Scale-space: A framework for handling image structures at multiple scales

Tony Lindeberg *KTH, S-100 44 Stockholm, Sweden*

Abstract

This article gives a tutorial overview of essential components of scale-space theory — a framework for multi-scale signal representation, which has been developed by the computer vision community to analyse and interpret real-world images by automatic methods.

1 The need for multi-scale representation of image data

An inherent property of real-world objects is that they only exist as meaningful entities over certain ranges of scale. A simple example is the concept of a branch of a tree, which makes sense only at a scale from, say, a few centimeters to at most a few meters, It is meaningless to discuss the tree concept at the nanometer or kilometer level. At those scales, it is more relevant to talk about the molecules that form the leaves of the tree, and the forest in which the tree grows, respectively. This 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. This general need is well understood in cartography, where maps are produced at different degrees of abstraction. Similarly in physics, phenomena are modelled at several scales, ranging from particle physics and quantum mechanics at fine scales, through solid mechanics and thermodynamics dealing with everyday phenomena, to astronomy and relativity theory at scales much larger than those we are usually dealing with. Notably, the form of description may be strongly dependent upon the scales at which the world is modelled, and this is in clear contrast to certain idealized mathematical concepts, such as 'point' and 'line', which are independent of the scale of observation.

Specifically, the need for multi-scale representation arises when designing methods for automatically analysing and deriving information from real-world measurements. To be able to extract any information from image data, one obviously has to interact with it using certain operators. The type of information that can be obtained is largely determined by the relationship between the size of the actual structures in the data and the size (resolution) of the operators (probes). Some of the very fundamental problems in image processing concern *what* operators to use, *where* to apply them and *how large* they should be. If these problems are not appropriately addressed, then the task of interpreting the operator response can be very hard.

In certain controlled situations, appropriate scales for analysis may be known *a priori*. For example, a desirable quality of a physicist is his intuitive ability to select proper 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 indirect properties of a three-dimensional world. To be able to cope with these problems, an essential tool is a formal theory for how to describe image structures at different scales.

2 Scale-space representation: Definition and basic ideas

Scale-space theory is a framework for early visual operations, which has been developed by the computer vision community (in particular by Witkin [21], Koenderink [11], Yuille and Poggio [23], Lindeberg [15] and Florack [4]) to handle the above-mentioned multi-scale nature of image data. A main argument behind its construction is that if no prior information is available about

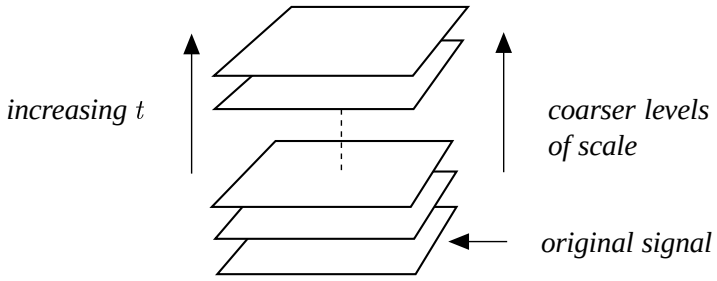

Figure 1: A multi-scale representation of a signal is an ordered set of derived signals intended to represent the original signal at different levels of scale.

what are the appropriate scales for a given data set, then the only reasonable approach for an uncommitted vision system is to represent the input data at multiple scales. This means that the original signal should be embedded into a one-parameter family of derived signals, in which fine-scale structures are successively suppressed (see figure 1). How should such an idea be carried out in practice? A crucial requirement is that structures at coarse scales in the multi-scale representation should constitute simplifications of corresponding structures at finer scales—they should not be accidental phenomena created by the method for suppressing fine-scale structures. This idea has been formalized in a variety of ways by different authors. A noteworthy coincidence is that similar conclusions can be obtained from several different starting points. A main result is that if rather general conditions are imposed on the types of computations that are to be performed, then convolution by the Gaussian kernel and its derivatives is singled out as a canonical class of smoothing transformations. The requirements (scale-space axioms) that specify the uniqueness are essentially linearity and spatial shift invariance, combined with different ways of formalizing the notion that new structures should not be created in the transformation from fine to coarse scales. In summary, for any N-dimensional signal $f\colon \mathbb{R}^N \to \mathbb{R}$, its *scale-space representation* $L\colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ is defined by

$$L(x;\ t) = \int_{\xi \in \mathbb{R}^N} f(x - \xi)\, g(\xi)\, d\xi \tag{1}$$

where $g\colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ denotes the Gaussian kernel

$$g(x;\ t) = \frac{1}{(2\pi\sigma^2)^{D/2}}\, e^{-(x_1^2 + \cdots + x_D^2)/2t} \tag{2}$$

and the variance t of this kernel is referred to as the *scale parameter*. Equivalently, the scale-space family can be obtained as the solution to the (linear) diffusion equation

$$\partial_t L = \frac{1}{2} \nabla^2 L \tag{3}$$

with initial condition $L(\cdot;\ t) = f$. Then, based on this representation, *scale-space derivatives* at any scale t are defined by

$$L_{x^\alpha}(\cdot;\ t) = \partial_{x_1^{\alpha_1} \cdots x_D^{\alpha_D}} L(\cdot;\ t) = (\partial_{x_1^{\alpha_1} \cdots x_D^{\alpha_D}} g(\cdot;\ t)) * f. \tag{4}$$

Figure 2(a) shows the result of applying Gaussian smoothing to a one-dimensional signal in this way. Notice how this successive smoothing captures the intuitive notion of fine-scale information being suppressed, and the signals becoming successively smoother. Figure 3 gives a corresponding example for a two-dimensional image. Here, to emphasize the local variations in

Figure 2: (a) The main idea of a scale-space representation is to generate a one-parameter family of derived signals in which the fine-scale information is successively suppressed. This figure shows a signal which has been successively smoothed by convolution with Gaussian kernels of increasing width. (b) Since new zero-crossings cannot be created by the diffusion equation in the one-dimensional case, the trajectories of zero-crossings in scale-space (here, zero-crossings of the second derivative) form paths across scales that are never closed from below.

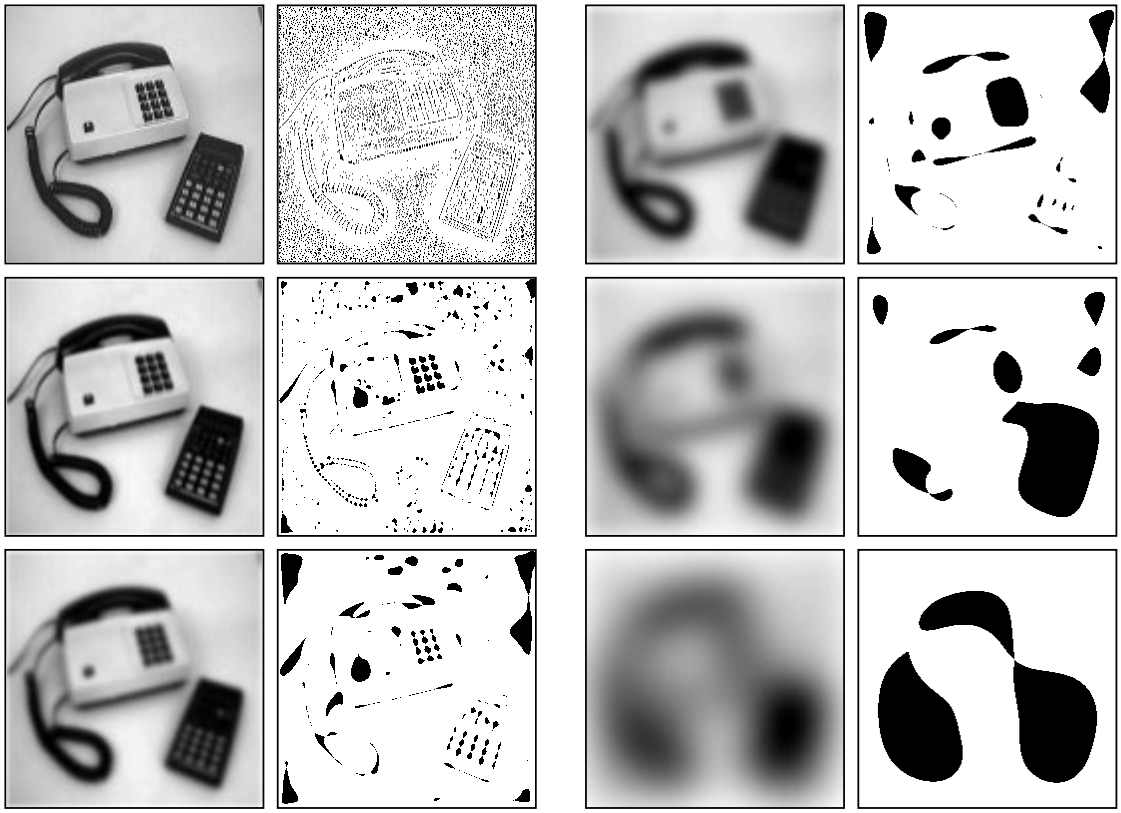

Figure 3: Different levels in the scale-space representation of a two-dimensional image at scale levels $t =$ 0, 2, 8, 32, 128 and 512 together with grey-level blobs indicating local minima at each scale.

the grey-level landscape, local minima in the grey-level images at each scale have been indicated by dark blobs (with spatial extent determined from a certain watershed analogy, which essentially describes how large a region associated with a local minimum can be filled with water, without water flooding over to regions associated with other local minima). As can be seen, mainly small blobs due to noise and texture are detected at fine scales. After a small amount of smoothing, the buttons on the keyboard manifest themselves as distinct minima, whereas at even coarser scales they merge to one unit (the keyboard). Also other dominant dark image structures (such as the calculator, the cord and the receiver) appear as single blobs at coarser scales. This example gives one illustration of the types of hierarchical shape decompositions that can be obtained by varying the scale parameter in the scale-space representation. The relations between image structures at different scales induced in this way is referred to as *deep structure* [11, 15].

3 Axiomatic scale-space formulations

For a reader not familiar with the scale-space literature, the task of designing a multi-scale signal representation may at first glance be regarded as somewhat arbitrary. Why make use of Gaussian smoothing, and why not just carry out *any* type of "smoothing operation"? To illuminate the special properties that have lead computer vision researchers to consider linear scale-space representation as a natural model for an uncommitted visual front-end, we shall in this section give a very brief review of a number of the major scale-space formulations. By necessity, this presentation will be somewhat technical, and the hasty reader may without loss of continuity proceed to section 4. More extensive reviews can be found in [14, 15, 17, 20].

Original formulation. When Witkin [21] introduced the term scale-space, he was concerned with one-dimensional signals and observed that new local extrema cannot be created in this family. Since differentiation commutes with convolution,

$$\partial_{x^n} L(\cdot;\ t) = \partial_{x^n} (g(\cdot;\ t) * f) = g(\cdot;\ t) * \partial_{x^n} f, \tag{5}$$

this non-creation property applies also to any n^{th}-order spatial derivative computed from the scale-space representation. Figure 2(b) illustrates this property, by showing zero-crossings of the second derivative of the smoothed signal at different scales. Note that the trajectories of zero-crossings in scale-space form paths across scales that are never closed from below. This property does, however, not extend to dimensions higher than one.

Causality. Witkins observation shows that Gaussian convolution satisfies certain sufficiency requirements for being a smoothing operation. The first proof of the *necessity* of Gaussian smoothing for generating a scale-space representation was given by Koenderink [11], who also gave a formal extension of the scale-space theory to higher dimensions. He introduced the concept of *causality*, which means that new level surfaces

$$\{(x, y;\ t) \in \mathbb{R}^2 \times \mathbb{R}\colon\ L(x, y;\ t) = L_0\} \tag{6}$$

must not be created in the scale-space representation when the scale parameter is increased. By combining causality with the notions of *isotropy* and *homogeneity*, which essentially mean that all spatial positions and all scale levels must be treated in a similar manner, he showed that the scale-space representation must satisfy the diffusion equation. Related formulations have been expressed by Yuille and Poggio [23] and by Hummel [9].

Non-creation of local extrema and semi-group structure. Lindeberg [15] considered the problem of characterizing those kernels in one dimension that share the property of not

introducing new local extrema in a signal under convolution. Such kernels have to be non-negative and unimodal. Moreover, they can be completely classified. He also imposed a *semi-group* structure on the family of kernels, which means that if two such kernels are convolved with each other, then the resulting kernel will be a member of the same family

$$h(\cdot;\ t_1) * h(\cdot;\ t_2) = h(\cdot;\ t_1 + t_2). \tag{7}$$

In particular, this condition ensures that the transformation from a fine scale to any coarse scale should be of the same type as the transformation from the original signal to any scale in the scale-space representation,

$$\begin{aligned} L(\cdot;\ t_2) &= \{\text{definition}\} = h(\cdot;\ t_2) * f \\ &= \{\text{semi-group}\} = (h(\cdot;\ t_2 - t_1) * h(\cdot;\ t_1)) * f \\ &= \{\text{associativity}\} = h(\cdot;\ t_2 - t_1) * (h(\cdot;\ t_1) * f) \\ &= \{\text{definition}\} = h(\cdot;\ t_2 - t_1) * L(\cdot;\ t_1). \end{aligned} \tag{8}$$

If this semi-group structure is combined with non-creation of local extrema and the existence of a *continuous scale parameter*, and if the kernels are required to be symmetric and satisfy a mild degree of smoothness in the scale direction, then it can be shown that the family is uniquely determined to consist of Gaussian kernels.

Non-enhancement of local extrema and infinitesimal generator. If the semi-group structure *per se* is combined with a strong continuity requirement with respect to the scale parameter, then it follows from well-known results in functional analysis [7] that the scale-space family must have an *infinitesimal generator*. In other words, if a transformation operator $\mathcal{T}_t$ from the input signal to the scale-space representation at any scale t is defined by $L(\cdot;\ t) = \mathcal{T}_t f$, then under reasonable regularity requirements there exists a limit case of this operator (the infinitesimal generator)

$$\mathcal{A} f = \lim_{h \downarrow 0} \frac{\mathcal{T}_h f - f}{h} \tag{9}$$

and the scale-space family satisfies the differential equation

$$\lim_{h \downarrow 0} \frac{L(\cdot, \cdot;\ t + h) - L(\cdot, \cdot;\ t)}{h} = \mathcal{A}(\mathcal{T}_t f) = \mathcal{A} L(\cdot;\ t). \tag{10}$$

Lindeberg [15, 17] showed that this structure implies that the scale-space family must satisfy the diffusion equation if combined with a slightly modified formulation of Koenderinks causality requirement expressed as *non-enhancement of local extrema*:

> *Non-enhancement of local extrema:* If for some scale level t_0 a point x_0 is a non-degenerate local maximum for the scale-space representation at that level (regarded as a function of the space coordinates only) then its value must not increase when the scale parameter increases. Analogously, if a point is a non-degenerate local minimum then its value must not decrease when the scale parameter increases.

Moreover, he showed that this scale-space formulation extends to discrete data as well as to non-symmetric temporal and spatio-temporal image domains.

Scale invariance. A formulation by Florack [4] and continued work by Pauwels *et al.* [19] show that the class of allowable scale-space kernels can be restricted under weaker conditions, essentially by combining the conditions about linearity, shift invariance, rotational invariance and

semi-group structure with *scale invariance*. It can be shown that for a scale invariant rotationally symmetric semi-group, the Fourier transform of the convolution kernel must be of the form

$$\hat{h}(\omega;\ \sigma) = \hat{H}(\omega\sigma) = e^{-\alpha\,|\omega\sigma|^p/2} \tag{11}$$

for some $\alpha > 0$ and $p > 0$, which gives a one-parameter class of possible semi-groups. Florack [4] proposed to use separability in Cartesian coordinates as an additional basic constraint. Except in the one-dimensional case, this fixates h to be a Gaussian. Pauwels *et al.* showed [19] that the corresponding multi-scale representations have *local infinitesimal generators* (basically meaning that the operator $\mathcal{A}$ in (9) is a differential operator) if and only if the exponent p is an even integer. Out of this countable set of choices, $p = 2$ is the only choice that corresponds to a *non-negative convolution kernel* (recall from above that non-creation of local extrema implies that the kernel has to be non-negative). Koenderink and van Doorn [12] carried out a closely related study, and showed that derivative operators are natural operators to derive from a scale-space representation given the assumption of scale invariance.

Relations to biological vision. Interestingly, the results of this computationally motivated analysis are in qualitative agreement with the results of biological evolution. Neurophysiological studies by Young [22] have shown that there are receptive field profiles in the mammalian retina and visual cortex, which can be well modelled by superpositions of Gaussian derivatives.

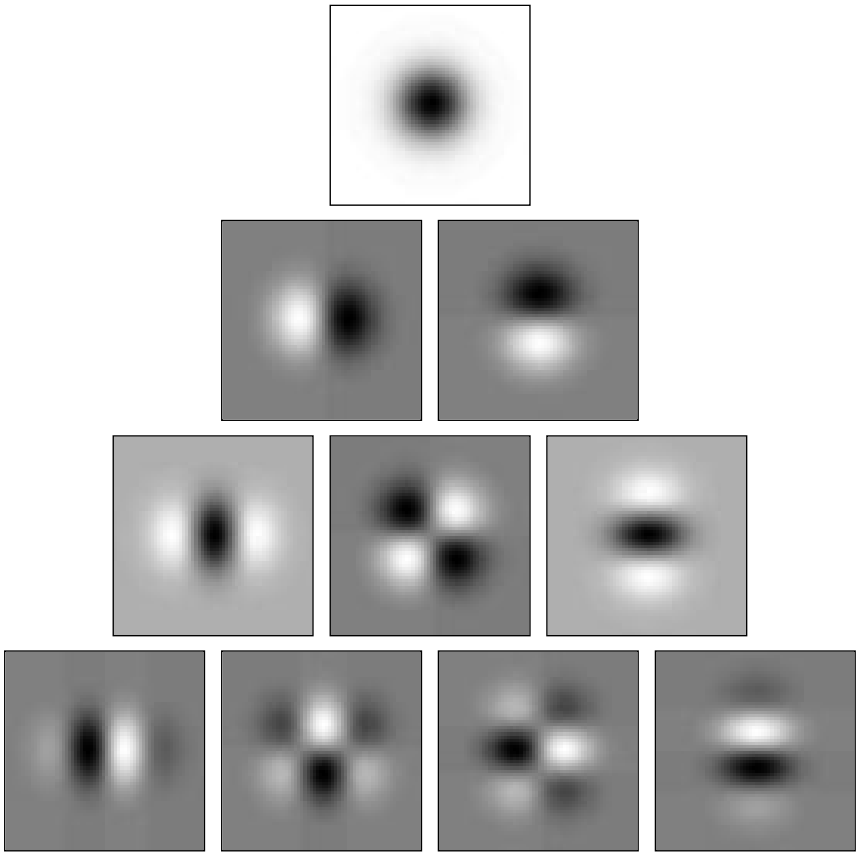

Figure 4: Gaussian derivative kernels up to order four the two-dimensional case.

4 Multi-scale feature detection

The above-mentioned results serve as a formal and empirical justification for using Gaussian filtering followed by derivative computations as initial steps in early processing of image data.

More important, a catalogue is provided of what smoothing kernels are natural to use, as well as a framework for relating filters of different types and at different scales. (Figure 4 shows a few examples of filter kernels from this *filter bank*.) Linear filtering, however, cannot be used as the only component in a vision system aimed at deriving symbolic representations from images; some non-linear processing steps must be introduced into the analysis. More concretely, some mechanism is required for combining the output of these *Gaussian derivative operators* of different orders and at different scales into more explicit descriptors of the image geometry. An approach that has been advocated by Koenderink and his co-workers is to describe image properties in terms of differential geometric descriptors, *i.e.*, different possibly non-linear combinations of derivatives. Since one would typically like image descriptors to possess invariance properties under certain transformations (typically, rotations, rescalings and affine or perspective deformations), this naturally leads to the study of differential invariants [6]. A major difference compared to traditional invariant theory, however, is that the primitive derivative operators in this case are smoothed derivatives computed from the scale-space representation.[1] In this section, a few examples will be given of how this framework of *multi-scale differential geometry* can be used for expressing various types of multi-scale feature detectors. The output from these feature detectors is in turn intended to be used as input to higher-level visual modules, for task such as object recognition, object reconstruction/manipulation and robot navigation.

4.1 Feature detectors expressed in terms of local directional derivatives

Edge detection. A notion of *gauge coordinates* which has been adopted in the computer vision community is to express image descriptors in terms of local directional derivatives defined from certain preferred coordinate systems. At any image point, introduce a local (u, v)-system such that the v-direction is parallel to the gradient direction $(\cos\alpha, \sin\alpha) = (L_x, L_y)/\sqrt{L_x^2 + L_y^2}$, and introduce directional derivative operators along these directions by

$$\partial_{\bar{u}} = \sin\alpha\, \partial_x - \cos\alpha\, \partial_y, \qquad \partial_{\bar{v}} = \cos\alpha\, \partial_x + \sin\alpha\, \partial_y. \tag{12}$$

Then, we can define an edge point as a point for which the gradient assumes a local maximum in the gradient direction, and restate this edge definition as

$$\begin{cases} L_{vv} = 0, \\ L_{vvv} < 0, \end{cases} \tag{13}$$

where L_{vv} and L_{vvv} denote second- and third-order directional derivatives in the v-direction. After expansion to Cartesian coordinates and simplification, this edge definition assumes the form

$$\begin{cases} \tilde{L}_{vv} = & L_v^2 L_{vv} = & L_x^2 L_{xx} + 2 L_x L_y L_{xy} + L_y^2 L_{yy} = 0, \\ \tilde{L}_{vvv} = & L_v^3 L_{vvv} = & L_x^3 L_{xxx} + 3 L_x^2 L_y L_{xxy} + 3 L_x L_y^2 L_{xyy} + L_y^3 L_{yyy} < 0. \end{cases} \tag{14}$$

Interpolating for zero-crossings of $\tilde{L}_{vv}$ within the sign-constraints of $\tilde{L}_{vvv}$ gives a straightforward method for sub-pixel edge detection. Figure 5(a) shows the result of applying this edge detector to an image of an arm at scale levels $t = 1.0$, 16.0 and 256.0. Observe how qualitatively different types of edge curves are extracted at the different scales. A characteristic behaviour is that most of the sharp edge structures corresponding to object boundaries give rise to edge curves at both fine and coarse scales. Moreover, the number of spurious edges due to noise is much larger at fine scales, whereas the localization of the edges can be poor at coarse scales. Notably, the shadow of the arm can only be extracted as a connected curve at a coarse scale. This example constitutes one illustration of the *need* for including image operators at coarse scales when extracting general classes of image structures from real-world data.

[1] In this respect, there is a high degree of similarity to Schwartz distribution theory [8], although in scale-space

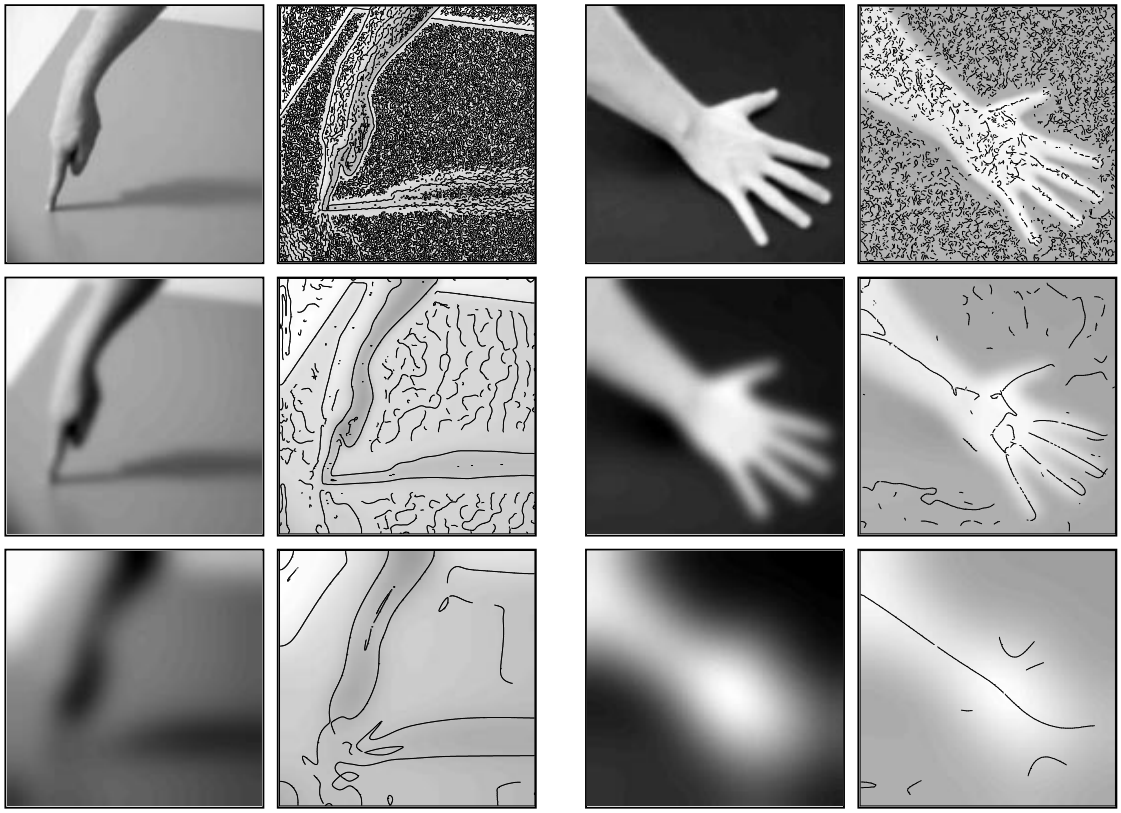

Figure 5: Edges and bright ridges detected at scale levels $t = 1.0$, 16.0 and 256.0, respectively.

Ridge detection. A ridge detector can be expressed in a conceptually similar way as follows: Introduce at any image point a local (p, q)-system aligned to the principal curvature directions such that the mixed second-order derivative is zero, *i.e.*, $L_{pq} = 0$. Then, we can define a bright (dark) ridge point as a point for which the intensity assumes a local maximum in the main principal curvature direction. In terms of the (p, q)-coordinates, this definition can be written

$$\left\{ \begin{array}{rcl} L_p & = & 0, \\ L_{pp} & < & 0, \\ |L_{pp}| & \geq & |L_{qq}|, \end{array} \right. \quad \text{or} \quad \left\{ \begin{array}{rcl} L_q & = & 0, \\ L_{qq} & < & 0, \\ |L_{qq}| & \geq & |L_{pp}|, \end{array} \right. \tag{15}$$

depending on whether the p- or the q-direction corresponds to the maximum absolute value of the principal curvatures. At points where the gradient does not vanish, this condition can equivalently be expressed as follows in the (u, v)-system and in terms of Cartesian partial derivatives

$$\left\{ \begin{array}{rl} L_{uv} & = L_x L_y \left(L_{xx} - L_{yy}\right) - (L_x^2 - L_y^2) L_{xy} \quad = 0, \\ L_{uu}^2 - L_{vv}^2 & = (L_y^2 - L_x^2) \left(L_{xx} - L_{yy}\right) - 4 L_x L_y L_{xy} \quad > 0. \end{array} \right. \tag{16}$$

Figure 5(b) shows the result of applying this ridge detector to an image of an arm. As can be seen, the types of ridge curves that are obtained are strongly strongly scale dependent. At very fine scales, the ridge detector responds mainly to noise and spurious fine-scale textures. Then, the fingers give rise to ridge curves at scale level $t = 16.0$, and the arm as a whole is extracted as a long ridge curve at $t = 256.0$. Notably, these ridge descriptors are much more sensitive to the choice of scale levels than the edge features in figure 5(a). In particular, no single scale level is appropriate for describing the dominant ridge structures in this image.

theory it is neither needed nor desired to approach the infinitesimal limit case when the support regions of the test functions associated with the generalized derivative operators tend to zero.

Corner detection and blob detection. A useful entity for blob detection is the curvature of level curves κ multiplied by the gradient magnitude L_v raised to the power of three

$$\tilde{\kappa} = L_v^3 \, \kappa = L_v^2 L_{uu} = L_x^2 L_{yy} + L_y^2 L_{xx} - 2 L_x L_y L_{xy}, \tag{17}$$

whereas zero-crossings and spatial maxima of the Laplacian

$$\nabla^2 L = L_{xx} + L_{yy} = L_{uu} + L_{vv} \tag{18}$$

can be used for stereo matching and blob detection, respectively.

5 Automatic scale selection

Although the scale-space theory presented so far provides a well-founded framework for *representing* and detecting image structures at multiple scales, it does not address the problem of how to *select* locally appropriate scales for further analysis. Whereas the problem of finding "the best scales" for handling a given real-world data set may be regarded as intractable unless further information is available, there are many situations in which a mechanism is required for generating hypotheses about interesting scales. A general methodology for feature detection with automatic scale selection has been proposed in [15, 16]. The approach is based on the evolution over scales of (possibly non-linear) combinations of *normalized derivatives* defined by

$$\partial_{\xi_i} = t^{\gamma/2} \, \partial_{x_i}, \tag{19}$$

where γ is a free parameter to be tuned to the task at hand. The basic idea proposed in the abovementioned sources is to apply the feature detector at all scales, and then *select scale levels from the scales at which normalized measures of feature strength assume local maxima with respect to scale*. Intuitively, this approach corresponds to the selection of the scales at which the operator response is as strongest. Moreover, it can be shown that the specific form of derivative normalization in (19) spans a large class of normalization approaches for which the scale selection mechanism has a desirable behaviour under size variations of the input pattern.

Feature type	Normalized strength measure for scale selection	Value of γ
Edge	$t^{\gamma/2} \, L_v$	1/2
Ridge	$t^{2\gamma} (L_{pp} - L_{qq})^2$	3/4
Corner	$t^{2\gamma} L_v^2 L_{uu}$	1
Blob	$t^{\gamma} \nabla^2 L$	1

Table I: Measures of feature strength used for feature detection with automatic scale selection.

Figures 6–15 show a few examples of integrating this scale selection mechanism with the differential geometric feature detectors considered in section 4. Here, the extracted features have been illustrated graphically in two ways; (i) as two-dimensional spatial projections onto the image plane, and (ii) as three-dimensional entities in scale-space, with the height over the image plane representing the selected scales. Figure 6–8 show the result of performing edge detection on the arm image and introducing a measure of significance by integrating the measure of edge strength along each connected edge curve. As can be seen, coarse scales are selected for the diffuse edge structures due to illumination effects, whereas fine scales are selected for the sharp edges due to object boundaries. Notably, the selected scale levels vary substantially along the shadow of the arm, which is necessary to obtain a good trade-off in the inherent conflict between detection and localization properties. Figures 9–11 show corresponding ridge curves detected from the other hand image. Observe how a coarse-scale ridge descriptor is obtained for the arm as a whole, whereas the individual fingers give rise to ridge curves at finer scales. Figures 12–13

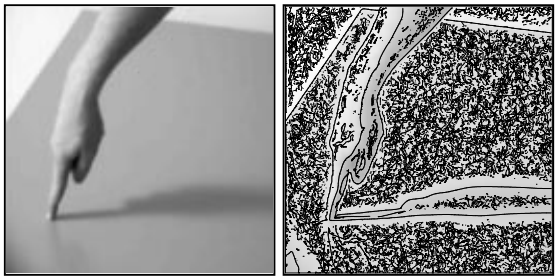

Figure 6: Results of edge detection with automatic scale selection on the arm image (all edges).

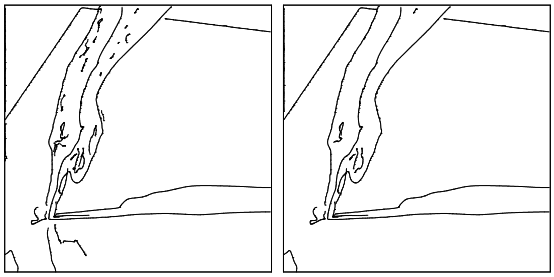

Figure 7: The 50 and 10 most significant edge curves as ranked on the integrated measure of edge strength along each connected edge curve.

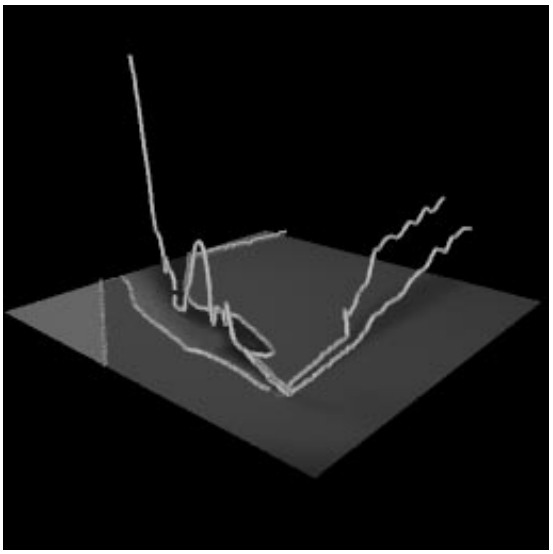

Figure 8: Three-dimensional view of the 10 most significant edge curves drawn in three-dimensional scale-space with the selected scale represented as the height over the image plane.

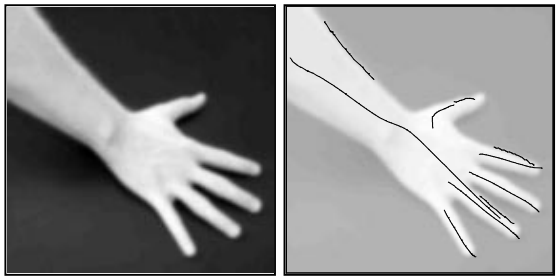

Figure 9: The 10 strongest bright ridge curves obtained by with automatic scale selection.

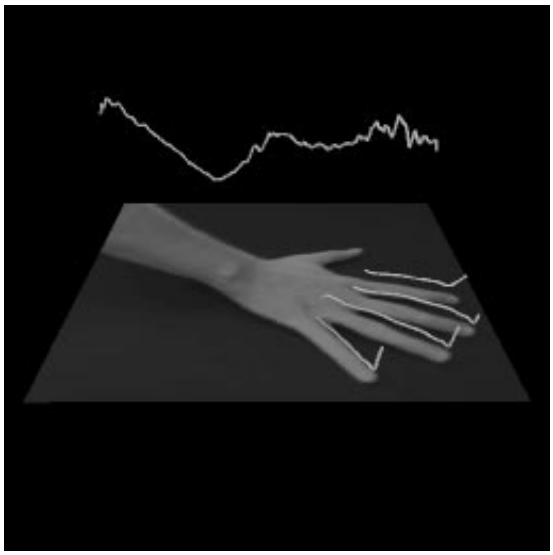

Figure 10: Three-dimensional view of the five strongest ridge curves in scale-space.

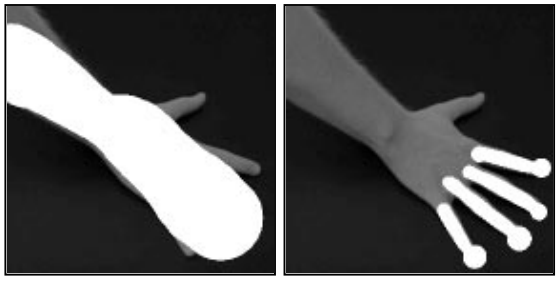

Figure 11: Backprojections of the five strongest ridge curves to the image domain in terms of unions of bright circles centered at the ridge curves and with the radius proportional to the selected scale.

show the result of detecting corners in a closely related way, by detecting points at which the normalized measure of corner strength assumes a simultaneous local maximum over space and scale. Note in figure 13 that a coarse scale response is obtained for the large size corner structure as a whole, whereas finer-scale responses are obtained for the superimposed corner structures of smaller spatial extent. Figures 14–15 show corresponding result of blob detection on a sunflower image, where we can see how the size variations in the image of the sunflowers are captured in the features extracted by the scale selection mechanism.

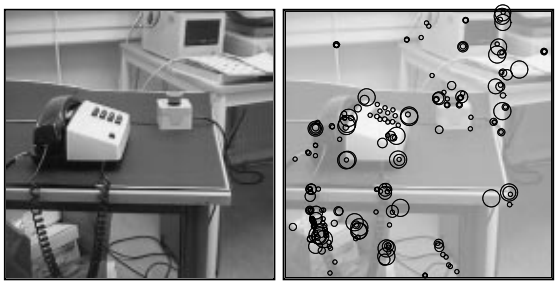

Figure 12: Results of corner detection with automatic scale selection on an office scene (200 strongest junction responses).

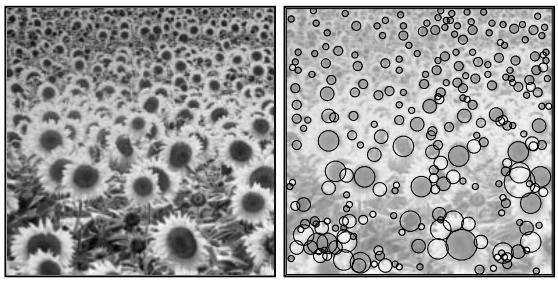

Figure 14: Results of blob detection detection with automatic scale selection on a sunflower image (200 strongest blob responses).

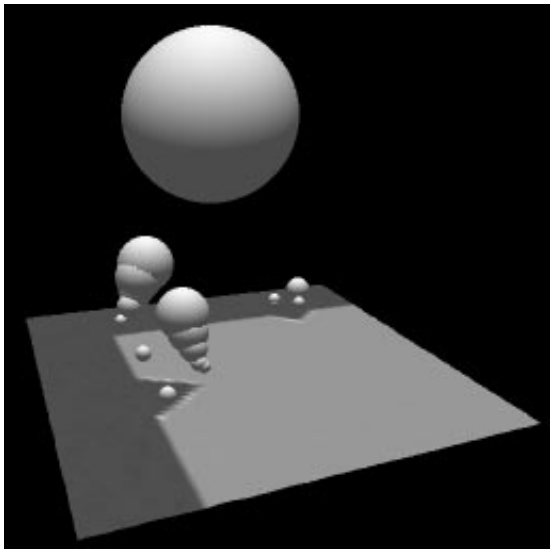

Figure 13: Three-dimensional view of corners detected from a synthetic image with corner structures of different spatial extent.

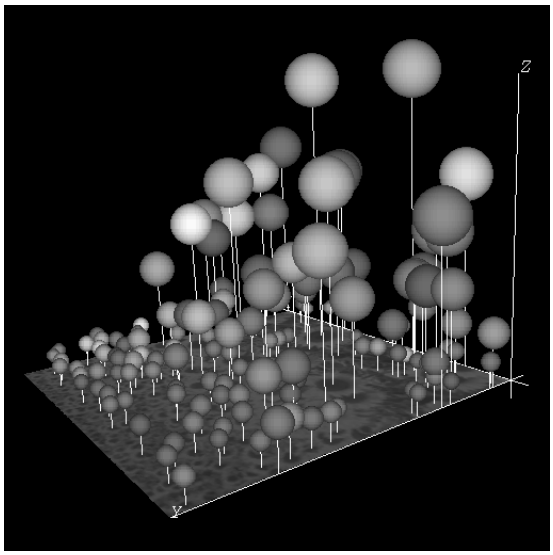

Figure 15: Three-dimensional view of the blob responses obtained from the sunflower image.

6 Summary and outlook

Scale-space theory provides a framework for modelling image structures at multiple scales, and the output from the scale-space representation can be used as input to a variety of early visual tasks. Operations like feature detection, feature classification and shape computation can be expressed directly in terms of (possibly non-linear) combinations of Gaussian derivatives at multiple scales. In this sense, the scale-space representation can serve as a basis for early vision. In the terminology of Kuhn [13], scale-space theory can also be seen as a promising seed to a new paradigm for computer vision. During the last few decades a number of other approaches to

multi-scale representations have been developed, which are more or less related to scale-space theory, notably the theories of *pyramids* [1, 2, 10], *wavelets* [18, 3] and *multi-grid methods* [5]. Despite their qualitative differences, the increasing popularity of each of these approaches indicates that the crucial notion of *scale* is increasingly appreciated by the computer vision community and by researchers in other fields. The goal of this presentation has been to provide a few selected pointers to central ideas behind the recently developed scale-space theory and to show examples of straightforward applications. Main issues of current research concern the incorporation of scale-space techniques into increasingly complex visual modules and the extension to non-linear scale-space concepts more committed to specific tasks at hand [20].

References

1 P. J. Burt. Fast filter transforms for image processing. *CVGIP*, 16:20–51, 1981.

2 J. L. Crowley. *A Representation for Visual Information*. PhD thesis, Carnegie-Mellon University, Robotics Institute, Pittsburgh, Pennsylvania, 1981.

3 I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.

4 L. M. J. Florack. *The Syntactical Structure of Scalar Images*. PhD thesis, Dept. Med. Phys. Physics, Univ. Utrecht, NL-3508 Utrecht, Netherlands, 1993.

5 W. Hackbush. *Multi-Grid Methods and Applications*. Springer-Verlag, New York, 1985.

6 D. Hilbert. Über die vollen Invariantensystemen. *Math. Annal.*, 42:313–373, 1893.

7 E. Hille and R. S. Phillips. *Functional Analysis and Semi-Groups*, volume XXXI. American Mathematical Society Colloquium Publications, 1957.

8 L. Hörmander. *Linear Partial Differential Operators*, volume 257 of *Grundlehren der mathematische Wissenshaften*. Springer-Verlag, 1963.

9 R. A. Hummel. The scale-space formulation of pyramid data structures. In L. Uhr, (ed.), *Parallel Computer Vision*, pages 187–223, New York, 1987. Academic Press.

10 J.-J. Jolion and A. Rozenfeld. *A Pyramid Framework for Early Vision*. Kluwer, 1994.

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

12 J. J. Koenderink and A. J. van Doorn. Generic neighborhood operators. *IEEE-PAMI*, 14(6):597–605, 1992.

13 T. S. Kuhn. *The Structure of Scientific Revolutions*. Univ. of Chicago Press, Chicago, 1962.

14 T. Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *J. Appl. Stat.*, 21(2):225–270, 1994. Suppl. *Adv. in Appl. Stat.: Statistics and Images: 2*.

15 T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer, Netherlands, 1994.

16 T. Lindeberg. Feature detection with automatic scale selection. Technical Report ISRN KTH/NA/P--96/18--SE, KTH, 1996.

17 T. Lindeberg. On the axiomatic foundations of linear scale-space. In J. Sporring, M. Nielsen, L. Florack, and P. Johansen, (eds.), *Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory*, Copenhagen, Denmark, 1996. Kluwer. (To appear).

18 Y. Meyer. *Ondolettes et Operateurs*. Hermann, 1988.

19 E. J. Pauwels, P. Fiddelaers, T. Moons, and L. J. van Gool. An extended class of scale-invariant and recursive scale-space filters. *IEEE-PAMI*, 17(7):691–701, 1995.

20 B. ter Haar Romeny, (ed.). *Geometry-Driven Diffusion in Computer Vision*. Kluwer, 1994.

21 A. P. Witkin. Scale-space filtering. In *8th IJCAI*, pages 1019–1022, 1983.

22 R. A. Young. The Gaussian derivative model for spatial vision: I. Retinal mechanisms. *Spatial Vision*, 2:273–293, 1987.

23 A. L. Yuille and T. A. Poggio. Scaling theorems for zero-crossings. *PAMI*, 8:15–25, 1986.

More material on this subject can be fetched from http://www.bion.kth.se and http://www.bion.kth.se/~tony.