_1)\right)^2 + \left(r_1^y(x, y, \sigma_1)\right)^2}$$

$$\theta_M(x, y, \sigma_1) = \arctan\left(r_1^y(x, y, \sigma_1)/r_1^x(x, y, \sigma_1)\right)$$

To be confident that a nonzero gradient exists at a point (x, y) in the image $I(x, y)$ we must consider the likelihood that the response of the gradient operator could be due to noise alone. This computation is complicated by the fact that the gradient operator is nonlinear, and so its response will not be normally distributed. However, it is possible to determine its distribution by exploiting a second elementary result from probability theory [26]:

PROPOSITION 2. *Let U be a random variable with pdf p_U, $U \in A \subseteq \Re$. Let $V = f(U)$, where f is diffeomorphic on A. Then*

$$p_V(v) = p_U\left(f^{-1}(v)\right)\left|\frac{d}{dv} f^{-1}(v)\right|, v \in f(A).$$

This proposition can be used to derive the distribution of the gradient response to noise. We let

$$p_U(u) = p\left[\left|r_1^x\right| = u\right] = p\left[\left|r_1^y\right| = u\right]$$

and

$$p_V(v) = p\left[\left(r_1^x\right)^2 = v\right] = p\left[\left(r_1^y\right)^2 = v\right].$$

If the image contains only Gaussian i.i.d. noise of standard deviation s_n, U will have a half-Gaussian distribution

$$p_U(u) = \frac{2}{\sqrt{2\pi}s_1} e^{-u^2/2s_1^2}, u \in [0, \infty)$$

where

$$s_1 = \left\|g_1(x, y, \sigma_1)\right\|_2 s_n \qquad \text{(Proposition 1)}$$

Since $f(u) = u^2$ is diffeomorphic on $[0, \infty)$, by Proposition 2,

$$p_V(v) = \frac{1}{\sqrt{2\pi v}s_1} e^{-v/2s_1^2}, v \in [0, \infty)$$

Now let

$$p_{V_1+V_2}(v) = p\left[\left(r_1^x\right)^2 + \left(r_1^y\right)^2 = v\right] = \int_0^v p_V(v')p_V(v - v')dv'$$

Solving the integral, we obtain

$$p_{V_1+V_2}(v) = \frac{1}{2s_1^2} e^{-v/2s_1^2}, v \in [0, \infty) \qquad (3)$$

To ensure a pointwise significance of α_p, we require a critical value c_1, below which responses are not considered reliable, satisfying

$$\int_{c_1^2}^{\infty} p_{V_1+V_2}(v)dv = \alpha_p.$$

Substituting from (3) and solving, we obtain

$$c_1 = s_1\sqrt{-2\ln(\alpha_p)},$$

where

$$s_1 = \left\|g_1(x, y, \sigma_1)\right\|_2 s_n.$$

Substituting $\alpha_p = 2.0 \times 10^{-7}$ (Section 5), we have $c_1 = 5.6 s_1$. The L_2 norm of the first derivative operator is given by

$$\left\|g_1^\theta(x, y, \sigma_1)\right\|_2 = 1/\left(2\sqrt{2\pi}\sigma_1^2\right),$$

hence, we have the following.

DEFINITION 1. *The critical value function $c_1(\sigma_1)$ of the nonlinear Gaussian gradient operator, tracing the threshold of statistical reliability ($\alpha = .05$) for the operator as a function of scale, is given by*

$$c_1(\sigma_1) = \frac{1.1 s_n}{\sigma_1^2}. \qquad (4)$$

For a given level of sensor noise and operator scale, the critical value function specifies the minimum response value that can be considered statistically reliable.

In order to relate the critical value function to edge detection, we must consider the response of the gradient operator to an edge. Given a blurred step edge along the y axis of amplitude A and blur parameter σ_b, the gradient magnitude is given by

$$r_1^x(x, y, \sigma_1) = Au(x) * g_1^x\left(x, y, \sqrt{\sigma_b^2 + \sigma_1^2}\right)$$

$$= \frac{A}{\sqrt{2\pi\left(\sigma_b^2 + \sigma_1^2\right)}} e^{-x^2/2\left(\sigma_b^2 + \sigma_1^2\right)} \qquad (5)$$

which attains its maximum on the y axis:

$$r_1^x(0, y, \sigma_1) = \frac{A}{\sqrt{2\pi\left(\sigma_b^2 + \sigma_1^2\right)}} \qquad (6)$$

Thus, while both the maximum gradient response to a blurred edge (6) and the critical value function (4) decrease with increasing estimation scale, the critical value function decreases faster, reflecting the improved signal detection properties of larger oriented operators. By combining (6) with (4) and solving for σ_1, we derive Proposition 3.

PROPOSITION 3. *For an imaging system with white sensor noise of standard deviation s_n, and an edge of amplitude A and blur parameter σ_b, there exists a minimum reliable scale $\hat{\sigma}_1$ at which the luminance gradient can be reliably detected:*

$$\hat{\sigma}_1^2 = \frac{s_n}{A}\left(5.4 + \sqrt{28.9 + 15.2\left(A\sigma_b/s_n\right)^2}\right)\text{pixels}^2.$$

This situation for an example edge is shown in Fig. 3a. The minimum reliable scale for estimating the gradient of the edge is defined by the scale at which the edge response just exceeds the significance threshold: at $\hat{\sigma}_1 = 2.1$ pixels in this case.
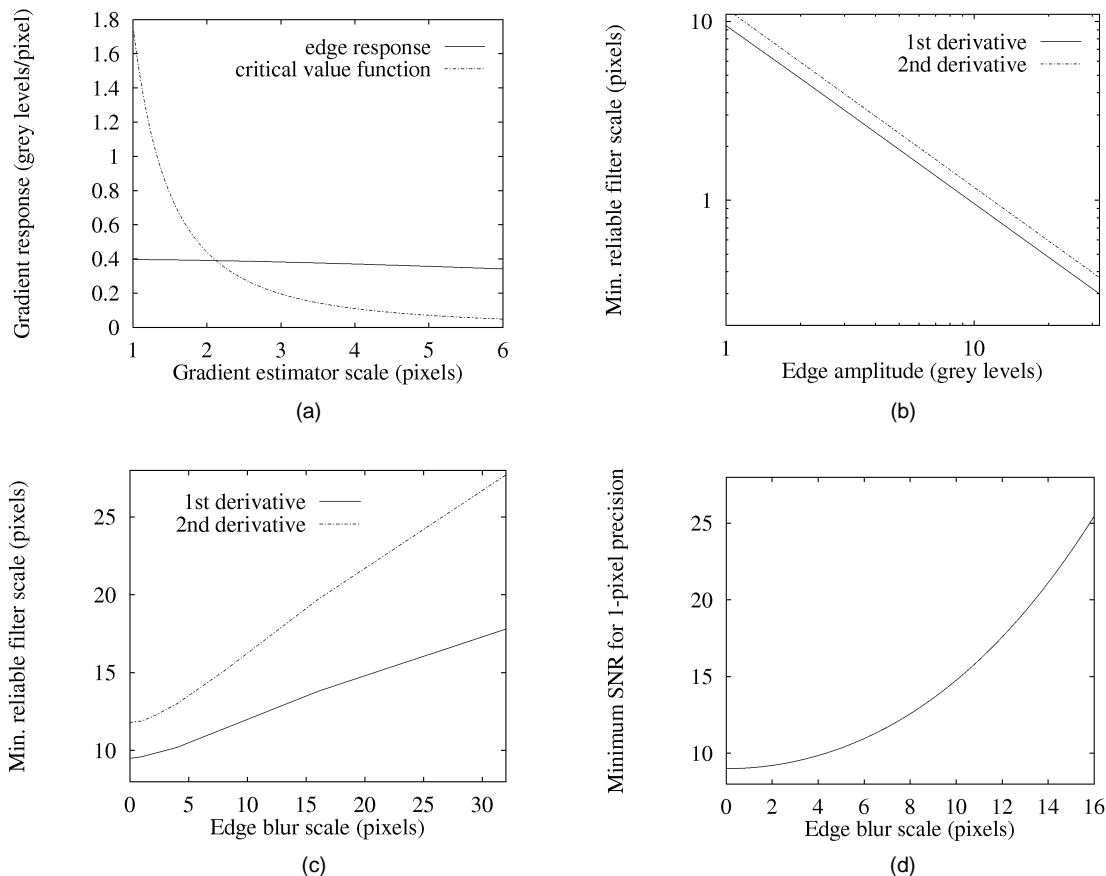
Fig. 3. Predicted performance characteristics of local scale control. (a) Local scale control for a simulated edge. Parameters are: A = 10 gray levels, B = 127 gray levels, σ_b = 10 pixels, s_n = 1.6 gray levels (SNR = 6.3). The intersection of the critical value function $c_1(\sigma_1)$ with the maximum gradient response to the edge $r_1^x(0, y, \sigma_1)$ determines the minimum reliable scale for gradient estimation. (b) Minimum reliable scales $\hat{\sigma}_1$ and $\hat{\sigma}_2$ to detect a sharp edge (σ_b = 0), as a function of edge amplitude A. (c) Minimum reliable scales $\hat{\sigma}_1$ and $\hat{\sigma}_2$ to detect a low-contrast edge (A = 1 gray level), as a function of edge blur σ_b. (d) Minimum SNR required to localize a high-precision edge to the nearest pixel (ϵ = 0.5), as a function of blur scale σ_b.

Note that estimating the minimum reliable scale accurately does not allow one to estimate the blur of the edge σ_b, since the amplitude A of the edge is also unknown. In our experiments, therefore, we attempt only to stay close to the minimum reliable scale by computing gradient estimates at octave intervals of scale, at each point using the smallest scale at which the gradient estimate exceeds the critical value function, i.e.,

$$\hat{\sigma}_1(x, y) = \inf\left\{\sigma_1 : r_1^{\theta_M}(x, y, \sigma_1) > c_1(\sigma_1)\right\}$$

The result of the gradient computation using local scale control for the image of the mannequin and shadow is shown in Fig. 4a. Six scales were used: $\sigma_1 \in \{0.5, 1, 2, 4, 8, 16\}$ pixels. Here the shade of gray indicates the smallest scale at which gradient estimates are reliable, black indicating σ_1 = 0.5 pixel, lighter shades indicating higher scales, and white indicating that no reliable estimate could be made. While the smallest scale is reliable for the contours of the mannequin, higher scales are required to fully recover the shadow.

Many edge detectors (e.g., [1]) define edges as local maxima in the gradient map. Fig. 5 shows why this approach cannot work. A one-dimensional cross-section from the penumbra of the shadow has been selected to examine the behavior of the derivative responses and minimum reliable scales for local estimation. Fig. 5b shows the luminance profile in cross-section, Fig. 5d shows the gradient magnitude along the cross-section, and Fig. 5c shows the minimum reliable scales at which the gradient was estimated. Note how the scale of estimation automatically adapts as the strength of the signal varies. Although this allows the gradient to be reliably detected as nonzero over this cross-section, the response is not unimodal: there are in fact five maxima in the gradient along the cross section of the edge. It is clear that selecting the maxima of the gradient function would lead to multiple separate responses to this single edge. While there is no local solution to this multiple response problem based upon the gradient map alone, in the next section we show how reliable estimation of the second derivative of the intensity function can be used to solve this problem.
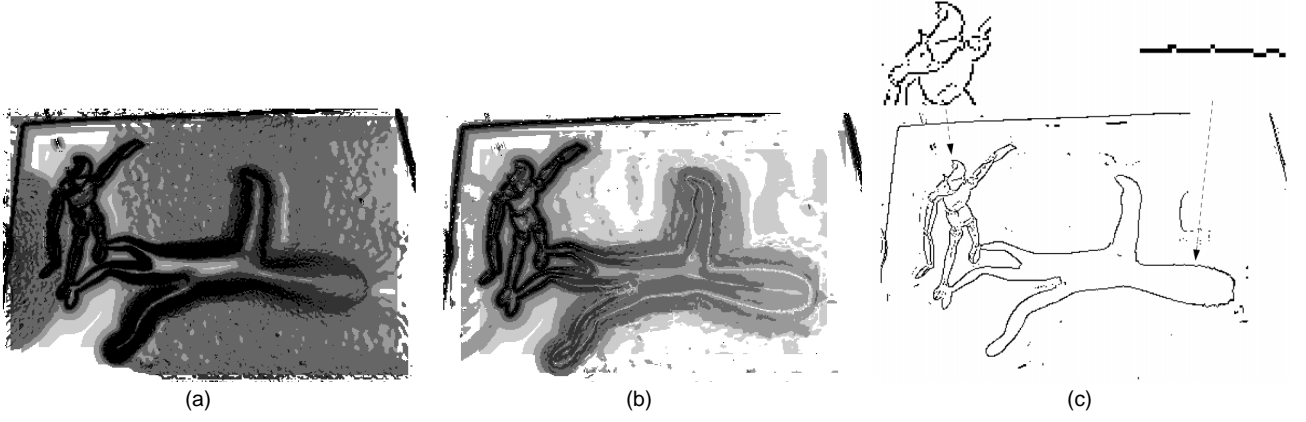
Fig. 4. Results of local scale control for image of mannequin with shadow. For scale maps, larger scales are rendered in lighter gray, white indicates that no reliable estimate could be made. (a) Map of minimum reliable scale for gradient estimation. (b) Map of minimum reliable scale for second derivative estimation. (c) Detected edges. Note that both the fine detail of the mannequin and the blurred, low-contrast shadow are reliably recovered.

7   LOCAL SCALE CONTROL AND SECOND DERIVATIVE ESTIMATION

The second derivative of the intensity function can be estimated with a steerable second derivative of Gaussian operator:

$$g_2^x(x, y, \sigma_2) = \frac{1}{2\pi\sigma_2^4}\left((x/\sigma_2)^2 - 1\right)e^{-(x^2+y^2)/2\sigma_2^2}$$

$$g_2^y(x, y, \sigma_2) = \frac{1}{2\pi\sigma_2^4}\left((y/\sigma_2)^2 - 1\right)e^{-(x^2+y^2)/2\sigma_2^2}$$

$$g_2^{xy}(x, y, \sigma_2) = \frac{xy}{2\pi\sigma_2^6}e^{-(x^2+y^2)/2\sigma_2^2}$$

and

$$g_2^\theta(x, y, \sigma_2) = \cos^2(\theta)g_2^x(x, y, \sigma_2)$$
$$- 2\cos(\theta)\sin(\theta)g_2^{xy}(x, y, \sigma_2)$$
$$+ \sin^2(\theta)g_2^y(x, y, \sigma_2)$$

We restrict our attention to the second derivative of the intensity function $g_2^{\theta_M}(x, y, \sigma_2)$ along the gradient direction θ_M. Since $g_2^{\theta_M}(x, y, \sigma_2)$ is linear, the derivation of the critical value function $c_2(\sigma_2)$ for second derivative estimation is relatively straightforward. Specifically, we require that

$$p\left(\left|r_2^{\theta_M}\right| > c_2\right) = \alpha_p \rightarrow erf\left(c_2/\sqrt{2}s_2\right) = \alpha_p$$
$$\rightarrow c_2 = \sqrt{2}s_2\,erf^{-1}(\alpha_p)$$

where

$$s_2 = \left\|g_2(x, y, \sigma_2)\right\|_2 s_n$$

The L_2 norm of the second derivative operator is given by

$$\left\|g_2(x, y, \sigma_2)\right\|_2 = \left(4\sqrt{\pi/3}\sigma_2^3\right)^{-1}.$$

Substituting, and setting $\alpha_p = 2.0 \times 10^{-7}$ (Section 5), we have

$$c_2 = \frac{1.8s_n}{\sigma_2^3} \qquad (7)$$

where s_n is the standard deviation of the noise and σ_2 is the scale of the second derivative Gaussian filter.

Since we are interested only in the luminance variation orthogonal to the edge, at each point in the image the second derivative is steered in the direction of the gradient estimated at the minimum reliable scale. The expected output of the second derivative operator to our local edge model (1) is given by:

$$r_2^x(x, y, \sigma_2) = Au(x) * g_2^x\left(x, y, \sqrt{\sigma_b^2 + \sigma_2^2}\right)$$
$$= \frac{-Ax}{\sqrt{2\pi}\left(\sigma_b^2 + \sigma_2^2\right)^{3/2}}e^{-x^2/2\left(\sigma_b^2 + \sigma_2^2\right)} \qquad (8)$$

As for the gradient, one can show that near an edge there exists a unique minimum scale at which the sign of the second derivative can be reliably determined (Section 8).

A second derivative map is thus obtained which describes, at each point in the image where a significant gradient could be estimated, how this gradient is changing in the gradient direction (if at all). Six scales are employed to estimate the second derivative, at octave intervals: $\sigma_2 \in \{0.5, 1, 2, 4, 8, 16\}$ pixels. The minimum reliable scale map for the mannequin image is shown in Fig. 4b. The coding scheme is as for the gradient scale map.

Fig. 6a illustrates how the second derivative distinguishes between edges and smooth shading gradients. The second derivative zero-crossing localizes the edge, while the flanking extrema of opposite sign indicate the sigmoidal nature of the luminance function, distinguishing it from a smoothly shaded ramp.

The importance of the second derivative in localizing blurred edges is illustrated in Fig. 5. Fig. 5f shows the estimated second derivative steered in the gradient direction, and Fig. 5e shows the minimum reliable scales for these estimates. Note again how scale automatically adapts as the signal varies in strength: Larger scales are needed near the center of the edge where the luminance function is nearly linear. Despite the rockiness of the gradient response, the adaptive second derivative response provides a unique zero-crossing to localize the edge. The key here is that lo-
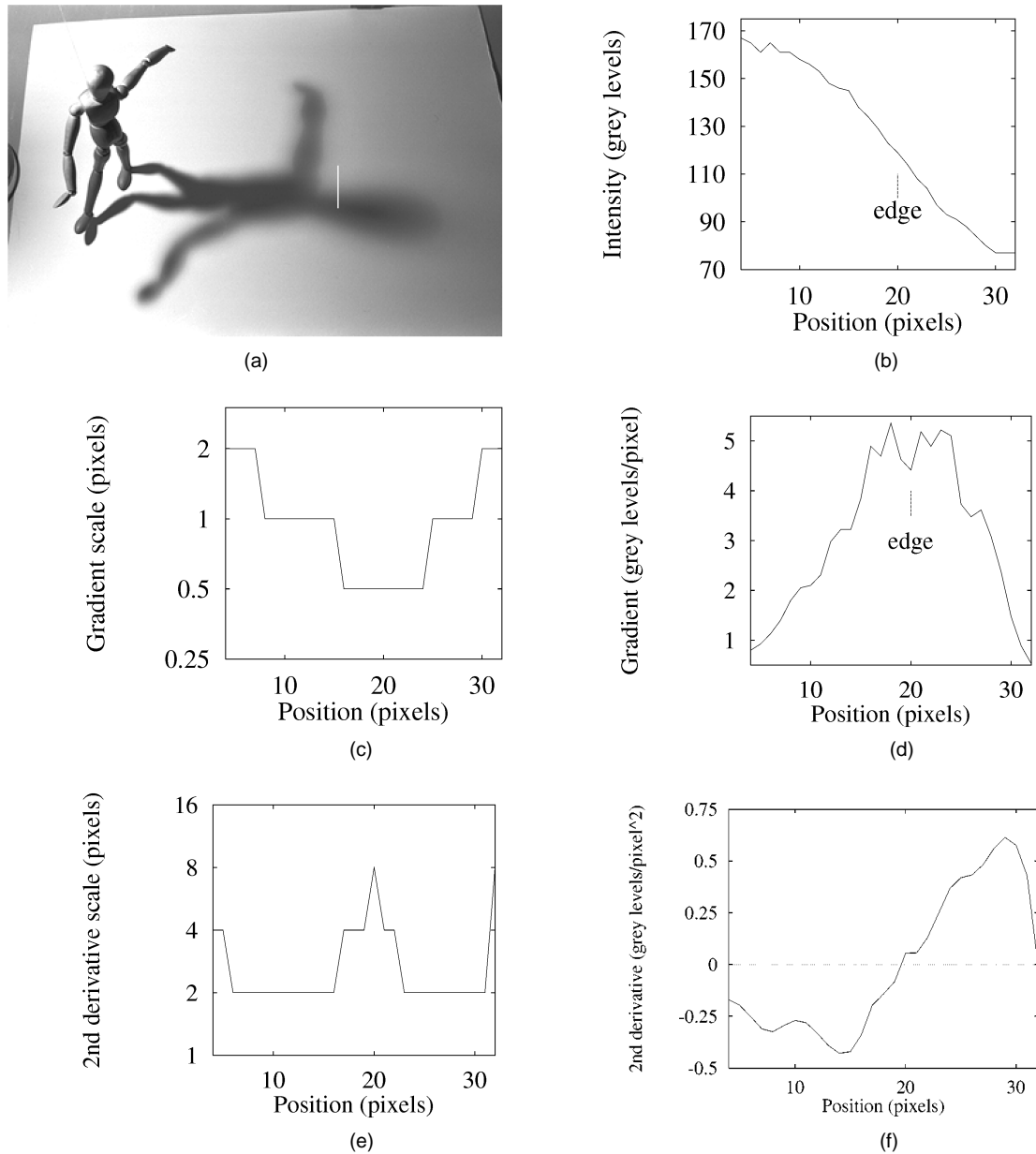
Fig. 5. Unique localization of blurred luminance transitions. (a) Original image with locus of one-dimensional cut used in figures (b)-(f). (b) Luminance function. (c) Minimum reliable scale for the gradient estimate. (d) Estimated gradient magnitude. Note that the signal is not unimodal, possessing five maxima. (e) Minimum reliable scale for the second derivative estimate. (f) Estimated directional second derivative. A unique zero-crossing localizes the edge. The location of the edge is shown by a vertical line segment in (b) and (d).

cal estimation at the minimum reliable scale guarantees that the sign of the second derivative estimate is reliable, and hence that the zero-crossing is unique. The number of peaks in the gradient response, on the other hand, depends on the blur of the edge, and is not revealed in the response of the operator at any single point: ensuring the uniqueness of a gradient maximum is not a local problem. Thus, the reliable detection and localization of blurred edges requires both gradient and second derivative information.

8 ANALYSIS OF DETECTION

As a first step in analyzing the performance of local scale control for edge detection, we can use the edge model of (1)

to predict the range of SNR and blur scale over which edges can be detected, and the range of filter scales required.

To detect the sigmoidal shape of an edge, we must at least reliably determine the sign of the second derivative in the gradient direction at its positive and negative extrema, which occur at the zero-crossings x_+ and x_- of the third derivative of the blurred step edge:

$$r_3^x(x, y, \sigma_2) = \frac{A}{\sqrt{2\pi(\sigma_2^2 + \sigma_b^2)}} \left(\frac{x^2}{\sigma_2^2 + \sigma_b^2} - 1 \right) e^{-\frac{x^2}{2(\sigma_2^2 + \sigma_b^2)}}$$

$$= 0$$

$$\rightarrow x_+ = \sqrt{\sigma_b^2 + \hat{\sigma}_2^2}, \quad x_- = -\sqrt{\sigma_b^2 + \hat{\sigma}_2^2} \qquad (9)$$
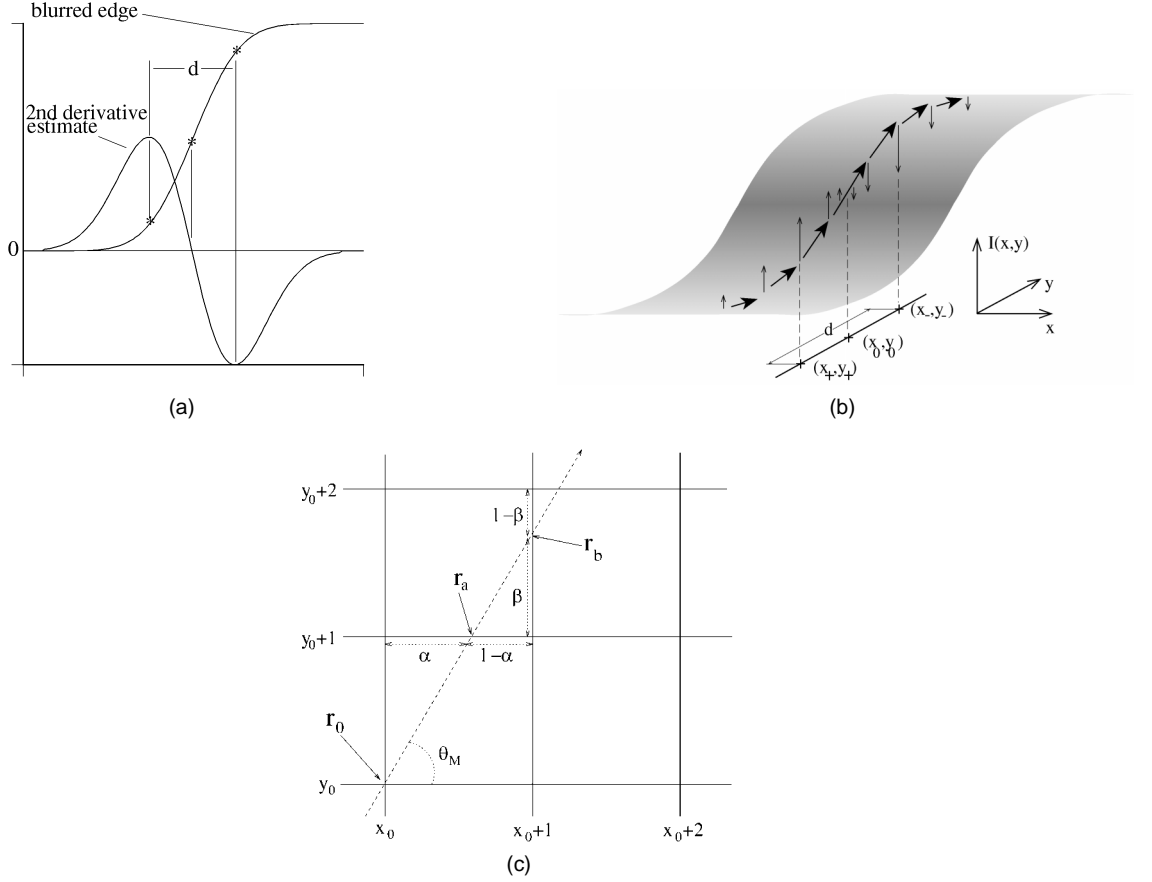
Fig. 6. (a) Ideal blurred step and second derivative response. The zero-crossing of the second derivative localizes the edge, while the distance between its extrema provides a measure of blur scale. (b) The gradient line segment. Thick arrows represent gradient estimates, thin arrows represent second derivative estimates in the gradient direction. (c) Derivative estimation along the gradient line using linear interpolation.

From (7) and (8), the minimum reliable second derivative scale $\hat{\sigma}_2$ must satisfy:

$$r_2^{\theta_M}\left(x_+\right) = -r_2^{\theta_M}\left(x_-\right) = \frac{A}{\sqrt{2\pi e}\left(\sigma_b^2 + \hat{\sigma}_2^2\right)} = \frac{1.8 s_n}{\hat{\sigma}_2^3}$$

and thus

$$A\hat{\sigma}_2^3 - 7.4 s_n\left(\hat{\sigma}_2^2 + \sigma_b^2\right) = 0$$

For a sharp edge where $\sigma_b = 0$, this reduces to

$$\hat{\sigma}_2 = \frac{7.4 s_n}{A} = \frac{7.4}{\text{SNR}},$$

where

$$\text{SNR} = \frac{A}{s_n}$$

To steer the second derivative at these points, we must also reliably detect the gradient of the signal. Thus, from (4) and (5), we obtain for the minimum reliable gradient scale $\hat{\sigma}_1$:

$$r_1^{\theta_M}\left(x_+\right) = r_1^{\theta_M}\left(x_-\right) = \frac{A}{\sqrt{2\pi\left(\sigma_b^2 + \hat{\sigma}_1^2\right)}} e^{-\left(\sigma_b^2 + \hat{\sigma}_2^2\right)/2\left(\sigma_b^2 + \hat{\sigma}_1^2\right)}$$

$$= \frac{1.1 s_n}{\hat{\sigma}_1^2}$$

For a sharp edge where $\sigma_b = 0$, this reduces to

$$\hat{\sigma}_1 e^{-\hat{\sigma}_2^2/2\hat{\sigma}_1^2} = \frac{1.1 s_n}{A} = \frac{1.1}{\text{SNR}}$$

While we cannot solve for $\hat{\sigma}_1$ and $\hat{\sigma}_2$ analytically in the general case, we can solve for $\hat{\sigma}_2$ using iterative techniques for specific values of s_n, A and σ_b, and then solve for $\hat{\sigma}_1$ using the obtained $\hat{\sigma}_2$. Assuming $s_n = 1.6$ gray levels, we have computed $\hat{\sigma}_1$ and $\hat{\sigma}_2$ for a sharp edge ($\sigma_b = 0$) at various amplitudes ($A \in \{1..32\}$ gray levels: Fig. 3b) and for a low-contrast edge ($A = 1$ gray level) at various blurs ($\sigma_b \in \{1..32\}$ pixels: Fig. 3c). Note that these curves represent the minimum scales required for *detection* of an edge. In the next section, we will show that localization precision may require higher second derivative scales (see also [27]).

9 ANALYSIS OF LOCALIZATION

In the example of Fig. 5, the contrast of the shadow was high enough to allow reliable estimation of the second derivative at each pixel along the gradient direction, so that the edge could be localized as a zero-crossing in the second derivative. We refer to such edges as *high-precision edges*. For very low contrast edges, the second derivative signal may be too weak to reliably determine the second derivative

sign near the zero-crossing, even though the second derivative extrema are detected. These edges may thus be detectable under the analysis of the previous section, but not localizable to one-pixel precision. We will refer to these edges as *low-precision edges*.

To understand this localization problem, we must consider the reliability criterion for the second derivative response near the zero-crossing. Defining ϵ as the distance (in pixels) between the actual edge location and the nearest point at which the second derivative sign can be reliably determined, from (7) and (8) we can write:

$$\frac{\epsilon}{\sqrt{2\pi}\left[(\sigma_b/\sigma_2)^2 + 1\right]^{3/2}} e^{-\epsilon^2/2(\sigma_b^2+\sigma_2^2)} = \frac{1.8 s_n}{A} = \frac{1.8}{\text{SNR}} \quad (10)$$

For $\epsilon \ll \sqrt{\sigma_b^2 + \sigma_2^2}$, we can develop a first-order Taylor series for the left-hand side of (10) around $\epsilon = 0$ to derive an approximate solution for ϵ:

$$\epsilon \approx \frac{4.5\left[(\sigma_b + \sigma_2)^2 + 1\right]^{3/2}}{SNR} \text{ pixels}$$

Note that precision improves monotonically as the maximum second derivative filter scale is increased. As $\sigma_2 \to \infty$, precision asymptotically approaches $\epsilon = 4.5/\text{SNR}$ pixels.

In theory, localization precision ϵ is not affected by the blur of the edge, only by the contrast of the edge relative to the sensor noise. To obtain precision to the nearest pixel, we need $\epsilon \approx 0.5$, and thus require that SNR > 9.0. For practical systems, there must be an upper bound on filter scale $\tilde{\sigma}_2$, and thus, to obtain precision to the nearest pixel, we require that

$$\text{SNR} > 9.0\left[(\sigma_b/\tilde{\sigma}_2)^2 + 1\right]^{3/2}$$

For a given maximum filter scale $\tilde{\sigma}_2$, this equation defines a bound on the contrast and blur of edges which can be localized to one-pixel precision. Fig. 3d shows this bound for our implementation, in which we use second derivative scales up to $\sigma_2 = 16$ pixels. SNR (A/s_n) must stay above this curve if an edge of a particular blur scale σ_b is to be localized to the nearest pixel.

When edges are detected but are not localizable to one-pixel precision, how should they be represented in an edge map? In our implementation, we represent the location of low-precision edges by pixels which approximately bisect the gradient line segment connecting the positive and negative extrema of the second derivative response. The exact algorithm used to detect and represent low- and high-precision edges is detailed in Section 11.

10 IMPLEMENTATION ON A DISCRETE GRID

While derivative estimates are initially made only at discrete pixel locations, edge detection requires derivative estimates along the gradient line of a potential edge point, at off-pixel locations. Fig. 6c illustrates how this is accomplished when making estimates in the gradient direction.

The first estimate r_a is made at the first intersection of the gradient line with the pixel grid, in this case between pixels $(x_0, y_0 + 1)$ and $(x_0 + 1, y_0 + 1)$, using linear interpolation: $r_a = (1 - \alpha)r(x_0, y_0 + 1) + \alpha r(x_0 + 1, y_0 + 1)$. The next intersection occurs between pixels $(x_0 + 1, y_0 + 1)$ and $(x_0 + 1, y_0 + 2)$, so that the next derivative estimate r_b is given by $r_b = (1 - \beta)r(x_0 + 1, y_0 + 1) + \beta r(x_0 + 1, y_0 + 2)$. Estimates along the gradient line in the direction opposite to the gradient are made in an analogous fashion.

In the following, we identify interpolated derivate estimates as $\tilde{r}_1^{\theta_M}(x, y)$ and $\tilde{r}_2^{\theta_M}(x, y)$, to be distinguished from derivative estimates $r_1^{\theta_M}(x, y, \sigma_1)$ and $r_2^{\theta_M}(x, y, \sigma_2)$ made at pixel locations. We do not associate a scale with the interpolated estimates, as they may be derived from on-pixel estimates made at two different scales.

11 SUMMARY OF EDGE CRITERIA

The proposed algorithm for edge detection consists of three stages:

1) Use local scale control to reliably estimate the intensity gradient at each image point.
2) Use local scale control to reliably estimate the second derivative of the intensity function in the gradient direction at each image point.
3) Localize edges at zero-crossings of the second derivative in the gradient direction.

While the basic structure of the algorithm is straightforward, there are details in the discrete implementation and in the handling of both high- and low-precision cases which require some attention. We provide these details in the next two subsections.

11.1 High-Precision Edge Criteria

To be labeled as a high-precision edge, a pixel (x_p, y_p) must be located at a zero-crossing in the second derivative of the intensity function. Defining (x_n, y_n) as the location of the first intersection of the pixel grid with the gradient line through (x_p, y_p) in the gradient direction, four specific criteria must be satisfied:

1) The gradient must be reliably detected at the point: $r_1^{\theta_M}(x_p, y_p, \sigma_1) > c_1(\sigma_1)$ for some σ_1.
2) The second derivative in the gradient direction must be reliably detected as positive at the point: $r_2^{\theta_M}(x_p, y_p, \sigma_2) > c_2(\sigma_2)$ for some σ_2.
3) The interpolated gradient must be detected as nonzero at the next estimation point in the gradient direction: $\tilde{r}_1^{\theta_M}(x_n, y_n) > 0$.
4) The interpolated second derivative in the gradient direction must be detected as negative at the next estimation point in the gradient direction: $\tilde{r}_2^{\theta_M}(x_n, y_n) < 0$.

The choice of representing an edge location at the dark side of the edge (where the second derivative is positive) is arbitrary. To localize edges more precisely, a subpixel representation must be employed [27].

Fig. 7. Testing the local scale control algorithm on a synthetic image. (a) The blur grade is linear, ranging from $\sigma_b = 1$ pixel to $\sigma_b = 26.6$ pixels. The contrast of the edge and the amplitude of the added Gaussian noise are fixed. Parameters (see Section 4): $A = B = 85$ gray levels, $\sigma_b \in [1, 26.6]$ pixels, $s_n = 1.6$ gray levels. (b) Detected edge. (c) Estimated vs. actual blur scale along edge.

11.2 Low-Precision Edge Criteria

To be labeled as a low-precision edge, a pixel (x_0, y_0) must be equidistant from second derivative extrema along the gradient line. This requires the definition of a gradient line segment running though (x_0, y_0) (Fig. 6b).

DEFINITION 2. *The* gradient line segment *of a point* (x_0, y_0) *extends in the gradient direction until either a positive second derivative estimate is detected, or until a negative estimate is followed by a point where no reliable estimate of the second derivative sign can be made. Similarly, the gradient line segment extends in the direction opposite to the gradient until either a negative second derivative estimate is detected, or until a positive estimate is followed by a point where no reliable estimate of the second derivative sign can be made. We define* (x_+, y_+) *as the location of the global maximum of the second derivative in the gradient direction on the gradient line segment, and* (x_-, y_-) *as the location of the global minimum of the second derivative in the gradient direction on the gradient line segment.*

Given this definition, a pixel (x_0, y_0) may be classified as a low-precision edge if it satisfies the following criteria:

1) The gradient is reliably detected as nonzero at all grid intersections of the gradient line segment through the point.
2) There is at least one grid intersection of the gradient line segment at which no reliable estimate of the second derivative in the gradient direction could be made.
3) There exists a point on the gradient line segment in the gradient direction where the second derivative in the gradient direction is reliably detected as negative.
4) There exists a point on the gradient line segment in the direction opposite to the gradient where the second derivative in the gradient direction is reliably detected as positive.
5) The pixel (x_0, y_0) lies within $\frac{1}{\sqrt{2}}$ pixels of the point which bisects the extrema locations (x_+, y_+) and (x_-, y_-).

The tolerance of $\frac{1}{\sqrt{2}}$ accounts for the worst-case error in representing the bisection of the extrema locations to the nearest pixel.

12 EXPERIMENTAL RESULTS

12.1 Synthetic Images

12.1.1 Blur Scale Experiment

We first tested our edge-detection method on the synthetic image of Fig. 7a, a vertical edge blurred by a space-varying, one-dimensional horizontal Gaussian blur kernel, with blur scale increasing linearly along the edge. Gaussian i.i.d. noise ($s_n = 1.6$ gray levels) was added to simulate sensor noise. Fig. 7b shows the edge points detected in this image. Note that the edge is reliably detected and localized over a wide range of blur, and that no artifactual edges are detected.

12.1.2 Contrast (SNR) Experiment

To evaluate edge-detection performance of the local scale control algorithm as a function of noise level, we ran the algorithm on synthetic images of a square in Gaussian white noise (Fig. 8). SNR is 2, 1, and 0.5, from top to bottom. In all cases, SNR is below the range required for one-pixel localization (Fig. 3d): The edges detected in these synthetic images are therefore low-precision edges.

The middle column of Fig. 8 shows the scale map for estimating the second derivative in the gradient direction of the intensity function at each point (larger scales in lighter gray, white indicates that no reliable estimate could be made). As predicted, in the immediate vicinity of the square's edges the second derivative signal is too weak to be estimated reliably, so that an explicit zero-crossing in the second derivative is not available. The right column shows the effect of this dropout in the second derivative signal on edge localization. Again, no artifactual edges are detected. While in all cases the square is entirely detected, as SNR decreases, the error in edge localization increases. One can also see an increased rounding in the corners of the square as SNR decreases. This is due to the increased blurring of the image at the larger scales required for estimation at higher noise levels.
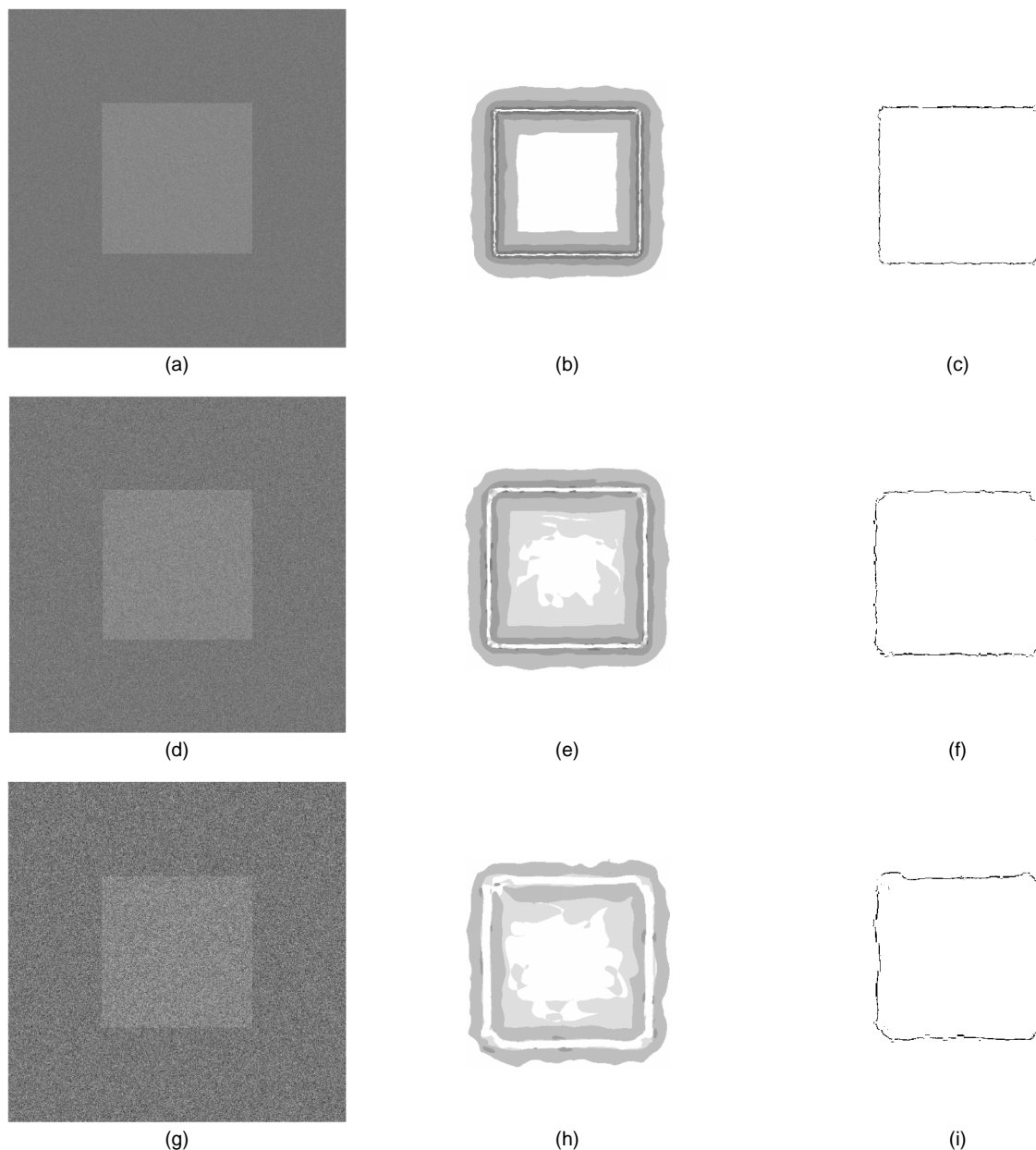
Fig. 8. SNR behavior of local scale control for edge localization. The left column shows synthetic images of a square in Gaussian noise. From top to bottom, SNR = 2, 1, 0.5. The middle column shows the variation in scale in estimating the sign of the second derivative of the intensity function in the gradient direction. Larger scales are rendered in lighter gray, white indicates that no reliable estimate could be made. At these noise levels, the sign of the second derivative near the edge cannot be estimated reliably at any scale. The right column shows the edges detected. Uncertainty in the second derivative at high noise levels leads to localization error.

12.2 Natural Images

Fig. 4c shows a map of the edge points in the image of the mannequin and shadow. Note that the contours of the image are recovered, without spurious responses to the smooth shading gradients behind the mannequin and in the foreground and background of the shadow. Observe also that both the fine detail of the mannequin and the complete contour of the shadow are resolved (compare with the results of the Canny/Deriche detector in Fig. 2). This is achieved by a single automatic system with no input parameters other than the second moment of the sensor noise.

Heath et al. have recently compared the performance of five edge-detection algorithms on a set of natural images. We will identify the five algorithms tested by the names of the authors: Bergholm [11], Canny [28], Iverson and Zucker [29], Nalwa and Binford [30], and Rothwell et al. [31]. In this study, algorithm performance was evaluated on the basis of subjective human visual judgment [32]. Since the performance of these algorithms depended upon the setting of up to three parameters, Heath et al. coarsely sampled each algorithm's parameter space to determine the parameter settings for each algorithm that maximized the mean performance over all images. Thresholding with hysteresis and nonmaximum suppression [28] were used for all but the Rothwell et al. detector.
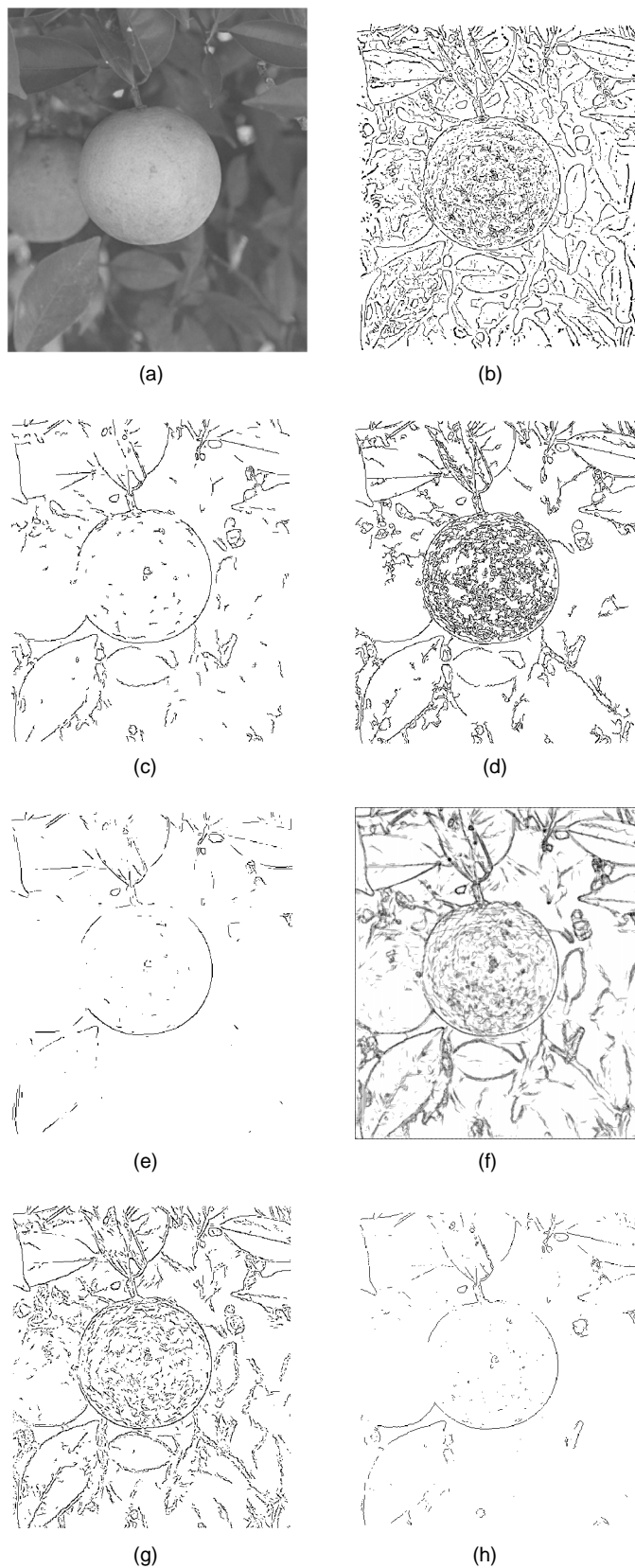
Fig. 9. Comparison of local scale control with previous edge-detection algorithms. (a) Original image. (b) Edges detected with local scale control. (c) Bergholm detector [11]. (d) Canny detector [28]. (e) Iverson detector [29] (with modifications by Heath et al. [32]). (f) Original Iverson detector [29] (g) Nalwa detector [30]. (h) Rothwell detector [31]. (a), (c), (d), (e), (g), and (h) courtesy of K. Bowyer, University of South Florida [32].

In their evaluation, Heath et al. made changes to some of the algorithms so that they would produce output in the same format for comparison purposes. In the case of the Iverson and Zucker algorithm, these changes were very significant, so we have also included the results of the original version of the Iverson and Zucker algorithm for comparison.

One of the test images used by Heath et al. is shown in Fig. 9a. The edge map computed by the local scale control algorithm is shown in Fig. 9b. The edge maps computed by the Bergholm, Canny, Iverson and Zucker (modified), Iverson and Zucker (original), Nalwa and Binford, and Rothwell et al. algorithms are shown in Figs. 9c, 9d, 9e, 9f, 9g, and 9h, respectively.

It is apparent how the local scale control algorithm detects much more of the lower-contrast, slightly blurred leaf structure than the other methods. While the original Iverson and Zucker detector and the Nalwa detector are second-best by this criterion, many of the leaf edges detected exhibit the multiple-response problem discussed in Section 1. For the Iverson and Zucker detector these contours appear as thickened or smudged; for the Nalwa detector they appear as laterally shifted echoes.

It may be argued that methods which detect only a fraction of the edges in an image may be useful if they detect only the most important edges and ignore the less important edges. However, the analysis of Section 2 suggests that making this distinction at an early stage is very difficult, since important edges may often be more blurry and of lower contrast than relatively unimportant edges. We argue, therefore, that the goal of edge detection should be to detect all of the edges in the image, over the broad range of contrasts and blur scales with which they occur. Of course, subsequent processing of the edge map is required before it can be directly used for higher-level tasks such as object recognition. This higher-level processing must include methods for discriminating texture edges from nontexture edges [33] and for grouping edges into bounding contours (e.g., [34], [35], [36], [37]).

13 LOCAL SCALE CONTROL FOR BLUR ESTIMATION

Fig. 10 shows a tangle of branches, photographed with a shallow depth of field ($f/3.5$). The connected cluster of twigs in focus at the center of the frame forms a clear subject, while the remaining branches appear as defocused background. Fig. 10b shows the edges detected for this image using local scale control for reliable estimation: both the in-focus and defocused branches are recovered. Note that although occlusions provide some clue as to the depth ordering of the contours, the immediate perceptual segmentation of foreground and background structure provided by the focal blur is lost.

Existing techniques for focal blur estimation typically assume surfaces varying slowly in depth [20], [38], [39]. This assumption fails for this image, in which any local neighborhood may contain many distinct depth discontinuities. This creates a dilemma: While employing small estimation filters will lead to larger errors in blur estimation due to sensor noise, employing large filters will increase error due

to interference between distinct structures which are nearby in the image, but far apart in depth (and, hence, in focal blur) [38]. The problem is thus to choose a compromise filter scale which is as small as possible, but large enough to avoid error due to sensor noise. Local scale control is naturally suited to this task.

For the estimation of focal blur, the Gaussian kernel of our edge model represents the point spread function of the lens system of the camera employed. From (9), the locations of the extrema in the estimated second derivative are expected to occur at $\pm\sqrt{\sigma_2^2 + \sigma_b^2}$ pixels to either side of the edge location, where σ_2 is the minimum reliable scale of the second derivative operator and σ_b is the blur scale of the edge. Defining d as the distance between second derivative extrema of opposite sign in the gradient direction (Fig. 6a), we obtain

$$\sigma_b = \sqrt{\left(d/2^2\right) - \sigma_2^2}$$

Thus, the blur due to defocus can be estimated from the measured thickness of the contours, after compensation for the blur induced by the estimation itself.

Fig. 7c shows a plot of the estimated and actual blurs of the synthetic test image of Fig. 7a. Error increases roughly linearly with blur scale, and can thus best be expressed as a percentage. While the estimation method appears to be approximately unbiased, with mean error of only 2.8 percent, the individual estimates are quite noisy, with an RMS error of 17.6 percent. While we are unaware of performance measures for competitive methods of estimating blur in a single image (as opposed to error in estimating range from two images of differing depth of field, for example), the just-noticeable difference in edge blur for the human visual system is known to be on the order of 13-20 percent [40], implying an ability to estimate the blur of a single edge at 9-14 percent.

We believe that the main source of error in our method is error in the localization of the second-derivative extrema of the edge. Note that the scale of the second derivative filter is selected to ensure that the sign of the second derivative is reliably detected: This does not guarantee that the extrema will be correctly localized, and in fact for blurred edges typically several extrema will exist. The problem is analogous to the problem in using the gradient maxima to localize the position of an edge (see Section 7). The correct solution is also analogous: to localize an edge we must reliably detect the zero-crossing of the second derivative in the gradient direction, to localize the second derivative extrema (and hence estimate blur), we must reliably detect the zero-crossing of the third derivative in the gradient direction. We are presently developing an improved method for blur estimation based upon this local scale control technique.

Despite the noise in our present method for blur estimation, it may still be employed usefully to segment image structure into distinct depth planes. For example, we can apply it to the image of Fig. 10 to replicate the perceptual segmentation between subject and background that we experience when viewing the original image. Fig. 10c and 10d show the extracted foreground (focused) and background (defocused) structures, respectively.
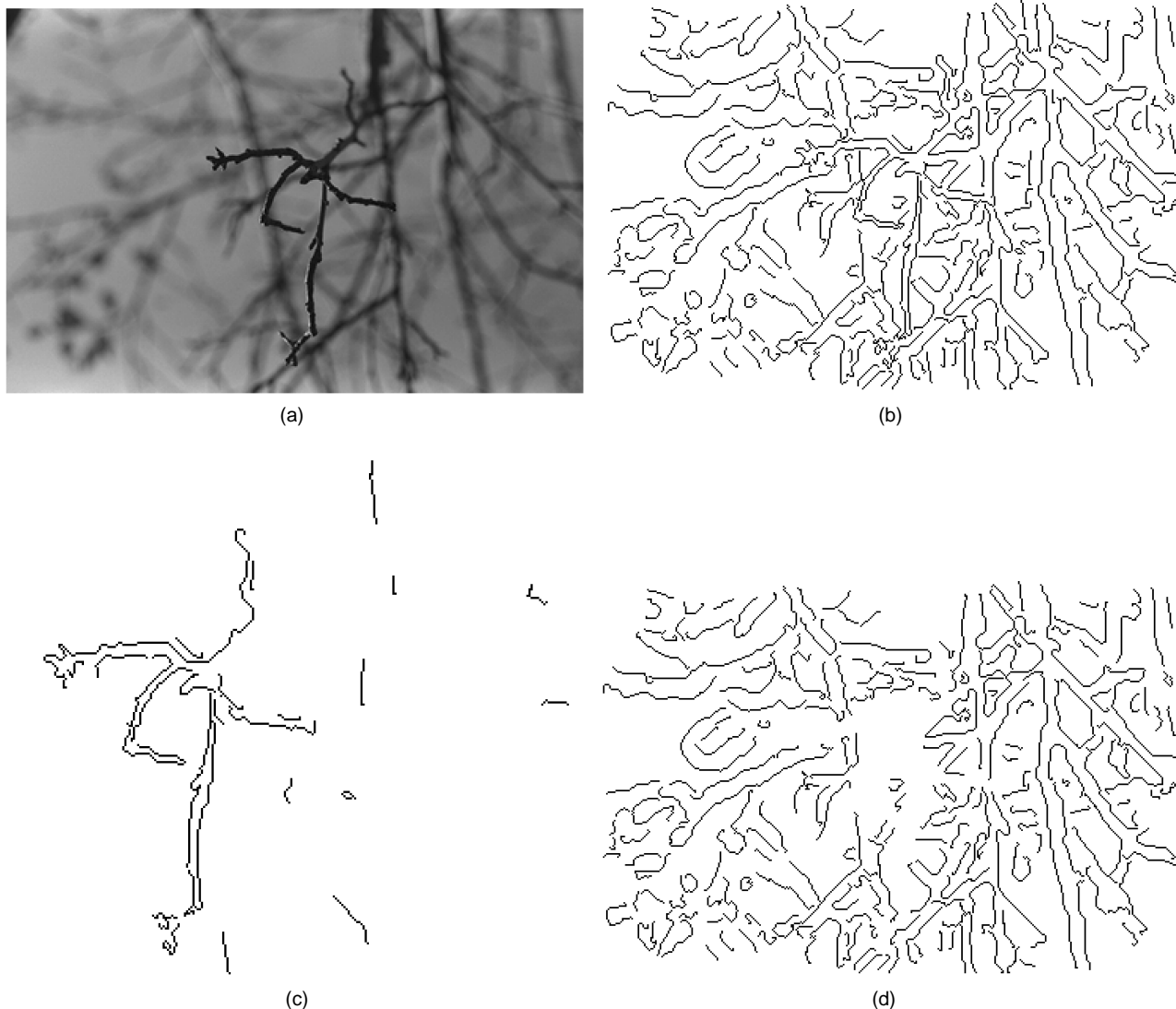
Fig. 10. Depth segmentation based on focal blur. (a) A photograph of tree branches with shallow depth of field (*f*/3.5) and near focus. (b) Edge map. (c) Foreground structure (focused contours). (d) Background structure (blurred contours).

14 SPACE CURVES FROM DEFOCUS AND CAST SHADOWS

While others have had some success in classifying contours as thin or diffuse [41], [42], [43], we show here that our method for estimating contour blur can provide dense estimates continuously along image contours to recover complete space curves from an image. As an example, consider the image of a model car (Fig. 11a) photographed with shallow depth of field (*f*/2.5). The lens was focused on the rear wheel of the car, so that the hood and front bumper are defocused. Fig. 11b shows the edges selected by the minimum reliable scale method. Note that, in spite of the severe defocus, the foreground and background structures are reliably detected and localized. Fig. 11c shows a three-dimensional plot of one of the main contours of the car. Here the vertical axis represents the focal blur σ_b, estimated as described in the previous section, and smoothed along the contour with a Gaussian blur kernel ($\sigma = 22$ pixels). The contour provides a continuous estimate of focal blur, related by a monotonic function to the distance from the plane of best focus, which in this case is at the rear wheel of the model car.

This method for blur estimation can also be used to estimate penumbral blur. Let us again consider the image of the mannequin casting a shadow (Fig. 2a). The blur of the shadow contour increases toward the head of the shadow. The results of penumbral blur estimation along the shadow contour (after Gaussian smoothing blur estimates along the contour, $\sigma = 22$ pixels) are shown in Fig. 11d.

As discussed in Section 1, the duality between defocus and cast shadows indicates that focal and penumbral blur cannot be distinguished by a local computation on a single image frame. Existing passive methods for estimating depth from defocus typically use two frames with different depths of field to distinguish focal blur from other types of blur [20], [38], [39]. This technique could also be applied to our method for blur estimation, allowing focal and penumbral blur to be decoupled and estimated separately.
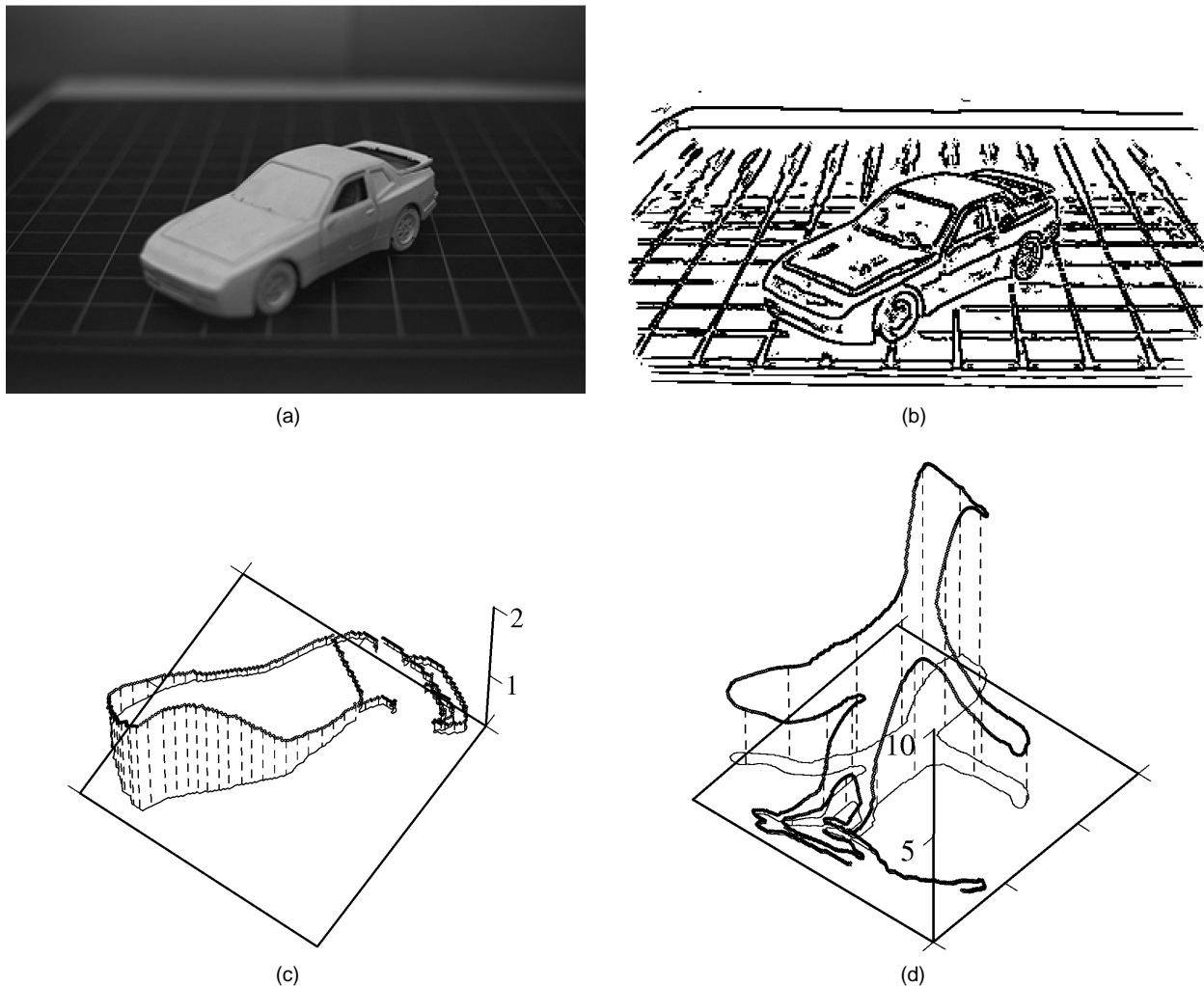
Fig. 11. Space curves from focal blur. (a) A photograph of a car model with shallow depth of field (*f*/2.5). The lens is focused on the left rear wheel. (b) Edges recovered using local scale control for reliable estimation. (c) Space curve of contour from car. (d) The Space curve of contour bounding shadow of mannequin (Fig. 2a). The vertical axis represents estimated penumbral blur scale in pixels.

15 OPEN QUESTIONS AND FUTURE WORK

There are a number of ways in which we are extending or plan to extend the present work:

1) Subpixel localization of edge points using local scale control. Preliminary results have already been reported [27].
2) Extension of scale adaptation to include orientation-tuning adaptation, allowing very low SNR and cluttered edges to be reliably detected.
3) Improvements in precision of blur estimation based on reliable estimation of the third derivative using local scale control.

One major open question is how we can best evaluate the degree to which our edge model and the proposed local scale control method accurately represents all of the edges in a natural image. To address this question, we have recently developed a means for inverting our edge representation to compute an estimate of the original image [27]. Each edge point is represented by the parameters of (1). From this representation, we show that an estimate of the original image can be computed which is perceptually nearly identical to the original image. The blur scale component of the model is found to be critical to achieving perceptually accurate reconstructions. These results suggest that the proposed model and edge-detection method accurately represent virtually all of the edges of the image.

Perhaps the most difficult open question facing this work is the validity of the assumption that only a single scale at each edge point suffices to characterize an image. Consider, for example, a shadow falling on a textured surface. At many points along the shadow, we may wish to identify two edges at two distinct orientations and scales: a small scale edge generated by an element of the texture, and a large scale edge generated by the soft shadow. In such cases, we may wish to search scale and orientation space at each edge point seeking potentially multiple edges at the point. For such an algorithm, the scales selected by local scale control would form only the lower envelope in scale space above which this search must be constrained. An alternative to this approach is to detect such a shadow edge by means of a "second-order" computation which detects sudden changes in the intensity statistics of previously detected texture edges.

16 CONCLUSIONS

Physical edges in the world generally project to a visual image as blurred luminance transitions of unknown blur scale and contrast. Detecting, localizing, and characterizing edges over this broad range of conditions requires a multiscale approach. Previous attempts at edge detection using local scale selection fail to distinguish real from artifactual edges and/or require heuristic global computations. In this paper, we have shown that, given prior knowledge of sensor noise and operator norms, real edges can be reliably distinguished from artifactual edges using a purely local computation. The proposed algorithm for local scale control detects and localizes edges over a broad range of blur scale and contrast and requires no input parameters other than the second moment of the sensor noise. This method for edge detection leads naturally to a method for estimating the local blur of image contours. This contour-based method for blur estimation was shown to be useful for complex images where the smoothness assumptions underlying Fourier methods for blur estimation do not apply.

ACKNOWLEDGMENTS

REFERENCES

[1] J. Canny, "Finding Edges and Lines in Images," Master's thesis, MIT Artificial Intelligence Laboratory, 1983.
[2] R. Deriche, "Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector," *Int'l J. Computer Vision*, vol. 1, no. 2, pp. 167-187, 1987.
[3] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, Mass.: MIT Press, 1987.
[4] Y. Leclerc and S. Zucker, "The Local Structure of Image Discontinuities in One Dimension," *IEEE Trans. Pattern Anaysis and Machine Intelligence*, vol. 9, no. 3, pp. 341-355, 1987.
[5] J. Elder, *The Visual Computation of Bounding Contours*. PhD thesis, McGill University, Dept. of Electrical Eng., 1995.
[6] D. Hubel and T. Wiesel, "Receptive Fields and Functional Architecture in the Cat's Visual Cortex," *J. Neuroscience*, vol. 160, pp. 106-154, 1962.
[7] D. Hubel and T. Wiesel, "Receptive Fields and Functional Architecture of Monkey Striate Cortex," *J. Neuroscience*, vol. 195, pp. 215-243, 1968.
[8] D. Marr and E. Hildreth, "Theory of Edge Detection," *Proc. Royal Soc. of London B*, vol. 207, pp. 187-217, 1980.
[9] A. Witkin, "Scale Space Filtering," *Proc. Int'l Joint Conf. Artificial Intelligence*, pp. 1,019-1,021, Karlsruhe, 1983.
[10] J. Koenderink, "The Structure of Images," *Biol. Cyberetics*, vol. 50, pp. 363-370, 1984.
[11] F. Bergholm, "Edge Focusing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, no. 6, pp. 726-741, 1987.
[12] T. Lindeberg, "Scale-Space for Discrete Signals," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 3, pp. 234-254, 1990.
[13] Y. Lu and R. Jain, "Reasoning About Edges in Scale Space," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 4, pp. 450–468, 1992.
[14] P. Perona and J. Malik, "Scale-Space and Edge Detection Using Anisotropic Diffusion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, pp. 629-639, 1990.
[15] T. Lindeberg, "Detecting Salient Blob-Like Image Structures and Their Scales With a Scale-Space Primal Sketch: A Method for Focus-of-Attention," *Int'l J. Computer Vision*, vol. 11, no. 3, pp. 283-318, 1993.
[16] H. Jeong and C. Kim, "Adaptive Determination of Filter Scales for Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 5, pp. 579-585, 1992.
[17] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Dordrecht, The Netherlands: Kluwer, 1994.
[18] T. Lindeberg, "Edge Detection and Ridge Detection With Automatic Scale Selection," *IEEE Conf. Computer Vision Pattern Recognition*, pp. 465-470, San Francisco, IEEE CS Press, June 1996.
[19] G. Healey and R. Kondepudy, "Modeling and Calibrating CCD Cameras for Illumination Insensitive Machine Vision," *Physics-Based Vision: Radiometry*, L. Wolff, S. Shafer, and G. Healey, eds. Boston, Mass.: Jones and Bartlett, 1992.
[20] A. Pentland, "A New Sense for Depth of Field," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 523-531, 1987.
[21] M. Subbarao, "Parallel Depth Recovery by Changing Camera Parameters," *Proc. Second Int'l Conf. Computer Vision*, pp. 149-155, Tampa, Fla., IEEE CS Press, 1988.
[22] A. Pentland, S. Scherock, T. Darrell, and B. Girod, "Simple Range Cameras Based on Focal Error," *J. Optical Soc. Am. A*, vol. 11, no. 11, pp. 2,925-2,934, 1994.
[23] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. New York, NY: McGraw-Hill, 1965.
[24] W. Freeman and E. Adelson, "The Design and Use of Steerable Filters," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891-906, 1991.
[25] P. Perona, "Deformable Kernels for Early Vision," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 488-499, 1995.
[26] E. Wong, *Introduction to Random Processes*. New York: Springer-Verlag, 1983.
[27] J. Elder and S. Zucker, "Scale Space Localization, Blur and Contour-Based Image Coding," *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pp. 27-34, San Francisco, IEEE CS Press, 1996.
[28] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 11, pp. 679-698, Nov. 1986.
[29] L. Iverson and S. Zucker, "Logical/Linear Operators for Image Curves," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 10, pp. 982-996, Oct. 1995.
[30] V. Nalwa and T. Binford, "On Detecting Edges," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 699-714, 1986.
[31] C. Rothwell, J. Mundy, W. Hoffman, and V. Nguyen, "Driving Vision by Topology," *Int'l Symp. Computer Vision*, pp. 395-400, Coral Gables, Fla., Nov. 1995.
[32] M. Heath, "A Robust Visual Method for Assessing the Relative Performance of Edge Detection Algorithms," Master's thesis, Univ. of South Florida, Dec. 1996.
[33] M. Heath, S. Sarkar, T. Sanocki, and K. Bowyer, "Comparison of Edge Detectors: A Methodology and Initial Study," *Proc. IEEE Conf. Computer Vision Pattern Recognition*, pp. 143-148, San Francisco, IEEE CS Press, June 1996.
[34] B. Dubuc and S. Zucker, "Indexing Visual Representations Through the Complexity Map," *Proc. Fifth Int'l Conf. Computer Vision*, pp. 142-149, Cambridge, Mass., June 1995.
[35] I. Cox, J. Rehg, and S. Hingorani, "A Bayesian Multiple-Hypothesis Approach to Edge Grouping and Contour Segmentation," *Int'l J. Comp. Vision*, vol. 11, no. 1, pp. 5-24, 1993.
[36] J. Elder and S. Zucker, "Computing Contour Closure," *Lecture Notes in Computer Science, Proc. Fourth European Conf. on Computer Vision*, pp. 399-412, New York, NY, Springer Verlag, 1996.
[37] D. Jacobs, "Finding Salient Convex Groups," *Partitioning Data Sets DIMACS (Series in Discrete Mathematics and Theoretical Computer Science)*, I. Cox, P. Hansen, and B. Julesz, eds., vol. 19, 1995.
[38] D. Lowe, "Three-Dimensional Object Recognition From Single Two-Dimensional Images," *Artificial Intelligence*, vol. 31, pp. 355-395, 1987.
[39] J. Ens and P. Lawrence, "Investigation of Methods for Determining Depth From Focus," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 15, no. 2, pp. 97-108, 1993.

[40] S. Nayar and N. Yasuo, "Shape From Focus," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824-831, 1994.

[41] R. Watt and M. Morgan, "The Recognition and Representation of Edge Blur: Evidence for Spatial Primitives in Human Vision," *Vision Research*, vol. 23, no. 2, pp. 1,465-1,477, 1983.

[42] D. Marr, "Early Processing of Visual Information," *Phil. Trans. Royal Soc. London*, vol. 275, pp. 97-137, 1976.

[43] A. Korn, "Toward a Symbolic Representation of Intensity Changes in Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, no. 5, pp. 610-625, 1988.
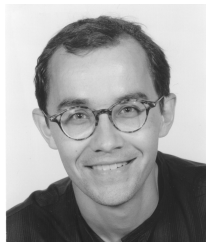
James H. Elder received the BASc degree in electrical engineering from the University of British Columbia in 1987 and the PhD degree (with honours) in electrical engineering from McGill University in 1995. From 1987 to 1989, he was with Bell-Northern Research in Ottawa, Canada. From 1995 to 1996, he was a senior research associate in the Computer Vision Group at the NEC Research Institute in Princeton, NJ. He joined the faculty of York University in 1996, where he is presently Assistant Professor of Psychology and Project Leader at the Human Performance Laboratory of the Centre for Research in Earth and Space Technology. His current research addresses problems in computer vision, human visual psychophysics, and virtual reality.
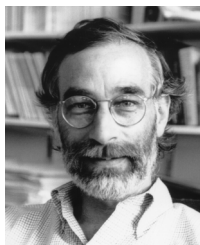
Steven W. Zucker is a professor of computer science and electrical engineering at Yale University. Before moving to Yale in 1996, he was professor of electrical engineering at McGill University, director of the program in artificial intelligence and robotics of the Canadian Institute for Advanced Research, and the codirector of the Computer Vision and Robotics Laboratory in the McGill Research Center for Intelligent Machines. He was elected a fellow of the Canadian Institute for Advanced Research (1983), a fellow of the IEEE (1988), and (by)fellow of Churchill College, Cambridge (1993).

Dr. Zucker obtained his education at Carnegie-Mellon University in Pittsburgh and at Drexel University in Philadelphia, and was a postdoctoral Research Fellow in Computer Science at the University of Maryland, College Park. He was Professor Inviteé at Institut National de Recherche en Informatique et en Automatique, Sophia-Antipolis, France, in 1989, a visiting professor of computer science at Tel Aviv University in January, 1993, and an SERC fellow of the Isaac Newton Institute for Mathematical Sciences, University of Cambridge.

Prof. Zucker has authored or coauthored more than 130 papers on computational vision, biological perception, artificial intelligence, and robotics, and serves on the editorial boards of eight journals.