On the axiomatic foundations of linear scale-space: Combining semi-group structure with causality vs. scale invariance

Tony Lindeberg

[1]Computational Vision and Active Perception Laboratory (CVAP)
Dept. of Numerical Analysis and Computing Science,
KTH, S-100 44 Stockholm, Sweden
http://www.nada.kth.se/~tony
Email: tony@nada.kth.se

Technical report ISRN KTH/NA/P–94/20–SE, Aug. 1994.

Revised version published as Chapter 6 in
J. Sporring, M. Nielsen, L. Florack, and P. Johansen (eds.)
Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory, (Copenhagen, Denmark, May 1996),
Kluwer Academic Publishers, 1997.

Abstract

Since the pioneering work by Witkin (1983) and Koenderink (1984) on the notion of "scale-space representation", a large number of different scale-space formulations have been stated, based on different types of assumptions (usually referred to as scale-space axioms). The main subject of this chapter is to provide a synthesis between these linear scale-space formulations and to show how they are related. Another aim is to show how the scale-space formulations, which were originally expressed for continuous data on spatial domains without preferred directions, can be extended to discrete data as well as to spatio-temporal domains with preferred directions. Connections will also be pointed out to approaches based on non-uniform smoothing.

Keywords: scale-space, Gaussian filtering, causality, diffusion, scale invariance, multi-scale representation, computer vision, signal processing

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 essential tool is a formal theory for describing structures at multiple scales. In image processing and computer vision, this insight has lead to the general methodology of representing data at multiple scales, using concepts such as quad-trees, pyramids, wavelets and scale-space representation. A common idea behind the creation of these concepts is that for any input image, a set of gradually smoothed or simplified images should be generated, in which fine scale structures are successively suppressed.

The subject of this chapter is to consider the axiomatic foundations of the special framework for multi-scale representation known as *linear scale-space theory* (Witkin 1983; Koenderink 1984; Babaud *et al.* 1986; Yuille and Poggio 1986; Koenderink and van Doorn 1992; Lindeberg 1990, 1993, 1994; Florack *et al.* 1992, 1994; Pauwels *et al.* 1995). Assuming that no *a priori* information is available about the measurement data, the linear *scale-space representation* of any given signal is according to this theory derived from the basic constraints that it should

- be generated by linear and shift-invariant operators (convolutions),
- possess a continuous scale parameter, and
- the transformation from any fine level in the scale-space representation to any coarser level must not introduce "spurious structures".

The aim of the last requirement is to guarantee that fine-scale structures should disappear monotonically with increasing scale, such that any coarse-scale representation in the multi-scale family can be regarded as a simplification of any finer-scale representation. Since the first scale-space formulations by Witkin (1983) and Koenderink (1984), this condition about "non-creation of structure" has been formalized in different ways by different authors (to be reviewed in section 2).

A notable coincidence between the different scale-space formulations that have been stated is that the Gaussian kernel arises as a unique choice for a large number of different combinations of underlying assumptions (scale-space axioms). In summary, linear scale-space theory states that a natural way to process a given N-dimensional input signal $f\colon \mathbb{R}^N \to \mathbb{R}$ is by convolving it with Gaussian kernels[1]

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

and their derivatives

$$g_{x^\alpha}(x;\ \sigma^2) = \partial_{x_1^{\alpha_1}\dots x_D^{\alpha_D}}\, g(x;\ \sigma^2) \tag{2}$$

of various widths σ. The output from these operations can then in turn be used as a basis for expressing a large number of early visual operations, such as feature detection, matching, optic flow and computation of shape cues. A particularly convenient framework for formalizing such processes is in terms of multi-scale differential geometric invariants and singularities of these (Koenderink and van Doorn 1987; Florack *et al.* 1993; Lindeberg 1993, 1994; Johansen 1993).[2]

A main purpose of this chapter is to discuss relations between the abovementioned scale-space formulations. Moreover, a complementary treatment will be given showing how a scale-space formulation previously expressed for discrete signals applies to continuous signals. Another main goal is to show how this selection of scale-space axioms, based on the assumption of a semi-group structure combined with a reformulation of the causality requirement in terms of non-enhancement of local extrema, relates to scale-space formulations based on scale invariance.

It will also be indicated how this approach can be generalized to the following types of scale-space concepts:

- affine Gaussian scale-space,
- spatio-temporal scale-space, and
- non-linear scale-space.

The presentation is organized as follows: Section 2 provides the necessary background by reviewing the notions of causality, non-enhancement of local extrema and scale invariance. Then, section 3 shows how assumptions about causality and a semi-group structure uniquely determine the smoothing kernel to be a Gaussian, if combined with rotational symmetry and certain regularity assumptions. In view of this result, it will be described how causality combines with scale invariance. Finally, section 4 gives a summary of the main results and section 5 points out connections to approaches based on non-uniform smoothing.

---

[1] Here, $x = (x_1, \dots, x_N) \in \mathbb{R}^N$ is standard vector notation for an D-dimensional variable, and $\alpha = (\alpha_1, \dots, \alpha_N) \in \mathbb{Z}^N$ constitutes so-called multi-index notation with $x^\alpha = x_1^{\alpha_1} \dots x_N^{\alpha_N}$.

[2] See also chapter 1 by ter Haar Romeny (1996) and chapter 8 by Salden (1996) in this volume.

2 Axiomatic formulations of linear scale-space

Since the pioneering work by Witkin (1983) and Koenderink (1984), a large number of scale-space formulations have been stated. This section gives a brief historical review of the development of linear scale-space theory with special emphasis on the scale-space formulations that we shall later build upon.

For further overviews, the reader is referred to (Lindeberg 1994, 1996; Lindeberg and ter Haar Romeny 1994) as well as the preface by Koenderink (1996) and chapter 5 by Florack (1996) in this volume.

2.1 Original formulation

When Witkin (1983) introduced the term "scale-space", he was concerned with one-dimensional signals and observed that new local extrema cannot be created under Gaussian convolution. Since differentiation commutes with convolution,

$$\partial_{x^n} L(\cdot;\ t) = \partial_{x^n} (g(\cdot;\ t) * f) = g(\cdot;\ t) * \partial_{x^n} f, \tag{3}$$

this non-creation property applies also to any n^{th}-order spatial derivative computed from the scale-space representation. Specifically, he applied this property to zero-crossings of the second derivative to construct so-called "fingerprints".

2.2 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 a scale-space representation was given by Koenderink (1984), 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

$$\{(x, y;\ t) \in \mathbb{R}^2 \times \mathbb{R}\colon\ L(x, y;\ t) = L_0\} \tag{4}$$

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

$$\partial_t L = \frac{1}{2} \nabla^2 L. \tag{5}$$

The technique used for proving this necessity result was by studying the level surface through any point in scale-space for which the grey-level function assumes a maximum with respect to the spatial coordinates. If no new level surface is to be created when increasing scale, the level surface should point with its concave side towards decreasing scales. This gives rise to a sign condition on the curvature of the level surface, which assumes the form (5) when expressed in terms of derivatives of the scale-space representation.

A similar result was given by Yuille and Poggio (1985, 1986). Related formulations have been expressed by Hummel (1986, 1987).

2.3 Non-creation of local extrema

Lindeberg (1990) 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. A kernel $h \in \mathbb{L}_1$ possessing the property that for *any* input signal $f_{in} \in \mathbb{L}_1$ the number of extrema (zero-crossings) in the convolved signal $f_{out} = h * f_{in}$ is always less than or equal to the number of local extrema (zero-crossings) in the original signal is termed a *scale-space kernel*.

Such kernels must be non-negative and unimodal both in the spatial and the frequency domain. Moreover they can be completely classified using classical results by Schoenberg (1950, 1953) (see also Hirschmann and Widder (1955) and Karlin (1968)). Besides trivial *translations* and *rescalings*, there are two primitive types of linear and shift-invariant smoothing transformations in the *continuous* case that never increase the number of extrema (zero-crossings):

- convolution with *Gaussian kernels*,

$$h(x) = e^{-\gamma x^2}, \tag{6}$$

- convolution with *truncated exponential functions*,

$$h(x) = \begin{cases} e^{-x/|\mu|} & x \geq 0, \\ 0 & x < 0, \end{cases} \qquad h(x) = \begin{cases} e^{x/|\mu|} & x \leq 0, \\ 0 & x > 0, \end{cases} \tag{7}$$

In fact, this theory states that *all* continuous scale-space kernels can be decomposed into (possibly infinite) compositions of these primitive smoothing operations.

Correspondingly, in the *discrete* case, there are besides rescaling and translation, three primitive types of smoothing transformations (where $f_{out} = h * f_{in}$):

- two-point weighted averaging or *generalized binomial smoothing*,

$$\begin{aligned} f_{out}(x) &= f_{in}(x) + \alpha_i \, f_{in}(x-1) \qquad (\alpha_i \geq 0), \\ f_{out}(x) &= f_{in}(x) + \delta_i \, f_{in}(x+1) \qquad (\delta_i \geq 0), \end{aligned} \tag{8}$$

- moving average or *first-order recursive filtering*,

$$\begin{aligned} f_{out}(x) &= f_{in}(x) + \beta_i \, f_{out}(x-1) \qquad (0 \leq \beta_i < 1), \\ f_{out}(x) &= f_{in}(x) + \gamma_i \, f_{out}(x+1) \qquad (0 \leq \gamma_i < 1), \end{aligned} \tag{9}$$

- *infinitesimal smoothing* (or diffusion smoothing) with the generating function of the smoothing kernel being of the form

$$H_{semi-group}(z) = \sum_{n=-\infty}^{\infty} h(n) \, z^n = e^{t(az^{-1}+bz)}. \tag{10}$$

Among these discrete kernels, the generalized binomial kernels provide a natural basis for constructing pyramid representations (Burt 1981; Crowley 1981), whereas recursive filters can be used for efficient implementations of smoothing operations (Deriche 1987). The interpretation and applications of the class of infinitesimal discrete smoothing kernels will be apparent in next section.

2.4 Semi-group structure and continuous scale parameter

A natural structure to impose on a scale-space representation is a *semi-group* structure *i.e.*, if every smoothing kernel is associated with a parameter value, and if two such kernels are convolved with each other, then the resulting kernel should be a member of the same family,

$$h(\cdot;\ t_1) * h(\cdot;\ t_2) = h(\cdot;\ t_1 + t_2). \tag{11}$$

This condition states that all scale-space transformations are of the same type. 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,

$$\begin{aligned} L(\cdot;\ t_2) &= \{\text{definition}\} = h(\cdot;\ t_2) * f \\ &= \{\text{semi-group}\} = (h(\cdot;\ t_2 - t_1) * h(\cdot;\ t_1)) * f \\ &= \{\text{associativity}\} = h(\cdot;\ t_2 - t_1) * (h(\cdot;\ t_1) * f) \\ &= \{\text{definition}\} = h(\cdot;\ t_2 - t_1) * L(\cdot;\ t_1). \end{aligned} \tag{12}$$

If this semi-group structure is combined with non-creation of local extrema, the existence of a *continuous scale parameter*, and if the kernels are required to be symmetric and normalized and to satisfy a mild degree of smoothness in the scale direction (Borel-measurability), then the family of smoothing kernels is uniquely determined to be a Gaussian (Lindeberg 1990)

$$h(x;\ t) = \frac{1}{\sqrt{2\pi\alpha t}} e^{-x^2/(2\alpha t)} \quad (t > 0 \quad \alpha \in \mathbb{R}^+). \tag{13}$$

If on the other hand the spatial symmetry requirements are relaxed, we obtain translated or *velocity-adapted Gaussian kernels*

$$h(x;\ \delta, t) = \frac{1}{\sqrt{2\pi\alpha t}} e^{-(x-\delta)^2/(2\alpha t)} \quad (t > 0, \alpha \in \mathbb{R}^+) \tag{14}$$

(with δ determined by local velocity information). This family of kernels has also been derived by Florack (1992), based on dimensional analysis (see section 2.6).

Corresponding arguments about a semi-group structure on a spatially symmetric discrete domain uniquely lead to the *discrete analogue of the Gaussian kernel*

$$L(x;\ t) = \sum_{n=-\infty}^{\infty} T(n;\ t)\, f(x-n) \quad \text{where} \quad T(n;\ t) = e^{-t} I_n(t), \tag{15}$$

where I_n are the modified Bessel functions of integer order (Abramowitz and Stegun 1964). This scale-space family corresponds to the result of letting $a = b$ in (10) and describes the solution to the *semi-discretized diffusion equation*

$$\partial_t L(x;\ t) = \frac{1}{2}(L(x+1;\ t) - 2L(x;\ t) + L(x-1;\ t)) = \frac{1}{2}\nabla_3^2 L(x;\ t), \tag{16}$$

where ∇_3^2 is the standard second-order difference operator. Indeed, the discrete analogue of the Gaussian kernel can be interpreted as the result of exponentiating the second-order difference operator

$$T(\cdot;\ t) = e^{\frac{t}{2}\nabla_3^2}. \tag{17}$$

When considering temporal image data in a real-time situation, we have to require the scale-space kernels to be *time-causal* and to not extend into the future (Koenderink 1988; Lindeberg and Fagerström 1996). Imposing this constraint on the infinitesimal smoothing filters in (10) uniquely gives rise to the *temporal scale-space* generated by convolution with the Poisson kernel (Lindeberg 1996)

$$L(x;\ t) = \sum_{n=-\infty}^{\infty} p(n;\ t)\, f(x-n) \quad \text{where} \quad p(n;\ \lambda) = e^{-\lambda}\, \frac{\lambda^n}{n!}. \tag{18}$$

In terms of differential equations, this scale-space family satisfies

$$\partial_\lambda L = -\delta_- L, \tag{19}$$

where δ_- denotes the backward difference operator $\delta_- L(t) = L(t) - L(t-1)$.

Despite the completeness of these results based on non-creation of extrema (or zero-crossings), however, they cannot be extended to higher dimensions, since in two (and higher) dimensions, since there are no non-trivial kernels guaranteed to never increase the number of local extrema in a signal (see (Lifshitz and Pizer 1990; Yuille and Poggio 1988; Lindeberg 1990, 1994) for counterexamples).

2.5 Non-enhancement and infinitesimal generator

If the semi-group structure is combined with a strong continuity requirement with respect to the scale parameter, then it follows from well-known results in functional analysis (Hille and Phillips 1957) 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

$$L(\cdot;\ t) = \mathcal{T}_t f, \tag{20}$$

then under reasonable regularity requirements there exists a limit case of this operator (the infinitesimal generator)

$$\mathcal{A} f = \lim_{h \downarrow 0} \frac{\mathcal{T}_h f - f}{h} \tag{21}$$

and the scale-space family satisfies the differential equation

$$\partial_t L(\cdot;\ t) = \lim_{h \downarrow 0} \frac{L(\cdot;\ t+h) - L(\cdot;\ t)}{h} = \mathcal{A}(\mathcal{T}_t f) = \mathcal{A} L(\cdot;\ t). \tag{22}$$

For discrete signals, Lindeberg (1990, 1991, 1994) showed that this structure implies that the scale-space family must satisfy a (semi-discretized) 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.

In summary, these conditions imply that the scale-space family $L\colon \mathbb{Z}^N \times \mathbb{R}^+ \to \mathbb{R}$ of a discrete signal $f\colon \mathbb{Z}^N \to \mathbb{R}$ must satisfy the semi-discrete differential equation

$$(\partial_t L)(x;\ t) = (\mathcal{A}_{ScSp} L)(x;\ t) = \sum_{\xi \in \mathbb{Z}^N} a_\xi L(x - \xi;\ t), \tag{23}$$

for some *infinitesimal scale-space generator* $\mathcal{A}_{ScSp}$ characterized by

- the *locality* condition $a_\xi = 0$ if $|\xi|_\infty > 1$,
- the *positivity* constraint $a_\xi \geq 0$ if $\xi \neq 0$,
- the *zero sum* condition $\sum_{\xi \in \mathbb{Z}^N} a_\xi = 0$, as well as
- the *symmetry* requirements
  - $a_{(-\xi_1, \xi_2, ..., \xi_N)} = a_{(\xi_1, \xi_2, ..., \xi_N)}$ and
  - $a_{P_k^N(\xi_1, \xi_2, ..., \xi_N)} = a_{(\xi_1, \xi_2, ..., \xi_N)}$

  for all $\xi = (\xi_1, \xi_2, ..., \xi_N) \in \mathbb{Z}^N$ and all possible permutations P_k^N of N elements.

The locality condition means that $\mathcal{A}_{ScSp}$ corresponds to the discretization of derivatives of order up to two. In one and two dimensions, (23) reduces to

$$\begin{aligned} \partial_t L &= \alpha_1 \nabla_3^2 L, \\ \partial_t L &= \alpha_1 \nabla_5^2 L + \alpha_2 \nabla_{\times^2}^2 L, \\ \partial_t L &= \alpha_1 \nabla_7^2 L + \alpha_2 \nabla_{+^3}^2 L + \alpha_3 \nabla_{\times^3}^2 L, \end{aligned} \tag{24}$$

for some constants $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$. Here, the symbols, ∇_5^2 and $\nabla_{\times^2}^2$, denote the common discrete approximations of the Laplacian operator in two dimensions (below the notation $f_{-1,1}$ stands for $f(x-1, y+1)$ etc.):

$$\begin{aligned} (\nabla_5^2 f)_{0,0} &= f_{-1,0} + f_{+1,0} + f_{0,-1} + f_{0,+1} - 4 f_{0,0}, \\ (\nabla_{\times^2}^2 f)_{0,0} &= 1/2 (f_{-1,-1} + f_{-1,+1} + f_{+1,-1} + f_{+1,+1} - 4 f_{0,0}). \end{aligned} \tag{25}$$

whereas ∇_7^2, $\nabla_{+^3}^2$ and $\nabla_{\times^3}^2$ represent corresponding discrete approximations of the Laplacian operator in three dimensions:

$$\begin{aligned} (\nabla_7^2 f)_{0,0,0} &= f_{-1,0,0} + f_{+1,0,0} + f_{0,-1,0} \\ &\quad + f_{0,+1,0} + f_{0,0,-1} + f_{0,0,+1} 6 f_{0,0,0}, \\ (\nabla_{+^3}^2 f)_{0,0,0} &= \tfrac{1}{4} (f_{-1,-1,0} + f_{-1,+1,0} + f_{+1,-1,0} + f_{+1,+1,0} \\ &\quad + f_{-1,0,-1} + f_{-1,0,+1} + f_{+1,0,-1} + f_{+1,0,+1} \\ &\quad + f_{0,-1,-1} + f_{0,-1,+1} + f_{0,+1,-1} + f_{0,+1,+1} - 12 f_{0,0,0}), \\ (\nabla_{\times^3}^2 f)_{0,0,0} &= \tfrac{1}{4} (f_{-1,-1,-1} + f_{-1,-1,+1} + f_{-1,+1,-1} + f_{-1,+1,+1} \\ &\quad + f_{+1,-1,-1} + f_{+1,-1,+1} + f_{+1,+1,-1} + f_{+1,+1,+1} - 8 f_{0,0,0}). \end{aligned}$$

If the spatial symmetry requirements underlying this formulation are relaxed, then a larger class of scale-space transformations will be obtained (see section 5.1). The structure required from such a non-isotropic smoothing operation to satisfy non-enhancement of local extrema is that the infinitesimal scale-space generator should satisfy the locality, positivity and zero sum conditions, essentially corresponding to the discretization of a second-order differential operator.

If we would like the Fourier transform of the associated convolution kernel to be real (corresponding to to a milder degree of symmetry for the smoothing kernels, such that they will be mirror symmetric along any axis through the origin), then a necessary requirement is that the filter coefficients should satisfy $a_\xi = a_{-\xi}$

2.6 Scale invariance

A formulation by Florack (1992, 1993) and continued work by Pauwels *et al.* (1995) show that the class of allowable scale-space kernels can be restricted under weaker conditions, essentially by combining the earlier mentioned conditions about linearity, shift invariance, rotational invariance and semi-group structure with *scale invariance*. The basic argument is taken from physics; physical laws must be independent of the choice of fundamental parameters. In practice, this corresponds to what is known as dimensional analysis; a function that relates physical observables must be independent of the choice of dimensional units. Notably, this condition comprises no direct measure of "structure" in the signal; the non-creation of new structure is only implicit in the sense that physically observable entities subject to scale changes should be treated in a self-similar manner.

In these scale-space formulations based on scale invariance, however, a further assumption is introduced concerning the semi-group structure. In sections 2.1–2.5, the scale parameter t associated with the semi-group (see equation (11)) was regarded as an *abstract ordering parameter* only. *A priori*, *i.e.* in the stage of formulating the axioms, there was no direct connection between this parameter and measurements of scale in terms of units of length. The only requirement was the qualitative (and essential) constraint that increasing values of the scale parameter should somehow correspond to representations at coarser scales. *A posteriori*, *i.e.* after deriving the shape of the convolution kernel, we could conclude that this parameter is related to scale as measured in units of length, *e.g. via* the standard deviation of the Gaussian kernel σ. The relationship turned out to be $t = \sigma^2/2$ (up to an unessential linear reparametrization of the scale parameter) and the semi-group operation to correspond to adding of σ-values in the Euclidean norm.

Restrictions from scale invariance: In this section, we shall assume that such a relationship exists already in the stage of formulating the axioms. Let σ be a scale parameter of dimension length associated with each layer in the scale-space representation, and introduce a monotonically increasing transformation function

$$t = \varphi(\sigma) \tag{27}$$

(with $\varphi(0) = 0$) such that the semi-group structure of the convolution kernel corresponds to mere adding of the scale values when measured in terms of t. For kernels parameterized by σ, the semi-group operation then assumes the form

$$h(\cdot;\ \sigma_1) * h(\cdot;\ \sigma_2) = h(\cdot;\ \varphi^{-1}(\varphi(\sigma_1) + \varphi(\sigma_2))). \tag{28}$$

The basic idea[3] is then to study the convolution operation in the Fourier domain

$$\hat{L}(\omega;\ \sigma) = \hat{h}(\omega;\ \sigma)\,\hat{f}(\omega). \tag{29}$$

where $\hat{h}$ represents the Fourier transform of the one-parameter family of smoothing kernels $h\colon \mathbb{Z}^N \times \mathbb{R}^+ \to \mathbb{R}$. Scale invariance implies that it must be possible to write this relation in terms of dimensionless variables. Here, we choose $\hat{L}/\hat{f} \in \mathbb{R}$ and $\omega\sigma \in \mathbb{R}^N$ and require the following relation to hold for some $\tilde{h}\colon \mathbb{R}^N \to \mathbb{R}$:

$$\frac{\hat{L}(\omega;\ \sigma)}{\hat{f}(\omega;\ \sigma)} = \hat{h}(\omega;\ \sigma) = \tilde{h}(\omega\sigma). \tag{30}$$

If in addition, this relation is to be independent of orientation, it follows that

$$\tilde{h}(\omega\sigma) = \hat{H}(|\omega\sigma|) \tag{31}$$

for some function $\hat{H}\colon \mathbb{R} \to \mathbb{R}$ with $\hat{H}(0) = 1$. The semi-group structure implies that $\hat{H}$ must obey

$$\hat{H}(|\omega\sigma_1|)\,\hat{H}(|\omega\sigma_2|) = \hat{H}(|\omega\,\varphi^{-1}(\varphi(\sigma_1) + \varphi(\sigma_2))|) \tag{32}$$

for all $\sigma_1, \sigma_2, \omega \in \mathbb{R}_+$, and it can be shown that φ must then be of the form

$$\varphi(\sigma) = C\,\sigma^p \tag{33}$$

for some arbitrary constants $C > 0$ and $p > 0$ (where we without loss of generality can take $C = 1$). Then, with $\tilde{H}(x^p) = \hat{H}(x)$, this constraint reduces to

$$\tilde{H}(|\omega\sigma_1|^p)\,\tilde{H}(|\omega\sigma_2|^p) = \tilde{H}(|\omega\sigma_1|^p + |\omega\sigma_2|^p), \tag{34}$$

which can be recognized as the definition of the exponential function ($\psi(\xi_1)\psi(\xi_2) = \psi(\xi_1 + \xi_2) \Rightarrow \psi(\xi) = a^{\xi}$ for some $a > 0$). In summary, for a scale-invariant rotationally symmetric semi-group, the Fourier transform must be of the form[4]

$$\hat{h}(\omega;\ \sigma) = \hat{H}(\omega\sigma) = \tilde{H}(|\omega\sigma|^p) = e^{-\alpha|\omega\sigma|^p} \tag{35}$$

for some $\alpha \in \mathbb{R}$. Requiring $\lim_{\sigma\to\infty} \hat{h}(\omega;\ \sigma) = 0$ gives $\alpha < 0$. Moreover, we can without loss of generality let $\alpha = -1/2$ to preserve consistency with the definition of the standard deviation of the Gaussian kernel σ in the case when $p = 2$.

[3] This derivation has been shortened substantially to save space. More detailed arguments showing how the assumption of scale invariance narrows down the class of smoothing kernels are presented in different forms in (Florack 1992; Pauwels *et al.* 1995; Lindeberg 1994).

[4] In a closely related work in chapter 7 in this volume, Nielsen (1996) arrives at filters of the same form from a slightly different starting point. He considers the problem of deriving optimal smoothing filters, and formulates an optimization problem in Euclidean norm. Thereby, the solution can be expressed as a linear filter, with the filter coefficients determined by a linear system of equations. Combined with shift-invariance, this gives rise to a convolution structure, and by requiring the filters to form a semi-group and to be associated with a scale parameter of dimension length raised to some power, it then follows that the filter must have a scale invariant Fourier transform which is additive under some self-similar reparametrization of the scale parameter. In other words, the Fourier transform must be of the form (35). In the spatial domain, this corresponds to regularization involving infinite orders of differentiation.

If on the other hand, the regularization functional is truncated at lower orders of differentiation, then a larger class of regularization filters is obtained, including the recursive filters studied by (Deriche 1987). It is interesting to note that these are also scale-space kernels in the sense that they are guaranteed to not increase the number of local extrema in a signal.

Additional conditions. Florack (1992) used separability in Cartesian coordinates as an additional basic constraint. Except in the one-dimensional case, this fixates h to be a Gaussian. Since, however, rotational symmetry combined with *separability per se* are sufficient to fixate the function to be a Gaussian, and the selection of orthogonal coordinate directions constitutes a very specific choice, it is illuminating to consider the effect of using other choices of p.

Pauwels *et al.* (1995) showed that the corresponding multi-scale representations generated by convolution kernels of the form (35) have *local infinitesimal generators* (basically meaning that the operator $\mathcal{A}$ in (21) 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 one that corresponds to a *non-negative convolution kernel* [5] (recall from section 2.3 that non-creation of local extrema implies that the kernel has to be non-negative).

Koenderink and van Doorn (1992) carried out a closely related study, where they showed that Gaussian derivative operators are natural operators to derive from a scale-space representation, given the assumption of scale invariance.

3 Semi-group and causality: Continuous domain

A main subject of this article is to extend the last two types of scale-space formulations in previous section. We shall first state explicitly how the scale-space formulation in section 2.5 applies to continuous signals. Then, as a corollary, it follows that for the scale-space formulation in section 2.6 only the specific choice $p = 2$ corresponds to a scale invariant semi-group satisfying causality requirements.

A main result we shall prove is that semi-group structure combined with the existence of a continuous scale parameter and non-enhancement of local extrema implies that the convolution kernel must be a Gaussian. This result will be obtained by an analogous way of reasoning as in a corresponding treatment for discrete signals (Lindeberg 1990, 1991, 1994). A certain number of technical modifications, however, have to be made due to the fact that the signals are continuous.

3.1 Assumptions

Given that the task is to state an axiomatic formulation of the first stages of visual processing, *the visual front-end*, a list of desired properties may be long:

> linearity, translational invariance, rotational symmetry, mirror symmetry, semi-group, causality, positivity, unimodality, continuity, differentiability, normalization, nice scaling behaviour, locality, rapidly decreasing for large x and t, existence of an infinitesimal generator, invariance with respect to certain grey-level transformations, etc.

Such a list will, however, contain redundancies, as does this one. Here, a (minimal) subset of these properties is taken as axioms. In fact, it can be shown that all the other above-mentioned properties follow from the subset we shall select.

[5] This follows directly from the well-known relation $\int_{x\in\mathbb{R}} x^n h(x)\,dx = (-i)^n \hat{h}^{(n)}(0)$ between moments in the spatial domain and derivatives in the frequency domain. It is straightforward to verify that the second moments of h are zero for any $p > 2$. Hence, the convolution kernel assumes both positive and negative values for all $p > 2$.

To begin, let us postulate that the scale-space representation should be generated by convolution with a one-parameter family of kernels such that $L(x;\ 0) = f(x)$ and for $t > 0$

$$L(x;\ t) = \int_{\xi \in \mathbb{R}^N} T(\xi;\ t)\, f(x - \xi). \tag{36}$$

This form of the smoothing formula corresponds to natural requirements about *linear shift-invariant smoothing* and the existence of a *continuous scale parameter*. Specifically, the assumption about linearity implies that all scale-space properties valid for the original signal will transfer to its derivatives. Hence, there is no commitment to certain aspects of image structure, such as the zero-order representation, or its first- or second-order derivatives. The assumption of shift-invariance reflects the desire to process all spatial points identically in the absence of further information, and the requirement about a continuous scale parameter makes it unnecessary to fixate any specific scale sampling in advance.

Initially, in the absence of any information, it is natural to require all coordinate directions to be handled identically. Therefore we assume that all kernels should be *rotationally symmetric*. Let us also impose a *semi-group* condition on the family T. This means that all scale levels will be treated similarly, that is, the smoothing operation should not depend on the scale value, and the transformation from a lower scale level to a higher scale level should always be given by convolution with a kernel from the family (see (12)).

As smoothing criterion the *non-enhancement* requirement for local extrema is taken (see section 2.5). It is convenient to express it as a sign condition on the derivative of the scale-space family with respect to the scale parameter. Hence, at any non-degenerate extremum point (extremum point where the determinant of the Hessian matrix is non-zero) we require the following conditions to hold

$$\begin{aligned} \partial_t L < 0 &\qquad \text{at a non-degenerate local maximum,} \\ \partial_t L > 0 &\qquad \text{at a non-degenerate local minimum.} \end{aligned} \tag{37}$$

In the one-dimensional case, this condition is equivalent to

$$\begin{aligned} \partial_t L < 0 &\qquad \text{if} \qquad L_{xx} < 0 \\ \partial_t L > 0 &\qquad \text{if} \qquad L_{xx} > 0 \end{aligned} \tag{38}$$

and since the Laplacian operator is negative (positive) at any non-degenerate local maximum (minimum) point, we can in the N-dimensional case require that

$$\operatorname{sign} \partial_t L = \operatorname{sign} \nabla^2 L \tag{39}$$

should hold at any extremum point where the Hessian matrix $\mathcal{H}L$ is either positive or negative definite. To ensure a proper statement of these conditions, where differentiability is guaranteed, we shall state a series of preliminary definitions, which will lead to the desired formulation.

3.2 Definitions

Let us summarize this (minimal) set of basic properties a family should satisfy to be a candidate family for generating a (rotationally symmetric) linear scale-space.

Definition 1 *(Pre-scale-space family of kernels)*
A one-parameter family of kernels $T\colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ *in* $\mathcal{L}_1$ *is said to be a (rotationally symmetric) pre-scale-space family of kernels if it satisfies*

- $T(\cdot;\ 0) = \delta(\cdot)$,
- *the semi-group property* $T(\cdot;\ s) * T(\cdot;\ t) = T(\cdot;\ s+t)$,
- *rotational symmetry*

$$T(x_1, x_2, \dots, x_N;\ t) = T(\sqrt{x_1^2 + x_2^2 + \dots + x_N^2}, 0, \dots, 0;\ t)$$

  for all $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ *and all* $t \in \mathbb{R}_+$, *and*
- *the continuity requirement* $\| T(\cdot;\ t) - \delta(\cdot) \|_1 \to 0$ *for any* $f \in \mathcal{L}_1$ *when* $t \downarrow 0$.

Definition 2 *(Pre-scale-space representation) Given a signal* $f\colon \mathbb{R}^N \to \mathbb{R}$, *let* $T\colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ *be a pre-scale-space family of kernels. Then, the one-parameter family of signals* $L\colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ *given by* (36) *is said to be the pre-scale-space representation of* f *generated by* T.

Provided that the input signal f is sufficiently regular, these conditions on T guarantee that L is differentiable with respect to the scale parameter and satisfies a system of linear evolution equations.

Lemma 3 *(A pre-scale-space representation is differentiable)*
Let $L\colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ *be the pre-scale-space representation of a signal* $f\colon \mathbb{R}^N \to \mathbb{R}$ *in* $\mathcal{L}_1$. *Then,* L *satisfies the evolution equation*

$$\partial_t L = \mathcal{A} L \tag{40}$$

for some linear and shift-invariant operator $\mathcal{A}$.

Proof: If f is sufficiently regular, *e.g.*, if $f \in L_1$, define a family of operators $\{\mathcal{T}_t, t > 0\}$, here from from L_1 to L_1, by $\mathcal{T}_t f = T(\cdot;\ t) * f$. Due to the conditions imposed on the kernels, the family satisfies the relation

$$\lim_{t \to t_0} \| (\mathcal{T}_t - \mathcal{T}_{t_0}) f \|_1 = \lim_{t \to t_0} \| (\mathcal{T}_{t-t_0} - \mathcal{I})(\mathcal{T}_{t_0} f) \|_1 = 0, \tag{41}$$

where $\mathcal{I}$ is the identity operator. Such a family is called a strongly continuous semi-group of operators (Hille and Phillips 1957: p. 58–59). A semi-group is often characterized by its *infinitesimal generator* $\mathcal{A}$ defined by

$$\mathcal{A} f = \lim_{h \downarrow 0} \frac{\mathcal{T}_h f - f}{h}. \tag{42}$$

The set of elements f for which $\mathcal{A}$ exists is denoted $\mathcal{D}(\mathcal{A})$. This set is not empty and never reduces to the zero element. Actually, it is even dense in L_1 (Hille and Phillips 1957: p. 307). If this operator exists then

$$\lim_{h \downarrow 0} \frac{L(\cdot, \cdot;\ t+h) - L(\cdot, \cdot;\ t)}{h} = \lim_{h \downarrow 0} \frac{\mathcal{T}_{t+h} f - \mathcal{T}_t f}{h}$$
$$= \lim_{h \downarrow 0} \frac{\mathcal{T}_h (\mathcal{T}_t f) - (\mathcal{T}_t f)}{h} = \mathcal{A}(\mathcal{T}_t f) = \mathcal{A} L(\cdot;\ t). \tag{43}$$

According to a theorem in (Hille and Phillips 1957: p. 308) strong continuity implies $\partial_t(\mathcal{T}_t f) = \mathcal{A}\mathcal{T}_t f = \mathcal{T}_t \mathcal{A} f$ for all $f \in \mathcal{D}(\mathcal{A})$. Hence, the scale-space family L must obey the differential equation $\partial_t L = \mathcal{A} L$ for some linear operator $\mathcal{A}$. Since L is generated from f by a convolution operation it follows that $\mathcal{A}$ must be shift-invariant. $\square$

This property makes it possible to formulate the previously indicated scale-space property in terms of derivatives of the scale-space representation with respect to the scale parameter—the grey-level value in a local maximum point must not increase with scale, whereas the grey-level value in every local minimum point must not decrease.

Definition 4 *(Pre-scale-space property: Non-enhancement of local extrema)*
A pre-scale-space representation $L\colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ of a smooth (infinitely continuously differentiable) signal is said to possess pre-scale-space properties, or equivalently not to enhance local extrema, if for every value of the scale parameter $t_0 \in \mathbb{R}_+$ it holds that if $x_0 \in \mathbb{R}^N$ is an extremum point for the mapping $x \mapsto L(x;\ t_0)$ at which the Hessian matrix is (positive or negative) definite, then the derivative of L with respect to t in this point satisfies

$$\operatorname{sign} \partial_t L = \operatorname{sign} \nabla^2 L. \tag{44}$$

Now we can state that a pre-scale-space family of kernels is a scale-space family of kernels if it satisfies this property for *any* input signal.

Definition 5 *(Scale-space family of kernels) A one-parameter family of pre-scale-space kernels $T\colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ is said to be a scale-space family of kernels if for any smooth (infinitely continuously differentiable) signal $f\colon \mathbb{R}^N \to \mathbb{R} \in L_1$ the pre-scale-space representation of f generated by T possesses pre-scale-space properties,* i.e., *if for any signal local extrema are never enhanced.*

Definition 6 *(Scale-space representation) A pre-scale-space representation $L\colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ of a signal $f\colon \mathbb{R}^N \to \mathbb{R}$ generated by a family of kernels $T\colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$, which are scale-space kernels, is said to be a scale-space representation of f.*

3.3 Necessity

We shall first show that these conditions by necessity imply that the scale-space family L satisfies the diffusion equation.

Theorem 7 *(Scale-space for continuous signals: Necessity)*
A scale-space representation $L\colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ of a signal $f\colon \mathbb{R}^N \to \mathbb{R}$ satisfies the differential equation

$$\partial_t L = \alpha \nabla^2 L \tag{45}$$

with initial condition $L(\cdot;\ 0) = f(\cdot)$ for some $\alpha > 0$.

Proof: The proof consists of two parts. The first part has already been presented in lemma 3, where it was shown that the requirements on pre-scale-space kernels imply that a pre-scale-space family obeys a linear evolution equation where the infinitesimal generator is shift-invariant. In the second part, counterexamples will be constructed from various simple test functions to delimit the class of possible operators.

B.1. The extremum point condition (39) combined with definitions 5-6 means that $\mathcal{A}$ must be a pure differential operator. This can be easily understood by studying the following class of counterexamples: Consider a smooth (infinitely continuously differentiable) function $f \colon \mathbb{R}^N \to \mathbb{R}$ such that f has a maximum point at the origin at which the Hessian matrix is negative definite and for some $\varepsilon > 0$ $f(x) = 0$ when $|x| \in [\frac{\varepsilon}{2}, \varepsilon]$. Split this function into two components

$$f = f_I + f_E \tag{46}$$

where

$$\begin{aligned} \nabla f_I(0) &= 0, \\ \nabla^2 f_I(0) &< 0, \\ f_I = 0 \qquad \text{when} \quad |x| &\geq \varepsilon/2, \\ f_E = 0 \qquad \text{when} \quad |x| &\leq \varepsilon. \end{aligned} \tag{47}$$

Assume first that $f_E = 0$. Then, evaluation of $\partial_t L = \mathcal{A} L$ at $t = 0$ gives $L(\cdot;\ 0) = f$ and $\partial_t f = \mathcal{A}(f_I + f_E) = \mathcal{A} f_I$. Hence, at $(x,;\ t) = (0;\ 0)$, we must require $\mathcal{A} f_I = C_1 < 0$. Fixate these $\mathcal{A}$ and f_I and consider then any f_E for which $\mathcal{A} f_E = C_2 \neq 0$. Then, for $f = f_I + \beta_1 f_E$ it holds that

$$\partial_t L = \mathcal{A} f_I + \beta_1 \mathcal{A} f_E = C_1 + \beta_1 C_2. \tag{48}$$

Obviously, the sign of this expression can be made positive and (39) be violated by a suitable choice of β_1. Hence, for any $\varepsilon > 0$ we have that $\mathcal{A} f_E$ must be identically zero for all functions that assume non-zero values outside the region $|x| < \varepsilon$. In other words, $\mathcal{A}$ must be a *local operator* and $\mathcal{A} f$ can only exploit information from f at the central point. Thus, for any smooth function, $\mathcal{A} f$ must be of the form

$$\mathcal{A} f = \sum_{\xi \in \mathbb{Z}_+^N} a_\xi L_{x^\xi} \tag{49}$$

where $\xi = (\xi_1, \xi_2, \ldots, \xi_N)$ is a multi-index, $a_\xi \in \mathbb{R} \forall \xi$, and $L_{x^\xi} = L_{x_1^{\xi_1} x_2^{\xi_2} \ldots x_N^{\xi_N}}$.

B.2. The extremum point condition (39) also means that $\mathcal{A} L$ must not contain any term proportional to L or derivatives of order higher than two. This can be seen by considering a test signal of the form

$$f(x) = x_1^2 + x_2^2 + \cdots + x_N^2 + \beta_2 x^\eta, \tag{50}$$

for some $\eta = (\eta_1, \eta_2, \ldots, \eta_N) \in \mathbb{Z}^N$ with $|\eta| = |\eta_1| + |\eta_2| + \cdots + |\eta_N| > 2$. If $a_\xi \neq 0$ for some $\xi \in \mathbb{Z}^N$ it is clear that a suitable choice of β_2 can make the sign of $\mathcal{A} f$ arbitrary and hence violate (39). Similarly, by considering a test signal of the form

$$f(x) = x_1^2 + x_2^2 + x_N^2 + \beta_3 \tag{51}$$

it follows that $a_0 = 0$. Thus, $\mathcal{A}$ can only contain derivatives of order one and two.

B.3. Since the scale-space kernels are required to be rotationally symmetric, if follows that the contributions from the first-order derivatives as well as the mixed second-order derivatives must be zero. Moreover, the contributions from the second-order derivatives along all coordinate directions must be similar. Hence, the only possibility is that $\mathcal{A}$ is a constant times the Laplacian operator. □

If the requirements about rotational symmetry are relaxed, then $\mathcal{A}$ will be a linear combination of first- and second-order derivatives, for which the coefficients of the second-order derivative terms correspond to a positive definite quadratic form and the coefficients of the first-order terms are arbitrary.

3.4 Sufficiency

The reverse statement of theorem 7, *i.e.* the fact that L generated by (45) satisfies definition 4, is obvious, since at extremum points where $\mathcal{H}L$ is positive or negative definite, we have $\nabla^2 L < 0$ at maxima and $\nabla^2 L > 0$ at minima.

3.5 Application to scale invariant semi-groups

By combining theorem 7 with the treatment in section 2.6, it follows that the Gaussian kernel is the only rotationally symmetric scale invariant semi-group that satisfies the causality requirement. Thus, also from this point of view, selecting $p = 2$ in (35) constitutes a very special choice.

Besides the fact that the additional requirements concerning locality of the infinitesimal generator and positivity of the convolution operator can be replaced by causality, it is worth noting that causality and semi-group structure *per se* imply that the infinitesimal generator must be local and the convolution kernel must be non-negative. More importantly, if we assume a semi-group structure and combine it with causality and rotational invariance, then scale invariance arises as a consequence and is not required as an axiom.

4 Summary and conclusions

We have seen how a scale-space formulation for continuous signals can be stated based on the essential assumptions about a semi-group and a causality requirement expressed as non-enhancement of local extrema. Combined with the assumption about convolution operations and a certain regularity assumptions along the scale direction (strong continuity; *i.e.* continuity in norm) the semi-group structure implies that the multi-scale representation is differentiable along the scale direction and has an infinitesimal generator. The causality requirement then, in turn, implies that the infinitesimal generator must be local (correspond to a differential operator) and be a linear combination of derivatives of orders one and two. The essence of these results is that the scale-space representation on a spatial domain is given by a (possibly semi-discretized) parabolic differential equation corresponding to a *second-order* differential operator with respect to the spatial coordinates, and a *first-order* differential operator with respect to the scale parameter.

Rotational symmetry implies that no first-order derivatives are allowed and that the second-order derivatives must occur in a combination such that the differential operator is a constant times the Laplacian. In this case, the scale-space family is generated by convolution with rotationally symmetric Gaussian kernels.

It has also been described how the causality requirement relates to scale-space formulations based on semi-group structure combined with scale invariance. Unless additional conditions are imposed, these assumptions give rise to a one-parameter family of smoothing kernels, and scale invariance *per se* does not uniquely single out the Gaussian kernel. Previously, it has been shown that the Gaussian arises as a unique choice if the scale-space family is required to have a local infinitesimal generator and the smoothing kernel is required to be positive. As a corollary, it follows that the Gaussian is a unique choice if the additional assumptions about locality and positivity are replaced by a causality assumption. In fact, causality implies locality and positivity. More importantly, when combined with the semi-group structure, the causality assumption gives rise to scale invariance.

Concerning scale-space representation for discrete signals, scale invariance obviously cannot be used if the discrete signal constitutes the only available data.[6] A perfectly scale invariant operator cannot be expressed on a discrete grid, which has a certain preferred scale given by the distance between adjacent grid points. The formulation based on non-enhancement/causality applies in both domains, provided that the definition of local maximum and the requirement about rotational symmetry are appropriately modified.

5 Extensions of linear scale-space

A natural question then arises: Does this approach constitute the *only* reasonable way to perform the low-level processing in a vision system, and are the Gaussian kernels and their derivatives the only smoothing kernels that can be used? Of course, this question is impossible to answer to without further specification of the purpose of the representation, and what tasks the visual system has to accomplish. In any sufficiently specific application it should be possible to design a smoothing filter that in some sense has a "better performance" than the Gaussian derivative model. For example, it is well-known that scale-space smoothing leads to shape distortions at edges by smoothing across object boundaries. Distortions arise also in algorithms for estimating local surface shape, such as shape-from-texture and shape-from-disparities. Hence, it should be emphasized that the linear scale-space model is rather aimed at describing the principles of the very first stages of low-level processing in an *uncommitted* visual system aimed at handling a large class of different situations, and for which no or very little a priori information is available.

Then, once initial hypotheses about the structure of the world have been generated

[6]If further information is available about the image formation process, *e.g.*, such that the continuous signal can be reconstructed exactly from the sampled data, then the discrete signal can be treated as equivalent to the original continuous signal, and an equivalent discrete scale-space model be expressed for the continuous scale-space representation of the reconstructed continuous signal. Chapter 9 by Åström and Heyden (1996) in this volume exploits this idea based on the sampling theorem and the assumption of an ideally sampled band limited signal.

within this framework, the intention is that it should be possible to invoke more refined processing, which can compensate for these effects and adapt to the current situation and the task at hand. These are the motivations for studying non-uniform scale-space concepts, such as affine Gaussian scale-space and non-linear diffusion techniques. From the viewpoint of such approaches, the linear scale-space model is intended to serve as a natural starting point.

5.1 Relaxing rotational symmetry

Affine Gaussian scale-space. A straightforward extension of the raw linear scale-space representation can be obtained by relaxing the requirement about rotational symmetry in definition 1. Then, the same way of reasoning as in sections 3.2–3.4 still applies. The only essential differences are that part B.3 in the proof of theorem 7 should be omitted and that the Laplacian operator in equation (45) should be replaced by an arbitrary linear and symmetric (elliptic) second-order differential operator.

In terms of convolution operations, the resulting (three-parameter) *affine Gaussian scale-space representation* is generated by non-uniform Gaussians defined by

$$g(x;\ \Sigma_t) = \frac{1}{(2\pi)^{D/2}\sqrt{\det \Sigma_t}} e^{-x^T \Sigma_t^{-1} x/2} \tag{52}$$

where Σ_t is a symmetric positive definite (covariance) matrix. If the covariance matrix is written $\Sigma_t = t\Sigma_0$ for some (constant) matrix Σ_0, then the shape-adapted multi-scale representation satisfies the diffusion equation

$$\partial_t = \frac{1}{2}\nabla^T \left(\Sigma_0 \nabla L\right). \tag{53}$$

This representation, considered in (Lindeberg 1994), satisfies all the scale-space properties listed in sections 2–3 except those specifically connected to rotational symmetry. For example, because of the linearity of this operation, all scale-space properties transfer to spatial derivatives of the scale-space representation as well as to linear combinations of these.[7] [8]

Actually, the affine Gaussian scale-space representation of any signal f is equivalent to the result of applying the rotationally symmetric linear scale-space concept to a transformed image. More specifically, it can be constructed by composing the following operations in cascade: (i) subject the original signal f to an affine transformation, (ii) apply the linear scale-space representation to the deformed signal,

[7] For the corresponding transformation kernels from the input signal f to a linear combination of scale-space derivatives, however, the semi-group structure is replaced by a cascade smoothing property. This means that any transformation kernel $h(\cdot;\ t)$ corresponds to the result of convolving some fixed kernel h_0 with a Gaussian kernel, *i.e.*, $h(\cdot\ \ t) = h_0 * g(\cdot;\ t)$. Hence, these kernels satisfy $h(\cdot\ \ s+t) = g(\cdot;\ s) * h(\cdot;\ t)$ where g denotes the Gaussian kernel.

[8] Moreover, as shown in chapter 12 by Griffin (1996), the classification of what scale-space singularities (bifurcations) can occur with increasing scale, transfers from the rotationally symmetric Gaussian scale-space to the affine Gaussian scale-space (see (Koenderink and van Doorn 1986; Lindeberg 1992, 1994) as well as chapter 11 by Damon (1996), chapter 10 by Johansen (1996) chapter 13 by Kalitzin (1996). Examples of image representations depending on this *deep structure of scale-space* can be found in (Lindeberg 1993) and in chapter 14 by Olsen (1996).

(iii) subject the smoothed signal to the inverse affine transformation.[9] By varying the affine transformations, we can in this way span the family of affine Gaussian scale-space representations.

A major motivation for considering the affine Gaussian scale-space representation is that it is *closed* under affine transformations. For two patterns f_L and f_R related by an invertible linear transformation $\eta = B\xi$

$$f_L(\xi) = f_R(B\xi) \tag{54}$$

the corresponding affine Gaussian scale-space representations

$$L(\cdot;\ \Sigma_L) = g(\cdot;\ \Sigma_L) * f_L(\cdot), \tag{55}$$

$$R(\cdot;\ \Sigma_R) = g(\cdot;\ \Sigma_R) * f_R(\cdot), \tag{56}$$

are related by

$$L(\xi;\ \Sigma_L) = R(\eta;\ \Sigma_R), \tag{57}$$

where

$$\Sigma_R = B \Sigma_L B^T. \tag{58}$$

Hence, for any matrix Σ_L there exists a matrix Σ_R such that the affine scale-space representations of f_L and f_R are equal (see the commutative diagram in figure 1), and any non-singular affine transformations can be captured exactly within this three-parameter multi-scale representation.[10] This extension of the linear scale-space concept is useful, for example when estimating local surface orientation from local deformations of surface patterns (Lindeberg and Gårding 1993; Gårding and Lindeberg 1994, 1996) and more generally, whenever computing affine image deformations, such as in first-order optic flow and first-order stereo matching. Chapter 5 by Florack (1996) in this volume exploits this idea further, by emphasizing the equivalence between deformations of images and filters.

$$\begin{array}{ccc}
L(\xi;\ \Sigma_L) & \text{—}\ \ \eta = B\xi\ \ \rightarrow & R(\eta;\ B\Sigma_L B^T) \\
\uparrow & & \uparrow \\
*g(\cdot;\ \Sigma_L) & & *g(\cdot;\ B\Sigma_L B^T) \\
| & & | \\
f_L(\xi) & \text{—}\ \ \eta = B\xi\ \ \rightarrow & f_R(\eta)
\end{array}$$

Figure 1: Commutative diagram of the non-uniform scale-space representation under linear transformations of the spatial coordinates in the original image.

[9]This essentially corresponds to the duality between transformations of image operators and image domains described in detail in chapter 5 by Florack (1996) in this volume.

[10]Compared to the affine invariant level curve evolution scheme proposed by (Alvarez *et al.* (1993) and Sapiro and Tannenbaum (1993), given as a one-parameter solution to a non-linear differential equation (see equation (66) in section 5.2), an obvious disadvantage of the affine Gaussian scale-space is that it gives rise to a three-parameter variation. The advantage is that commutative properties can still be preserved within a family of linear transformations.

Discrete affine Gaussian scale-space. The discrete counterpart of the affine Gaussian scale-space is obtained by relaxing the symmetry requirements of the infinitesimal generator in (23), while preserving the locality, positivity and zero sum constraints. In addition, to avoid spatial shifts, the Fourier transform should be required to be real, implying $a_{i,j} = a_{-i,-j}$. This gives rise to the *discrete affine Gaussian scale-space* generated by an infinitesimal generator having a computational molecule of the form

$$\mathcal{A} = \begin{pmatrix} -C/2 & B & C/2 \\ A & -2A-2B & A \\ C/2 & B & -C/2 \end{pmatrix} + \alpha \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}. \tag{59}$$

This representation can be interpreted as a second-order discretization of the diffusion equation (53) associated with the continuous affine Gaussian scale-space

$$\partial_t L = A L_{xx} + B L_{yy} + 2C L_{xy} \tag{60}$$

where $A > 0$ and $AB - C^2 > 0$ are necessary requirements for the operator to be elliptic. The free parameter α (which controls the addition of a discretization of the mixed fourth-order derivative L_{xxyy}) should be in the interval $C/2 \leq \alpha \leq \min(A, B)/2$ to ensure that all non-central coefficients are non-negative.

As a general tool, the affine Gaussian scale-space constitutes a useful framework for adapting the shape of the smoothing kernels in situations where further information is available, and the process of performing a one-parameter variation of the shapes of the smoothing kernels to the image data is referred to as *shape adaptation* (Lindeberg 1994; Lindeberg and Gårding 1994). Alternatively, it can be seen as a way of obtaining invariance to affine transformations by expanding the data into a three-parameter scale-space representation instead of the more common one-parameter approach.

In chapter 2 by Almansa and Lindeberg (1996), this concept is used for expressing an *adaptation of linear smoothing operations to image data*, guided by a non-linear image descriptor (a second moment matrix) and a non-linear scale selection algorithm. Non-linear smoothing approaches with a large number of structural similarities are considered in chapter **??**6 by Weickert (1996). An extension of the affine Gaussian scale-space concept to a more general *spatio-temporal scale-space representation*, allowing for *velocity adaptation*, is presented in (Lindeberg 1996).

5.2 Non-linear smoothing

Let us finally discuss how the scale-space formulation based on non-enhancement of local extrema applies to a number of non-linear scale-space approaches. Clearly, any evolution scheme of the form

$$\partial_t L = \nabla^T (c(x;\ t) \nabla L) \tag{61}$$

satisfies the non-enhancement/causality requirement as long as the conduction coefficient $c(x;\ t)$ is non-negative. (Since $\partial_t L = \nabla^T c(x;\ t)\, \nabla L + c(x;\ t)\, \nabla^2 L = c(x;\ t) \nabla^2 L$ at local extrema it follows that $\partial_t L$ has the same sign as $\nabla^2 L$.) In contrast to linear (and affine) scale-space representation, however, this property does not necessarily extend to spatial derivatives of the scale-space representation.

Anisotropic diffusion. For the edge enhancing anisotropic diffusion scheme proposed by Perona and Malik (1990), where

$$c(x;\ t) = h(|\nabla L(x;\ t)|) \tag{62}$$

for some function $h\colon \mathbb{R}_+ \to \mathbb{R}_+$, the causality violation in the first-order derivative can be illustrated as follows: Consider, for simplicity, the one-dimensional case

$$\partial_t L = \partial_x (h(|L_x|) L_x) = h_x(|L_x|) L_x^2 + h(|L_x|) L_{xx} \tag{63}$$

and introduce $\phi(L_x) = h(|L_x|) L_x$. Then, following Whitaker and Pizer (1993), the evolution equation can be written $\partial_t L = \partial_x(\phi(L_x)) = \phi'(L_x) L_{xx}$ and the gradient magnitude satisfies

$$\partial_t L_x = \phi''(L_x) L_{xx}^2 + \phi'(L_x) L_{xxx}. \tag{64}$$

For a local gradient maximum with $L_{xx} = 0$ and $L_{xxx} < 0$ it holds that $\partial_t L_x > 0$ if $\phi'(L_x) < 0$. For the conductivity function used by Perona and Malik (1990)

$$h(|\nabla L|) = e^{-|\nabla L|^2/k^2}, \tag{65}$$

where k is a free parameter, we have $\phi'(L_x) < 0$ if $L_x > k/\sqrt{2}$. In other words, gradients that are sufficiently strong will be enhanced, and gradients that are sufficiently weak will be suppressed (no matter what are the spatial extents of the image structures). Moreover, this evolution equation depends upon an external parameter and is not scale invariant.

Affine invariant scale-space. Recently, Alvarez *et al.* (1993) and Sapiro and Tannenbaum (1993) have presented an affine invariant (one-parameter) scale-space representation which essentially commutes with affine transformations. It is generated by the evolution equation

$$\partial_t L = (|\nabla L|^3\, \kappa(L))^{1/3} \tag{66}$$

where $\kappa(K) = (L_x^2 L_{yy} + L_y^2 L_{xx} - 2 L_x L_y L_{xy})/(L_x^2 + L_y^2)^{3/2}$ is the curvature of a level curve of the intensity distribution. In Alvarez *et al.* (1992), it is shown that this representation is essentially unique if semi-group structure and existence of a local infinitesimal generator are combined with invariance under monotone intensity transformations as well as affine transformations of the spatial domain.[11] From the affine invariance, it follows that the evolution equation is scale invariant. Regarding causality, however, $\partial_t L$ is exactly zero at any local extremum.

Concerning other non-linear scale-space approaches, the reader is referred to (Nordström 1990 1990; Nitzberg and Shiota 1992; Florack *et al.* 1993; Whitaker and Pizer 1993), the book edited by ter Haar Romeny (1994), as well as chapter 15 by van den Boomgaard (1996) and chapter 16 by Weickert (1996) in this volume.

Acknowledgments

The support from the Swedish Research Council for Engineering Sciences, TFR, is gratefully acknowledged. Earlier versions of this manuscript have been presented in in (Lindeberg 1994) and (Lindeberg 1997).

[11] Note the difference in terminology in Alvarez *et al.* (1993), where the semi-group structure is referred to as causality.

References

Abramowitz, M. & Stegun, I. A., eds (1964), *Handbook of Mathematical Functions*, Applied Mathematics Series, 55 edn, National Bureau of Standards.

Alvarez, L., Guichard, F., Lions, P.-L. & Morel, J.-M. (1993), 'Axioms and fundamental equations of image processing', *Arch. for Rational Mechanics* **123**(3), 199–257.

Babaud, J., Witkin, A. P., Baudin, M. & Duda, R. O. (1986), 'Uniqueness of the Gaussian kernel for scale-space filtering', *IEEE Trans. Pattern Analysis and Machine Intell.* **8**(1), 26–33.

Burt, P. J. (1981), 'Fast filter transforms for image processing', *Computer Vision, Graphics, and Image Processing* **16**, 20–51.

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

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

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

Florack, L. M. J., ter Haar Romeny, B. M., Koenderink, J. J. & Viergever, M. A. (1992*a*), Families of tuned scale-space kernels, *in* 'Proc. European Conf. on Computer Vision', Santa Margherita Ligure, Italy, pp. 19–23.

Florack, L. M. J., ter Haar Romeny, B. M., Koenderink, J. J. & Viergever, M. A. (1992*b*), 'Scale and the differential structure of images', *Image and Vision Computing* **10**(6), 376–388.

Florack, L. M. J., ter Haar Romeny, B. M., Koenderink, J. J. & Viergever, M. A. (1993*a*), 'Cartesian differential invariants in scale-space', *J. of Mathematical Imaging and Vision* **3**(4), 327–348.

Florack, L. M. J., ter Haar Romeny, B. M., Koenderink, J. J. & Viergever, M. A. (1993*b*), Non-linear diffusion by metrical affinity. In preparation.

Florack, L. M. J., ter Haar Romeny, B. M., Koenderink, J. J. & Viergever, M. A. (1994*a*), 'General intensity transformations and differential invariants', *J. of Mathematical Imaging and Vision* **4**, 171–187.

Florack, L. M. J., ter Haar Romeny, B. M., Koenderink, J. J. & Viergever, M. A. (1994*b*), 'Linear scale-space', *J. of Mathematical Imaging and Vision* **4**(4), 325–351.

Gårding, J. & Lindeberg, T. (1994), Direct estimation of local surface shape in a fixating binocular vision system, *in* J.-O. Eklundh, ed., 'Proc. 3rd European Conf. on Computer Vision', Vol. 800 of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden, pp. 365–376.

Gårding, J. & Lindeberg, T. (1996), 'Direct computation of shape cues using scale-adapted spatial derivative operators', *Int. J. of Computer Vision* **17**(2), 163–191.

Hille, E. & Phillips, R. S. (1957), *Functional Analysis and Semi-Groups*, Vol. XXXI, American Mathematical Society Colloquium Publications.

Hirschmann, I. I. & Widder, D. V. (1955), *The Convolution Transform*, Princeton University Press, Princeton, New Jersey.

Hummel, R. A. (1986), Representations based on zero crossings in scale space, *in* 'Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition', pp. 204–209.

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

Johansen, P. (1994), 'On the classification of toppoints in scale space', *J. of Mathematical Imaging and Vision* **4**, 57–67.

Karlin, S. (1968), *Total Positivity*, Stanford Univ. Press.

Koenderink, J. J. (1984), 'The structure of images', *Biological Cybernetics* **50**, 363–370.

Koenderink, J. J. (1988), 'Scale-time', *Biological Cybernetics* **58**, 159–162.

Koenderink, J. J. & van Doorn, A. J. (1986), 'Dynamic shape', *Biological Cybernetics* **53**, 383–396.

Koenderink, J. J. & van Doorn, A. J. (1987), 'Representation of local geometry in the visual system', *Biological Cybernetics* **55**, 367–375.

Koenderink, J. J. & van Doorn, A. J. (1992), 'Generic neighborhood operators', *IEEE Trans. Pattern Analysis and Machine Intell.* **14**(6), 597–605.

Lifshitz, L. & Pizer, S. (1990), 'A multiresolution hierarchical approach to image segmentation based on intensity extrema', *IEEE Trans. Pattern Analysis and Machine Intell.* **12**(6), 529–541.

Lindeberg, T. (1990), 'Scale-space for discrete signals', *IEEE Trans. Pattern Analysis and Machine Intell.* **12**(3), 234–254.

Lindeberg, T. (1991), Discrete Scale-Space Theory and the Scale-Space Primal Sketch, Doctoral dissertation,, Dept. of Numerical Analysis and Computing Science, KTH, Stockholm, Sweden. 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.

Lindeberg, T. (1992), 'Scale-space behaviour of local extrema and blobs', *J. of Mathematical Imaging and Vision* **1**(1), 65–99.

Lindeberg, T. (1993*a*), 'Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention', *Int. J. of Computer Vision* **11**(3), 283–318.

Lindeberg, T. (1993*b*), 'Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction', *J. of Mathematical Imaging and Vision* **3**(4), 349–376.

Lindeberg, T. (1994*a*), On the axiomatic foundations of linear scale-space: Combining semi-group structure with causality vs. scale invariance, report, ISRN KTH/NA/P--94/20--SE, Dept. of Numerical Analysis and Computing Science, KTH, Stockholm, Sweden. 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.

Lindeberg, T. (1994*b*), Scale-space for N-dimensional discrete signals, *in* Y. O. Ying, A. Toet & H. Heijmanns, eds, 'Shape in Picture: Proc. NATO workshop on Shape in Picture', Vol. 126 of *NATO ASI Series F*, Springer Verlag, New York, pp. 571–590. (Also available in Tech. Rep. ISRN KTH/NA/P–92/26–SE from Royal Inst of Technology).

Lindeberg, T. (1994*c*), 'Scale-space theory: A basic tool for analysing structures at different scales', *Journal of Applied Statistics* **21**(2), 225–270. Supplement *Advances in Applied Statistics: Statistics and Images: 2*.

Lindeberg, T. (1994*d*), *Scale-Space Theory in Computer Vision*, The Kluwer International Series in Engineering and Computer Science, Kluwer Academic Publishers, Dordrecht, Netherlands.

Lindeberg, T. (1996*a*), Linear spatio-temporal scale-space, Technical report, Dept. of Numerical Analysis and Computing Science, KTH.

Lindeberg, T. (1996*b*), On the axiomatic foundations of linear scale-space, *in* J. Sporring, M. Nielsen, L. Florack & P. Johansen, eds, 'Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory', Kluwer Academic Publishers, Copenhagen, Denmark.

Lindeberg, T. (1996*c*), Scale-space theory: A framework for handling image structures at multiple scales, *in* 'Proc. CERN School of Computing', Egmond aan Zee, The Netherlands, pp. 27–38. Tech. Rep. CERN 96-08.

Lindeberg, T. & Fagerström, D. (1996), Scale-space with causal time direction, *in* 'Proc. 4th European Conf. on Computer Vision', Vol. 1064, Springer Verlag, Berlin, Cambridge, UK, pp. 229–240.

Lindeberg, T. & Gårding, J. (1993), Shape from texture from a multi-scale perspective, *in* H.-H. N. et. al., ed., 'Proc. 4th Int. Conf. on Computer Vision', IEEE Computer Society Press, Berlin, Germany, pp. 683–691.

Lindeberg, T. & Gårding, J. (1994), Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D structure, *in* J.-O. Eklundh, ed., 'Proc. 3rd European Conf. on Computer Vision', Vol. 800 of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden, pp. 389–400.

Lindeberg, T. & ter Haar Romeny, B. (1994), Linear scale-space I: Basic theory, *in* B. ter Haar Romeny, ed., 'Geometry-Driven Diffusion in Computer Vision', Series in Mathematical Imaging and Vision, Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 1–41.

Nitzberg, M. & Shiota, T. (1992), 'Non-linear image filtering with edge and corner enhancement', *IEEE Trans. Pattern Analysis and Machine Intell.* **14**(8), 826–833.

Nordström, N. (1990), 'Biased anisotropic diffusion: A unified regularization and diffusion approach to edge detection', *Image and Vision Computing* **8**(4), 318–327.

Pauwels, E. J., Fiddelaers, P., Moons, T. & van Gool, L. J. (1995), 'An extended class of scale-invariant and recursive scale-space filters', *IEEE Trans. Pattern Analysis and Machine Intell.* **17**(7), 691–701.

Perona, P. & Malik, J. (1990), 'Scale-space and edge detection using anisotropic diffusion', *IEEE Trans. Pattern Analysis and Machine Intell.* **12**(7), 629–639.

Sapiro, G. & Tannenbaum, A. (1993), 'Affine invariant scale-space', *Int. J. of Computer Vision* **11**(1), 25–44.

Schoenberg, I. J. (1950), 'On Pòlya frequency functions. ii. Variation-diminishing integral operators of the convolution type', *Acta Sci. Math. (Szeged)* **12**, 97–106.

Schoenberg, I. J. (1953), 'On smoothing operations and their generating functions', *Bull. Amer. Math. Soc.* **59**, 199–230.

ter Haar Romeny, B., ed. (1994), *Geometry-Driven Diffusion in Computer Vision*, Series in Mathematical Imaging and Vision, Kluwer Academic Publishers, Dordrecht, Netherlands.

Whitaker, R. T. & Pizer, S. M. (1993), 'A multi-scale approach to nonuniform diffusion', *Computer Vision, Graphics, and Image Processing* **57**(1).

Witkin, A. P. (1983), Scale-space filtering, *in* 'Proc. 8th Int. Joint Conf. Art. Intell.', Karlsruhe, West Germany, pp. 1019–1022.

Yuille, A. L. (1988), The creation of structure in dynamic shape, *in* 'Proc. 2nd Int. Conf. on Computer Vision', Tampa, Florida, pp. 685–689.

Yuille, A. L. & Poggio, T. (1985), 'Fingerprint theorems for zero crossings', *J. of the Optical Society of America (A)* **2**, 683–692.

Yuille, A. L. & Poggio, T. A. (1986), 'Scaling theorems for zero-crossings', *IEEE Trans. Pattern Analysis and Machine Intell.* **8**, 15–25.