

Feature 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.

<http://www.nada.kth.se/~tony>

Email: tony@nada.kth.se

*Technical report ISRN KTH/NA/P-96/18-SE, May 1996, Revised August 1998.
Int. J. of Computer Vision, vol 30, number 2, 1998. (In press).*

Abstract

The fact that objects in the world appear in different ways depending on the scale of observation has important implications if one aims at describing them. It shows that the notion of *scale* is of utmost importance when processing unknown measurement data by automatic methods. In their seminal works, Witkin (1983) and Koenderink (1984) proposed to approach this problem by *representing* image structures at different scales in a so-called scale-space representation. Traditional scale-space theory building on this work, however, does not address the problem of how to *select* local appropriate scales for further analysis.

This article proposes a systematic methodology for dealing with this problem. A framework is proposed for *generating hypotheses about interesting scale levels in image data*, based on a general principle stating that *local extrema over scales* of different combinations of γ -normalized derivatives are likely candidates to correspond to interesting structures. Specifically, it is shown how this idea can be used as a major mechanism in algorithms for automatic scale selection, which adapt the local scales of processing to the local image structure.

Support for the proposed approach is given in terms of a general theoretical investigation of the behaviour of the scale selection method under rescalings of the input pattern and by experiments on real-world and synthetic data. Support is also given by a detailed analysis of how different types of feature detectors perform when integrated with a scale selection mechanism and then applied to characteristic model patterns. Specifically, it is described in detail how the proposed methodology applies to the problems of blob detection, junction detection, edge detection, ridge detection and local frequency estimation.

In many computer vision applications, the poor performance of the low-level vision modules constitutes a major bottle-neck. It will be argued that the inclusion of mechanisms for automatic scale selection is essential if we are to construct vision systems to analyse complex unknown environments.

Keywords: scale, scale-space, scale selection, normalized derivative, feature detection, blob detection, corner detection, frequency estimation, Gaussian derivative, scale-space, multi-scale representation, computer vision

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 figure 5 and figure 11 have been generated with the kind assistance of Pascal Grostabussiat.

Contents

1	Introduction	1
1.1	Outline of the presentation	2
2	Scale-space representation: Review	3
3	Normalized derivatives and intuitive idea for scale selection	3
4	Proposed methodology for scale selection	5
4.1	General scaling property of local maxima over scales	6
4.2	The scale selection mechanism in practice	8
4.3	Experiments: Scale-space signatures from real data	9
4.4	Simultaneous detection of interesting points and scales	9
5	Blob detection with automatic scale selection	11
5.1	Analysis of scale-space maxima for idealized model patterns	11
5.2	Comparisons with fixed-scale blob detection	14
5.3	Applications of blob detection with automatic scale selection	15
6	Junction detection with automatic scale selection	15
6.1	Selection of detection scales from normalized scale-space maxima	16
6.2	Analysis of scale-space maxima for diffuse junction models	18
6.3	Experiments: Scale-space signatures in junction detection	19
7	Feature localization with automatic scale selection	21
7.1	Corner localization by local consistency	21
7.2	Automatic selection of localization scales	23
7.3	Experiments: Choice of localization scale	25
7.4	Composed scheme for junction detection and localization	27
7.5	Further experiments	28
7.6	Applications of corner detection with automatic scale selection	32
7.7	Extensions of the junction detection method	32
7.8	Extensions to edge detection	33
8	Dense frequency estimation	34
9	Analysis and interpretation of normalized derivatives	38
9.1	Interpretation of γ -normalized derivatives in terms of L_p -norms	38
9.2	Interpretation in terms of self-similar Fourier spectrum	38
9.3	Relations to previous work	40
10	Summary and discussion	40
10.1	Technical contributions	41
A	Appendix	42
A.1	Necessity of the form of the γ -parameterized derivative operator	42
A.2	L_p -normalization interpretation of γ -normalized derivatives	44
A.3	Normalized derivative responses to self-similar power spectra	44
A.4	Discrete implementation of the scale selection mechanisms	45

1 Introduction

One of the very fundamental problems that arises when analysing real-world measurement data originates from the fact that objects in the world may appear in different ways depending upon the scale of observation. This fact is well-known in physics, where phenomena are modelled at several levels of scale, ranging from particle physics and quantum mechanics at fine scales, through thermodynamics and solid mechanics dealing with every-day phenomena, to astronomy and relativity theory at scales much larger than those we are usually dealing with. Notably, the type of physical description that is obtained may be strongly dependent on the scale at which the world is modelled, and this is in clear contrast to certain idealized mathematical entities, such as “point” or “line”, which are independent of the scale of observation.

In certain controlled situations, appropriate scales for analysis may be known *a priori*. For example, a desirable property of a good physicist is his intuitive ability to select appropriate scales to model a given situation. Under other circumstances, however, it may not be obvious at all to determine in advance what are the proper scales. One such example is a vision system with the task of analysing unknown scenes. Besides the inherent multi-scale properties of real world objects (which, in general, are unknown), such a system has to face the problems that the perspective mapping gives rise to size variations, that noise is introduced in the image formation process, and that the available data are two-dimensional data sets reflecting only indirect properties of a three-dimensional world. To be able to cope with these problems, an image representation that explicitly incorporates the notion of scale is a crucially important tool whenever we attempt to interpret sensory data, such as images, by automatic methods.

In computer vision and image processing, these insights have lead to the construction of multi-scale representations of image data, obtained by embedding any given signal into a one-parameter family of derived signals (Burt 1981; Crowley 1981; Witkin 1983; Koenderink 1984; Yuille and Poggio 1986; Florack *et al.* 1992; Lindenberg 1994d; Haar Romeny 1994). This family should be parameterized by a scale parameter and be generated in such a way that fine-scale structures are successively suppressed when the scale parameter is increased. A main intention behind this construction is to obtain a separation of the image structures in the original image, such that fine scale image structures only exist at the finest scales in the multi-scale representation. Thereby, the task of operating on the image data will be simplified, provided that the operations are performed at sufficiently coarse scales where unnecessary and irrelevant fine-scale structures have been suppressed. Empirically, this idea has proved to be extremely useful, and multi-scale representations such as pyramids, scale-space representation and non-linear diffusion methods are commonly used as preprocessing steps to a large number of early visual operations, including feature detection, stereo matching, optic flow, and the computation of shape cues.

A multi-scale representation by itself, however, contains no explicit information about what image structures should be regarded as significant or what scales are appropriate for treating those. Hence, unless early judgements can be made about what image structures should be regarded as important, we obtain a substantial expansion of the amount of data to be interpreted by later stage processes. In most previous works, this problem has been handled by formulating algorithms which rely on the information present in the data at a small set of manually chosen scales (or even a single scale). Alternatively, coarse-to-fine algorithms have been expressed, which start at a given coarse scale and propagate down to a given finer scale. Determining such

scales in advance, however, leads to the introduction of free parameters. If one aims at autonomous algorithms which are to operate in a complex environment without need for external parameter tuning, we therefore argue that it is essential to *complement traditional multi-scale processing by explicit mechanisms for scale selection*. Notably, image descriptors can be highly unstable if computed at inappropriately chosen scales, whereas a proper tuning of the scale parameter can improve the quality of an image descriptor substantially. As will be demonstrated later, local scale information can also constitute an important to in its own right.

Early work addressing this problem was presented in (Lindeberg 1991, 1993a) for blob-like image structures. The basic idea was to study the behaviour of image structures over scales, and to measure the saliency of image structures from the stability properties and the lifetime of these structures in scale-space. Scale levels were selected from the scales at which a measure of blob strength assumed local maxima over scales and significant image structures from the stability of the blob structures in scale-space. Experimentally, it was shown that this approach could be used for extracting regions of interest with associated scale levels, which in turn could serve as to guide various early visual processes.

The subject of this article is to address the problem of automatic scale selection in a more general setting, for wider classes of image descriptors. We shall be concerned with the problem of extracting image features and computing filter-like image descriptors, and present a scale selection principle for image descriptors which can be expressed in terms of Gaussian derivative filters. The general idea for scale selection that will be proposed is to study the *evolution properties over scales* of normalized differential descriptors. Specifically, it will be suggested that *local extrema over scales* of such *normalized differential entities*, which arise in this way, are likely to correspond to interesting image structures. By theoretical considerations and experiments it will be shown that this approach gives rise to intuitively reasonable results in different situations and that it provides a unified framework for scale selection for detecting image features such as blobs, corners, edges and ridges.

1.1 Outline of the presentation

The presentation is organized as follows: Section 2 reviews the main concepts from scale-space theory we build upon. Section 3 introduces the notion of normalized derivatives and illustrates how maxima over scales of normalized Gaussian derivatives reflect the frequency content in sine wave patterns. This material serves as a preparation for section 4, which presents the proposed scale selection methodology and shows how it applies generally to a large class of differential descriptors. Section 4 also proposes a general extension of the common idea of defining features as zero-crossings of spatial differential descriptors. If a scale selection mechanism is integrated into such a feature detector, this corresponds to adding another zero-crossing requirement over the scale dimension in the differential feature definition.

Then, section 5 and section 6 show in detail how these ideas can be used for formulating blob detectors and corner detectors with automatic scale selection. Section 8 shows an example of how this approach applies to the computation of dense feature maps. Section 9 describes different ways of interpreting the normalized derivative concept. Finally, section 10 summarizes the main results and ideas of the approach. In a complementary paper (Lindeberg 1996a) it is developed in detail how this approach applies to edge detection and ridge detection.

Earlier presentations of different parts of this material have appeared elsewhere (Lindeberg 1993b, 1994a, 1994d, 1996b) as well as applications of the general ideas to

various problem domains (Lindeberg and Gårding 1993, 1997; Gårding and Lindeberg 1996; Lindeberg and Li 1995, 1997; Bretzner and Lindeberg 1998, 1997; Almansa and Lindeberg 1996; Wiltshi *et al.* 1997; Lindeberg 1997). The subject of this paper is to present a coherent description of the proposed scale selection methodology in journal form, including the developments and refinements that have been performed since the earliest presented manuscripts.

2 Scale-space representation: Review

Given any continuous signals $f: \mathbb{R}^D \rightarrow \mathbb{R}$, the (linear) scale-space representation $L: \mathbb{R}^D \times \mathbb{R}_+ \rightarrow \mathbb{R}$ of f is defined as the solution to the diffusion equation

$$\partial_t L = \frac{1}{2} \nabla^2 L = \frac{1}{2} \sum_{i=1}^D \partial_{x_i x_i} L \quad (1)$$

with initial condition $L(\cdot; 0) = f(\cdot)$. Equivalently, this family can be defined by convolution with Gaussian kernels of various width t

$$L(\cdot; t) = g(\cdot; t) * f(\cdot), \quad (2)$$

where $g: \mathbb{R}^D \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is given by

$$g(x; t) = \frac{1}{(2\pi t)^{N/2}} e^{-(x_1^2 + \dots + x_D^2)/(2t)}, \quad (3)$$

and $x = (x_1, \dots, x_D)^T$. There are several mathematical results (Koenderink 1984; Babaud *et al.* 1986; Yuille and Poggio 1986; Lindeberg 1990, 1994d, 1994b; Koenderink and van Doorn 1990, 1992; Florack *et al.* 1992; Florack 1993; Florack *et al.* 1994; Pauwels *et al.* 1995) stating that within the class of linear transformations the Gaussian kernel is the unique kernel for generating a scale-space. The conditions that specify the uniqueness are essentially linearity and shift invariance combined with different ways of formalizing the notion that new structures should not be created in the transformation from a finer to a coarser scale.

Interestingly, the results from these theoretical considerations are in qualitative agreement with the results of biological evolution. Neurophysiological studies by (Young 1985, 1987) have shown that there are receptive fields in the mammalian retina and visual cortex, whose measured response profiles can be well modelled by Gaussian derivatives up to order four. In these respects, the scale-space representation with its associated *Gaussian derivative operators* (where α denotes the order of differentiation)

$$L_{x^\alpha}(\cdot; t) = (\partial_{x^\alpha} L)(\cdot, \cdot; t) = \partial_{x^\alpha} (g * f) = (\partial_{x^\alpha} g) * f = g * (\partial_{x^\alpha} f), \quad (4)$$

can be seen as a canonical idealized model of a visual front-end. It is for this multi-scale representation concept we will develop the scale selection methodology.

3 Normalized derivatives and intuitive idea for scale selection

A well-known property of the scale-space representation is that the amplitude of spatial derivatives

$$L_{x^\alpha}(\cdot; t) = \partial_{x^\alpha} L(\cdot; t) = \partial_{x_1^{\alpha_1}} \dots \partial_{x_D^{\alpha_D}} L(\cdot; t) \quad (5)$$

in general *decrease with scale*, *i.e.*, if a signal is subject to scale-space smoothing, then the numerical values of spatial derivatives computed from the smoothed data can be expected to decrease. This is a direct consequence of the non-enhancement property of local extrema, which states that the value at a local maximum cannot increase, and the value at a local minimum cannot decrease. In practice, it means that the amplitude of the variations in a signal will always decrease with scale.

As a simple example of this, consider a sinusoidal input signal¹ of some given frequency ω_0 ; for simplicity in one dimension,

$$f(x) = \sin \omega_0 x. \quad (6)$$

It is straightforward to show that the solution of the diffusion equation is given by

$$L(x; t) = e^{-\omega_0^2 t/2} \sin \omega_0 x. \quad (7)$$

Hence, the amplitude of the scale-space representation, L_{max} , as well as the amplitude of the m th order smoothed derivative, $L_{x^m, max}$, decrease exponentially with scale

$$L_{max}(t) = e^{-\omega_0^2 t/2}, \quad L_{x^m, max}(t) = \omega_0^m e^{-\omega_0^2 t/2}. \quad (8)$$

Let us next introduce a γ -normalized derivative operator defined by

$$\partial_{\xi, \gamma\text{-norm}} = t^{\gamma/2} \partial_x, \quad (9)$$

which corresponds to the change of variables

$$\xi = \frac{x}{t^{\gamma/2}}. \quad (10)$$

In the special case when $\gamma = 1$, these ξ -coordinates and their associated normalized derivative operator are *dimensionless*. The property of perfect *scale invariance* has been used by (Florack *et al.* 1992) as a main requirement in an axiomatic scale-space formulation (see also (Pauwels *et al.* 1995; Lindeberg 1994b)). As we shall see later, however, values of $\gamma < 1$ will be highly useful when formulating scale selection mechanisms for edge detection and ridge detection.

For the sinusoidal signal, the amplitude of an m th order normalized derivative as function of scale is then given by

$$L_{\xi^m, max}(t) = t^{m\gamma/2} \omega_0^m e^{-\omega_0^2 t/2}, \quad (11)$$

i.e., it first increases and then decreases. Moreover, it assumes a unique maximum at $t_{max, L_{\xi^m}} = \gamma m / \omega_0^2$. If we define a scale parameter σ of dimension length by $\sigma = \sqrt{t}$ and introduce the wavelength λ_0 of the signal by $\lambda_0 = 2\pi / \omega_0$, we can see that the scale at which the amplitude of the γ -normalized derivative assumes its maximum over scales is *proportional* to the wavelength, λ_0 , of the signal:

$$\sigma_{max, L_{\xi^m}} = \frac{\sqrt{\gamma m}}{2\pi} \lambda_0. \quad (12)$$

The maximum value over scales is

$$L_{\xi^m, max}(t_{max, L_{\xi^m}}) = \frac{(\gamma m)^{\gamma m/2}}{e^{\gamma m/2}} \omega_0^{(1-\gamma)m}. \quad (13)$$

¹An analysis of scale-space like responses to sine waves corresponding to the case when $\gamma = 1$ in this section has also been performed in wavelet analysis by (Mallat and Hwang 1992); see section 9.3.

In the case when $\gamma = 1$, this maximum value is *independent* of the frequency of the signal (see figure 1), and the situation is highly symmetric, *i.e.*, given any scale t_0 , the maximally amplified frequency is given by $\omega_{max} = \sqrt{m/t_0}$, and for any ω_0 the scale with maximum amplification is $t_{max} = m/\omega_0^2$. In other words, for normalized derivatives with $\gamma = 1$ it holds that sinusoidal signals are treated in a similar (scale invariant) way independent of their frequency (see figure 1). The situation is a bit different when $\gamma \neq 1$. We shall return to this subject in section 4.1.

4 Proposed methodology for scale selection

The example above shows that the scale at which a normalized derivative assumes its maximum over scales is for a sinusoidal signal proportional to the wavelength of the signal. In this respect, maxima over scales of normalized derivatives reflect the scales over which spatial variations take place in the signal.

This operation corresponds to an interesting computational structure, since it constitutes a way of *estimating length* based on local measurements performed at only a *single spatial point* in the scale-space representation, and without explicitly laying out a ruler. Moreover, compared to a local windowed Fourier transform there is no need for making any explicit settings of window size for computing the Fourier transform. Instead, the propagation of length information over space is performed via the diffusion equation, and the decisions about the contents in the data are made by studying the output of derivative operators as the diffusion process evolves.

Alternatively, we can view such a measurement procedure as a pattern matcher, which matches Gaussian derivative kernels of different size to the given image pattern, based on a specific normalization of the primitive templates. By using the proposed γ -normalized derivative concept for normalization, we obtain one-to-one correspondence between the matching response of the Gaussian derivative kernels and the wavelength of the signal. Selecting the scale at which the maximum over scale is assumed corresponds to selecting the pattern (or the scale) for which the operator response is strongest.

This property is, however, not restricted to sine wave patterns or to image measurements in terms of linear derivative operators of a certain order. Contrary, it applies to a large class of image descriptors which can be formulated as multi-scale differential invariants expressed in terms of Gaussian derivatives (this notion will be made more precise next). A main message of this article is that this property can be used as

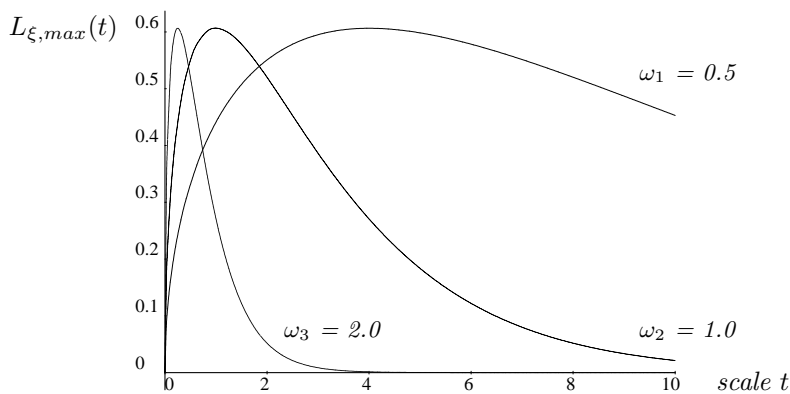


Figure 1: The amplitude of first order normalized derivatives as function of scale for sinusoidal input signals of different frequencies ($\omega_1 = 0.5$, $\omega_2 = 1.0$ and $\omega_3 = 2.0$).

a major mechanism in algorithms for automatic scale selection, which automatically adapt the local scales of processing to image data. Let us hence generalize the above-mentioned observation to more complex signals and state the following principle for scale selection, to be applied in situations when no other information is available. In its most general form, it can be expressed as follows:

Principle for scale selection:

In the absence of other evidence, assume that a scale level, at which some (possibly non-linear) combination of normalized derivatives assumes a local maximum over scales, can be treated as reflecting a characteristic length of a corresponding structure in the data.

This principle is closely related to although not equivalent to the method for scale selection in previously proposed in (Lindeberg 1991, 1993a), where interesting scale levels were determined from maxima over scales of a normalized blob measure. It can be theoretically justified under a number of different assumptions and for a number of specific brightness models (see next). Its general usefulness, however, must be verified empirically, and with respect to the type of problem it is to be applied to.

4.1 General scaling property of local maxima over scales

A basic justification for the abovementioned arguments can be obtained from the fact that for a large class of (possibly non-linear) combinations of normalized derivatives it holds that maxima over scales have a nice behaviour under rescalings of the intensity pattern. If the input image is rescaled by a constant scaling factor s , then the scale at which the maximum is assumed will be multiplied by the same factor (if measured in units of $\sigma = \sqrt{t}$). This is a fundamental requirement on a scale selection mechanism, since it guarantees that the image operations will commute with size variations.

Transformation properties under rescalings: To give a formal characterization of this scaling property, consider two signals f and f' related by

$$f(x) = f'(sx), \quad (14)$$

and define the scale-space representations of f and f' in the two domains by

$$L(\cdot; t) = g(\cdot; t) * f, \quad (15)$$

$$L'(\cdot; t') = g(\cdot; t') * f', \quad (16)$$

where the spatial variables and the scale parameters are transformed according to

$$x' = sx, \quad (17)$$

$$t' = s^2 t. \quad (18)$$

Then, L and L' are related by

$$L(x; t) = L'(x'; t'), \quad (19)$$

and the m th order spatial derivatives satisfy

$$\partial_{x^m} L(x; t) = s^m \partial_{x'^m} L'(x'; t'). \quad (20)$$

Finally, for γ -normalized derivatives defined in the two domains by

$$\partial_{\xi} = t^{\gamma/2} \partial_x, \quad (21)$$

$$\partial_{\xi'} = t'^{\gamma/2} \partial_{x'}, \quad (22)$$

we have that

$$\partial_{\xi^m} L(x; t) = s^{m(1-\gamma)} \partial_{\xi'^m} L'(x'; t'). \quad (23)$$

Perfect scale invariance when $\gamma = 1$: From this relation it can be seen that, when $\gamma = 1$ the normalized derivative concept leads to perfect scale invariance. The normalized derivatives are equal in the two domains, provided that the scale parameters and the spatial positions are matched according to (17) and (18). More specifically, local maxima over scales are always assumed at corresponding positions, and this scaling property holds for any differential expression defined from the local N -jet.

Sufficient scale invariance when $\gamma \neq 1$: The case when $\gamma \neq 1$ leads to a different type of structure, since we cannot preserve a scaling property for arbitrary combinations of normalized derivatives. Let us hence restrict the analysis to polynomial differential invariants which are homogeneous in the sense that the sum of the orders of differentiation is the same for each term in the polynomial. To express this notion compactly, introduce multi-index notation for derivatives by $L_{x^\alpha} = L_{x_1^{\alpha_1} x_2^{\alpha_2} \dots x_D^{\alpha_D}}$ where $x = (x_1, x_2, \dots, x_D)$, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_D)$ and $|\alpha| = \alpha_1 + \alpha_2 + \dots + \alpha_D$. Then, consider a homogeneous polynomial differential invariant $\mathcal{D}L$ of the form

$$\mathcal{D}L = \sum_{i=1}^I c_i \prod_{j=1}^J L_{x^{\alpha_{ij}}}, \quad (24)$$

where the sum of the orders of differentiation in a certain term

$$\sum_{j=1}^J |\alpha_{ij}| = M \quad (25)$$

does not depend on the index i of that term. For a differential expression of this form, the corresponding normalized differential expression in each domain is given by

$$\mathcal{D}_{\gamma\text{-norm}} L = t^{M\gamma/2} \mathcal{D}L, \quad (26)$$

$$\mathcal{D}'_{\gamma\text{-norm}} L' = t'^{M\gamma/2} \mathcal{D}'L'. \quad (27)$$

From (23) it follows that these normalized differential expressions are related by

$$\mathcal{D}_{\gamma\text{-norm}} L = s^{M(1-\gamma)} \mathcal{D}'_{\gamma\text{-norm}} L'. \quad (28)$$

Clearly, by γ -normalization with $\gamma = 1$, the magnitude of the derivative is not scale invariant. Local maxima over scales will, however, still be preserved, since

$$\partial_t (\mathcal{D}_{\gamma\text{-norm}} L) = 0 \quad \Leftrightarrow \quad \partial_{t'} (\mathcal{D}'_{\gamma\text{-norm}} L') = 0 \quad (29)$$

and the type of critical points are preserved under this transformation. Hence, even when $\gamma \neq 1$, we can achieve sufficient scale invariance to support the proposed scale selection methodology.

Scale compensated magnitude measures when $\gamma \neq 1$: When basing a scale selection methodology on $\gamma \neq 1$, there is, however, a minor complication which needs attention. When performing feature detection in practice, it is natural to associate a measure of feature strength with each detected feature. Specifically, the magnitude of the response at the local maximum over scales constitutes a natural entity to include in

such a measure. From the transformation property (23), it is, however, apparent that this magnitude measure will be strongly dependent on the scale at which the maximum over scales is assumed. Hence, the magnitude measure will depend on the feature size. In view of the scale invariant magnitude measure obtained using $\gamma = 1$, it is, however, straightforward to correct for this phenomenon by multiplying the response by a correction factor and to define a compensated magnitude measure by

$$\mathcal{M}_{\gamma\text{-norm}}L = t^{M(1-\gamma)/2} \mathcal{D}_{\gamma\text{-norm}}L. \quad (30)$$

Then, the magnitude measures in the two domains will satisfy

$$\mathcal{M}_{\gamma\text{-norm}}L(x; t) = \mathcal{M}'_{\gamma\text{-norm}}L'(x'; t'). \quad (31)$$

Necessity of the γ -normalization: More generally, one may ask what choices of normalization factors are possible, provided that we would like to state this scaling property as a fundamental constraint on a scale selection mechanism based on local maxima over scales of normalized differential entities. Then, in fact, it can be shown that the γ -normalized derivative concept according to (9) arises by *necessity*.

In other words, the γ -normalized derivative concept comprises the most general class of normalization factors for which detection of local maxima over scales commutes with rescalings of the input pattern. A more precise formulation of this statement as well as the details of the necessity proof can be found in appendix A.1.

Summary: Scale selection properties: To conclude, this analysis shows that if a γ -normalized homogeneous differential expression assumes a maximum over scales at $(x_0; t_0)$ in the scale-space representation of f , then there will be a corresponding maximum over scales in the scale-space representation of f' at $(s x_0; s^2 t_0)$. Moreover, although the magnitude of a normalized derivative at a local maximum over scales is not scale invariant unless $\gamma = 1$, it is possible to compensate for this phenomenon and to define scale invariant magnitude descriptors also when $\gamma \neq 1$.

4.2 The scale selection mechanism in practice

So far we have proposed a general methodology for scale selection by detecting local maxima in feature responses over scales. Whereas this approach constitutes an extension of the traditional way in which spatial features are detected from spatial maxima of feature responses, there is a fundamental difference. Since the image operators at different scales by necessity have to be of different size, the problem of normalizing the filter responses is of crucial importance. In section 4.1, we analysed this problem in detail and investigated the feasibility of capturing image structures under size variations. Specifically, we characterized general classes of differential invariants as well as normalization approaches, which allow for scale selection within the proposed computational structure. A fundamental problem that remains to be solved in this context concerns what differential expressions to use. Is any differential invariant feasible? Here, we shall not attempt to answer this question. Let us instead contend that the differential expression should at least be determined so as to capture the types of image structures under consideration.

The general approach to scale selection that will be proposed is to use these maximal responses over scales in the stage of *detecting image features*, i.e., when *establishing the existence* of different types of image structures. Basically, the scale at which a maximum over scales is attained will be assumed to give information about *how large* a feature is, in analogy with the common approach of taking the spatial

position at which the maximum operator response is assumed as an estimate of the spatial location of a feature. In certain situations, and as we shall see more specific examples of later, this implies that image features may be detected at quite coarse scales, and the localization properties may not be the best. Therefore, we propose to complement this framework by a second processing stage, in which more refined processing is invoked for computing more accurate localization estimates. In this respect, the suggested framework naturally gives rise to *two-stage algorithms*, with feature detection at coarse scales followed by feature localization at finer scales. Whereas coarse-to-fine approaches are common practice in several computer vision problems, a notable aspect of this approach will be that we include explicit mechanisms for automatic selection of all scale parameters.

In the following, we shall shift attention to the application of the abovementioned general ideas to specific problems. A series of theoretical and experimental results will be presented showing how the proposed approach applies to different types of feature detectors expressed as polynomial combinations of Gaussian derivatives.

4.3 Experiments: Scale-space signatures from real data

Figure 2 shows the variations over scales of two simple differential expressions formulated in terms of normalized derivatives. It shows the result of computing the trace and the determinant of the normalized Hessian matrix by (with $\gamma = 1$)

$$\text{trace } \mathcal{H}_{\text{norm}} L = t^\gamma \nabla^2 L = t^\gamma (L_{xx} + L_{yy}), \quad (32)$$

$$\det \mathcal{H}_{\text{norm}} L = t^{2\gamma} (L_{xx}L_{yy} - L_{xy}^2), \quad (33)$$

for two details in an image of a field of sunflowers. (To avoid the sensitivity to sign of these entities, and hence the polarity of the signal, $\text{trace } \mathcal{H}_{\text{norm}} L$ and $\det \mathcal{H}_{\text{norm}} L$ have been squared before presentation.) These graphs are called the *scale-space signatures* of $(\text{trace } \mathcal{H}_{\text{norm}} L)^2$ and $(\det \mathcal{H}_{\text{norm}} L)^2$, respectively.

As can be seen, the maximum over scales in the top row of figure 2 is assumed at a finer scale than in the bottom row. A more detailed examination the ratio between the scale values² where the graphs attain their maxima over scales shows that when the scale parameter is measured in dimension length this scale ratio is roughly equal to the ratio of the diameters of the sunflowers in the centers of the two images, respectively. This example illustrates that results in agreement with the proposed scale selection principle can be obtained also for real-world data (and for signals having a much richer frequency content than a single sine wave).

The reason why these particular differential expressions have been selected here is because they constitute differential entities useful for blob detection; see *e.g.* (Marr 1982; Voorhees and Poggio 1987; Blostein and Ahuja 1989). Before we turn to the problem of expressing an integrated blob detector with automatic scale selection, however, let us describe a further extension of the general scale selection idea.

4.4 Simultaneous detection of interesting points and scales

In figure 2, the signatures of the normalized differential entities were computed at the central point in each image. These points were deliberately chosen to coincide

² In the graphs in figure 2 the scale parameter (on the horizontal axis) is measured in terms of *effective scale*, τ . For continuous signals, this parameter is essentially the logarithm of the scale parameter $\tau = C_1 \log t + C_2$ for some $C_1, C_2 > 0$. To avoid the singularity at zero scale, however, all experiments are based on an effective scale concept especially developed for discrete signals and defined such that $\tau \sim \log t$ at coarse scales and $\tau \sim t$ at fine scales see (Lindeberg 1994d).

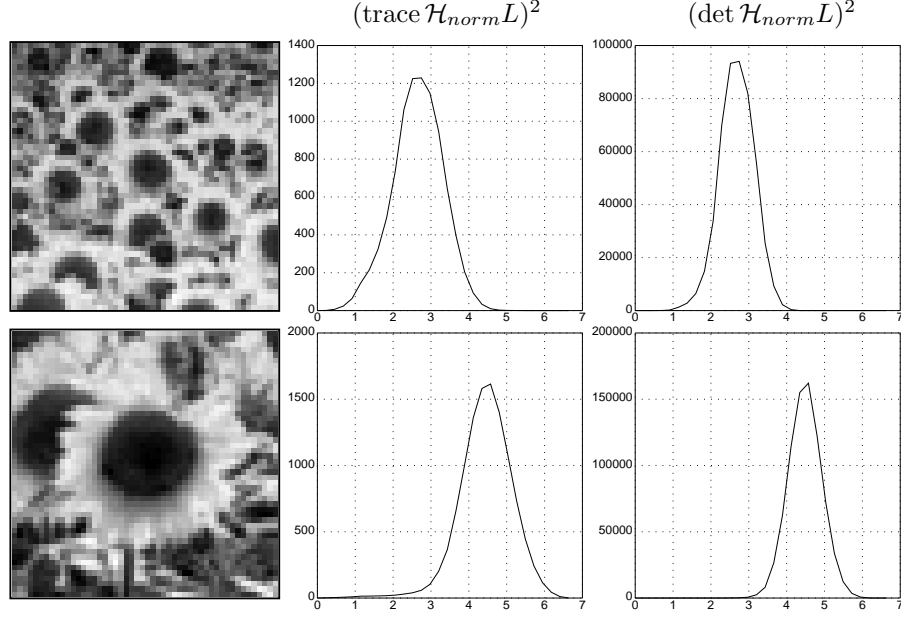


Figure 2: Scale-space signatures of the trace and the determinant of the normalized Hessian matrix computed for two details of a sunflower image; (left) grey-level image, (middle) signature of $(\text{trace } \mathcal{H}_{\text{norm}} L)^2$, (right) signature of $(\det \mathcal{H}_{\text{norm}} L)^2$. (The signature have been computed at the central point in each image. The horizontal axis shows effective scale (essentially the logarithm of the scale parameter), whereas the scaling of the vertical axis is linear in the normalized operator response.)

with the centers of the sunflowers, where the blob response can be expected to be maximal under spatial perturbations. In a real-world vision situation, however, we cannot assume such points to be known *a priori*. Moreover, we can expect that the spatial maximum of the operator response is assumed at different positions at different scales. This is one example of the well-known fact that scale-space smoothing leads to shape distortions.

Therefore, a more general approach to scale selection from local extrema in the scale-space signature is by accumulating the signature of any normalized differential entity $\mathcal{D}_{\text{norm}} L$ along the path $r: \mathbb{R}_+ \rightarrow \mathbb{R}^N$ that a local extremum in $\mathcal{D}_{\text{norm}} L$ describes across scales. The mathematical framework for describing such paths is described in (Lindeberg 1994d). Formally, an *extremum path* of a differential entity $\mathcal{D}_{\text{norm}} L$ is defined (from the implicit function theorem) as a set of points $(r(t); t) \in \mathbb{R}^N \times \mathbb{R}_+$ in scale-space such that for any $t \in \mathbb{R}_+$ the point $r(t)$ is a local extremum of the mapping $x \mapsto (\mathcal{D}_{\text{norm}} L)(x; t)$,

$$\{(x; t) \in \mathbb{R}^N \times \mathbb{R}_+\} = \{(r(t); t) \in \mathbb{R}^N \times \mathbb{R}_+ : (\nabla(\mathcal{D}_{\text{norm}} L))(r(t); t) = 0\}.$$

At the point at which an extremum in the signature is assumed, the derivative along the scale direction is zero as well. Hence, it is natural to define a *normalized scale-space extremum* of a differential entity $\mathcal{D}_{\text{norm}} L$ as a point $(x_0; t_0) \in \mathbb{R}^N \times \mathbb{R}_+$ in scale-space which is *simultaneously* a local extremum with respect to both the spatial

coordinates and the scale parameter.³ In terms of derivatives, such points satisfy

$$\begin{cases} (\nabla(\mathcal{D}_{norm}L))(x_0; t_0) = 0, \\ (\partial_t(\mathcal{D}_{norm}L))(x_0; t_0) = 0. \end{cases} \quad (34)$$

These normalized scale-space extrema constitute natural generalizations of extrema with respect to the spatial coordinates, and can serve as natural interest points for feature detectors formulated in terms of spatial maxima of differential operators, such as blob detectors, junction detectors, symmetry detectors, etc. Specific examples of this idea will be worked out in more detail in the following sections.⁴

Referring to the invariance properties of local maxima over scales under rescalings of the input signal, we can observe that they transfer trivially to scale-space maxima. Hence, if a normalized scale-space maximum is assumed at $(x_0; t_0)$ in the scale-space representation of a signal f , then in a rescaled signal f' defined by $f'(sx) = f(x)$, a corresponding scale-space maximum is assumed at $(sx_0; s^2t_0)$ in the scale-space representation of f' .

5 Blob detection with automatic scale selection

Figure 3 shows the result of detecting normalized scale-space extrema of the normalized Laplacian in an image of a sunflower field. Every scale-space maximum has been graphically illustrated by a circle centered at the point at which the spatial maximum is assumed, and with the size determined such that the radius (measured in pixel units) is proportional to the scale at which the maximum is assumed (measured in dimension length). To reduce the number of blobs, a threshold on the maximum normalized response has been selected such that the 250 blobs having the maximum normalized responses according to (30) remain.

The bottom row shows the result of superimposing these circles onto a bright copy of the original image, as well as corresponding results for the normalized scale-space extrema of the square of the determinant of the Hessian matrix. Corresponding experiments for a synthetic pattern (analysed in section 5.1) are given in figure 4. Observe how these conceptually very simple differential geometric descriptors give a very reasonable description of the blob-like structures in the image (in particular concerning the blob size) considering how little information is used in the processing.

Figure 5 shows a three-dimensional illustration of the multi-scale blob descriptors computed from the sunflower image. Here, each scale-space maximum has been visualized by a sphere centered at the position $(x_0; t_0)$ in scale-space at which the maximum was assumed, with the radius proportional to the selected scale, and the brightness increasing with the significance of the blob. Observe how the size variations in the image data are reflected in the spatial variations of the image descriptors.

5.1 Analysis of scale-space maxima for idealized model patterns

Whereas the theoretical analysis in section 4.1 applies generally to large classes of differential invariants and input signals, one may ask how the scale selection method for blob detection performs in specific situations. In this section, we shall study two

³When detecting scale-space maxima in practice, there is, of course, no need to explicitly track the extrema along the extremum path in scale-space. It is sufficient to detect three-dimensional maxima over space and scale (as described in more detail in section A.4.3).

⁴Further extensions of this idea are also explored in (Lindeberg 1996a), where differential definitions of edges and ridges are expressed in such a way that scale selection constitutes an integrated part of the feature definition.

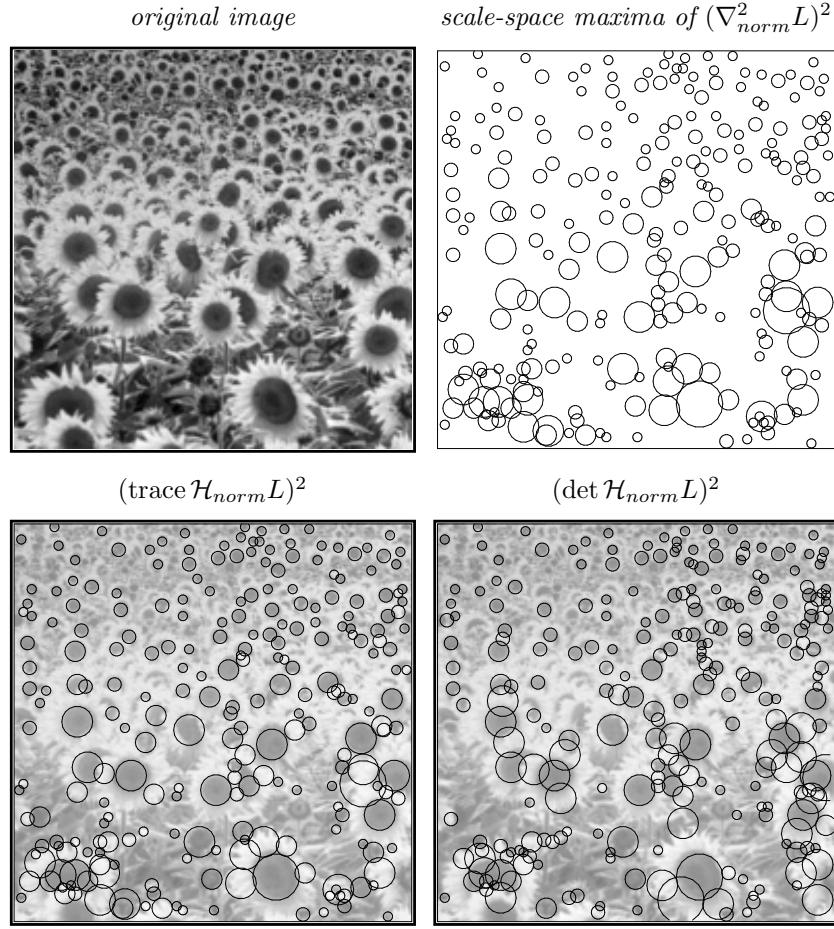


Figure 3: Normalized scale-space maxima computed from an image of a sunflower field: (top left): Original image. (top right): Circles representing the 250 normalized scale-space maxima of $(\text{trace } \mathcal{H}_{\text{norm}} L)^2$ having the strongest normalized response. (bottom left): Circles representing scale-space maxima of $(\text{trace } \mathcal{H}_{\text{norm}} L)^2$ superimposed onto a bright copy of the original image. (bottom right): Corresponding results for scale-space maxima of $(\det \mathcal{H}_{\text{norm}} L)^2$.

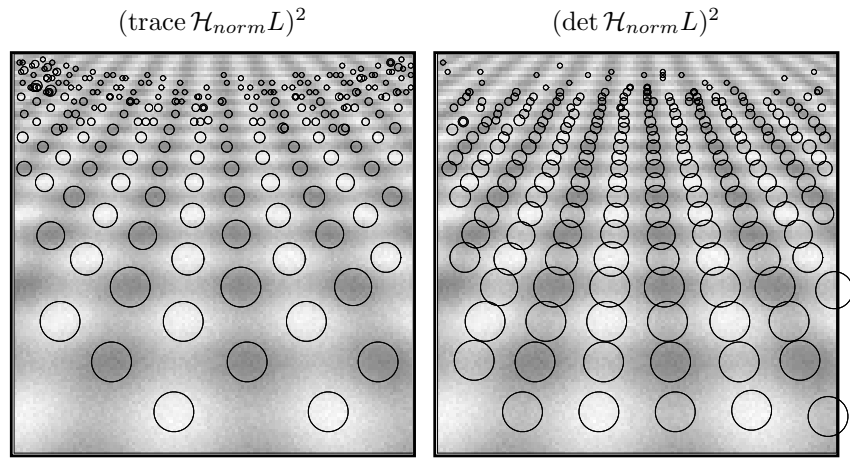


Figure 4: The 250 most significant normalized scale-space extrema detected from the perspective projection of a sine wave of the form (with 10% added Gaussian noise).

model patterns for which a closed-form solution of diffusion equation can be calculated and a complete analytical study hence is feasible.

Example 1. Consider first a non-symmetric Gaussian function

$$f(x_1, x_2) = g(x_1; t_1) g(x_2; t_2) = \frac{1}{\sqrt{2\pi t_1}} e^{-x_1^2/2t_1} \frac{1}{\sqrt{2\pi t_2}} e^{-x_2^2/2t_2}$$

as a model of a two-dimensional blob with characteristic lengths $\sqrt{t_1}$ and $\sqrt{t_2}$ along the coordinate directions. From the semi-group property of the Gaussian kernel $g(\cdot; t_A) * g(\cdot; t_B) = g(\cdot; t_A + t_B)$ it follows that the scale-space representation L of f is

$$L(x_1, x_2; t) = g(x_1; t_1 + t) g(x_2; t_2 + t). \quad (35)$$

After a few algebraic manipulations it can be shown that for any $t_1, t_2 > 0$ there is a unique maximum over scales in

$$|(\nabla_{norm}^2 L)(0, 0; t)| = \frac{t(t_1 + t_2 + 2t)}{2\pi(t_1 + t)^{3/2}(t_2 + t)^{3/2}}. \quad (36)$$

In the case when $t_1 = t_2 = t_0$, this maximum over scales is given by

$$\partial_t(\nabla_{norm}^2 L)(0, 0; t) = 0 \iff t = t_0. \quad (37)$$

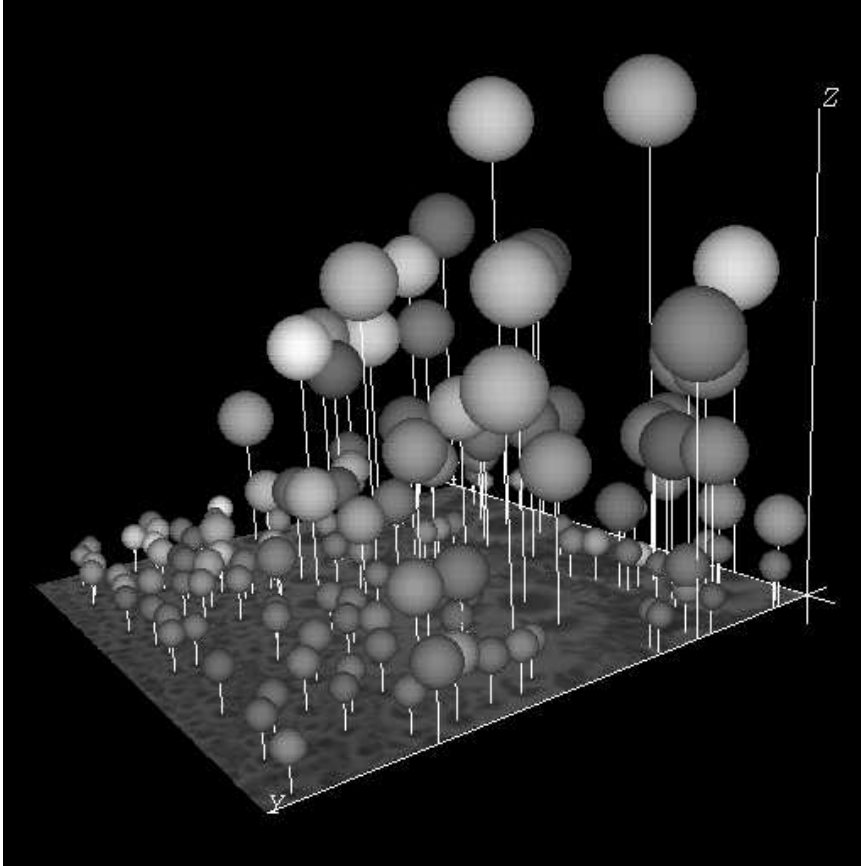


Figure 5: Three-dimensional view of the 150 strongest scale-space maxima of the square of the normalized Laplacian of the Gaussian computed from the sunflower image. (A dark copy of the original grey-level image is shown in the ground plane, and the vertical dimension represents scale.)

The closed-form solution for the maximum over scales in

$$|\det \mathcal{H}_{norm}(0, 0; t)| = \frac{t^2}{4\pi^2(t_1 + t)^2(t_2 + t)^2} \quad (38)$$

is simple. It is assumed at

$$t_{\det \mathcal{H}L} = \sqrt{t_1 t_2}, \quad (39)$$

verifying that for both trace $\mathcal{H}_{norm}L$ and $\det \mathcal{H}_{norm}L$ the scale at which the scale-space maximum is assumed reflects a characteristic size of the blob.

Example 2. Another interesting special case to consider is a periodic signal defined as the sum of two perpendicular sine waves,

$$f(x, y) = \sin \omega_1 x + \sin \omega_2 y \quad (\omega_1 \leq \omega_2). \quad (40)$$

Its scale-space representation is

$$L(x, y; t) = e^{-\omega_1^2 t/2} \sin \omega_1 x + e^{-\omega_2^2 t/2} \sin \omega_2 y, \quad (41)$$

and $\nabla_{norm}^2 L$ and $\det \mathcal{H}_{norm}L$ assume their spatial maxima at $(\pi/2, \pi/2)$. Setting the derivative $\partial_t |(\nabla_{norm}^2 L)(\frac{\pi}{2}, \frac{\pi}{2}; t)| = \partial_t (t(\omega_1^2 e^{-\omega_1^2 t/2} + \omega_2^2 e^{-\omega_2^2 t/2}))$ to zero gives

$$\omega_1^2(2 - \omega_1^2 t) e^{-\omega_1^2 t/2} + \omega_2^2(2 - \omega_2^2 t) e^{-\omega_2^2 t/2} = 0. \quad (42)$$

There is a unique solution when the ratio ω_2/ω_1 is close to one, and three solutions when the ratio is sufficiently large. Hence, there is a unique maximum over scales when ω_2/ω_1 is close to one, and two maxima when the ratio is sufficiently large. (The bifurcation occurs when $\omega_2/\omega_1 \approx 2.4$.) In the special case when $\omega_1 = \omega_2 = \omega_0$, the maximum over scales is assumed at

$$t_{\text{trace } \mathcal{H}L} = \frac{2}{\omega_0^2}. \quad (43)$$

Similarly, setting $\partial_t |(\det \mathcal{H}_{norm}L)(\pi/2, \pi/2; t)| = \partial_t (t^2 \omega_1^2 e^{-\omega_1^2 t/2} \omega_2^2 e^{-\omega_2^2 t/2})$ to zero gives that the maximum over scales in $\det \mathcal{H}_{norm}L$ is assumed at

$$t_{\det \mathcal{H}L} = \frac{4}{\omega_1^2 + \omega_2^2}. \quad (44)$$

Hence, for both the Gaussian blob model and the periodic sine waves, these specific results agree with the suggested general scale selection principle. When the scale parameter is measured in units of $\sigma = \sqrt{t}$, the scale levels, at which the maxima over scales are assumed, reflect the characteristic length of the structures in the signal.

5.2 Comparisons with fixed-scale blob detection

In view of the results presented so far, it is interesting to compare this blob detector with automatic scale selection to a standard multi-scale blob detector operating at a fixed scale. Figure 6 shows the result of computing spatial maxima at different scales in the response of the Laplacian operator from the sine wave pattern in figure 4. At each scale, the 50 strongest responses have been extracted.

As can be seen, small blobs are given the highest relative ranking at fine scales, whereas large blobs are given the highest relative ranking at coarse scales. Hence, a

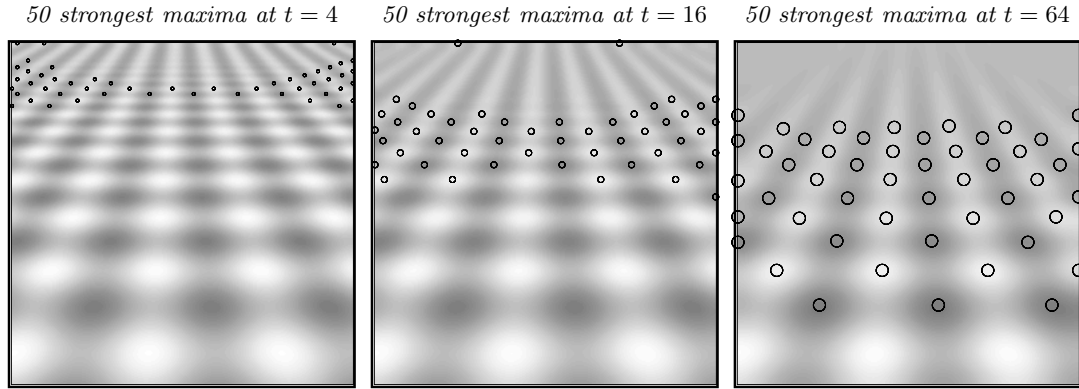


Figure 6: The 50 strongest spatial responses to the Laplacian operator computed at the scale levels: (a) $t = 4.0$, (b) $t = 16.0$, and (c) $t = 64.0$. Observe how this blob detector leads to a bias towards image structures of a certain size.

blob detector of this type (operating at a single predetermined scale) induces a bias towards image structures of a certain size. On the other hand, if we use the proposed methodology for blob detection based on the detection of scale-space maxima, we obtain a conceptually clean way of handling image structures of all sizes (between the inner scale and the outer scale of the analysis) in a similar manner. (As was shown above, the associated measure of blob strength is strictly scale invariant.)

5.3 Applications of blob detection with automatic scale selection

Following the previously presented arguments, we argue that a scale selection mechanism is an essential complement to any blob detector aimed at handling large size variations in the image structures. In addition, scale information associated with such adaptively computed image descriptors may serve as an important cue in its own right.

In (Bretzner and Lindeberg 1996, 1998) an application to feature tracking is presented, where (i) the scale information constitutes a key component in the criterion for matching image features over time, and (ii) the scale selection mechanism is essential for the vision system to be able to capture objects under large size variations over time.

In (Lindeberg and Gårding 1993; Gårding and Lindeberg 1996) an extension of this general blob detection idea is presented, where: (i) each scale-space maximum is used for guiding the computation of a regional image texture descriptor (a second moment matrix) as a pre-processing stage to shape-from-texture, (ii) the shape of each blob is represented by an ellipse with its shape determined from the local statistics of image gradient directions, and (iii) the scale information is used as a cue to three-dimensional surface shape when it can be assumed that the texture elements on the surface have the same size.

6 Junction detection with automatic scale selection

A similar approach as was used for blob detection in previous section can be used for detecting corners in grey-level images. In this section, it will be shown how a multi-scale junction detector can be formulated in terms of the scale-space maxima of a normalized differential invariant.

6.1 Selection of detection scales from normalized scale-space maxima

A commonly used entity for junction detection is the curvature of level curves in intensity data multiplied by the gradient magnitude (Kitchen and Rosenfeld 1982; Dreschler and Nagel 1982; Koenderink and Richards 1988; Noble 1988; Deriche and Giraudon 1990; Blom 1992; Florack *et al.* 1992; Lindeberg 1994d). A special choice is to multiply the level curve curvature by the gradient magnitude raised to the power of three. This is the smallest value of the exponent that leads to a polynomial expression

$$\tilde{\kappa} = L_v^2 L_{uu} = L_{x_2}^2 L_{x_1 x_1} - 2L_{x_1} L_{x_2} L_{x_1 x_2} + L_{x_1}^2 L_{x_2 x_2}. \quad (45)$$

Moreover, the spatial maxima of this operator are invariant under affine transformations. The corresponding normalized differential expression is obtained by replacing each derivative operator ∂_{x_i} by $t^{\gamma/2} \partial_{x_i}$, which gives

$$\tilde{\kappa}_{norm} = t^{2\gamma} \tilde{\kappa}. \quad (46)$$

Figure 7 shows the results of applying this operator to a grey-level image at a number of different scales. The results are displayed in two ways; (i) in terms of grey-level images showing the scale-space representation L as well as the junction response $\tilde{\kappa}^2$ computed at each scale, and (ii) in terms of the 50 strongest spatial maxima of $\tilde{\kappa}^2$, respectively, extracted at the same scale levels. As can be seen, qualitatively different responses are obtained at different scales. At fine scales, the strongest responses are obtained for the sharp corners and for a number of spurious fine-scale perturbations along edges. Then, with increasing values of the scale parameter, the selectivity to junction-like structures increases. In particular, the diffuse (non-sharp) and strongly rounded corners only give rise to strong responses at coarse scales. In summary, this example illustrates the following fundamental aspects of multi-scale corner detection:

- If we are only interested in *sharp ideal corners*, *i.e.*, corners which can be well approximated by straight lines and for which the intensity contrast across the edge is high and corresponds to a step edge, then it is often sufficient to use a fine scale in the detection stage. The main motivation for using coarser scales on such data is to reduce the number of false positives.
- If we are interested in capturing rounded corners and corners for which the intensity variations across the edges around the junction are slow (diffuse corners), then it is essentially necessary to use a coarse scale if we want to have strong spatial maxima in the response of $\tilde{\kappa}^2$ at such image structures.

Specifically, this example shows that if no *a priori* information is available about what can be expected to be in the scene, then a mechanism for automatic scale selection is essential to capture corner structures at different scales.

Figure 8 shows the result of including such a scale selection mechanism in the junction detector. It shows the result of detecting the 50 strongest normalized scale-space maxima of $\tilde{\kappa}_{norm}^2$ from the same grey-level image. Each scale-space maximum has been graphically illustrated by a circle centered at the point at which the maximum is assumed, and with the size determined such that the radius is proportional to the scale at which the maximum over scales was assumed (measured in dimension length). To reduce the number of junction candidates, the scale-space maxima have been sorted with respect to a saliency measure. This measure has been determined as the magnitude of the normalized response according to (30) multiplied by the scale parameter measured in dimension area, so as to approximate the area of the spatial

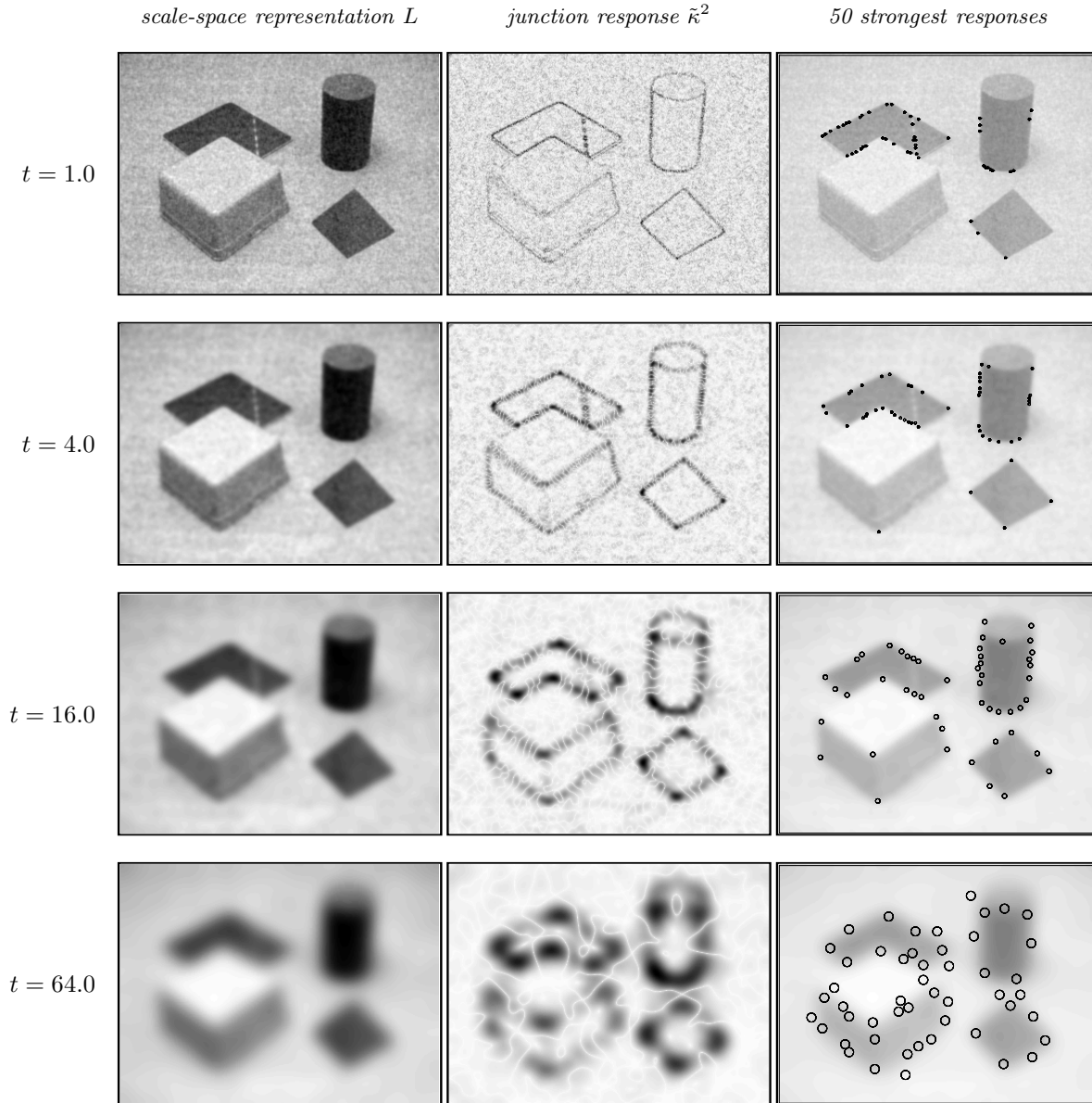


Figure 7: Junction responses at different scales computed from a noisy image containing a number of ideal sharp corners as well as rounded and diffuse corners. As can be seen, different types of junction structures give rise to different types of responses at different scales. Notably, certain diffuse junction structures fail to give rise to dominant responses at the finest levels of scale. (Image size: 320*240 pixels.)

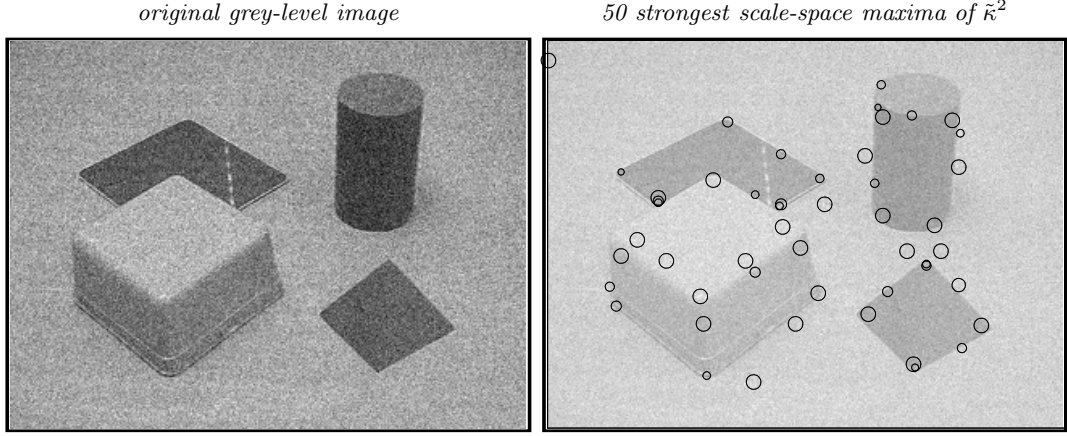


Figure 8: Junction detection with automatic scale selection: The result of computing the 50 most significant normalized scale-space extrema of $\tilde{\kappa}_{norm}^2$ from a grey-level image containing sharp straight edges as well as diffuse and rounded edges (with $\gamma = 1$). Compare with figure 7 and observe how corner structures at different scales are captured by this operation.

support region of each scale-space maximum. Finally, the 50 most significant blobs according to this ranking have been displayed.

Of course, thresholding on the magnitude of the operator response constitutes a coarse selective mechanism for feature detection. Nevertheless, note that this operation gives rise to a set of junction candidates with reasonable interpretations in the scene. Moreover, observe that the circles representing the scale-space extrema constitute natural regions of interest around the candidate junctions. In particular, the selected scales reflect the diffuseness and the spatial extent of the corners, such that coarser scales are, in general, selected for the diffuse corners than for sharp ones.

6.2 Analysis of scale-space maxima for diffuse junction models

To obtain an intuitive understanding of the qualitative behaviour of the scale selection method in this case, let us analyse a simple junction model for which a closed-form analysis can be carried out without too much effort.

Diffuse step junction. Consider

$$f(x_1, x_2) = \Phi(x_1; t_0) \Phi(x_2; t_0) \quad (47)$$

as a simple model of a *diffuse L-junction*, where $\Phi(\cdot; t_0)$ describes a *diffuse step edge*

$$\Phi(x_i; t_0) = \int_{x'=-\infty}^{x_i} g(x'; t_0) dx' \quad (48)$$

with *diffuseness* t_0 . From the semi-group property of the Gaussian kernel it follows that the scale-space representation L of f is

$$L(x_1, x_2; t) = \Phi(x_1; t_0 + t) \Phi(x_2; t_0 + t). \quad (49)$$

After differentiation, and using the fact that $L_{x_1 x_1} = 0$ and $L_{x_2 x_2} = 0$ at the origin, as well as $\Phi(0; t) = 1/2$, $\forall t$ and $g(0; t) = 1/\sqrt{2\pi t}$, we obtain

$$|\tilde{\kappa}_{norm}(0, 0; t)| = |2t^{2\gamma} L_{x_1} L_{x_2} L_{x_1 x_2}| = \frac{t^{2\gamma}}{8\pi^2(t_0 + t)^2}. \quad (50)$$

When $\gamma = 1$, this entity increases *monotonically* with scale, whereas for $\gamma \in]0, 1[$, $\tilde{\kappa}_{norm}(0, 0; t)$ assumes a unique maximum over scales at

$$t_{\tilde{\kappa}} = \frac{\gamma}{1 - \gamma} t_0. \quad (51)$$

Non-uniform Gaussian blob. A limitation of the abovementioned analysis is that the signature is computed at a fixed point, whereas the maximum in $\tilde{\kappa}^2$ can be expected to drift due to scale-space smoothing. Unfortunately, the equation that determines the position of the spatial maximum in $\tilde{\kappa}^2$ over scales is non-trivial to handle (it contains a non-linear combination of the Gaussian function, the primitive function of the Gaussian, and polynomials). Carrying out a closed-form analysis along non-vertical extremum paths is, however, straightforward for the previously treated non-uniform Gaussian blob model. This function can be regarded as a coarse model of the behaviour at so coarse scales in scale-space that the shape distortions are substantial and the overall shape of a finite-size object is severely affected. From (35) we have that the scale-space representation of the non-uniform Gaussian blob is

$$L(x_1, x_2; t) = g(x_1; t_1 + t) g(x_2; t_2 + t). \quad (52)$$

Differentiation and insertion into (45) shows that the absolute value of the rescaled level curve curvature assumes its spatial maximum

$$|\tilde{\kappa}_{norm}|_{max} = \frac{t^{2\gamma}}{12e\pi^3(t_1 + t)^{5/2}(t_2 + t)^{5/2}} \quad (53)$$

on the ellipse $\frac{3x_1^2}{2(t_1+t)} + \frac{3x_2^2}{2(t_2+t)} = 1$. In the special case when $\gamma = 1$, the maximum over scales is assumed at

$$t_{\tilde{\kappa}} = \frac{(t_1 + t_2)}{12} \left(\sqrt{1 + \frac{96 t_1 t_2}{(t_1 + t_2)^2}} - 1 \right), \quad (54)$$

whereas when $t_1 = t_2 = t_0$

$$t_{\tilde{\kappa}} = \frac{2\gamma}{5 - 2\gamma} t_0. \quad (55)$$

Interpretation of the qualitative behaviour. To conclude, the junction response $\tilde{\kappa}_{norm}^2$ can for $\gamma = 1$ be expected to increase with scales when a single corner model of infinite extent constitutes a reasonable approximation. On the other hand, $\tilde{\kappa}_{norm}^2$ can be expected to decrease with scales when so much smoothing is applied that the overall shape of the object is substantially distorted (and neighbouring junctions interfere with each other or disappear altogether).

Hence, selecting scale levels (and spatial points) where $\tilde{\kappa}_{norm}^2$ assumes maxima over scales can be expected to give rise to scale levels in the intermediate scale range (where a finite extent junction model constitutes a reasonable approximation). In particular, this approach will lead to larger scale values for corners having large spatial extent, and prevent too fine scales from being selected at rounded junctions.

6.3 Experiments: Scale-space signatures in junction detection

Figure 9 illustrates these effects for synthetic *L*-junctions with varying degrees of diffuseness. It shows simulation experiments with scale-space signatures of $\tilde{\kappa}_{norm}$ accumulated in two different ways: (i) a *vertical signature* obtained by computing $\tilde{\kappa}_{norm}$

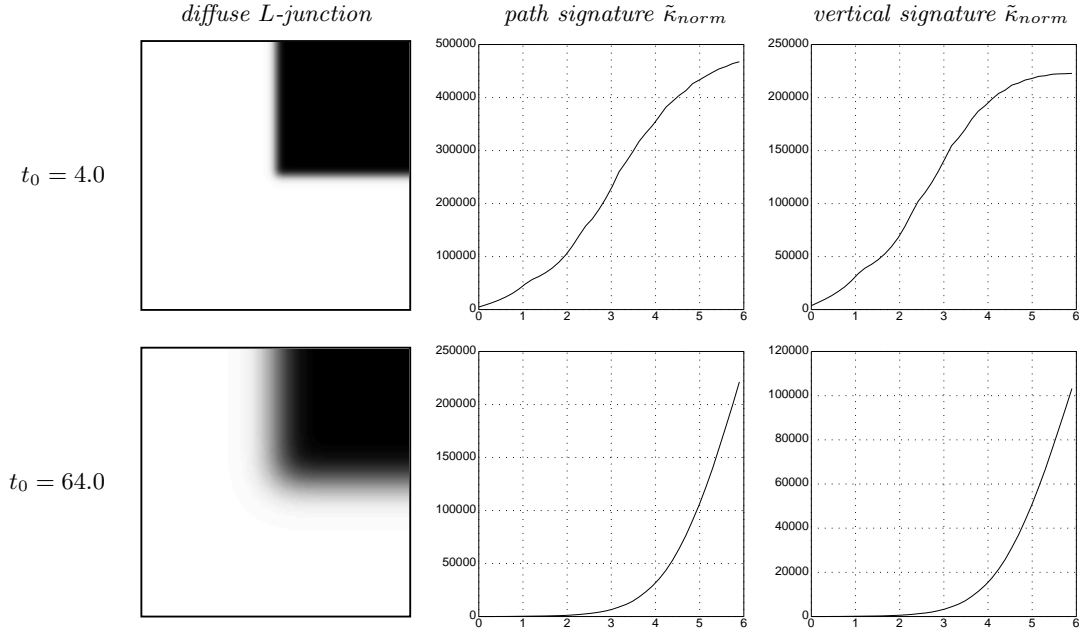


Figure 9: Scale-space signatures of $\tilde{\kappa}_{norm}$ for synthetic L -junctions with different degrees of diffuseness (top $t = 4.0$, bottom $t = 64.0$). (left) original grey-level image, (middle) path signature of $\tilde{\kappa}_{norm}$ accumulated by tracking a spatial maximum in $\tilde{\kappa}_{norm}$ across scales, (right) vertical signature of $\tilde{\kappa}_{norm}$ accumulated at the central point.

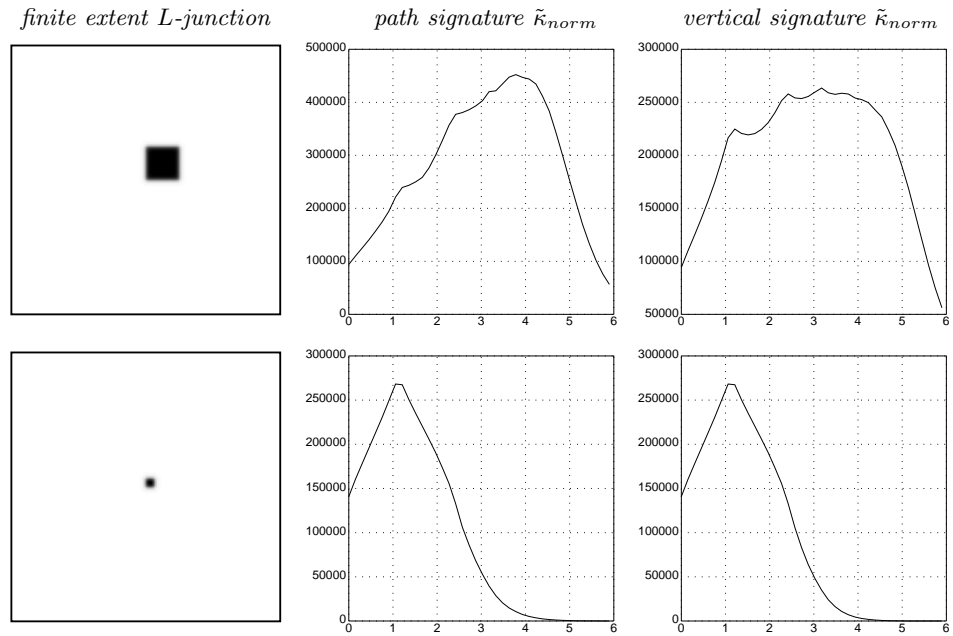


Figure 10: Scale-space signatures of $\tilde{\kappa}_{norm}$ for diffuse L -junctions ($t_0 = 1.0$) of different spatial extent ($1/4$ and $1/16$ of the image size). (left) original grey-level image, (middle) path signature of $\tilde{\kappa}_{norm}$ accumulated by tracking a spatial maximum in $\tilde{\kappa}_{norm}$ across scales, (right) vertical signature of $\tilde{\kappa}_{norm}$ accumulated at the central point.

at the fixed central point at different scales, and (ii) a *path signature* obtained by tracking the spatial extremum in $\tilde{\kappa}_{norm}$ across scales. As can be seen, the qualitative behaviour is in agreement with the approximate analysis in previous section—with increasing degree of diffuseness the values of $\tilde{\kappa}_{norm}$ become smaller at fine scales.

Figure 10 shows the result of replacing the infinite extent L -junction model by junction models of finite size. Observe that the peak in the signature is assumed at finer scales when the spatial extent of the junction is decreased. In other words, the scale at which the maximum over scales is assumed indicates the spatial extent (the size) of the region for which a junction model is consistent with the grey-level data (in agreement with the suggested scale selection principle).

Figure 11 gives a three-dimensional illustration of this junction detector with automatic scale selection. It shows scale-space maxima of $\tilde{\kappa}_{norm}^2$ computed from a synthetic image containing corner structures at different scales. The original grey-level image is shown in the ground plane, and each scale-space maximum has been graphically visualized by a sphere centered at the position $(x_0; t_0)$ in scale-space at which the scale-space maximum was assumed. (Hence, the height over the image plane reflects the selected scale.) Observe how the large scale corner as a whole gives rise to a response at coarse scales, whereas the superimposed corner structures of smaller size give rise to scale-space maxima at finer scales.

More results on corner detection, including a complementary mechanism for accurate corner localization, are presented in section 7.

7 Feature localization with automatic scale selection

The scale selection methodology presented so far applies to the detection of image features, and the role of the scale selection mechanism is to estimate the approximate size of the image structures the feature detector responds to. Whereas this approach provides a conceptually simple way to express various feature detectors, such as a junction detector, which automatically adapts its scale levels to the local image structure, it is not guaranteed that the spatial positions of the scale-space maxima constitute accurate estimates of the corner locations. The local maxima over scales may be assumed at rather coarse scales, where the drift due to scale-space smoothing is substantial and adjacent features may interfere with each other. For this reason, it is natural to complement the initial feature detection step by an explicit *feature localization* stage.

The subject of this section is show how mechanism for automatic scale selection can be formulated in this context, *by minimizing normalized measures of inconsistency over scales*.

7.1 Corner localization by local consistency

Second stage computation of localization estimate. Given an approximate estimate x_0 of the location and the size s of a corner (computed according to section 6), an improved estimate of the corner position can be computed as follows: Following (Förstner and Gülch 1987), consider at every point $x' \in \mathbb{R}^2$ in a neighbourhood of a junction candidate x_0 , the line $l_{x'}$ perpendicular to the gradient vector $(\nabla L)(x') = (L_{x_1}, L_{x_2})^T(x')$ at that point:

$$D_{x'}(x) = ((\nabla L)(x'))^T(x - x') = 0. \quad (56)$$

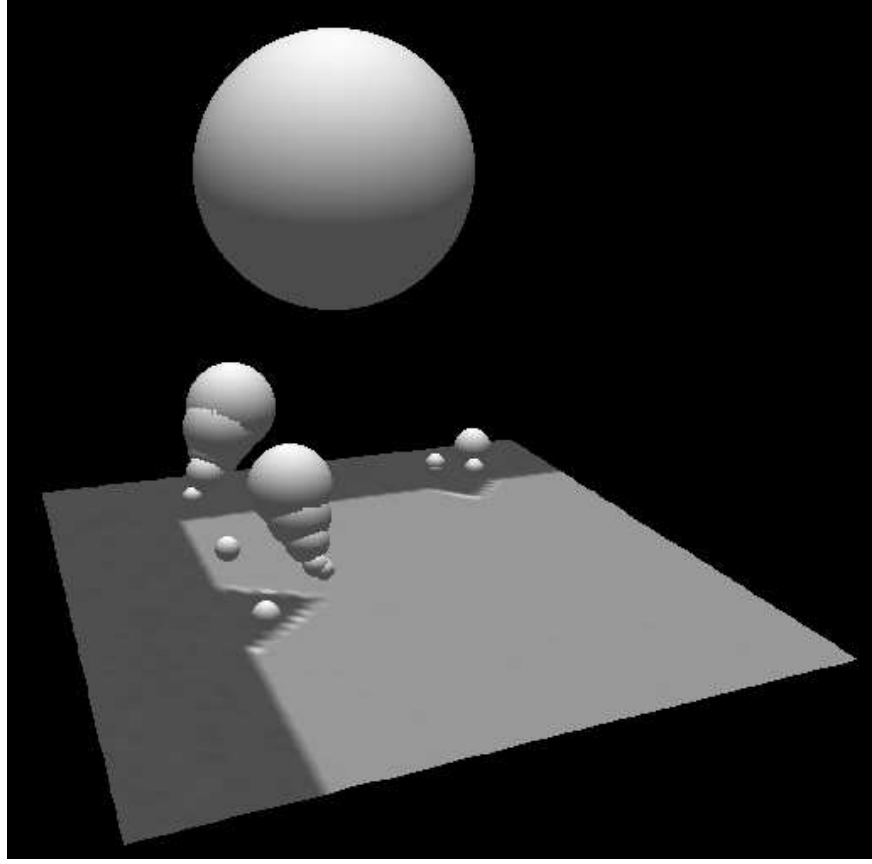


Figure 11: Three-dimensional view of scale-space maxima of $\tilde{\kappa}_{norm}^2$ computed for a large scale corner with superimposed corner structures at finer scales. Observe that a coarse scale response is obtained for the large scale corner structure as a whole, whereas the superimposed corner structures of smaller size give rise to scale-space maxima at finer scales.

Then, minimize the perpendicular distance to all lines $l_{x'}$ in a neighbourhood of x_0 , *i.e.* determine the point $x \in \mathbb{R}^2$ that minimizes

$$\min_{x \in \mathbb{R}^2} \int_{x' \in \mathbb{R}^2} (D_{x'}(x))^2 w_{x_0}(x') dx' \quad (57)$$

for some window function $w_{x_0}: \mathbb{R}^2 \rightarrow \mathbb{R}$ centered at the candidate junction x_0 . Minimizing this expression corresponds to finding the point x that minimizes the weighted integral of the squares of the distances from x to all $l_{x'}$ in the neighbourhood, see figure 12. ($D_{x'}(x)$ is distance from x to $l_{x'}$ multiplied by the gradient magnitude, and the window function implies that stronger weights are given to points in a neighbourhood of x_0 .) The overall intention of this formulation is that for an image pattern containing a junction, the point x that minimizes (57) should constitute a better estimate of the projection of the physical junction than x_0 .

Explicit solution in terms of local image statistics. An attractive property of the formulation in (57) is that it allows for a compact closed-form solution. After expansion, it can be written

$$\min_{x \in \mathbb{R}^2} \int_{x' \in \mathbb{R}^2} (x - x')^T ((\nabla L)(x')) ((\nabla L)(x'))^T (x - x') w_{x_0}(x') dx', \quad (58)$$

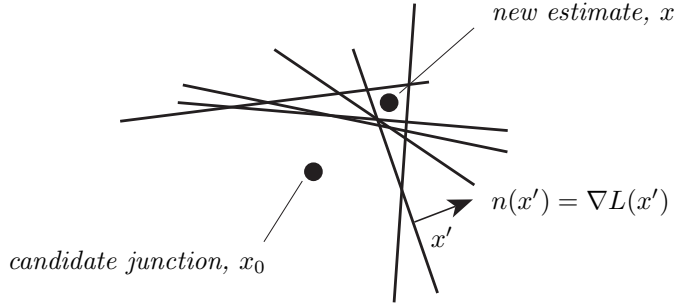


Figure 12: Minimizing (57) basically corresponds to finding the point x that minimizes the distance to all edge tangents in a neighbourhood of the given candidate junction point x_0 .

and the minimization problem be expressed as a standard least squares problem

$$\min_{x \in \mathbb{R}^2} x^T A x - 2 x^T b + c \iff A x = b, \quad (59)$$

where $x = (x_1, x_2)^T$, and A , b , and c are determined by the local statistics of the gradient directions in a neighbourhood of x_0 ,

$$A = \int_{x' \in \mathbb{R}^2} (\nabla L)(x') (\nabla L)^T(x') w_{x_0}(x') dx', \quad (60)$$

$$b = \int_{x' \in \mathbb{R}^2} (\nabla L)(x') (\nabla L)^T(x') x' w_{x_0}(x') dx', \quad (61)$$

$$c = \int_{x' \in \mathbb{R}^2} x'^T (\nabla L)(x') (\nabla L)^T(x') x' w_{x_0}(x') dx'. \quad (62)$$

Provided that the 2×2 matrix A is non-singular, the minimum value is given by

$$d_{min} = \min_{x \in \mathbb{R}^2} x^T A x - 2 x^T b + c = c - b^T A^{-1} b, \quad (63)$$

and the point x that minimizes (57) is $x = A^{-1}b$. Hence, an improved localization estimate can be computed directly from image measurements.

7.2 Automatic selection of localization scales

The formulation in previous section however, leaves two major problems open: How to choose the window function w_{x_0} , and the scale(s) for computing the gradient vectors.

- The problem of choosing the weighting function is a special case of a common scale problem in least squares estimation: Over what spatial region should the fitting be performed? Clearly, it should be large enough such that statistics of gradient directions is accumulated over a sufficiently large neighbourhood around the candidate junction. Nevertheless, the region must not be so large that interfering structures corresponding to other junctions are included.
- The second scale problem, on the other hand, is of a slightly different nature than the previous ones—it concerns what scales should be used for *localizing image structures*. Previously, in this paper, only the problem of *detecting* image structures has been treated.

Here, the following solutions are proposed:

Selection of window function and spatial points from the detection step: When computing A , b , and c above, let the window function w_{x_0} be a Gaussian function centered at the point x_0 at which $\tilde{\kappa}_{norm}^2$ assumed its scale-space maximum. Moreover, let the scale value of this window function be proportional to the detection scale t_0 at which the maximum over scales in $\tilde{\kappa}_{norm}^2$ was assumed.

The idea behind this approach is that the detection scale should reflect a representative region around the candidate junction, such that larger regions are selected for corners with large spatial extent than for corners with small extent. Experimentally, this has been demonstrated to be the case in a large number of situations (compare also with the qualitative results in sections 6.2–6.3). (An more general approach for defining the support region of the junction feature is described in section 7.7.)

Selection of localization scale: Minimize the normalized residual. Clearly, the gradient estimates used for computing A , b , and c must be computed at a certain scale. To determine this *localization scale*, it is natural to *select the scale that minimizes the normalized residual d_{min} in (65) over scales*.

This scale selection criterion corresponds to extending the minimization problem (59) from a single scale to optimization over multiple scales

$$\begin{aligned} \min_{t \in \mathbb{R}_+} \min_{x \in \mathbb{R}^2} \frac{x^T A x - 2 x^T b + c}{\text{norm}(t)} = \\ \min_{t \in \mathbb{R}_+} \min_{x \in \mathbb{R}^2} \frac{c - b^T A^{-1} b}{\text{trace } A} \end{aligned} \quad (64)$$

where the normalization factor $\text{norm}(t)$ has been introduced to relate minimizations at different scales. The particular choice of $\text{norm}(t) = \text{trace } A$ implies that the normalized residual

$$\tilde{d} = \min_{x \in \mathbb{R}^2} \frac{\int_{x' \in \mathbb{R}^2} |((\nabla L)(x'))^T (x - x')|^2 w_{x_0}(x') dx'}{\int_{x' \in \mathbb{R}^2} |(\nabla L)(x')|^2 w_{x_0}(x') dx'} \quad (65)$$

has dimension $[\text{length}]^2$ can be interpreted as a weighted estimate of the localization error. Specifically, scale selection according by minimizing the normalized residual \tilde{r} (65) over scales, corresponds to selecting the scale that minimizes the estimated inaccuracy in the localization estimate.

This principle for selecting localization scales implies that we take as *localization scale* the scale that gives the *maximum consistency between the distribution of gradient directions* in a neighbourhood of x_0 and a local (qualitative) junction model. More specific motivations behind this choice can also be expressed as follows:

At very fine scales, where a large amount of noise and interfering fine-scale structures can be expected to be present, the first-order derivative operators will respond mainly to such structures. Hence, the gradient directions can be expected to be roughly randomly distributed, and the residual d_{min} will, in general, be large. At coarser scales in scale-space, such fine-scale structures will be suppressed and the locally computed gradient directions will be better aligned to the underlying corner structure. Thus, when smoothing is necessary, the residual will decrease. On the other hand, if too much smoothing is applied, then the shape distorting effects of scale-space smoothing will be dominant and the residual can be expected to increase again. Hence, selecting the minimum gives a natural trade-off between these two effects.

Behaviour at ideal sharp junctions of polygon-type. Note, in particular, that for an ideal (sharp) step junction, the localization scale given by this method will always be zero in the noise free case. This can be easily understood by observing that for an ideal polygon-type junction (consisting of regions of uniform grey-level delimited by straight edges), all edge tangents meet at the junction point, which means that the residual \tilde{d}_{min} is exactly zero. Thus, any amount of smoothing increases the residual, and the minimum value will be assumed at zero scale.

7.3 Experiments: Choice of localization scale

Figure 13 and figure 14 show the result of applying this scale selection mechanism to a sharp and a diffuse corner with different amounts of added white Gaussian noise. As can be seen, the results agree with a the qualitative discussion above. There is a clear minimum over scales in each scale-space signature, and the minimum over scales is assumed at coarser scales when the noise level is increased. Moreover, slightly coarser scales are selected for the diffuse junction than for the sharp one.

Junction localization with automatic scale selection: T-junction						
Noise level	Selected scale	Normalized residual	Absolute error	Reference scale	Normalized residual	Absolute error
0.01	1.2	1.0	0.9	1.0	1.0	0.9
0.03	2.8	2.3	1.4	2.1	2.5	1.4
0.1	8.1	6.7	5.1	6.3	6.8	4.9
0.3	18.8	13.9	16.6	13.4	15.0	14.7
1.0	50.3	28.5	116.0	28.2	39.2	96.3

Table 1: The result of applying the junction localization method to a synthetic T -junction with different amounts of added white Gaussian noise. For each noise level, this table gives the scale at which the normalized residual assumes its minimum over scales, as well as the scale at which the estimate with the minimum absolute error over scales is obtained. Moreover, numerical values of the two error measures are given at these scales. As can be seen, the selected scales increase with the noise level, and the scale at which the normalized residual assumes its minimum over scales serves as a reasonable estimate of a scale at which a near optimal localization estimate over scales is obtained.

Table 1 gives a numerical illustration of basic properties of this scale selection method for junction localization. It shows the result of applying one iteration of the junction localization method to a T junction with 90 degree opening angles, and the results are shown in terms of the following six measures as function of the noise level:

- the selected scale $t_{d_{min}}$ obtained by minimizing the normalized residual over scales,
- the normalized residual at the selected scale,
- the absolute error in the localization estimate at the selected scale,
- the scale t_{abs} at which the localization estimate with the minimum absolute error is obtained,
- the normalized residual at t_{abs} ,
- the minimum actual error of the localization estimates computed at all scales.

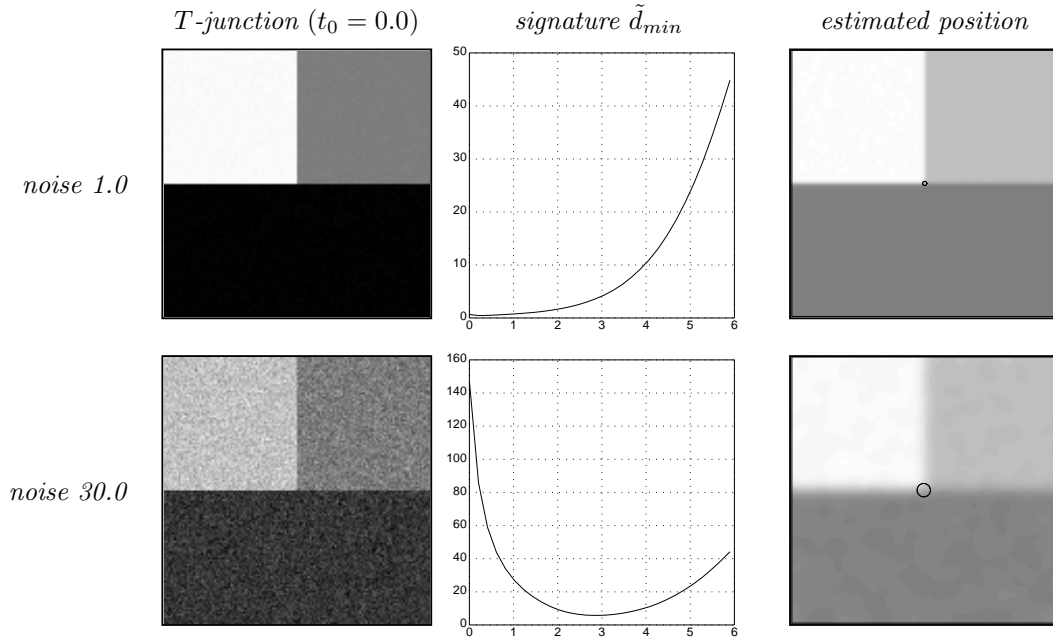


Figure 13: Scale-space signatures of the normalized residual at a synthetic sharp T -junction ($t_0 = 0.0$) for different amounts of added white Gaussian noise ($\nu = 1.0$ and 30.0): (left) grey-level image, (middle) signature of \tilde{d}_{min} accumulated at the central point, (right) localization estimate computed at the scale at which \tilde{d}_{min} assumes its minimum over scales (illustrated by a circle overlayed onto a bright copy of the image smoothed to that scale).

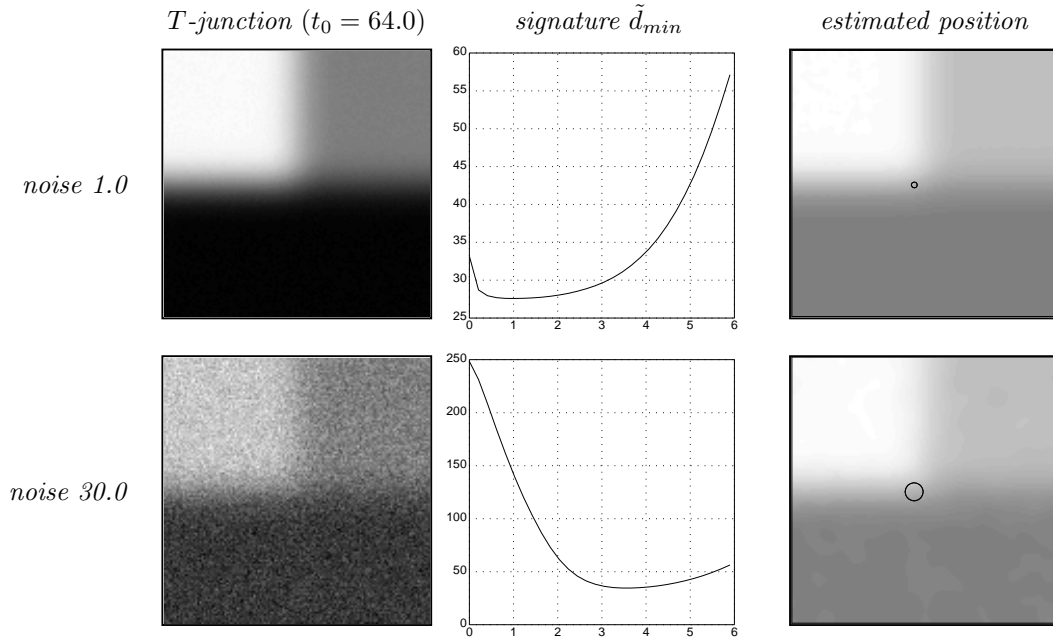


Figure 14: Corresponding results for a synthetic diffuse T -junction ($t_0 = 64.0$).

All descriptors have been computed at a position with 10 pixels horizontal and vertical offset from true corner position.

The results show that the normalized residual serves as an estimate of the inaccuracy in the corner localization estimate, and specifically that the scale at which the minimum over scales in \tilde{d}_{min} is assumed is a reasonable estimate of the scale at which we have the localization estimate with the minimum absolute error. Whereas the correspondence between the two error measures is not perfect, the absolute error computed at the scale at which the normalized residual assumes its minimum over scales is only slightly higher than the minimum absolute error over all scales. In this respect, minimization of \tilde{d}_{norm} over scales gives a near optimal localization estimate (without knowledge about the true junction position).

Moreover, whereas the error estimates assume rather high values for a single application of the junction localization scheme when the noise level is high, the localization error can be decreased substantially by applying the junction localization scheme iteratively.

Figure 15 shows the result of applying the composed junction localization stage to the junction candidates in figure 8. For each scale-space maximum, an individual scale selection process has been invoked consisting of the following processing steps:

- The signature of the normalized residual has been accumulated using a window function with scale value equal to the detection scale of the scale-space maximum.
- The minimum over scales in the signature of \tilde{d}_{min} has been detected, and a new localization estimate has been computed using $x = A^{-1}b$.
- This procedure has been repeated iteratively until either the difference between two successive localization estimates is less than one pixel or the number of iterations has reached an upper bound (here 3 iterations).
- Junction candidates for which the new localization estimates fall outside the support region of the original scale-space maximum have been classified as “diverged” and been suppressed.
- Each remaining junction candidate has been illustrated by a circle with radius proportional to the detection scale or the localization scale.

Observe how the localization is improved by this postprocessing step, and moreover, that the selected localization scales serve as estimates of the spatial localization error.

7.4 Composed scheme for junction detection and localization

To summarize, the composed two-stage scheme for junction detection and junction localization consists of the following processing steps:⁵

1. *Detection.* Detect scale-space maxima in the square of the normalized rescaled level curve curvature

$$\tilde{\kappa}_{norm} = t^{2\gamma} \tilde{\kappa} = t^{2\gamma} (L_{x_2}^2 L_{x_1 x_1} - 2L_{x_1} L_{x_2} L_{x_1 x_2} + L_{x_1}^2 L_{x_2 x_2})$$

(or some other suitable normalized differential entity). This generates a set of junction candidates.

⁵Besides the general descriptions given in previous sections, further details concerning algorithms and discrete implementation can be found in appendix A.4 and in (Lindeberg 1994c).

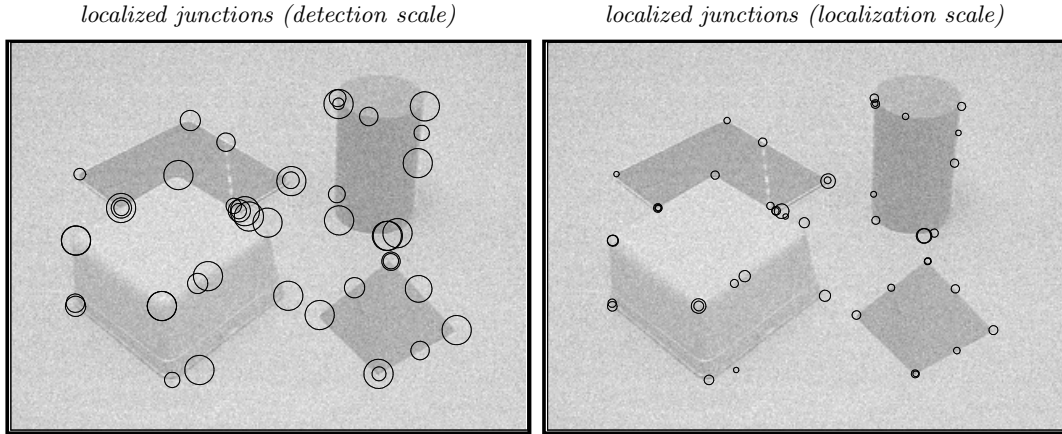


Figure 15: Improved localization estimates for the junction candidates in figure 8. Each junction has been graphically illustrated by a circle centered at the new location estimate. In the left image, the size reflects the detection scale, whereas in the right image, the size reflects the localization scale.

2. *Localization.* For each junction candidate, accumulate the scale-space signature of the normalized residual

$$\tilde{d}_{min} = \frac{c - b^T A^{-1} b}{\text{trace } A}$$

with A , b , and c computed according to (60)–(62), and using window function reflecting the support region of the scale-space maximum at the detection scale.

Then, at the scale at which the minimum in \tilde{d}_{min} is assumed, compute an improved localization estimate using

$$x = A^{-1}b.$$

3. *Iterations.* Optionally, repeat step 2 until the increment is sufficiently small.

7.5 Further experiments

Figures 16–17 show the result of applying the composed two-state junction detection scheme to four indoor images containing different types of junctions. As can be seen, very reasonable sets of junction candidates are obtained, and the support regions of the scale-space blobs again serve as natural regions of interest around the features.

Concerning the number of junction candidates to be processed and passed on to later processing stages, we have not made any attempts in this work to decide automatically how many of the extracted junction candidates correspond to physical junctions in the world. We argue that such decisions require integration with higher-level reasoning and verification processes, and may be extremely hard to make at the earliest processing stages unless additional information is available about the external conditions.⁶ For this reason, this module only aims at computing an early ranking of image features in order of significance, which can be used by a vision system for processing features in decreasing order of significance.

⁶An integrated vision system for analysing junctions by actively zooming in to interesting structures is presented in (Brunnström *et al.* 1992; Lindeberg 1993a).

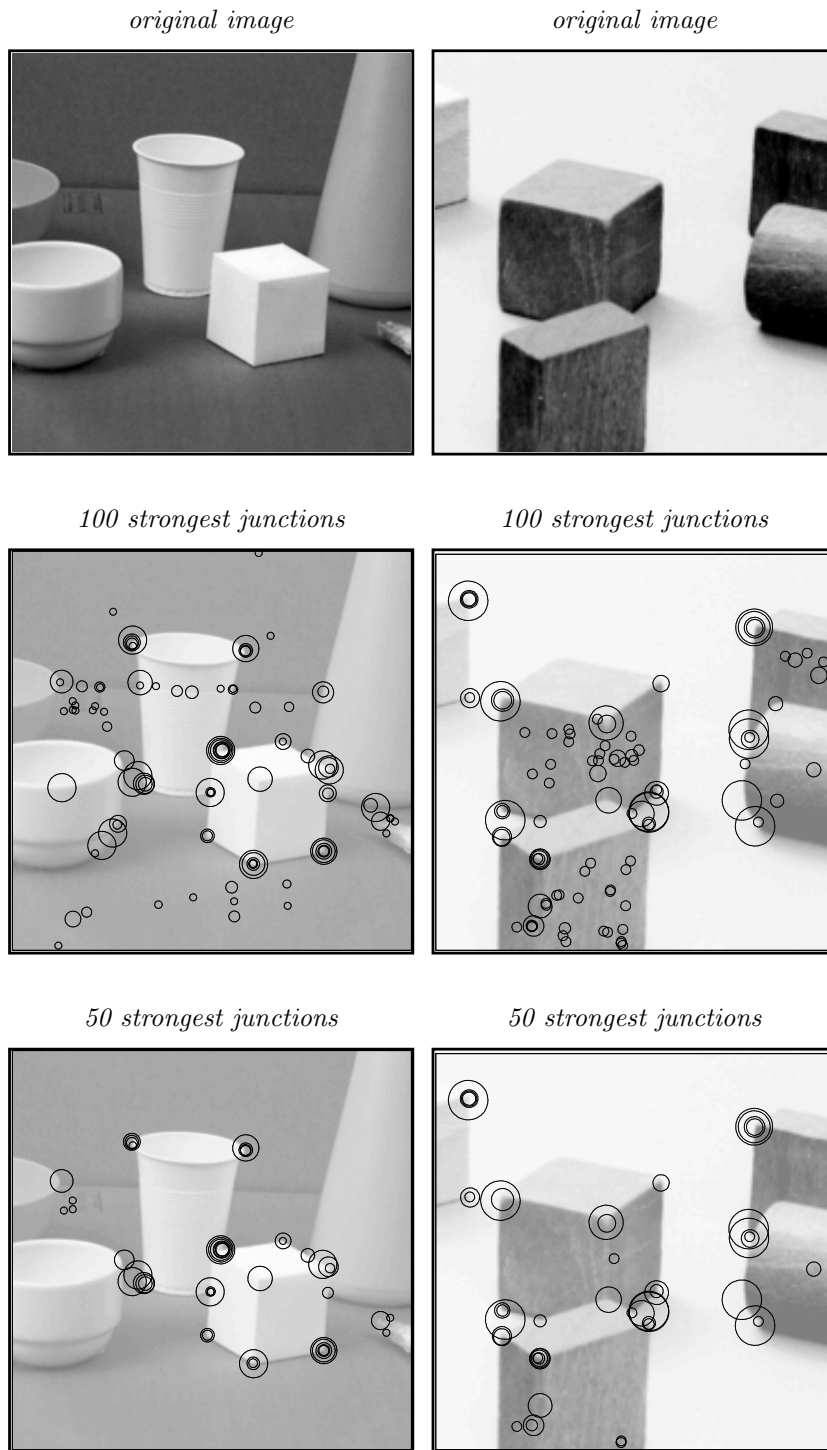


Figure 16: Results of composed two-stage junction detection followed by junction localization for two different grey-level images. (top row) original grey-level image, (middle and bottom rows) the N strongest junction candidates for different values of N .

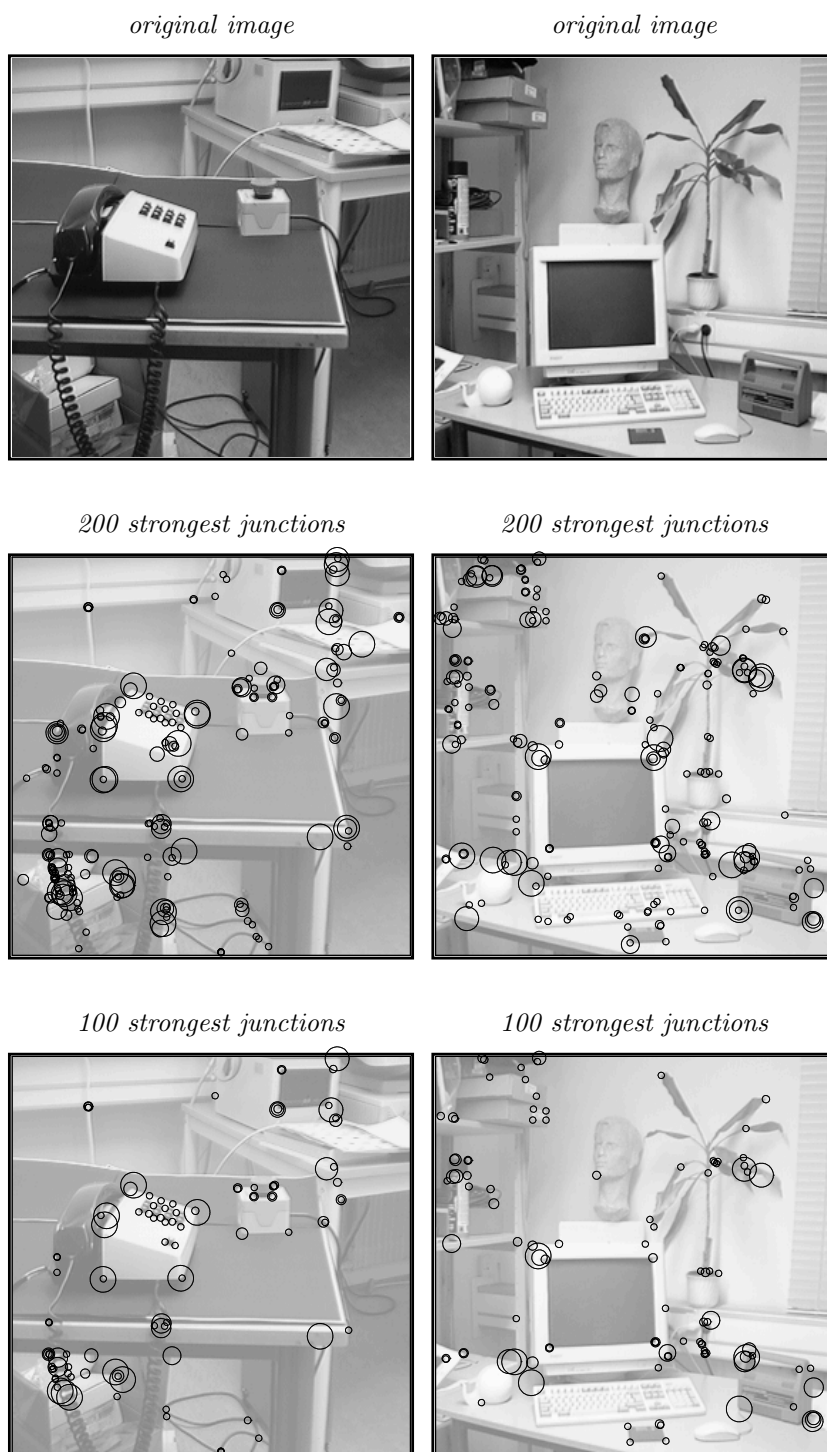


Figure 17: Results of composed two-stage junction detection followed by junction localization for two different grey-level images. (top row) original grey-level image, (middle and bottom rows) the N strongest junction candidates for different values of N .

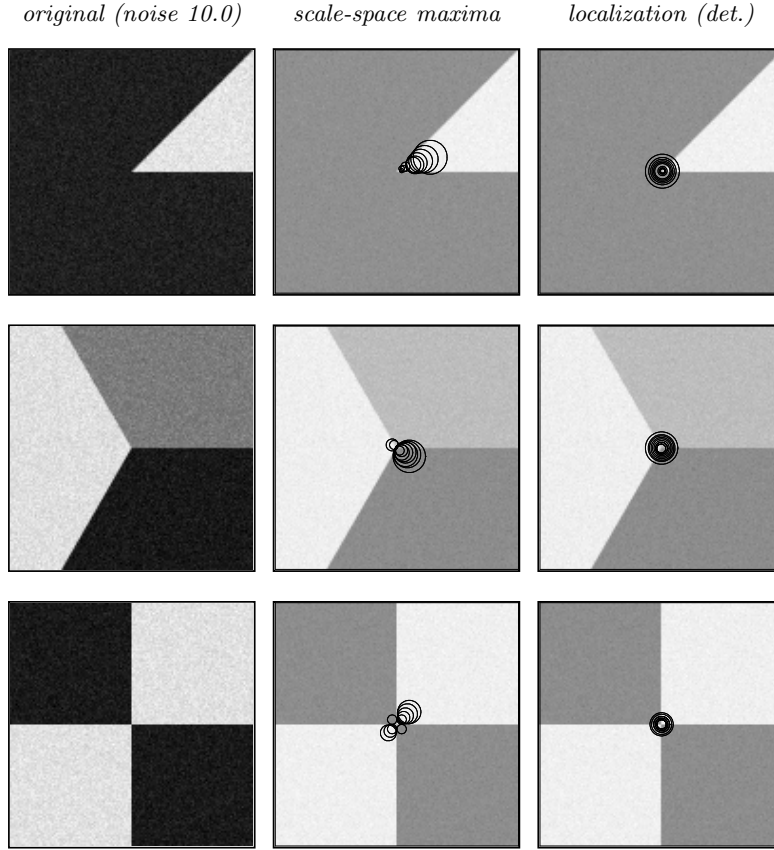


Figure 18: The result of applying the composed junction detection scheme to synthetic junction models with added Gaussian noise ($\nu = 10.0$).

Localization errors (noise level $\nu = 10.0$)			
	detection	localization	iteration
L-junction	3.81	0.78	0.43
Y-junction	2.12	0.35	0.35
4-junction	3.53	0.28	0.07

Figure 19: Localization errors in the different processing stages of the composed junction detection scheme applied to the synthetic junction models in figure 18: (i) in the detection stage, (ii) after one localization step, (iii) after convergence of the iterative procedure.

In line with this idea, the results are shown in terms of the N strongest junction candidates for different (manually chosen) values of N . In figure 16, which contains only a few objects, the 50 and the 100 strongest junctions responses are shown. In figure 17, which shows corresponding examples for more cluttered scenes, the number of junctions displayed has been increased to 100 and 200. Notably, this number of junction candidates constitutes the only essential tuning parameter of the composed algorithm. In particular, no external setting of the scale parameters is needed.

Figure 18 shows two examples of applying the composed method to synthetic polygon-type junctions with added Gaussian noise. Here, the 10 most significant junctions have been processed. As can be seen, higher order junctions are handled in a similar way as lower order junctions. The table in figure 19 shows numerical values exemplifying how large the localization errors can be in the different processing stages.

7.6 *Applications of corner detection with automatic scale selection*

In (Lindeberg and Li 1995, 1997) it is shown how the support region associated with each junction allows for conceptually simple matching between junctions and edges based on spatial overlap only and without any need for providing externally determined thresholds on *e.g.* distance. Then, the matching relations between edges and junction cues that arise in this way are used in a pre-processing stage for classifying edges into straight and curved. More generally, such relations between edges and junctions are also useful for other problems relating to object recognition (Lindeberg and Olofsson 1995).

In (Bretzner and Lindeberg 1996) it is demonstrated how these support regions can be used for simplifying matching of junctions over time in tracking algorithms. Specifically, it is shown that the scale selection mechanism in the junction detector is essential to capture junctions that undergo large size changes. Moreover, the scale information associated with each junction can be used as an important matching cue in its own right and be included in the matching criterion.

In (Lindeberg 1995a, 1996d) a scale selection principle for stereo matching and flow estimation is presented, which also involves the extension of a fixed scale least squares estimation problem to optimization over multiple scales.

7.7 *Extensions of the junction detection method*

The main purpose of the presentation in this section has been to make explicit how a scale selection mechanism can be incorporated into a junction detector. When building a stand-alone junction detector, there are a few additional mechanisms which are natural to include if the aim is to construct a stand-alone junction detector.

Concerning the ranking on significance, we can conceive linking the maxima of the junction responses across scales in a similar way as done in the scale-space primal sketch (Lindeberg 1993a), register scale-space events such as bifurcations, and include the scale-space lifetime of each junction response into the significance measure.

Concerning the region of interest associated with each junction candidate, we have throughout this work represented the support region of a scale-space maximum by a circle with area reflecting the detection scale. A possible limitation of this approach is that nearby junctions may lead to interference effects in operations such as the localization stage. If we want to reduce this problem, a natural extension is to define the support region of a spatial maximum in a differential invariant in the same way as grey-level blobs are defined from spatial maxima in (Lindeberg 1993a).

7.8 Extensions to edge detection

Concerning more general applications of the proposed methodology, it should be noted that the scale selection method for junction localization applies to edge detection as well. If we compute the scale-space signature of the normalized residual \tilde{d}_{min} according to (63) at an edge point, then we can interpret the scale at which the minimum over scales in \tilde{d}_{min} as the scale for which a local edge (or corner) model is maximally consistent with data. This means that the fine scale fluctuations of the edge normals can be expected to be small.

Figure 20 shows the result of applying this approach to different types of edge data. The columns show from left to right; (i) the local grey-level pattern, (ii) the signature of \tilde{d}_{min} computed at the central point, and (iii) (unthresholded) edges detected at the scale $t_d = \operatorname{argmin} \tilde{d}_{min}$ at which the minimum over scales in \tilde{d}_{min} was assumed.

In the first row, we can see that when performing edge detection at $\operatorname{argmin} \tilde{d}_{min}$ we obtain coherent edge descriptor corresponding to the dominant edge structure in this region. In the second row, a large amount of white Gaussian noise has been added to the grey-level image, and the minimum over scales is assumed at a much

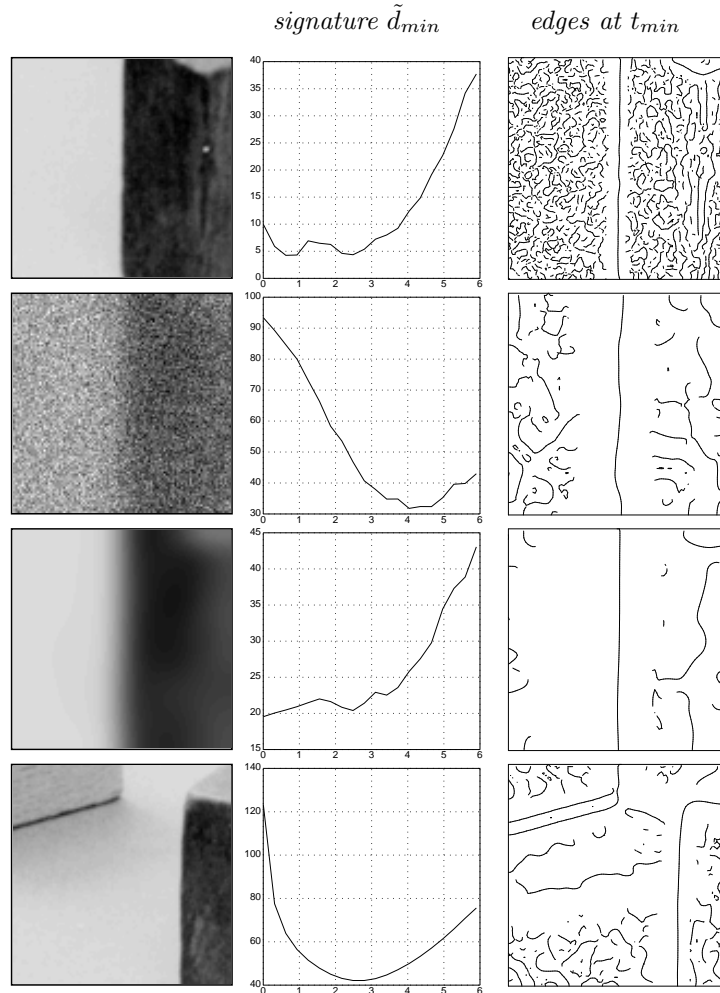


Figure 20: Illustration of the result of applying minimization over scales of the normalized residual \tilde{d}_{min} to different types of edge structures: (left) grey-level image, (middle) scale-space signature of \tilde{d}_{min} accumulated at the central point, (right) (unthresholded) edges computed at the scale where \tilde{d}_{min} assumes its minimum over scales.

coarser scale. (Note that a coherent edge descriptor is nevertheless obtained at the new argmin \tilde{d}_{min} , whereas edge detection at the same scale as in the first row gives rise to severely fragmented edges.) In the third row, where the grey-level image has been subject to a large amount of presmoothing, no additional smoothing is necessary and the minimum is assumed at zero scale. Finally, the fourth row shows an interesting example where the minimization of \tilde{d}_{min} over scales leads to the selection of a scale level suitable for capturing a very faint shadow edge.

Concerning these experiments it should be pointed out that they are mainly intended to demonstrate the potential in applying the proposed method for selecting localization scales to the problem of edge detection, and that further processing steps are needed to give a complete algorithm. A complementary scale selection method applicable in the edge detection stage is presented in (Lindeberg 1996a).

8 Dense frequency estimation

So far, we have seen how the scale selection methodology can be applied to the detection of sparse feature points. In certain situations, however, one is also interested in computing *dense* image descriptors.

An obvious problem that arises if we would base a scale selection mechanism for computing dense image descriptors on a partial derivative of the intensity function, such as the Laplacian operator is that there would be large spatial variations in the operator response. (Hence, the spatial variations in the selected scales could be large.) A common methodology in signal processing for reducing this so-called phase dependency is by using *quadrature filter pairs* defined (from a Hilbert transform) in such a way that the Euclidean sum of the filter responses will be constant for any sine wave. Since, however, the Hilbert transform of a Gaussian derivative kernel is not within the Gaussian derivative family, one may be interested in operators of small support which can be expressed within the scale-space framework.

Given the normalized derivative concept, there is, a straightforward way of combining Gaussian derivatives into an entity that gives an approximately constant operator response at the scale given by the scale selection mechanism. At any scale t in the scale-space representation L of a one-dimensional signal f , define the following *quasi quadrature* entity in terms of normalized derivatives based on $\gamma = 1$ ⁷ by

$$\mathcal{Q}L = L_\xi^2 + C L_{\xi\xi}^2 = t L_x^2 + C t^2 L_{xx}^2 \quad (66)$$

where C is a free parameter (to be determined later). By differentiating (7), it is follows that for a signal of the form $f(x) = \sin \omega_0 x$ the quasi quadrature measure assumes the form

$$(\mathcal{Q}L)(x; t) = t \omega_0^2 e^{-\omega_0^2 t} (1 + (C t \omega_0^2 - 1) \sin^2 \omega_0 x). \quad (67)$$

As can be seen, the spatial variations in $\mathcal{Q}L$ will be large when $t \omega_0^2$ is either much smaller or much larger than one, whereas the relative oscillations decrease to zero when t approaches $1/(C \omega_0^2)$. (As will be shown below, this scale value is of the same order of magnitude as the scales that maximize $\mathcal{Q}L$ over scales; compare also with section 3.) In this respect, $\mathcal{Q}L$ serves as an approximate quadrature pair leading to small relative spatial variations near the scales given by the scale selection procedure.

⁷Since $\mathcal{Q}L$ is an inhomogeneous differential expression, $\gamma = 1$ is a necessary requirement for the scale selection procedure to commute with size variations in the input pattern (see section 4.1).

For this reason, we propose to use QL as an entity to maximize over scales when computing dense image descriptors. For two-dimensional data, we can instead consider

$$QL = L_\xi^2 + L_\eta^2 + L_{\xi\xi}^2 + 2L_{\xi\eta}^2 + L_{\eta\eta}^2 = t(L_x^2 + L_y^2) + C t^2 (L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2) \quad (68)$$

defined to be rotationally symmetric and equal to the one-dimensional quadrature measure in any direction.

Analysis for sine wave patterns. The free parameter C determines the relative weight between the information in the first- and second-order derivatives. To obtain an intuitive understanding of how it affects the scale selection procedure, it is instructive to analyse what scales are obtained by maximizing QL over scales for different values of C . Straightforward differentiation of (67) gives that selected scale as function of spatial position is given by

$$t_{QL}(x) = \frac{1}{\omega_0^2} \left(1 + \frac{2C \sin^2(\omega_0 x)}{\cos^2(\omega_0 x) + \sqrt{\cos^4(\omega_0 x) + 4C^2 \sin^4(\omega_0 x)}} \right) \quad (69)$$

and that the extreme values at $\omega_0 x = 0$ and $\frac{\pi}{2}$ are independent of C

$$t_{Q,0} = t_{QL}|_{\omega_0 x=0} = \frac{1}{\omega_0^2}, \quad t_{Q,\frac{\pi}{2}} = t_{QL}|_{\omega_0 x=\frac{\pi}{2}} = \frac{2}{\omega_0^2}. \quad (70)$$

Graphs showing this variation for a few values of C are displayed in figure 21. Given the form of these curves, a natural symmetry requirement can be stated as

$$t_{QL}|_{\omega_0 x=\frac{\pi}{4}} = \frac{1}{2} \left(t_{QL}|_{\omega_0 x=0} + t_{QL}|_{\omega_0 x=\frac{\pi}{2}} \right) \quad \Rightarrow \quad C = \frac{2}{3} \approx 0.6667. \quad (71)$$

In this respect, $C = \frac{2}{3}$ gives the most symmetric variation of selected scale levels with respect to the information contents in the first-order and second-order derivatives.⁸

Another interesting factor to analyse is the variation in magnitude at the selected scales. Insertion of the scale values according to (69) into the quasi quadrature measure (66) gives spatial variations of as displayed in figure 22. To determine C , a simple minimum-ripple condition is to require that

$$QL|_{\substack{\omega_0 x=0 \\ t=t_{Q,0}}} = QL|_{\substack{\omega_0 x=\frac{\pi}{2} \\ t=t_{Q,\frac{\pi}{2}}}} \quad \Rightarrow \quad C = \frac{e}{4} \approx 0.6796. \quad (72)$$

In other words, also a determination of C based on small spatial variations in the magnitude measure computed at the selected scales gives rise to an approximately similar value of C as the abovementioned symmetry requirement.

Experimental results. Figure 23 gives a three-dimensional illustration of the result of applying this operation to the perspective image of a sine wave pattern with large size variations. The results are shown as a surface plot of the magnitude of QL computed at different positions and scales along a vertical cross-section of the image. Moreover, the position of the first local maximum over scales has been indicated at each spatial point. Observe how the size variations in the vertical direction are captured and that the spatial variations in QL at the selected scales are minor compared to response of an operator such as the squared Laplacian.

⁸If we redefine the quasi quadrature measure as $QL = (1-\alpha) L_\xi^2 + \alpha L_{\xi\xi}^2$, then $C = 2/3$ corresponds to the relative weights $1 - \alpha = 3/5$ and $\alpha = 2/5$.

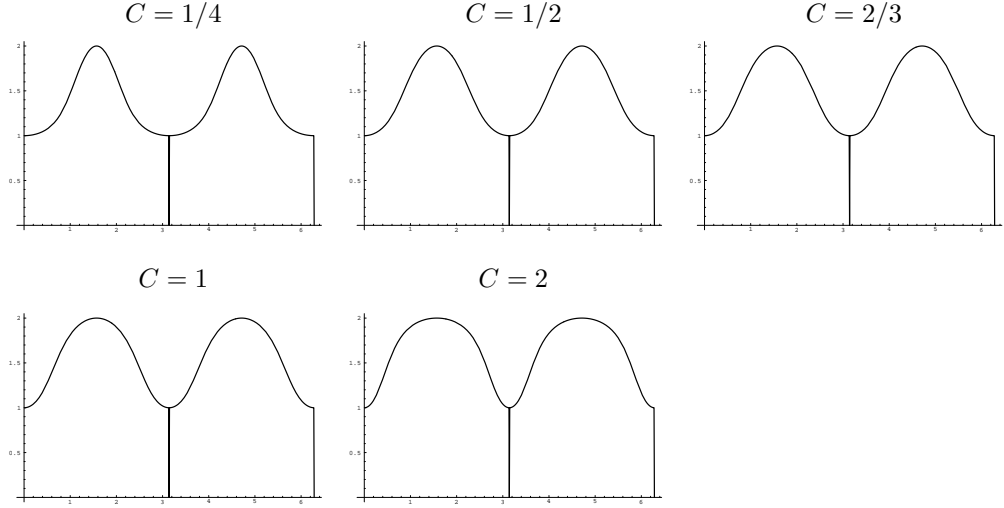


Figure 21: Spatial variation of the selected scale levels when maximizing the quasi quadrature entity (66) over scales for different values of the free parameter C using a one-dimensional sine wave of unit input frequency as input pattern. Observe that $C = 2/3$ gives rise to the most symmetric variations in the selected scale values.

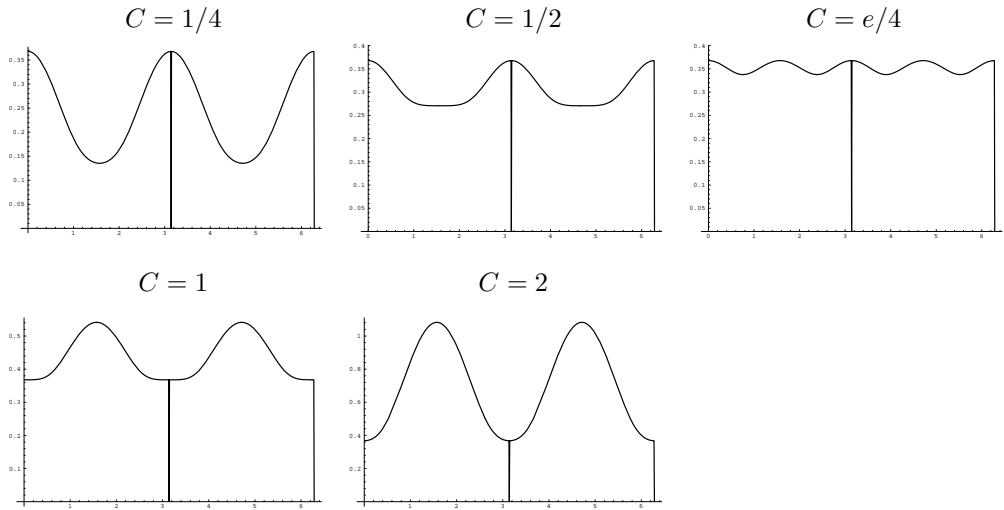


Figure 22: Spatial variation of the maximum value over scales of the quasi quadrature entity (66) computed for different values of the free parameter C using a one-dimensional sine wave of unit input frequency as input pattern. As can be seen, the smallest spatial variations in the amplitude of the maximum response are obtained for $C = e/4$.

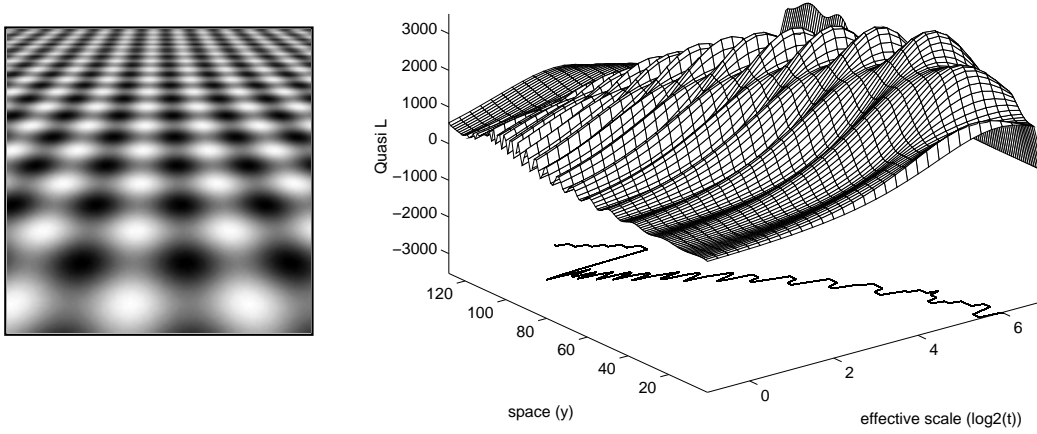


Figure 23: Dense scale selection by maximizing the quasi quadrature measure (68) over scales: (left) Original grey-level image. (right) The variations over scales of the quasi quadrature measure QL computed along a vertical cross-section through the center of the image. The result is visualized as a surface plot showing the variations over scale of the quasi quadrature measure as well as the position of the first local maximum over scales.

Extensions and applications. In (Almansa and Lindeberg 1996, 1998) this approach is applied for estimating the distance between ridges in fingerprint images. For that specific problem, the following extensions were found to be highly useful:

- The extension of the renormalized one-dimensional quadrature entity in (73) to two dimensions is done using the ridge strength measure (Lindeberg 1996b)

$$\mathcal{A}_{\gamma_2\text{-norm}}L = t^{2\gamma_2} ((L_{xx} - L_{yy})^2 + 4L_{xy}).$$

This second order differential entity is also invariant under rotations, but has more selective response properties to elongated ridge-like structures than $L_{xx}^2 + 2L_{xy}^2 + L_{yy}^2$, which rather constitutes a measure of the total amount of second order information.

- The quasi quadrature entity (66) is multiplied by a uniform self-similar scaling factor

$$\begin{aligned} \mathcal{Q}'L &= t^{-\Gamma} \mathcal{Q}L \\ &= t^{-\Gamma} (L_{\xi}^2 + L_{\eta}^2). \\ &= \left(t^{(1-\Gamma)/2} L_x \right)^2 + \left(t^{(1-\Gamma)/2} L_{xx} \right)^2. \end{aligned} \quad (73)$$

This operation does not affect the homogeneity of the general class of differential expressions (24), and with the special choice $\Gamma = \frac{1}{2}$, the quasi quadrature entity can be written

$$\begin{aligned} \mathcal{Q}'L &= \left(t^{\gamma_1/2} L_x \right)^2 + \left(t^{\gamma_2} L_{xx} \right)^2 \\ &= \mathcal{G}_{\gamma_1\text{-norm}}L + \mathcal{A}_{\gamma_2\text{-norm}}L. \end{aligned} \quad (74)$$

In view of results presented in a companion paper (Lindeberg 1996c, 1996b), this scale renormalized quasi quadrature measure can be interpreted as the linear

combination of an edge strength measure $\mathcal{G}_{\gamma_1-norm}L$ with scale normalization parameter $\gamma_1 = \frac{1}{2}$ and a ridge strength measure $\mathcal{A}_{\gamma_2-norm}L$ with scale normalization parameter $\gamma_2 = \frac{3}{4}$. In the abovementioned sources, these specific entities and normalization parameters are shown to be useful for edge detection and ridge detection with automatic scale selection.

A main motivation for this renormalization is that it gives more pronounced peaks over scales, and prevents the monotone increase with scale that occurs for edges if the scale normalization parameter γ is equal to one.

9 Analysis and interpretation of normalized derivatives

9.1 Interpretation of γ -normalized derivatives in terms of L_p -norms

Concerning the interpretation of the γ -normalized derivatives, it is straightforward to show (see appendix A.2) that in the D -dimensional case, the variation over scales of the L_p -norm of an m th order normalized Gaussian derivative kernel is given by

$$\|g_{\xi^m}(\cdot; t)\|_p = \sqrt{t}^{m(\gamma-1)+D(1/p-1)} \|g_{\xi^m}(\cdot; t)\|_p. \quad (75)$$

In other words, the L_p -norm of the m th order Gaussian derivative kernel is constant over scales if and only if

$$p = \frac{1}{1 + \frac{m}{D}(1 - \gamma)}. \quad (76)$$

Hence, the γ -normalized derivative concept can be interpreted as an L_p -normalization of the Gaussian derivative kernels over scales for a specific value of p , which depends upon γ , the dimension as well as the order m of differentiation.

Notably, the perfectly scale invariant case $\gamma = 1$ gives $p = 1$ for all orders m and corresponds to L_1 -normalization of the Gaussian derivative kernels. For orders up to four, the kernel norms are then in the one-dimensional case given by

$$N_1 = \int_{-\infty}^{\infty} |g_{\xi}(u; t)| du = \sqrt{\frac{2}{\pi}} \approx 0.797885, \quad (77)$$

$$N_2 = \int_{-\infty}^{\infty} |g_{\xi^2}(u; t)| du = \sqrt{\frac{8}{\pi e}} \approx 0.967883, \quad (78)$$

$$N_3 = \int_{-\infty}^{\infty} |g_{\xi^3}(u; t)| du = \sqrt{\frac{2}{\pi}} \left(1 + \frac{4}{e^{3/2}}\right) \approx 1.51003, \quad (79)$$

$$N_4 = \int_{-\infty}^{\infty} |g_{\xi^4}(u; t)| du = \frac{4\sqrt{3}}{e^{3/2+\sqrt{3}/2}\sqrt{\pi}} (\sqrt{3 - \sqrt{6}}e^{\sqrt{6}} + \sqrt{3 + \sqrt{6}}) \approx 2.8006. \quad (80)$$

9.2 Interpretation in terms of self-similar Fourier spectrum

Another useful interpretation of normalized derivatives can be obtained in the context of signals having a self-similar Fourier spectrum.

Consider a D -dimensional signal $f: \mathbb{R}^D \rightarrow \mathbb{R}$ having a self-similar power spectrum of the form

$$\begin{aligned} S_f(\omega) &= S_f(\omega_1, \dots, \omega_D) = (\hat{f}\hat{f}^*)(\omega) \\ &= |\omega|^{-2\beta} = (\omega_1^2 + \dots + \omega_D^2)^{-\beta}, \end{aligned} \quad (81)$$

and the following class of energy measures concerning the amount of information in the m th order γ -normalized Gaussian derivatives

$$E_m = \int_{x \in \mathbb{R}^D} \sum_{|\alpha|=m} t^{m\gamma} |L_{x^\alpha}|^2 dx. \quad (82)$$

where α represents multi-index notation. These differential energy measures are related to m th order spectral moments by

$$E_m = \frac{t^{m\gamma}}{(2\pi)^D} \int_{\omega \in \mathbb{R}^D} |\omega|^{2m} |\hat{L}(\omega; t)|^2 d\omega. \quad (83)$$

Specifically, for derivatives up to order three in the two-dimensional case, this class of energy measures includes the following descriptors

$$E_0 = \int_{x \in \mathbb{R}^2} L(x; t)^2 dx, \quad (84)$$

$$E_1 = \int_{x \in \mathbb{R}^2} t^\gamma (L_\xi^2 + L_\eta^2) dx, \quad (85)$$

$$E_2 = \int_{x \in \mathbb{R}^2} t^{2\gamma} (L_{\xi\xi}^2 + 2L_{\xi\eta}^2 + L_{\eta\eta}^2) dx, \quad (86)$$

$$E_3 = \int_{x \in \mathbb{R}^2} t^{3\gamma} (L_{\xi\xi\xi}^2 + 3L_{\xi\xi\eta}^2 + 3L_{\xi\eta\eta}^2 + L_{\eta\eta\eta}^2) dx. \quad (87)$$

It is rather straightforward to show (see appendix A.3) that the variation over scales of these γ -normalized energy measures are given by

$$E_m(\cdot; t) \sim t^{\beta-D/2-m(1-\gamma)}. \quad (88)$$

This expression is scale independent if and only if

$$\beta = \frac{D}{2} + m(1 - \gamma). \quad (89)$$

In other words, the normalized derivative model is *neutral* with respect to power spectra of the form

$$S_f(\omega) = |\omega|^{-D-2m(1-\gamma)}. \quad (90)$$

In the special case when $D = 2$ and $\gamma = 1$, this corresponds to power spectra of the form $|\omega|^{-2}$.

It is well-known that natural images often show a qualitative behaviour similar to this (Field 1987). This is in fact a direct consequence of scale invariance; the power spectrum variation

$$S(\omega) \sim |\omega|^{-D}, \quad (91)$$

where D is the dimension of the signal, can be easily derived⁹ directly from the

⁹To derive the self-similar power spectrum, consider an D -dimensional signal with power spectrum $S(\omega)$, and parameterize the D -dimensional frequency space using the D -dimensional correspondence to spherical coordinates, $(r; \varphi_1, \dots, \varphi_{D-1})$, where $r = |\omega| \in [0, \infty]$ and $\varphi_1, \dots, \varphi_{D-1}$ are suitably selected angles in some domain Ω . To analyse the energy contribution from each range of frequencies, consider a volume element dV defined by $r_0 \leq |\omega| \leq r_0(1 + d\rho)$ for some r_0 . Since the area of an D -dimensional hypersphere of radius r_0 is proportional to r_0^{D-1} , the volume of a scale-invariant element can be written $dV = C_D r_0^{D-1} r_0 d\rho$ for some constant C_D . If we want the signal to contain the same amount of energy for all frequencies (where the total energy is measured by $E = \int_{\omega \in \mathbb{R}^D} S(\omega) d\omega$), it follows by necessity that $dE = \int_{dV} (\int_{\varphi \in \Omega} S(r; \varphi) d\varphi) dV$ must be independent of ω , which in turn implies that $(\int_{\varphi \in \Omega} S(r; \varphi) d\varphi) dV$ must be proportional to $d\rho$, and the power spectrum must be of the form $S(\omega) \sim |\omega|^{-D}$.

assumption that the power spectrum should contain the same amount of energy for all frequencies.

9.3 Relations to previous work

Such L_1 -normalized kernels of first order have been used, for example, in edge detection and edge classification by (Korn 1988), (Mallat and Zhong 1992), and (Zhang and Bergholm 1993), and in pyramids by (Crowley and Parker 1984). More generally, evolution properties across scales of wavelet transforms have been used by (Mallat and Hwang 1992) for characterizing local Lipschitz exponents of singularities. (Mallat and Hwang 1992) also proposed the notion of “general maxima” of wavelet transforms for estimating the frequency of local oscillations. This idea is closely related to the notion of scale-space maximum considered here and to the scale selection mechanism in (Lindeberg 1991, 1993a) based on local maxima over scales of blob responses computed along extremum paths in scale-space. There is also a connection to the “top point” representation proposed by (Johansen *et al.* 1986) in the sense that the points in the scale-space at which bifurcations occur serve as to delimit extremum paths with different topology. A main difference between the scale selection mechanism suggested here and the work in (Lindeberg 1991) and (Mallat and Hwang 1992), however, is that here it is shown how these notions can be applied to large classes of non-linear differential invariants computed in a scale-space representation. Moreover, feature detection algorithms have been formulated with integrated scale selection mechanisms and it has been shown how different derivative normalization approaches lead to different classes of differential expressions for which the scale selection mechanism commutes with rescalings of the input pattern. Specifically, it has been shown how L_1 -normalization is special in terms of scale invariance properties.¹⁰

10 Summary and discussion

We have argued that the subject of scale selection is essential to many problems in computer vision and automated image analysis. Specifically, we have outlined how the evolution properties over scales of normalized Gaussian derivatives provide important cues in this context—for generating hypotheses about interesting scales (and associated spatial points or regions) for further analysis. A general scale selection principle has been presented stating that in the absence of other evidence, coarse estimates of the size of image structures can be computed from the scales at which normalized differential geometric descriptors assume maxima over scales. In particular, it has been suggested that this approach can be used for adaptively choosing the scales for feature detection. Support for this idea has been provided in terms of a theoretical analysis of the general scaling property of local maxima over scales in the scale-space signature, and by a detailed analysis of the behaviour of the scale selection method when integrated with feature detection algorithms and applied to characteristic model patterns; see table 2 for an overview. The main support of the methodology is, however, experimental; it has been demonstrated that intuitively reasonable and quantitatively accurate results can be obtained by applying the proposed scheme to the problems of blob detection, junction detection and frequency estimation.¹¹

For a problem such as junction detection, the methodology naturally gives rise to two-stage feature detection algorithms, where features are first detected at locally

¹⁰ Applications of scale selection based on L_p -normalization with $p < 1$ are developed in more detail in (Lindeberg 1996a).

¹¹ See (Lindeberg 1996a) concerning scale selection mechanisms for detecting edges and ridges.

Feature type	Differential entity for scale selection	γ -value	L_p -norm	Fourier β
Edge	$t^{\gamma/2} L_v$	1/2	4/5	3/2
Ridge	$t^{2\gamma} (L_{pp} - L_{qq})^2$	3/4	4/5	3/2
Corner	$t^{2\gamma} L_v^2 L_{uu}$	1	1	1
Blob	$t^\gamma \nabla^2 L$	1	1	1

Table 2: Measures of feature strength and normalization parameters used for different types of feature detectors with automatic scale selection (including results from a companion paper (Lindeberg 1996c, 1996b)). For each feature detector, a preferred γ -value is specified as well as the p -value for which the L_p -norm of the Gaussian derivatives is constant over scales (76) and the β -value for which the energy of a self-similar Fourier spectrum is constant over scales (89).

adapted coarse scales, and then localized to finer scales in a second stage processing stage. Whereas the general advantages of such a two-stage approach to feature detection are well-known in the literature, a major contribution here is that explicit mechanisms are provided for automatic selection of the detection scales as well as the localization scales. Moreover, these processing stages are integrated into algorithms which are essentially free from other tuning parameters that the number of features of interest.

Of course, the task of selecting “the best scale” for handling real-world image data (about which usually no or very little *a priori* information is available) is intractable if treated as a pure mathematical problem. Therefore, the proposed scale selection principle should not be interpreted as any “optimal solution”, but rather as a systematic method for generating initial hypotheses in situations where no or very little information is available about what can be expected to be in the scene.

10.1 Technical contributions

At a technically more detailed level some of the main contributions are that:

- It is emphasized how the evolution properties over scales of normalized scale-space derivatives differ from those of traditional spatial derivatives. Whereas the magnitude of a traditional scale-space derivative always decreases with scale, peaks over scales can be expected in the scale-signatures of normalized derivatives computed from data containing dominant information at certain scales.
- A general scale selection principle for scale selection has been proposed stating that extrema over scales in the signature of normalized differential entities are useful in the stage of detecting image features. In particular, it has been theoretically analyzed and experimentally demonstrated how extrema over scales of the following differential entities can be used in feature detection;
 - maxima and minima in the rescaled level curve curvature are highly useful for *junction detection*,
 - maxima and minima of the trace and the determinant of the Hessian matrix can serve as qualitative descriptors for *blob detection*, and
 - local maxima over scales of the pseudo quadrature measure can be used for local *frequency estimation*.

The problem of junction detection is treated more extensively, and the resulting method is the first junction detection algorithm with automatic scale selection.

- It is shown that the γ -normalized derivative concept arises by necessity given natural scale invariance requirements on a scale selection mechanism.
- It is explained how the localization of junction candidates can be improved by invoking a modified Förstner operator adapted to the local image structure. Specifically, it is shown how localization scales can be selected automatically by minimizing a certain normalized residual across scales.

The same mechanism can be used for selecting localization scales in edge detection, which constitute a trade-off between the diffuseness of the edge and the noise level.

- It is shown how the normalized derivative concept can be discretized in a consistent manner (see appendix A.4).

A Appendix

A.1 Necessity of the form of the γ -parameterized derivative operator

In this appendix it is shown that the form of the γ -normalized derivative operator

$$\partial_{\xi} = t^{\gamma/2} \partial_x \quad (92)$$

can be derived by necessity, given natural assumptions on a scale selection procedure based on local maxima over scales. Let us follow the general methodology proposed in section 4 and state the following requirements on a normalized derivative concept:

- The main idea of introducing a normalized derivative operator should be to compensate for the general decrease in amplitude caused by scale-space smoothing, and to ensure that image structures are processed by the vision system in such a way that the results are not critically dependent upon how large the image structures actually are.

In the absence of further information, the main source of information for normalizing the derivative operator should be the value of the scale parameter. Moreover, this normalization should allow a differential descriptor to be expressed in terms of normalized derivatives in a similar way as traditional image descriptors are expressed in terms of unnormalized derivatives. Thereby, it is natural to perform the normalization as a multiplicative factor.

Motivated by these qualitative requirements, the m th order normalized derivative operator ∂_{ξ^m} at a certain scale t should therefore be defined from the ordinary spatial derivative operator ∂_{x^m} by

$$\partial_{\xi^m} = \varphi(t) \partial_{x^m} \quad (93)$$

where $\varphi: \mathbb{R}_+ \rightarrow \mathbb{R}$ is a smooth function.

- To allow for scale selection based on local maxima over scales, the normalized derivative concept must preserve local maxima over scales. Specifically, if a normalized differential entity assumes a local maximum over scales at a certain point (x_0, t_0) in scale-space, then after a rescaling of the input signal by a factor of s , the maximum over scales should be assumed at $(s x_0, s^2 t_0)$.

Given these requirements, it follows that the normalization must be of the form

$$\varphi(t) = A t^B \quad (94)$$

for some constants A and B .

Proof: Consider two signals f and f' related by a scaling factor s such that $f(x) = f'(sx)$, and define the normalized m th order derivatives by

$$L_{\xi^m} = t^{\gamma/2} L_{x^m}, \quad (95)$$

$$L_{\xi'^m} = t'^{\gamma/2} L_{x'^m}. \quad (96)$$

From (20) we have that these derivatives at corresponding points $(x'; t') = (sx, s^2t')$ are related by

$$L_{\xi^m}(x; t) = s^m \frac{\varphi(t)}{\varphi(s^2t)} L_{\xi^m}(sx; s^2t). \quad (97)$$

If a local maximum over scales is to be preserved, we must require that

$$\partial_t(L_{\xi^m}(x; t)) = 0 \quad \Leftrightarrow \quad \partial_{t'}(L_{\xi'^m}(x'; t')) = 0. \quad (98)$$

Differentiating (97) with respect to t gives

$$\partial_t(L_{\xi^m}(x; t)) = s^m \left(\partial_t \left(\frac{\varphi(t)}{\varphi(s^2t)} \right) + \partial_t(L_{\xi^m}(sx; s^2t)) \right) \quad (99)$$

and application of (98) results in the necessary requirement:

$$\partial_t \left(\frac{\varphi(t)}{\varphi(s^2t)} \right) = 0. \quad (100)$$

Rewrite φ as

$$\varphi(t) = e^{\psi(\log t)} \quad (101)$$

and introduce $u = \log t$. Moreover, make use of the fact that (100) implies that the ratio $\varphi(t)/\varphi(s^2t)$ cannot depend on t and must be a function of s only. With $v = 2 \log s$, (100) can hence be rewritten as

$$\frac{e^{\psi(\log t)}}{e^{\psi(\log(s^2t))}} = \tilde{h}(s) \quad (102)$$

for some function \tilde{h} , and be reduced to the relation

$$\psi(u) - \psi(u + v) = h(v) \quad (103)$$

for some other (arbitrary) function h . If we differentiate with respect to u and v

$$\begin{cases} \psi'(u) - \psi'(u + v) = 0, \\ \psi'(u + v) + h'(v) = 0, \end{cases} \quad (104)$$

and add up the equations

$$\psi'(u) + h'(v) = 0, \quad (105)$$

we find that ψ' (and h') must be constant. In other words, $\psi(u) = C_1 u + C_2$ and

$$\varphi(t) = e^{C_1 \log t + C_2} = A t^B. \quad (106)$$

A.2 L_p -normalization interpretation of γ -normalized derivatives

In this appendix it is shown how the γ -normalized derivative concept can be interpreted in terms of L_p -normalization of the Gaussian derivative kernels over scales. The L_p -norm of the m th order γ -normalized Gaussian derivative kernel is given by

$$\|g_{\xi^m}(\cdot; t)\|_p^p = \int_{x=-\infty}^{\infty} |t^{m\gamma/2} g_x^m(x; t)|^p dx. \quad (107)$$

From the well-known relation between the derivatives of the unnormalized Gaussian kernel and the Hermite polynomials H_n $\partial_{x^m}(e^{-x^2}) = (-1)^m H_m(x) e^{-x^2}$ (Abramowitz and Stegun 1964) it follows that the m th order Gaussian derivative kernel can be written

$$\partial_{x^m} g(x; t) = (-1)^m \frac{1}{(2t)^{m/2}} H_m\left(\frac{x}{\sqrt{2t}}\right) g(x; t). \quad (108)$$

Insert this expression into (107) and make the substitution $x = \sqrt{t}u$. Then, by exploiting the fact that $g(x; t) = \frac{1}{\sqrt{t}} g(\frac{x}{\sqrt{t}}; 1)$, we obtain

$$\begin{aligned} \|g_{\xi^m}(\cdot; t)\|_p^p &= \int_{u=-\infty}^{\infty} \left| \frac{t^{m(\gamma-1)/2}}{2^{m/2}} H_m\left(\frac{u}{\sqrt{2}}\right) \frac{1}{\sqrt{t}} g(u; 1) \right|^p \sqrt{t} du \\ &= t^{p(m\gamma-m-1)/2+1/2} \int_{u=-\infty}^{\infty} \left| \frac{1}{2^{m/2}} H_m\left(\frac{u}{\sqrt{2}}\right) g(u; 1) \right|^p \sqrt{t} du \\ &= t^{p(m\gamma-m-1)/2+1/2} \|g_{\xi^m}(\cdot; 1)\|_p^p. \end{aligned} \quad (109)$$

In other words, the L_p -norm of the Gaussian derivative kernel at scale level t is related to the L_p -norm at unit scale by

$$\|g_{\xi^m}(\cdot; t)\|_p = \sqrt{t}^{m(\gamma-1)-1+1/p} \|g_{\xi^m}(\cdot; 1)\|_p. \quad (110)$$

Concerning L_p -norms of Gaussian derivatives in higher dimensions, we can make use of the separability of the normalized Gaussian derivative to separate the D -dimensional integral in the L_p -norm definition into a product of one-dimensional integrals of the form (109). Hence, if we let $|m| = m_1 + \dots + m_D$ denote the total order of differentiation, it follows that the variation over scales of the L_p -norm of the D -dimensional normalized Gaussian derivative kernel will be of the form

$$\|g_{\xi^m}(\cdot; t)\|_p = \sqrt{t}^{|m|(\gamma-1)+D(1/p-1)} \|g_{\xi^m}(\cdot; 1)\|_p. \quad (111)$$

In other words, the L_p -norm of this kernel is constant over scales if and only if $m(\gamma-1) + D(1/p-1) = 0$. Specifically, p is independent of m if and only if $\gamma = 1$.

A.3 Normalized derivative responses to self-similar power spectra

In this appendix, a closed-form expression is derived for the variation over scales of the following class of energy measures

$$E_m = \int_{x \in \mathbb{R}^D} \sum_{|\alpha|=m} t^{m\gamma} |L_{x^\alpha}|^2 dx. \quad (112)$$

when computed at different scales t in the scale-space representation L of a two-dimensional signal f with a self-similar power spectrum of the form

$$\begin{aligned} S_f(\omega_1, \dots, \omega_D) &= (\hat{f} \hat{f}^*)(\omega) \\ &= |\omega|^{-2\beta} = (\omega_1^2 + \omega_2^2)^{-\beta}. \end{aligned} \quad (113)$$

Using Plancherel's relation

$$\int_{\omega \in \mathbb{R}^D} \hat{h}_1(\omega) \hat{h}_2^*(\omega) d\omega = (2\pi)^D \int_{x \in \mathbb{R}^2} h_1(x) h_2^*(x) dx, \quad (114)$$

where $\hat{h}_i(\omega)$ is the Fourier transform of $h_i(x)$ and by letting $h_1 = h_2 = L_{x^\alpha} = L_{x_1^{\alpha_1} \dots x_D^{\alpha_D}}$, we obtain

$$\int_{x \in \mathbb{R}^2} L_{x_1^{\alpha_1} \dots x_D^{\alpha_D}}^2(\cdot; t) dx = \frac{1}{(2\pi)^D} \int_{\omega \in \mathbb{R}^D} \omega_1^{2\alpha_1} \dots \omega_D^{2\alpha_D} \hat{g}^2(\omega; t) |\omega|^{-2\alpha} d\omega \quad (115)$$

where \hat{g} denotes the Fourier transform of the Gaussian kernel. By adding (115) over all possible multi-indices α with $|\alpha| = \alpha_1 + \dots + \alpha_D = m$, and by using the definition (82) we obtain the rotationally invariant descriptor

$$E_m = \frac{t^{m\gamma}}{(2\pi)^D} \int_{\omega \in \mathbb{R}^D} |\omega|^{2m} \hat{g}^2(\omega; t) |\omega|^{-2\beta} d\omega \quad (116)$$

Let us next introduce the D -dimensional correspondence to spherical coordinates, $(\rho; \varphi_1, \dots, \varphi_{D-1})$, with the volumetric element $d\omega = C_D \rho^{D-1} d\rho$

$$E_m = \frac{t^{m\gamma}}{(2\pi)^D} \int_{\rho=0}^{\infty} \rho^{2m-2\beta+D-1} e^{-\rho^2 t} d\rho \quad (117)$$

Then, using $\int_0^\infty x^m e^{-ax^2} dx = \frac{\Gamma((m+1)/2)}{2a^{(m+1)/2}}$, we get

$$E_m = \frac{C_D t^{m\gamma}}{(2\pi)^D} \frac{\Gamma(m - \beta + D/2)}{2 t^{m-\beta+D/2}}, \quad (118)$$

and the variation over scales is of the form

$$E_m(t) \sim t^{\beta-D/2-m(1-\gamma)}. \quad (119)$$

A.4 Discrete implementation of the scale selection mechanisms

Discretizing the normalized derivative operators leads to two discretization problems; (i) how to discretize the scale-space derivatives such that the scale-space properties are preserved, and (ii) how to discretize the normalization factor.

A.4.1 Computing discrete derivative approximations

The first problem can be solved by using the scale-space concept for discrete signals (Lindeberg 1994d), given by $L(\cdot, \cdot; t) = T(\cdot, \cdot; t) * f(\cdot, \cdot)$, where $T(m, n; t) = T_1(m; t) T_1(n; t)$ and $T_1(m; t) = e^{-t} I_m(t)$ is the one-dimensional discrete analogue of the Gaussian kernel defined from the modified Bessel functions I_n .

The scale-space properties of L transfer to any discrete derivative approximations $L_{x_1^i x_2^j}$ defined as the result of applying difference operators $\delta_{x_1^i x_2^j}$ to L . In the implementations presented here, the first- and second-order derivatives are approximated by the operators $(\delta_x L)(x; t) = \frac{1}{2}(L(x+1; t) - L(x-1; t))$ and $(\delta_{xx} L)(x; t) = (L(x+1; t) - 2L(x; t) + L(x-1; t))$, respectively.

A.4.2 Normalization in the discrete case

In view of the results presented in section 9.1, it is natural to normalize the discrete derivative approximation kernels $\delta_{x_1 x_2}^i T$ such that their discrete l_1 -norms will be constant over scales. Of course, it is not necessary to construct the normalized derivative approximation kernels explicitly. Concerning *e.g.* first order derivatives, discrete approximations to L_{x_1} and L_{x_2} can first be computed according to section A.4.1. Then, the result can be multiplied by the discrete normalization factor

$$\alpha_1(t) = \frac{\sqrt{2}}{\sqrt{\pi}(T_1(0; t) + T_1(1; t))}, \quad (120)$$

which has been determined such that

$$\sum_{x=-\infty}^0 \alpha_1(t) (\delta_x T)(x; t) = \int_{x=-\infty}^0 \sqrt{t} (\partial_x g)(x; t) dx = \frac{1}{\sqrt{2\pi}}. \quad (121)$$

Using an asymptotic expression for the modified Bessel functions of integer order (Abramowitz and Stegun 1964) $I_n(t) = \frac{e^t}{\sqrt{2\pi t}} (1 - \frac{4n^2-1}{8t} + O(\frac{1}{t^2}))$, it can be verified that $\alpha_1(t)$ approaches the continuous normalization factor \sqrt{t} when the scale parameter t increases. Observe, however, that $\alpha_1(t)$ assumes a non-zero value when $t = 0$, in contrast to the continuous normalization factor \sqrt{t} , which forces any normalized derivative to be zero in unsmoothed data. This property of the L_p -normalization approach is important if we want to capture peaks at fine scales in the scale-space signatures. In (Lindeberg 1995b) it is shown that discrete L_p -normalization also has better performance than variance based normalization when expressing scale selection mechanisms in subsampled multi-scale representations.

A.4.3 Detection of scale-space maxima in discrete data

Given the abovementioned discretization methods for computing normalized differential descriptors based on the local N -jet representation, it is straightforward to express algorithms for detecting scale-space maxima. In summary, the implementations underlying this work have been performed as follows:

1. Given a discrete image f (here: of size between 128×128 or 256×256 pixels), select a scale range for the analysis (here: $t_{min} = 2$ and $t_{max} = 256$). Within this range, distribute a set of scale levels t_k (here: 20 or 40 levels) such that the ratio between successive scale levels t_{k+1}/t_k is approximately constant.¹²
2. For each scale t_k , compute the scale-space representation of f according to section A.4.1. Then, for each point at each scale, compute discrete derivative approximations according to section A.4.1 and normalize them according to section A.4.2. Finally, compute the normalized differential expression by pointwise combination of these entities.
3. In the three-dimensional volume so generated, detect local maxima (as points whose values are greater than or equal to the values of their 26 discrete neighbours).¹³ Optionally, select the N points having the strongest normalized responses.

¹²Specifically, the scale levels have been determined such that the difference $\tau_{k+1} - \tau_k$ in effective scale between adjacent scales t_{k+1} and t_k is constant (see footnote 2).

¹³In other words, a point (x_0, y_0, t_{k_0}) is regarded as a discrete scale-space maximum of a normalized differential entity $\mathcal{D}_{norm} L$ if and only if $(\mathcal{D}_{norm} L)(x_0, y_0, t_{k_0}) \geq (\mathcal{D}_{norm} L)(x_0 + i, y_0 + j, t_{k_0+k})$ holds for all 26 neighbouring points $(i, j, k) \in \{-1, 0, 1\}$.

A.4.4 *Implementing junction localization*

For each scale-space maximum $(x_0; t_0)$ detected according to the methodology in section A.4.3, the subsequent junction localization step is performed as follows:

- Compute A , b and c according to the definitions in (60), (61) and (62) using a Gaussian window function $w_{x_0}(x') = g(x' - x_0; t_0)$ with integration scale value t_0 equal to the detection scale t_0 of the scale-space maximum and with the center at the candidate junction x_0 . At a number of scales (here: 5–10 levels), uniformly distributed between a lower scale (here: 0.01) and the detection scale t_0 , vary the local scale at which derivatives are computed and select the local scale that minimizes the normalized residual (63) over scales.
- At this scale, the new localization estimate is $\hat{x} = A^{-1}b$.
- Iterate the abovementioned steps until either the increment is sufficiently small (here: within the same pixel) or an upper bound (here: 3 iterations) has been reached. Suppress all points for which the scheme diverges (here: when the total update is larger than the detection scale measured in dimension [length]).

References

- M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. Applied Mathematics Series. National Bureau of Standards, 55 edition, 1964.
- A. Almansa and T. Lindeberg. “Enhancement of Fingerprint Images by Shape-Adapted Scale-Space Operators”. In J. Sporring; M. Nielsen; L. Florack, and P. Johansen, editors, *Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory*, Copenhagen, Denmark, May. 1996. Kluwer Academic Publishers.
- A. Almansa and T. Lindeberg. “Fingerprint Enhancement by Shape Adaptation of Scale-Space Operators with Automatic Scale-Selection”. Technical Report ISRN KTH/NA/P--98/03--SE, Dept. of Numerical Analysis and Computing Science, KTH, Stockholm, Sweden, Apr. 1998.
- J. Babaud; A. P. Witkin; M. Baudin, and R. O. Duda. “Uniqueness of the Gaussian Kernel for Scale-Space Filtering”. *IEEE Trans. Pattern Analysis and Machine Intell.*, 8(1):26–33, 1986.
- J. Blom. *Topological and Geometrical Aspects of Image Structure*. PhD thesis. , Dept. Med. Phys. Physics, Univ. Utrecht, NL-3508 Utrecht, Netherlands, 1992.
- D. Blostein and N. Ahuja. “A Multiscale Region Detector”. *Computer Vision, Graphics, and Image Processing*, 45:22–41, 1989.
- L. Bretzner and T. Lindeberg. “Feature Tracking with Automatic Selection of Spatial Scales”. Technical Report ISRN KTH/NA/P--96/21--SE, Dept. of Numerical Analysis and Computing Science, KTH, Stockholm, Sweden, Jun. 1996.
- L. Bretzner and T. Lindeberg. “On the Handling of Spatial and Temporal Scales in Feature Tracking”. In B. M. ter Haar Romeny; L. M. J. Florack; J. J. Koenderink, and M. A. Viergever, editors, *Scale-Space Theory in Computer Vision: Proc. First Int. Conf. Scale-Space’97*, volume 1252 of *Lecture Notes in Computer Science*, pages 128–139, Utrecht, The Netherlands, July 1997. Springer Verlag, New York.
- L. Bretzner and T. Lindeberg. “Feature tracking with automatic selection of spatial scales”. *Computer Vision and Image Understanding*, 1998. (To appear).
- K. Brunnström; T. Lindeberg, and J.-O. Eklundh. “Active detection and classification of junctions by foveation with a head-eye system guided by the scale-space primal sketch”. In G. Sandini, editor, *Proc. 2nd European Conf. on Computer Vision*, volume 588 of *Lecture Notes in Computer Science*, pages 701–709, Santa Margherita Ligure, Italy, May. 1992. Springer-Verlag.
- P. J. Burt. “Fast Filter Transforms for Image Processing”. *Computer Vision, Graphics, and Image Processing*, 16:20–51, 1981.
- J. L. Crowley and A. C. Parker. “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, 1984.
- J. L. Crowley. *A Representation for Visual Information*. PhD thesis. , Carnegie-Mellon University, Robotics Institute, Pittsburgh, Pennsylvania, 1981.
- R. Deriche and G. Giraudon. “Accurate Corner Detection: An Analytical Study”. In *Proc. 3rd Int. Conf. on Computer Vision*, pages 66–70, Osaka, Japan, 1990.
- L. Dreschler and H.-H. Nagel. “Volumetric Model and 3D-Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a Street Scene”. *Computer Vision, Graphics, and Image Processing*, 20(3):199–228, 1982.
- D. J. Field. “Relations between the statistics of natural images and the response properties of cortical cells”. *J. of the Optical Society of America*, 4:2379–2394, 1987.
- L. M. J. Florack. *The Syntactical Structure of Scalar Images*. PhD thesis. , Dept. Med. Phys. Physics, Univ. Utrecht, NL-3508 Utrecht, Netherlands, 1993.

- L. M. J. Florack; B. M. ter Haar Romeny; J. J. Koenderink, and M. A. Viergever. "Scale and the Differential Structure of Images". *Image and Vision Computing*, 10(6):376–388, Jul. 1992.
- L. M. J. Florack; B. M. ter Haar Romeny; J. J. Koenderink, and M. A. Viergever. "Linear scale-space". *J. of Mathematical Imaging and Vision*, 4(4):325–351, 1994.
- W. A. Förstner and E. Gülch. "A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centers of Circular Features". In *Proc. Intercommission Workshop of the Int. Soc. for Photogrammetry and Remote Sensing*, Interlaken, Switzerland, 1987.
- J. Gårding and T. Lindeberg. "Direct computation of shape cues using scale-adapted spatial derivative operators". *Int. J. of Computer Vision*, 17(2):163–191, 1996.
- B. ter Haar Romeny, editor. *Geometry-Driven Diffusion in Computer Vision*. Series in Mathematical Imaging and Vision. Kluwer Academic Publishers, Dordrecht, Netherlands, 1994.
- P. Johansen; S. Skelboe; K. Grue, and J. D. Andersen. "Representing Signals by Their Top Points in Scale-Space". In *Proc. 8:th Int. Conf. on Pattern Recognition*, pages 215–217, Paris, France, Oct. 1986.
- L. Kitchen and A. Rosenfeld. "Gray-Level Corner Detection". *Pattern Recognition Letters*, 1(2):95–102, 1982.
- J. J. Koenderink and W. Richards. "Two-Dimensional Curvature Operators". *J. of the Optical Society of America*, 5:7:1136–1141, 1988.
- J. J. Koenderink and A. J. van Doorn. "Receptive Field Families". *Biological Cybernetics*, 63:291–298, 1990.
- J. J. Koenderink and A. J. van Doorn. "Generic neighborhood operators". *IEEE Trans. Pattern Analysis and Machine Intell.*, 14(6):597–605, Jun. 1992.
- J. J. Koenderink. "The structure of images". *Biological Cybernetics*, 50:363–370, 1984.
- A. F. Korn. "Toward a Symbolic Representation of Intensity Changes in Images". *IEEE Trans. Pattern Analysis and Machine Intell.*, 10(5):610–625, 1988.
- T. Lindeberg and J. Gårding. "Shape from Texture from a Multi-Scale Perspective". In H.-H. Nagel et. al., editor, *Proc. 4th Int. Conf. on Computer Vision*, pages 683–691, Berlin, Germany, May. 1993. IEEE Computer Society Press.
- T. Lindeberg and J. Gårding. "Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D structure". *Image and Vision Computing*, 15:415–434, 1997.
- T. Lindeberg and M. Li. "Segmentation and classification of edges using minimum description length approximation and complementary junction cues". In G. Borgefors, editor, *Proc. 9th Scandinavian Conference on Image Analysis*, pages 767–776, Uppsala, Sweden, June 1995. Swedish Society for Automated Image Processing.
- T. Lindeberg and M. Li. "Segmentation and classification of edges using minimum description length approximation and complementary junction cues". *Computer Vision and Image Understanding*, 67(1):88–98, 1997.
- T. Lindeberg and G. Olofsson. "The Aspect Feature Graph in Recognition by Parts". draft manuscript, 1995.
- T. Lindeberg. "Scale-Space for Discrete Signals". *IEEE Trans. Pattern Analysis and Machine Intell.*, 12(3):234–254, Mar. 1990.
- T. Lindeberg. *Discrete Scale-Space Theory and the Scale-Space Primal Sketch*. Ph. D. dissertation. Ph. D. dissertation, Dept. of Numerical Analysis and Computing Science, KTH, Stockholm, Sweden, May. 1991. ISRN KTH/NA/P--91/08--SE. An extended and revised version published as book "Scale-Space Theory in Computer Vision" in The Kluwer International Series in Engineering and Computer Science.

- T. Lindeberg. "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, Dec. 1993.
- T. Lindeberg. "On Scale Selection for Differential Operators". In K. Heia K. A. Høgdra, B. Braathen, editor, *Proc. 8th Scandinavian Conf. on Image Analysis*, pages 857–866, Tromsø, Norway, May. 1993. Norwegian Society for Image Processing and Pattern Recognition.
- T. Lindeberg. "Junction detection with automatic selection of detection scales and localization scales". In *Proc. 1st International Conference on Image Processing*, volume I, pages 924–928, Austin, Texas, Nov. 1994. IEEE Computer Society Press.
- T. Lindeberg. "On the Axiomatic Foundations of Linear Scale-Space: Combining Semi-Group Structure with Causality vs. Scale Invariance". Technical Report ISRN KTH/NA/P--94/20--SE, Dept. of Numerical Analysis and Computing Science, KTH, Stockholm, Sweden, Aug. 1994. Extended version to appear in J. Sporring and M. Nielsen and L. Florack and P. Johansen (eds.) *Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory*, Copenhagen, Denmark, Kluwer Academic Publishers, May 1996.
- T. Lindeberg. "Scale Selection for Differential Operators". Technical Report ISRN KTH/NA/P--94/03--SE, Dept. of Numerical Analysis and Computing Science, KTH, Stockholm, Sweden, Jan. 1994. Extended version in *Int. J. of Computer Vision*, vol 30, number 2, 1998. (In press).
- T. Lindeberg. *Scale-Space Theory in Computer Vision*. The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, Dordrecht, Netherlands, 1994.
- T. Lindeberg. "Direct Estimation of Affine Deformations of Brightness Patterns Using Visual Front-End Operators with Automatic Scale Selection". In *Proc. 5th International Conference on Computer Vision*, pages 134–141, Cambridge, MA, June 1995.
- T. Lindeberg. "On scale selection in subsampled (hybrid) multi-scale representations". 1995. draft manuscript.
- T. Lindeberg. "Edge detection and ridge detection with automatic scale selection". 1996. (Submitted).
- T. Lindeberg. "Edge detection and ridge detection with automatic scale selection". In *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition, 1996*, pages 465–470, San Francisco, California, June 1996. IEEE Computer Society Press.
- T. Lindeberg. "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, Stockholm, Sweden, Jan. 1996. Also in *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition, 1996*. Extended version in *Int. J. of Computer Vision*, vol 30, number 2, 1998. (In press).
- T. Lindeberg. "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, Stockholm, Sweden, Apr. 1996. to appear in *Image and Vision Computing*.
- T. Lindeberg. "On Automatic Selection of Temporal Scales in Time-Casual Scale-Space". In G. Sommer and J. J. Koenderink, editors, *Proc. AFPAC'97: Algebraic Frames for the Perception-Action Cycle*, volume 1315 of *Lecture Notes in Computer Science*, pages 94–113, Kiel, Germany, September 1997. Springer Verlag, Berlin.
- S. G. Mallat and W. L. Hwang. "Singularity Detection and Processing with Wavelets". *IEEE Trans. Information Theory*, 38(2):617–643, 1992.
- S. G. Mallat and S. Zhong. "Characterization of Signals from Multi-Scale Edges". *IEEE Trans. Pattern Analysis and Machine Intell.*, 14(7):710–723, 1992.
- D. Marr. *Vision*. W.H. Freeman, New York, 1982.

- J. A. Noble. "Finding Corners". *Image and Vision Computing*, 6(2):121–128, 1988.
- E. J. Pauwels; P. Fiddelaers; T. Moons, and L. J. van Gool. "An extended class of scale-invariant and recursive scale-space filters". *IEEE Trans. Pattern Analysis and Machine Intell.*, 17(7):691–701, 1995.
- H. Voorhees and T. Poggio. "Detecting Textons and Texture Boundaries in Natural Images". In *Proc. 1st Int. Conf. on Computer Vision*, London, England, 1987.
- K. Wiltschi; A. Pinz, and T. Lindeberg. "Classification of Carbide Distributions using Scale Selection and Directional Distributions". In *Proc. 4th International Conference on Image Processing*, volume II, pages 122–125, Santa Barbara, California, October 1997. IEEE.
- A. P. Witkin. "Scale-space filtering". In *Proc. 8th Int. Joint Conf. Art. Intell.*, pages 1019–1022, Karlsruhe, West Germany, Aug. 1983.
- R. A. Young. "The Gaussian Derivative Theory of Spatial Vision: Analysis of Cortical Cell Receptive Field Line-Weighting Profiles". Technical Report GMR-4920, Computer Science Department, General Motors Research Lab., Warren, Michigan, 1985.
- R. A. Young. "The Gaussian derivative model for spatial vision: I. Retinal mechanisms". *Spatial Vision*, 2:273–293, 1987.
- A. L. Yuille and T. A. Poggio. "Scaling Theorems for Zero-Crossings". *IEEE Trans. Pattern Analysis and Machine Intell.*, 8:15–25, 1986.
- W. Zhang and F. Bergholm. "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 H.-H. Nagel et. al., editor, *Proc. 4th Int. Conf. on Computer Vision*, pages 183–191, Berlin, Germany, May. 1993. IEEE Computer Society Press.