International Journal of Computer Vision 34(2/3), 97–122 (1999)

Are Edges Incomplete?

JAMES H. ELDER
Centre for Vision Research, York University, 4700 Keele Street, North York, Ontario, Canada M3J 1P3
jelder@yorku.ca

Abstract. We address the problem of computing a general-purpose early visual representation that satisfies two criteria. 1) Explicitness: To be more useful than the original pixel array, the representation must take a significant step toward making important image structure explicit. 2) Completeness: To support a diverse set of high-level tasks, the representation must not discard information of potential perceptual relevance. The most prevalent representation in image processing and computer vision that satisfies the completeness criterion is the wavelet code. In this paper, we propose a very different code which represents the location of each edge and the magnitude and blur scale of the underlying intensity change. By making edge structure explicit, we argue that this representation better satisfies the first criterion than do wavelet codes. To address the second criterion, we study the question of how much visual information is lost in the representation. We report a novel method for inverting the edge code to reconstruct a perceptually accurate estimate of the original image, and thus demonstrate that the proposed representation embodies virtually all of the perceptually relevant information contained in a natural image. This result bears on recent claims that edge representations do not contain all of the information needed for higher level tasks.

Keywords: edge detection, image reconstruction, scale space, diffusion, blur, deblurring, denoising, diffusion

1. Introduction

Visual perception may be seen as a loose collection of algorithms to solve specific, well-defined problems, e.g. estimating the pose of a known object. From this perspective, algorithms and representations may be designed to rely heavily on the specific constraints and knowledge available for the restricted task at hand. This approach was adopted by some of the earliest machine vision systems (e.g. Roberts, 1965), and remains popular, and often effective, today.

An alternative to this approach is to seek more general algorithms and representations which impose coherence over a diversity of visual tasks (e.g. Barrow and Tenenbaum, 1981; Marr, 1982; Zucker, 1986). This approach is motivated in part by what we know of biological vision systems. While there is ample evidence for modularity in the primate visual cortex, it is also clear that these modules depend upon common computations and representations in the retina, lateral geniculate nucleus and striate cortex at least. It is this type of early, general-purpose visual computation which concerns us here.

It is widely assumed that a general-purpose visual system must entail multiple types of representation which share data through a variety of computations. This paper focuses on the specific problem of computing an initial, early, general-purpose representation directly from the image; a representation which will replace the original image in providing all of the data required for a diversity of higher-level computations.

Lacking a narrow high-level goal, this early visual representation must be designed according to general criteria. We propose the following evaluative measures:

- *Generality*: The diversity of structures found in natural images must be detectable and representable in the proposed encoding.
- *Reliability*: Features of the image which correspond to scene features must be reliably represented (low false negative rate). At the same time, artifacts of

the sensing process should not be represented (low false positive rate).
- *Precision*: Features in the image should be localized precisely.
- *Concision*: The encoding should reduce pointless redundancy present in the original image (e.g. Barlow (1961)).
- *Explicitness*: The encoding should represent important image structure more explicitly than does the original pixel array.
- *Completeness*: Given the need to support a diversity of higher level visual tasks, the encoding should contain all visual information of potential perceptual relevance.

Many early visual representations have been proposed, but most do not attempt to satisfy the completeness criterion, which is critical if the representation is to support all higher-level computation. An exception to this is the large class of wavelet-based image codes (e.g. Daubechies, 1991; Mallat, 1989; Mallat and Zhong, 1992; Simoncelli et al., 1992), which are typically mathematically complete or over-complete.

In itself, a wavelet encoding is not a concise description of an image: at best, an orthogonal code requires the same storage as the original image. However, it has been shown that wavelet encoding leads to a decreased correlation in the second-order statistics of natural images, and hence a reduction in entropy in the first-order statistics. This reduction in entropy can in turn be exploited to compress the image (Vetterli, 1984; Adelson et al., 1987).

So what is wrong with the wavelet code as a general-purpose early vision representation? In our view, the wavelet code falls short of meeting the explicitness criterion. Fundamentally, a wavelet code is simply a linear transformation of the original image into a set of new images. The pixels of these new images represent the strength of response to a small set of linear filters. Beyond that, no interpretation or inference has been made, and the structures and features of the image are not explicitly represented, but remain implicitly coded as collections of pixels. To use a distinction due to Adelson (1991), the wavelet code represents "stuff" not "things". This property makes the criteria of reliability, generality and precision meaningless, since these criteria assume that explicit features are being represented.

The search for a more symbolic yet complete early visual code can be traced to Marr and Hildreth (1980), who conjectured that an image may be completely represented by zero-crossing data over scale space. There have been a number of subsequent attempts to design complete or nearly-complete early visual codes based upon zero-crossings or edges (e.g. Carlsson, 1984, 1988; Hummel and Moniot, 1989; Mallat and Zhong, 1992; Grattoni and Guioucci, 1990; Cumani et al., 1991), however none of these codes are able to jointly satisfy the key criteria for a general-purpose early visual representation. In particular, high fidelity (near completeness) has been achieved, but only at the expense of reliability and concision and/or explicitness (these claims are justified in Section 4).

In this paper we introduce a representation which comes much closer to meeting all of these criteria. The proposed representation is an edge code: information is represented only at pixels where an edge has been identified and localized. At each of these edge points, four parameters of a local edge model are represented: two parameters describing the asymptotic intensity on either side of the edge (alternatively coded as brightness and contrast), one parameter representing the blur scale of the edge, and a fourth parameter representing the gradient (alternatively, the tangent) direction of the edge. For the purposes of reconstruction, the gradient direction parameter need only be represented to 90° accuracy, i.e. 2-bit resolution. In contrast to prior approaches (e.g. Yuille and Poggio, 1985; Hummel and Moniot, 1989; Mallat and Zhong, 1992), each edge is *not* represented at multiple scales. Rather, estimation scale is adapted over the image, so that each edge is estimated at a single, optimum scale, and is represented only once. Thus the proposed representation is far more concise than prior multiscale approaches.

Unlike wavelet codes, the edge code makes important features explicit. At any particular image location, an edge either exists or does not exist: edges are "things", not "stuff". Edges signal and localize important changes in scene properties (object boundaries, surface creases, pigment changes, shadows, etc...). Moreover, the parameters of the edge model make explicit information which will be critical in computing scene structure in higher-level computations. For example, while not sufficient for positive identification, edge blur is an important cue to shadows and shading (Elder, 1999). Quantitative blur estimation can also be useful for depth from defocus (e.g. Pentland, 1987; Nayar and Yasuo, 1994) and depth from shadows (Kersten et al., 1997; Elder and Zucker, 1998; Elder et al., 1998). Edge tangent information is crucial for

the refinement and grouping of global image contours (e.g. Zucker et al., 1977; Sha'ashua and Ullman, 1988; David and Zucker, 1990; Cox et al., 1993; Elder and Zucker, 1996a).

The specific method we propose to detect edges and estimate model parameters has been reported in detail elsewhere (Elder and Zucker, 1996b, 1998). In this paper we will only briefly summarize the method, and focus on those points which demonstrate the generality and reliability of the algorithm. We have also demonstrated methods to localize edges to a precision of less than 1/20 pixel (Elder and Zucker, 1996c); we will not discuss this localization algorithm further here. We are left with the criteria of completeness and concision, which will be our focus for the remainder of this paper.

It is sometimes argued (e.g. Belhumeur and Kriegman, 1996) that edges fail to capture all of the information in an image important for higher-level computations. To address this claim, we have developed a novel method for inverting our edge representation to reconstruct a perceptually accurate estimate of the original image. Through a series of experiments on natural images, we demonstrate that the proposed edge code is perceptually nearly complete, that is, it contains nearly all of the visually perceptible information in a natural image. The encoding and reconstitution of the edge blur information is seen to be critical to the high fidelity of these reconstructions. We believe this to be the first purely edge-based representation demonstrated to yield high-quality reconstructions as well as high potential compactness.

These results are important for a number of reasons. First, they will hopefully help to correct the mistaken view that edges do not carry enough information to support higher level tasks. Second, reconstruction allows us to evaluate, both perceptually and objectively, any errors or inaccuracies that may be present in the representation. Finally, there are a number of potential applications for such an invertible representation, some of which we have begun to explore (Elder and Goldberg, 1998).

2. Organization of Paper

In the next section, we briefly review the local scale control algorithm used to compute the proposed edge code. More details of this work can be found in Elder and Zucker (1996b, 1998). In Section 4 we address the problem of information loss in the proposed edge code. We begin by reviewing previous theoretical and empirical results on the reconstructability of images from zero-crossings, and then develop a novel algorithm for image reconstruction from the proposed edge code. We demonstrate the perceptual content of the code through a series of experiments on a diverse set of natural images. In Section 5 we analyse the sources of error in our reconstructions and how these may be reduced. In Section 6 we discuss possible applications of the proposed representation.

3. Computing the Edge Representation

While edge detectors are typically designed to recover step discontinuities in an image (e.g. Canny, 1983; Leclerc and Zucker, 1987), the boundaries of physical structures in the world generally do not project to the image as step discontinuities, but as blurred transitions corrupted by noise (Fig. 1). Variations in reflectance and lighting generate a broad range of contrasts, while defocus, shadows and shading generate a broad range of local blur.

Consider the image shown in Fig. 2 (left). The middle figure shows the edge map generated by the Canny edge detector (Canny, 1983), where the scale parameter has been tuned to detect the details of the mannequin. At this relatively small scale, the contour of the shadow cannot be reliably resolved. On the right is shown the edge map generated when scale is tuned to detect the contour of the shadow. At this larger scale, the details of the mannequin are blurred out, and the contour of the shadow is fragmented at the section of high curvature under one arm.

This example demonstrates that detecting and modeling the edges in a natural image requires a multi-scale approach. Multi-scale theories generally require

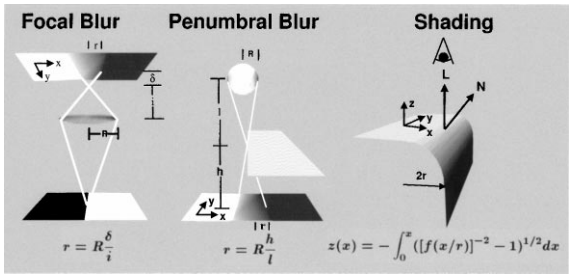

Figure 1. Edges in the world typically project to the image as spatially blurred. From left to right: focal blur due to finite depth-of-field; penumbral blur at the edge of a shadow; shading blur at a smoothed object edge.

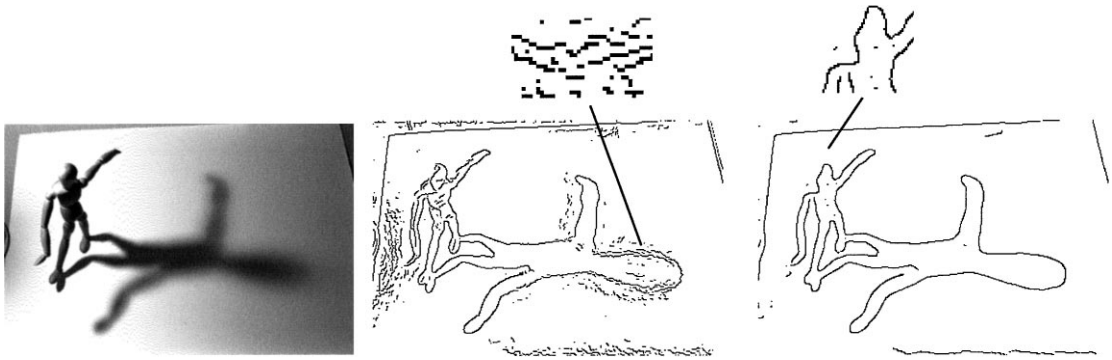

Figure 2. The problem of local estimation scale. Different structures in a natural image require different spatial scales for local estimation. The original image contains edges over a broad range of contrasts and blur scales. In the middle are shown the edges detected with a Canny/Deriche operator tuned to detect structure in the mannequin. On the right is shown 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 (Deriche, 1987) for details of the Deriche detector.

the integration of filter responses over many scales (Marr and Hildreth, 1980; Canny, 1983) or feature-tracking through a continuous 3-D scale space, in which image contours are coded as two-dimensional surfaces (Witkin, 1983; Koenderink, 1984; Grattoni and Guiducci, 1990; Lindeberg, 1990). Here we describe a theory for local scale control that determines a unique scale for local estimation at each point in an image, thus avoiding the problem of combining filter responses or tracking features through scale space. Alternative methods for scale selection based on normalized derivatives have been proposed (Lindeberg, 1996); however we argue that the local scale control method offers advantages in its simplicity, robustness to sensor noise and minimal distortion in localizing edges (Elder and Zucker, 1996b, 1998).

Edge detection typically involves the estimation of first and perhaps second derivatives of the luminance function, followed by selection of zero crossings or extrema (e.g. Marr and Hildreth, 1980; Canny, 1983). If these derivative estimates are unreliable, the selected edges will be incorrect. Local scale control is based on the observation that sensor noise properties and operator norms can be used to determine a unique minimum scale at which derivative estimates can be used to make reliable local logical inferences. We call this unique scale the *minimum reliable scale* for the estimation.

Ensuring reliability in terms of sensor noise 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. By selecting the minimum of the set of reliable scales, we ensure reliability relative to sensor noise, while minimizing distortion from nearby image events and deviations from our local edge model. By reliable here we mean that the likelihood of committing at least one Type I (false positive) error over an entire image is less than a standard tolerance of 5%.

The key to the local scale control algorithm is the prior computation of a critical value function $c(\sigma)$ which determines the lower bound on filter response as a function of filter scale σ, below which the response cannot be used reliably. The critical value function depends on the L_2 norm of the filter, the statistics of the sensor noise, and the property of the response that we wish to exploit.

Our method for edge detection depends upon the reliable detection of a non-zero gradient $r_1^{\theta_M}(x, y, \sigma_1)$ in the luminance function, and the reliable inference of the *sign* of the second derivative $r_2^{\theta_M}(x, y, \sigma_2)$ of the luminance function in the gradient direction θ_M. These derivative estimates are computed using steerable Gaussian derivative filters (Freeman and Adelson, 1991; Perona, 1995), based on unit-integral Gaussian smoothing kernels. The gradient estimates are computed using the two basis functions

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

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

where σ_1 denotes the scale of the first derivative smoothing kernel. The second derivative estimates are

computed using the three basis functions

$$g_2^x(x, y, \sigma_2) = \frac{1}{2\pi\sigma_2^4}((x/\sigma_2)^2 - 1)e^{-(x^2+y^2)/2\sigma_2^2}$$
$$g_2^y(x, y, \sigma_2) = \frac{1}{2\pi\sigma_2^4}((y/\sigma_2)^2 - 1)e^{-(x^2+y^2)/2\sigma_2^2}$$
$$g_2^{xy}(x, y, \sigma_2) = \frac{xy}{2\pi\sigma_2^6}e^{-(x^2+y^2)/2\sigma_2^2}$$

where σ_2 denotes the scale of the second derivative smoothing kernel. Details of the methods used for derivative estimation may be found in Elder and Zucker (1998).

Each of these inferences determines a specific critical value function. The critical value functions for asserting a non-zero gradient, and for determining the sign of the second derivative in the gradient direction are (Elder and Zucker, 1998)

$$c_1(\sigma_1) = 1.1s_n/\sigma_1^2 \qquad c_2(\sigma_2) = 1.3s_n/\sigma_2^3$$

where s_n is the standard deviation of the sensor noise (assumed to be Gaussian i.i.d.), and σ_1 and σ_2 are the scales of the Gaussian first and second derivative filters, respectively.[1]

3.1. *Edge Detection*

To apply local scale control to the detection of blurred edges, we model an edge as a step function $(I_l - I_d)u(x) + I_d$ from a dark intensity I_d to a bright intensity I_l. The edge is aligned, for convenience, with the y-axis of our coordinate frame. The blur of this edge is modelled by a Gaussian blur kernel, $g(x, y, \sigma_b) = \frac{1}{2\pi\sigma_b^2}e^{-(x^2+y^2)/2\sigma_b^2}$ of unknown scale constant σ_b. Sensor noise $n(x, y)$ is modeled as a stationary, additive, zero-mean white noise process. The complete edge model is thus

$$\frac{I_l - I_d}{2}(erf(x/\sqrt{2}\sigma_b) + 1) + I_d + n(x, y). \quad (1)$$

Figure 3 illustrates the model with no sensor noise added.

The prior construction of critical value functions allows the minimum reliable scale for each estimation to be determined at each point as the image is processed. The situation for gradient estimation at an edge is shown in Fig. 4 (top left). Although both the gradient response $r_1^{\theta_M}$ and the critical value function $c(\sigma_1)$ decline with increasing scale, the critical value function declines much more quickly. The scale at which the two curves intersect is the minimum reliable scale for gradient estimation. In our experiments we attempt only to stay close to the minimum reliable scale by computing estimates at octave intervals of scale, at each point using the smallest scale at which the estimate exceeds the critical value function, i.e., for gradient estimation, $\hat{\sigma}_1(x, y) = \inf\{\sigma_1 : r_1^{\theta_M}(x, y, \sigma_1) > c_1(\sigma_1)\}$.

The importance of the second derivative in localizing blurred edges is also illustrated in Fig. 4. Figure 4 (top right) shows the luminance profile through the edge of the mannequin's shadow. Figure 4 (middle right) shows the gradient magnitude along the cross-section, and Fig. 4 (middle left) 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 non-zero over this cross-section, the response is not unimodal: there are in fact 5 maxima in the gradient along the cross section of the edge. Marking edges at extrema in the gradient function would clearly lead to multiple separate responses to this single edge.

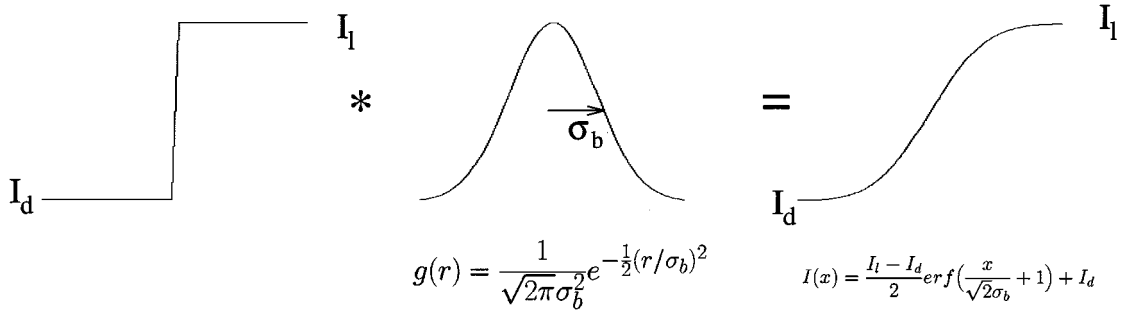

Figure 3. The edge model used for edge detection and representation.

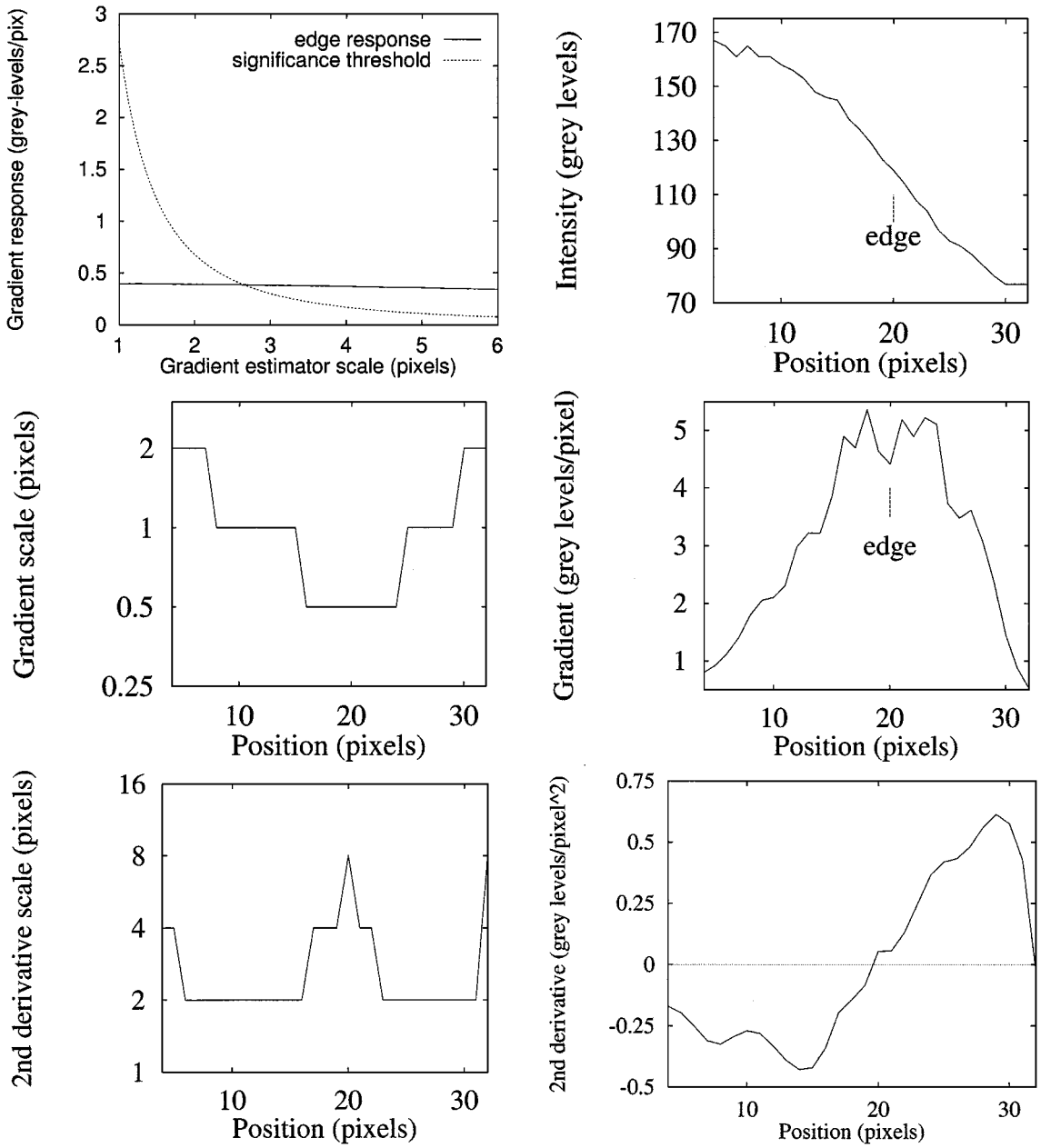

Figure 4. *Top left*: The intersection of the critical value function $c_1(\sigma_1)$ with the gradient response $r_1^{\theta_M}(\sigma_1)$ determines the minimum reliable scale for gradient estimation. *Top right*: Cross section of image across shadow boundary. *Middle left*: Minimum reliable scale for the gradient estimate. *Middle right*: Estimated gradient magnitude. *Bottom left*: Minimum reliable scale for the second derivative estimate *Bottom right*: Estimated directional second derivative. A unique zero-crossing localizes the edge.

Figure 4 (bottom right) shows the estimated second derivative steered in the gradient direction, and Fig. 4 (bottom left) 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 centre 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 local estimation at

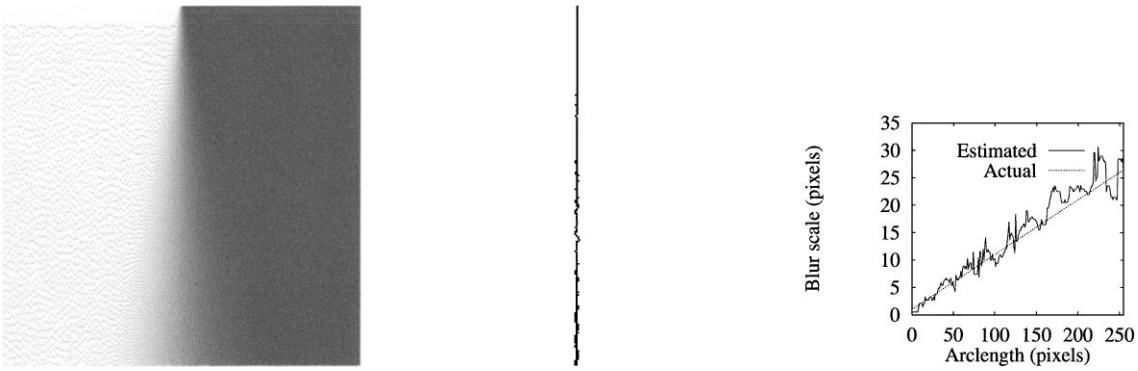

Figure 5. *Left*: The blur grade is linear, ranging from $\sigma_b = 1$ pixel to $\sigma_b = 26.6$ pixels. Parameters (see Eq. 1): $I_l = 170$ grey levels, $I_d = 85$ grey levels, $\sigma_b \in [1, 26.6]$ pixels, $\sigma_n = 1.6$ grey levels. *Middle*: Edges detected by local scale control. *Right*: Estimated vs actual blur scale σ_b.

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 estimation.

The ability of the local scale control algorithm to detect edges over a broad range of blurs is shown in Fig. 5. The test image is a vertical edge blurred by a space-varying Gaussian kernel, corrupted by Gaussian i.i.d. noise. Five scales were used for gradient estimation ($\sigma_1 \in \{0.5, 1, 2, 4, 8\}$ pixels), six for second derivative estimation ($\sigma_2 \in \{0.5, 1, 2, 4, 8, 16\}$ pixels). Figure 5 (middle) shows the edges detected by the local scale control algorithm. The edges are reliably and uniquely detected over a wide range of blur. Figure 6 shows the edges detected in the image of the mannequin and shadow. Both the fine detail of the mannequin and the complete contour of the shadow are resolved, without spurious responses to the smooth shading gradients on the ground surface (compare with the results of the Canny detector in Fig. 2). We emphasize that this is achieved by a single system with no input parameters other than the second moment of the sensor noise. The key to the performance of the system is the automatic identification of a unique minimum reliable scale for estimation at each point in the image.

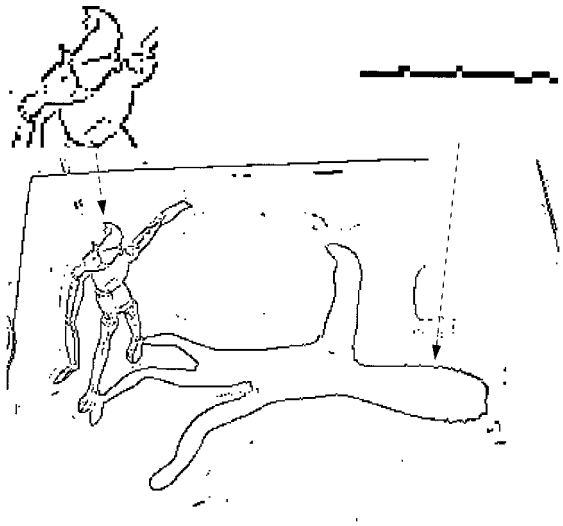

Figure 6. Results of local scale control for image of mannequin with shadow. Both the fine detail of the mannequin and the blurred, low-contrast shadow are reliably recovered.

3.2. Estimation of Blur and Intensity

Local scale control may also be applied to the estimation of focal blur. While excellent passive techniques for blur estimation have been developed, these typically require densely-textured surfaces varying slowly in depth (e.g. Pentland, 1987; Ens and Lawrence, 1993, Nayar and Yasuo, 1994) and are therefore not suited for complex images, where each local neighbourhood may contain many depth discontinuities. In such cases, blur estimates can only be made where structure exists,

and must be computed as locally as possible. Contour-based blur estimation at the minimum reliable scale is naturally suited to the task.

In our local scale control method, blur scale σ_b is estimated by computing the distance d between extrema of opposite sign in the second derivative response along the gradient direction, and then correcting for the blur induced by the estimation itself, yielding the equation (Elder and Zucker, 1996b, 1998)

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

Figure 5 (right) shows a plot of the estimated and actual blurs of the synthetic test image. While the resulting pointwise blur estimates are noisy, they provide an approximately unbiased estimate of the blur scale of the edge.

Estimation of the parameters I_l and I_d of our edge model Eq. (1) is straightforward. We first estimate the intensity values $I(\pm\sigma_b)$ at a distance $\pm\sigma_b$ from the edge location along the gradient line, using bilinear interpolation. From the model Eq. (1), we can write

$$I(\pm\sigma_b) \approx \frac{I_l - I_d}{2}(\pm erf(1/\sqrt{2}) + 1) + I_d$$

Solving for I_l and I_d, we have

$$I_l \approx \frac{I(\sigma_b) + I(-\sigma_b)}{2} + \frac{I(\sigma_b) - I(-\sigma_b)}{2erf(1/\sqrt{2})}$$

and

$$I_d \approx \frac{I(\sigma_b) + I(-\sigma_b)}{2} - \frac{I(\sigma_b) - I(-\sigma_b)}{2erf(1/\sqrt{2})}$$

3.3. Generality and Reliability

In terms of an edge code, the generality criterion for an early visual representation demands that the diversity of edge types that occur in natural images be identified and localized. In our approach, this is achieved through the generalization of the edge model to encompass edges over a broad range of blur scales and contrasts. The examples shown in Figs. 5 and 6 demonstrate that this broad range of naturally-occurring edges are indeed detected.

The reliability criterion demands that real edges corresponding to structure in the scene be consistently detected, while false edges caused by sensor noise are ignored. This is achieved by the local scale control algorithm, which automatically adapts estimation scale to determine the minimum scale for reliable estimation. Estimation at the minimum reliable scale reduces the false-positive rate to a very low, pre-determined level, while at the same time minimizing distortion and missed (false-negative) edges.

There are other reliability criteria against which an early visual representation could be evaluated. One example is stability with respect to illumination conditions. It has been pointed out (Adini et al., 1997; Belhumeur and Kriegman, 1996) that when lighting changes, the locations of shading and shadow edges change, so that two edge maps taken under different illumination conditions may look very different. These observations have motivated "appearance-based" methods for object recognition.

When considering this criticism, it is important to maintain the distinction between special-purpose and general-purpose early representations. Suppose we restrict our computational goal to the recognition of objects with fixed pose that have been previously viewed under multiple illumination conditions. Then the restricted nature of the problem introduces significant constraints that we may take advantage of in designing representations, and an edge code may not be the first choice. However, these constraints are not available for a general-purpose system, which may be viewing a novel object, or a novel view, or which may not be supporting just recognition, but a computation of scene layout, requiring interpretation of shadows and shading.

One promising property of our proposed representation for such higher-level tasks is the explicit encoding of edge blur scale. We have previously noted that edge blur is an important cue to shadows and shading. More than this, the degree of blur is a useful predictor of stability with respect to variation in the illuminant. For example, the blur scale of a cast shadow edge measures the width of the shadow penumbra. If the dominant light source moves, the cast shadow moves also, and the distance it moves is roughly proportional to the blur scale of the shadow edge. Similarly, the blur scale of an attached shadow edge varies inversely with the local curvature of the surface. Consequently, the shift in location of attached shadow edges as a function of the illuminant direction also increases as a function of the blur of the attached shadow edge. Thus shadow and shading edge blur may be used by higher-level algorithms to estimate and account for instability with respect to illumination conditions.

4. Are Edges Incomplete?

The idea that the primary purpose of early visual coding is to detect and represent contour dates to the experiments of Mach and the proposition of lateral inhibition. However, this idea does not explain our perception of brightness, colour and shading between contours. 60 years ago, Werner provided some possible clues, in what is generally regarded as the first "metacontrast" experiment (Werner, 1935). Werner showed that by interrupting the formation of a complete bounding contour, one could prevent observers from perceiving the colour of the object which this contour bounds. Many subsequent metacontrast experiments have elaborated this finding (e.g. Paradiso and Nakayama, 1991), leading to the hypothesis of a contour-based "filling-in" process responsible for our perception of surface brightness and colour.

In a parallel development, computational vision researchers have been studying the degree to which an image can be reconstructed from its edges or zero-crossings. This research has produced two classes of techniques:

1. Scale space algorithms which produce high-quality reconstructions but which require large, highly redundant representations of edges over scale space. These representations thus require an expansion, rather than a compression, of the original image, and thus fail to satisfy our concision criterion for an early visual representation (Section 1).
2. Hybrid algorithms which code only a subset of edges, and use a sub-sampled, lowpass image to carry the low frequency image information (e.g. shading). These algorithms, while more compact, produce reconstructions with substantial artifact, and thus do not satisfy our completeness criterion. Also, the implicit representation of low-pass structure through sub-sampled images fails to satisfy our explicitness criterion (Section 1).

Here, we describe an algorithm for reconstructing an approximation of the original image from the edge representation computed by the local scale control algorithm. We will show empirically that nearly flawless reconstructions are possible from this purely edge-based representation. Since the local scale control algorithm selects a unique scale for estimation at each edge point, the representation is potentially very compact.

We begin our analysis of the reconstruction problem by reviewing prior theoretical and experimental results on the reconstructability of images from edges[2] (see Hummel and Moniot, 1989 for a more detailed review).

4.1. Previous Theoretical Results

Logan (1977) proved that the zero-crossings of a one-dimensional, strictly bandpass one-octave signal form a complete representation of the signal. Based on this result, Marr and Hildreth (1980) conjectured that an image may be completely represented by the zero-crossings and gradient magnitudes of the image convolved with a Laplacian of Gaussian operator at multiple scales.

Yuille and Poggio (1985) proved the theoretical completeness of the zero-crossing locations over scale space for a one-dimensional finite polynomial signal. Curtis et al. (1987) proved the theoretical completeness of zero-crossing locations for a restricted class of band-limited two-dimensional signals. However, Hummel and Moniot (1989) have argued that these theorems are unlikely to lead to stable reconstructions in practice, and have noted that theoretical results for polynomial or otherwise-restricted signals may be of little relevance to the problem of reconstructing more general continuous functions. In fact there are many examples of substantially different continuous functions which have the same zero-crossings at all scales in Gaussian scale space (Mallat and Zhong, 1992). Thus it is clear that the zero-crossing locations alone must be incomplete.

Hummel and Moniot (1989) have shown that if the zero-crossing locations over scale space of an n-dimensional signal are supplemented with gradient data at the zero crossings, then the original signal can, in principle, be exactly reconstructed. They noted, however, that their method for exact reconstruction is provably unstable.

4.2. Previous Empirical Results

Carlsson (1984, 1988) introduced an algorithm which computes an approximate reconstruction of an image from information coded at the image edges. The specific edge representation consists of edge locations and the grey level values on either side of the encoded edges. The reconstruction algorithm computes

an optimally smooth interpolation of the grey level image between these edge points. Carlsson employed the two-dimensional Laplacian operator as a measure of smoothness, and used a multi-resolution technique to solve Laplace's equation with boundary conditions imposed by the grey-level values at the edge points.

In Carlsson's system, candidate edge points are identified by first locating the zero-crossings of the image convolved with a small-scale Laplacian of Gaussian operator. These edge points are then thresholded based upon the output of simple non-steerable horizontal and vertical gradient filters. Carlsson found that the edges thus detected contain many small gaps, which he attributed to the anisotropy in his gradient detection scheme. He therefore followed this thresholding stage by a gap-filling stage designed to reconnect the fragmented contours.

In his original algorithm (Carlsson, 1984), Carlsson employed this edge representation to produce reconstructions which were intelligible but contained significant perceptual flaws. Texture and shading were largely absent, certain features were badly smudged, and defocused structures appeared unnaturally sharp.

In a later version of the algorithm (Carlsson, 1988), Carlsson augmented the edge representation with a subsampled low-pass residual image, which was expanded and added to the edge-based reconstruction to correct for low spatial-frequency errors. This produced a slight improvement in the representation of shaded and defocused structures, although many perceptually significant errors remained.

Zeevi and Rotem (1986) developed an image representation in which the image is decomposed into multiple bandpass sub-images which tile frequency space. The binary signum function *sgn* of these bandpass subimages is then encoded. The low-frequency information is also encoded as a lowpass sub-sampling of the original image. Reconstruction proceeds as a sequence of iterative one-dimensional horizontal and vertical reconstructions of the bandpass sub-images, which are then recombined with the low-frequency sub-image. While the iterative technique used for reconstructing the sub-bands is not guaranteed to converge (Hummel and Moniot, 1989), results of the reconstruction are generally quite good.

Hummel and Moniot (1989) developed an algorithm for stable image reconstruction from zero-crossing locations and gradient data at zero-crossings over scale space, using a method based on minimizing equation error. The reconstructions are quite good, although there remain perceptual differences between the reconstruction and the original image. Also, since the algorithm requires both the zero-crossing locations and gradient vectors along the zero-crossings over the entire scale space, the representation is highly redundant and significantly larger than the original image.

Grattoni and Guiducci (1990) developed a representation of images based upon a model of an edge as a Gaussian-blurred step discontinuity, encoding edge location, brightness, contrast, blur and 'contour width'. The latter parameter specifies the spatial extent, in the normal direction to the contour, over which the model is considered valid. Edges are detected at the finest scale (filter scale of 0.5 pixels), and edge parameters are estimated by fitting the model to the gradient operator response as a function of filter scale. Images are reconstructed by rendering the edge models within their domain of validity, and interpolating using the Laplace equation at remaining image points. The approach was later extended to colour images (Cumani et al., 1991). Results are reasonable, although images appear to lack fine detail and have a blurred appearance. It should be noted that at a scale of 0.5 pixels, edge density is typically very high (on the order of 30%), and hence most pixels in the image are adjacent to at least one edge pixels. Thus reasonable reconstructions are not terribly surprising.

Mallat and Zhong (1992) have developed an algorithm for reconstructing an image from the locations, amplitudes and directions of the gradient maxima of the image, sampled at octave scales over Gaussian scale space. They report that reconstructions are perceptually indistinguishable from the original image. As with the approach of Hummel and Moniot (1989), however, this representation is highly redundant, since edges are represented at many scales, and at the finest scales edge density is very high. Using techniques similar to those of Carlsson (1984), Mallat and Zhong have achieved a more compact code by thresholding the edges detected at the finest scales, and encoding the low-frequency variations as a sub-sampled grey-scale image. While this representation is far more compact, the reconstructions lack texture detail and some defocused structures are not recovered.

4.3. Analysis

On the theoretical side, the most recent results (Hummel and Moniot, 1989) suggest that exact

reconstruction of a two-dimensional image from zero-crossing locations and gradient data at zero-crossings over scale space may be theoretically possible, but stability problems make exact reconstruction unlikely in practice.

Recent empirical results are more encouraging. The highest-quality reconstructions appear to come from Mallat and Zhong (1992). However these high-quality reconstructions require a highly redundant encoding of the edges over scale space. In order to achieve reasonable reconstructions from more compact representations, Zeevi and Rotem (1986), Carlsson (1988) and Mallat and Zhong (1992) have all resorted to hybrid representations, which combine the encoding of a subset of edges (hopefully, the most important ones) with a low pass sub-sampled grey-level image.

In our view, the main impediments to developing reconstructible image representations which are purely edge-based are limitations in the methods employed for detecting and characterizing local edge structure. For example, Carlsson's results clearly suffer from the limitations of the edge detection algorithm he employs. Anisotropies in linear filtering, and a failure to adapt filter scale to local signal-to-noise ratios are two of the more serious problems. Most importantly, the use of a fixed spatial scale and fixed threshold creates a dilemma: if a high threshold is chosen, then many important edges are missed and the reconstruction fails to represent important image structure. If a lower threshold is chosen, blurred gradients due to shadows, shading and defocused edges are detected, but when reconstructed these edges appear unnaturally sharp, leading to perceptually objectionable artifacts. In Carlsson's words, "contours due to illumination variation have a tendency to stand out in front of the smooth surround" (Carlsson, 1988).

Mallat and Zhong (1992) avoid these problems by encoding the structure at each edge point at many scales. As noted above, this yields a large, highly-redundant representation. There are two additional problems with this representation for computer vision applications. First, although inclusion of the finest-scale zero-crossings over the entire image allows small-scale detail to be reconstructed, many of these edges are in fact artifact (false positives) due to sensor noise, which violates our reliability criterion and creates problems for higher-level computations. Second, each edge point is represented as a distributed code over many scales, and the correspondence between representations of the same edge at different scales is not computed. Thus there is no explicit compact characterization of an edge which is rapidly available for further inference, and the representation fails to satisfy our explicitness criterion.

The representation proposed by Grattoni and Guiducci (1990) has the advantage that edge blur is represented explicitly, rather than as a distributed code over scale space. However, as with Mallat and Zhong (1992), the representation of all edges detected at the finest scale (most of which are artifact) violates our reliability and concision criteria. Also, dependence upon a 'contour width' parameter introduces significant complexity, requiring the tracking of edge points through scale space, and a spatially heterogeneous reconstruction method that would be difficult to parallelize. It is hard to argue that contour width (as distinct from blur) is a critical perceptual parameter that should be explicitly represented.

In this paper, we will show that by addressing these problems, the local scale control algorithm (Elder and Zucker, 1996b, 1998) produces a compact, purely edge-based representation which is perceptually nearly complete. Each edge point is explicitly represented only once, not as a distributed code over scale space. No additional lowpass image is required to capture the low frequency grey-level variations. The representation is based on the hypothesis that the most critical perceptual parameters of an edge are its location and the magnitude and blur of the underlying change in brightness (Elder and Zucker, 1995). These are the parameters which are estimated at each edge point, and which are used in the reconstruction algorithm.

The proposed representation is more compact and explicit than the multi-scale representation employed by previous methods (e.g. Hummel and Moniot, 1989; Mallat and Zhong, 1992) in that the local scale control algorithm chooses a single most appropriate scale (the *minimum reliable scale*) for estimation at each edge point. An edge is thus represented by a point in scale space, with curves in the image sweeping out space curves in scale space, producing a one-dimensional encoding of the two-dimensional image. In contrast, typical multi-scale techniques represent each edge point as a curve lying in the scale space volume, so that image curves sweep out surfaces in scale space, producing a two-dimensional encoding of the two-dimensional image.

The use of steerable filters (Freeman and Adelson, 1991; Perona, 1995) ensures that edges are detected and represented with isotropic accuracy, and the use of the

minimum reliable scale at each point ensures that while actual brightness changes in the scene are accurately characterized, artifactual changes due to sensor noise are ignored. The explicit encoding of blur scale is useful for the interpretation of shadows, shading and focal blur (Elder and Zucker, 1998). No additional 'contour width' parameter is required, and reconstruction is accomplished with a spatially homogeneous multiorid algorithm that maps directly to parallel architectures.

4.4. The Reconstruction Algorithm

The edge representation produced by local scale control represents not only the estimated position of the edge, but also the intensity values I_l and I_d on either side of the edge and the estimated blur scale σ_b of the edge. A very coarse encoding of the gradient direction at each edge point into the 4 diagonal directions is also required for reconstruction. No attempt is made to compress this information further: at this time our main interest is to determine the quality of reconstructions that are achievable from a purely one-dimensional edge-based representation where edges are represented at only a single scale at each edge point.

Reconstruction proceeds in two stages. In the first stage, the intensity function over the image is reconstructed using only the position, intensity and 2-bit gradient direction values at each edge point. The method employed is very similar to that used by Carlsson (1984, 1988). The fundamental assumption underlying the reconstruction is that the Laplacian $(\frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2})$ of the original image is zero at all non-edge points. Under this assumption, a unique interpolation of the intensity function between the edges is computed by minimizing the Laplacian over the image. The problem is thus one of solving Laplace's equation with boundary conditions imposed by the intensity values at the detected edges. Reflection boundary conditions are used at the frame of the image and the initial conditions are zero at all non-edge points.

The most straightforward solution to Laplace's equation is to solve the heat equation by Gaussian diffusion until it reaches steady state. The only complication arises from the fact that each boundary point (i.e. each edge point) represents two different boundary conditions: one for the light side of the edge (I_l) and one for the dark side of the edge (I_d). Using a small 4-tap

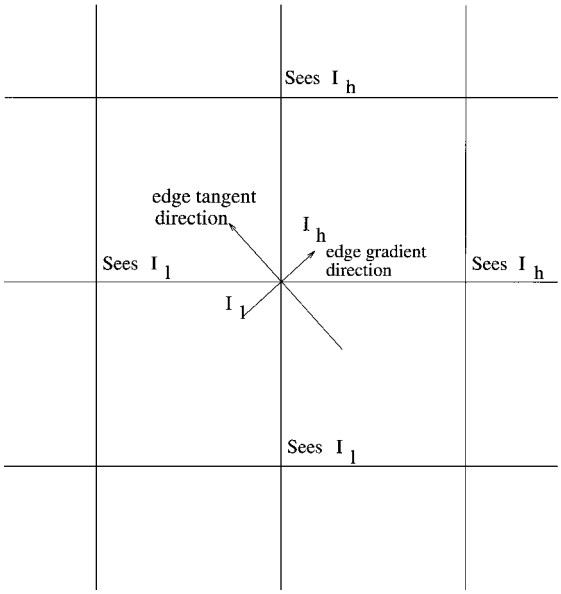

Figure 7. In solving the Laplace equation, boundary conditions at edge points are determined by either the dark intensity I_d or the light intensity I_l, depending upon the dot product of the gradient direction at the edge and the location of the kernel centre relative to the edge.

smoothing kernel given by

$$\begin{pmatrix} 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 \end{pmatrix},$$

this problem is resolved by selecting the appropriate luminance value at each edge point based upon the 2-bit gradient direction at the edge point and the position of the kernel centre with respect to the edge point (Fig. 7). We define the kernel displacement vector as the vector from the edge point to the kernel centre. At points where the dot product of the kernel displacement vector with the gradient vector is positive, I_l is used as the boundary condition at the edge. At points where the dot product is negative, I_d is used as the boundary condition.

Figure 8 (middle left) shows the solution to the heat equation using this algorithm for a lower resolution (256×384) image of the mannequin figure at an intermediate stage (625 smoothing iterations).

This simple diffusion algorithm is far too slow to be practical, requiring tens of thousands of iterations to converge. Carlsson (1988) used a multiresolution method in which the edge representation

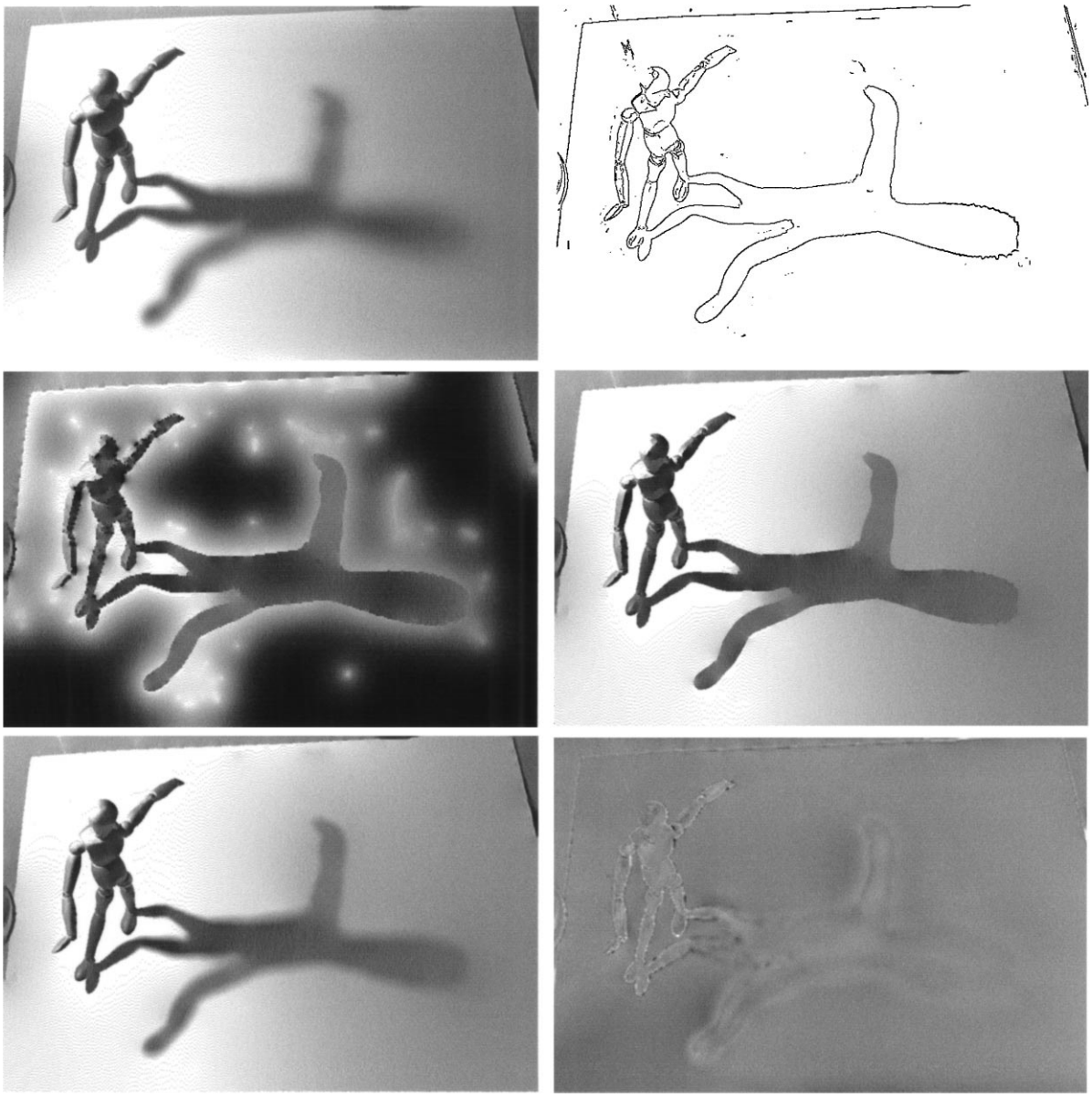

Figure 8. *Top left*: Original image. *Top right*: Detected edge locations. *Middle left*: Intermediate solution to the heat equation. *Middle right*: Reconstructed luminance function. *Bottom left*: Reblurred result. *Bottom right*: Error map (reblurred result—original). Bright indicates overestimation of intensity, dark indicates underestimation. Edge density is 1.7%. RMS error is 10.1 grey levels, with a 3.9 grey level DC component, and an estimated 1.6 grey levels due to noise removal.

is sub-sampled and Laplace's equation solved on a small grid and then used as an initial condition for solution on a grid twice the size. This procedure is repeated until Laplace's equation is solved at full resolution. In our approach, we have implemented a full multigrid algorithm (Press et al., 1992, pp. 871–881) in which Laplace's equation is solved at successively greater resolution by using coarser resolutions to compute a correction term from the residual at the current scale.

The Laplacian kernel used to compute the residual is implemented in finite difference form as

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

The corresponding smoothing kernel is

$$\begin{pmatrix} 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 \end{pmatrix}$$

Standard red-black Gauss-Seidel relaxation is used for smoothing. Downsampling is performed by averaging high resolution pixels in blocks of 4. Upsampling is performed by copying each low resolution pixel into 4 high resolution pixels. The main design decision in implementing the multigrid solution is how to subsample (restrict) the boundary conditions given by the dark and light intensity estimates at the edge pixels. We have taken the following approach: if one or more edges exist in a block of 4 pixels to be subsampled, the highest contrast edge is selected and copied to the low resolution grid. In all other respects, the reconstruction is a standard full multigrid algorithm.

Convergence of the algorithm can either be assessed by measuring the maximum residual, or by visually inspecting successive reconstructions. By the first measure, iteration is halted when the maximum residual is less than one gray level. By the second measure, iteration is halted when two successive images are perceptually indistinguishable. We find empirically that by either measure, convergence is achieved roughly twice as fast as for Carlsson's method, requiring 3-4 cycles of the multigrid algorithm, with 6 smoothing iterations at each cycle. This takes roughly 15 s on a 200 MHz Pentium Pro computer for a 512×512 image.

The result of this algorithm on the image of the mannequin casting a shadow is shown in Fig. 8 (middle right). While the brightness values appear roughly correct, note that the cast shadow appears highly unnatural. Many observers perceive it as a hole or a painted figure on the ground surface rather than as a cast shadow. Also note that the mannequin surface appears less smooth than in the original image. Figure 9 illustrates the cause of these problems. Although the solution to Laplace's equation may have produced a perceptually reasonable interpolation of brightness values between edges, the boundary conditions at the edges are such that all edges have been completely deblurred and appear perfectly sharp.

It is clearly important that a representation capture the blur characteristics at the edges of the original image in order to approach perceptual completeness. However, restoring the local blur information to the reconstructed image is problematic since we only have blur estimates at edge points, yet we need to smoothly reblur the image not just at these points, but in the general neighbourhood of each point, and the size of this neighbourhood depends upon the local degree of blur.

There is an elegant solution to this problem (Fig. 10). We again solve Laplace's equation over the image, but this time blur scale, rather than luminance, is used as the diffusive quantity, and the boundary conditions are the local estimates of blur scale along the detected contours. In this way, a blur surface is constructed which assigns a blur scale to each pixel in the image. The reconstructed luminance function can then be reblurred with an isotropic Gaussian blur kernel with space-varying blur scale, where the blur scale at each pixel is selected from the corresponding point on the reconstructed blur surface. If we let $I(i, j)$ represent the initial reconstructed intensity function at pixel (i, j), $\sigma_b(i, j)$ represent the estimated blur scale at pixel (i, j), and $N(i, j)$ represent the support of a truncated isotropic Gaussian kernel at pixel (i, j), then the reconstructed intensity function $\hat{I}(i, j)$ after reblurring is given by

$$\hat{I}(a, b) = \sum_{(i,j)\in N(a,b)} \frac{1}{2\pi\sigma_b^2(a, b)} e^{-\frac{(a-i)^2+(b-j)^2}{2\sigma_b^2(a,b)}} \times I(i, j)$$

The results of the complete reconstruction algorithm are shown in Fig. 8 (bottom left). Note that the shadow appears much more realistic, and the mannequin has been restored to its original smoothness. We stress that this reconstruction is computed without resorting to a low-frequency sub-image and that each edge is represented only once, not over the entire scale space. Edges account for only 1.7% of the pixels in this image.

In practice we find that for realistic rendering of blur, the isotropic blur kernel must extend to approximately $\pm 2\sigma_b$. Efficiency can be improved by pre-computing the blur kernels: empirically we find that blur scales can be sampled in 0.5 pixel intervals without perceptible artifact. Reblurring is found to take roughly 5 s

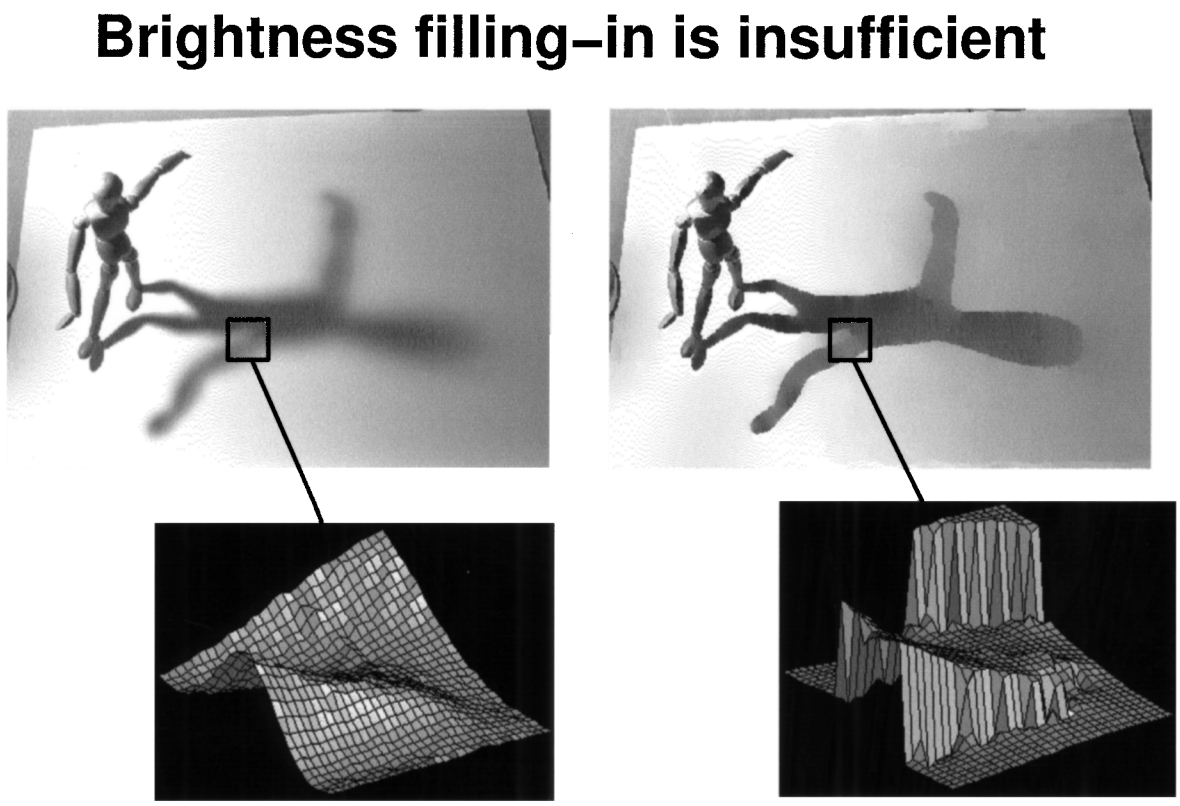

Figure 9. Reconstruction of brightness alone leads to perceptually significant artifact. *Left*: Original image, with 3D rendering of intensity function for indicated region. *Right*: Reconstruction obtained by solving Laplace's equation for brightness, with 3D rendering of intensity function for indicated region. Deblurring of intensity edges leads to artifactual rendering of shadows and shading.

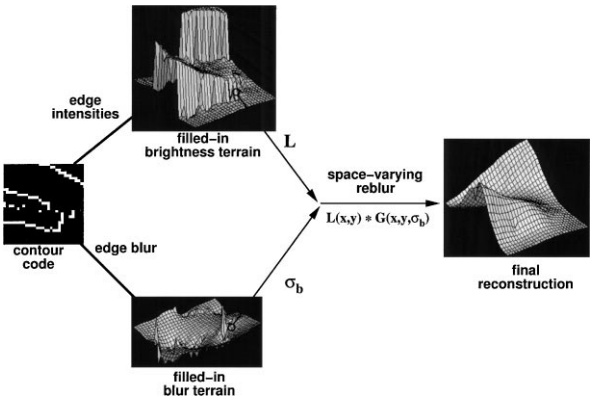

Figure 10. The complete reconstruction algorithm. Brightness and blur reconstructions are computed in parallel by solving Laplace's equation. These two functions are then recombined with a space-varying convolution to compute the final reconstruction.

on a 200 MHz Pentium Pro computer for a typical 512×512 image.

The RMS error of the reconstruction is 10.1 grey levels for this image. The DC component of this error is 3.9 grey levels, the reconstruction being slightly darker than the original image. This DC error may be due to the fact that considerable parts of the original image were saturated at a luminance of 255. Since estimated luminances were thresholded at 255, this may have introduced a systematic negative bias in luminance estimation.

In addition to the DC component, roughly 1.6 grey levels can be attributed to the sensor noise in the original image which has been removed in the reconstruction. This leaves an RMS error of roughly 9.2 grey levels due to other factors. We discuss sources of error in Section 5.

It should be noted that the perceptual fidelity of the reconstruction is better than might be predicted by the RMS accuracy. In other words, although the edge code is technically lossy, it appears to retain the perceptually critical image features, while discarding only unimportant (largely imperceptible) information. What appears to matter most to perceptual fidelity is accurate representation of edge information. This underscores the

advantage of a representation that represents edges explicitly, rather than approximately and indirectly over a compressed wavelet code.

4.5. Reconstruction Results

Given that the proposed representation is entirely edge-based, it is important to test our encoding and decoding algorithms on images which are not necessarily readily described by edges. Figure 11 shows the results of our algorithm on an image of low-pass filtered Gaussian i.i.d. noise. Certainly edges have nothing to do with the construction of this image, and yet, unavoidably, edges emerge, as zero-crossings in the second derivative of the intensity function, steered in the gradient direction. The edge code is a potentially very compact representation of the image (edge density = 6%), and reconstruction results are excellent. RMS error is 7.1 grey-levels, with a 1.4 grey-level DC component.

The natural image of a cloud shown in Fig. 12 also has an amorphous quality which might not seem readily described by edges. Yet in fact an edge encoding is seen to be both highly compact (edge density = 1.8%) and nearly complete: reconstruction results are perceptually very accurate. The amorphous quality of the cloud is captured by the edge code in terms of a high degree of edge blur and a lack of good contour continuity. RMS error is 7.1 grey-levels, with a 2.3 grey-level DC component.

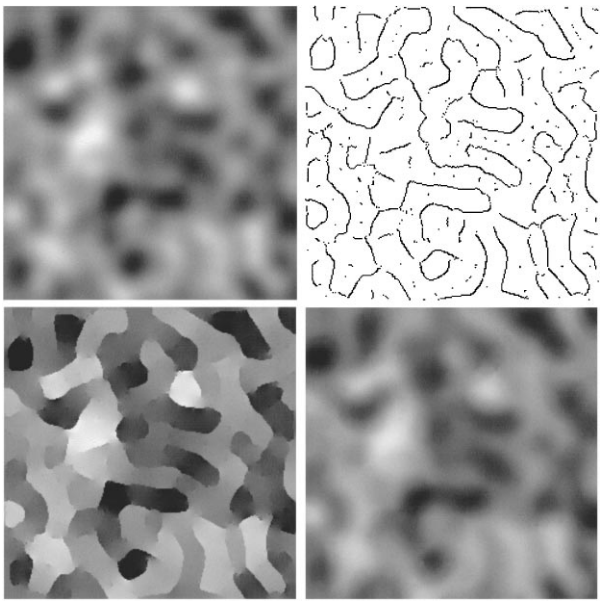

Figure 11. *Top left*: Image of low-pass filtered Gaussian i.i.d. noise. *Top right*: Detected edges. *Bottom left*: Reconstructed intensity function. *Bottom right*: Reblurred result. Edge density is 6%, and RMS error for this reconstruction is 7.1 grey levels, including a 1.4 grey-level DC component.

Figure 13 shows the edge representation and reconstruction for a scene with shallow depth-of-field, and hence a substantial range of focal blur. Observe that reconstruction of the luminance signal alone leads to a highly artifactual rendering, which appears more as a painting than as a photograph. It is apparent that focal blur can be an important perceptual component of a natural image. Figure 13(d) shows how restitution of the blur signal yields a reconstruction which is perceptually nearly flawless. Edge density is 12% and RMS error is 6.7 grey levels, including a 1.0 grey-level DC component and an estimated 1.6 grey levels due to noise removal.

Figure 14 affords an opportunity to examine how well the proposed edge code captures information about shading and highlights. Note again how reconstructing the brightness signal alone produces an unrealistic, painting-like effect. However, restitution of the blur signal allows shading structure and highlights to be faithfully restored. The sensor noise in the original image is higher than the previous images, with a standard deviation of roughly 2.5 grey levels. One can clearly see the noise-removing function of the algorithm in comparing the surface of the bell pepper in the foreground in the original and reconstructed images. For this image, edge density is 10% and RMS error is 10.0 grey levels, including a 1.0 grey-level DC component, and an estimated 2.5 grey levels due to noise removal.

Given the high sensitivity of the human visual system to the details of human faces, portrait images are a good test of the perceptual accuracy of an edge code. Figure 15 shows how reconstructing the brightness alone fails to properly render the smooth facial shading. Once the blur component is restored, however, the rendering is perceptually nearly flawless, aside from the removal of noise. Note, however, that the rendering of the detailed texture in the hat, scarf and hair requires many edge points: edge density is 18%. In order to apply the proposed edge code to the compression of such images, it is clear that some method must be found for efficient coding of this texture information. The demonstrated near-invertibility of the proposed edge code raises the interesting possibility of developing compact texture coding schemes entirely

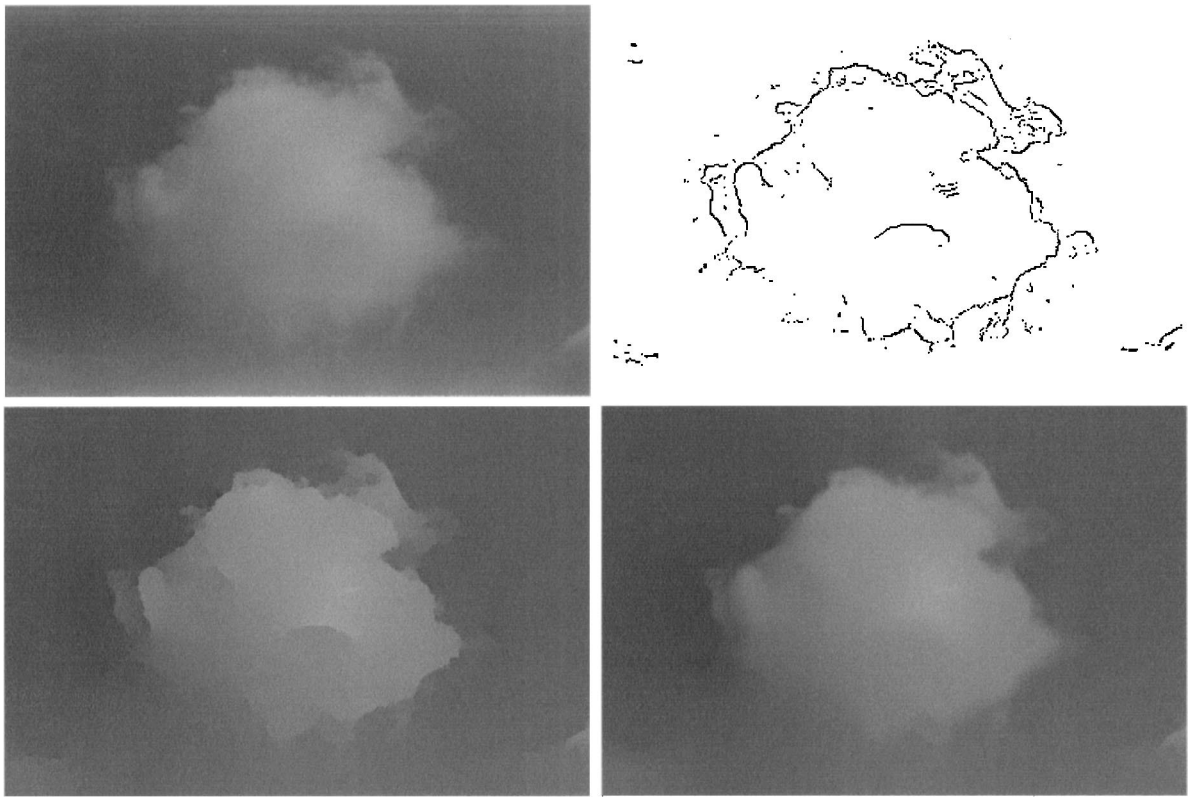

Figure 12. *Top left*: Image of cloud. *Top right*: Detected edges *Bottom left*: Reconstructed intensity function. *Bottom right*: Reblurred result. Edge density is 1.8%, and RMS error for this reconstruction is 7.1 grey levels, including a 2.3 grey-level DC component, and an estimated 1.6 grey levels due to noise removal.

in the edge domain. RMS error for this reconstruction is 5.0 grey levels, including a 0.1 grey-level DC component, and an estimated 1.7 grey levels due to noise removal.

Accurate reconstruction with simultaneous noise removal will of course become more challenging as noise levels increase. Figure 16 shows reconstruction for an image where the standard deviation of the noise is estimated at 3.3 grey levels. Results are still very good perceptually. Edge density is 10%, and RMS error for this reconstruction is 8.4 grey levels, including a 0.1 grey-level DC component, and 3.3 grey levels due to noise removal.

At even higher levels of noise, artifacts of the reconstruction become more apparent. Figure 17 shows an image where the standard deviation of the noise is estimated at 5.8 grey levels. This high level of noise causes the local scale control algorithm to automatically select larger filter scales, preventing smaller signals from being detected. This is most evident in the reconstruction of the young girl's legs: failure to encode the attached shadows have produced an eerie shapeless and transparent quality to their reconstruction. Edge density for this image is 5%, and RMS error for this reconstruction is 11.1 grey levels, including a 2.7 grey-level DC component, and an estimated 5.8 grey levels due to noise removal.

The tradeoff between noise removal and feature loss is most apparent in Fig. 18, where the noise standard deviation is estimated at 7.0 grey levels. While the smooth quality of the skin has been restored quite nicely, detail has been lost in the hair and blouse. Edge density for this image is 5%, and RMS error for this reconstruction is 12.2 grey levels, including a 0.7 grey-level DC component, and an estimated 7.0 grey levels due to noise removal.

Figure 19 shows the effects of errors in estimating the sensor noise variance. When noise variance is underestimated, reconstruction is actually improved, since fine details in the image are better represented. However, this improvement comes at the expense of representing and reconstructing a large number of false

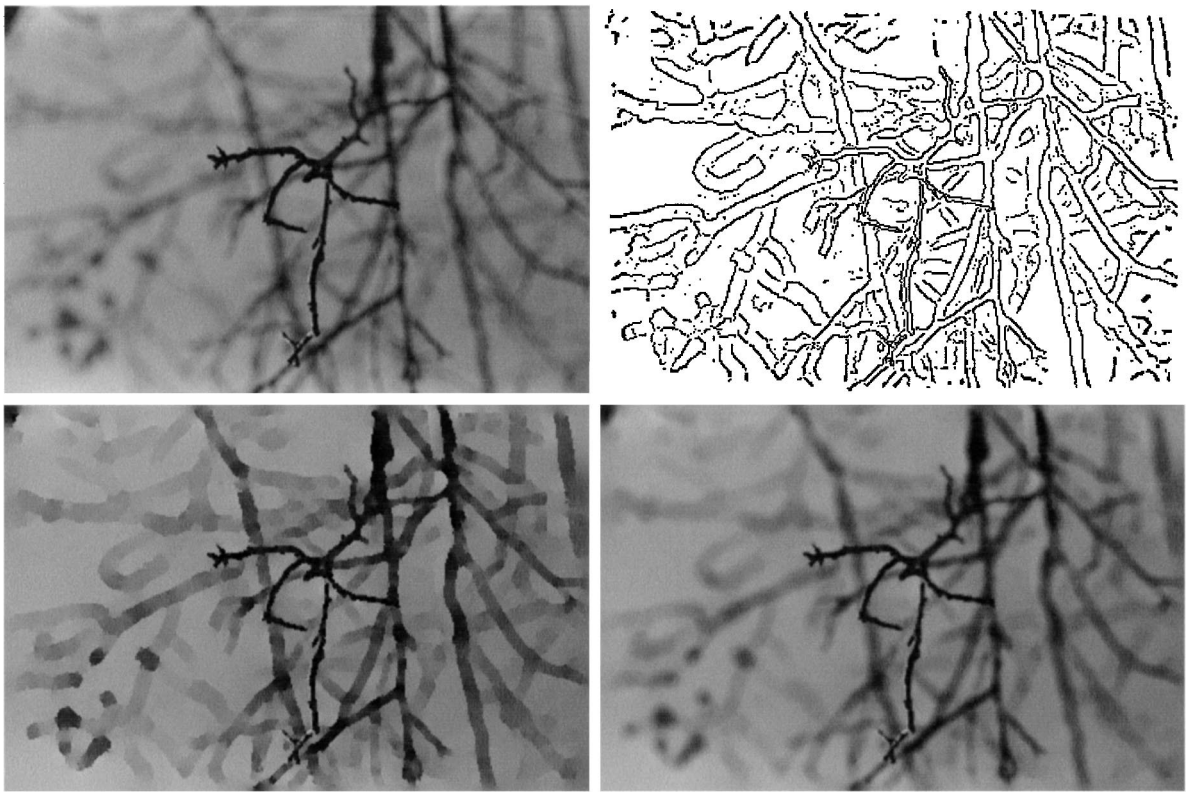

Figure 13. *Top left*: A photograph of tree branches with small depth of field ($f/3.5$) and near focus. *Top right*: Edge map computed by local scale control. *Bottom left*: Reconstructed luminance terrain. *Bottom right*: Reconstructed, reblurred result. Note the importance of reconstructing the blur information for images with finite depth–of-field. Edge density is 12% and RMS error is 6.7 grey levels, including a 1.0 grey-level DC component and an estimated 1.6 grey levels due to noise removal.

positive edges due solely to sensor noise (edge density is 25%). Overestimating noise variance does not perceptually improve the denoising behaviour of the local scale control algorithm, but does lead to further loss of fine detail, i.e. false negatives (edge density is 3%). Both types of errors lead to a decrease in the reliability of the edge code.

5. Analysis of Errors

Figure 8 (lower right) shows an error map for the reconstructed mannequin image, where intensity is proportional to the difference between the final reconstruction and the original image (bright for overestimates of intensity, dark for underestimates). The error map indicates that most of this error is concentrated near the edges of the image, particularly near the sharp edges. There are at least two plausible explanations for this general pattern of errors:

1. Small errors in localization of edges result in local shifts of the reconstructed intensity function. These shifts produce intensity error proportional to intensity gradient.
2. Small errors in the edge model and/or estimation of model parameters produce unbiased intensity errors, which are attenuated by averaging at points more distant from the edges.

There are many specific potential sources of error in our representation and reconstruction algorithm. We discuss each of these in turn.

5.1. Texture Flows

Our representation is based upon a model of an edge as a Gaussian-blurred step discontinuity. The rationale for this representation is that edges signal important changes in scene properties (object boundaries, surface creases, pigment changes, shadows, etc...), and the step discontinuity model appears to be the simplest possible model to describe these events.

The step edge model fails when edges are located very close together to form stripes or texture flows, for

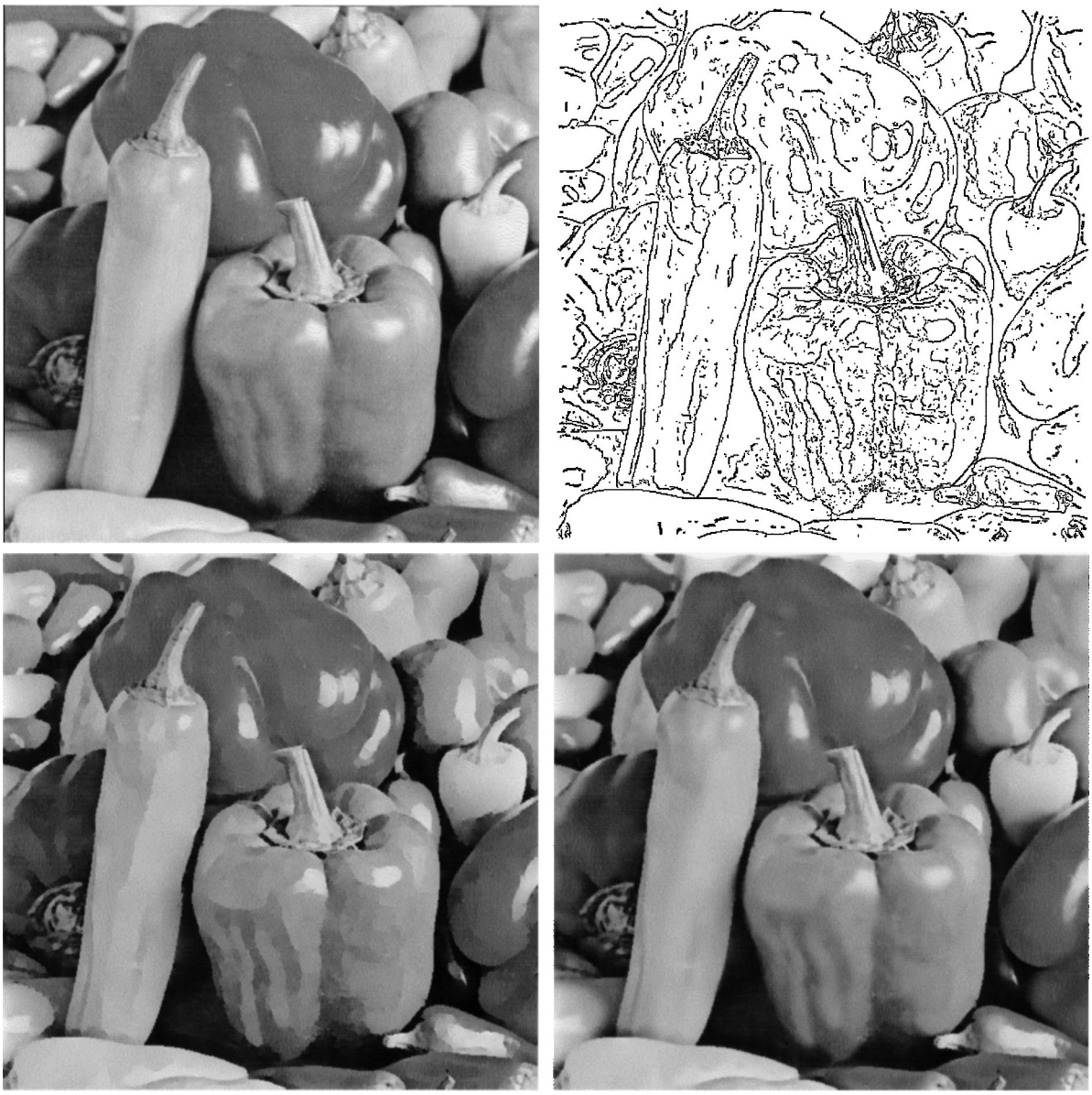

Figure 14. *Top left*: Original image. *Top right*: Detected edges. *Bottom left*: Reconstructed luminance function. *Bottom right*: Reblurred result. Note the importance of reconstructing the blur function for accurate rendering of shading structure. Edge density is 10% and RMS error is 10.0 grey levels, including a 1.0 grey-level DC component, and an estimated 2.5 grey levels due to noise removal.

example in hair or corduroy texture. The primary cause of failure is a decrease in signal-to-noise ratio caused by interference between these neighbouring edges. The evidence for these errors is blur and loss of detail in reconstructions of hair and fine texture.

There are numerous methods in the literature for detecting thin curvilinear structures in images (e.g. Fischler et al., 1981, Cox et al., 1990; Iverson and Zucker, 1995). We are presently studying methods which could be integrated with the local scale control approach.

5.2. Pixel Aliasing

Edges are localized to the nearest pixel in our representation. This results in "jagginess" (pixel aliasing) in oblique contours which are relatively sharp. While

Figure 15. *Top left*: Original image. *Top right*: Detected edges. *Bottom left*: Reconstructed luminance function. *Bottom right*: Reblurred result. Reblurring is critical to capturing the smooth shading of the skin. Edge density is 18%, and RMS error for this reconstruction is 5.0 grey levels, including a 0.1 grey-level DC component, and an estimated 1.7 grey levels due to noise removal.

we have developed methods for sub-pixel localization using local scale control, we have not yet integrated these into our reconstruction method to eliminate this problem.

5.3. Kurtotic Noise

The local scale control algorithm assumes that the probability distribution function of the sensor noise is

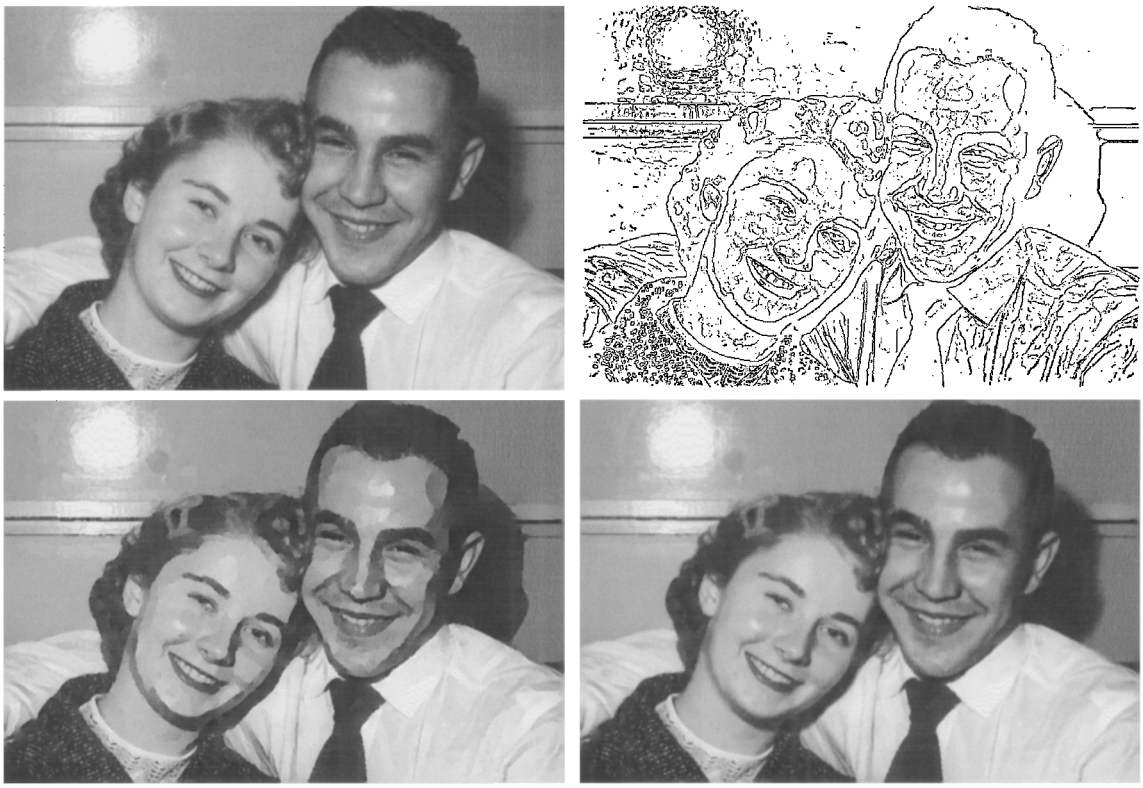

Figure 16. Top left: Original image. Top right: Detected edges. Bottom left: Reconstructed luminance function. Bottom right: Reblurred result. Edge density for this image is 10%, and RMS error for this reconstruction is 8.4 grey levels, including a 0.1 grey-level DC component, and an estimated 3.3 grey levels due to noise removal.

Gaussian, and requires an estimate of the variance of the noise in order to guide scale adaptation. Counter to this assumption, we have found in practice that many images contain noise with a kurtotic distribution. In such cases, the longer tails of these actual distributions lead to a greater number of false positives than predicted by the Gaussian model.

Since these false edges are typically of low contrast and often blurred, their effect on reconstruction may be minimal. However, these edges do decrease the potential compression of the representation, and pose a problem for higher-level visual inference.

To solve this problem, we require a more complex noise model tuned to match not only measured variance, but also measured kurtosis. This model could then be used to derive the appropriate critical value functions for local scale control.

5.4. Blur Model Errors

The proposed edge representation uses a Gaussian model for local image blur, even though geometric optics would predict a "pillbox" model for common focal and penumbral blur scenarios (Elder and Zucker, 1998). Differences between images blurred with pillbox and Gaussian blur kernels of similar size are apparent mainly because of the Mach bands visible in the former but missing in the latter. However, the more critical problem may be the systematic overestimation in contrast that results from applying the Gaussian model to blur which is closer to that predicted by geometric optics (Elder, 1997). This error in contrast estimation can lead to large RMS errors in reconstruction.

We have developed a more general model for local image blurring which can in principle be used to estimate not only the amount of local image blurring, but also the shape of the blur kernel (Elder, 1997). We have not yet fully integrated this new model into our detection and reconstruction method.

5.5. Reconstruction Models

Reconstruction is based upon the approximation that the Laplacian of the intensity function is zero at all

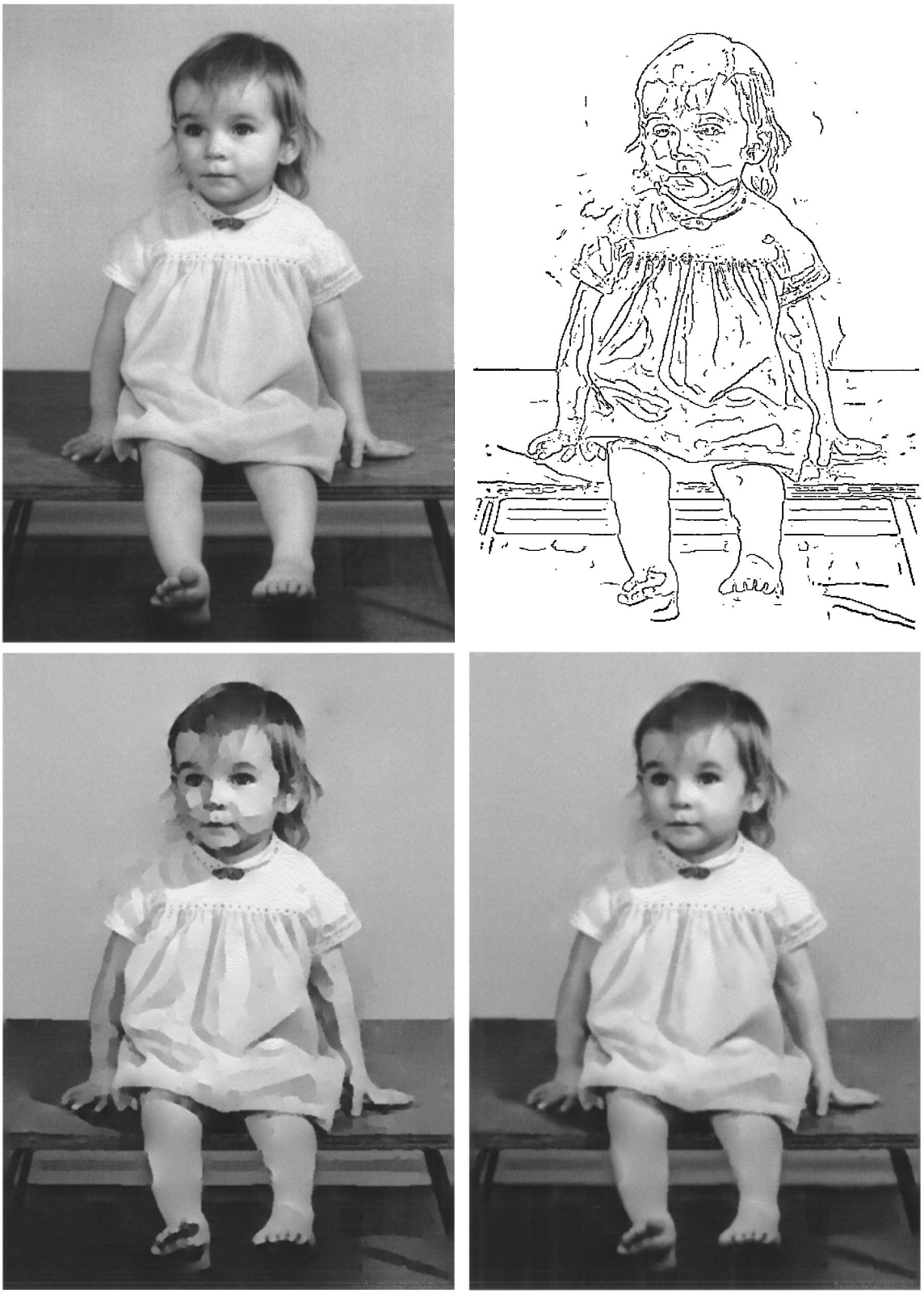

Figure 17. *Top left*: Original image. *Top right*: Detected Edges. *Bottom left*: Reconstructed luminance function. *Bottom right*: Reblurred result. Due to higher noise levels, minimum reliable scales for estimation are generally greater, so that some smaller signals are not detected. Note that the failure to encode the attached shadows on the legs results in an artifactual reconstruction. Edge density for this image is 5%, and reconstruction is 11.1 grey levels, including a 2.7 grey-level DC component, and an estimated 5.8 grey levels due to noise removal.

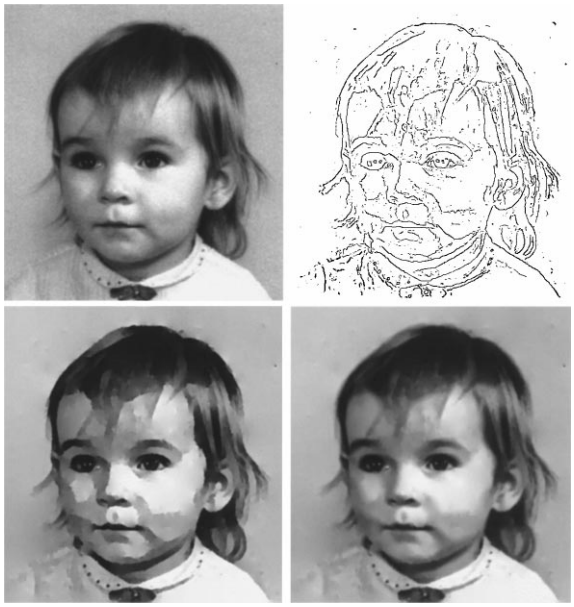

Figure 18. *Top left*: Original image. *Top right*: Detected edges. *Bottom left*: Reconstructed luminance function. *Bottom right*: Reblurred result. The local scale control algorithm is able to remove large amounts of uncorrelated noise; however when noise levels are high, fine features (e.g. in the hair and on the blouse) are also eliminated. RMS error for this reconstruction is 12.2 grey levels, including a 0.7 grey-level DC component, and an estimated 7.4 grey levels due to noise removal.

non-edge points.[3] In differential geometric terms, Laplace's equation constrains the intensity function to be either planar or hyperbolic at all non-edge points. The curvatures of the intensity surface at hyperbolic points must be such that the Laplacian is zero (this does not mean that the mean curvature is zero, in general). Our algorithm can therefore be viewed as a reconstruction of planar and hyperbolic image regions from edge data. In contrast, Barth et al. (1993) have suggested that little information is contained in "low dimensional" planar and parabolic regions of the intensity surface for natural images, and that such regions can be reconstructed from data at "high dimensional" elliptic and hyperbolic regions.

We chose Laplace's equation as an interpolation model because of its simplicity and prior application (Carlsson, 1988), not for any evidence that it is in some way optimal. It would be worthwhile to examine various candidate interpolation models and measure empirically which of these leads to lowest error, both in RMS and perceptual terms.

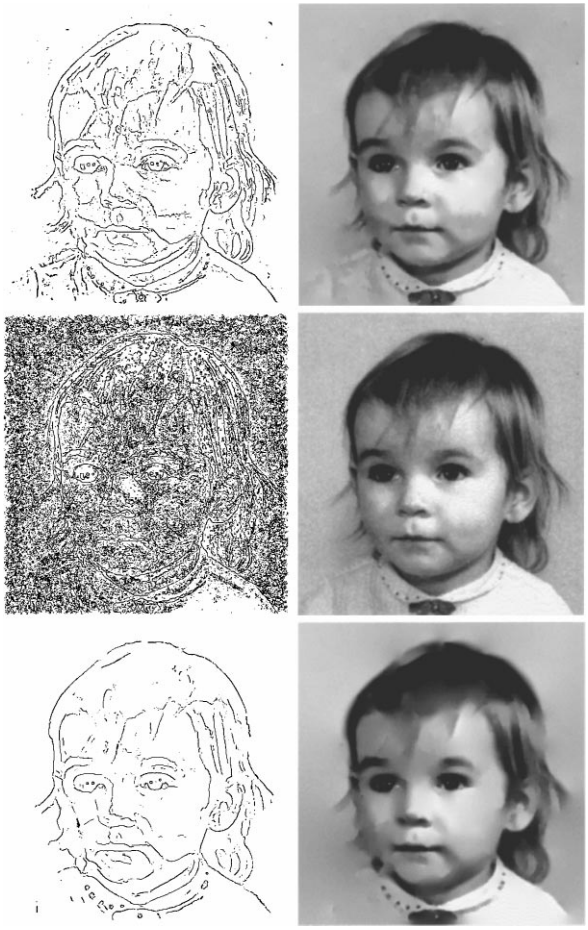

Figure 19. Effects of error in sensor noise input parameter. Standard deviation of sensor noise measured in original image is $s_n = 7.0$ grey levels. *Top*: Edge map and reconstruction when input sensor noise parameter is correct ($s_n = 7.0$ grey levels). Edge density is 5%. *Middle*: Edge map and reconstruction when input sensor noise parameter is underestimated ($s_n = 2.0$ grey levels). Fine image details are much better represented (compare with original image in Fig. 18), but at the expense of representing many false edges produced by sensor noise. Edge density is 25%. *Bottom*: Edge map and reconstruction when input sensor noise parameter is overestimated ($s_n = 20.0$ grey levels). Overestimating the sensor noise does not improve denoising, and leads to additional loss of detail in image structure. Edge density is 3%.

6. Applications

The approximate invertibility of our representation suggests a number of possible applications.

Image Compression. While other invertible edge codes represent each edge at many scales, our

approach detects each edge at only one scale (the minimum reliable scale) and represents each edge by a simple local model, the parameters of which are likely to be highly redundant along edgel chains. The representation thus could form a very good basis for compression. In order to fully evaluate this potential, methods for exploiting the redundancy in edgel position and model parameters must be developed. As for many other methods, a limiting factor in compressibility will be the encoding of texture, which in the edge domain may produce a dense field of edges which are not easily linked into redundant chains.

Image Editing. Image editing systems are essentially pixel-based. The problem with this is that pixels are essentially artifacts of the sensing process, having nothing to do with image content. Edges and contours, on the other hand, have a direct correspondence to important events in the image and scene, and are known to be very important to human perception. The reconstructability of our edge code suggests the possibility of using edges as the primitive working unit in an image editing system. In such a system, artists and designers manipulate edges and contours directly, and the effects their changes produce are seen by reconstructing the resulting modified edge representation. We have recently developed a prototype image editing method called ICE (Interactive Contour Editing) which demonstrates this concept (Elder and Goldberg, 1998).

Deblurring. The initial stage of reconstruction produces an intensity function which is approximately correct, except that all edges are perfectly sharp. This could be considered a perfectly deblurred image. In practical applications, one may wish to remove a general focal blur, while retaining blur variation in the image due to shadows and shading. This could be accomplished by estimating the minimum blur over all edge points, and then correcting the blur of each edge point to remove this common blur factor.

Denoising. The local scale control method is designed to achieve an extremely low false positive rate (less than 5% probability that one or more false edges are detected in an image). The result is an edge representation in which each edge reflects real image structure. The reconstructed image is essentially noise free.

7. Conclusion

To support a diversity of higher-level computations, an early visual representation must make important information explicit and discard only redundant or perceptually irrelevant information. We have argued that an edge code computed by local scale control satisfies these requirements. This edge code represents position, intensity change, blur and gradient direction at each edge point in an image. Edges are detected reliably over a broad range of contrast and blur scale, and edge model parameters may be estimated accurately even for complex, highly discontinuous images. A simple and efficient method for inverting the edge code to reconstruct an estimate of the original image was presented. Restitution of local blur information was seen to be critical to achieving high fidelity reconstructions. The perceptual accuracy of these reconstructions is evidence that the edge code captures the information needed for higher-level visual inference. Immediate applications of the representation include compression, image editing, deblurring and denoising.

Acknowledgments

The author thanks David Jacobs, Michael Langer, John Oliensis and Steven Zucker for helpful comments on the work.

Note

1. The equation for the critical value function $c_2(\sigma_2)$ published in (Elder, 1998) contains a typographical error: the constant 1.8 should be replaced by 1.3.
2. There has been considerable work on the problem of reconstructing an image from two-dimensional image regions, selected by some form of significance criteria (e.g., Barth et al., 1993; Dron, 1977). We will not review these here.
3. While initial intensity reconstruction satisfies this constraint, the image will not obey Laplace's equation exactly after the reblurring stage.

References

Adelson, E. 1991. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, M. Landy and J. Movshon (Eds.), MIT Press: Cambridge, MA.

Adelson, E., Simoncelli, E., and Hingorani, R. 1987. Orthogonal pyramid transforms for image coding. In *Proc. of SPIE*, Vol. 845, Cambridge, MA, pp. 50–58.

Adini, Y., Moses, Y., and Ullman, S. 1997. Face recognition: the problem of compensating for illumination changes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 721–732.

Barlow, H. 1961. The coding of sensory messages. In *Current Problems in Animal Behavior*, W. Thorpe and O. Zangwill (Eds.), Cambridge U. Press, pp. 331–360.

Barrow, H. and Tenenbaum, J. 1981. Computational vision. *Proc. IEEE*, 69:572–595.

Barth, E., Caelli, T., and Zetzsche, C. 1993. Image encoding, labeling and reconstruction from differential geometry. *CVGIP: Graphical Models and Image Processing*, 55(6):428–446.

Belhumeur, P. and Kriegman, D. 1996. What is the set of images of an object under all possible lighting conditions? In *Proc. IEEE Conf. Computer Vision Pattern Recognition*, San Francisco. IEEE Computer Society, IEEE Computer Society Press, pp. 270–277.

Canny, J. 1983. Finding edges and lines in images. Master's thesis, MIT Artificial Intelligence Laboratory.

Carlsson, S. 1984. Sketch based image coding. In *Proc. of Premier Colloque Image*, Biarritz, France, pp. 71–77.

Carlsson, S. 1988. Sketch based coding of grey level images. *Signal Processing*, 15:57–83.

Cox, I., Boie, R., and Wallach, D. 1990. Line recognition. In *Proc. Int. Conf. on Pattern Recognition*, Atlantic City, NJ, pp. 639–645.

Cox, I., Rehg, J., and Hingorani, S. 1993. A Bayesian multiple-hypothesis approach to edge grouping and contour segmentation. *Int. J. Comp. Vision*, 11(1):5–24.

Cumani, A., Grattoni, P., and Guiducci, A. 1991. An Edge-Based Description of Color Images. *CVGIP: Graphical Models and Image Processing*, 53(4):313–323.

Curtis, S., Shitz, S., and Oppenheim, A. 1987. Reconstruction of nonperiodic two-dimensional signals from zero crossings. *IEEE Trans. Acoust., Speech, Signal Processing*, 35:890–893.

Daubechies, I. 1991. Ten lectures on wavelets. In *CBMS-NSF Series Appl. Math., SIAM*.

David, C. and Zucker, S. 1990. Potentials, valleys and dynamic global coverings. *Int. J. Computer Vision*, 5:219–238.

Deriche, R. 1987. Using Canny's criteria to derive a recursively implemented optimal edge detector. *Int. J. Computer Vision*, 1(2): 167–187.

Dron, L. 1977. The multiscale veto model: a two-stage analog network for edge detection and image reconstruction. *Int. J. Comp. Vision*, 56:487–510.

Elder, J. 1997. Brightness filling-in of natural images. *European Conf. on Visual Perception*, Helsinki, Finland, In Perception, 26(Suppl.) 57.

Elder, J., Beniaminov, D., and Pintilie, G. 1999. Edge classification in natural images. *J. Invest. Opthalm. Visual Sci.* 40(4):1897.

Elder, J. and Goldberg, R. 1998. Interactive contour editing. In *Proc. IEEE Conf. Computer Vision Pattern Recognition*, Santa Barbara, CA, IEEE Computer Society, IEEE Computer Society Press, pp. 374–381.

Elder, J., Trithart, S., Pintilie, G., and MacLean, D. 1998. Rapid processing of cast and attached shadows. *J. Invest. Opthalm. Visual Sci.*, 39(4):S853.

Elder, J. and Zucker, S. 1995 The local character of generalized luminance transitions. *J. Invest. Opthalm. Visual Sci.*, 36(4):S836.

Elder, J. and Zucker, S. 1998. Local scale control for edge detection and blur estimation. *IEEE Pattern Anal. Machine Intell.*, 20(7): 699–716.

Elder, J. and Zucker, S. 1996a. Computing contour closure. In *Proc 4th European Conf. on Computer Vision*, Lecture Notes in Computer Science, Springer Verlag, New York, pp. 399–412.

Elder, J. and Zucker, S. 1996b. Local scale control for edge detection and blur estimation. In *Proc. 4th European Conf. on Computer Vision*, Lecture Notes in Computer Science, Springer Verlag, New York, pp. 57–69.

Elder, J. and Zucker, S. 1996c. Scale space localization, blur and contour-based image coding. In *Proc. IEEE Conf. Computer Vision Pattern Recognition*, San Francisco. IEEE Computer Society, IEEE Computer Society Press, pp. 27–34.

Ens, J. and Lawrence, P. 1993. Investigation of methods for determining depth from focus. *IEEE Trans. Pattern Anal. Machine Intell.*, 15(2):97–108.

Fischler, M., Tenenbaum, J., and Wolf, H. 1981. Detection of roads and linear structures in low-resolution aerial imagery using a multi-source knowledge integration technique. *Computer Graphics and Image Processing*, 18(4):201–223.

Freeman, W. and Adelson, E. 1991. The design and use of steerable filters. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(9):891–906.

Grattoni, P. and Guiducci, A. 1990. Contour Coding for Image Description. *Pattern Recognition Letters*, 11:95–105.

Hummel, R. and Moniot, R. 1989. Reconstructions from zero crossings in scale space. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 37(12):2111–2130.

Iverson, L. and Zucker, S. 1995. Logical/linear operators for image curves. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(10):982–996.

Kersten, D., Mamassian, P., and Knill, D. 1997. Moving cast shadows induce apparent motion in depth. *Perception*, 26(2):171–192.

Koenderink, J. 1984. The structure of images. *Biol. Cybern.*, 50: 363–370.

Leclerc, Y. and Zucker, S. 1987. The local structure of image discontinuities in one dimension. *IEEE Trans. Pattern Anal. Machine Intell.*, 9(3):341–355.

Lindeberg, T. 1990. Scale-space for discrete signals. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(3):234–254.

Lindeberg, T. 1996. Edge detection and ridge detection with automatic scale selection. In *IEEE Conf. Computer Vision Pattern Recognition*, San Francisco, IEEE Computer Society, IEEE Computer Society Press, pp. 465–470.

Logan, B.F. 1977. Information in the zero-crossings of bandpass signals. *Bell Syst. Tech. J.*, 56:487–510.

Mallat, S. 1989. Multifrequency channel decompositions of images and wavelet models. *IEEE Trans. Inform. Theory*, 37(12):2091–2110.

Mallat, S. and Zhong, S. 1992. Characterization of signals from multiscale edges. *IEEE Trans. Pattern Anal. Machine Intell.*, 14: 710–732.

Marr, D. 1982. *Vision*. W.H. Freeman:New York.

Marr, D. and Hildreth, E. 1980. Theory of edge detection. *Proc. R. Soc. Lond. B*, 207:187–217.

Nayar, S. and Yasuo, N. 1994. Shape from focus. *IEEE Trans. Pattern Anal. Machine Intell.*, 16(8):824–831.

Paradiso, M. and Nakayama, K. 1991. Brightness perception and filling-in. *Vision Res.*, 7/8:1221–1236.

Pentland, A. 1987. A new sense for depth of field. *IEEE Trans. Pattern Anal. Machine Intell.*, 9(4):523–531.

Perona, P. 1995. Deformable kernels for early vision. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(5):488–499.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. 1992. *Numerical Recipes in C* (2 edition)., Cambridge University Press, chap 19, pp. 871–882.

Roberts, L. 1965. Machine perception of 3-dimensional solids. In *Optical and Electro-Optical Information Processing*, J. Tippett (Ed.), MIT Press: Cambridge, MA.

Sha'ashua, A. and Ullman, S. 1988. Structural saliency: the detection of globally salient structures using a locally connected network. In *Proc. 2nd Int. Conf. on Computer Vision*, Tampa, Florida. IEEE Computer Soc. Press, pp. 321–327.

Simoncelli, E., Freeman, W., Adelson, E., and Heeger, D. 1992. Shiftable multiscale transforms. *IEEE Trans. on Inf. Theory*, 38(2):587–607.

Vetterli, M. 1984. Multidimensional subband coding: some theory and algorithms. *Signal Processing*, 6(2):97–112.

Werner, H. 1935. Studies on contour: I. Qualitative analyses. *Amer. J. Psychol.*, 47:40–64.

Witkin, A. 1983. Scale space filtering. In *Proc. Int. Joint Conf. on Artif. Intell.*, Karlsruhe, pp. 1019–1021.

Yuille, A. and Poggio, T. 1985. Fingerprints theorems for zero crossings. *J. Opt. Soc. Am. A*, 2(5):683–692.

Zeevi, Y. and Rotem, D. 1986. Image reconstruction from zero crossings. *IEEE Trans. Acoust., Speech, Signal Processing*, 34:1269–1277.

Zucker, S. 1986. Early vision. In *The Encyclopedia of Artificial Intelligence*, S. Shapiro (Ed.), John Wiley.

Zucker, S., Hummel, R., and Rosenfeld, A. 1977. An application of relaxation labeling to line and curve enhancement. *IEEE Trans. Comput.*, 26:394–403.