

On Scale Selection for Differential Operators

Tony Lindeberg

Computational Vision and Active Perception Laboratory (CVAP)*
Royal Institute of Technology (KTH), Stockholm, Sweden

Abstract: Although traditional scale-space theory provides a well-founded framework for dealing with image structures at different scales, it does not directly address the problem of how to *select* appropriate scales for further analysis.

This paper introduces a new tool for dealing with this problem. A heuristic principle is proposed stating that *local extrema over scales* of different combinations of *normalized scale invariant derivatives* are likely candidates to correspond to interesting structures. Support is given by theoretical considerations and experiments on real and synthetic data.

The resulting methodology lends itself naturally to two-stage algorithms; feature detection at coarse scales followed by feature localization at finer scales. Experiments on blob detection, junction detection and edge detection demonstrate that the proposed method gives intuitively reasonable results.

1 Introduction

An inherent property of objects in the world and details in images is that they only exist as meaningful entities over certain ranges of scale. The *scale-space representation*, introduced by Witkin [28] and Koenderink [13], provides a methodology for handling such size or scale variations in image data. The basic idea is to embed any measured signal into a one-parameter family of gradually smoothed and simplified signals, in which the fine scale information is successively suppressed.

For continuous signals $f : \mathbb{R}^N \rightarrow \mathbb{R}$ the scale-space $L : \mathbb{R}^N \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is defined as the solution to the diffusion equation

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

*I would like to thank Jonas Gårding for collaboration on parallel work on shape-from-texture and stereo analysis, as well as valuable suggestions about this manuscript. The support from the Swedish National Board for Industrial and Technical Development, NUTEK, is gratefully acknowledged.
Address: NADA, KTH, S-100 44 Stockholm, Sweden.
Email: tony@bion.kth.se

with initial condition $L(\cdot; 0) = f(\cdot)$, or equivalently, by convolution with the Gaussian kernel $L(\cdot; t) = g(\cdot; t) * f(\cdot)$, where $g : \mathbb{R}^N \times \mathbb{R}_+ \rightarrow \mathbb{R}$ is given by

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

and $x = (x_1, \dots, x_N)^T$. There are several mathematical results (see e.g. [13, 15, 1, 29, 17, 22, 10]) indicating that within the class of linear transformations this scale-space theory describes the canonical way to formulate a multi-scale representation. In fact, assuming that the first stages of visual processing, the *visual front end*, are to perform linear operations and be invariant to translations, rotations and rescalings in space, it can be shown that the Gaussian kernels and their associated smoothed derivatives

$$L_{x^\alpha}(\cdot; t) = \partial_{x^\alpha} (g * f) = (\partial_{x^\alpha} g) * f = g * (\partial_{x^\alpha} f), \quad (3)$$

at various scales arise by necessity. By combining the output from such *Gaussian derivative* operators at any specific scale, smoothed differential descriptors can be defined at that scale. Defining such descriptors at *all* scales gives a multi-scale differential geometric representation of the signal; a type of representation that is useful for a large number of early vision tasks.

Although this (traditional) scale-space theory provides a well-founded framework for handling image structures at different scales, it does not directly address the problem of *selecting* appropriate scales and structures from the scale-space representation for further analysis. Early work in this direction was performed by Bischof and Caelli [3] concerning the zero-crossings of the Laplacean operator. Another approach was developed by Lindeberg [18, 19], who considered blob-like structures at different scales in scale-space and constructed a multi-scale tree-like representation called the *scale-space primal sketch*. A significance measure was postulated as the volume that certain primitives of the representation, called scale-space blobs, occupy in scale-space. The scale levels in turn were determined from scales where the scale-space blobs assumed their maximum (normalized) blob response over scales. Experimentally it was

demonstrated that the 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 [18, 7].

In this paper a related but not identical parameter variation principle will be introduced for finding interesting scales — the evolution properties over scales of different nonlinear combinations of *normalized scale-invariant derivatives* of low order computed from the scale-space representation. The heuristic principle to be employed is that *local maxima over scale* of these entities correspond to interesting structures. By theoretical considerations and numerical experiments it will be demonstrated that the proposed methodology gives useful and intuitively reasonable results in different types of situations.

2 Basic idea for scale selection

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

$$L_{x^\alpha}(\cdot; t) = \partial_{x^\alpha} L(\cdot; t) = \partial_{x_1^{\alpha_1}} \dots \partial_{x_N^{\alpha_N}} L(\cdot; t) \quad (4)$$

in general *decrease with scale*, i.e. if a signal is subjected to scale-space smoothing, then the numerical values of spatial derivatives computed from the smoothed data can be expected to decrease. As an example of this, consider, say, a sinusoidal input signal of some given frequency ω_0 ; for simplicity in one dimension,

$$f(x) = \sin \omega_0 x. \quad (5)$$

It is straightforward to show that in this case the solution to the (one-dimensional) diffusion equation is given by

$$L(x; t) = e^{-\omega_0^2 t/2} \sin \omega_0 x. \quad (6)$$

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

$$L_{max}(t) = e^{-\omega_0^2 t/2}, \quad L_{x^m, max}(t) = \omega_0^m e^{-\omega_0^2 t/2}. \quad (7)$$

An alternative formulation of the scale-space concept is in terms of *normalized* (dimensionless) *coordinates*, $\xi = x/\sigma = x/\sqrt{t}$. One motivation for introducing such a coordinate system is *scale invariance*; see Florack et al [10]. In these coordinates the *normalized* (dimensionless) *derivative operator* is

$$\partial_\xi = \sqrt{t} \partial_x. \quad (8)$$

For the sinusoidal signal the amplitude of a normalized derivative as function of scale is given by

$$L_{\xi^m, max}(t) = t^{m/2} \omega_0^m e^{-\omega_0^2 t/2}, \quad (9)$$

i.e., it first increases and then decreases. It assumes a unique maximum at $t_{max, L_{\xi^m}} = m/\omega_0^2$. Introducing $\lambda_0 = 2\pi/\omega_0$ shows that the σ -value ($\sigma = \sqrt{t}$) for which $L_{\xi^m, max}(t)$ assumes its maximum is *proportional* to the wavelength, λ_0 , of the signal:

$$\sigma_{max, L_{\xi^m}} = \frac{\sqrt{m}}{2\pi} \lambda_0. \quad (10)$$

Note that the maximum value

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

is *independent* of the frequency of the signal. Note also the symmetry in the situation, 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 these normalized derivatives it holds that sinusoidal signals are treated in a similar (scale invariant) way independent of their frequency (see Figure 1).

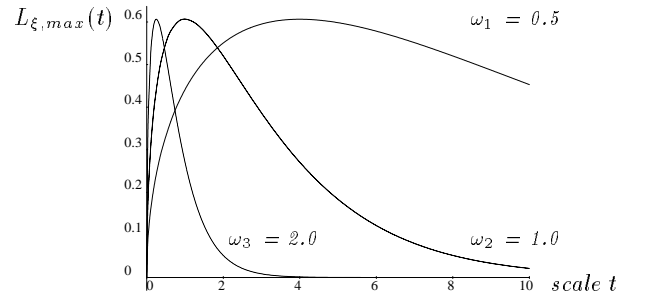


Figure 1: The amplitude of first order normalized derivatives as function of scale for sinusoidal input signals of different frequency ($\omega_1 = 0.5$, $\omega_2 = 1.0$ and $\omega_3 = 2.0$).

Note that although there is an intuitive similarity between this scale response and a local Fourier transform, there are two fundamental differences; (i) the normalization factor, and (ii) this method allows for local estimates of the frequency content without any explicit setting of a window size.

3 Proposed method for scale selection

As shown above, the scale at which a normalized derivative assumes its maximum is in the case of sinusoidal signals proportional to the wavelength of the

signal. Now, I propose to generalize this observation to more complex signals, leading to the following heuristic principle, which is to be applied in situations when no other information is available. *In the absence of other evidence, a scale level at which some (possibly non-linear) combination of normalized derivatives assumes a local maximum can be treated as reflecting the characteristic length of a corresponding structure in the data.*

This principle is similar although not equivalent to the parameter variation method in [18, 19], where interesting scale levels are determined from maxima over scales of a (normalized) blob measure. The underlying motivation behind using maxima over scales for scale selection is to select scale levels where the operator response is at its strongest, and other interfering structures have been suppressed.

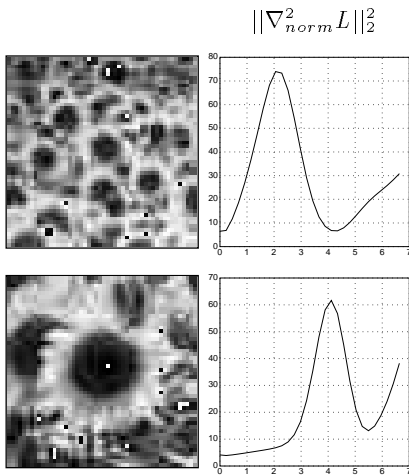


Figure 2: Scale-space signatures of the (absolute value of the) normalized Laplacean for two details of a sunflower image; (left) grey-level image, (right) signature of $|\nabla^2_{norm} L|$ computed at the central point.

Figure 2 illustrates the variation over scale of a simple measure formulated in terms of normalized spatial derivatives; it displays the scale variation of the absolute value of the normalized Laplacean computed at two different points. These graphs are called the *scale-space signatures* of $|\nabla^2_{norm} L| = |t \nabla^2 L|$.

Clearly, the maximum over scales in the top row of Figure 2 is obtained at a finer scale than in the bottom row. An examination of the ratio between the scale levels where the graphs attain their maxima shows that this value is roughly equal to the ratio of the sizes of the sunflowers in the centers of the two images respectively (in agreement with the heuristic principle).

4 Blob detection

The reason why this particular differential expression has been selected here is because it constitutes an entity commonly used in blob detection; see e.g. [25, 6, 5, 27]. Figure 3 shows the result of extending this approach to multi-scale blob detection. It shows *scale-space extrema* (points that are local extrema both in space and scale) of $|\nabla^2_{norm} L|$. Every extremum is graphically illustrated by an ellipse indicating the local directional statistics in the image¹.

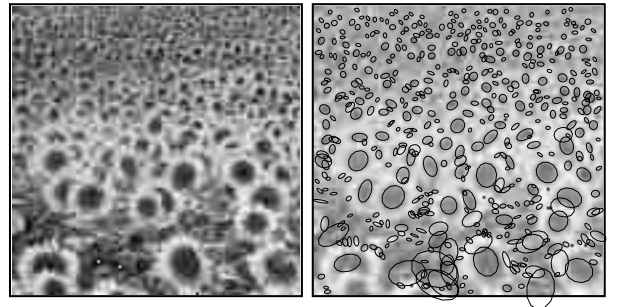


Figure 3: Multi-scale blob detection using normalized scale-space extrema of the squared Laplacean; (left) grey-level image, (right) ellipses illustrating the scale-space extrema superimposed onto a bright copy of the original grey-level image (the size of each ellipse is proportional to scale at which the maximum is assumed, while the shape of the ellipse is given by the local directional statistics in a neighbourhood of the maximum). (Adapted from [23]).

Note the ability of the method to adaptively zoom in to different scales, and also how well the computed ellipses describe the blobs in the image, considering how little information is used in the processing. In order to study this behaviour analytically, consider

$$f(x_1, x_2) = g(x_1, x_2; t_0) = \frac{1}{2\pi t} e^{-(x_1^2 + x_2^2)/(2t)} \quad (12)$$

as a simple model of a two-dimensional blob with characteristic length $\sqrt{t_0}$. From the semi-group property of the Gaussian kernel $g(\cdot; t) * g(\cdot; t) = g(\cdot; t_0 + t)$ it follows that the scale-space representation L of f is

$$L(x_1, x_2; t) = g(x_1, x_2; t_0 + t). \quad (13)$$

Clearly, the spatial maximum of $|\nabla^2 L|$ is assumed at $(x_1, x_2)^T = (0, 0)^T$. The corresponding normalized

¹ More precisely, each ellipse represents a second moment matrix (a matrix similar to A in (27)) computed using a Gaussian window function with scale value proportional to the scale at which the scale-space maximum of $\nabla^2_{norm} L$ is assumed; see [23] for a detailed description and theoretical analysis of this method.

entity is

$$t(\nabla^2 L)(0, 0; t) = \frac{t}{\pi(t_0 + t)^2}. \quad (14)$$

Differentiation with respect to t gives that the maximum is given by

$$\partial_t(\nabla_{norm}^2 L)(0, 0; t) = 0 \iff t = t_0, \quad (15)$$

which verifies that the maximum is assumed at a scale proportional to a characteristic length of the blob.

5 Junction detection

It will now be described how the above scale selection method can be used also for scale selection concerning other feature detectors formulated in terms of non-linear combinations of smoothed derivatives.

5.1 Selection of detection scale

As a first example, consider an entity commonly used for junction detection — the curvature of level curves in intensity data (see e.g. [12, 14]), which in terms of spatial derivatives can be expressed as

$$\kappa = \frac{L_{x_2}^2 L_{x_1 x_1} - 2L_{x_1} L_{x_2} L_{x_1 x_2} + L_{x_1}^2 L_{x_2 x_2}}{(L_{x_1}^2 + L_{x_2}^2)^{3/2}}. \quad (16)$$

In order to give a stronger response near edges, this entity is usually multiplied by the gradient magnitude raised to some power, k . A natural choice is $k = 3$. This leads to a polynomial expression (see also [4, 7])

$$\tilde{\kappa} = L_{x_2}^2 L_{x_1 x_1} - 2L_{x_1} L_{x_2} L_{x_1 x_2} + L_{x_1}^2 L_{x_2 x_2}. \quad (17)$$

Figure 4 shows the result of accumulating the signature of the corresponding normalized entity $\tilde{\kappa}_{norm} = t^2 \tilde{\kappa}$ at two different details of a toy block image, and then computing the curvature descriptor at the scale where the maximum is assumed. Note that the scale maximum in the top row is obtained at a finer scale than the maximum in the bottom row, reflecting the fact that the corner is much sharper (and neighboring junctions are much closer) in the first case than in the second case. In order to study this behaviour analytically, consider

$$f(x_1, x_2) = E_{t_0}(x_1) E_{t_0}(x_2) \quad (18)$$

as a simple model of a *diffuse L-junction*, where E_{t_0} describes a diffuse step edge

$$E_{t_0}(x_i) = \int_{x'=-\infty}^{x_i} g(x'; t_0) dx'. \quad (19)$$

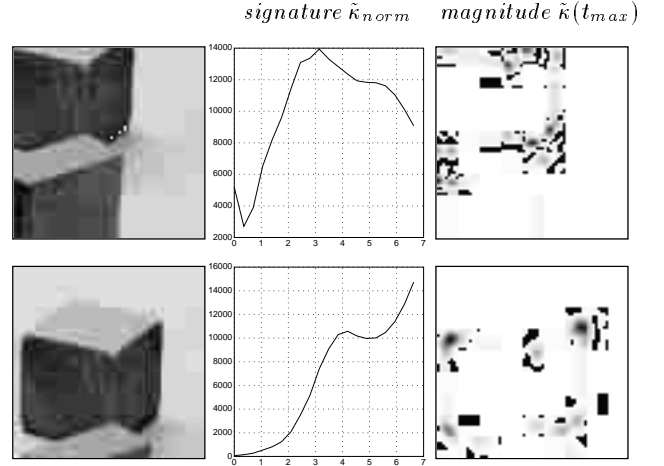


Figure 4: Scale-space signature of (the absolute value of) the rescaled level curve curvature $\tilde{\kappa}_{norm} = t^2 \tilde{\kappa}$ computed at two different details of a toy block image; (left) grey-level image, (middle) signature of $\tilde{\kappa}_{norm}$ accumulated at the central point, (right) $\tilde{\kappa}_{norm}$ computed at the scale at which the maximum is assumed.

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

$$L(x_1, x_2; t) = E_{t_0+t}(x_1) E_{t_0+t}(x_2). \quad (20)$$

Differentiation with respect to x_1 and x_2 gives

$$\begin{aligned} L_{x_1} &= g(x_1; t_0 + t) E_{t_0+t}(x_2), \\ L_{x_1 x_1} &= (-x_1/t) g(x_1; t_0 + t) E_{t_0+t}(x_2), \\ L_{x_1 x_2} &= (x_1 x_2 / t^2) g(x_1; t_0 + t) g(x_2; t_0 + t), \end{aligned} \quad (21)$$

etc. Insertion of these expressions into (17) gives that $\tilde{\kappa}_{norm}$ at the central point is

$$\tilde{\kappa}_{norm}(0, 0; t) = \frac{t^2}{16\pi^2(t_0 + t)^4}. \quad (22)$$

This expression is of the same form as (14). Hence, the maximum over scales is assumed at a scale level proportional to the diffuseness of the junction,

$$\partial_t(\tilde{\kappa}_{norm}(0, 0; t)) = 0 \iff t = t_0. \quad (23)$$

It turns out that selecting scales where the normalized rescaled level curve curvature assumes a maximum is an applicable method when *detecting* features, here junction candidates. This method may, however, lead to poor *localization*, since in general the scale maximum will be assumed at a rather coarse scales, where the drift due to scale-space smoothing may be substantial, and also adjacent features may begin to interfere with each other. Therefore, some postprocessing is necessary in order to improve the localization.

5.2 Selection of localization scale

A simple method for improving the localization is as follows: Following Förstner and Gülch [11] consider at every point $x' \in \mathbb{R}^2$ in a neighbourhood of a junction candidate the line $l_{x'}$ perpendicular to the gradient vector $(\nabla L)(x') = (L_{x_1}, L_{x_2})^T(x')$ at that point. The equation of this line is

$$D_{x'}(x) = ((\nabla L)(x'))^T (x - x') = 0. \quad (24)$$

A simple estimate of the location of the junction candidate is obtained from the point $x \in \mathbb{R}^2$ that minimizes

$$\min_x \int_{x' \in \mathbb{R}^2} (D_{x'}(x))^2 w_{x_0}(x') dx' \quad (25)$$

for some window function $w_{x_0} : \mathbb{R}^2 \rightarrow \mathbb{R}$ (centered at the point x_0 at which the signature is computed). Minimizing this expression corresponds to minimizing the weighted integral of the squares of the distances from x to all $l_{x'}$ in the neighbourhood. (Note that $D_{x'}(x)$ describes the distance from x to the line $l_{x'}$ multiplied by the gradient magnitude. Hence, every distance is given a weight proportional to the square gradient magnitude multiplied by the window function). After expansion (25) can be expressed as

$$\min_x x^T A x - 2 x^T b + c \iff A x = b \quad (26)$$

where $x = (x_1, x_2)^T$, and

$$A = \int_{x' \in \mathbb{R}^2} (\nabla L)(x') (\nabla L)^T(x') w(x') dx', \quad (27)$$

$$b = \int_{x' \in \mathbb{R}^2} (\nabla L)(x') (\nabla L)^T(x') x' w(x') dx', \quad (28)$$

$$c = \int_{x' \in \mathbb{R}^2} x'^T (\nabla L)(x') (\nabla L)^T(x') x' w(x') dx'. \quad (29)$$

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

$$d_{min} = \min_x x^T A x - 2 x^T b + c = c - b^T A^{-1} b, \quad (30)$$

with the associated localization estimate $x = A^{-1} b$. A common problem in least squares estimation concerns how to choose the region over which to perform the fitting. Here, the following method is proposed:

- **Selection of window function and spatial points:** When computing A , b and c above, let the window function w_{x_0} be a Gaussian function centered at the point x_0 where $\tilde{\kappa}_{norm}$ assumes a spatial maximum, and let the scale value of this window function be proportional to the (detection) scale at which the scale maximum in $\tilde{\kappa}_{norm}$ was assumed,

- **Criterion for selecting local scale:** Compute the gradient estimates $(\nabla L)(x')$ from smoothed grey-level data, and set the (localization) scale to the scale that minimizes the residual as function of scale. In practice, in order to obtain an expression of dimension $[length]^2$, determine the minimum over scales of d_{min} given by (30) divided by $\text{trace}(A)$ (this denominator is the weighted average of the gradient magnitude in the neighbourhood), i.e.,

$$\tilde{d}_{min} = \frac{d_{min}}{\text{trace } A} = \frac{c - b^T A^{-1} b}{\text{trace } A}. \quad (31)$$

The basic ideas behind this approach are:

- Firstly, that the detection scale should give a representative region around the candidate junction. Experimentally, this has been demonstrated to be the case in a large number of situations.
- Secondly, the idea of setting the localization scale to the scale at which the residual is minimized is as follows: At too fine a scale where noise is present, the gradient directions will be randomly distributed, which in turn means that the residual error will increase. At too coarse a scale the scale-space smoothing will lead to increasing shape distortions and hence increase the error. Selecting the minimum gives a natural trade-off between these two effects.

Note, in particular, that for an ideal (sharp) junction, the localization scale given by this method will always be zero in the noise free case.

Figure 5 illustrates the result of applying this method to the examples in Figure 4. It shows (smaller) windows around the interest points together with the scale-space signature of \tilde{d}_{min} , and the rescaled level curve curvature $\tilde{\kappa}$ computed at the scale at which the minimum was assumed.

As an illustration of the stability of the method with respect to noise and violations of the corner-type junction model, Figure 8 shows the result of applying it to a rounded corner with and without added Gaussian noise (standard deviation 64 compared to the grey-level range $[0..255]$). Note the similarity of the scale-space signatures of $\tilde{\kappa}_{norm}$ at coarse scales, and that the minimum in the signature of \tilde{d}_{min} is obtained at a finer scale in the top row than in the bottom row, reflecting the much lower noise level in the first case, and hence the reduced need for smoothing.

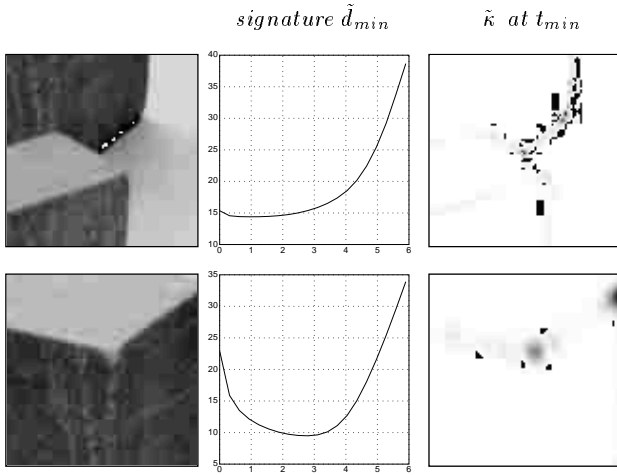


Figure 5: Scale-space signature of the least squares residual, \tilde{d}_{min} (computed using a Gaussian window function with scale level proportional to the scale at which $\tilde{\kappa}_{norm}$ assumed its maximum over scales in Figure 4); (left) grey-level image, (middle) scale-space signature of \tilde{d}_{min} accumulated at the central point, (right) $\tilde{\kappa}_{norm}$ computed at the scale where \tilde{d}_{min} assumes its minimum. Note that the minimum in the top row is assumed at a finer scale than in the bottom row, reflecting the fact that the junction in the top row is sharper than the junction in the bottom row.

6 Edge detection

A similar approach for scale selection can also be applied in edge detection. The method that will be proposed is similar to the edge focusing method developed by Bergholm [2], in the sense that edges are detected at a coarse scale, and the followed to finer scales. A main difference compared to that work is that here a method for selecting those scale levels is addressed.

A natural way to define edges from a two-dimensional continuous image is as the points for which the gradient magnitude assumes a maximum in the gradient direction, see e.g. Canny [8], or Korn [16]. In differential geometric terms the condition for a point to be an edge point may be stated as

$$\begin{cases} L_{\bar{v}\bar{v}} = 0, \\ L_{\bar{v}\bar{v}\bar{v}} < 0, \end{cases} \quad (32)$$

where $\partial_{\bar{v}} = \cos \phi \partial_{x_1} + \sin \phi \partial_{x_2}$ denotes the directional derivative operator in the (normalized) gradient direction $(\cos \phi, \sin \phi)^T = (L_{x_1}, L_{x_2})^T / (L_{x_1}^2 + L_{x_2}^2)^{1/2}$.

6.1 Selection of detection scale

A natural measure of the strength of the edge response is the normalized gradient magnitude $L_{\bar{v},norm} =$

$\sqrt{t} L_{\bar{v}}$. A qualitative difference between this measure and the earlier treated blob and junction strength measures, $\nabla_{norm}^2 L$ and $\tilde{\kappa}_{norm}$ respectively, however, is that this entity cannot be expected to decrease in the same way for sufficiently large scale values. As an example of this, consider the response of a *diffuse step edge*

$$f(x) = E_{t_0}(x) = \int_{x'=-\infty}^x g(x'; t_0) dx', \quad (33)$$

for which the variation over scales of the normalized gradient magnitude increases monotonically

$$(\partial_{\xi} L_{t_0})(0; t) = \sqrt{t} g(0; t_0 + t) \sim \frac{\sqrt{t}}{\sqrt{t_0 + t}}. \quad (34)$$

A similar result holds for edges in the scale-space representation $L(x_1, x_2; t) = g(x_1, x_2; t_0 + t)$ of a *Gaussian blob* $f(x_1, x_2) = g(x_1, x_2; t_0)$. It is straightforward to show that the edge position is given by

$$\tilde{L}_{\bar{v}\bar{v}} = \frac{x^2 + y^2 - t_0 - t}{(t_0 + t)^2} g(x_1, x_2; t_0 + t) = 0, \quad (35)$$

and that the normalized gradient magnitude is constant over scales at the edge points

$$\sqrt{t} L_{\bar{v}}(x_1, x_2; t) \Big|_{L_{\bar{v}\bar{v}}=0} = e^{-1/2}. \quad (36)$$

Of course, there are also situations when $L_{\bar{v},norm}$ decreases. As an example of this, consider one of the edges at a *double asymmetric step edge*, which disappears by annihilation of a minimum-maximum pair in the gradient magnitude. A simple local model of this situation is the polynomial

$$L_{\bar{v}} = x_1^3 + 3x_1(t - t_b), \quad (37)$$

which represents the *fold unfolding* (see e.g. Poston and Stewart [26]), and also satisfies the diffusion equation. Setting $L_{\bar{v}\bar{v}} = 3(x_1^2 + t - t_b) = 0$, gives that the variation over scales at the edge point $x_{1,edge}(t) = (t_b - t)^{1/2}$ ($t \leq t_b$) follows

$$L_{\bar{v},norm} = \sqrt{t} L_{\bar{v}} = 4\sqrt{t}(t_b - t)^{3/2}. \quad (38)$$

Clearly, $L_{\bar{v}}$ decreases with scale when t approaches the bifurcation scale t_b .

Figure 6 shows the result of accumulating the signature of the normalized gradient magnitude at two different details of the previously treated toy block image. Note that the maximum at the step edge (top row) is assumed at much a coarser scale than is the maximum at the double edge (bottom row). Also, the

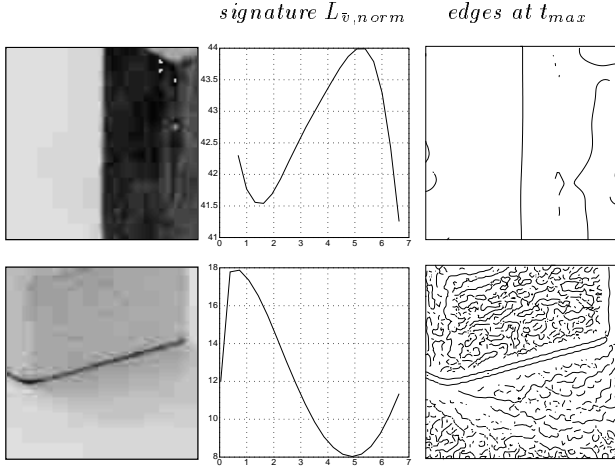


Figure 6: Scale-space signature of the normalized gradient magnitude $L_{\bar{v}, norm} = \sqrt{t} L_{\bar{v}}$ at a straight edge (top row) and a double edge (bottom row); (left) grey-level image, (middle) signature of $L_{\bar{v}, norm}$ accumulated at the central point, (right) (unthresholded) edges detected at the scale at which the maximum was assumed.

relative variation over scales of the normalized gradient magnitude is much larger in the second case.

To summarize, scale selection using maxima in $L_{\bar{v}}$ can be expected to give rise to rather coarse scale levels, delimited from above by the bifurcation scales at which the edges disappear. Notably, the scales where these maxima are assumed do not depend directly on the diffuseness t_0 of the original edge. At double edges (and staircase edges) the behaviour is different, and relatively finer scale levels will be selected (proportional to the distance between adjacent edges).

6.2 Selection of localization scale

Clearly, this type of scale selection may lead to poor localization, in particular at curved edges, where the total drift due to scale-space smoothing may be substantial. Therefore it is necessary to apply a second stage selection of localization scale. It turns out that this selection can be performed analogously to the previous method concerning junctions, although the results become better if the directional statistics is based on the normal directions to the edges than the gradient vectors. Hence, for every point x' in a neighbourhood of an edge, consider the line $l_{x'}$ given by

$$n^T(x - x') = 0, \quad (39)$$

where $n = (n_1, n_2)^T$ denotes the normal direction to a level curve of $L_{\bar{v}}$, which at any edge point is a normal direction to the edge. In terms of spatial derivatives

the components of n can be written

$$\begin{aligned} n_1 &= L_x(L_x L_{xxx} + 2(L_{xx} L_{xx} + L_{xy}^2)) + \\ &\quad L_y(L_y L_{yyy} + 2(L_x L_{xxy} + L_{xy}(L_{xx} + L_{yy}))), \\ n_2 &= L_y(L_y L_{yyy} + 2(L_{xy}^2 + L_{yy} L_{yyy})) + \\ &\quad L_x(L_x L_{xxx} + 2(L_y L_{xyy} + L_{xy}(L_{xx} + L_{yy}))), \end{aligned}$$

where above the notation $x = (x_1, x_2)^T$ has been temporarily replaced by $(x, y)^T$. Following previous section, consider the point x that minimizes

$$\min_x \int_{x' \in \mathbb{R}^2} (D_{x'}(x))^2 w_{x_0}(x') dx', \quad (40)$$

for some window function w_{x_0} . In order to preserve the weighting with respect to gradient magnitude, it is natural to define $D_{x'}$ by

$$D_{x'}(x) = \tilde{n}^T(x - x') = 0, \quad (41)$$

with the gradient vector replaced by

$$\tilde{n} = |(\nabla L)(x')| \frac{n}{|n|}. \quad (42)$$

The minimization problem (40) is of the same type as (26) with A , b and c defined as in (27)–(29), and with $(\nabla L)(x')$ replaced by \tilde{n} . From motivations analogous to previous section, it is natural to select the scale that minimizes the normalized residual

$$\tilde{d}_{min} = \frac{d_{min}}{\text{trace } A} = \frac{c - b^T A^{-1} b}{\text{trace } A}. \quad (43)$$

Since, however, A may degenerate to a rank one matrix at ideal straight edges, in actual computations A^{-1} above is replaced by the pseudo inverse of A .

Applying this method to the top row example in Figure 6 gives the result shown in Figure 7. Note that a rather fine scale is selected indicating the sharpness of the edge. As a test of the stability of the method, the second and third rows show corresponding results for a smoothed version of the image in the first row, with and without added (white Gaussian) noise (standard deviation 25% of the grey-level range). Observe that the minimum in the second row is assumed at a much coarser scale indicating the higher noise level, and hence the increased need for smoothing. The fourth row shows an interesting example, where a coarse scale is selected at a shadow edge. Figure 9 shows a stability tests in the neighbourhood of a curved junction.

Finally it should be remarked that in order to introduce as few commitments into the processing as possible, no thresholding on gradient magnitude has been performed in these experiments (except in Figure 9). The results of scale adaption should therefore be interpreted as locally valid in a neighbourhood of the edge through the central point.

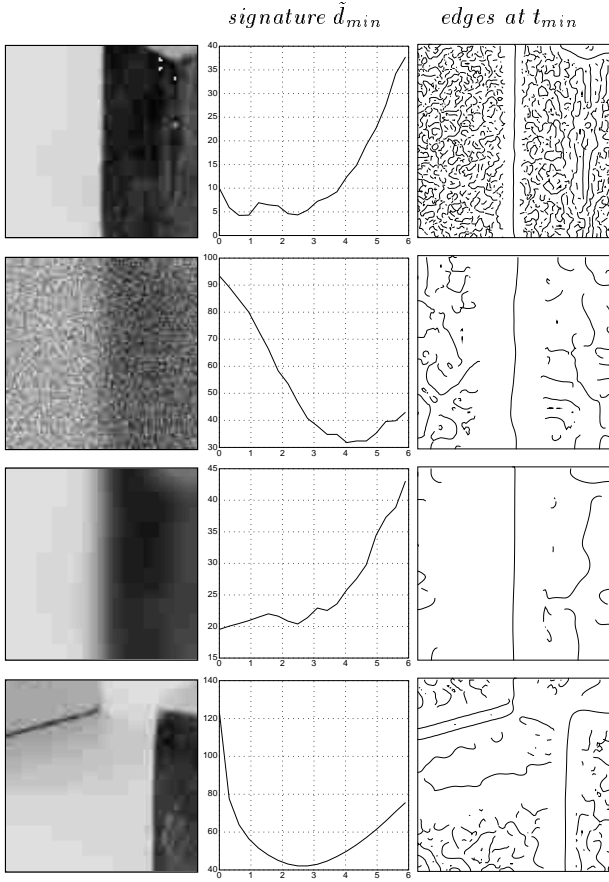


Figure 7: Scale-space signature of the least squares residual, \tilde{d}_{min} (computed using a Gaussian window function with scale level proportional to the scale at which $L_{\tilde{v},norm}$ assumed its maximum over scales in Figure 6); (left) grey-level image, (middle) scale-space signature of \tilde{d}_{min} accumulated at the central point, (right) (unthresholded) edges computed at the scale where \tilde{d}_{min} assumes its minimum.

7 Discrete implementation

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

The first problem can be solved by using the scale-space concept for discrete signals [17], which is given by

$$L(\cdot, \cdot; t) = T(\cdot, \cdot; t) * f(\cdot, \cdot), \quad (44)$$

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 in terms of the modified Bessel functions I_n . The scale-space properties of L transfer to any discrete derivative approximations

$L_{x_1^i x_2^j}$ that are defined as the result of applying difference operators $\delta_{x_1^i x_2^j}$ to L [22].

Concerning the normalization problem, it is straightforward to verify that the normalized derivative operator given by (8) corresponds to rescaling the Gaussian derivative kernels such that the integral of the positive part of the kernel remains constant over scales. Such kernels of first order have been used, for example, in edge detection and edge classification by Korn [16], and Zhang and Bergholm [30], as well as in pyramids by Crowley and his co-workers [9].

Given this analogy, it is natural to normalize the discrete derivative approximation kernels $\delta_{x_1^i x_2^j} T$ such that the l_1 norm remains constant over scales. Of course, it is not necessary to perform any explicit convolution with a normalized derivative approximation kernel. In practice, concerning e.g. first order derivatives, the discrete approximations to L_{x_1} and L_{x_2} can be multiplied by the discrete normalization constant,

$$\alpha_1(t) = \frac{\sqrt{2}}{\sqrt{\pi}(T_1(0; t) + T_1(1; t))}, \quad (45)$$

which can be shown to approach \sqrt{t} when t increases.

8 Summary and discussion

The main purpose of this paper has been to outline how the evolution properties over scales of normalized Gaussian derivatives can be used for finding interesting scales for further analysis.

A heuristic principle has been proposed stating that, in the absence of other evidence, local extrema over scales of normalized differential geometric descriptors are candidates for representing characteristic lengths of corresponding structures in the image. Once such a detection scale (outer scale) has been selected, a complementary second stage selection of localization scale (inner scale) is applied, based on the mutual consistency of directional information computed in a local neighbourhood of the selected point.

Experimentally, the methodology has been demonstrated to give useful results both analytically and by experiments on real and synthetic data.

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

Choice of differential invariants: An open problem that remains to be addressed concerns what differential properties are useful. Of course, the differential entity should be selected in order to reflect the type of structures to be studied. More generally, it is straightforward to show that the scale invariant properties of differential expressions transfer to all homogeneous differential entities that can be expressed on the form

$$\mathcal{D}L = \sum_{i=1}^I c_i \prod_{j=1}^J L_{\bar{u}^{\alpha_{ij}}}, \quad (46)$$

where $L_{\bar{u}^\alpha} = L_{\bar{u}^m \bar{v}^n}$ denotes a mixed directional derivative of order $|\alpha| = m + n$, \bar{v} is the gradient direction, \bar{u} the perpendicular direction, $|\alpha_{ij}| > 0$ for all $i = [1..I]$ and $j = [1..J]$, and

$$\sum_{j=1}^J |\alpha_{ij}| = N \quad (47)$$

for all $i \in [1..I]$. Assume that a local extremum over scales in the the corresponding normalized entity $\mathcal{D}_{norm}L = t^{N/2} \mathcal{D}L$ is assumed at scale t_0 . Then, under rescalings of the original signal f , i.e. $f'(x) = f(sx)$, the extremum over scales in $\mathcal{D}_{norm}L'$ is assumed at $t'_0 = t_0/s^2$.

Linking structures across scales: Because of simplicity of presentation and implementation, the scale-space signatures have here been accumulated at points which are assumed to be given (and fixed). In practice, concerning, e.g. blob detection and junction detection, it is natural to require all feature points to be spatial maxima with respect to the differential entity considered. Hence, it is necessary to impose some additional constraint, e.g. by either (i) detecting *scale-space maxima*, that are points which are maxima with respect to both scale and space of the differential entity considered (one such approach concerning blob detection is developed in [23]), or (ii) *linking* features across scales and registering the signature along such a path through scale-space. An approach of the latter type, concerning the extrema of the grey-level landscape (blobs), is developed in [18, 19], where the lifetime of the path is used as an important component for formulating a significance measure. More generally, the linking approach can be applied also to other entities, like extrema of the Laplacean or the rescaled level curve curvature. In [20, 21] it is described how the implicit function theorem allows for a formal definition of scale linking concerning features formulated as differential singularities, that are entities which are zero-crossings of expressions of the form (46).

References

- [1] J. Babaud, A.P. Witkin, M. Baudin, R. O. Duda (1986) "Uniqueness of the Gaussian kernel for scale-space filtering", *IEEE-PAMI*, 8:1, 26–33.
- [2] F. Bergholm F (1987) "Edge Focusing", *IEEE-PAMI*, 9:6, 726–741.
- [3] W.F. Bischof, T. Caelli (1988) "Parsing scale-space and spatial stability analysis", *CVGIP*, 42, 192–205.
- [4] J. Blom (1992) *Topological and geometrical aspects of image structure*. PhD thesis, Dept. Med. Phys. Physics, Univ. Utrecht, NL-3508 Utrecht, Netherlands.
- [5] D. Blostein, N. Ahuja (1989) "A multiscale region detector", *CVGIP*, 45, 22–41, 1989.
- [6] D. Blostein, N. Ahuja (1989) "Shape from texture: integrating texture element extraction and surface estimation", *IEEE-PAMI*, 11, 1233–1251.
- [7] K. Brunnström, T. Lindeberg, J.-O. Eklundh (1992) "Active detection and classification of junctions by foveation with a head-eye system guided by the scale-space primal sketch", *Proc. 2nd ECCV*, (Italy), 701–709.
- [8] J. Canny (1986) "A computational approach to edge detection" *IEEE-PAMI*, 8:6, 679–698.
- [9] J.L. Crowley A.C. Parker A.C. (1984) "A Representation of Shape Based on Peaks and Ridges in the Difference of Low-Pass Transform", *IEEE-PAMI*, 6:2, 156–169.
- [10] L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, M.A. Viergever (1992), "Scale and the differential structure of images", *Image Vision Comp.*, 10, 376–388.
- [11] M.A. Förstner, E. Gülch (1987) "A fast operator for detection and precise location of distinct points, corners and centers of circular features", *ISPRS Intercomm. Workshop*.
- [12] L. Kitchen, A. Rosenfeld (1982) "Gray-level corner detection", *PRL*, 1:2, 95–102.
- [13] J.J. Koenderink (1984) "The structure of images", *Biol. Cyb.*, 50, 363–370.
- [14] J.J. Koenderink, W. Richards (1988) "Two-dimensional curvature operators", *J. Opt. Soc. Am.*, 5:7, 1136–1141.
- [15] J.J. Koenderink, A.J. van Doorn (1990) "Receptive field families", *Biol. Cyb.*, 63, 291–298.
- [16] A.F. Korn (1988) "Toward a symbolic representation of intensity changes in images", *IEEE-PAMI*, 10:5, 610–625.
- [17] T. Lindeberg (1990) "Scale-space for discrete signals", *IEEE-PAMI*, 12, 234–254.
- [18] T. Lindeberg (1991) *Discrete scale space theory and the scale space primal sketch*. PhD thesis, Royal Inst. of Technology, Stockholm, Sweden. An extended version to appear in Kluwer Int. Ser. in Engineering and Computer Science.
- [19] T. Lindeberg, J.-O. Eklundh (1992) "Scale-space primal sketch: Construction and experiments", *Image Vision Comp.*, 10, 3–18.
- [20] T. Lindeberg (1992) "Scale-space behaviour of local extrema and blobs", *J. Math. Imaging Vision*, 1, 65–99.
- [21] T. Lindeberg (1992) "Scale-space behaviour and invariance properties of differential singularities" In: Ying-Lie et al. (eds) *Proc. NATO ARW on Shape in Picture*, (Driebergen, Netherlands), Springer-Verlag, in press.
- [22] T. Lindeberg (1993) "Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction", *J. Math. Imaging Vision*, in press.
- [23] T. Lindeberg, J. Gårding (1993) "Shape from texture from a multi-scale perspective", *Proc. 4th ICCV*, (Berlin, Germany), Extended version available as technical report.
- [24] D. Marr (1976) "Early processing of visual information", *Phil. Trans. Royal Soc (B)*, 27S, 483–524.

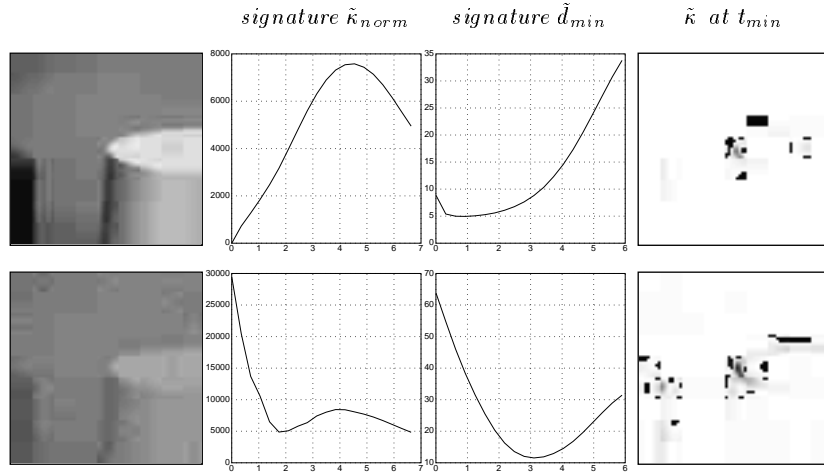


Figure 8: Junction detection: Scale-space signatures of $\tilde{\kappa}_{norm}$ and \tilde{d}_{min} computed with and without added (white Gaussian) noise. (left) grey-level image. (middle left) signature of $\tilde{\kappa}_{norm}$ at the central point, (middle right) signature of \tilde{d}_{min} at the central point, (right) $\tilde{\kappa}_{norm}$ computed at the scale where \tilde{d}_{min} assumes its minimum. Note that the minimum in the top row in \tilde{d}_{min} is assumed at a finer scale than in the bottom row, reflecting the much lower noise level in the first case. The effect of the noise on the signature of $\tilde{\kappa}_{norm}$ is only minor, except for a new peak that appears at finer scales. (Noise standard deviation 64, grey-level range [0..255]).

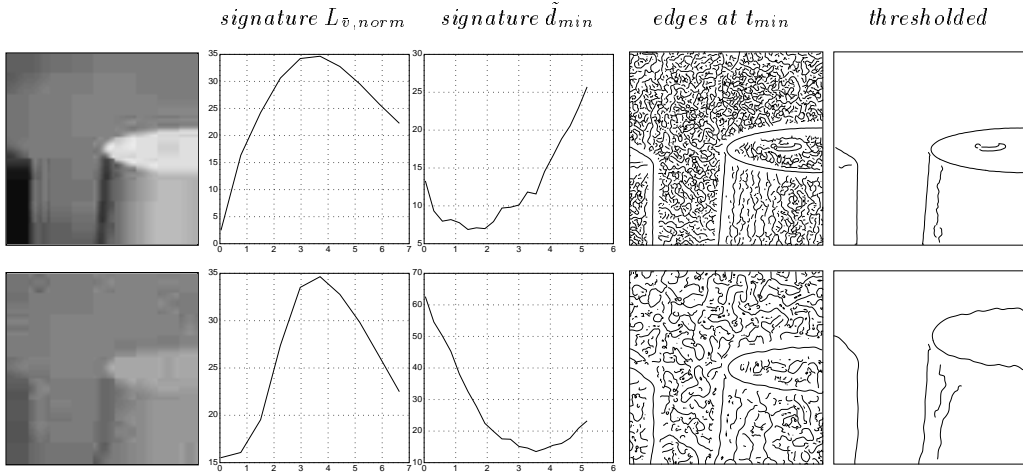


Figure 9: Edge detection: Scale-space signatures of $L_{\tilde{v},norm}$ and \tilde{d}_{min} computed with and without added (white Gaussian) noise. (left) grey-level image; (middle left) signature of $L_{\tilde{v},norm}$ at the central point, (middle) signature of \tilde{d}_{min} at the central point, (middle right) unthresholded edges detected at the scale where \tilde{d}_{min} assumes its minimum, (right) edges after hysteresis thresholding on the gradient magnitude (low threshold 2.0, high threshold 4.0). Note that the minimum in the top row in \tilde{d}_{min} is assumed at a finer scale than in the bottom row, reflecting the much lower noise level in the first case. The effect of the noise on the signature of $L_{\tilde{v},norm}$ is minor, the only difference is that the value is larger at finer scales. (Noise standard deviation 64, grey-level range [0..255]).

- [25] D. Marr (1982) *Vision*. W.H. Freeman, New York.
- [26] T. Poston, I. Stewart (1978) *Catastrophe theory and its applications*, Pitman Publishing, London.
- [27] H. Voorhees, T. Poggio (1987) "Detecting textons and texture boundaries in natural images", *1st ICCV*, (London).
- [28] A.P. Witkin (1983) "Scale-space filtering", *Proc. 8th IJCAI*, (Karlsruhe, West Germany), 1019–1022.

- [29] A.L. Yuille, T. A. Poggio (1986) "Scaling theorems for zero-crossings", *IEEE-PAMI*, 8, 15–25.
- [30] W. Zhang, F. Bergholm (1991) "An extension of Marr's "signature" based edge classification", *Proc. 7th SCIA*, (Aalborg, Denmark), 435–443.
- [31] W. Zhang, F. Bergholm (1993) "On ... recent developments of edge focusing", *Proc. 8th SCIA*, (Tromsø, Norway).