



Edge Detection and Ridge Detection with Automatic Scale Selection

TONY LINDEBERG

*Computational Vision and Active Perception Laboratory (CVAP), Department of Numerical Analysis
and Computing Science, KTH (Royal Institute of Technology), S-100 44 Stockholm, Sweden*

tony@nada.kth.se

Received June 1, 1996; Revised July 10, 1998; Accepted July 30, 1998

Abstract. When computing descriptors of image data, the type of information that can be extracted may be strongly dependent on the scales at which the image operators are applied. This article presents a systematic methodology for addressing this problem. A mechanism is presented for automatic selection of scale levels when detecting one-dimensional image features, such as edges and ridges.

A novel concept of a *scale-space edge* is introduced, defined as a connected set of points in scale-space at which: (i) the gradient magnitude assumes a local maximum in the gradient direction, and (ii) a normalized measure of the strength of the edge response is locally maximal over scales. An important consequence of this definition is that it allows the scale levels to vary along the edge. Two specific measures of edge strength are analyzed in detail, the gradient magnitude and a differential expression derived from the third-order derivative in the gradient direction. For a certain way of normalizing these differential descriptors, by expressing them in terms of so-called γ -normalized derivatives, an immediate consequence of this definition is that the edge detector will adapt its scale levels to the local image structure. Specifically, sharp edges will be detected at fine scales so as to reduce the shape distortions due to scale-space smoothing, whereas sufficiently coarse scales will be selected at diffuse edges, such that an edge model is a valid abstraction of the intensity profile across the edge.

Since the scale-space edge is defined from the intersection of two zero-crossing surfaces in scale-space, the edges will by definition form closed curves. This simplifies selection of salient edges, and a novel significance measure is proposed, by integrating the edge strength along the edge. Moreover, the scale information associated with each edge provides useful clues to the physical nature of the edge.

With just slight modifications, similar ideas can be used for formulating ridge detectors with automatic selection, having the characteristic property that the selected scales on a *scale-space ridge* instead reflect the width of the ridge.

It is shown how the methodology can be implemented in terms of straightforward visual front-end operations, and the validity of the approach is supported by theoretical analysis as well as experiments on real-world and synthetic data.

Keywords: edge detection, ridge detection, scale selection, diffuseness, normalized derivative, Gaussian derivative, scale-space, multiscale representation, feature detection, computer vision

1. Introduction

One of the most intensively studied subproblems in computer vision concerns how to detect edges from grey-level images. The importance of edge information for early machine vision is usually motivated from the observation that under rather general assumptions

about the image formation process, a discontinuity in image brightness can be assumed to correspond to a discontinuity in either depth, surface orientation, reflectance or illumination. In this respect, edges in the image domain constitute a strong link to physical properties of the world. A representation of image information in terms of edges is also compact in the sense

that the two-dimensional image pattern is represented by a set of one-dimensional curves. For these reasons, edges have been used as main features in a large number of computer vision algorithms.

A nontrivial aspect of edge-based analysis of image data, however, concerns what should be meant by a discontinuity in image brightness. Real-world image data are inherently discrete, and for a function defined on a discrete domain, there is no natural notion of “discontinuity”. This means that there is no inherent way to judge what are the edges in a given discrete image. Therefore, the concept of an image edge is only *what we define it to be*.

From this viewpoint, it is easy to understand that a large number of approaches have been developed for detecting edges. The earliest schemes focused on the detection of points at which the gradient magnitude is high. Derivative approximations were computed either from the pixels directly, using operators such as Robert’s cross operator (Roberts, 1965), the Sobel operator (Pingle, 1969) and the Prewitt operator (Prewitt, 1970), or from local least-squares fitting (Haralick, 1984). An important step was then taken by the introduction of *multiscale techniques*. Torre and Poggio (1980) motivated the need for linear filtering as a pre-processing step to differentiation, in order to regularize ill-posed differentiation into well-posed operators. The Marr-Hildreth operator (Marr and Hildreth, 1980) was motivated by the rotational symmetry of the Laplacian operator and the special property of the Gaussian kernel as being the only real function that minimizes the product of the variances of the filter in the spatial and the Fourier domains.¹ Other early techniques with a multiscale character were presented by Rosenfeld and Thurston (1971) and Marr (1976).

Canny (1986) considered the problem of determining an “optimal smoothing filter” of finite support for detecting step edges. The methodology was to maximize a certain performance measure constituting a trade-off between detection and localization properties given a restriction on the probability of obtaining multiple responses to a single edge. He also showed that the resulting smoothing filter could be well approximated by the first-order derivative of a Gaussian kernel. Deriche (1987) extended this approach to filters with infinite support, and proposed a computationally efficient implementation using recursive filters.² Canny (1986) also introduced the notions of nonmaximum suppression and hysteresis thresholding. Similar concepts were developed independently by Korn (1988).

These ideas have then been furthered by several authors, and a large literature has evolved on the design of “optimal edge detectors” for different types of edge models, noise models, and optimality criteria, see, for example, (Nalwa and Binford, 1986; Boyer and Sarkar, 1991; Petrou and Kittler, 1991; Wilson and Bhalerao, 1992).³ Another subject, which has been given large attention during recent years, is the replacement of the linear smoothing operation by a nonlinear smoothing step, with the goal of avoiding smoothing across object boundaries, see (Perona and Malik, 1990; Saint-Marc et al., 1991; Nitzberg and Shiota, 1992; Harr Romeny, 1994) for examples.

A trade-off problem, which is common for all these edge detection schemes, and, in fact, arises for any multi-scale feature detector, concerns how much smoothing to use in the smoothing step. A larger amount of smoothing will, in general, have the desirable effect of increasing the suppression of noise and other interfering fine-scale structures. This, in turn, can be expected to simplify the detection problem at the cost of possible poor localization. A smaller amount of smoothing, on the other hand, can be expected to improve the localization properties at the cost of a lower signal-to-noise ratio. To circumvent this problem, Bergholm (1987) proposed to detect edges at coarse scales, and to follow these to finer scales using edge focusing. He did, however, not present any method for determining the scale level for the detection step, or to what scales the edges should be localized. Hence, edge focusing serves mainly as a selection procedure, which among all the edges at the finer (localization) scale selects those who can be traced to corresponding edges at the coarser (detection) scale.

The subject of this article is to address the general problem of automatically selecting local appropriate scales for detecting edges in a given data set. As will be illustrated by examples later in Section 4, the choice of scale levels can crucially affect the result of edge detection. Moreover, different scale levels will, in general, be needed in different parts of the image. Specifically, coarse scale levels are usually necessary to capture diffuse edges, due to shadows and other illumination phenomena.

To cope with this problem, we will propose an extension of the notion of nonmaximum suppression, with the scale dimension included already in the edge definition. This approach builds upon previous work on feature detection with automatic scale selection (Lindeberg, 1993c, 1994a), based on the idea that in

the absence of further evidence, scale levels for feature detection can be selected from the scales at which normalized measures of feature strength assume local maxima over scales. It will be shown that for two specific measures of edge strength, an immediate consequence of this definition is that fine scale levels will be (automatically) selected for sharp edges, and coarse scales for diffuse edges.

In addition, an edge detector based on this approach computes a diffuseness estimate at each edge point. Such attribute information constitutes an important cue to the physical nature of the edge. Another important property is that the scale levels are allowed to vary along the edge, which is essential to capture edges for which the degree of diffuseness varies along the edge.

In this respect, the approach we will arrive at will have certain properties in common with methods for estimating edge diffuseness (Mallat and Zhong, 1992; Zhang and Bergholm, 1993). With just a slight modification, it can also be used for formulating ridge detection methods with automatic scale selection. Let us, however, defer from discussing these relationships until we have described our own methodology. The presentation begins with a hands-on demonstration of the need for a scale selection mechanism in edge detection. After this, a more detailed outline will be given of how the article is organized.

2. The Need for Automatic Scale Selection in Edge Detection

To illustrate the need for an explicit mechanism for automatic scale selection when detecting edges from image data about which no a priori information is available, let us consider the effects of performing edge detection at different scales.

Figure 1 shows the result of applying a standard edge detector (Lindeberg, 1993b)⁴ to two images, which have been smoothed by convolution with Gaussian kernels of different widths. (To isolate the behavior of the edge detector, no thresholding has been performed on the gradient magnitude.) As can be seen, different types of edge structures give rise to edge curves at different scales. In the left image of the office scene, the sharp edges due to object boundaries give rise to connected edge segments at both fine and coarse scales. For these edges, the localization is best at fine scales, since an increased amount of smoothing mainly results in destructive shape distortions. At the finest scales, however, there is a large number of

other edge responses due to noise and other spurious structures. Moreover, the diffuse edge structures fail to give rise to connected edge curves at the finest scales. Hence, coarser scale levels are *necessary* to capture the shadow edges (as well as other edge diffuse edge structures). This effect is even more pronounced for the synthetic fractal image in the right column, for which it can be clearly seen how the edge detector responds to qualitatively different types of edge structures at different scales.

Figure 2 shows a further example of this behavior for an image of an arm, with and without 10% added white Gaussian noise. Here, it can be clearly seen that the sharp edge structures (the outline of the arm and the boundaries of the table) give rise to edge curve responses both at fine and coarse scales, whereas the diffuse edge structures (the shadow on the arm, the cast shadow on the table, and the reflection of the hand on the table) only give rise to connected edge curves only at coarser scales. When noise is added (the images in the right column), the diffuse edges are much more sensitive to these local perturbations than the sharp edges. Furthermore, if as much smoothing is applied to the cast shadow on the table as is necessary to capture the widest parts of its boundaries as closed edge curves, the shape distortions will be substantial near the fingertip. Hence, to capture this shadow with a reasonably trade-off between detection and localization properties, we must allow the amount of smoothing to vary along the edge.

In summary, these experiments illustrate how an edge detector responds to different types of edge structures in qualitatively different ways depending on the physical nature of the edge and the scale level at which the edge detector operates. A natural requirement on any edge detector intended to work robustly in a complex unknown environment is that it must be able to make these qualitatively different types of behaviors explicit for further processes. Whereas a straightforward approach could be to detect edges at multiple scales, and then send these as input to higher-level processing modules, such a representation would not only be highly redundant. It may also be very hard to interpret, since the representation of the edge behavior is only implicit. For this reason, and in view of the fact that the choice of scale levels crucially affects the performance of the edge detector, (and different scale levels will, in general, be required in different parts of the image), we argue that it is essential that the edge detection module is complemented by an *explicit*

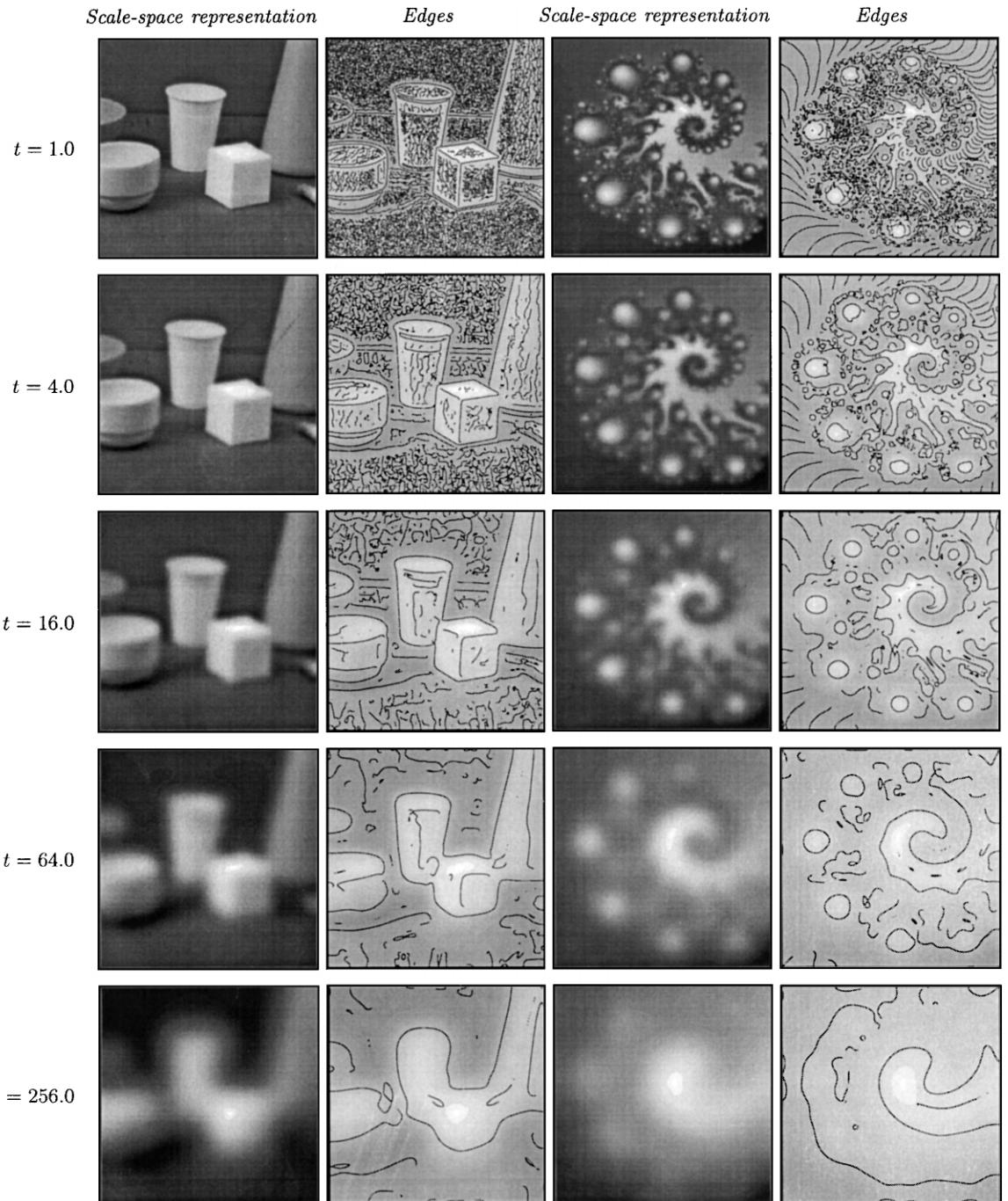


Figure 1. Edges computed at different scales from an image of an indoor office scene and a synthetic fractal image. Observe how different types of edge structures are extracted at different scales, and specifically how certain diffuse edge structures fail to give rise to connected edge curves at the finest levels of scale. (Image size: 256 * 256 and 182 * 182 pixels.)

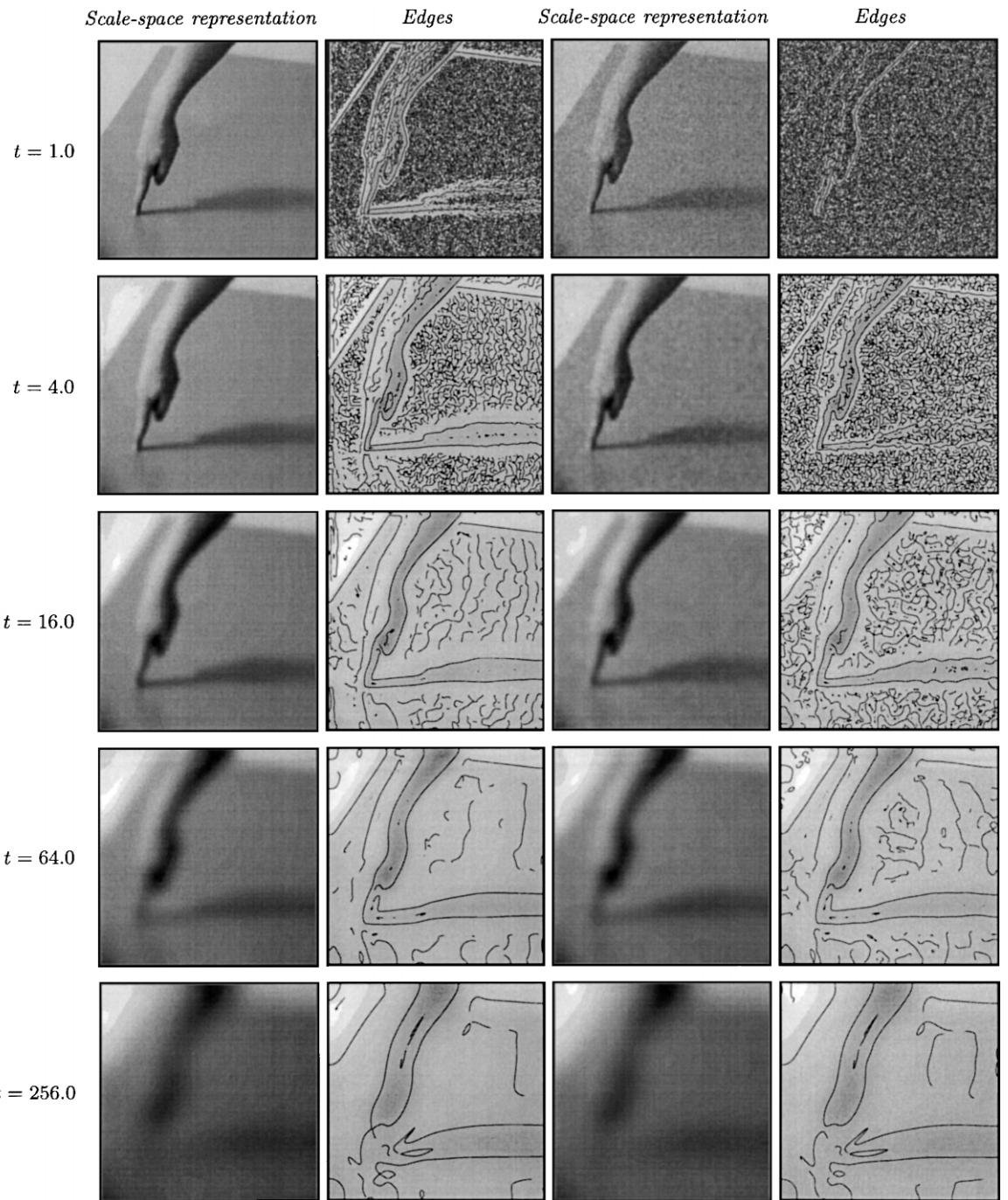


Figure 2. Edges computed at different scales from an image of arm. The left column shows results computed from the original image, and the right column corresponding results after adding 10% white Gaussian noise to the raw grey-level image. (Image size: 256 * 256 pixels.)

*mechanism which automatically adapts the scale levels to the local image structure.*⁵

The requirements on such a mechanism are (i) to output a more compact representation of the edge information than a raw multiscale edge representation, and (ii) to produce edge curves with an appropriate trade-off between detection and localization properties. Specifically, some of the desirable properties of this mechanism are to detect sharp edges at sufficiently fine scales so as to minimize the negative influence of the smoothing operation. For diffuse edges, the scale selection mechanism should select sufficiently coarse scales, such that a smooth edge model is a valid abstraction of the local intensity variations. Moreover, to capture edges with variable degree of diffuseness, it will in general be necessary to allow the scale levels to vary along the edge. The main subject of this article is to develop a framework for edge detection, in which the scale dimension is taken into account already the edge definition, and the abovementioned properties *follow as consequences* of the proposed definitions.

The presentation is organized as follows: Section 3 reviews some of the main results of scale-space theory as well as the main steps in a general methodology for automatic scale selection developed in (Lindeberg, 1993c, 1994a). In Section 4, this methodology is adapted to the edge detection problem. It is shown how the notion of nonmaximum suppression can be extended to the scale dimension, by maximizing a suitable measure of edge strength over scales. A theoretical analysis of the behavior of the resulting scale selection method is presented for a number of characteristic edge models. Experimental results are also presented for different types of real-world and synthetic images. Then, Section 5 shows how these ideas, in a slightly modified version, can be applied to the ridge detection problem, and be used for formulating a ridge detector with automatic scale selection. Section 6 explains the relations to previous works on these subjects, and Section 7 summarizes the main properties of the proposed approach. Finally, Section 8 outlines natural extensions of this methodology, and Section 9 concludes with the implications of this work with respect to the previously proposed scale selection principle.

3. Scale-Space Representation and Automatic Scale Selection: A Brief Review

A fundamental problem that arises whenever computing descriptors from image data concerns what image

operators to use. A systematic approach that has been developed for dealing with this problem is provided *scale-space theory* (Witkin, 1983; Koenderink, 1984; Yuille and Poggio, 1986; Lindeberg, 1994c; Florack et al., 1992). It focuses on the basic fact that image structures, like objects in the world, exist as meaningful entities over certain ranges of scale, and that one, in general, cannot expect to know in advance what scales are appropriate for describing those.

The essence of the results from scale-space theory is that if one assumes that the first stages of visual processing should be as uncommitted as possible, and have no particular bias, then convolution with Gaussian kernels and their derivatives of different widths is singled out as a canonical class of low-level operators. Hence, given any image $f : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$, its *scale-space representation* $L : \mathbb{R}^2 \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is defined by

$$L(\cdot; t) = g(\cdot; t) * f(\cdot) \quad (1)$$

where $g : \mathbb{R}^2 \times \mathbb{R}_+ \rightarrow \mathbb{R}$ denotes the Gaussian kernel given by

$$g(x; t) = \frac{1}{2\pi t} e^{-(x^2 + y^2)/(2t)}. \quad (2)$$

and t is the scale parameter. From this representation, scale-space derivatives are then defined by

$$L_{x^\alpha y^\beta}(\cdot; t) = \partial_{x^\alpha y^\beta} L(\cdot; t) = g_{x^\alpha y^\beta}(\cdot; t) * f(\cdot),$$

where (α, β) denotes the order of differentiation. The output from these operators can be used as a basis for expressing a large number of visual operations, such as feature detection, matching, and computation of shape cues. A particularly convenient framework for expressing these is in terms of multiscale differential invariants (Koenderink and van Doorn, 1992; Florack et al., 1992), or more specifically, as singularities (maxima or zero-crossings) of such entities (Lindeberg, 1993b).

3.1. Scale Selection from Local Maxima over Scales of γ -Normalized Derivatives

A basic problem that arises for any such feature detection method expressed within a multiscale framework concerns how to determine at what scales the image features should be extracted, or if the feature detection is performed at several scales simultaneously, what image features should be regarded as significant.

Early work addressing this problem for blob-like image structures was presented in (Lindeberg, 1991, 1993a). The basic idea was to study the behavior of image structures over scales, and to measure saliency from the stability of image structures across scales. Scale levels were, in turn, selected from the scales at which a certain normalized blob measure assumed local maxima over scales.

Then, in (Lindeberg, 1993c, 1994a) an extension was presented to other aspects of image structure. A general heuristic principle was proposed stating that *local maxima over scales* of (possibly nonlinear) combinations of *normalized derivatives*,

$$\partial_\xi = \sqrt{t} \partial_x, \quad (3)$$

serve as useful indicators reflecting the spatial extent of corresponding image structures. It was suggested that this idea could be used as a major guide for scale selection algorithms, which automatically adapt the local scale of processing to the local image structure. Specifically, it was proposed that feature detection algorithms with automatic scale selection could be formulated in this way. Integrated algorithms were presented, applying this idea to blob detection and corner detection with automatic scale selection, and early results were shown for the edge detection problem.

In this article, we shall develop more generally how this scale selection principle applies to the detection of one-dimensional image features, such as edges and ridges. For reasons that will become apparent later, we shall also extend the notion of normalized derivatives, and introduce a γ -parameterized *normalized derivative* by (Lindeberg, 1996a)

$$\partial_{x,\gamma\text{-norm}} = t^{\gamma/2} \partial_x. \quad (4)$$

When $\gamma = 1$, this definition is identical to the previous notion of normalized derivative in (3). As we shall see in Sections 4.5 and 5.6, however, the additional degree of freedom in the parameter γ will be essential when formulating scale selection mechanisms for detecting one-dimensional features such as edges and ridges.

4. Edge Detection with Automatic Scale Selection

In this section, we shall first review basic notions for expressing a differential geometric edge detector in terms of local directional derivatives. Then, we turn to the

formulation of a mechanism for automatic scale selection, and illustrate its performance by theoretical analysis and experiments.

4.1. Local Directional Derivatives

A convenient framework to use when computing image features from differential invariants, is to define the feature descriptors in terms of local directional derivatives relative to some preferred coordinate system. A natural way to construct such a system suitable for edge detection is as follows: At any point (x_0, y_0) in a two-dimensional image, introduce a local coordinate system (u, v) with the v -axis parallel to the gradient direction at (x_0, y_0) and the u -direction perpendicular, i.e.,

$$\begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} = \frac{1}{\sqrt{L_x^2 + L_y^2}} \begin{pmatrix} L_x \\ L_y \end{pmatrix} \Big|_{(x_0, y_0)}. \quad (5)$$

Directional derivatives in this local (u, v) -system are related to partial derivatives in the Cartesian coordinate system by

$$\partial_u = \sin \alpha \partial_x - \cos \alpha \partial_y, \quad (6)$$

$$\partial_v = \cos \alpha \partial_x + \sin \alpha \partial_y, \quad (7)$$

and the (u, v) -system is characterized by the fact that the first-order derivative L_u is zero.

4.2. Differential Geometric Edge Definition

According to the notion of nonmaximum suppression, an edge point is defined as a point at which the gradient magnitude assumes a maximum in the gradient direction (Canny, 1986; Korn, 1988). In terms of directional derivatives, such a definition can be expressed as the second-order directional derivative in the v -direction L_{vv} being zero, and the third-order directional derivative in the same direction L_{vvv} being negative:

$$\begin{aligned} L_{vv} &= 0, \\ L_{vvv} &< 0. \end{aligned} \quad (8)$$

Since only the sign information is important, this condition can be restated as⁶

$$\begin{aligned}\tilde{L}_{vv} &= L_v^2 L_{vv} \\ &= L_x^2 L_{xx} + 2L_x L_y L_{xy} + L_y^2 L_{yy} = 0 \\ \tilde{L}_{vvv} &= L_v^3 L_{vvv} \\ &= L_x^3 L_{xxx} + 3L_x^2 L_y L_{xxy} \\ &\quad + 3L_x L_y^2 L_{xyy} + L_y^3 L_{yyy} < 0.\end{aligned}\tag{9}$$

If we interpret L as the scale-space representation of an image f at a certain scale t , (9) defines the edges of f at that scale.

Here, we shall base the analysis on this definition, since (i) it is differential geometric and makes theoretical analysis possible, and (ii) it can be easily reformulated in a discrete setting and be implemented in terms of conceptually simple and general-purpose visual front-end operations (Lindeberg, 1993b).

4.3. Scale Selection: Selection of Edge Curves on the Edge Surface

If the edge definition (8) is applied at all scales in the scale-space representation of an image, the edge curves will sweep out a surface in scale-space. This surface will be referred to as the *edge surface in scale-space*.

In view of the scale selection principle reviewed in Section 3.1, a natural extension of the notion of non-maximum suppression is to define an edge as a curve on the edge surface in scale-space such that some suitably selected measure of edge strength is locally maximal with respect to scale on this curve. Thus, given such a (normalized) measure of edge strength $\mathcal{E}_{\text{norm}}L$, let us define a *scale-space edge* as the intersection of the edge surface in scale-space with the surface defined by $\mathcal{E}_{\text{norm}}L$ being maximal over scales. In differential geometric terms, this scale-space edge is thus defined as a connected set of points $\{(x, y; t) \in \mathbb{R}^2 \times \mathbb{R}_+\}$ (a curve Γ) which satisfies

$$\begin{aligned}\partial_t(\mathcal{E}_{\text{norm}}L(x, y; t)) &= 0, \\ \partial_{tt}(\mathcal{E}_{\text{norm}}L(x, y; t)) &< 0, \\ L_{vv}(x, y; t) &= 0, \\ L_{vvv}(x, y; t) &< 0.\end{aligned}\tag{10}$$

Of course, there are several possible ways of expressing the notion that $\mathcal{E}_{\text{norm}}L$ should assume local maxima over scales on the edge curve. In (9), this condition is

formulated as zero-crossings of the partial derivative of $\mathcal{E}_{\text{norm}}L$ with respect to the scale parameter (i.e., a directional derivative in the vertical scale direction). A natural alternative is to consider directional derivatives in a direction in the tangent plane to the edge surface, and to choose this direction as the steepest ascent direction of the scale parameter. In other words, let

$$\begin{aligned}N &= (N_x, N_y, N_t) \\ &= \nabla_{(x, y; t)}(\mathcal{E}_{\text{norm}}L)|_{P_0} \\ &= (\partial_x(\mathcal{E}_{\text{norm}}L), \partial_y(\mathcal{E}_{\text{norm}}L), \partial_t(\mathcal{E}_{\text{norm}}L))\end{aligned}\tag{11}$$

denote the (unnormalized) normal direction of the edge surface in scale-space, and define the (normalized) steepest ascent direction of the scale parameter by

$$T = (T_x, T_y, T_t) = \frac{(-N_x N_t, -N_y N_t, N_x^2 + N_y^2)}{\sqrt{(N_x^2 + N_y^2)(N_x^2 + N_y^2 + N_t^2)}}.$$

Then, with the directional derivative operator in the T -direction,

$$\partial_T = T_x \partial_x + T_y \partial_y + T_t \partial_t,\tag{12}$$

an alternative definition of a scale-space edge is as a connected set of points $\Gamma = \{(x, y; t) \in \mathbb{R}^2 \times \mathbb{R}_+\}$ that satisfies

$$\begin{aligned}\partial_T(\mathcal{E}_{\text{norm}}L(x, y; t)) &= 0, \\ \partial_{TT}(\mathcal{E}_{\text{norm}}L(x, y; t)) &< 0,\end{aligned}\tag{13}$$

$$\begin{aligned}L_{vv}(x, y; t) &= 0, \\ L_{vvv}(x, y; t) &< 0.\end{aligned}\tag{14}$$

The general scale selection method we propose for edges (and more generally also for other one-dimensional features), is to extract the features from the intersection between the feature surface in scale-space with the surface defined by a measure of (normalized) feature strength being locally maximal with respect to scale.

What remains to turn this idea into an operationally useful definition for detecting edges is to define the measure of edge strength. Based on the γ -parameterized normalized derivative concept introduced in (4), we shall here consider the following two γ -normalized differential entities:

$$\begin{aligned}\mathcal{G}_{\gamma\text{-norm}}L &= L_{v, \gamma\text{-norm}}^2 \\ &= t^\gamma(L_x^2 + L_y^2),\end{aligned}\tag{15}$$

$$\begin{aligned}\mathcal{T}_{\gamma\text{-norm}}L &= -\tilde{L}_{vvv,\gamma\text{-norm}} \\ &= -t^{3\gamma}(L_x^3 L_{xxx} + 3L_x^2 L_y L_{xxy} \\ &\quad + 3L_x L_y^2 L_{xyy} + L_y^3 L_{yyy}).\end{aligned}\quad (16)$$

The first entity, the gradient magnitude, is the presumably simplest measure of edge strength to think of in edge detection. (Here, we have squared it to simplify the differentiations to be performed in next section.). The second entity originates from the sign condition in the edge definition (9). As we shall see later, both these strength measures are useful in practice, but have slightly different properties.

4.4. Derivatives of Edge Strength with Respect to Scale

In the definition of scale-space edge, expressions such as $\partial_t(\mathcal{E}_{\text{norm}}L)$ and $\partial_{tt}(\mathcal{E}_{\text{norm}}L)$ occur. In this section, we shall present a methodology for rewriting such differential expressions involving derivatives of L with respect to the scale parameter t in terms of spatial derivatives only.

Since the scale-space representation L satisfies the diffusion equation, the spatial derivatives of L satisfy this equation as well. This implies that we can trade derivatives with respect to the scale parameter by spatial derivatives according to

$$\begin{aligned}\partial_t(L_{x^\alpha y^\beta}) &= \frac{1}{2} \nabla_{(x,y)}^2(L_{x^\alpha y^\beta}) \\ &= \frac{1}{2} (L_{x^{\alpha+2} y^\beta} + L_{x^\alpha y^{\beta+2}}).\end{aligned}\quad (17)$$

In Appendices A.2.1 and A.2.2, this idea is applied to the first- and second-order derivatives of $\mathcal{G}_{\gamma\text{-norm}}L$ and $\mathcal{T}_{\gamma\text{-norm}}L$ with respect to the scale parameter. Note that the differential expressions obtained contain spatial derivatives of L up to order three and five concerning $\mathcal{G}_{\gamma\text{-norm}}L$, and spatial derivatives of L up to order five and seven concerning $\mathcal{T}_{\gamma\text{-norm}}L$. Whereas one could raise a certain scepticism concerning the applicability of derivatives of such high order to real-world image data, we shall in Section 4.7 demonstrate that these descriptors can indeed be used for computing highly useful edge features, when integrated into a scale selection mechanism.

Before turning to experiments, however, let us first investigate the properties of the proposed scale selection method theoretically.

4.5. Closed-Form Theoretical Analysis for Characteristic Edge Models

To understand the consequences of using local maxima over scales of $\mathcal{G}_{\gamma\text{-norm}}L$ and $\mathcal{T}_{\gamma\text{-norm}}L$ for selecting scale levels for edge detection, we shall in this section analyze a set of characteristic model signals, for which a closed-form analysis is tractable.

4.5.1. Diffuse Gaussian Step Edge. Consider first a diffuse Gaussian step edge, defined as the primitive function of a Gaussian kernel,

$$\begin{aligned}f_{t_0}(x, y) &= \Phi(x; t_0) \\ &= \int_{x'=-\infty}^x g(x'; t_0) dx'.\end{aligned}\quad (18)$$

Here, g denotes the one-dimensional Gaussian kernel $g(x; t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}$, and t_0 represents the degree of diffuseness of the edge. From the semi-group property of the Gaussian kernel, it follows that the scale-space representation of this signal is

$$\begin{aligned}L(x, y; t) &= \Phi(x; t_0 + t) \\ &= \int_{x'=-\infty}^x g(x'; t_0 + t) dx'.\end{aligned}\quad (19)$$

On the edge, i.e., at $x = 0$, the γ -normalized first-order derivative is

$$\begin{aligned}t^{\gamma/2} L_x(0, y; t) &= t^{\gamma/2} g(x; t_0 + t)|_{x=0} \\ &= \frac{1}{\sqrt{2\pi}} \frac{t^{\gamma/2}}{(t_0 + t)^{1/2}},\end{aligned}\quad (20)$$

and the γ -normalized third-order derivative

$$\begin{aligned}t^{3\gamma/2} L_{xxx}(0, y; t) &= t^{\gamma/2} \frac{(x^2 - t - t_0)}{(t + t_0)^2} g(x; t_0 + t)\Big|_{x=0} \\ &= -\frac{1}{\sqrt{2\pi}} \frac{t^{3\gamma/2}}{(t_0 + t)^{3/2}}.\end{aligned}\quad (21)$$

Scale Selection Based on $\mathcal{G}_{\gamma\text{-norm}}L$. By differentiating the first-order measure of edge strength

$$\begin{aligned}\partial_t((\mathcal{G}_{\gamma\text{-norm}}L)(0, y; t)) &= \partial_t(t^\gamma L_x^2(0, y; t)) \\ &\sim \partial_t\left(\frac{t^\gamma}{t_0 + t}\right) \\ &= \frac{t^{\gamma-1}}{(t_0 + t)^2} (\gamma t_0 - (1 - \gamma)t),\end{aligned}\quad (22)$$

we see that when $0 < \gamma < 1$ this measure assumes a unique maximum⁷ over scales at

$$t_{\mathcal{G}_{\gamma\text{-norm}}} = \frac{\gamma}{1-\gamma} t_0. \quad (23)$$

Requiring the maximum to be assumed at $t_{\mathcal{G}_{\gamma\text{-norm}}} = t_0$ gives $\gamma = \frac{1}{2}$. If we then insert (23) into (19) we see that the edge strength measure at the selected scale is given by

$$L_{\frac{1}{2}}(0, y; t_{\mathcal{G}_{\gamma\text{-norm}}}) = \frac{1}{2\sqrt{\pi}} t_0^{-(1-\gamma)/2}. \quad (24)$$

Scale Selection Based on $\mathcal{T}_{\gamma\text{-norm}} L$. For the third-order measure of edge strength, the γ -normalized magnitude is proportional to the third power of $\mathcal{G}_{\gamma\text{-norm}} L$:

$$\begin{aligned} (\mathcal{G}_{\gamma\text{-norm}} L)(0, y; t) &= -t^{3\gamma} L_x^3(0, y; t) L_{xxx}(0, y; t) \\ &\sim \left(\frac{t^\gamma}{t_0 + t} \right)^3. \end{aligned} \quad (25)$$

Hence, for this edge model, the two scale selection methods have the same effect.

4.5.2. Analysis for a Gaussian Blob. Consider next the scale-space representation

$$L(x, y; t) = g(x, y; t_0 + t) \quad (26)$$

of a Gaussian blob $f(x, y) = g(x, y; t_0)$. For this signal, the edges are given by

$$L_{vv} = \frac{x^2 + y^2 - t_0 - t}{(t_0 + t)^2} g(x, y; t_0 + t) = 0, \quad (27)$$

i.e., they are located on a circle $x^2 + y^2 = t + t_0$, and the radius increases with scale.

Scale Selection Based on $\mathcal{G}_{\gamma\text{-norm}} L$. To analyse the behavior of the scale selection method on this curved edge model, let us insert L according to (26) into the expression for $\partial_t((\mathcal{G}_{\gamma\text{-norm}} L)(x, y; t))$ in (15). Then, straightforward calculations give

$$\begin{aligned} &\frac{\partial_t((\mathcal{G}_{\gamma\text{-norm}} L)(x, y; t))}{t^{\gamma-1} (x^2 + y^2)} \\ &= \frac{t(x^2 + y^2 - 4(t + t_0)) + \gamma(t + t_0)^2}{4\pi^2(t + t_0)^6 e^{(x^2+y^2)/(t+t_0)}}, \end{aligned}$$

and on the edge this expression reduces to

$$\begin{aligned} &\partial_t((\mathcal{G}_{\gamma\text{-norm}} L)(x, y; t))|_{L_{vv}=0} \\ &= \frac{t^{\gamma-1} (\gamma t_0 - (3 - \gamma)t)}{4e\pi^2(t + t_0)^4}. \end{aligned}$$

Hence, the selected scale level is

$$t_{\mathcal{G}_{\gamma\text{-norm}}} = \frac{\gamma}{3 - \gamma} t_0 = \left\{ \begin{array}{l} \text{if } \gamma = \frac{1}{2} \\ \frac{1}{5} t_0, \end{array} \right\} \quad (28)$$

which is significantly finer than the scale selected for a straight edge having the same degree of diffuseness.

Scale Selection Based on $\mathcal{T}_{\gamma\text{-norm}} L$. Similar insertion of L according to (26) into the expression for $\partial_t((\mathcal{T}_{\gamma\text{-norm}} L)(x, y; t))$ in (16), followed by a restriction to the edge, gives

$$\begin{aligned} &\partial_t((\mathcal{T}_{\gamma\text{-norm}} L)(x, y; t))|_{L_{vv}=0} \\ &= \frac{t^{3\gamma-1} (6\gamma t_0 - (13 - 6\gamma)t)}{16e^2\pi^4(t + t_0)^8}, \end{aligned}$$

and the selected scale level

$$t_{\mathcal{T}_{\gamma\text{-norm}}} = \frac{6\gamma}{13 - 6\gamma} t_0 = \left\{ \begin{array}{l} \text{if } \gamma = \frac{1}{2} \\ \frac{3}{10} t_0 \end{array} \right\}$$

is again much finer than for a straight edge having the same degree of diffuseness.

Approximate Maximization of Edge Strength over Scales. Since the curved edges in this edge model have a nonzero edge drift in scale-space, we can use this model for comparing scale selection based on zero-crossings of partial derivatives of the strength measures computed in the (vertical) scale direction with the alternative approach of computing such directional derivatives along the edge surface.

Insertion of (26) into the definition of $\mathcal{G}_{\gamma\text{-norm}} L$ in (15) gives the following expression for the first-order measure of edge strength in scale-space

$$(\mathcal{G}_{\gamma\text{-norm}} L)(x, y; t) = \frac{t^\gamma (x^2 + y^2)}{4\pi^2(t + t_0)^4} e^{-(x^2+y^2)/(2(t+t_0))},$$

which on the edge surface reduces to

$$(\mathcal{G}_{\gamma\text{-norm}} L)(x, y; t)|_{L_{vv}=0} = \frac{t^\gamma}{4e\pi^2(t + t_0)^3}$$

and assumes its maximum over scales at

$$t_{\mathcal{G}_{\gamma\text{-norm}},s} = \frac{\gamma}{3-\gamma} t_0 = \left\{ \begin{array}{l} \text{if } \gamma = \frac{1}{2} \\ \end{array} \right\} = \frac{1}{5} t_0. \quad (29)$$

For the third-order measure of edge strength, similar insertion of (26) into (16) gives

$$(\mathcal{T}_{\gamma\text{-norm}} L)(x, y; t)|_{L_{vv}=0} = \frac{t^{3\gamma}}{8e^2\pi^4(t+t_0)^7}$$

on the edge surface, and the maximum over scales is assumed at

$$t_{\mathcal{T}_{\gamma\text{-norm}},s} = \frac{3\gamma}{7-3\gamma} t_0 = \left\{ \begin{array}{l} \gamma = \frac{1}{2} \\ \end{array} \right\} = \frac{3}{11} t_0. \quad (30)$$

Hence, for this model, differentiation of $\mathcal{G}_{\gamma\text{-norm}}$ along the edge surface gives the same result as in (28), whereas there is a minor difference for $\mathcal{T}_{\gamma\text{-norm}}$ compared to Section 4.5.2.

4.5.3. Bifurcation. So far, we have considered isolated edges. In practice, however, interference effects between adjacent structures can often be as important. To illustrate the behavior of the scale selection mechanisms in such a situation, let us consider a double symmetric step edge, for which one of two edges disappears when the scale parameter becomes sufficiently large. This singularity corresponds to the annihilation of a maximum-minimum pair in gradient magnitude. A simple model of the local variations in gradient magnitude near this bifurcation can be obtained from the following polynomial, corresponding to a scale-space representation of the form $L(x; t) = \frac{1}{4}x^4 + \frac{3}{2}x^2(t - t_b) + \frac{3}{4}(t - t_b)^2$:

$$L_v(x; t) = L_x(x; t) = x^3 + 3x(t - t_b), \quad (31)$$

which represents the canonical type of singularity in a one-parameter family of functions, *the fold unfolding* (Poston and Stewart, 1978), and also has the attractive property that it satisfies the diffusion equation (here, t_b denotes the bifurcation scale). Differentiation gives

$$L_{vv}(x; t) = L_{xx}(x; t) = 3(x^2 + t - t_b) = 0,$$

and the position of the edge point as function of scale follows

$$x_{1,E}(t) = (t_b - t)^{1/2} \quad t \leq t_b. \quad (32)$$

Scale Selection Based on $\mathcal{G}_{\gamma\text{-norm}} L$. Insertion of $x_{1,E}(t)$ according to (32) into (31) gives that the variation over scales of the γ -normalized first-order strength measure on edge surface⁸ follows

$$(\mathcal{G}_{\gamma\text{-norm}} L)(x_{1,E}(t); t) = 4t^\gamma(t_b - t)^3, \quad (33)$$

with the (global) maximum over scales at $t_{\mathcal{G}_{\gamma\text{-norm}},\text{surf}} = \frac{\gamma}{3+\gamma} t_b = \left\{ \text{if } \gamma = \frac{1}{2} \right\} = \frac{1}{7} t_b$.

Scale Selection Based on $\mathcal{T}_{\gamma\text{-norm}} L$. Similar insertion of L_x and L_{xxx} into (16) shows that on the edge surface the third-order strength measure varies is

$$(\mathcal{T}_{\gamma\text{-norm}} L)(x_{1,E}(t); t) = 48t^{3\gamma}(t_b - t)^5, \quad (34)$$

with the maximum over scales at $t_{\mathcal{T}_{\gamma\text{-norm}},\text{surf}} = \frac{3\gamma}{5+3\gamma} t_b = \left\{ \text{if } \gamma = \frac{1}{2} \right\} = \frac{3}{13} t_b$.

Remarks. For this edge model, the first-order measure of edge strength results in the selection of a significantly finer scale than scale selection based on the third-order edge strength measure. In this context, however, a remark is necessary. Since the polynomial edge model constitutes a *local* approximation in a small neighborhood of the singularity, the global behavior may not be that relevant. The crucial point is instead the qualitative behavior at the singularity, which means that both the measures of edge strength decrease with scale when the scale parameter t approaches the bifurcation scale t_b . This property is important, since it prevents the scale selection mechanism from selecting scale levels near bifurcations.⁹

4.6. Measures of Edge Saliency

As is well-known, and as can be seen from the results in Fig. 2, edge detection based on nonmaximum suppression without thresholding, does, in general, lead to a large number of edge curves. Whereas the experiments in next section will show that the proposed scale selection scheme usually delivers a much smaller set of edge curves than fixed-scale edge detection performed over the same number of scales, there will nevertheless be a need for some selective mechanism for generating hypotheses about what edge curves should be regarded as significant. To reduce the number of edges, without introducing any global thresholds, such as (hysteresis) thresholds on gradient magnitude, we shall in this section introduce a measure of edge saliency to each

scale-space edge, which will be used for generating a (coarse) ranking on significance.

4.6.1. Integrated Edge Strength along the Curve. A straightforward way of constructing such a saliency measure in the context of the proposed edge detection framework is by integrating the measure of edge strength along the edge curve. Hence, for any connected edge curve Γ , and for each of the measures $\mathcal{G}_{\gamma\text{-norm}}L$ and $\mathcal{T}_{\gamma\text{-norm}}L$, we define these significance measures (based on $\gamma = 1$)¹⁰ by

$$G(\Gamma) = \int_{(x;t) \in \Gamma} \sqrt{(\mathcal{G}_{\gamma\text{-norm}}L)(x; t)} ds, \quad (35)$$

$$T(\Gamma) = \int_{(x;t) \in \Gamma} \sqrt[4]{(\mathcal{T}_{\gamma\text{-norm}}L)(x; t)} ds, \quad (36)$$

where the integration is performed over the projection of the edge curve onto the image plane, i.e., $ds^2 = dx^2 + dy^2$, and the edge strength measures have been transformed to be proportional to the image brightness at each point before integration.¹¹

Whereas this type of construction can be highly sensitive to spurious fragmentation,¹² it will be demonstrated in next section that significance values computed in this way are indeed be highly useful for suppressing spurious edge responses and serve as a help for selecting intuitively reasonable subsets of edge curves (see also Section 8.2).

4.7. Experiments

Let us now apply the proposed scale selection methodology and compute scale-space edges from different types of real-world and synthetic images. In brief, the edge features will be extracted by first computing the differential descriptors occurring in the definition in (9) at a number of scales in scale-space.¹³ Then, a polygon approximation is constructed of the intersections of the two zero-crossing surfaces of \tilde{L}_{vv} and $\partial_t(\mathcal{E}_{\gamma\text{-norm}})$ that satisfy the sign conditions $\tilde{L}_{vvv} < 0$ and $\partial_t(\mathcal{E}_{\gamma\text{-norm}}) < 0$. Finally, significance measures are computed according to (35) and (36), and the N edges having the strongest saliency measures been extracted. (A detailed description of the algorithm can be found in Appendix A.1.)

4.7.1. Scale Selection Based on the First-Order Edge Strength Measure.

Figure 3 shows the result of applying this edge detection scheme based on local max-

ima over scales of $\mathcal{G}_{\gamma\text{-norm}}L$ to four different images. The left middle column shows all the scale-space edges extracted based on the definition in (9), and the right column shows the 100 edge curves having the strongest saliency measures according to (35). As can be seen, a large number of edges is obtained, corresponding to object boundaries, shadow edges, as well as spurious structures in the smooth background.

For the arm image in the first row, we can observe that the sharp edges corresponding to the object boundaries are extracted, as well as the shadow edges on the arm, the cast shadow on the table, and the reflection on the table. In other words, the scale selection procedure leads to a qualitatively very reasonable extraction of edges with the scale levels adapted to the local image structure. (Recall from Fig. 2 that for this image it is impossible to capture the entire shadow edge at one scale without introducing severe shape distortions at the finger tip.)

The second and the third rows show corresponding results for an indoor office scene and an outdoor image of a statue. As can be seen, the major edges of the objects are extracted, as well as the occlusion shadow on the cylindrical object in the right part of the image. For the outdoor image, the outline of the statue is extracted, some parts of the shadows on the ground, the outline of the forest at the horizon, as well as an outline of the clouds in the sky.

The fourth row shows the result of applying the edge detection procedure to a fractal image. This example is interesting, since the image contains structures of different types and at a large number of different scales. As can be seen, the edge detector captures a large number of different features, and among the 100 strongest edges we find the boundaries of the bright blob structures and a subset of the boundary edges of highest contrast.

Of course, the number of edges selected for display is arbitrary, and in an integrated vision system, some mechanism is required for evaluating how many of the edges correspond to meaningful image structures in a given situation. We argue, however, that the significance values provide important information for making such decisions. Figure 4 illustrates this property, by showing the 50, 10 and 5 strongest edges, respectively, extracted from the arm image. As can be seen, the outlines of the arm, the table and the cast shadow are among the most significant edges.

Figure 5 gives a three-dimensional illustration of how the selected scale levels vary along the edges.

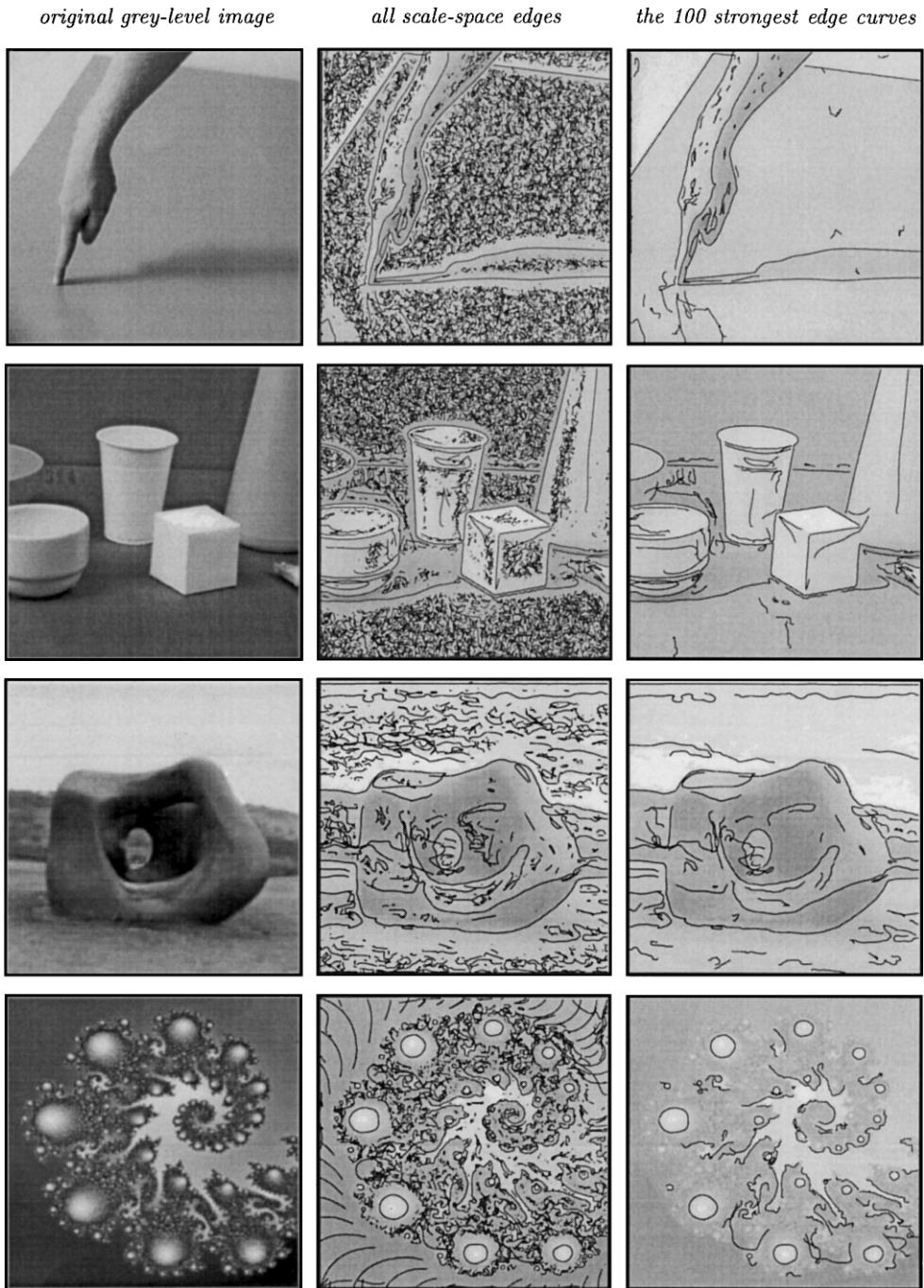


Figure 3. The result of edge detection with automatic scale selection based on local maxima over scales of the first-order edge strength measure $G_{\gamma\text{-norm}} L$ with $\gamma = \frac{1}{2}$. The middle column shows all the scale-space edges, whereas the right column shows the 100 edge curves having the highest significance values (according to (35)). Image size: 256×256 pixels for the two top-most images, 143×143 pixels for the statue image, and 182×182 pixels for the fractal image.

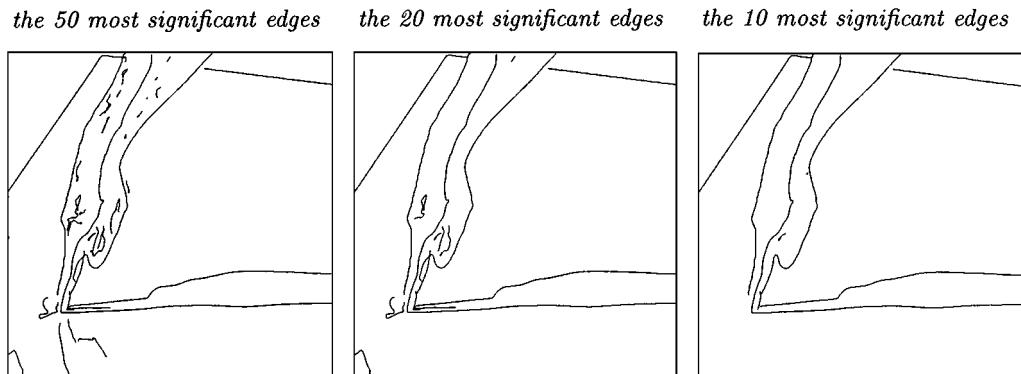


Figure 4. Illustration of the ranking on saliency obtained from the integrated γ -normalized gradient magnitude along the scale-space edges. Here, the 50, 20, and 10 most significant edges, respectively, have been selected from the arm image.

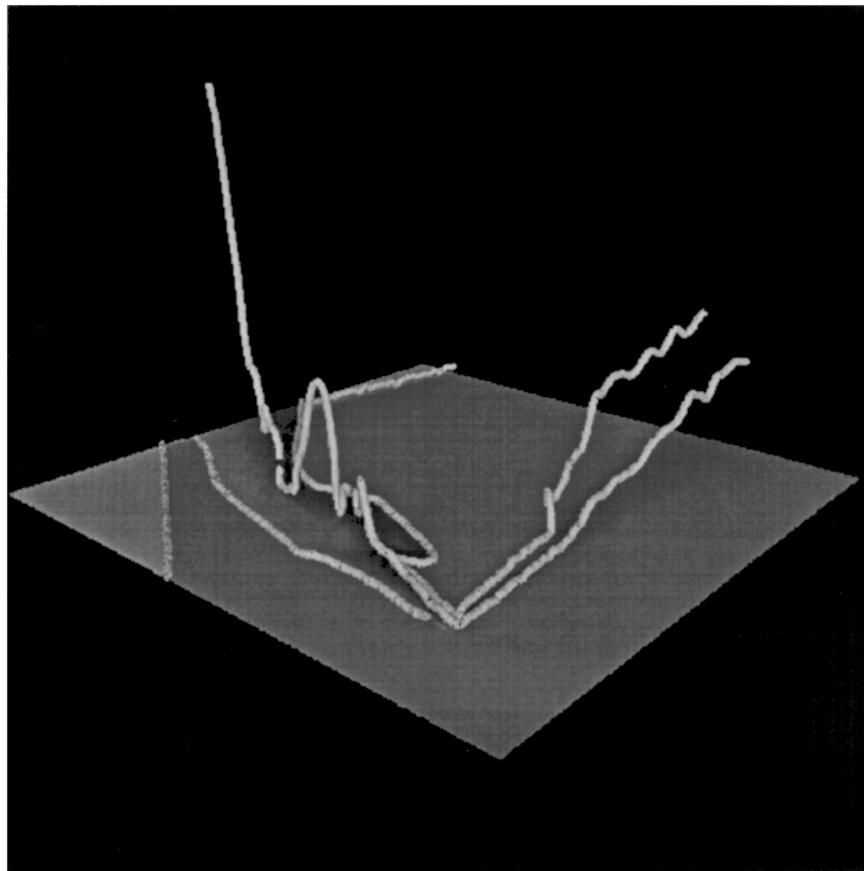


Figure 5. Three-dimensional view of the 10 most significant scale-space edges extracted from the arm image. From the vertical dimension representing the selected scale measured in dimension length (in units of \sqrt{t}), it can be seen how coarse scales are selected for the diffuse edge structures (due to illumination effects) and that finer scales are selected for the sharp edge structures (the object boundaries).

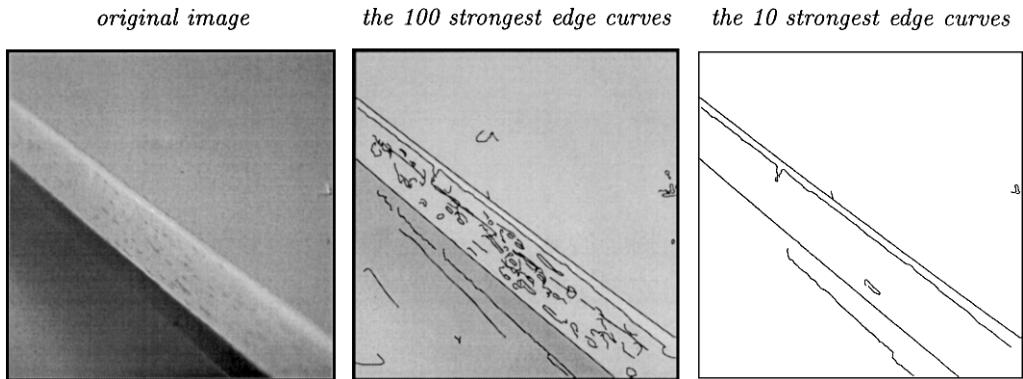


Figure 6. Corresponding results of applying the edge detection method with automatic scale selection to an image of a detail of a table (containing strong effects of focus blur). Here, the 100 and the 10 strongest edge responses, respectively, have been extracted.

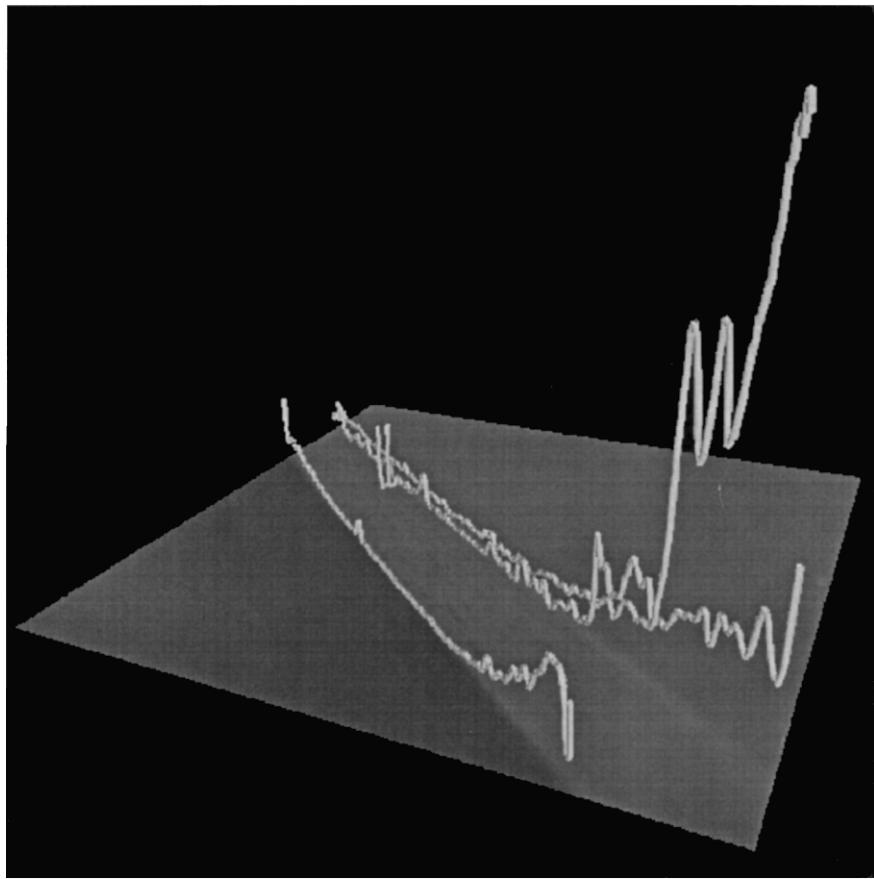


Figure 7. Three-dimensional view of the three strongest scale-space edges extracted from the image in Fig. 5 (showing a detail of a table registered with a narrow depth of field). Observe how the selected scale levels (graphically represented by the height of the curves over the image plane) reflect the variation in the amount of focus blur along the edge.

In this figure, the 10 most salient scale-space edges have been extracted from the arm image and visualized as one-dimensional curves embedded in the three-dimensional scale-space representation. These curves in scale-space have been overlayed on top of a low-contrast copy of the original grey-level image, and are seen from an oblique view with the height over the image plane representing the selected scale measured in dimension length (in units of \sqrt{t}). From this illustration, it can be seen how fine scales have been selected at the object boundaries, and that coarser scales are selected with increasing degree of diffuseness.

Figure 6 shows another illustration of how diffuseness estimates are obtained from this edge detector. It shows edges detected from an image of a detail of table, for which the effects of focus blur are strong. Note how the selected scale levels capture the varying degree of diffuseness along the edges (see Fig. 7).

4.7.2. Scale Selection Based on the Third-Order Edge Strength Measure. Figure 8 shows corresponding results of edge detection with scale selection based on local maxima over scales of $T_{\gamma,\text{norm}}L$. To a first approximation, the results are qualitatively similar. At the more detailed level, however, we can observe that the performance is slightly better in the respect that more responses are obtained for the shadow edges in the indoor office image and for the outdoor statue image. An intuitive explanation of why the edge strength measures differ in this respect, is that the third-order derivative operator has more narrow response properties to edges. Therefore, the magnitude of this response will start to decrease earlier with scale, when interference effects between neighboring edges affect the edge responses.

Figure 9 shows the performance of this scale selection method when applied to two images containing a large amount of fine-scale structures. From a first view, these results may look very similar to the result of traditional edge detection at a fixed (very fine) scale. A more detailed study, however, reveals that a number of shadow edges are extracted, which would be impossible to detect at the same scale as the dominant fine-scale information. In this context, it should be noted that this detection of edges at very fine scales is not the result of any manual setting of tuning parameters. It is a direct consequence of the definition of the scale-space edge concept, and is the result of applying the same mechanism as selects coarse scale levels for diffuse image structures.

4.8. Summary

To conclude,¹⁴ for both these measures of edge strength, the proposed scale selection scheme has the desirable property of adapting the scale levels to the local image structure such that *the selected scales reflect the degree of diffuseness of the edge*.

5. Ridge Detection with Automatic Scale Selection

In most current feature based computer vision algorithms, edges are used as the main type of image features. This historical heritage should, however, not exclude the use of other features. For example, blob descriptors can deliver important hypotheses about the existence of objects, signalling that “there might be something there of about that size—now some other processing module could take a closer look” (Lindeberg, 1993a). A ridge feature can be seen as a refined version of such a descriptor, which in addition provides an approximate symmetry axis of the candidate object. Psychophysical support for this idea have been presented by Burbeck and Pizer (1995).

When to define ridges from intensity data, there are several possible approaches. In topography, a ridge is defined as a separator between regions from which water flows in different directions (to different sinks). The precise mathematical formulation of this property has, however, lead to confusions. A historic account of this development is given by Koenderink and van Doorn (1994). In computer vision, early approaches to ridge detection were proposed by Haralick (1983), who defined bright (dark) ridges as points for which the main eigenvalue of the Hessian matrix assumes a maximum (minimum) in the main eigendirection of the Hessian matrix, and by Crowley and Parker (1984), who considered directional maxima in bandpass filtered images.

During more recent years, the ridge detection problem has been intensively studied by Pizer and his co-workers (Pizer et al., 1994). Gauch and Pizer (1993) define ridges from topographical watersheds computed in a scale-space representation of the image data. Morse et al. (1994) compute “core” descriptors in a multiscale fashion by propagating a measure of edge strength from each edge point and then detecting peaks in the measure of “medialness” so obtained. A more extensive discussion of different types of ridge detectors is presented by Eberly et al. (1994), including extensions to higher dimensions. Related applications of

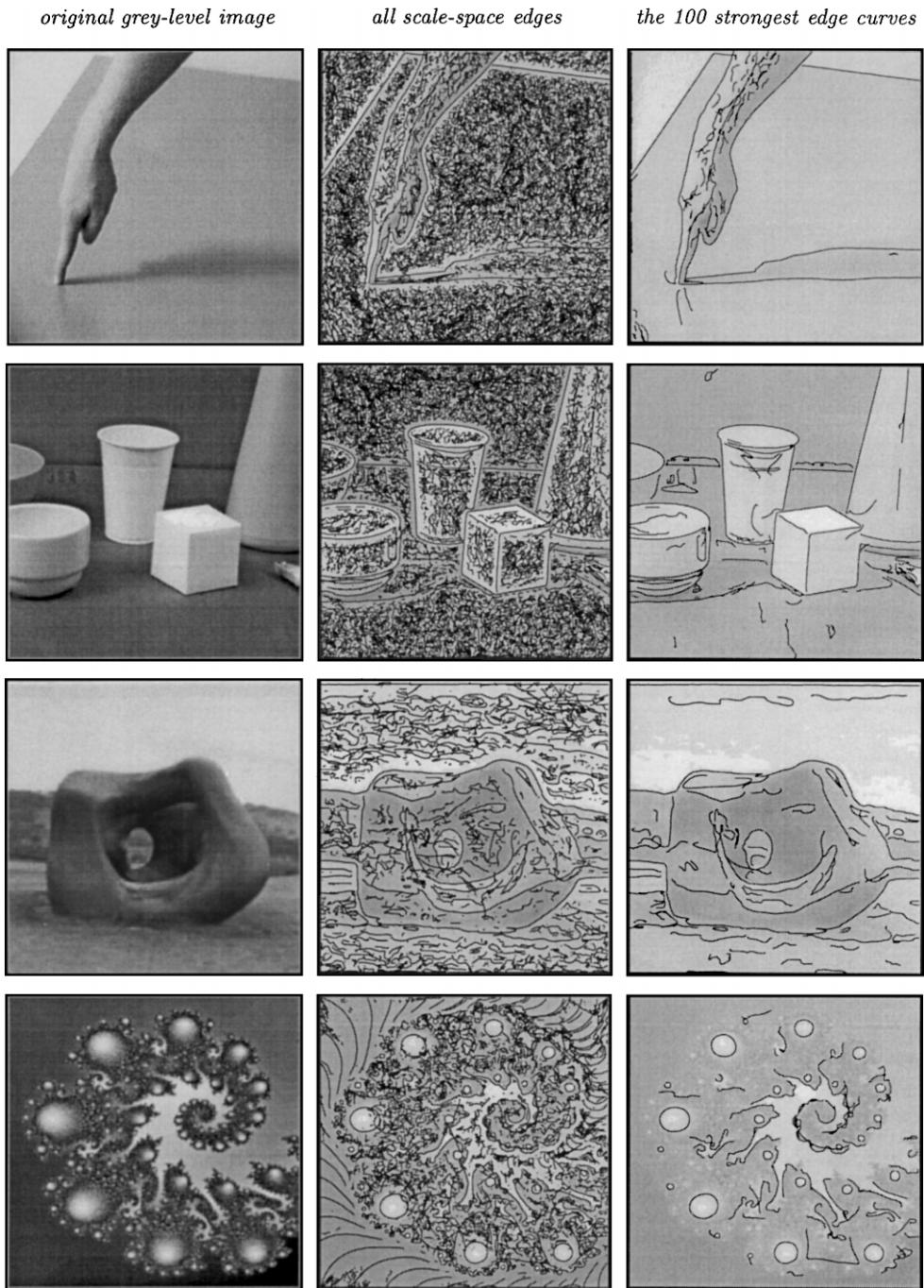


Figure 8. The result of edge detection with automatic scale selection based on local maxima over scales of the third-order edge strength measure $T_{\gamma\text{-norm}} L$ with $\gamma = \frac{1}{2}$. The middle column shows all the scale-space edges, whereas the right column shows the 100 edge curves having the highest significance values (according to (36)). Image size: 256 × 256 pixels for the two top-most images, 143 × 143 pixels for the statue image, and 182 × 182 pixels for the fractal image.

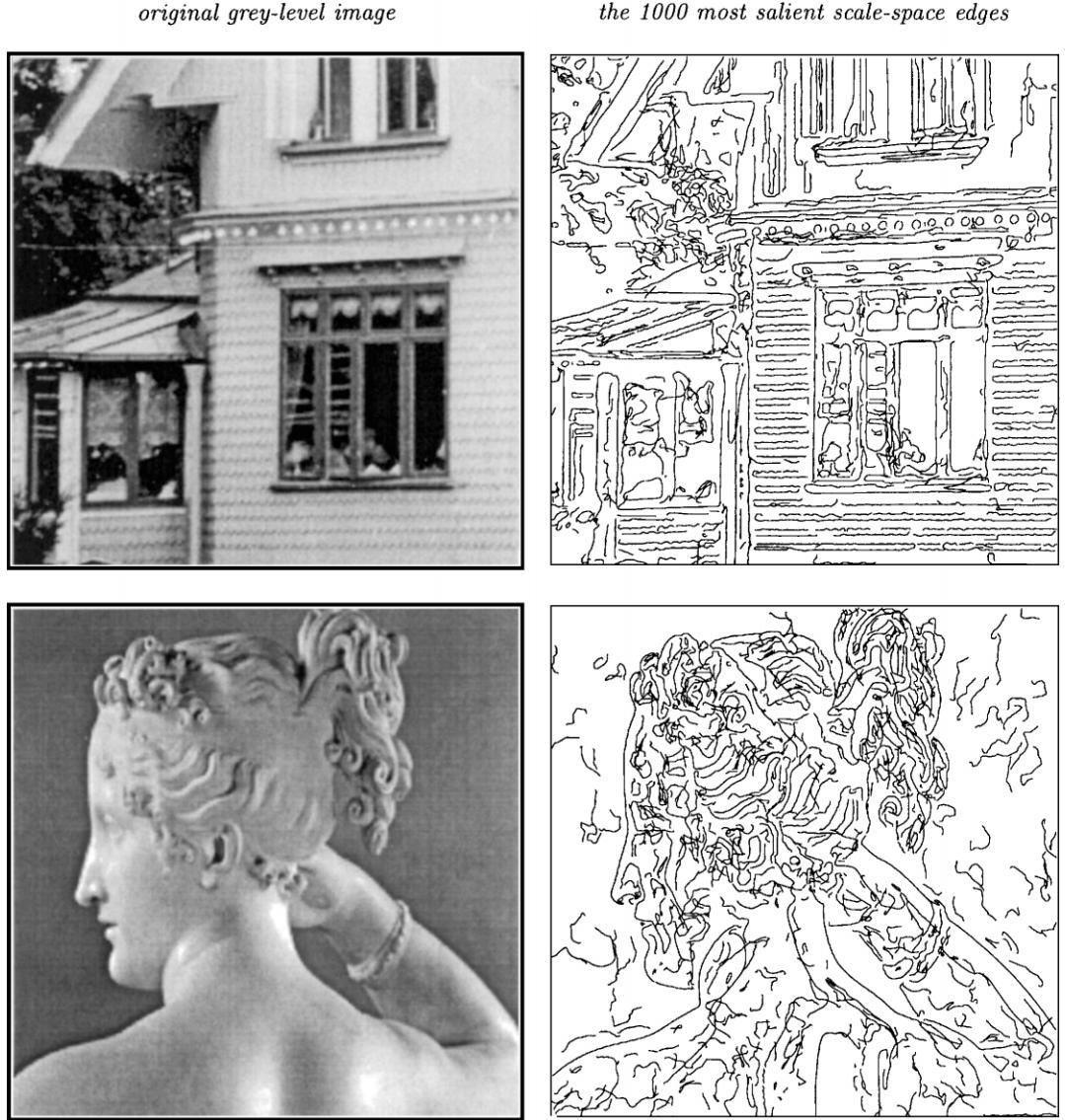


Figure 9. The result of edge detection with automatic scale selection using local maxima over scales of $\mathcal{T}_{\gamma\text{-norm}}L$ with $\gamma = \frac{1}{2}$ for two images containing a large number of fine-scale details. Observe how well the fine-scale structures are captured and resolved. (Image size: 256×256 pixels for the Godthem Inn image and 240×240 for the Paolina image.)

similar ideas to medical images have been presented by Griffin et al. (1992), Monga et al. (1994), and Koller et al. (1995).

For binary data, the related notion of “skeletons” can be derived from the medial axis (Blum and Nagel, 1978) and be computed by (grass-fire-like) distance transforms (Arcelli and Baja, 1992). It is, however, well-known that features extracted in this way can be highly sensitive to small perturbations of the boundary. To reduce these problems, Ogniewicz and Kübler

(1995) proposed a hierarchical skeleton concept. Alternatively, this sensitivity can be reduced by grey-level based multiscale techniques.

In this section, we shall show how the framework for edge detection developed in previous section, with just minor modifications, can be used for formulating a ridge detector with automatic scale selection. In analogy with the treatment in Section 4, we shall first express a differential geometric ridge detector in terms of local directional derivatives at a fixed scale in

scale-space. Then, we turn to the problem of including a mechanism for automatic scale selection.

5.1. Local Directional Derivatives

At any image point (x_0, y_0) , introduce a local (p, q) -system aligned to the eigendirections of the Hessian matrix of the brightness function. To express directional derivatives in these coordinates, which are characterized by the mixed second-order derivative being zero, $L_{pq} = 0$, we can rotate the coordinate system by an angle β defined by

$$\cos \beta|_{(x_0, y_0)} = \sqrt{\frac{1}{2} \left(1 + \frac{L_{xx} - L_{yy}}{\sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}} \right)} \quad (37)$$

$$\begin{aligned} \sin \beta|_{(x_0, y_0)} &= (\text{sign } L_{xy}) \\ &\times \sqrt{\frac{1}{2} \left(1 - \frac{L_{xx} - L_{yy}}{\sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}} \right)} \quad (38) \end{aligned}$$

and define unit vectors in the p - and q -directions by $e_p = (\sin \beta, -\cos \beta)$ and $e_q = (\cos \beta, \sin \beta)$ with associated directional derivative operators

$$\partial_p = \sin \beta \partial_x - \cos \beta \partial_y, \quad (39)$$

$$\partial_q = \cos \beta \partial_x + \sin \beta \partial_y. \quad (40)$$

Then, it is straightforward to verify that this definition implies that

$$\begin{aligned} L_{pq} &= \partial_p \partial_q L \\ &= (\cos \beta \partial_x + \sin \beta \partial_y)(\sin \beta \partial_x - \cos \beta \partial_y)L \\ &= \cos \beta \sin \beta (L_{xx} - L_{yy}) \\ &\quad - (\cos^2 \beta - \sin^2 \beta)L_{xy} = 0. \quad (41) \end{aligned}$$

5.2. Differential Geometric Ridge Definition

As mentioned in the introduction to this section, there are several ways to define ridges from intensity data. A natural way to formulate a ridge concept in terms of

local differential geometric properties of image brightness is by defining a bright (dark) ridge as a connected set of points for which the intensity assumes a local maximum (minimum) in the main eigendirection of the Hessian matrix. When expressed in the (p, q) -system, this requirement for point to be a bright ridge point can be written

$$\begin{aligned} L_p &= 0, & L_q &= 0, \\ L_{pp} &< 0, & \text{or} & L_{qq} < 0, \\ |L_{pp}| &\geq |L_{qq}|, & |L_{qq}| &\geq |L_{pp}|, \end{aligned} \quad (42)$$

depending on whether the p - or the q -direction corresponds to the maximum absolute eigenvalue of the Hessian matrix. This idea, which goes back to (Saint-Venant, 1852), is closely related to the approaches in (Haralick, 1983; Eberly et al., 1994).¹⁵

In (Lindeberg, 1994b) it is shown that in terms of the (u, v) -system described in Section 4.1, this condition can for nondegenerate L equivalently be written

$$\begin{aligned} L_{uv} &= 0, \\ L_{uu}^2 - L_{vv}^2 &> 0, \end{aligned} \quad (43)$$

where the sign of L_{uu} determines the polarity; $L_{uu} < 0$ corresponds to bright ridges, and $L_{uu} > 0$ to dark ridges.

5.3. The Need for Automatic Scale Selection in Ridge Detection

Figure 10 shows the result of computing ridges defined in this way at different scales in scale-space for an aerial image of a suburb and an image of an arm, respectively. Observe how different types of ridge structures give rise to ridge curves at different scales. For example, the main roads in the aerial image appear at $t \approx 64$, the fingers give rise to ridge curves at $t \approx 16$, and the arm as a whole is extracted as a long ridge curve at $t \approx 256$. Moreover, note that these ridge descriptors are much more sensitive to the choice of the scale levels than the edge features in Fig. 2. In particular, no single scale is appropriate for extracting ridges over the entire image. Hence, a mechanism for automatic scale selection is necessary in order to compute these ridge descriptors from image data about which no a priori information is available.

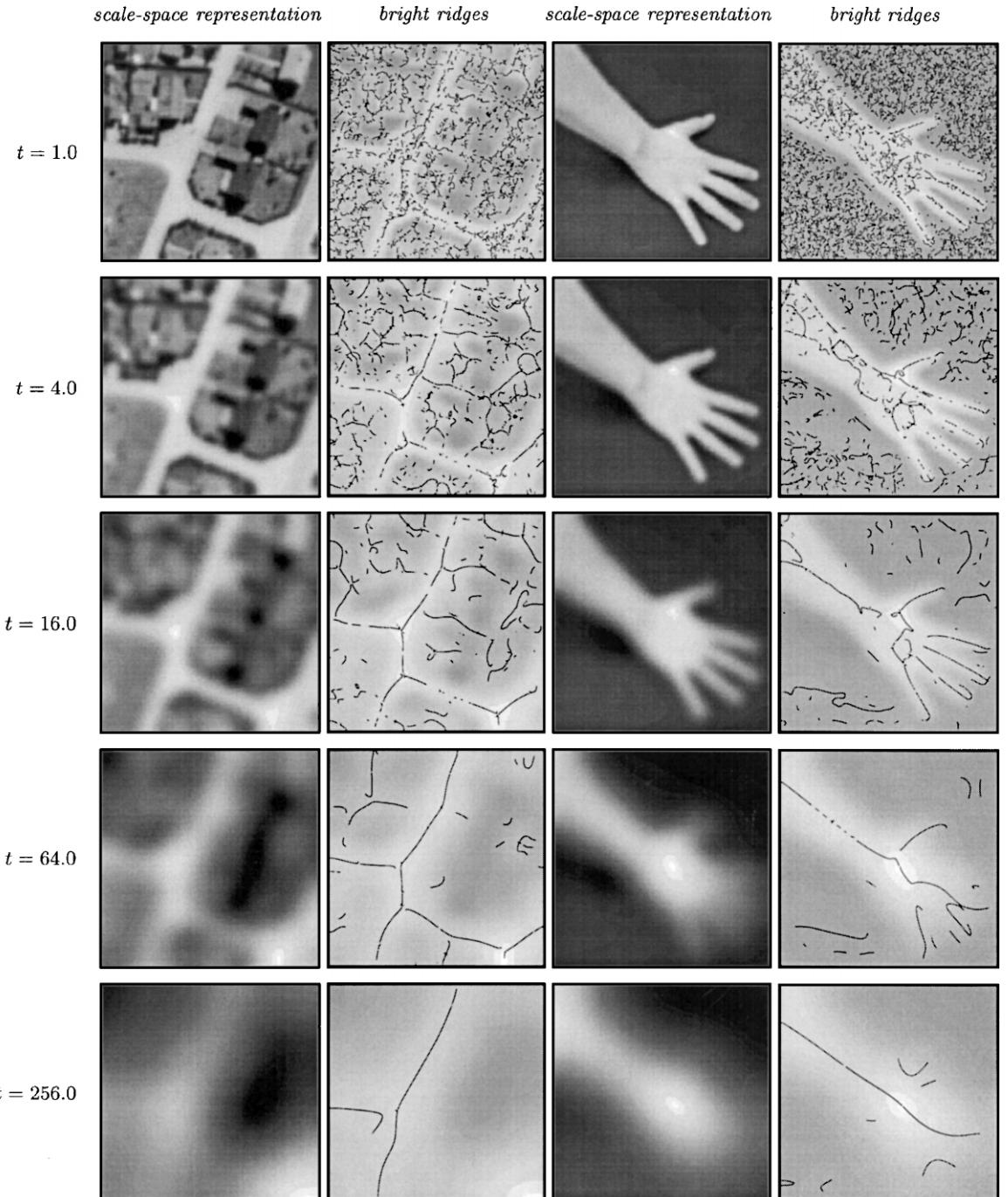


Figure 10. Ridges computed at different scales for an aerial image and an image of a hand using the ridge definition in (43). Notably, different types of image structures give rise to different ridge curves at different scales. In particular, no single scale is appropriate for capturing all major ridges. (Image size: 128 * 128 and 140 * 140 pixels.)

5.4. Scale Selection: Selection of Ridge Curves on the Ridge Surface

If the ridge definition (42) is applied at all scales in scale-space, it will sweep out a surface in scale-space. This surface will be referred to as the *ridge surface* in scale-space. To formulate a scale selection method for ridge detection, let us assume that we can associate a normalized measure of ridge strength $\mathcal{R}_{\text{norm}}L$ to each point in scale-space. Then, in analogy with Section 4.3, we can define a *scale-space ridge* as the intersection of the ridge surface with the surface defined by $\mathcal{R}_{\text{norm}}L$ being locally maximal over scales.

Assume, for simplicity, that we in some region can locally rename¹⁶ the p - and q -directions such that the p -direction corresponds to the maximum value of the eigenvalues of the Hessian matrix at each point. Then, a scale-space ridge is defined as a connected set of points $\Gamma = \{(x, y; t) \in \mathbb{R}^2 \times \mathbb{R}_+\}$ that satisfies

$$\begin{aligned}\partial_t(\mathcal{R}_{\text{norm}}L(x, y; t)) &= 0, \\ \partial_{tt}(\mathcal{R}_{\text{norm}}L(x, y; t)) &< 0, \\ L_p(x, y; t) &= 0, \\ L_{pp}(x, y; t) &< 0.\end{aligned}\tag{44}$$

Alternatively, we can consider directional derivatives of $\mathcal{R}_{\text{norm}}L$ computed in the tangent plane of the ridge surface, in analogy with the definition in Eq. (13).

What remains to turn this definition into an operational method for detecting ridges, is to define the measure of ridge strength. In the following sections, we will consider the consequences of using three such strength measures.

5.5. Measures of Ridge Strength

Given the ridge definition in (42), the presumably first choice to consider as measure of ridge strength is the maximum absolute eigenvalue of the Hessian matrix

$$\mathcal{ML} = \max(|L_{pp}|, |L_{qq}|).\tag{45}$$

If we again introduce normalized derivatives parameterized by a parameter γ such that $\partial_\xi = t^{\gamma/2}\partial_x$, we obtain the *γ -normalized maximum eigenvalue of the Hessian*

$$\begin{aligned}\mathcal{M}_{\gamma\text{-norm}}L &= \max(|L_{pp,\gamma\text{-norm}}|, |L_{qq,\gamma\text{-norm}}|) \\ &= t^\gamma \max(|L_{pp}|, |L_{qq}|),\end{aligned}\tag{46}$$

where the explicit expressions for $L_{pp,\gamma\text{-norm}}$ and $L_{qq,\gamma\text{-norm}}$ are

$$L_{pp,\gamma\text{-norm}} = \frac{t^\gamma}{2} \left(L_{xx} + L_{yy} - \sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2} \right),\tag{47}$$

$$L_{qq,\gamma\text{-norm}} = \frac{t^\gamma}{2} \left(L_{xx} + L_{yy} + \sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2} \right).\tag{48}$$

A negative property of this entity, however, is that it is not specific to ridge-like structures, and gives strong responses to other image structures, such as blobs. (Consider, for example, the behavior when $L_{xx} = L_{yy}$ and $L_{xy} = 0$.) For this reason, we shall also consider the following differential expression, which originates from the alternative formulation of the ridge definition in (43). This ridge strength measure will be referred to as the *square of the γ -normalized square eigenvalue difference*

$$\mathcal{N}_{\gamma\text{-norm}}L = (L_{pp,\gamma\text{-norm}}^2 - L_{qq,\gamma\text{-norm}}^2)^2.\tag{49}$$

In contrast to $\mathcal{M}_{\gamma\text{-norm}}L$, this entity assumes large values only when the eigenvalues of the Hessian matrix are significantly different, i.e., for elongated structures. Moreover, there is no logical “or” operation in its differential expression in terms of spatial derivatives

$$\begin{aligned}\mathcal{N}_{\gamma\text{-norm}}L &= ((L_{pp,\gamma\text{-norm}} + L_{qq,\gamma\text{-norm}}) \\ &\quad \times (L_{pp,\gamma\text{-norm}} - L_{qq,\gamma\text{-norm}}))^2 \\ &= t^{4\gamma} (L_{xx} + L_{yy})^2 \\ &\quad \times ((L_{xx} - L_{yy})^2 + 4L_{xy}^2).\end{aligned}\tag{50}$$

If we want to have a ridge strength measure that completely suppresses the influence of the Laplacian blob response $(\nabla^2 L)^2 = (L_{xx} + L_{yy})^2$, a natural third alternative to consider is the *square of the γ -normalized eigenvalue difference*.

$$\begin{aligned}\mathcal{A}_{\gamma\text{-norm}}L &= (L_{pp,\gamma\text{-norm}} - L_{qq,\gamma\text{-norm}})^2 \\ &= t^{2\gamma} ((L_{xx} - L_{yy})^2 + 4L_{xy}^2).\end{aligned}\tag{51}$$

In Appendix A.3, explicit expressions are derived for the first- and second-order derivatives of these ridge strength measures with respect to the scale parameter.

5.6. Qualitative Properties of Different Ridge Strength Measures

Concerning the qualitative behavior of ridge detectors based on these ridge strength measures, we can first make the general observation that the behavior is the same for cylindric image patterns, i.e., image patterns of the form $f(x, y) = h(ax + by + c)$ for some $h: \mathbb{R} \rightarrow \mathbb{R}$ and some constants $a, b, c \in \mathbb{R}$.¹⁷ For image structures without such symmetry, however, the qualitative behavior may be different.

5.6.1. Cylindrical Gaussian Ridge. To study the behavior for a cylindric ridge in more detail, consider a one-dimensional Gaussian blob with variance t_0 , extended cylindrically in the perpendicular direction

$$f(x, y) = g(x; t_0), \quad (52)$$

where g here denotes the one-dimensional Gaussian kernel

$$g(x; t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}. \quad (53)$$

From the semi-group property of Gaussian smoothing, it follows that the scale-space representation of f is given by

$$L(x, y; t) = g(x; t_0 + t). \quad (54)$$

Here, the ridge coincides with the y -axis, and on this ridge we have

$$\begin{aligned} (\mathcal{M}_{\gamma\text{-norm}} L)(0, y; t) &= t^\gamma |g_{xx}(0, y; t)| \\ &= \frac{1}{\sqrt{2\pi}} \frac{t^\gamma}{(t_0 + t)^{3/2}}. \end{aligned} \quad (55)$$

Differentiation with respect to t gives

$$\begin{aligned} \partial_t (\mathcal{M}_{\gamma\text{-norm}} L)(0, y; t) \\ = \frac{1}{2\sqrt{2\pi}} \frac{t^{\gamma-1}(2\gamma(t+t_0)-3t)}{(t_0+t)^{5/2}}, \end{aligned}$$

and setting this derivative $\partial_t (\mathcal{M}_{\gamma\text{-norm}} L)(0, y; t)$ to zero gives

$$t_{\mathcal{M}_{\gamma\text{-norm}}} = \frac{2\gamma}{3-2\gamma} t_0. \quad (56)$$

Clearly, $0 < \gamma < \frac{3}{2}$ is a necessary condition to give a local maximum of the ridge strength measure over scales. Moreover, $\gamma = 1$ corresponds to $t_{\mathcal{M}_{\gamma\text{-norm}}} = 2t_0$. If we want the selected scale level to reflect the width of the ridge such that $t_{\mathcal{M}_{\gamma\text{-norm}}} = t_0$, then we should select $\gamma = \frac{3}{4}$.

5.6.2. Simulation Experiments. Because of the complexity of the differential expressions for these ridge strength measures, it is hard to find representative ridge models that allow for compact closed-form analysis of curved ridges. For this reason, let us instead illustrate the qualitative behavior by simulations.

Figures 11 and 12 show the result of computing the three ridge strength measures, $\mathcal{M}_{\gamma\text{-norm}} L$, $\mathcal{N}_{\gamma\text{-norm}} L$ and $\mathcal{A}_{\gamma\text{-norm}} L$, at different scales for an aerial image of a suburb and an image of a hand, respectively. For all these descriptors different types of image structures give rise to different types of responses at different scales. Specifically, strong responses are obtained when the standard deviation of the Gaussian kernel is approximately equal to the width of the ridge structure. (Observe that the fixed-scale ridge detector in Fig. 10 extracts nice ridge curves at these scales.)

It can also be seen that the ridge descriptors have qualitative different behaviors in the way they respond to curved ridge structures, ridges of finite length, and to the extent they generate spurious responses at blob structures and edge structures. Notably, $\mathcal{M}_{\gamma\text{-norm}} L$ and $\mathcal{A}_{\gamma\text{-norm}} L$ give strong responses at edges, and $\mathcal{M}_{\gamma\text{-norm}} L$ also comparably strong blob responses. In this respect, $\mathcal{N}_{\gamma\text{-norm}} L$ appears to have the most ridge-specific response properties of these descriptors.

5.6.3. Gaussian Ridge of Delimited Extent. A preference to the ridge strength measure $\mathcal{N}_{\gamma\text{-norm}} L$ can also be obtained by studying the response properties to a ridge modeled as an elongated Gaussian blob of delimited extent

$$f(x, y) = g(x; t_1) g(y; t_2) \quad (57)$$

where g again denotes the one-dimensional Gaussian kernel. The scale-space representation of this signal is $L(x, y; t) = g(x; t_1 + t) g(y; t_2 + t)$ and we assume that $t_1 > t_2$ such that the ridge coincides with the x -axis.

Figure 13 shows the result of computing $\partial_t (\mathcal{N}_{\gamma\text{-norm}} L)$ and $\partial_t (\mathcal{A}_{\gamma\text{-norm}} L)$ as function of scale t in a horizontal cross-section of this ridge model. As can be seen, for scale selection based on $\mathcal{N}_{\gamma\text{-norm}} L$,

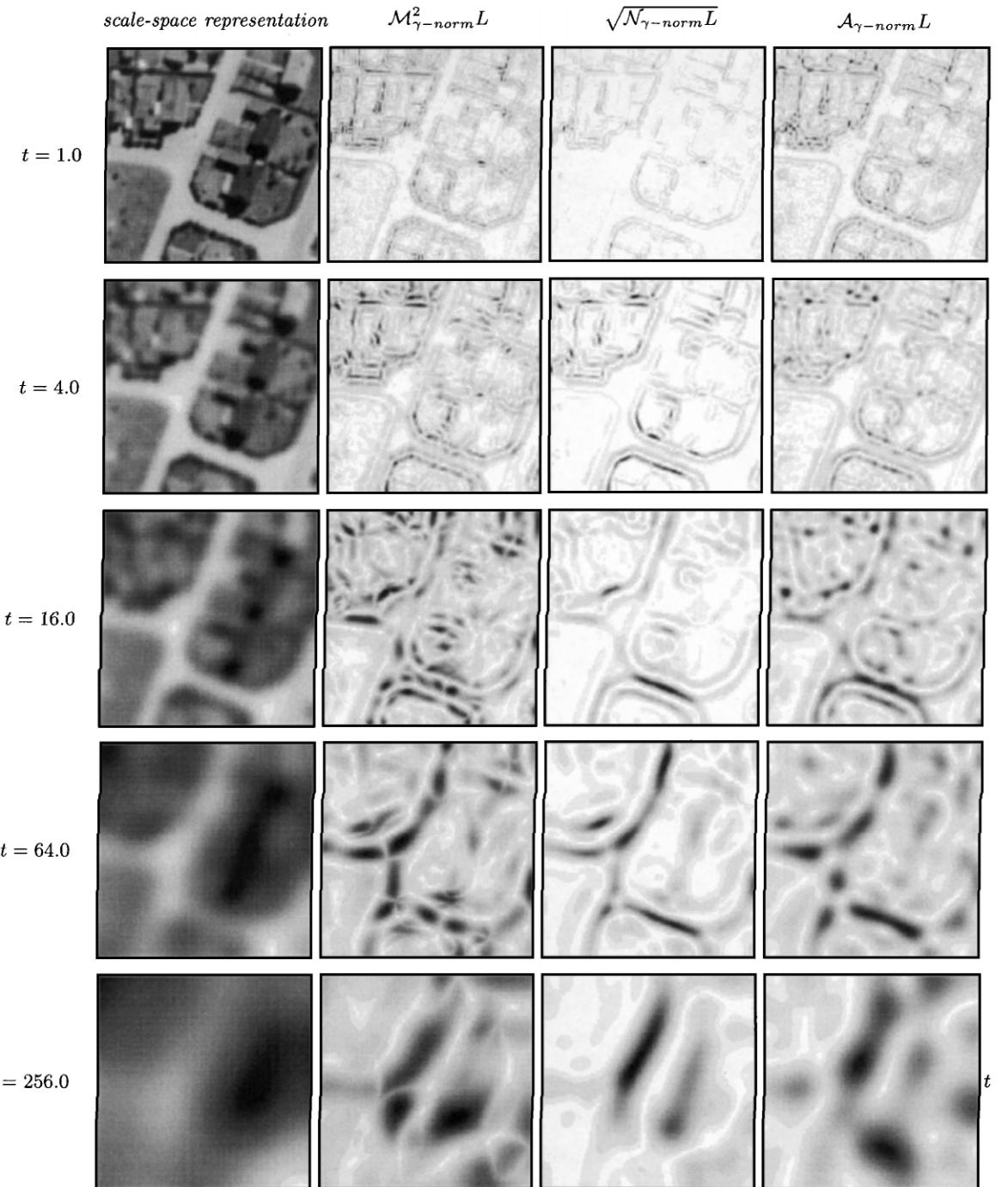


Figure 11. Ridge strength measures computed different scales for an aerial image of a suburb. Notably, different types of ridge structures give rise to strong responses at different scales. Moreover, there are qualitative differences in the response properties of the ridge descriptors to curved ridges and ridges of finite extent. The ridge strength measures also differ in terms of the extent to which they give spurious responses to blob-like and edge-like structures.

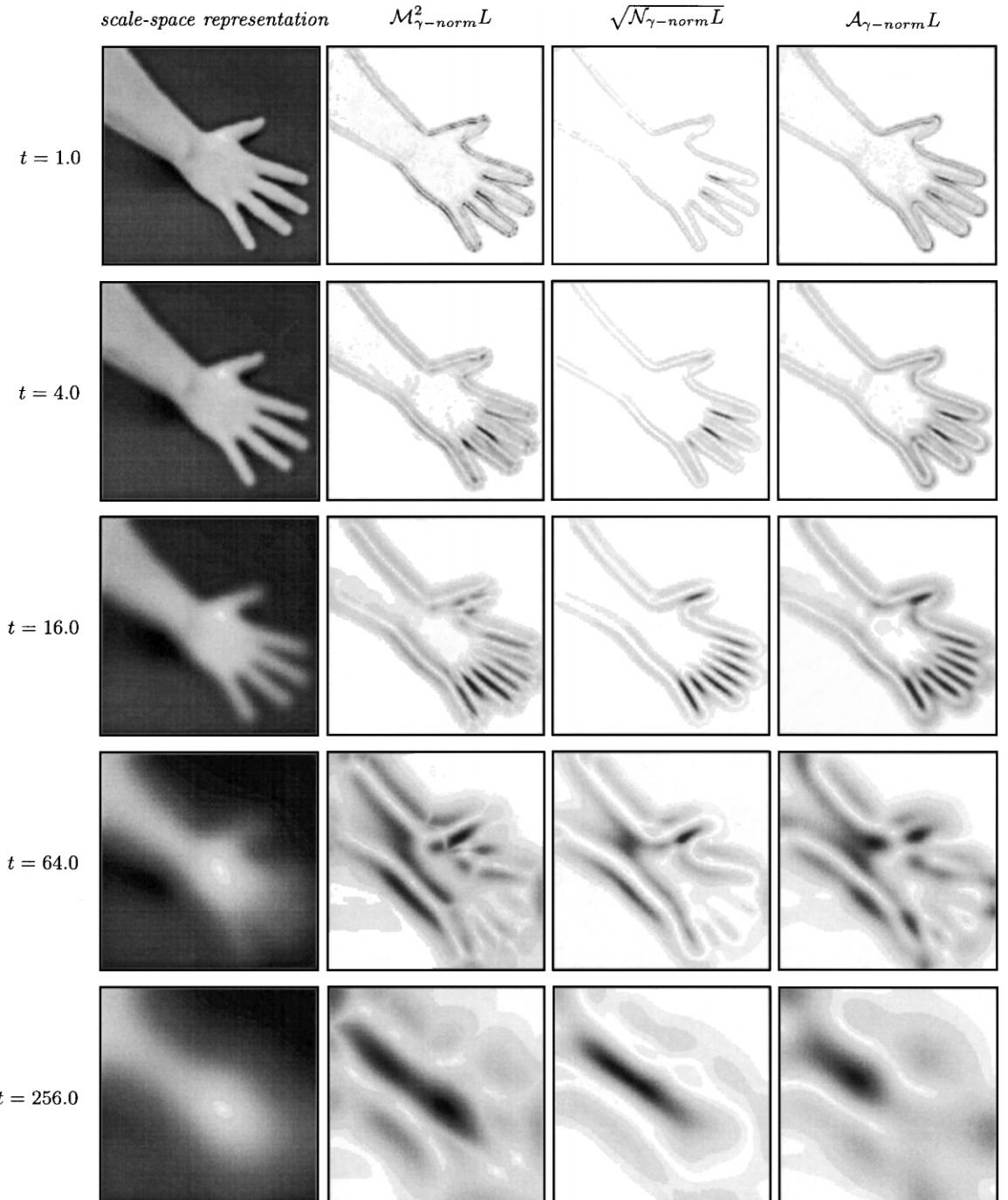


Figure 12. Ridge strength measures computed different scales for an image of a hand. Note that the fingers and the arm give rise to strong ridge strength responses at the same scales ($t \approx 16$ and $t \approx 256$) as the fixed-scale ridge detector in Fig. 10 succeeds in extracting corresponding ridge curves. Moreover, observe that $\mathcal{N}_{\gamma-\text{norm}} L$ has more ridge-specific response properties than $\mathcal{M}_{\gamma-\text{norm}} L$ and $\mathcal{A}_{\gamma-\text{norm}} L$.

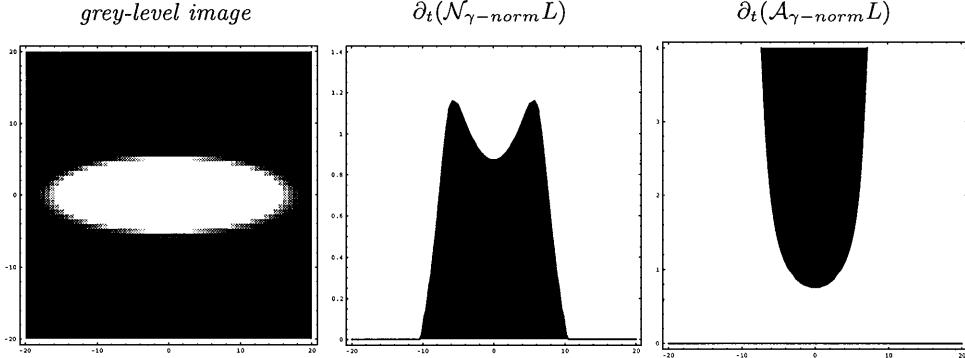


Figure 13. Selected scale levels along an elongated Gaussian ridge (of the form (57) with $t_1 = 10$ and $t_2 = 1$). (left) Grey-level image. (middle and right) Zero-crossings of $\partial_t(\mathcal{N}_{\gamma\text{-norm}}L)$ and $\partial_t(\mathcal{A}_{\gamma\text{-norm}}L)$, respectively, in a horizontal cross-section along the ridge. Observe that (once the distance has become sufficiently large) the selected scale levels decrease with the distance from the center in the first case, while the selected scales increase with the distance from the center in the second case.

the selected scale levels decrease with the distance from the center of the ridge (once the distance has become sufficiently large), while for scale selection based on $\mathcal{A}_{\gamma\text{-norm}}L$, the selected scale levels increase strongly towards the periphery. The reason why $\mathcal{N}_{\gamma\text{-norm}}L$ and $\mathcal{A}_{\gamma\text{-norm}}L$ have different behavior is that the eigenvalues L_{pp} and L_{qq} are of opposite sign in the periphery, which gives a relatively higher response of $\mathcal{A}_{\gamma\text{-norm}}L = (L_{pp,\gamma\text{-norm}} - L_{qq,\gamma\text{-norm}})^2$ than of $\mathcal{N}_{\gamma\text{-norm}}L = (L_{pp,\gamma\text{-norm}}^2 - L_{qq,\gamma\text{-norm}}^2)^2$.

5.7. Experiments

Let us now show the result of applying integrated ridge detectors with automatic scale selection to different types of real-world images.

In analogy with Section 4.7, we first compute the differential descriptors occurring in the definition of a scale-space ridge (43) at a number of scales in scale-space.¹⁸ Then, polygons are constructed to approximate the intersections of the two zero-crossing surfaces of L_p and $\partial_t(\mathcal{R}_{\gamma\text{-norm}})$ that satisfy the sign conditions $L_{pp} < 0$ ¹⁹ and $\partial_t(\mathcal{R}_{\gamma\text{-norm}}) < 0$. Finally, a significance measure is computed for each ridge curve in a way analogous to Section 4.6.1.

5.7.1. Measures of Ridge Saliency. To have the significance measures proportional to the local brightness contrast, the strength measures are again transformed before integration. For any scale-space ridge Γ , and for each measure of ridge strength, the saliency measure

is defined by:

$$M(\Gamma) = \int_{(x;t) \in \Gamma} (\mathcal{M}_{\gamma\text{-norm}})(x; t) ds, \quad (58)$$

$$N(\Gamma) = \int_{(x;t) \in \Gamma} \sqrt[4]{(\mathcal{N}_{\gamma\text{-norm}})(x; t)} ds, \quad (59)$$

$$A(\Gamma) = \int_{(x;t) \in \Gamma} \sqrt{(\mathcal{A}_{\gamma\text{-norm}})(x; t)} ds, \quad (60)$$

where the integration is again performed by projecting the scale-space ridge onto the image plane $ds^2 = dx^2 + dy^2$.

5.7.2. Scale Selection Based on $\mathcal{N}_{\gamma\text{-norm}}$ and $\mathcal{A}_{\gamma\text{-norm}}$. Figures 14 and 15 show the result of detecting bright ridges from the aerial image of the suburb and the image of the arm in Fig. 10, using scale selection based on maxima over scales of $\mathcal{N}_{\gamma\text{-norm}}$ and $\mathcal{A}_{\gamma\text{-norm}}$, respectively. As can be seen, the major roads are extracted from the aerial image. For the image of the arm, ridge descriptors are extracted for each one of the fingers. In addition, a coarse-scale ridge is obtained for the arm as a whole.

Figure 16 shows a three-dimensional illustration of the result from the arm image. Here, the five most significant scale-space ridges have been drawn as three-dimensional curves in scale-space with the height over the image plane representing the selected scale at each point. Note how the selected scale levels reflect the widths of ridges.

Figure 17 shows another illustration of this data, where each ridge curve has been represented by a

original grey-level image the 100 strongest bright ridges the 10 strongest bright ridges

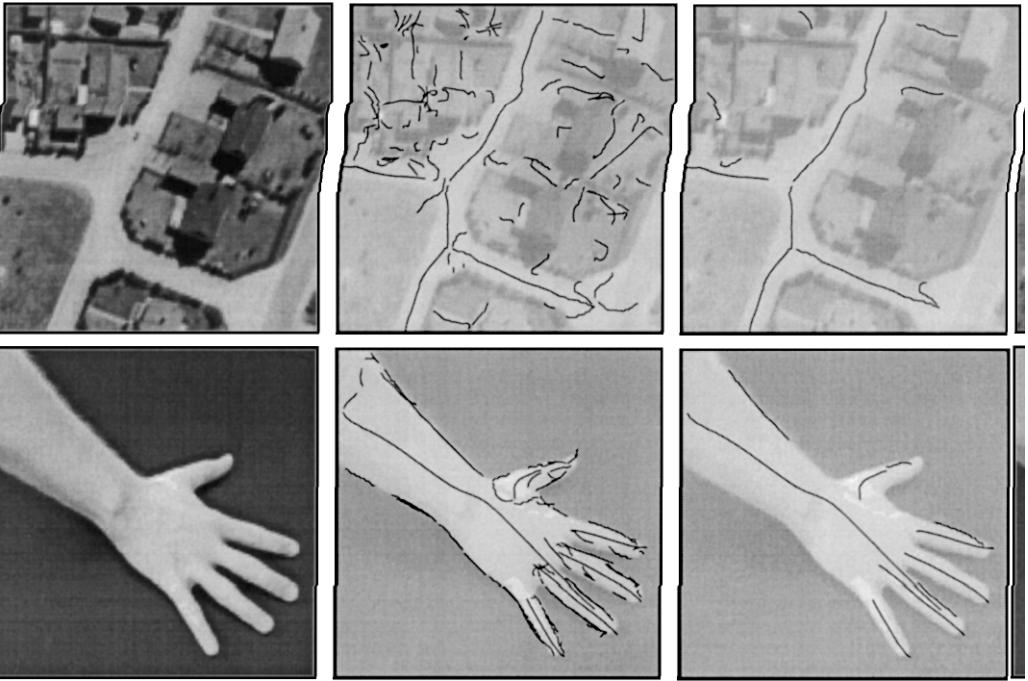


Figure 14. The result of detecting bright ridges with automatic scale selection based on local maxima over scales of $\mathcal{N}_{\gamma\text{-norm}}$ with $\gamma = \frac{3}{4}$. The middle and the right column show the 100 and 10 bright ridge curves having the highest significance values (according to (60)). Image size: 128×128 pixels in the top row, and 140×140 pixels in the bottom row.

original grey-level image the 100 strongest bright ridges the 10 strongest bright ridges

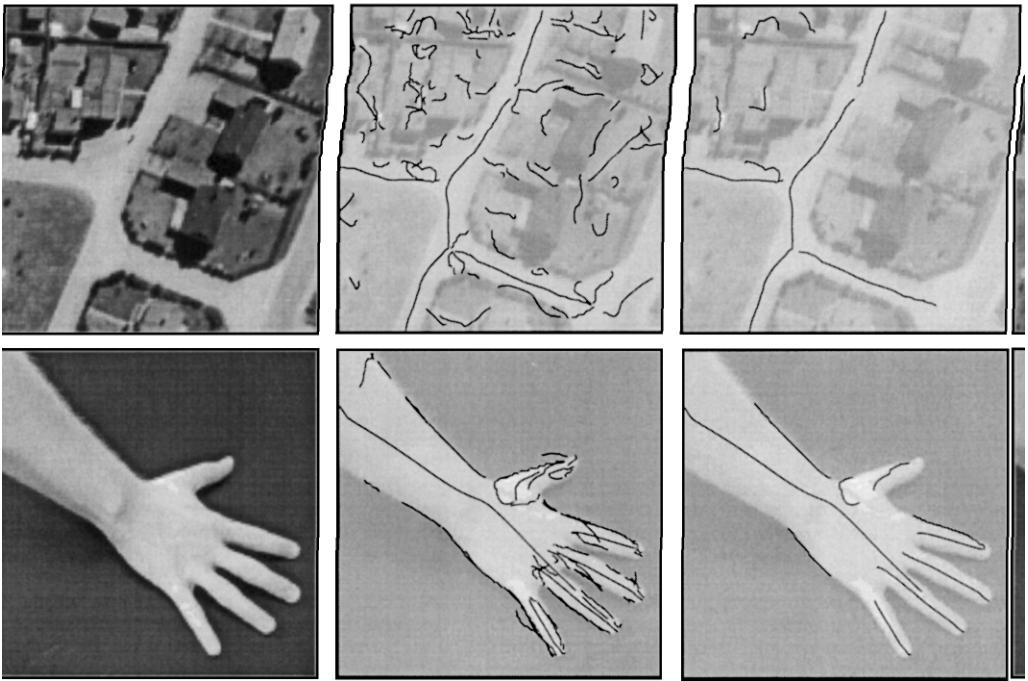


Figure 15. The result of detecting bright ridges with automatic scale selection based on local maxima over scales of $\mathcal{A}_{\gamma\text{-norm}}$ with $\gamma = \frac{3}{4}$. The middle and the right column show the 100 and 10 bright ridge curves having the highest significance values (according to (60)). Image size: 128×128 pixels in the top row, and 140×140 pixels in the bottom row.

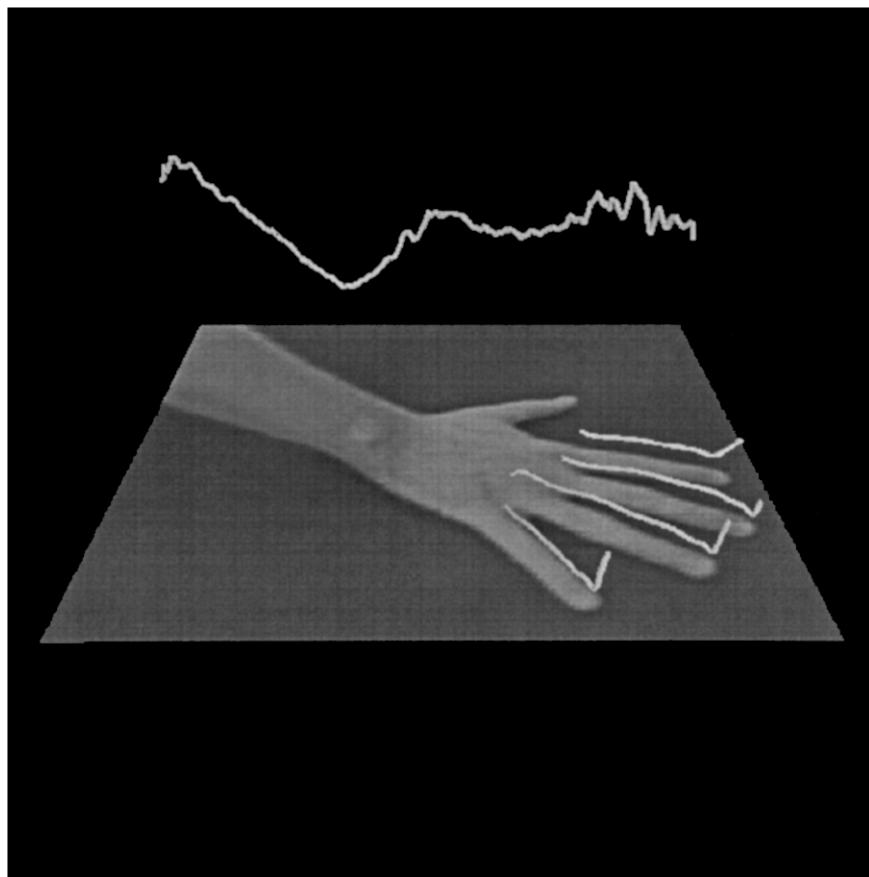


Figure 16. Three-dimensional view of the five strongest scale-space ridges extracted from the image of the arm in Fig. 14. Observe that a coarse-scale descriptor is extracted for the arm as a whole and that the individual fingers appear as ridge curves at finer scales.

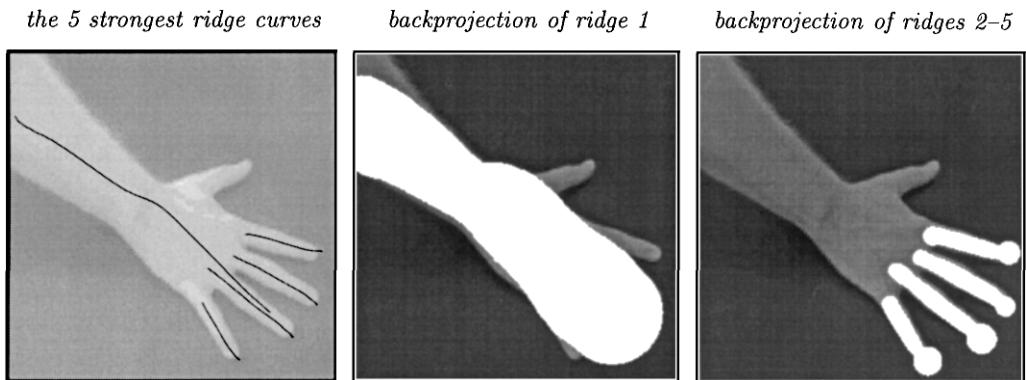


Figure 17. Alternative illustration of the five strongest scale-space ridges extracted from the image of the arm in Fig. 14 (and shown in Fig. 16). (left) The five strongest scale-space ridges drawn as dark ridge curves on a bright copy of the original grey-level image. (middle) The largest-scale ridge backprojected onto a dark copy of the original image by the union of a set of bright circles. Each circle is centered on the ridge curve with its radius proportional to the selected scale at that point. (right) Corresponding backprojections of the next four strongest scale-space ridges corresponding to the fingers.

region, constructed from the union of circles centered at the points on the ridge curve, and with the radius proportional to the selected scales measured in dimension length.

5.8. Summary

To conclude,²⁰ the proposed scale selection scheme has the desirable property of adapting the scale levels to the local image structure such that *the selected scales reflect the width of the ridge*.

6. Relations to Previous Works

Because of the range of topics spanned by this article, there are several links to previous works in addition to those already pointed out in the introductory sections.

The normalized gradient magnitude (with $\gamma = 1$) has been used also by Korn (1988), for relating thresholds on gradient magnitude at different scales. Inspired by the early work by Marr (1976), who observed how variations of the step size in a finite difference operator gives rise to different types of signatures depending on the edge type, Zhang and Bergholm (1993) studied ways of classifying edges based on the evolution properties over scales of the normalized gradient magnitude. Their approach was to first accumulate the signature and then to fit a predefined model to the data. Related techniques have been presented by Mallat and Zhong (1992), who characterize edges from the Lipschitz exponents of wavelet descriptors, Lu and Jain (1989) who reason about the behavior of Laplacian zero-crossings in scale-space, and Rohr (1992) who fit parametric junction and edge models to the image data to estimate the degree of diffuseness.

The notion of scale-space edge proposed here unifies the ideas of such diffuseness estimation schemes with the standard method for edge detection by non-maximum suppression proposed by Canny (1986) and Korn (1988). Moreover, the use of normalized derivatives in edge detection (Korn, 1988; Zhang and Bergholm, 1993), and the general idea of using maxima over scales of normalized derivatives for scale selection for feature detectors (Lindeberg, 1993c, 1994a), are here furthered to the notion of scale selection based on γ -normalized derivatives. In addition, a third-order

measure of edge strength is introduced, and an integrated edge detection scheme is presented in terms of the intersection of zero-crossing surfaces in scale-space. The local maximization of the edge strength measures across scales also provides a way to adapt the scale levels to the local image structures in such a way that the scale levels may vary along the edge, and diffuseness estimates be obtained directly when detecting edges and based on local operations only. Compared to the approach by Zhang and Bergholm (1993) there is hence no need for explicitly accumulating the signature by linking edges across scales, or to fit a model to the data to estimate the degree of diffuseness. Corresponding differences hold in relation to (Mallat and Zhong, 1992; Lu and Jain, 1989; Rohr, 1992).

The ridge detectors proposed by Pizer et al. (1994), Eberly et al. (1994), and Koller et al. (1995) also involve the maximization of certain strength measures across scales. In this respect, they are closely related to the scale selection principle used in this work. The methods in (Pizer et al., 1994; Koller et al., 1995), however, differ in the respect that they are nonlocal and construct ridges by propagating edge information from corresponding boundaries. The ridge detection scheme in (Eberly et al., 1994) is closer to this approach in the sense that it is local. This work differs, however, in the detailed ridge definition in the choice of ridge strength measures, and the explicit construction of ridges as intersections of zero-crossing surfaces. The use of γ -normalized derivatives instead of ordinary normalized derivatives also allows for a more direct control of the scale levels to be selected.²¹

The strongest connection to other works, however, is that this work makes explicit how the scale selection principle proposed in (Lindeberg, 1991, 1993c, 1994a) can be used for constructing integrated algorithms for detecting edges and ridges.

7. Summary and Discussion

We have presented a framework for automatic scale selection for detecting one-dimensional features, such as edges and ridges, from image data. Compared to a traditional approach of defining and extracting such features from a grey-level image at a fixed level of scale (or a set of such fixed scales), we have proposed that it is more natural to define the concepts of edges and ridges as one-dimensional curves in the three-dimensional

scale-space representation of the image (spanned by the space and scale dimensions). A basic reason why such a definition is more useful is that in general situations it will hardly ever be the case that a vision system can know in advance what scales are appropriate for analyzing the image structures in a given image. Therefore, a mechanism for automatic scale selection is a necessary complement to traditional multiscale processing in general, and feature detection in particular.

As we have illustrated by examples, a single scale level will usually not be appropriate for extracting the relevant edges in a real-world image of a scene of moderate complexity. For example, at distinct edge structures, such as sharp edges due at object boundaries with high contrast, the amount of smoothing should usually be as small as possible, to not affect the localization of the edges more than necessary. Conversely, at more diffuse edges structures, such as edges due to illumination effects and edge structures subject to out-of-focus blur, a substantial amount of smoothing may be required before the edge detector is able to produce coherent edge curves. Comparably coarse scales for edge detection will also be necessary if the noise level is high, or if other interfering fine-scale structures are present, such as surface texture. To achieve an appropriate trade-off in the well-known conflict between detection and localization, we must therefore allow the scale levels for edge detection to vary over the image. Specifically, to capture edges with varying degree of diffuseness, such as cast shadows, it will be necessary to let the scale levels vary along the edge. Similar problems arise when extracting ridges from image data. The general conclusion is that the scale level should be of the same order as the width of the ridge structure, and that a ridge detector is usually much more sensitive to the choice of scale levels than an edge detector.

7.1. Edge Detection: Automatic Scale Selection

To cope with the scale problem in edge detection, we proposed an extension of the notion of nonmaximum suppression to include the scale dimension already in the edge definition. In addition to the usual requirement, that the gradient magnitude should assume a maximum in the gradient direction (which gives rise to an edge surface in scale-space), we proposed to define a *scale-space edge* by simultaneous maximization

of a (normalized) measure of edge strength over scales.

Specifically, we considered the consequences of using two such measures of edge strength in more detail—the γ -normalized square gradient magnitude, $G_{\gamma\text{-norm}}$, and the negative γ -normalized third-order directional derivative in the gradient direction multiplied by the gradient magnitude raised to the power of three, $T_{\gamma\text{-norm}}$. By theoretical analysis for a set of characteristic edge models and by experiments on real-world data, we demonstrated that the qualitative behavior of the composed edge detection method is to select fine scales for sharp edge structures and to select coarse scales when extracting diffuse edges. Hence, the resulting scale selection mechanism has the desirable property of adapting the scale levels for edge detection to the local image structure. By such automatic selection of local appropriate scales for edge detection, we have not only eliminated the need for external choice of scale levels for the edge detection module. We have also obtained a methodology for edge detection, which can be expected to be more robust to changes in the environment. Notably, the behavior of scale selection method is not the result of any rule-based system. Instead, it follows as a direct *consequence* of the proposed edge definition.

Another important property of this approach is that the edge detector simultaneously determines an estimate of edge diffuseness as attribute to each edge point. Such information provides useful clues to the physical nature of the edge, which, for example, can be used for distinguishing object edges from shadow edges, and for estimating relative depth based on focus blur.

7.2. Ridge Detection: Automatic Scale Selection

Then, we turned to the problem of ridge detection. It was shown that by minor modifications, similar ideas could be used for formulating a ridge detector with automatic scale selection. A traditional fixed-scale differential geometric ridge definition, which defines ridges at any scale as points at which the main eigenvalue of the Hessian matrix assumes an extremum in the main eigendirection of the Hessian, was extended to the scale dimension, and the concept of a *scale-space ridge* was introduced, by maximizing a (normalized) measure of ridge strength over scales.

Specifically, three measures of ridge strength were considered in more detail, the γ -normalized main

absolute eigenvalue, $\mathcal{M}_{\gamma\text{-norm}}$ the γ -normalized square eigenvalue difference, $\mathcal{N}_{\gamma\text{-norm}}$, and the square of the γ -normalized eigenvalue difference, $\mathcal{A}_{\gamma\text{-norm}}$. For ridges with one-dimensional symmetry, all these entities have the same scale selection properties, and give rise to scale levels corresponding to the width of the ridge. Hence, in addition to automatic adaptation of the scale levels for ridge detection to the local image structure, this feature detector returns an estimate of the width of the ridge associated with each ridge point. This attribute information is important when mapping ridge features to the image regions they arose from.

7.3. Scale-Space Derivatives of High Order

The scale selection methodologies presented here are based on the zero-crossings of differential invariants containing derivatives up to order five, and the sign of differential invariants containing derivatives up to order seven. Sometimes it has been argued that for practical purposes it is very hard or even impossible to compute derivatives of order two or higher on real-world image data. With the results presented in this work, we have on the other hand demonstrated that highly useful results can indeed be obtained for much higher orders of differentiation, provided that the derivative approximations are computed in a multiscale framework and combined with an explicit mechanism for scale selection as well as a carefully chosen discrete implementation.

8. Extensions and Further Work

8.1. Multiple Feature Responses at Different Scales

An important property of feature detectors designed according to this framework is that they have the ability to respond to different types of image structures at different scales, which is impossible for a fixed-scale edge detector operating at a single level of scale. A fundamental problem in this context, however, concerns how to *combine* features extracted at different scales for further processing. For certain types of image data, we can expect qualitatively different types of interpretations to be possible depending on the scale of the analysis. One example is the classical abstraction hierarchy: forest-tree-branch-leaf. In this work, we have not made any effort to address this problem by integrating edge descriptors at different scales. It is worth pointing out,

however, that if this scheme is applied to such image data, it will respond to the outline of the tree at some coarse scale, and to the outlines of the leaves at some finer scales.²²

8.2. Selective Mechanisms

Another basic problem concerns the design of selective mechanisms. Whereas this scale selection methodology produces a much more compact set of edge descriptors than a fixed-scale edge detection at a corresponding set of scales, there is a need for ranking edges on saliency to suppress spurious responses and to allow subsets of the most salient edges to be selected. One of the most common ways of suppressing spurious edge responses is by (hysteresis) thresholding on the gradient magnitude. Of course, the approach presented here is compatible with such thresholding techniques (preferably based on the normalized measures of feature strength). The inherent limitations of such techniques, however, suggest that richer sources of information should be used already in the earliest processing stages. The saliency measures introduced in Section 5.5, constitute a step in this direction. More generally, one could conceive including stability properties of the features in scale-space, such as their scale-space lifetime, based on the explicit registration of bifurcation events (Lindeberg, 1993a).

8.3. Alternative Approaches to Feature Detection

Throughout this work we have focused on the detection of edge and ridge features represented as one-dimensional curves, following the dominant approach to machine vision today, based on sparse image features of low dimension. Extracting curve descriptors explicitly, however, leads to a number of technical complications, and with respect to a number of visual tasks one could conceive using only responses from differential operators (direct computation by dense feature descriptors), or to represent edges and ridges by regions instead of curves. In principle, there is nothing that prevents the ideas behind this approach from being extended in such ways.²³

Concerning the specific implementations of feature detection algorithms presented, they have been based on standard techniques, and the measures of feature strength chosen because of their intimate connections to the differential geometric feature detectors used. Whereas we by these experiments have shown that the

resulting methodology allows highly useful image descriptors to be extracted in a completely automatic way, the design choices made here should be interpreted as excluding the possibility that future work may show that other differential descriptors could improve the performance further.²⁴ The most important contribution of this work in this respect is that it opens up a general framework for scale selection, in which further developments and refinement of early visual modules can be performed, by integration with explicit scale selection mechanisms.

9. Conclusion: Scale Selection Principles

To conclude, this work combined with the earlier works in (Lindeberg, 1993c, 1994a) shows that *detection of local maxima over scales of normalized differential entities provides a consistent framework for generating hypothesis about local appropriate scales for detecting image features such as blobs, corners, edges and ridges.*

A related methodology for scale selection, for computing local image deformations such as optic flow and for performing image matching, is presented in (Lindeberg, 1995, 1996d). There, hypotheses about appropriate scales are generated from the scale levels at which *normalized error measures assume minima over scales*.

These general principles, implying that in the absence of further evidence, scale levels are selected from the scales at which the normalized response is as strongest, or the scale levels at which the error in the computed estimates are as smallest, may be far more general than the actual implementations presented so far.

Appendix

A.1. Detailed Algorithmic Description

According to the definitions in Sections 4.3 and 5.4, scale-space edges and scale-space ridges, respectively, are defined as the intersections of the zero-crossing surfaces of two differential invariants $\mathcal{Z}_1 L$ and $\mathcal{Z}_2 L$ that satisfy the sign conditions $\mathcal{S}_1 L < 0$ and $\mathcal{S}_2 L < 0$ for two other differential descriptors $\mathcal{S}_1 L$ and $\mathcal{S}_2 L$.

To compute these features from image data, we have made a rather straightforward extension of the single-scale edge detection method developed in (Lindeberg,

1993b), based on the following types of visual front-end operations:

- smoothing by (large support) convolution with a discrete Gaussian kernel,
- discrete derivative approximations from (small support) finite differences,
- (pointwise) combination of these descriptors to differential invariants, and
- interpolation for zero-crossings (by nearest neighbor comparisons).

In summary, the composed algorithms for multiscale edge and ridge detection consists of the following major processing steps:

A.1.1. Computation of Differential Invariants. In an initial processing step, differential geometric descriptors are computed based on the discrete scale-space framework in (Lindeberg, 1994c):

- *Choice of scale range:* Given a discrete image f of a certain size, select a scale range for the analysis.²⁵ Within this interval, distribute a set of scale levels t_k (in this implementation, we have used 40 scales), such that the ratio between successive scale values t_{k+1}/t_k is approximately constant. A natural way to distribute these scales, is such that the difference in effective scale between adjacent scales $t_{k+1} - t_k$ is approximately constant.
- *Scale-space smoothing:* For each scale t_k , compute the scale-space representation of f by separable convolution with the discrete analogue of the Gaussian kernel T : $L(\cdot; t_k) = T(\cdot; t_k) * f$. Alternatively, use the nonseparable discrete analogue, which allows for higher degree of rotational invariance.
- *Discrete approximation of γ -normalized differential invariants:* For each point at each scale, compute discrete derivative approximations of $L(\cdot; t_k)$ by central differences, combine these into discrete approximations of the differential invariants, and multiply the result by appropriate powers of t_k . Alternatively, normalize the discrete derivative approximation kernels to be constant over scales in appropriately selected l_p -norms, based on the constant L_p -norm interpretation of the γ -normalized derivative concept described in a companion paper (Lindeberg, 1996a).

A.1.2. Intersection of the Zero-Crossing Surfaces.

Given access to the differential invariants defining the zero-crossing surfaces, the next step is to construct a polygon approximation of the intersection of these surfaces.

In this implementation, we traverse all voxels in a way closely related to the marching cubes algorithm (Lorensen and Cline, 1987) and the marching lines algorithm (Thirion and Gourdon, 1993). For each voxel, a test is made if there the intersection of iso-surfaces passes through that voxel. Then, on top of this construction, a tracking algorithm links line segments into polygons, by following the end points of each detected line segment to the next voxel. In this way, connected polygons are obtained without any need for further combinatorial processing. At an overview level, this surface intersection algorithm works as follows:

- For each voxel:
 - If not this voxel was visited previously:
 - * Mark the voxel as visited and investigate if there is an intersection of iso-surfaces passing through this voxel.
 - If so, initiate a search procedure at each end point, which follows the line segment to the neighboring voxel.
 - If this voxel has not been visited previously, mark it as visited and investigate if there is an intersection passing through that voxel, etc.

Concerning the treatment of each individual voxel, a line segment representing the part of the intersection in this voxel is constructed as follows:

- Make sure that both \mathcal{S}_1 and \mathcal{S}_2 are negative in at least one of the corners of this voxel. If not, regard this voxel as having no surface intersection through it.
- For each edge of the voxel, and for each differential invariant \mathcal{Z}_1 and \mathcal{Z}_2 , test if this function has different signs in the corners corresponding to this segment. If so, estimate the position of the zero-crossing on this edge by linear interpolation.
- Then, for each face of the voxel, connect corresponding zero-crossings for each differential invariant \mathcal{Z}_1 and \mathcal{Z}_2 by straight lines, provided that this can be done in a nonambiguous way. In the current implementation, connections are constructed only if there are zero-crossings on exactly two edges.

If both \mathcal{Z}_1 and \mathcal{Z}_2 have zero-crossings on the same face, and if the straight lines representing these zero-crossings intersect, then an intersection between \mathcal{Z}_1 and \mathcal{Z}_2 is registered on this face.

- Finally, for this voxel, connect the intersection points by one (or by several) line segment(s), provided that this can be done in a nonambiguous way.

In the current implementation, a line segment is created only if there are exactly two intersection points on the faces of the voxel.

This algorithm is conservative in the sense that no attempts are made to resolve situations which are regarded as ambiguous by the three-step linear interpolation. For this reason, small gaps can be expected in the resulting edge segments.

In this implementation, a simple post-processing step has been applied, which connects adjacent polygons if the distance between the end points is sufficiently small²⁶ and the tangential directions at the end points are sufficiently similar.²⁷ Alternatively, one could enforce connectivity by using techniques such as tri-linear interpolation.²⁸

A.2. Derivatives of Edge Strength with Respect to Scale

In this appendix it will be shown how explicit expressions for the first- and second-order derivatives of $\mathcal{G}_{\gamma\text{-norm}} L$ with respect to the scale parameter can be expressed in terms of spatial derivatives only, based on the fact that all derivatives of the scale-space representation satisfy the diffusion Eq. (16).

A.2.1. Derivatives of $\mathcal{G}_{\gamma\text{-norm}} L$ with Respect to Scale.

If we differentiate $(\mathcal{G}_{\gamma\text{-norm}} L)(x, y, t)$ in (15) with respect to the scale parameter, we obtain

$$\partial_t (\mathcal{G}_{\gamma\text{-norm}} L) = \gamma t^{\gamma-1} (L_x^2 + L_y^2) + 2t^\gamma (L_x L_{xt} + L_y L_{yt}). \quad (61)$$

Substituting $L_{xt} = (L_{xxx} + L_{xyy})/2$ and $L_{yt} = (L_{xxy} + L_{yyy})/2$ in (61) then gives the following expression in terms of spatial derivatives only

$$\partial_t (\mathcal{G}_{\gamma\text{-norm}} L) = \gamma t^{\gamma-1} (L_x^2 + L_y^2) + t^\gamma (L_x (L_{xxx} + L_{xyy}) + L_y (L_{xxy} + L_{yyy})). \quad (62)$$

To have a compact notation for differential invariants, we shall often express them in terms of directional

derivatives in a local preferred coordinate system (as described in Section 4.1), by making the formal replacement $(x, y) \mapsto (u, v)$, and using the fact that the partial derivative L_u is zero in this coordinate system.²⁹ Then, (62) assumes the form

$$\partial_t(\mathcal{G}_{\gamma\text{-norm}}L) = t^{\gamma-1}L_v(\gamma L_v + t(L_{uuv} + L_{vvv})). \quad (63)$$

For the second-order derivative, corresponding calculations give

$$\begin{aligned} \partial_{tt}(\mathcal{G}_{\gamma\text{-norm}}L) &= \gamma(\gamma - 1)t^{\gamma-2}(L_x^2 + L_y^2) \\ &+ 2\gamma t^{\gamma-1}(L_x(L_{xyy} + L_{xxx}) + L_y(L_{xxy} + L_{yyy})) \\ &+ \frac{t^\gamma}{2}((L_{xxx} + L_{xyy})^2 + (L_{xxy} + L_{yyy})^2 \\ &+ L_x(L_{xxxxx} + 2L_{xxxxy} + L_{yyyyy}) \\ &+ L_y(L_{xxxxy} + 2L_{xxyyy} + L_{yyyyy})) \end{aligned} \quad (64)$$

in Cartesian coordinates. In the (u, v) -system, this expression reduces to

$$\begin{aligned} \partial_{tt}(\mathcal{G}_{\gamma\text{-norm}}L) &= \gamma(\gamma - 1)t^{\gamma-2}L_v^2 + 2\gamma t^{\gamma-1}L_v(L_{uuv} + L_{vvv}) \\ &+ \frac{t^\gamma}{2}((L_{uuu} + L_{uvv})^2 + (L_{uuv} + L_{vvv})^2 \\ &+ L_v(L_{uuuuv} + 2L_{uuvvv} + L_{vvvvv})). \end{aligned} \quad (65)$$

A.2.2. Derivatives of $\mathcal{T}_{\gamma\text{-norm}}L$ with Respect to Scale.

Similar calculations for the first-order derivative of $(\mathcal{T}_{\gamma\text{-norm}}L)(x, y; t)$ in (16) give

$$\begin{aligned} \partial_t(\mathcal{T}_{\gamma\text{-norm}}L) &= -3\gamma t^{3\gamma-1}(L_x^3 L_{xxx} + 3(L_x^2 L_y L_{xxy} + L_x L_y^2 L_{xyy}) \\ &+ L_y^3 L_{yyy}) - \frac{t^{3\gamma}}{2}(3(L_{xxx} + L_{xyy}) \\ &\times (L_x^2 L_{xxx} + 2L_y L_x L_{xxy} + L_y^2 L_{xyy}) \\ &+ 3(L_{yyy} + L_{xxy})(L_x^2 L_{xyy} + 2L_y L_x L_{xxy} \\ &+ L_y^2 L_{yyy}) + 3(L_x^2 L_y(L_{xxxxy} + L_{xxyyy}) \\ &+ L_x L_y^2(L_{xxyyy} + L_{xyyyy})) + L_x^3(L_{xxxxx} \\ &+ L_{xxxxy}) + L_y^3(L_{xxyyy} + L_{xyyyy})) \end{aligned} \quad (66)$$

in terms of Cartesian coordinates, and

$$\begin{aligned} \partial_t(\mathcal{T}_{\gamma\text{-norm}}L) &= -t^{3\gamma-1}L_v^2 \left(6\gamma L_v L_{vvv} + \frac{t}{2}L_v(L_{uuvvv} + L_{vvvvv}) \right. \\ &\left. + \frac{3t}{2}(L_{uuu}L_{uvv} + L_{uvv}^2 + L_{uuv}L_{vvv} + L_{vvv}^2) \right) \end{aligned} \quad (67)$$

in the local (u, v) -system. The Cartesian expression for the second-order derivative much larger. (In unfactorized form, the polynomial expression contains 251 partial derivatives of L .) When expressed in terms of the local (u, v) -system, it reduces to

$$\begin{aligned} \partial_{tt}(\mathcal{T}_{\gamma\text{-norm}}L) &= -3\gamma(3\gamma - 1)t^{3\gamma-2}L_v^3 L_{vvv} \\ &- 3\gamma t^{3\gamma-1}L_v^2(L_v(L_{vvvvv} + L_{uuvvv}) \\ &+ 3(L_{vvv}^2 + L_{uvv}^2 + L_{vvv}L_{uuv} + L_{uvv}L_{uuu})) \\ &- \frac{t^{3\gamma}}{4}(12L_v L_{uvv}(L_{vvv} + L_{uuv})(L_{uvv} + L_{uuu}) \\ &+ 6(L_v(L_{vvv}(L_{vvv} + L_{uuv})^2 \\ &+ L_{uuv}(L_{uvv} + L_{uuu})^2) \\ &+ L_v^2((L_{vvv} + L_{uuv})(L_{vvvvv} + L_{uuvvv}) \\ &+ (L_{uvv} + L_{uuu})(L_{uvvvv} + L_{uuuvv})) \\ &+ L_v^3(L_{vvvvvv} + 2L_{uuvvvvv} + L_{uuuuvvv}) \\ &+ 3L_v^2(L_{vvv}(L_{vvvvv} + 2L_{uuvvv} + L_{uuuuv}) \\ &+ L_{uvv}(L_{uvvvv} + 2L_{uuvvv} + L_{uuuuv}))). \end{aligned} \quad (68)$$

A.3. Derivatives of Ridge Strength with Respect to Scale

In this appendix, we derive explicit expressions the first- and second-order scale derivatives of the ridge strength measures defined in Section 5.5.

A.3.1. Derivatives of $\mathcal{M}_{\gamma\text{-norm}}L$ with Respect to Scale.

Following the methodology in Section 4.4 and Appendix A.2.1, let us use the diffusion equation for replacing derivatives of L with respect to t by derivatives of L with respect to x and y . Applying this procedure to the first- and second-order derivatives of $L_{pp,\gamma\text{-norm}}$

with respect to scale gives:³⁰

$$\begin{aligned}\partial_t(L_{pp,\gamma\text{-norm}}) &= \gamma t^{\gamma-1} \left(L_{xx} + L_{yy} - \sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2} \right) \\ &\quad + \frac{t^\gamma}{2} \left(L_{xxxx} + 2L_{xxyy} + L_{yyyy} - \frac{(L_{xx} - L_{yy})(L_{xxxx} - L_{yyyy}) + 4L_{xy}(L_{xxyy} + L_{yyyy})}{\sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}} \right). \end{aligned}\quad (69)$$

$$\begin{aligned}\partial_{tt}(L_{pp,\gamma\text{-norm}}) &= \gamma(\gamma-1)t^{\gamma-2} \left(L_{xx} + L_{yy} - \sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2} \right) + \frac{t^\gamma}{4} S_1 \\ &\quad + \gamma t^{\gamma-1} \left(L_{xxxx} + 2L_{xxyy} + L_{yyyy} - \frac{(L_{xx} - L_{yy})(L_{xxxx} - L_{yyyy}) + 4L_{xy}(L_{xxyy} + L_{yyyy})}{\sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}} \right) \end{aligned}\quad (70)$$

where

$$\begin{aligned}S_1 &= L_{xxxxxx} + 3(L_{xxxxyy} + L_{xxyyyy}) + L_{yyyyyy} + \frac{(4L_{xy}(L_{xxyy} - L_{yyyy}) + (L_{xx} - L_{yy})(L_{xxxx} - L_{yyyy}))^2}{((L_{xx} - L_{yy})^2 + 4L_{xy}^2)^{3/2}} \\ &\quad + \frac{(L_{xxxx} - L_{yyyy})^2 + (L_{xx} - L_{yy})(L_{xxxxxx} + L_{xxxxyy}L_{xxyyyy} - L_{yyyyyy})}{\sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}} \\ &\quad - \frac{4((L_{xxyy} + L_{yyyy})^2 + L_{xy}(L_{xxxxxy} + 2L_{xxyyy} + L_{yyyyyy}))}{\sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}}. \end{aligned}\quad (71)$$

These expressions simplify somewhat if we rewrite them in terms of local directional derivatives in the (p, q) -system. If we, in addition, apply the simplifying assumption $L_{pp} > L_{qq}$, which suppresses all occurrences of $\text{sign}(L_{pp} - L_{qq})$, then $\partial_t(L_{pp,\gamma\text{-norm}})$ and $\partial_{tt}(L_{pp,\gamma\text{-norm}})$ reduce to the form

$$\partial_t(L_{pp,\gamma\text{-norm}}) = 2\gamma t^{\gamma-1} L_{pp} + t^\gamma (L_{pppp} + L_{ppqq}), \quad (72)$$

$$\begin{aligned}\partial_{tt}(L_{pp,\gamma\text{-norm}}) &= 2\gamma(\gamma-1)t^{\gamma-2} L_{pp} + 2\gamma t^{\gamma-1} (L_{pppp} + L_{ppqq}) \\ &\quad + \frac{t^\gamma}{2} \left(L_{ppppqq} + 2L_{ppqqqq} + L_{qqqqqq} \right. \\ &\quad \left. - \frac{(L_{pppp} + L_{qqqq})^2 + 2(L_{pppq} + L_{pqqq})^2}{L_{pp} - L_{qq}} \right). \end{aligned}\quad (73)$$

A.3.2. Derivatives of $\mathcal{N}_{\gamma\text{-norm}} L$ with Respect to Scale.

For the first-order derivatives of $\mathcal{N}_{\gamma\text{-norm}} L$, corresponding calculations followed by simplifications to the

(p, q) -system give

$$\begin{aligned}\partial_t(\mathcal{N}_{\gamma\text{-norm}} L) &= 2t^{4\gamma} (L_{pp}^2 - L_{qq}^2) (L_{pp}(L_{pppp} + L_{ppqq}) \\ &\quad - L_{qq}(L_{ppqq} + L_{qqqq})), \end{aligned}\quad (74)$$

$$\begin{aligned}\partial_{tt}(\mathcal{N}_{\gamma\text{-norm}} L) &= 2t^{4\gamma} (L_{pp}^2 - L_{qq}^2) (L_{pppp} - L_{qqqq}) \\ &\quad \times (L_{qqqq} + 2L_{ppqq} + L_{pppp}) + \frac{t^{4\gamma}}{2} S_2, \end{aligned}\quad (75)$$

where

$$\begin{aligned}S_2 &= (L_{pp} - L_{qq})^2 ((L_{qqqq} + 2L_{ppqq} + L_{pppp})^2 \\ &\quad + (L_{pp} + L_{qq})(L_{qqqqqq} \\ &\quad + 3(L_{ppqqqq} + L_{ppppqq}) + L_{pppppp}) \\ &\quad + (L_{pp} + L_{qq})^2 (4(L_{pppq} + L_{pqqq})^2 \\ &\quad + (L_{pppp} - L_{qqqq})^2 \\ &\quad + (L_{pp} - L_{qq})(-L_{qqqqqq} - L_{ppqqqq} \\ &\quad + L_{ppppqq} + L_{pppppp})) \end{aligned}\quad (76)$$

A.3.3. Derivatives of \mathcal{A}_γ -norm L with Respect to Scale.

Finally, for \mathcal{A}_γ -norm L , the first- and second-order derivatives with respect to scale are

$$\begin{aligned} \partial_t(\mathcal{A}_\gamma\text{-norm } L) \\ = t^{2\gamma-1}(L_{pp} + L_{qq})(2\gamma(L_{pp} - L_{qq}) \\ + t(L_{pppp} - L_{qqqq})), \end{aligned} \quad (77)$$

$$\begin{aligned} \partial_{tt}(\mathcal{A}_\gamma\text{-norm } L) \\ = 2\gamma(2\gamma - 1)t^{2\gamma-2}(L_{pp} + L_{qq})^2 \\ + 4\gamma t^{2\gamma-1}(L_{pp} + L_{qq})(L_{pppp} - L_{qqqq}) \\ + \frac{t^{2\gamma}}{2}((L_{pppp} - L_{qqqq})^2 + 4(L_{pqqq} + L_{pppq})^2 \\ + (L_{pp} + L_{qq})(L_{pppppp} + L_{ppppqq} \\ - L_{ppqqqq} - L_{qqqqqq})). \end{aligned} \quad (78)$$

Acknowledgments

This work was partially performed under the ESPRIT-BRA project InSight and the ESPRIT-NSF collaboration Diffusion. The support from the Swedish Research Council for Engineering Sciences, TFR, is gratefully acknowledged.

The three-dimensional illustrations in Figs. 4, 6 and 16 have been produced with the kind assistance of Pascal Grootabuissat.

Earlier versions of the article have been presented in (Lindeberg, 1996b, 1996c).

Notes

1. As is well-known, however, this operator gives poor localization at curved structures, and contains no mechanism for distinguishing “false” from “true” edges.
2. In two dimensions, however, such recursive filters may be strongly biased to the orientation of image grid.
3. An important aspect to keep in mind concerning such “optimal edge detectors”, however, is that the optimality is relative to the model, and not necessarily with respect to the performance of the edge detector when applied to real-world data.
4. This specific implementation of nonmaximum suppression expressed in a scale-space setting will be described in more detail in Section 4.2.
5. Concerning the common use of a single fixed (often rather fine) scale level for edge detection in many computer vision applications, let us point out that such a strategy may be sufficient for scenes that are sufficiently simple, and the external conditions are under sufficiently good control, such that the contrast between the interesting objects and the background is high, and the edge structures are sharp. For such data, a large range of scales

can often be used for extracting connected edge curves, and any scale in that range may therefore be an acceptable choice. The complications we want to address in this paper are those arising when a single scale is not sufficient and/or such a scale cannot be known in advance. Note that (unlike sometimes implicitly assumed in previous works) these types of problems are not eliminated by a coarse-to-fine approach.

6. After expanding the expressions for L_{vv} and L_{vvv} into Cartesian derivatives using (6), and selecting the numerators from the resulting rational expressions.
7. For the singular boundary case $\gamma = 0$, corresponding to no derivative normalization, the monotone decrease of edge strength with scale is consistent with the general smoothing effect of the diffusion equation, which means that the amplitude of local variations decreases. In the other boundary case $\gamma = 1$, corresponding to traditional derivative normalization according to (3), the edge strength increases monotonically with scale, and asymptotically approaches the height of the diffuse step edge. A scale selection method based on such derivative normalization, would thus select an infinitely large scale for this edge model, corresponding to the infinite extent of the ideal step edge. By introducing the γ -parameterized derivative normalization, we make the scale selection method respond to the diffuseness of the edge instead of the spatial extent of the edge model.
8. Unfortunately, insertion of the polynomial expression for L into the partial scale derivatives of the edge strength measures (62) and (66) result in complicated polynomial expressions. For this reason, we will instead analyze this singularity by differentiation along the edge surface instead of in the vertical scale direction.
9. Besides such edges being highly sensitive to the choice of scale levels (the drift velocity may tend to infinity at bifurcations), edge features near bifurcations are not very likely to be significant for a vision system, since they will be extremely sensitive to local perturbations. When performing edge detection at a fixed scale, there is always a certain probability that the edge detector responds to such edges. By including a scale selection mechanism in the edge detector, we have obtained a way to suppress such features, based on a local analysis of the behavior of the edges across scales.
10. The reason we prefer to use $\gamma = 1$ in this case, is that γ -normalized scale-space derivatives are not perfectly scale invariant under rescalings of the input pattern unless $\gamma = 1$, which implies that edges of different degrees of diffuseness would be treated differently. Specifically, a systematic bias would be induced with preference to sharp edges as is apparent from the analysis of the step edge model in Section 4.5.1 (see Eq. (24)). Note, however, that if we (because of some deliberately chosen reason) would like to introduce such a bias, it holds that *relative ratios* of such saliency measures will still be preserved under rescalings of the input pattern, which implies that the relative ranking of the edge features will be invariant to size changes. This property is a direct consequence of the fact that the γ -normalized scale-space derivatives transform according to power laws (see Section 4.1 in the companion paper (Lindeberg, 1996a)), and that these saliency measures are self-similar functions of homogenous differential expressions.
11. With reference to the invariance properties discussed in the previous footnote, it is worth noting that the integration measure

- in these saliency measures is not scale invariant. Relative ratios will, however, be preserved under size changes and thus the relative ranking between features. If for some reason comparisons must be made between images of a scene in which objects have changed size (e.g., between different time frames in a time sequence), it is worth noting that such effects can be easily compensated for given the size information in the image features and the self-similar transformation property of this saliency measure. Of course, this property holds irrespective of the value of γ .
12. Fragmentation means a spurious loss of edge points, which destroys the connectivity of the edge curves. Note, however, that by definition, the scale-space edges form connected curves. Hence, with this edge concept, the fragmentation will be less severe than for algorithmically defined edge trackers.
 13. Here, 40 scale levels uniformly distributed between $t_{\min} = 0.1$ and $t_{\max} = 256$.
 14. A more extensive summary and discussion is given in Section 7.1.
 15. In terms of details, however, the approach by Eberly et al. (1994) differs from the approach taken here (and in (Lindeberg, 1994b)) in the sense that Eberly et al. (1994) compute derivatives in the p - and q -directions by differentiating along the *curved* trajectories of the (p, q) -system, whereas we here compute directional derivatives in the *tangential* directions of this curvi-linear coordinate system. For curved ridges, these two approaches will, in general, have different properties.
 16. Globally, however, we cannot expect to be able to reorient the coordinate system in this way. Hence, when implementing this ridge detection scheme in practice, the logical “or” operation occurring in (42) is always necessary. The simplifying assumption is introduced here with the only purpose of simplifying the presentation and shortening the algebraic expressions.
 17. For such image patterns, one of the eigenvalues of the Hessian matrix is zero, and the ridge strength measures are all proportional to the other eigenvalue of the Hessian raised to some power.
 18. Here, 40 scale levels uniformly distributed between $t_{\min} = 1$ and $t_{\max} = 512$.
 19. This condition concerns bright ridges. For dark ridges, it is, of course, changed to $L_{pp} > 0$.
 20. A more extensive summary and discussion is given in Section 7.2.
 21. Note that (as shown in Section 5.6.1) the use of ordinary normalized derivatives (with $\gamma = 1$) leads to the selection of much coarser scale levels than with this approach (using $\gamma = \frac{3}{4}$).
 22. Provided that there is a brightness contrast relative to the background and the resolution is sufficiently high for the leaves to be resolved.
 23. Compare, for example, with the dense pseudo-quadrature responses in (Lindeberg, 1994a), and the multiscale blob extraction in (Lindeberg, 1993a).
 24. For example, an interesting path to follow is to investigate if more explicit local estimates of signal to noise ratios can be included in the measures of feature strength, analogously to the minimization of error measures in the second localization step in (Lindeberg, 1994a).
 25. Here, for images of size up to 256×256 pixels, we have used $t_{\min} = 0.1$ and $t_{\max} = 256$ when detecting edges and $t_{\min} = 1$ and $t_{\max} = 512$ when detecting ridges. These scale ranges can, of course, be narrowed if further a priori information is available.

26. Of the same order as the distance between adjacent voxels.
27. In practice, such a processing stage can be efficiently implemented if an intermediate data structure is created for storing the end points on a grid having a slightly coarser resolution than the original grey-level image.
28. It seems plausible that such an extension would provide an conceptually cleaner way of reducing the fragmentation.
29. This corresponds to rotating the coordinate system such that the Cartesian coordinate system is aligned with the local gradient direction.
30. $\partial_t^{\alpha}(L_{qq, \gamma\text{-norm}})$ can be obtained from these expressions by changing the sign of each square root.

References

- Arcelli, C. and Sanniti di Baja, G. 1992. Ridge points in Euclidean distance maps. *Pattern Recognition Letters*, 13(4):237–242.
- Bergholm, F. 1987. Edge focusing. *IEEE Trans. Pattern Analysis and Machine Intell.*, 9(6):726–741.
- Blum, H. and Nagel, H.-H. 1978. Shape description using weighted symmetry axis features. *Pattern Recognition*, 10:167–180.
- Boyer, K.L. and Sarkar, S. 1991. On optimal infinite impulse response edge detection. *IEEE Trans. Pattern Analysis and Machine Intell.*, 13(11):1154–1171.
- Burbeck, C.A. and Pizer, S.M. 1995. Object representation by cores: Identifying and representing primitive spatial regions. *Vision Research*, in press.
- Canny, J. 1986. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intell.*, 8(6):679–698.
- Crowley, J.L. and Parker, A.C. 1984. A representation for shape based on peaks and ridges in the difference of low-pass transform. *IEEE Trans. Pattern Analysis and Machine Intell.*, 6(2):156–170.
- Deriche, R. 1987. Using Canny’s criteria to derive a recursively implemented optimal edge detector. *Int. J. of Computer Vision*, 1:167–187.
- Eberly, D., Gardner, R., Morse, B., Pizer, S., and Scharlach, C. 1994. Ridges for image analysis. *J. of Mathematical Imaging and Vision*, 4(4):353–373.
- Florack, L.M.J., ter Haar Romeny, B.M., Koenderink, J.J., and Viergever, M.A., 1992. Scale and the differential structure of images. *Image and Vision Computing*, 10(6):376–388.
- Gauch, J.M. and Pizer, S.M. 1993. Multiresolution analysis of ridges and valleys in grey-scale images. *IEEE Trans. Pattern Analysis and Machine Intell.*, 15(6):635–646.
- Griffin, L.D., Colchester, A.C.F., and Robinson, G.P. 1992. Scale and segmentation of images using maximum gradient paths. *Image and Vision Computing*, 10(6):389–402.
- ter Haar Romeny, B. (Ed.) 1994. *Geometry-Driven Diffusion in Computer Vision*. Kluwer Academic Publishers: Netherlands.
- Haralick, R.M. 1983. Ridges and valleys in digital images. *Computer Vision, Graphics, and Image Processing*, 22:28–38.
- Haralick, R.M. 1984. Digital step edges from zero-crossings of second directional derivatives. *IEEE Trans. Pattern Analysis and Machine Intell.*, 6.
- Koenderink, J.J. 1984. The structure of images. *Biological Cybernetics*, 50:363–370.

- Koenderink, J.J. and van Doorn, A.J. 1992. Generic neighborhood operators. *IEEE Trans. Pattern Analysis and Machine Intell.*, 14(6):597–605.
- Koenderink, J.J. and van Doorn, A.J. 1994. Two-plus-one-dimensional differential geometry. *Pattern Recognition Letters*, 15(5):439–444.
- Koller, T.M., Gerig, G., Székely, G., and Dettwiler, D. 1995. Multiscale detection of curvilinear structures in 2-D and 3-D image data. In *Proc. 5th Int. Conf. on Computer Vision*, Cambridge, MA, pp. 864–869.
- Korn, A.F. 1988. Toward a symbolic representation of intensity changes in images. *IEEE Trans. Pattern Analysis and Machine Intell.*, 10(5):610–625.
- Lindeberg, T. 1990. Scale-space for discrete signals. *IEEE Trans. Pattern Analysis and Machine Intell.*, 12(3):234–254.
- Lindeberg, T. 1991. Discrete scale-space theory and the scale-space primal sketch. Ph.D. Dissertation, Dept. of Numerical Analysis and Computing Science, KTH, ISRN KTH/NA/P-91/08-SE.
- Lindeberg, T. 1993a. Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention. *Int. J. of Computer Vision*, 11(3):283–318.
- Lindeberg, T. 1993b. Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction. *J. of Mathematical Imaging and Vision*, 3(4):349–376.
- Lindeberg, T. 1993c. On scale selection for differential operators. In *Proc. 8th Scandinavian Conf. on Image Analysis*, K. Heia, K.A. Høgda, B. Braathen (Eds.), Tromsø, Norway.
- Lindeberg, T. 1994a. Scale selection for differential operators. Technical Report ISRN KTH/NA/P-94/03-SE, Dept. of Numerical Analysis and Computing Science, KTH.
- Lindeberg, T. 1994b. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21(2):225–270. Supplement *Advances in Applied Statistics: Statistics and Images*: 2.
- Lindeberg, T. 1994c. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers: Netherlands.
- Lindeberg, T. 1995. Direct estimation of affine deformations of brightness patterns using visual front-end operators with automatic scale selection. In *Proc. 5th Int. Conf. on Computer Vision*, Cambridge, MA, pp. 134–141.
- Lindeberg, T. 1996a. Feature detection with automatic scale selection. Technical Report ISRN KTH/NA/P-96/18-SE, Dept. of Numerical Analysis and Computing Science, KTH. Revised version in *Int. Jnl. Comp. Vision*, 30(2), 1998.
- Lindeberg, T. 1996b. Edge detection and ridge detection with automatic scale selection. Technical Report ISRN KTH/NA/P-96/06-SE, Dept. of Numerical Analysis and Computing Science, KTH.
- Lindeberg, T. 1996c. Edge detection and ridge detection with automatic scale selection. In *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, San Francisco, California, pp. 465–470.
- Lindeberg, T. 1996d. A scale selection principle for estimating image deformations. Technical Report ISRN KTH/NA/P-96/16-SE, Dept. of Numerical Analysis and Computing Science, KTH; *Image and Vision Computing*, in press.
- Lorensen, W.E. and Cline, H.E. 1987. Marching cubes: A high resolution 3-D surface construction algorithm. *Computer Graphics*, 21(4):163–169.
- Lu, Y. and Jain, R.C. 1989. Reasoning about edges in scale-space. *IEEE Trans. Pattern Analysis and Machine Intell.*, 14(4):450–468.
- Mallat, S.G. and Zhong, S. 1992. Characterization of signals from multi-scale edges. *IEEE Trans. Pattern Analysis and Machine Intell.*, 14(7):710–723.
- Marr, D. and Hildreth, E. 1980. Theory of edge detection. *Proc. Royal Soc. London*, 207:187–217.
- Marr, D.C. 1976. Early processing of visual information. *Phil. Trans. Royal Soc. (B)*, 27S:483–524.
- Monga, O., Lengagne, R., and Deriche, R. 1994. Extraction of the zero-crossings of the curvature derivatives in volumic 3-D medical images: A multi-scale approach. In *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, Seattle, Washington, pp. 852–855.
- Morse, B.S., Pizer, S.M., and Liu, A. 1994. Multi-scale medial analysis of medical images. *Image and Vision Computing*, 12(6):327–338.
- Nalwa, V.S. and Binford, T.O. 1986. On detecting edges. *IEEE Trans. Pattern Analysis and Machine Intell.*, 8(6):699–714.
- Nitzberg, M. and Shiota, T. 1992. Non-linear image filtering with edge and corner enhancement. *IEEE Trans. Pattern Analysis and Machine Intell.*, 14(8):826–833.
- Ogniewicz, R.L. and Kübler, O. 1995. Hierarchic voronoi skeletons. *Pattern Recognition*, 28(3):343–359.
- Perona, P. and Malik, J. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Analysis and Machine Intell.*, 12(7):629–639.
- Petrou, M. and Kittler, J. 1991. Optimal edge detectors for ramp edges. *IEEE Trans. Pattern Analysis and Machine Intell.*, 13(5):483–491.
- Pingle, K.K. 1969. Visual perception by computer. In *Automatic Interpretation and Classification of Images*, A. Grasselli (Ed.), Academic Press: New York, pp. 277–284.
- Pizer, S.M., Burbeck, C.A., Coggins, J.M., Fritsch, D.S., and Morse, B.S. 1994. Object shape before boundary shape: Scale-space medial axis. *J. of Mathematical Imaging and Vision*, 4:303–313.
- Poston, T. and Stewart, I. 1978. *Catastrophe Theory and its Applications*. Pitman: London.
- Prewitt, J.M.S. 1970. Object enhancement and extraction. In *Picture Processing and Psychophysics*, A. Rosenfeld and B.S. Lipkin (Eds.), Academic Press: New York, pp. 75–149.
- Roberts, L.G. 1965. Machine perception of three-dimensional solids. In *Optical and Electro-Optical Information Processing*, J.T.ipp et al. (Eds.), MIT Press; Cambridge, MA, pp. 159–197.
- Rohr, K. 1992. Modelling and identification of characteristic intensity variations. *Image and Vision Computing*, (2):66–76.
- Rosenfeld, A. and Thurston, M. 1971. Edge and curve detection for visual scene analysis. *IEEE Trans. Computers*, 20(5):562–569.
- Saint-Marc, P., Chen, J.-S., and Medioni, G. 1991. Adaptive smoothing: A general tool for early vision. *IEEE Trans. Pattern Analysis and Machine Intell.*, 13(6):514–529.
- De Saint-Venant. 1852. Surfaces à plus grande pente constituées sur des lignes courbes. *Bull. Soc. Philomath. Paris*.
- Thirion, J.-P. and Gourdon, A. 1993. The marching lines algorithm: New results and proofs. Technical Report RR-1881-1, INRIA, Sophia-Antipolis, France.
- Torre, V. and Poggio, T.A. 1980. On edge detection. *IEEE Trans. Pattern Analysis and Machine Intell.*, 8(2):147–163.
- Wilson, R. and Bhalerao, A.H. 1992. Kernel design for efficient multiresolution edge detection and orientation estimation. *IEEE Trans. Pattern Analysis and Machine Intell.*, 14(3):384–390.

- Witkin, A.P. 1983. Scale-space filtering. In *Proc. 8th Int. Joint Conf. Art. Intell.*, Karlsruhe, West Germany, pp. 1019–1022.
- Yuille, A.L. and Poggio, T.A. 1986. Scaling theorems for zero-crossings. *IEEE Trans. Pattern Analysis and Machine Intell.*, 8:15–25.
- Zhang, W. and Bergholm, F. 1993. An extension of Marr's signature based edge classification and other methods for determination of diffuseness and height of edges, as well as line width. In *Proc. 4th Int. Conf. on Computer Vision*, H.-H. Nagel et al. (Eds.), Berlin, Germany, pp. 183–191, IEEE Computer Society Press.