Scale-Space Theory: A Basic Tool for Analysing Structures at Different Scales

1 author:

Tony Lindeberg
KTH Royal Institute of Technology
156 PUBLICATIONS **13,963** CITATIONS

Some of the authors of this publication are also working on these related projects:

Neurogenerator View project

Time-Causal and Time-Recursive Spatio-Temporal Receptive Fields for Video Analysis View project

Scale-space theory: A basic tool for analysing structures at different scales

Tony Lindeberg

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

Abstract

An inherent property of objects in the world is that they only exist as meaningful entities over certain ranges of scale. If one aims at describing the structure of unknown real-world signals, then a multi-scale representation of data is of crucial importance.

This chapter gives a tutorial review of a special type of multi-scale representation, *linear scale-space representation*, which has been developed by the computer vision community in order to handle image structures at different scales in a consistent manner. The basic idea is to embed the original signal into a one-parameter family of gradually smoothed signals, in which the fine scale details are successively suppressed.

Under rather general conditions on the type of computations that are to performed at the first stages of visual processing, in what can be termed *the visual front end*, it can be shown that the Gaussian kernel and its derivatives are singled out as the only possible smoothing kernels. The conditions that specify the Gaussian kernel are, basically, linearity and shift-invariance combined with different ways of formalizing the notion that structures at coarse scales should correspond to simplifications of corresponding structures at fine scales — they should not be accidental phenomena created by the smoothing method. Notably, several different ways of choosing scale-space axioms give rise to the same conclusion.

The output from the scale-space representation can be used for 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.

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*, *wavelets* and *multi-grid methods*. 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 related fields.

An interesting similarity with biological vision is that the scale-space operators closely resemble receptive field profiles registered in neurophysiological studies of the mammalian retina and visual cortex.

1 Introduction

We perceive objects in the world as meaningful entities only 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 the kilometer level. At those scales it is more relevant to talk about the molecules that form the leaves of the tree, or the forest in which the tree grows. Similarly, it is only meaningful to talk about a cloud over a certain range of coarse scales. At finer scales it is more appropriate to consider the individual droplets, which in turn consist of water molecules, which consist of atoms, which consist of protons and electrons etc.

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 scale concept and the notion of multi-scale representation are of crucial importance. These general needs are well-understood, for example, in cartography; maps are produced at different degrees of abstraction. A map of the world contains the largest countries and islands, and possibly, some of the major cities, whereas towns and smaller islands appear at first in a map of a country. In a city guide, the level of abstraction is changed considerably to include streets and buildings, etc. An atlas can be seen as a symbolic multi-scale representation of the world around us, constructed manually and with very specific purposes in mind.

In physics, phenomena are modelled at several levels of scales, 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, a physical description may depend strongly upon the scale at which the world is modelled. This is in clear contrast to certain idealized mathematical entities, such as 'point' or 'line', which appear in the same way independent of the scale of observation.

Specifically, the need for multi-scale representation arises when to design methods for automatically analysing and deriving information from signals that are the results of real-world measurements. It is clear that to extract any type of information from data it is necessary to interact with it using certain operators. The type of information that can be obtained is to a large extent 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 signal processing concern *what* type of 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 responses can be very hard.

In certain controlled situations, appropriate scales for analysis may be known *a priori*. For example, a characteristic property of a physicist is the intuitive ability to select proper scales for modelling a problem. Under other circumstances, for example, in applications dealing with automated signal processing, however, it may not be obvious at all how to determine in advance what are the proper scales. One such example is a vision system with the task to analyse 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 describing structures at multiple scales.

The goal with this presentation is to review some fundamental results concerning a theory for multi-scale representation, called *scale-space theory*. It is a general framework that has been developed by the computer vision community for representing image data and the multi-scale nature of it at the very earliest stages in the chain of visual processing that is aims at understanding (perception). We shall deal with the operations that are performed directly on raw image data by the processing modules termed the *visual front-end*.

Although this treatment, from now on, will be mainly concerned with the analysis of visual data, the general notion of scale-space representation is of much wider applicability and arises in several contexts where measured data are to be analyzed and interpreted by automatic methods.

1.1 Scale-space theory for computer vision

Vision deals with the problem of deriving meaningful and useful information about the world from the light reflected from it. What should be meant by "meaningful and useful information" is, of course, strongly dependent on the goal of the analysis, that is, the underlying *purpose* why we want to make use of visual information and process it with automatic methods. One reason may be that of machine vision— the desire to provide machines and robots with visual abilities. Typical tasks to be solved are object recognition, object manipulation, and visually guided navigation. Other common applications of techniques from computer vision can be found in image processing, where one can mention image enhancement, visualization and analysis of medical data, as well as industrial inspection, remote sensing, automated cartography, data compression, and the design of visual aids, etc.

A subject of utmost importance in automated reasoning is the notion of internal representation. Only by representation can information be captured and made available to decision processes. The purpose of a representation is to make certain aspects of the information content *explicit*, that is, immediately accessible without any need for additional processing. This article deals with the basic aspects of early image representation—how to register the light that reaches the retina, and how to make explicit important aspects of it that are useful for later stages processes. This is the processing performed in the visual front-end. If the operations are to be local, then they have to preserve the topology at the retina; for this reason the processing can be termed *retinotopic processing*.

An obvious problem concerns what information should be extracted and what computations should be performed at these levels. Is *any* type of operation feasible? An axiomatic approach that has been adopted in order to restrict the space of possibilities, is to assume that the very first stages of visual processing should be able to function without any direct knowledge about what can be expected to be in the scene. As a natural consequence, the first processing stages should be as uncommitted and make as few irreversible decisions or choices as possible. Specifically, given that no prior information can be expected about what scales are appropriate, the only reasonable approach is to consider representations at *all* scales. This directly gives rise to the notion of multi-scale representation.

Moreover, the Euclidean nature of the world around us and the perspective mapping to images impose natural constraints on a visual system. Objects move rigidly,

the illumination varies, the size of objects at the retina changes with the depth from the eye, view directions may change etc. Hence, it is natural to require early visual operations to be unaffected by certain primitive transformations.

In this article, we shall show that these constraints, in fact, substantially restrict the class of possible low-level operations. For an uncommitted vision system, the scale-space theory states that the natural operations of perform in a visual front-end are convolutions with Gaussian kernels and their derivatives at different scales. The output from these operations can then, in turn, be used as input to a large number of other visual modules. What types of combinations of Gaussian derivatives are natural to perform can to a large extent be determined from basic invariance properties. An attractive property of this type of approach is that it gives a uniform structure on the first stages of computation.

With these remarks I would like close this philosophical introduction with the hope that it should motivate the more technical treatment that follows below.

2   Multi-scale representation of image data

The basic idea behind a multi-scale representation is to embed the original signal into a one-parameter family of derived signals. How should such a representation be constructed? 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 smoothing method.

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

This property has been formalized in a variety of ways by different authors. A noteworthy coincidence is that the same conclusion can be reached from several different starting points. The main result we will arrive at is that if rather general conditions are posed on the types of computations that are to be performed at the first stages of visual processing, then the Gaussian kernel and its derivatives are singled out as the only possible smoothing kernels. The requirements, or axioms, that specify the uniqueness of the Gaussian kernel are basically linearity and spatial shift invariance combined with different ways of formalizing the notion that structures at coarse scales should be related to structures at finer scales in a well-behaved manner; new structures should not be created by the smoothing method.

Why should one represent a signal at multiple scales when all information is anyway in the original data? The major reason for this is to explicitly represent the multi-scale aspect of real-world images. Another aim is to simplify further processing

by removing unnecessary and disturbing details. More technically, the latter motivation reflects the common need for smoothing as a pre-processing step to many numerical algorithms as a means of noise suppression.

Of course, there exists a variety of possible ways to construct a one-parameter family of derived signals from a given signal. The terminology that will be adopted[1] here is to refer to as "multi-scale representation" any one-parameter family for which the parameter has a clear interpretation in terms of spatial scale. The term "scale-space" will be reserved for the multi-scale representation constructed by convolution with Gaussian kernels of increasing width, or equivalently by solving the diffusion equation.

This presentation aims at giving an overview of some of the main results from the theory of scale-space representation, as it has been developed to its current state. Only few indications of proofs will be given; the reader is referred to the original sources concerning details and further explanation. In order to give a wider background to the problem, a number of outlooks will also be given concerning early and related work on other types of multi-scale representations. Inevitably, a short summary like this one will be biased towards certain aspects of the problem. Therefore, I would like to apologize for any reference that has been left out.

As a guide to the reader it should be remarked that in order keep the length of the presentation reasonably short, some descriptions of original work have been reduced to very condensed summaries. In those cases, a reader unfamiliar with the subject is recommended to proceed and to interpret the presentation merely as an overview that could serve as introduction to a more thorough study of the original sources.

3 Early multi-scale representations

The general idea of representing a signal at multiple scales is not entirely new. Early work in this direction was performed by Rosenfeld and Thurston 1971, who observed the advantage of using operators of different sizes in edge detection. Related approaches were considered by Klinger (1971), Uhr (1972), Hanson and Riseman (1974), and Tanimoto and Pavlidis (1975) concerning image representations using different levels of spatial resolution, i.e., different amounts of subsampling. These ideas have then been furthered, mainly by Burt and by Crowley, to one of the types of multi-scale representations most widely used today, the *pyramid*. A brief overview of some main results concerning this concept is given below.

3.1 Quad-tree

One of the earliest types of multi-scale representations of image data is the *quad tree* introduced by Klinger (1971). It is a tree-like representation of image data, where the image is recursively divided into smaller regions.

The basic idea is as follows: Consider, for simplicity, a discrete image f of size $2^K \times 2^K$ for some $K \in \mathbb{Z}$, and define a measure Σ of the grey-level variation in any region. This measure may e.g. be the standard deviation of the grey-level values.

Let $f^{(K)} = f$. If $\Sigma(f^{(K)})$ is greater than some pre-specified threshold α, then split $f^{(K)}$ into sub-images $f_j^{(K-1)}$ $(j = 1..p)$ according to some rule. Then, apply the

---

[1]In some literature the term "scale-space" is used for denoting any type of multi-scale representation. Using that terminology, the scale-space concept developed here should be called "(linear) diffusion scale-space".

procedure recursively to all sub-images until convergence is obtained. A tree of degree p is generated, in which each leaf $f_j^{(k)}$ is a homogeneous block with $\Sigma(f_j^{(k)}) < \alpha$. In the worst case, each pixel may correspond to an individual leaf. On the other hand, if the image contains a small number of regions with a relatively uniform grey-level, then a substantial data reduction can be obtained in this way.

Concerning grey-level data, this representation has been used in simple segmentation algorithms for image processing. In the "split-and-merge" algorithm, a splitting step is first performed according to the above scheme. Then, adjacent regions are merged if the variation measure of the union of the two regions is below the threshold. Another application (when typically $\alpha = 0$) concerns objects defined by uniform grey-levels, e.g. binary objects; see e.g. the book by Tanimoto and Klinger (1980) for more references on this type representation.
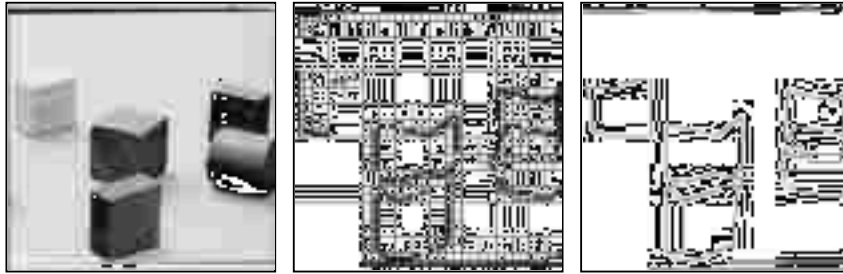

Figure 2: Illustration of the quad-tree concept and the split-and-merge segmentation algorithm; (left) original grey-level image, (middle) the leaves of the quad-tree, i.e., the regions after the split step that have a standard deviation below the given threshold, (right) regions after the merge step.

3.2 Pyramids

Pyramid representations are representations of grey-level data that combine the subsampling operation with a smoothing step (see figure 3 and figure 4). To illustrate the idea, assume again, for simplicity, that the size of the input image f is $2^K \times 2^K$, and let $f^{(K)} = f$. The representation of $f^{(K)}$ at a coarser level $f^{(K-1)}$ is defined by a reduction operator. For simplicity, assume that the smoothing filter is separable, and that the number of filter coefficients along one dimension is odd. Then, it is sufficient to study the following one-dimensional situation.

$$
\begin{aligned}
f^{(k-1)} &= \text{REDUCE}(f^{(k)}) \\
f^{(k-1)}(x) &= \sum_{n=-N}^{N} c(n) \, f^{(k)}(2x - n),
\end{aligned} \tag{1}
$$

where $c : \mathbb{Z} \to \mathbb{R}$ denotes a set of filter coefficients. This type of *low-pass pyramid* representation was proposed almost simultaneously by Burt (1981) and in a thesis by Crowley (1981). A main advantage with this representation it is that the image size decreases exponentially with the scale level, and hence also the amount of computations required to process the data. If the filter coefficients $c(n)$ are chosen properly, then the representations at coarser scale level (smaller k) will correspond to coarser scale structures in the image data. Some of the most obvious design criteria that have been proposed for selecting the filter coefficients are
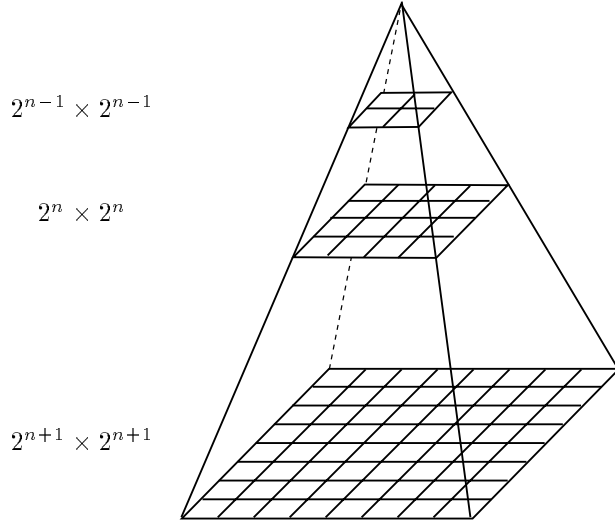
Figure 3: A pyramid representation is obtained by successively reducing the image size by combined smoothing and subsampling.

- positivity: $c(n) \geq 0$,

- unimodality: $c(|n|) \geq c(|n + 1|)$,

- symmetry: $c(-n) = c(n)$, and

- normalization: $\sum_{n=-N}^{N} c(n) = 1$.

Another natural condition is that all pixels should contribute with equal amounts to all levels. In other words, any point that has an odd coordinate index should contribute equally much to the next coarser level as any point having an even coordinate value. Formally, this can be expressed as

- equal contribution: $\sum_{n=-N}^{N} c(2n) = \sum_{n=-N}^{N} c(2n + 1)$.

Equivalently, this condition means that the kernel $(1/2, 1/2)$ of width two should occur as at least one factor in the smoothing kernel.

Whereas most authors agree on these criteria, mutually exclusive conditions have been stated in the frequency domain. Motivated by the sampling theorem, Meer et al. (1987) proposed to approximate an ideal low-pass filter as closely as possible. Since there is no non-trivial finite support kernels with ideal low-pass properties, an approximation is constructed which minimizes the error in the frequency domain. When using the $\mathbb{L}_\infty$ norm, this approach is termed "equiripple design". An alternative is to require the kernel to be positive and unimodal also in the frequency domain. Then, any high frequency signal is guaranteed to be suppressed more than any lower frequency signal.

The choice of N gives rise to a trade-off problem. A larger value of N increases the number of degrees of freedom in the design, at the cost of increased computational work. A natural choice when $N = 1$ is the binomial filter

$$(\frac{1}{4}, \quad \frac{1}{2}, \quad \frac{1}{4}), \tag{2}$$
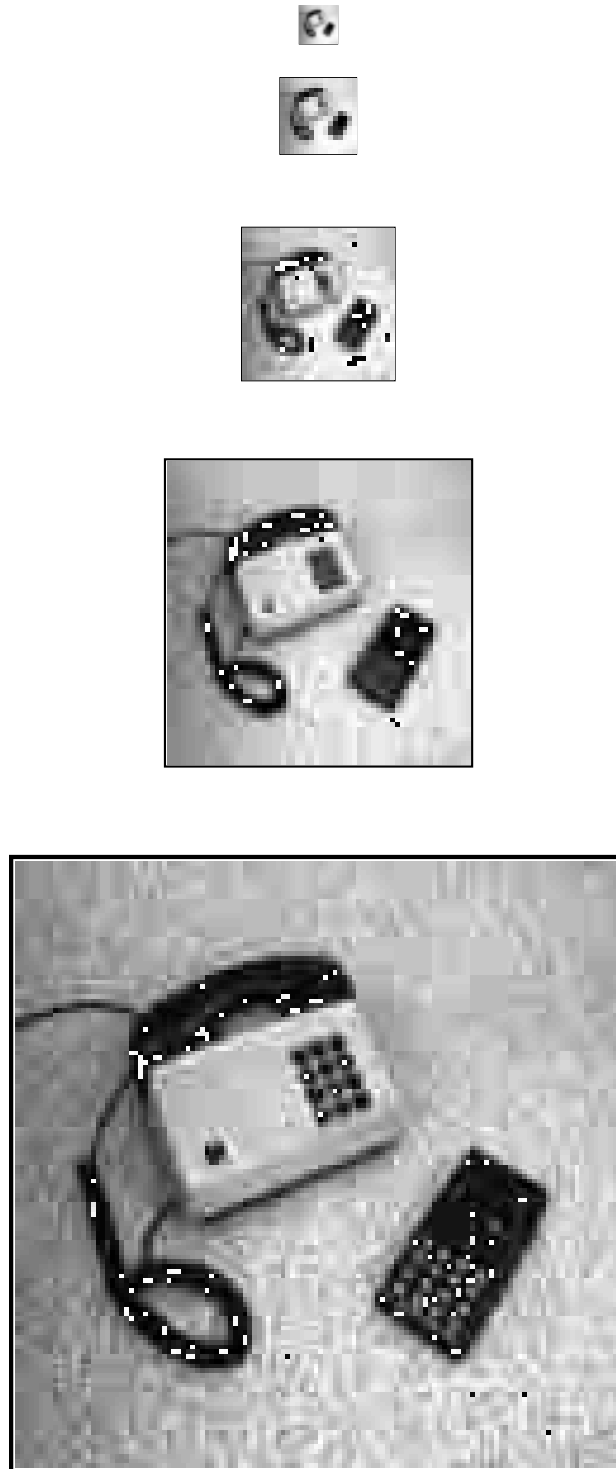
Gaussian (low-pass) pyramid

Figure 4: A Gaussian (low-pass) pyramid is obtained by successive smoothing and subsampling. This pyramid has been generated by the general reduction operator in equation (1) using the binomial filter from equation (2).
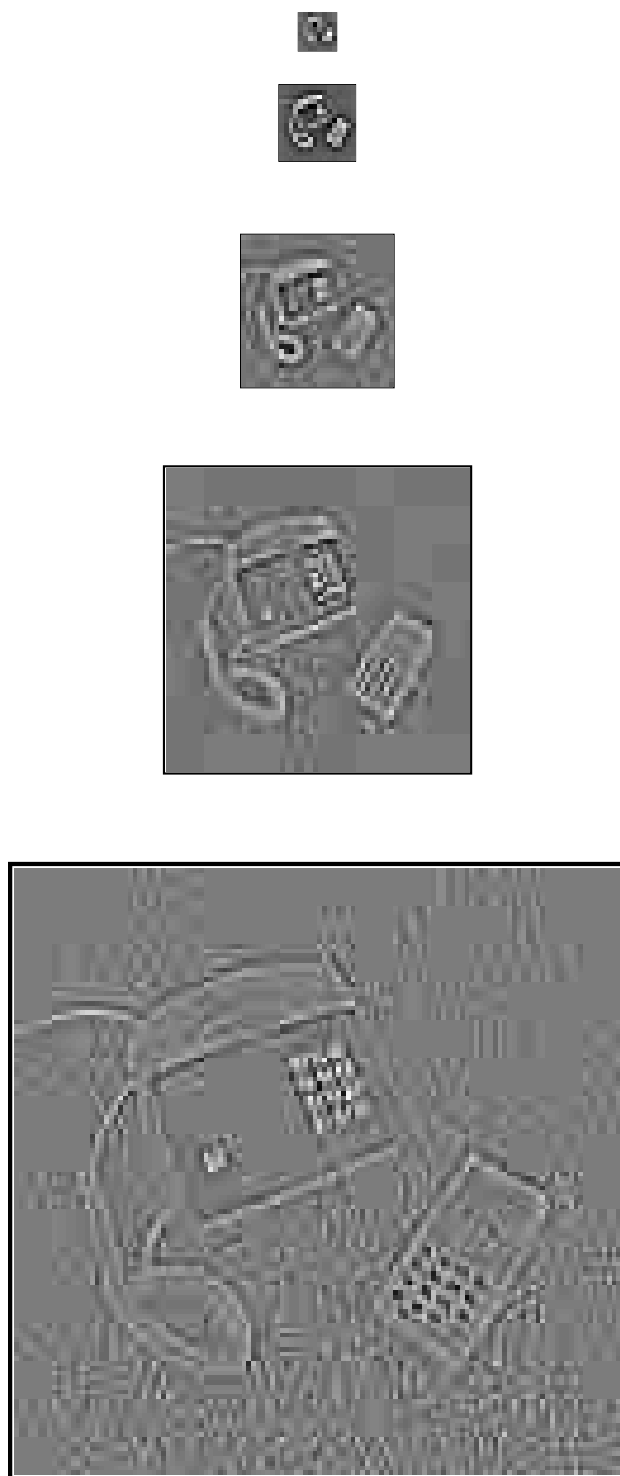
Laplacian (band-pass) pyramid

Figure 5: A Laplacian (low-pass) pyramid is obtained from by forming differences between adjacent levels in the Gaussian pyramid (equation (5)).
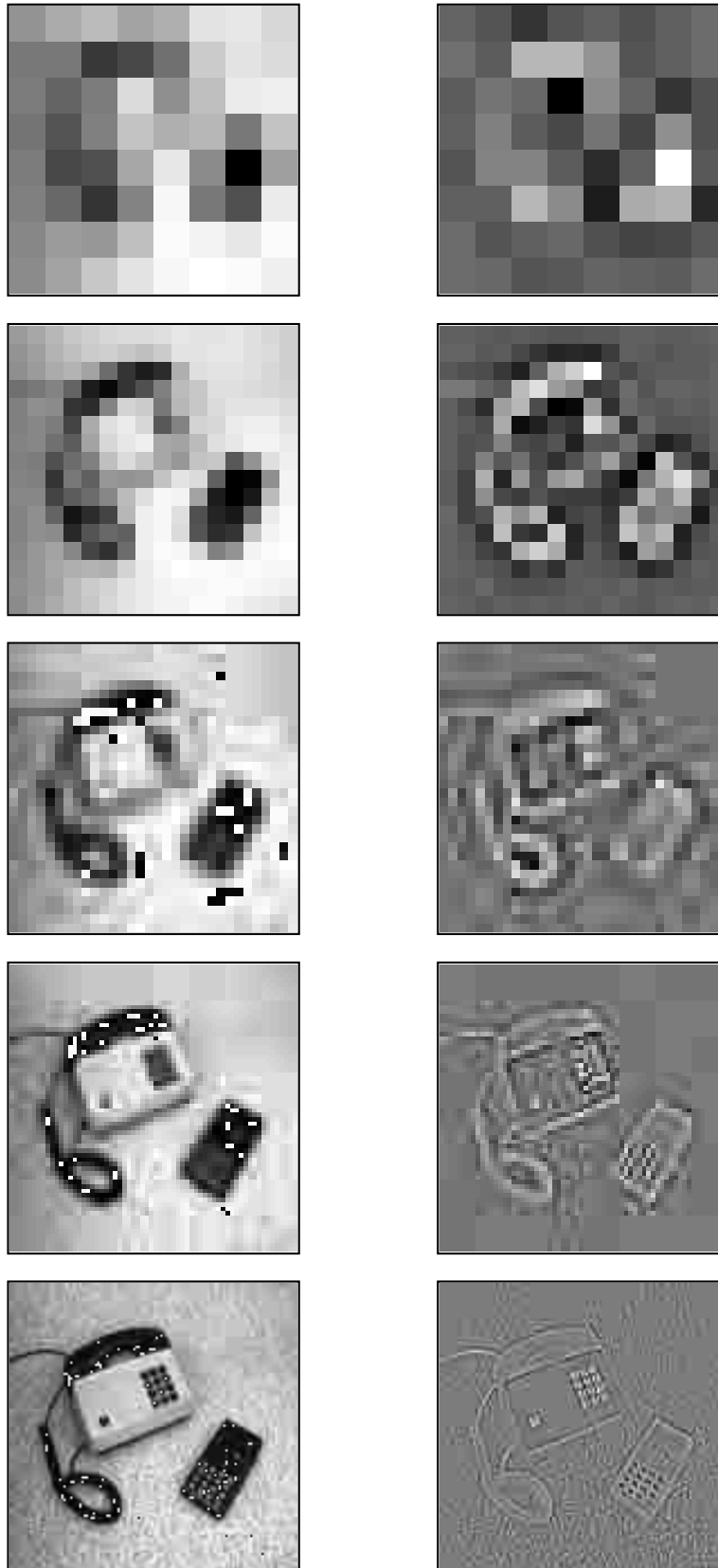
Figure 6: Alternative illustration of the Gaussian and Laplacian pyramids from figures 4–5. Here, the images at the different levels of resolution are rescaled to the same size.

which is the unique filter of width 3 that satisfies the equal contribution condition. It is also the unique filter of width 3 for which the Fourier transform $\psi(\theta) = \sum_{n=-N}^{N} c(n) \exp(-in\theta)$ is zero at $\theta = \pm\pi$. A negative property of this kernel, however, is that when applied repeatedly, the equivalent convolution kernel, which corresponds to the combined effect of iterated smoothing and subsampling, tends to a triangular function.

Of course, there is a large class of other possibilities. Concerning kernels of width 5, the previously stated conditions in the spatial domain imply that the kernel has to be of the form

$$(\frac{1}{4} - \frac{a}{2}, \quad \frac{1}{4}, \quad a, \quad \frac{1}{4}, \quad \frac{1}{4} - \frac{a}{2}). \tag{3}$$

Burt and Adelson (1983) argued that a should be selected such that the equivalent smoothing function should be as similar to a Gaussian as possible. Empirically, they selected the value $a = 0.4$.

By considering a representation defined as the difference between two adjacent levels in a low-pass pyramid, one obtains a *bandpass pyramid*, termed "Laplacian pyramid" by Burt, and DOLP (Difference Of Low Pass) by Crowley. It is defined by

$$L^{(k)} = f^{(k)} - \text{EXPAND}(f^{(k-1)})$$
$$L^{(0)} = f^{(0)}, \tag{4}$$

where EXPAND is an interpolation operator that constitutes the reverse of REDUCE. Formally, it can be expressed as

$$\tilde{f}^{(k)} = \text{EXPAND}(f^{(k-1)})$$
$$\tilde{f}^{(k)}(x) = 2 \sum_{n=-N}^{N} c(n) f^{(k-1)}(\frac{x-n}{2}), \tag{5}$$

where only the terms for which $x - n$ is even are to be included in the sum. This interpolation procedure means that the same weights are used for propagating grey-levels from a coarser to a finer sampling as were used when subsampling the signal.

The bandpass pyramid representation has been used for feature detection and data compression. Among features that can be detected are blobs (maxima), and peaks and ridges etc (Crowley et al 1984, 1987).

The idea behind using such pyramids for data compression is that a bandpass filtered signal will be decorrelated, and have its grey-level histogram centered around the origin. If a coarse quantization of the grey-levels is sufficient for the purpose in mind (typically display), then a data reduction can be obtained by e.g. coding the quantized grey-levels by variable length encoding. From the set of coded bandpass images $\{\tilde{L}^{(k)}\}$, an approximation to the original image $\tilde{f}^{(K)}$ can then be reconstructed by essentially reversing the construction in (4),

$$\tilde{f}^{(0)} = \tilde{L}^{(0)}$$
$$\tilde{f}^{(k)} = \tilde{L}^{(k)} + \text{EXPAND}(\tilde{f}^{(k-1)}). \tag{6}$$

To summarize, the main advantages of the pyramid representations are that they lead to a *rapidly decreasing image size*, which reduces the computational work both in the actual computation of the representation and in the subsequent processing.

The memory requirements are small, and there exist commercially available implementations of pyramids in hardware. The main disadvantage with pyramids is that they correspond to quite a coarse quantization along the scale direction, which makes it algorithmically complicated to relate (match) structures across scales. Pyramids are not translationally invariant.

There is a large literature on further work of pyramid representations; see e.g. the book by Jolion and Rosenfeld JolRos94-book (1994), the books edited by Rosenfeld (1984), Cantoni and Levialdi (1986) and the special issue of IEEE-TPAMI edited by Tanimoto (1989). A selection of recent developments can also be found in the articles by Chehikian and Crowley (1991), Knudsen and Christensen (1991), and Wilson and Bhalerao (1992), for a selection of recent developments. An interesting approach is the introduction of "oversampled pyramids", in which not every smoothing step is followed by a subsampling operation, and hence, a denser sampling along the scale direction can be obtained.

It is worth noting that pyramid representations show a high degree of similarity with a type of numerical methods called multi-grid methods; see the book by Hackbusch (1985) for an extensive treatment of the subject.

4   Scale-space representation

Scale-space representation is a special type of multi-scale representation that comprises a *continuous scale parameter* and preserves the *same spatial sampling* at all scales. As Witkin (1983) introduced the concept, the scale-space representation of a signal is an *embedding* of the original signal into a one-parameter family of derived signals constructed by convolution with a one-parameter family of Gaussian kernels of increasing width. Later, we shall see that the Gaussian kernel is, in fact, a unique choice.

Formally, the linear scale-space representation of a continuous signal is constructed as follows. Let $f : \mathbb{R}^N \to \mathbb{R}$ represent any given signal. Then, the scale-space representation $L : \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ is defined by $L(\cdot; 0) = f$ and

$$L(\cdot; t) = g(\cdot; t) * f, \tag{7}$$

where $t \in \mathbb{R}_+$ is the scale parameter, and $g : \mathbb{R}^N \times \mathbb{R}_+ \backslash \{0\} \to \mathbb{R}$ is the Gaussian kernel; in arbitrary dimensions it is written

$$g(x; t) = \frac{1}{(2\pi t)^{N/2}} e^{-x^T x/(2t)} = \frac{1}{(2\pi t)^{N/2}} e^{-\sum_{i=1}^{N} x_i^2/(2t)} \quad (x \in \mathbb{R}^N, x_i \in \mathbb{R}). \tag{8}$$

The square root of the scale parameter, $\sigma = \sqrt{t}$, is the standard deviation of the kernel g, and is a natural measure of spatial scale in the smoothed signal at scale t. The scale-space family L can equivalently be defined as the solution to the diffusion equation

$$\partial_t L = \frac{1}{2} \nabla^T \nabla L = \frac{1}{2} \sum_{i=1}^{N} \partial_{x_i^2} L. \tag{9}$$

with initial condition $L(\cdot; 0) = f$, which is the well-known physical equation that describes how a heat distribution L evolves over time t in a homogeneous medium with uniform conductivity, given an initial heat distribution $L(\cdot; 0) = f$; see e.g. Widder (1975), or Strang (1986).

Figure 7(a) shows the result of smoothing a one-dimensional signal to different scales in this way. Notice how this successive smoothing captures the intuitive notion of the signals becoming gradually smoother. A two-dimensional example is presented in Figure 8. Here, the scale levels have been determined such that the standard deviation of the Gaussian kernel is equal to the standard deviation of the corresponding pyramid generation kernel.

From this scale-space representation, multi-scale spatial derivatives can be defined by

$$L_{x^n}(\cdot; \ t) = \partial_{x^n} L(\cdot; \ t) = g_{x^n}(\cdot; \ t) * f, \tag{10}$$

where g_{x^n} denotes a (possibly mixed) derivative of some order[2] n. In terms of explicit integrals, the convolution operation (10) is written

$$L_{x^n}(x; \ t) = \int_{x' \in \mathbb{R}^N} g_{x^n}(x - x'; \ t) \, f(x') \, dx' = \int_{x' \in \mathbb{R}^N} g_{x^n}(x'; \ t) \, f(x - x') \, dx'. \tag{11}$$

This representation has a strong regularizing property. If f is bounded by some polynomial, e.g. if there exist some constants $C_1, C_2 \in \mathbb{R}_+$ such that

$$|f(x)| \leq C_1 \left(1 + x^T x\right)^{C_2} \quad (x \in \mathbb{R}^N), \tag{12}$$

then the integral is guaranteed to converge for any $t > 0$. Hence, (10) provides a well-defined way to construct multi-scale derivatives of a function f, although the function itself may not be differentiable of any order[3]. Thus, the scale-space representation given by (7) can for every $t > 0$ be treated as infinitely differentiable ($\mathbb{C}^\infty$).

4.1 Continuous signals: Original formulation

The main idea behind the construction of this scale-space representation is that the fine scale information should be suppressed with increasing values of the scale parameter. Intuitively, when convolving a signal by a Gaussian kernel with standard deviation $\sigma = \sqrt{t}$, the effect of this operation is to suppress[4] most of the structures in the signal with a characteristic length less than σ; see Figure 7(a).

When Witkin 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,

$$\partial_{x^n} L(\cdot; \ t) = \partial_{x^n} (g(\cdot; \ t) * f) = g(\cdot; \ t) * \partial_{x^n} f, \tag{13}$$

this non-creation property applies also to any nth order spatial derivative computed from the scale-space representation.

---

[2]Here, n should be interpreted as a multi-index $n = (n_1, ..., n_N)^T \in \mathbb{Z}^N$ where $n_i \in \mathbb{Z}$. In other words, $\partial_{x^n} = \partial_{x_1^{n_1}} ... \partial_{x_N^{n_N}}$, where $x = (x_1, ..., x_N)^T \in \mathbb{R}^N$ and $x_i \in \mathbb{R}$.

[3]In this sense, the scale-space representation of a signal shows a high degree of similarity with Schwartz distribution theory (1951), although it is neither needed nor desired to explicitly compute the limit case when the (scale) parameter t tends to zero. (See also Florack et al (1993)).

[4]This property does, however, not hold exactly. As we shall see later, adjacent structures (e.g. extrema) can be arbitrary close after arbitrary large amounts of smoothing, although the likelihood for the distance between two adjacent structures to be less than some value ϵ decreases with increasing scale.
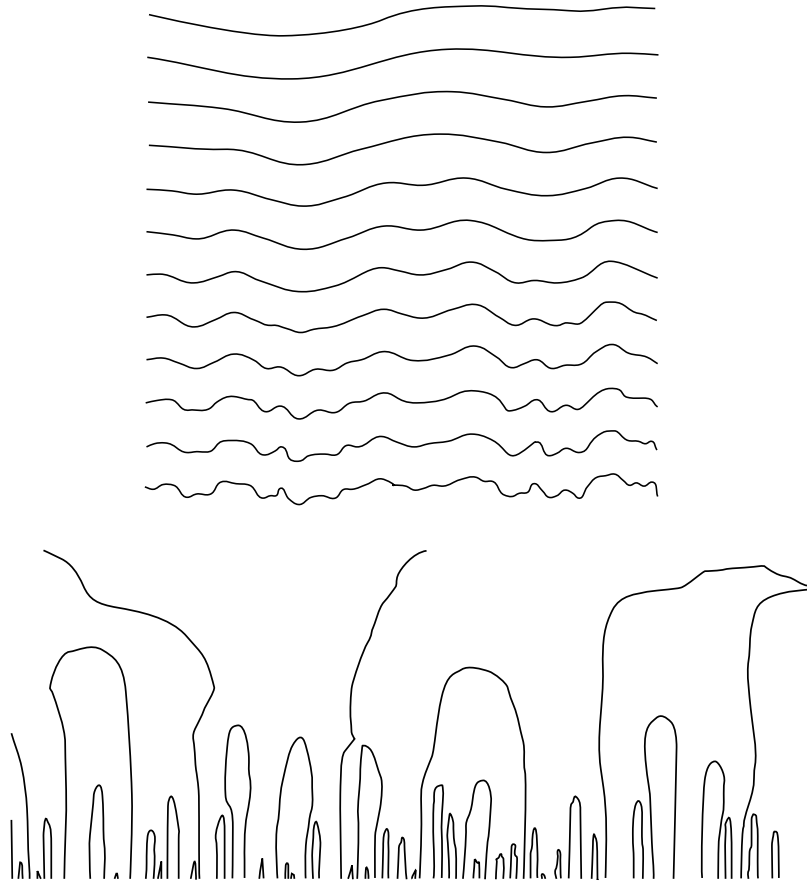
Figure 7: (left) The main idea with a scale-space representation of a signal is to generate a family of derived signals in which the fine-scale information is successively suppressed. This figure shows a one-dimensional signal that has been smoothed by convolution with Gaussian kernels of increasing width. (right) Under this transformation, the zero-crossings of the second derivative form paths across scales, that are never closed from below. (Adapted from Witkin (1983)).

Recall that an extremum in L is equivalent to a zero-crossing in L_x. The non-creation of new local extrema means that the zero-crossings in any derivative of L form closed curves across scales, which will never be closed from below; see Figure 7(b). Hence, in the one-dimensional case, the zero-crossings form paths across scales, with a set of inclusion relations that gives rise to a tree-like data structure, termed "interval tree".

An interesting empirical observation made by Witkin was that he noted a marked correspondence between the length of the branches in the interval tree and perceptual saliency:

> ... intervals that survive over a broad range of scales tend to leap out to the eye ...

In later work by Lindeberg (1991, 1992) it has been demonstrated that this observation can extended to a principle for actually detecting significant image structures from the scale-space representation.
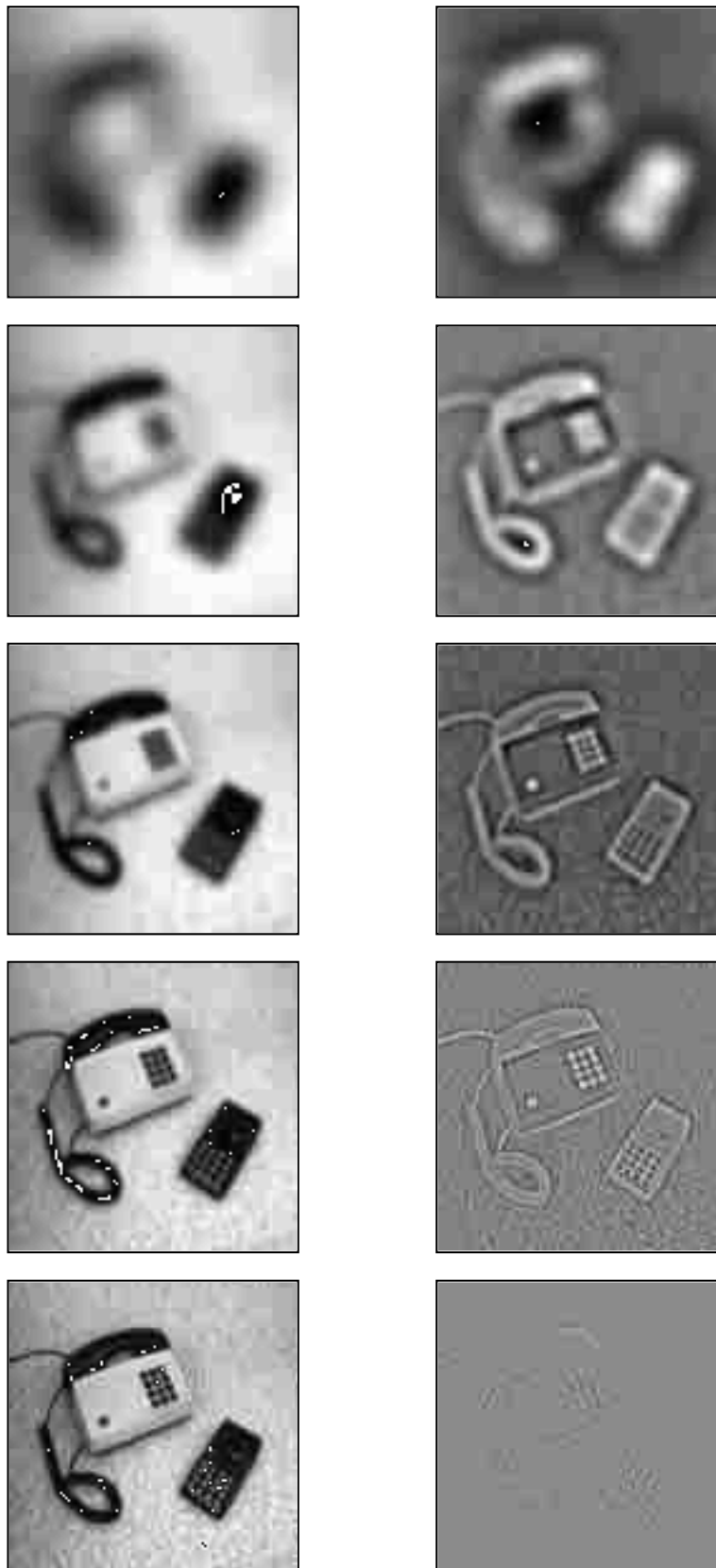
Figure 8: A few slices from the scale-space representation of the image used for illustrating the pyramid concept. The scale levels have been selected such that the scale level of the Gaussian kernel is approximately equal to the standard deviation of the equivalent convolution kernel corresponding to the combined effect of smoothing and subsampling. For comparison, the result of applying the Laplacian operator to these images is shown as well. Observe the differences and similarities compared to Fig. 6.

Gaussian smoothing has been used also before Witkin proposed the scale-space concept, e.g. by Marr and Hildreth (1980), who considered zero-crossings the Laplacian in images convolved with Gaussian kernels of different standard deviation. One of the most important contributions with Witkin's scale-space formulation, however, was the systematic way to *relate* and *interconnect* such representations and image structures at different scales, in the sense that a *scale dimension* should be added to the scale-space representation, so that the behaviour of structures across scales can be studied.

4.2   Causality

Koenderink (1984) emphasized that the problem of scale must be faced in any imaging situation. Any real-world image has a limited extent determined by two scales, the *outer scale* corresponding to the finite size of the image, and the *inner scale* given by the resolution; for a digital image the inner scale is determined by the pixel size, and for a photographic image by the grain size in the emulsion.

As described in the introduction, similar properties apply to objects in the world, and hence also to image features. The outer scale of an object or a feature may be said to correspond to the (minimum) size of a window that completely contains the object or the feature, while the inner scale may be loosely be said to correspond the scale at which substructures of the object or the feature begin to appear. The scale parameter in the scale-space representation determines the minimum scale, or the inner scale, that can be reached in an observation of an image at that scale.

Referring to the analogy with cartography given in the introduction, let us note that an atlas usually contains a set of maps covering some region of interest. Within each map the outer scale typically scales in proportion with the inner scale. A single map is, however, usually not sufficient for us in order to find our way around the world. We need the ability to zoom in to structures at different scales; i.e., decrease and increase the inner scale of the observation according to the type of situation at hand.

Koenderink also stressed that if there is no a priori reason for looking at specific image structures, then the visual system must be able to handle image structures at *all* scales. Pyramid representations approach this problem by successive smoothing and subsampling of images. However,

> The challenge is to understand the image really on all these levels *simultaneously*, and not as unrelated set of derived images at different levels of blurring ...

Adding a scale dimension onto the original data set, as is done in the one-parameter embedding, provides a formal way to express this interrelation.

The observation that new local extrema cannot be created with increasing scale shows that Gaussian convolution satisfies certain sufficiency requirements for being a smoothing operation. The first proof of the *necessity* of Gaussian smoothing for 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 $\{(x, y; \ t) \in \mathbb{R}^2 \times \mathbb{R} : L(x, y; \ t) = L_0\}$ 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

$$\partial_t L = \frac{1}{2}\nabla^2 L.\tag{14}$$

Since the Gaussian kernel is the Green's function of the diffusion equation at an infinite domain, it follows that the Gaussian kernel is the unique kernel for generating the scale-space. A similar result holds in one dimension, and as we shall see later, also in higher dimensions.

Figure 9: The causality requirement means that level surfaces in scale-space must point with their concave side towards finer scale (a); the reverse situation (b) must never occur.

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 with increasing scale, then the level surface must point with its concave side towards decreasing scales. This gives rise to a sign condition on the curvature of the level surface, which when expressed in terms of derivatives of the scale-space representation with respect to the spatial and scale coordinates assumes the form (14). Since the points at which the extrema are attained cannot be assumed to be known a priori, this condition must hold in any point, which proves the result.

In the *one-dimensional* case, this level surface condition becomes a level curve condition, and is equivalent to the previously stated non-creation of local extrema. Since any nth order derivative of L also satisfies the diffusion equation

$$\partial_t L_{x^n} = \frac{1}{2}\nabla^2 L_{x^n},\tag{15}$$

it follows that new zero-crossing curves in L_x cannot be created with increasing scale, and hence, no new maxima.

A similar result was given by Yuille and Poggio (1986) concerning the zero-crossings of the Laplacian of the Gaussian. Related formulations have also been expressed by Babaud et al (1986), and by Hummel (1986, 1987).

4.3 Decreasing number of local extrema

Lindeberg (1990, 1991) considered the problem of characterizing those kernels who 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 in the convolved signal $f_{out} = h * f_{in}$ is always less than or equal to the number of local extrema in the original signal is termed a *scale-space kernel*:

- scale-space kernel: $\#_{extrema}(h * f_{in}) \leq \#_{extrema}(f_{in}) \quad \forall f_{in} \in \mathbb{L}_1$.

16

From similar arguments as in Section 4.1, this definition implies that the number of extrema (or zero-crossings) in any nth order derivative is guaranteed to never increase. Hence, convolution with a scale-space kernel has a strong smoothing property.

Such kernels can be easily shown to be positive and unimodal both in the spatial domain and in the frequency domain. These properties may provide a formal justification for some of the design criteria listed in Section 3.2 concerning the choice of filter coefficients for pyramid generation.

Provided that the notion of local maximum or zero-crossing is defined in a proper manner to cover also non-generic functions, it holds that the scale-space kernels can be completely classified using classical results by Schoenberg (1950). It can be shown that a continuous kernel h is a scale-space kernel if and only if it has a bilateral Laplace-Stieltjes transform of the form

$$\int_{x=-\infty}^{\infty} h(x)\, e^{-sx} dx = C\, e^{\gamma s^2 + \delta s} \prod_{i=1}^{\infty} \frac{e^{a_i s}}{1 + a_i s} \qquad (-c < \mathrm{Re}(s) < c) \qquad (16)$$

for some $c > 0$, where $C \neq 0$, $\gamma \geq 0$, δ and a_i are real, and where $\sum_{i=1}^{\infty} a_i^2$ is convergent; see also the excellent books by Hirschmann and Widder (1955), and by Karlin (1968).

Interpreted in the spatial domain, this result means that for continuous signals there are four primitive types of linear and shift-invariant smoothing transformations; convolution with the *Gaussian kernel*,

$$h(x) = e^{-\gamma x^2} \qquad (17)$$

convolution with the *truncated exponential functions*,

$$h(x) = \begin{cases} e^{-|\lambda| x} & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad h(x) = \begin{cases} e^{|\lambda| x} & x \leq 0 \\ 0 & x > 0 \end{cases} \qquad (18)$$

as well as trivial *translation* and *rescaling*.

The product form of the expression (16) reflects a direct consequence of the definition of scale-space kernel; the convolution of two scale-space kernels is a scale-space kernel. Interestingly, the characterization means that the reverse holds; a shift-invariant linear transformation is a smoothing operation if and only if it can be decomposed into these primitive operations.

4.4 Semi-group 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,

$$h(\cdot;\, t_1) * h(\cdot;\, t_2) = h(\cdot;\, t_1 + t_2). \qquad (19)$$

In particular, this condition implies that the transformation from a fine scale level to any coarse scale level should be of the same type as the transformation from the original signal to the scale-space representation. Algebraically, this property can be written

$$L(\cdot;\, t_2) = \{\text{definition}\} = T(\cdot;\, t_2) * f = \{\text{semi-group}\} =$$

$$= (T(\cdot;\ t_2 - t_1) * T(\cdot;\ t_1)) * f = \{\text{associativity}\} =$$

$$= T(\cdot;\ t_2 - t_1) * (T(\cdot;\ t_1) * f) = \{\text{definition}\} = T(\cdot;\ t_2 - t_1) * L(\cdot;\ t_1). \quad (20)$$

If a semi-group structure is imposed on a one-parameter family of scale-space kernels that satisfy a mild degree of smoothness (Borel-measurability) with respect to the parameter, and if the kernels are required to be symmetric and normalized, then the family of smoothing kernels is uniquely determined (Lindeberg, 1990)

$$h(x;\ t) = \frac{1}{\sqrt{2\pi\alpha t}} e^{-x^2/(2\alpha t)} \qquad (t > 0 \quad \delta \in R). \quad (21)$$

In other words, when combined with the semi-group structure, the non-creation of new local extrema means that the smoothing family is uniquely determined.

Despite the completeness of these results, they cannot be extended directly to higher dimensions, since in two (and higher) dimensions there are no non-trivial kernels guaranteed to never increase the number of local extrema in a signal. One example of this, originating from an observation by Lifshitz and Pizer (1990), is presented below; see also Yuille (1988):

> Imagine a two-dimensional image function consisting of two hills, one of them somewhat higher than the other one. Assume that they are smooth, wide, rather bell-shaped surfaces situated some distance apart clearly separated by a deep valley running between them. Connect the two tops by a narrow sloping ridge without any local extrema, so that the top point of the lower hill no longer is a local maximum. Let this configuration be the input image. When the operator corresponding to the diffusion equation is applied to the geometry, the ridge will erode much faster than the hills. After a while it has eroded so much that the lower hill appears as a local maximum again. Thus, a new local extremum has been created.

Notice however, that this decomposition of the scene is intuitively quite reasonable. The narrow ridge is a fine-scale phenomenon, and should therefore disappear before the coarse scale peaks. The property that new local extrema can be created with increasing scale is inherent in two and higher dimensions.

4.5 Scale invariance

A recent formulation by Florack et al (1992) shows that the uniqueness of the Gaussian kernel for scale-space representation can be derived under weaker conditions, essentially by combining the earlier mentioned linearity, shift invariance and semi-group conditions 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 physical observable entities that are subjected to scale changes should be treated in a self-similar manner.

Some more technical requirements must be used in order to prove the uniqueness. The solution must not tend to infinity when the scale parameter increases. Moreover,

either rotational symmetry or separability in Cartesian coordinates needs to be imposed in order to guarantee uniform treatment of the different coordinate directions. Since the proof of this result is valid in arbitrary dimensions and not very technical, I will reproduce a simplified[5] version of it, which nevertheless contains the basic steps.

4.5.1 Necessity proof from scale invariance

Recall that any linear and shift-invariant operator can be expressed as a convolution operator. Hence, assume that the scale-space representation $L : \mathbb{R}^N \times R_+ \to \mathbb{R}$ of any signal $f : \mathbb{R}^N \to \mathbb{R}$ is constructed by convolution with some one-parameter family of kernels $h : \mathbb{R}^N \times R_+ \to \mathbb{R}$

$$L(\cdot;\ t) = h(\cdot;\ t) * f. \tag{22}$$

In the Fourier domain ($\omega \in \mathbb{R}^N$), this can be written

$$\hat{L}(\omega;\ t) = \hat{h}(\omega;\ t)\hat{f}(\omega). \tag{23}$$

A result in physics, called the Pi-theorem, states that if a physical process is scale independent, then it should be possible to express the process in terms of dimensionless variables. Here, the following dimensions and variables occur

$$
\begin{array}{ll}
\text{Luminance:} & \hat{L},\ \hat{f} \\
\text{Length}^{-1}: & \omega,\ 1/\sqrt{t}.
\end{array}
$$

Natural dimensionless variables to introduce are hence, $\hat{L}/\hat{f}$ and $\omega\sqrt{t}$. Using the Pi-theorem, a necessary requirement for scale invariance is that (23) can be expressed on the form

$$\frac{\hat{L}(\omega;\ t)}{\hat{f}(\omega;\ t)} = \hat{h}(\omega;\ t) = \hat{H}(\omega\sqrt{t}) \tag{24}$$

for some function $\hat{H} : \mathbb{R}^N \to \mathbb{R}$. A necessary requirement on $\hat{H}$ is that $\hat{H}(0) = 1$. Otherwise $\hat{L}(\omega;\ 0) = \hat{f}(\omega)$ would be violated.

If h is required to be a semi-group with respect to the scale parameter, then the following relation must hold in the Fourier domain

$$\hat{h}(\omega;\ t_1)\,\hat{h}(\omega;\ t_2) = \hat{h}(\omega;\ t_1 + t_2), \tag{25}$$

and consequently in terms of $\hat{H}$,

$$\hat{H}(\omega\sqrt{t_1})\,\hat{H}(\omega\sqrt{t_2}) = \hat{H}(\omega\sqrt{t_1 + t_2}). \tag{26}$$

Assume first that $\hat{H}$ is rotationally symmetric, and introduce new variables $v_i = u_i^T u_i = (\omega\sqrt{t_i})^T(\omega\sqrt{t_i}) = \omega^T\omega t_i$. Moreover, let $\tilde{H} : \mathbb{R} \to \mathbb{R}$ be defined by $\tilde{H}(u^T u) = \hat{H}(u)$. Then, (26) assumes the form

$$\hat{H}(v_1)\,\hat{H}(v_2) = \hat{H}(v_1 + v_2). \tag{27}$$

---

[5] Here, it is assumed that the semi-group is of the form $g(\cdot;\ t_1) * g(\cdot;\ t_2) = g(\cdot;\ t_1 + t_2)$, and that the scale values are measured in terms of t should be added by regular summation. This is a so-called *canonical semi-group*. More generally, Florack et al (1992) consider semi-groups of the form $g(\cdot;\ \sigma_1^p) * g(\cdot;\ \sigma_2^p) = g(\cdot;\ \sigma_1^p + \sigma_2^p)$ for some $p \geq 1$, where the scale parameter σ is assumed to have dimension length. By combining rotational symmetry and separability in Cartesian coordinates, they show that is uniquely fixates p to be two.

This expression can be recognized as the definition of the exponential function, which means that

$$\hat{H}(v) = \exp(\alpha v) \tag{28}$$

for some $\alpha \in \mathbb{R}$, and

$$\hat{h}(\omega;\ t) = \hat{H}(\omega\sqrt{t}) = e^{\alpha\,\omega^T \omega\, t}. \tag{29}$$

Concerning the sign of α, it is natural to require $\lim_{t\to\infty} \hat{h}(\omega;\ t) = 0$ rather than $\lim_{t\to\infty} \hat{h}(\omega;\ t) = \infty$. This means that α must be negative, and we can without loss of generality set $\alpha = -1/2$, in order to preserve consistency with the previous definitions of the scale parameter t. Hence, the Fourier transform of the smoothing kernel is uniquely determined as the Fourier transform of the Gaussian kernel

$$\hat{g}(\omega;\ t) = e^{-\omega^T \omega t/2}. \tag{30}$$

Alternative, the assumption about rotational invariance can be replaced by separability. Assume that $\hat{H}$ in (26) can be expressed on the form

$$\hat{H}(u) = \hat{H}(u^{(1)}, u^{(2)}, ..., u^{(N)}) = \prod_{i=1}^{N} \bar{H}(u^{(i)}) \tag{31}$$

for some function $\bar{H} : \mathbb{R} \to \mathbb{R}$. Then, (26) assumes the form

$$\prod_{i=1}^{N} \bar{H}(v_1^{(i)}) \prod_{i=1}^{N} \bar{H}(v_2^{(i)}) = \prod_{i=1}^{N} \bar{H}(v_1^{(i)} + v_2^{(i)}), \tag{32}$$

where new coordinates $v_j^{(i)} = (u_j^{(i)})^2$ have been introduced. Similarly to above, it must hold for any $\omega \in \mathbb{R}^n$, and hence under independent variations of the individual coordinates, i.e.,

$$\bar{H}(v_1^{(i)})\, \bar{H}(v_2^{(i)}) = \bar{H}(v_1^{(i)} + v_2^{(i)}), \tag{33}$$

for any $v_1^{(i)}, v_n^{(i)} \in \mathbb{R}$. This means that $\bar{H}$ must be an exponential function, and that $\hat{h}$ must be the Fourier transform of the Gaussian kernel. Q.E.D.

Remark. Here, it has been assumed that the semi-group is of the form

$$g(\cdot;\ t_1) * g(\cdot;\ t_2) = g(\cdot;\ t_1 + t_2),$$

and that the scale values measured in terms of t should be added by regular summation. This is a so-called *canonical semi-group*. More generally, Florack *et al.* (1992b) consider semi-groups of the form $g(\cdot;\ \sigma_1^p) * g(\cdot;\ \sigma_2^p) = g(\cdot;\ \sigma_1^p + \sigma_2^p)$ for some $p \geq 1$, where the scale parameter σ is assumed to have dimension length. By combining rotational symmetry with separability in Cartesian coordinates, they show that these conditions uniquely fixate the exponent p to be two.

For one-dimensional signals though, this parameter will be undetermined. This gives rise to a one-parameter family of one-parameter smoothing kernels having Fourier transforms of the form

$$\hat{h}(\omega) = e^{-\frac{|\omega|^p}{2}\, t},$$

where p is the free parameter. An analysis by Pauwels *et al.* (1994) shows that only a discrete subset of the corresponding multi-scale representations generated by these kernels can be described by differential equations (corresponding to infinitesimal generators in analogy with the (discrete) equation (51)), namely those for which p is an even integer. Out of these, only the specific choice $p = 2$ corresponds to a non-negative convolution kernel. Similarly, $p = 2$ is the unique choice for which the multi-scale representation satisfies the causality requirement (corresponding to non-enhancement of local extrema).

4.5.2 Operators derived from the scale-space representation

Above, it was shown that the Gaussian kernel is the unique kernel for generating a (linear) scale-space. An interesting problem, concerns what operators are natural to apply to the output from this representation.

In early work, Koenderink and van Doorn (1987) advocated the use of the multi-scale N-jet signal representation, that is the set of spatial derivatives of the scale-space representation up to some (given) order N. Then, in (Koenderink and van Doorn, 1992) they considered the problem of deriving linear operators from the scale-space representation, which are to be invariant under scaling transformations. Inspired by the relation between the Gaussian kernel and its derivatives, here in one dimension,

$$\partial_{x^n} g(x;\ t) = (-1)^n \frac{1}{(2t)^{n/2}} H_n(\frac{x}{\sqrt{2t}}) g(x;\ t), \tag{34}$$

which follows from the well-known relation between the derivatives of the Gaussian kernel and the Hermite polynomials H_n

$$\partial_{x^n}(e^{-x^2}) = (-1)^n H_n(x) e^{-x^2}, \tag{35}$$

they considered the problem of deriving operators with a similar scaling behaviour. Starting from the *ansatz*

$$\psi^{(\alpha)}(x;\ t) = \frac{1}{(2t)^{\alpha/2}} \varphi^{(\alpha)}(\frac{x}{\sqrt{2t}}) g(x;\ t), \tag{36}$$

where the superscript (α) describes the "order" of the function, they considered the problem of determining all functions $\varphi^{(\alpha)} : \mathbb{R}^N \to \mathbb{R}$ such that $\psi^{(\alpha)} : \mathbb{R}^N \to \mathbb{R}$ satisfies the diffusion equation. Interestingly, $\varphi^{(\alpha)}$ must then satisfy the time-independent Schrödinger equation

$$\nabla^T \nabla \varphi(\xi) + ((2\alpha + N) - \xi^T \xi)\varphi(\xi) = 0, \tag{37}$$

where $\xi = x/\sqrt{2t}$. This is the physical equation that governs the quantum mechanical free harmonic oscillator. It is well-known from mathematical physics that the solutions $\varphi^{(\alpha)}$ to this equation are the Hermite functions, that is Hermite polynomials multiplied by Gaussian functions. Since the derivative of a Gaussian kernel is a Hermite polynomial times a Gaussian kernel, it follows that the solutions $\psi^{(\alpha)}$ to the original problem are the derivatives of the Gaussian kernel.

This result provides a formal statement that Gaussian derivatives are natural operators to derive from scale-space. As pointed out above, these derivatives satisfy the diffusion equation, and obey scale-space properties, for example the *cascade smoothing property*

$$g(\cdot;\ t_1) * g_{x^n}(\cdot;\ t_2) = g_{x^n}(\cdot;\ t_2 + t_1). \tag{38}$$

The latter result is a special case of the more general statement

$$g_{x^m}(\cdot;\; t_1) * g_{x^n}(\cdot;\; t_2) = g_{x^{m+n}}(\cdot;\; t_2 + t_1),\tag{39}$$

whose validity follows directly from the commutative property of convolution and differentiation.

4.6 Special properties of the Gaussian kernel

Let us conclude this treatment concerning continuous signals by listing a number of other special properties of the Gaussian kernel. The definition of the Fourier transform of any function $h : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}$ that has been used in this paper is

$$\hat{h}(\omega) = \int_{x \in \mathbb{R}^N} h(x)\, e^{-i\omega^T x}\, dx\,.\tag{40}$$

Consider for simplicity the one-dimensional case and define the normalized second moments (variances) Δx and $\Delta \omega$ in the spatial and the Fourier domain respectively by

$$\Delta x = \frac{\int_{x \in \mathbb{R}} x^T x |h(x)|^2 dx}{\int_{x \in \mathbb{R}} |h(x)|^2 dx}\tag{41}$$

$$\Delta \omega = \frac{\int_{\omega \in \mathbb{R}} \omega^T \omega |\hat{h}(\omega)|^2 d\omega}{\int_{\omega \in \mathbb{R}} |\hat{h}(\omega)|^2 d\omega}\tag{42}$$

These entities the "spread" of the distributions h and $\hat{h}$. Then, the uncertainty relation states that

$$\Delta x \Delta \omega \geq \frac{1}{2}\,.\tag{43}$$

A remarkable property of the Gaussian kernel is that it is the only real kernel that gives equality in this relation. Moreover, the Gaussian kernel is the only rotationally symmetric kernel that is separable in Cartesian coordinates.

The Gaussian kernel is also the frequency function of the normal distribution. The central limit theorem in statistics states that under rather general requirements on the distribution of a stochastic variable, the distribution of a sum of a large number of such stochastic variables asymptotically approaches a normal distribution, when the number of terms tend to infinity.

4.7 Discrete signals: No new local extrema

The treatment so far has been concerned with continuous signals. However, real-world signals obtained from standard digital cameras are discrete. An obvious problem concerns how scale-space theory should be discretized, while still maintaining the scale-space properties.

For one-dimensional signals it turns out to be possible to develop a complete discrete theory based on a discrete analogy to the treatment in Section 4.3. Following Lindeberg (1990, 1991), define a discrete kernel $h \in l_1$ to be a *discrete scale-space kernel* if for any signal f_{in} the number of local extrema in $f_{out} = h * f_{in}$ does not exceed the number of local extrema in f_{in}.

Using classical results (mainly by Schoenberg (1953); see Karlin (1968) for a comprehensive summary) it is possible to completely classify those kernels that satisfy this definition. A discrete kernel is a scale-space kernel if and only if its generating function $\varphi_h(z) = \sum_{n=-\infty}^{\infty} h(n)z^n$ is of the form

$$\varphi_K(z) = c\, z^k\, e^{(q_{-1}z^{-1} + q_1 z)} \prod_{i=1}^{\infty} \frac{(1 + \alpha_i z)(1 + \delta_i z^{-1})}{(1 - \beta_i z)(1 - \gamma_i z^{-1})}, \tag{44}$$

where $c > 0$, $k; \in Z$, $q_{-1}, q_1, \alpha_i, \beta_i, \gamma_i, \delta_i \geq 0$, $\beta_i, \gamma_i < 1$ and $\sum_{i=1}^{\infty}(\alpha_i + \beta_i + \gamma_i + \delta_i) < \infty$.

The interpretation of this result is that there are five primitive types of linear and shift-invariant smoothing transformations, of which the last two are trivial;

- two-point weighted average or *generalized binomial smoothing*

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

- moving average or *first order recursive filtering*

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

- infinitesimal smoothing or *diffusion smoothing* (see below for an explanation),

- rescaling, and

- translation.

It follows that a discrete kernel is a scale-space kernel if and only if it can be decomposed into the above primitive transformations. Moreover, the only non-trivial smoothing kernels of *finite support* arise from generalized binomial smoothing.

If this definition is combined with a requirement that the family of smoothing transformations must obey a *semi-group* property over scales and possess a *continuous scale parameter*, then the result is that there is in principle only one way to construct a scale-space for discrete signals. Given a signal $f : \mathbb{Z} \to \mathbb{R}$ the scale-space representation $L : \mathbb{Z} \times \mathbb{R}_+ \to \mathbb{R}$ is given by

$$L(x;\, t) = \sum_{n=-\infty}^{\infty} T(n;\, t) f(x - n), \tag{45}$$

where $T : \mathbb{Z} \times \mathbb{R}_+ \to \mathbb{R}$ is a kernel termed *the discrete analogue of the Gaussian kernel.* It is defined in terms one type of Bessel functions, the modified Bessel functions I_n (see Abramowitz and Stegun 1964):

$$T(n;\, t) = e^{-\alpha t} I_n(\alpha t). \tag{46}$$

This kernel satisfies several properties in the discrete domain that are similar to those of the Gaussian kernel in the continuous domain; for example, it tends to the discrete delta function when $t \to 0$, while for large t it approaches the continuous Gaussian.

The scale parameter t can be related to spatial scale from the second moment of the kernel, which when $\alpha = 1$ is

$$\sum_{n=-\infty}^{\infty} n^2\, T(n;\, t) = t. \tag{47}$$

The term "diffusion smoothing" can be understood by noting that the scale-space family L satisfies a semi-discretized version of the diffusion equation:

$$\partial_t L(x;\, t) = \frac{1}{2}(L(x+1;\, t) - 2L(x;\, t) + L(x-1;\, t)) = \frac{1}{2}(\nabla_2^2 L)(x;\, t) \tag{48}$$

with initial condition $L(x;\, 0) = f(x)$, i.e., the equation that is obtained if the continuous one-dimensional diffusion equation is discretized in space using the standard second difference operator $\nabla_2^2 L$, but the continuous scale parameter is left untouched.

A simple interpretation of the discrete analogue of the Gaussian kernel is as follows: Consider the time discretization of (48) using Euler's explicit method

$$L^{(k+1)(i)} = \frac{\Delta t}{2}L^{(k)}(i+1) + (1 - \Delta t)L^{(k)}(i) + \frac{\Delta t}{2}L^{(k)}(i-1), \tag{49}$$

where the superscript (k) denotes iteration index. Assume that the scale-space representation of L at scale t is to be computed by applying this iteration formula using n steps with step size $\Delta t = t/n$. Then, the discrete analogue of the Gaussian kernel is the limit case of the equivalent convolution kernel

$$(\frac{t}{2n}, \quad 1 - \frac{t}{n}, \quad \frac{t}{2n})^n, \tag{50}$$

when n tends to infinity, i.e., when the number of steps increases and each individual step becomes smaller.

Despite the completeness of these results, and their analogies to the continuous situation, the extension to higher dimensions fails because of arguments similar to the continuous case; there are no non-trivial kernels in two or higher dimensions that are guaranteed to never introduce new local extrema. Hence, a discrete scale-space formulation in higher dimensions must be based on other axioms.

4.8   Non-enhancement of local extrema

It is clear that the continuous scale-space formulations in terms of causality and scale invariance cannot be transferred directly to discrete signals; there are no direct discrete correspondences to level curves and differential geometry in the discrete case. Neither can the scaling argument be carried out in the discrete situation if a continuous scale parameter is desired, since the discrete grid has a preferred scale given by the distance between adjacent grid points. An alternative way to express the causality requirement in the continuous case is as follows: *If for some scale level t_0 a point x_0 is a 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 local minimum then its value must not decrease when the scale parameter increases.

It is clear that this formulation is equivalent to the formulation in terms of level curves for continuous images, since if the grey-level value at a local maximum (minimum) would increase (decrease) then a new level curve would be created. Conversely,

if a new level curve is created then some local maximum (minimum) must have increased (decreased). An intuitive description of this requirement is that it prevents local extrema from being enhanced and from "popping up out of nowhere". In fact, it is closely related to the maximum principle for parabolic differential equations (see, e.g., Widder (1975) and also Hummel (1987)).

If this requirement is combined with a semi-group structure, spatial symmetry, and continuity requirements with respect to the scale parameter (strong continuity; see Hille and Phillips (1957)), then it can be shown that the scale-space family $L : \mathbb{Z}^N \times \mathbb{R}^+ \to \mathbb{R}$ of a discrete signal $f : \mathbb{Z}^N \to \mathbb{R}$ must satisfy the semi-discrete differential equation

$$(\partial_t L)(x;\, t) = (\mathcal{A}_{ScSp} L)(x;\, t) = \sum_{\xi \in \mathbb{Z}^N} a_\xi L(x - \xi;\, t), \tag{51}$$

for some *infinitesimal scale-space generator* $\mathcal{A}_{ScSp}$, which is 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.

In one and two dimensions respectively (51) reduces to

$$\partial_t L = \alpha_1 \nabla_3^2 L, \tag{52}$$

$$\partial_t L = \alpha_1 \nabla_5^2 L + \alpha_2 \nabla_{\times^2}^2 L, \tag{53}$$

for some constants $\alpha_1 \geq 0$ and $\alpha_2 \geq 0$. The symbols, ∇_5^2 and $\nabla_{\times^2}^2$ denote two common discrete approximations of the Laplace operator; they are defined by (below the notation $f_{-1,1}$ stands for $f(x - 1, y + 1)$ etc.):

$$(\nabla_5^2 f)_{0,0} = f_{-1,0} + f_{+1,0} + f_{0,-1} + f_{0,+1} - 4f_{0,0},$$

$$(\nabla_{\times^2}^2 f)_{0,0} = 1/2(f_{-1,-1} + f_{-1,+1} + f_{+1,-1} + f_{+1,+1} - 4f_{0,0}).$$

In the particular case when $\alpha_2 = 0$, the two-dimensional representation is given by convolution with the one-dimensional Gaussian kernel along each dimension. On the other hand, using $\alpha_1 = 2\alpha_2$ corresponds to a representation with maximum spatial isotropy in the Fourier domain.

Concerning operators derived from the discrete scale-space representation, it holds that the scale-space properties transfer to any discrete derivative approximation defined by spatial linear filtering of the scale-space representation. In fact, the converse result is true as well (Lindeberg 1993); if derivative approximation kernels are to satisfy the cascade smoothing property,

$$\delta_{x^n} T(\cdot;\, t_1) * T(\cdot;\, t_2) = \delta_{x^n} T(\cdot;\, t_1 + t_2), \tag{54}$$
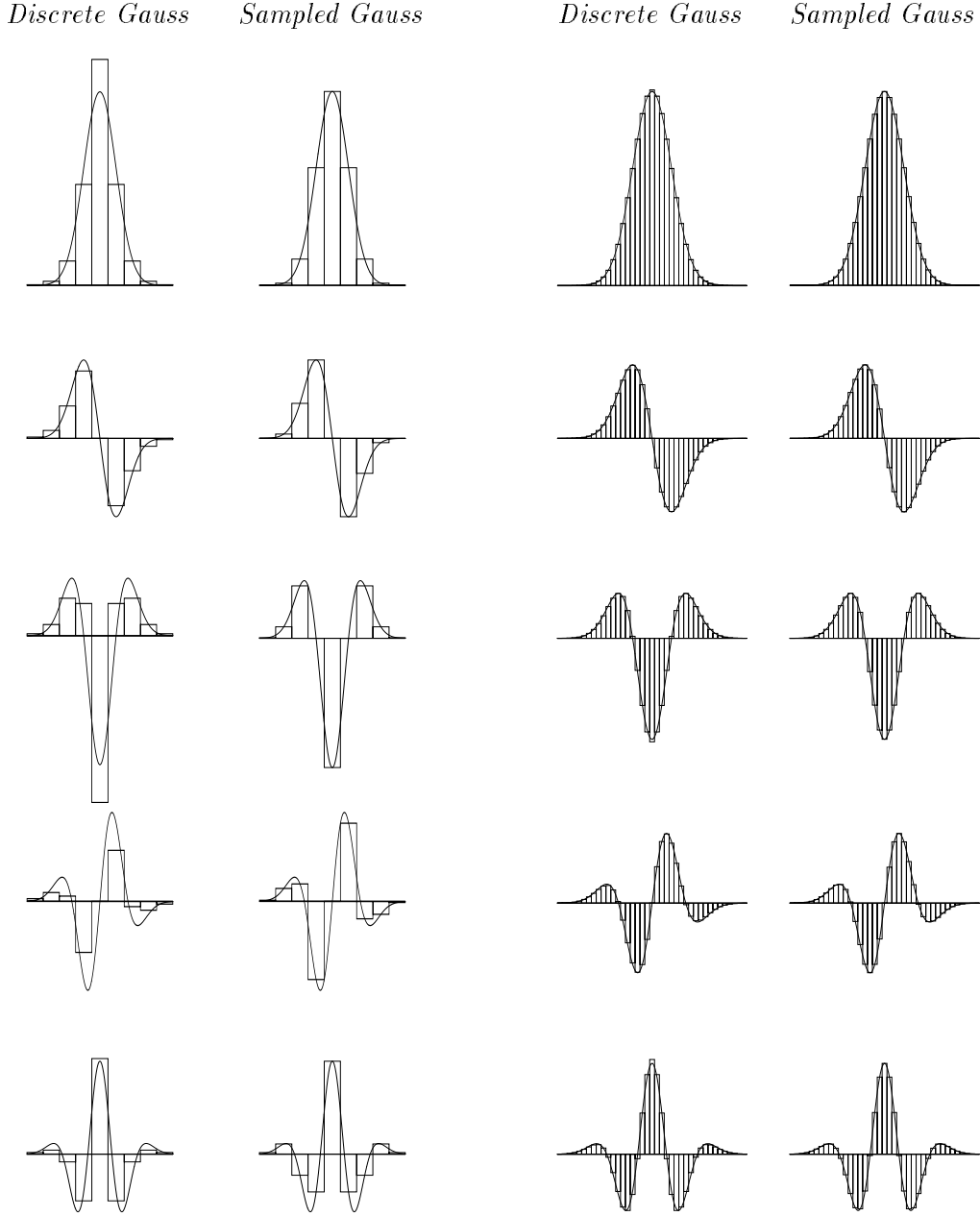
Figure 10: Graphs of the one-dimensional Gaussian derivative kernels $\partial_{x^n} g(x; \ t)$, and their corresponding discrete analogues $\delta_{x^n} T(x; \ t)$ up to order $n = 4$. The scale value in the left column is $t = 1.0$ and in the right column $t = 16.0$. The derivative/difference order increases from top to bottom. The upper row shows the raw smoothing kernel. Then follow the first, second, third and fourth order derivative/difference kernels. The block diagrams indicate the discrete kernels and the smooth curve the continuous Gaussian.
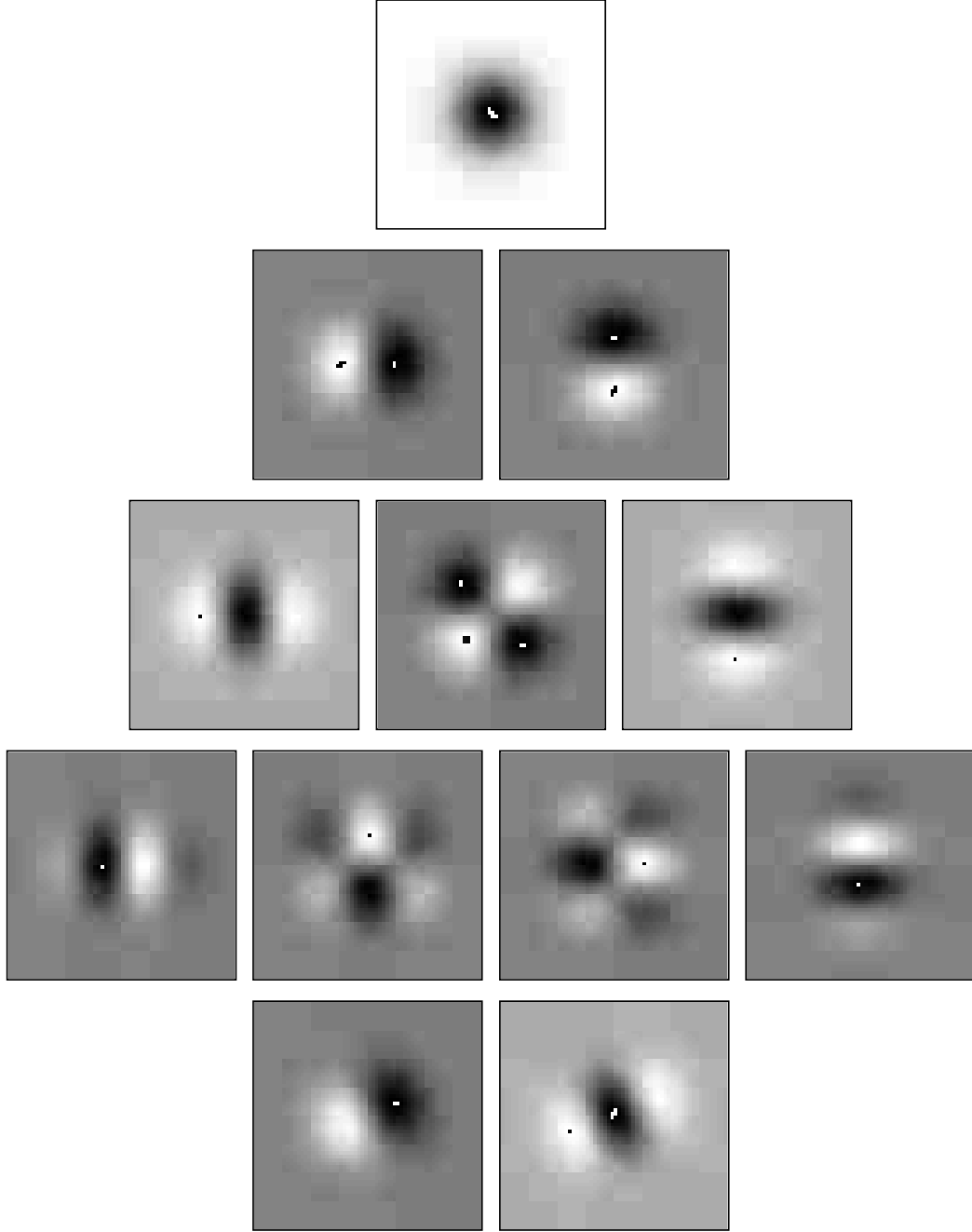
Figure 11: Grey-level illustrations of the *equivalent* two-dimensional discrete derivative approximation kernels up to order three (in the separable case corresponding to $\gamma = 0$). **(row 1)** (a) Zero-order smoothing kernel, T, (inverted). **(row 2)** (b-c) First order derivative approximation kernels, $\delta_x T$ and $\delta_y T$. **(row 3)** (d-f) Second order derivative approximation kernels $\delta_{xx} T$, $\delta_{xy} T$, $\delta_{yy} T$. **(row 4)** (g-j) Third order derivative approximation kernels $\delta_{xxx} T$, $\delta_{xxy} T$, $\delta_{xyy} T$, $\delta_{yyy} T$. **(row 5)** (k-l) First and second order directional derivative approximation kernels computed from the well-known expression for the directional derivative operator in direction α, $\partial_\alpha = \cos \alpha \partial_x + \sin \alpha \partial_y$, here in the direction 22.5°. (Scale level $t = 64.0$, image size 127×127 pixels).

and if similar continuity requirements concerning scale variations are imposed, then by necessity also the derivative approximations must satisfy the semi-discretized diffusion equation (51). The specific choice of operators δ_{x^n} is however arbitrary; any linear operator satisfies this relation. Graphs of these kernels at a few levels of scale and for the lowest orders of differentiation are shown in Figure 10 and Figure 11.

To summarize, there is a unique and consistent way to define a scale-space representation and discrete analogues to smoothed derivatives for discrete signals, which to a large extent preserves the algebraic structure of the multi-scale N-jet representation in the continuous case.

4.9 Summary and retrospective

As we have seen, the uniqueness of the Gaussian kernel for scale-space representation can be derived in a variety of different ways, non-creation of new level curves in scale-space, non-creation of new local extrema, non-enhancement of local extrema, and scale invariance. Similar formulations can be stated both in the spatial domain and in the frequency domain. The essence of these results is that the scale-space representation 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.

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 any further specification of the purpose of the representation, and what tasks the visual system are 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 proposed Gaussian derivative model. For example, it is well-known that scale-space smoothing leads to shape distortions at edges by smoothing across object boundaries, and also in surface orientation estimates computed by algorithms like shape from texture. Hence, it should be emphasized that the theory developed here 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 in which no or very little a priori information is available. Then, once initial hypotheses about the structure of the world have been generated within this framework, the intention is that it should be possible to invoke more refined processing, which can compensate for this, and adapt to current situation and the task at hand; see Section 10 for some indications about work in this direction. From the viewpoint of such non-uniform scale-space approaches, the linear scale-space model based on the rotationally symmetric Gaussian kernel provides a natural starting point for such analysis.

In fact, a certain degree of agreement[6] can be obtained with the result from this solely theoretical analysis and the experimental results of biological evolution.

---

[6]Another interesting similarity concerns the spatial layout of receptive fields over the visual field. If the scale-space axioms are combined with the assumption of a fixed readout capacity from the visual front end, then it is straightforward to show that there is a natural distribution of receptive fields (of different scales and different spatial position) over the retina such that the minimum receptive field size grows linearly with eccentricity, that is the distance from the center of the visual field (Lindeberg and Florack 1992). There are several results in psychophysics, neuroanatomy and electro-physiology in agreement with such a linear increase (Koenderink and van Doorn 1978; van de Grind et al. 1986; Bijl 1991).

Neurophysiological studies by Young (1985, 1987) have shown that there are receptive fields in the mammalian retina and visual cortex, whose measured response profiles can be very well modelled by Gaussian derivatives. For example, Young models cells in the mammalian retina by kernels termed differences of offset Gaussians (DOOG), which basically correspond to the Laplacian of the Gaussian with an added Gaussian offset term. He also reports cells in the visual cortex, whose receptive field profiles agree with Gaussian derivatives up to order four.

Of course, far-reaching conclusions should not be drawn from such a qualitative similarity, since there are also other functions, like Gabor functions that satisfy the recorded data up to the tolerance of the measurements. Nevertheless, it is interesting to note that operators similar to the Laplacian of the Gaussian have been reported to be dominant in the retina. A possible explanation concerning the construction of derivatives of other orders from the output of these operators can be obtained from the observation that the original scale-space representation can always be reconstructed from this data if Laplacian derivatives are available at all other scales. If the scale-space representation tends to zero at infinite scale, then it follows from the diffusion equation that

$$L(x;\ t) = -(L(x;\ \infty) - L(x;\ t)) = - \int_{t'=t}^{\infty} \partial_t L(x;\ t')dt' = - \int_{t'=t}^{\infty} \nabla^2 L(x;\ t')dt'. \tag{55}$$

Observe the similarity with the method (6) for reconstructing the original signal from a bandpass pyramid.

What remains to be understood is if there are any particular theoretical advantages of computing the Laplacian of the Gaussian in the first step. Of course, such an operation suppresses any linear illumination gradients. We summarize by contending that spatial derivatives of the Gaussian can be approximated by differences of Gaussian kernels at different spatial position, and it is therefore, at least in principle, possible to construct any spatial derivative from this representation. Remaining questions concerning the plausibility concerning biological vision are left to the reader's speculation.

It will now be described, first how the scale-space model relates to some other types of representations with multi-scale interpretation, and then how the scale-space model can be used for expressing early visual operations.

5 Related multi-scale representations

5.1 Wavelets

A type of multi-scale representation that has attracted a great interest in both signal processing, numerical analysis, and mathematics during recent years is *wavelet representation*, which dates back to Strömberg (1983) and Meyer (1988). A (two-parameter) family of translated and dilated (scaled) functions

$$h_{a,b}(x) = |a|^{-1/2} h(\frac{x-b}{a}) \quad a, b \in \mathbb{R}, a \neq 0 \tag{56}$$

defined from a single function $h : \mathbb{R} \to \mathbb{R}$ is called a *wavelet*. Provided that h satisfies certain admissibility conditions

$$\int_{\omega=-\infty}^{\infty} \frac{|\hat{h}(\omega)|^2}{|\omega|} d\omega < \infty, \tag{57}$$

then the representation $\mathcal{W}f : \mathbb{R}\backslash\{0\} \times \mathbb{R} \to \mathbb{R}$ given by

$$(\mathcal{W}f)(a,b) = < f, h_{a,b} > = |a|^{-1/2} \int_{x \in \mathbb{R}} f(x) h(\frac{x-b}{a}) dx \qquad (58)$$

is called the *continuous wavelet transform* of $f : \mathbb{R} \to \mathbb{R}$. From this background, scale-space representation can be considered as a *special case of continuous wavelet representation*, where the scale-space axioms imply that the function h *must* be selected as a derivative of the Gaussian kernel. In traditional wavelet theory, the zero order derivative is not permitted; it does not satisfy the admissibility condition, which in practice implies that

$$\int_{x=-\infty}^{\infty} h(x) \, dx = 0. \qquad (59)$$

There are several developments of this theory concerning different special cases. A particularly well studied problem is the construction of *orthogonal wavelets* for discrete signals, which permit a compact non-redundant multi-scale representation of the image data. This representation was suggested for image analysis by Mallat (1989, 1992). We will not attempt to review any of that theory here. Instead, the reader is referred to the chapters by Kay (1993), and by Donoho and Johnston (1993) in this volume.

5.2 *Regularization*

According to Hadamard, a problem is said to be well-posed if: (i) a solution exists, (ii) the solution is unique, and (iii) the solution depends continuously on the input data. It is well-known that several problem in computer vision are ill-posed; one example is differentiation. A small disturbance in a signal, $f(x) \mapsto f(x) + \varepsilon \sin \omega x$, where ε is small and ω is large, can lead to an arbitrarily large disturbance in the derivative $f_x(x) \mapsto f_x(x) + \omega \varepsilon \cos \omega x$, provided that ω is sufficiently large relative to $1/\epsilon$.

Regularization is a technique that has been developed for transforming ill-posed problems into well-posed ones, see Tikhonov and Arsenin, (1977) for an extensive treatment of the subject. Torre and Poggio (1986) describe this issue with application to one of the most intensely studied subproblems in computer vision, edge detection, and develop how regularization can be used in this context. One example of regularization concerning the problem "given an operator $\mathcal{A}$ and data y find z such that $\mathcal{A}z = y$" is the transformed problem "find z that minimizes the following functional"

$$\min_z \; (1-\lambda)||\mathcal{A}z - y||^2 + \lambda||\mathcal{P}z||^2, \qquad (60)$$

where $\mathcal{P}$ is a stabilizing operator, and $\lambda \in [0,1]$ is a regularization parameter controlling the compromise between the degree of regularization of the solution and closeness to the given data. Variation of the regularization parameter gives solutions with different degree of smoothness; a large value of λ may give rise a smooth solution, while a small value increases the accuracy at the cost of larger variations in the estimate. Hence, this parameter has a certain interpretation in terms of spatial scale in the result. (It should be observed, however, that the solution to the regularized problem

is in general not a solution to the original problem, not even in the case of ideal noise-free data.)

In the special case when $\mathcal{P} = \partial_{xx}$, and the measured data points are discrete, the solution of the problem of finding $S : \mathbb{R} \to \mathbb{R}$ that minimizes

$$\min_{S} \ (1 - \lambda) \sum (f_i - S(x_i))^2 + \lambda \int |S_{xx}(x_i)|^2 dx \tag{61}$$

given a set of measurements f_i is given by approximating cubic splines; see de Boor (1978) for an extensive treatment of the subject. Interestingly, this result was first proved by Schönberg (1946), who also proved the classification of Pólya frequency functions and sequences, which are the natural concepts in mathematics that underlie the scale-space kernels considered in previous sections. Torre and Poggio made the observation that the corresponding smoothing filters are very close to Gaussian kernels.

The strong regularization property of scale-space representation can be appreciated in the introductory example. Under a small high-frequency disturbance in the original signal $f(x) \mapsto (x) + \varepsilon \cos \omega x$, the propagation of the disturbance to the first order derivative of the scale-space representation is given by $L_x(x; \ t) \mapsto L_x(x; \ t) + \varepsilon \, \omega e^{\omega^2 t/2} \cos \omega x$. Clearly, this disturbance can be made arbitrarily small provided that the derivative of the signal is computed at a sufficiently coarse scale t in scale-space.

6 Multi-scale feature detection in scale-space

The above treatment gives formal justifications for using linear filtering as an initial step in early processing of image data. More importantly, it provides a catalogue of what filter kernels are natural to use, as well as an extensive theoretical explanation of how different kernels of different order and at different scales can be related. This forms the basis of a theoretically well-founded modelling of the smoothing operation.

Of course, linear filtering cannot be used as the only component in a vision system aimed at deriving symbolic representations from images; some non-linear steps must be introduced into the analysis. More concretely, some mechanism is required for combining the output from the Gaussian derivative operators of different order and at different scales into some more explicit descriptors of the image geometry.

6.1 Differential geometry and differential invariants

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. A basic motivation for this position is that differential equations and differential geometry constitute natural frameworks for expressing both physical processes and geometric properties. More technically, and as we have seen in section 4.5.2, it can also be shown that spatial derivatives are natural operators to derive from the scale-space representation.

When using such descriptors, it should be observed that a single partial derivative, e.g. L_{x_1}, does not represent any geometrically meaningful information, since its value is crucially dependent on the arbitrary choice of coordinate system. In other words, it is essential to base the analysis on descriptors that do not depend on the actual coordinatization of the spatial and intensity domains. Therefore, it is natural to require

the representation to be invariant with respect to primitive transformations such as translations, rotations, scale changes, and certain intensity transformations. In fact, it would be desirable to directly compute features that are invariant under perspective transformations. Since, however, this problem is known to be much harder, most work has so far been restricted to invariants of two-dimensional Euclidean operations and natural linear extensions thereof, like uniform rescaling and affine transformations of the spatial coordinates.

As we shall also see indications of below, quite a few types of low-level operations can expressed in terms of such multi-scale differential invariants defined from (non-linear) combinations of Gaussian derivatives at multiple scales. Examples of these are feature detectors, feature classification methods, and primitive shape descriptors. In this sense, the scale-space representation can serve as a useful basis for expressing a large number of early visual operations.

Florack et al (1992, 1993) and Kanatani (1990) have pursued this approach of deriving differential invariants in an axiomatic manner, and considered image properties defined in terms of directional derivatives along certain preferred coordinate directions. If the direction, along which a directional derivative is computed, can be uniquely defined from the intensity pattern, then rotational invariance is obtained automatically, since the preferred direction follows any rotation of the coordinate system. Similarly, any derivative is translationally invariant. These properties hold both concerning transformations of the original signal f and the scale-space representation L of f generated by smoothing with the rotationally symmetric Gaussian.

Detailed studies of differential geometric properties of two-dimensional and three-dimensional scalar images are presented by Saldens et al (1992), who makes use of classical techniques from differential geometry (Spivak 1975; Koenderink 1990), algebraic geometry, and invariant theory (Grace and Young 1965; Weyl 1946) for classifying geometric properties of the N-jet of a signal at a given scale in scale-space.

Here, a short description will be given concerning some elementary results. Although the treatment will be restricted to the two-dimensional case, the ideas behind it are general and can be easily extended to higher dimensions.

6.1.1 Local directional derivatives

One choice of preferred directions is to introduce a local orthonormal coordinate system (u, v) at any point P_0, where the v-axis is parallel to the gradient direction at P_0, and the u-axis is perpendicular, i.e. $e_v = (\cos \alpha, \sin \alpha)^T$ and $e_u = (\sin \alpha, -\cos \alpha)^T$, where

$$e_v|_{P_0} = \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} = \frac{1}{\sqrt{L_x^2 + L_y^2}} \begin{pmatrix} L_x \\ L_y \end{pmatrix} \Bigg|_{P_0}. \tag{62}$$

In terms of Cartesian coordinates, which arise frequently in standard digital images, these local directional derivative operators can be written

$$\partial_{\bar{u}} = \sin \alpha \partial_x - \cos \alpha \partial_y. \quad \partial_{\bar{v}} = \cos \alpha \partial_x + \sin \alpha \partial_y, \tag{63}$$

This coordinate system is characterized by the fact that one of the first order directional derivatives, $L_{\bar{u}}$, is zero.

Another natural choice of coordinate system is a (p, q) system aligned to the principal curvature directions. To express directional derivative operators along these

coordinates, which are characterized by $L_{\bar{p}\bar{q}} = 0$, we can choose to rotate the coordinate system by an angle β defined by

$$
\begin{aligned}
\cos\beta|_{P_0} &= \left. \sqrt{\frac{1}{2}\left(1 + \frac{L_{xx} - L_{yy}}{\sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}}\right)}\,\right|_{P_0}, \\[2ex]
\sin\beta|_{P_0} &= \left. \sqrt{\frac{1}{2}\left(1 - \frac{L_{xx} - L_{yy}}{\sqrt{(L_{xx} - L_{yy})^2 + 4L_{xy}^2}}\right)}\,\right|_{P_0},
\end{aligned}
\tag{64}
$$

and define explicit expressions for unit vectors by $e_{\bar{p}} = (\cos\beta, \sin\beta)$ and $e_{\bar{q}} = (\sin\beta, -\cos\beta)$. It is straightforward to verify that this definition implies that

$$
\begin{aligned}
L_{\bar{p}\bar{q}} &= \partial_{\bar{p}}\partial_{\bar{q}} L = (\cos\beta\,\partial_x + \sin\beta\,\partial_y)(\sin\beta\,\partial_x - \cos\beta\,\partial_y)L \\
&= \cos\beta\sin\beta(L_{xx} - L_{yy}) - (\cos^2\beta - \sin^2\beta)L_{xy} = 0.
\end{aligned}
\tag{65}
$$

6.1.2 *Monotonic intensity transformations*

One approach to deriving differential invariants is by requiring the differential entities to be invariant with respect to arbitrary *monotonic intensity transformations*. Then, any property that can be expressed in terms of the level curves of the signal is guaranteed to be invariant. A classification by Florack et al (1992) and Kanatani (1990), which goes back to the classical classification of polynomial invariants by Hilbert (1893), shows that concerning derivatives up to order two, there are only two irreducible differential expressions that are invariant to these transformations.

$$
\kappa = \frac{L_x^2 L_{yy} + L_y^2 L_{xx} - 2L_x L_y L_{xy}}{(L_x^2 + L_y^2)^{3/2}},
\tag{66}
$$

$$
\mu = \frac{(L_x^2 - L_y^2)L_{xy} - L_x L_y(L_{yy} - L_{xx})}{(L_x^2 + L_y^2)^{3/2}}.
\tag{67}
$$

Here, κ is the curvature of level curves in the smoothed signal, and μ the curvature of the integral paths of the gradient vectors. A general scheme for extending this technique to higher order derivatives and arbitrary dimensions has been proposed by Florack et al (1993).

6.1.3 *Affine intensity transformations*

Another approach is restrict the invariance to affine intensity transformations. Then, the class of invariants becomes larger. A natural condition to impose is that a differential expression $\mathcal{D}L$ should (at least) be a *relative invariant* with respect to scale changes, i.e., under a rescaling of the spatial coordinates, $L'(x)\hat{=}L(sx)$, the differential entity should transform as $\mathcal{D}L' = s^k \mathcal{D}L$ for some k. Trivially, this relation holds for any product of mixed directional derivatives, and extends to sums (and rational functions) of such expressions provided that the sum of the orders of differentiation is the same for any product of derivatives constituting one term in a sum.

In order to give a formal description of this, let $L_{\bar{u}^m \bar{v}^n} = L_{\bar{u}^\alpha}$ denote a mixed directional derivative of order $|\alpha| = m + n$, and let $\mathcal{D}$ be a (possibly non-linear) homogeneous differential expression of the form

$$\mathcal{D}L = \sum_{i=1}^{I} c_i \prod_{j=1}^{J} L_{\bar{u}^{\alpha_{ij}}}, \tag{68}$$

where $|\alpha_{ij}| > 0$ for all $i = [1..I]$ and $j = [1..J]$, and

$$\sum_{j=1}^{J} |\alpha_{ij}| = N \tag{69}$$

for all $i \in [1..I]$. Then, $\mathcal{D}L$ is invariant with respect to translations, rotations, and affine intensity transformations, and relative invariant to uniform rescalings of the spatial coordinates. (Expressions of this form can be more compactly written using tensor notation; see Florack et al).

6.2 Feature detection from differential singularities

The *singularities* (zero-crossings) of such expressions play an important role (Lindeberg 1992). This is a special case of a more general principle of using zero-crossings of differential geometric expressions for describing geometric features; see e.g. Bruce and Giblin (1984) for an excellent tutorial. If a feature detector can be expressed as a zero-crossing of such a differential expression (or a combination), then the feature will also be absolute invariant to uniform rescalings of the spatial coordinates, i.e. size changes. Formally, this invariance property can be expressed as follows:

Let $\mathcal{S}_{\mathcal{D}}L$ denote the *singularity set* of a differential operator of the form (68), i.e. $\mathcal{S}_{\mathcal{D}}L = \{(x;\, t) \in \mathbb{R}^2 \times \mathbb{R}_+ : \mathcal{D}L(x;\, t) = 0\}$, and let $\mathcal{G}$ be the Gaussian smoothing operator, i.e., $L = \mathcal{G}f$. Under these transformations of the spatial domain (represented by $x \in \mathbb{R}^2$) and the intensity domain (represented by either the unsmoothed f or the smoothed L) the singularity sets[7] transform as follows:

Transformation	Definition			Invariance		
translation	$(\mathcal{T}L)(x;\, t)$	$=$	$L(x + \Delta x;\, t)$	$\mathcal{S}_{\mathcal{D}}\,\mathcal{G}\,\mathcal{T}\,f =$	$\mathcal{S}_{\mathcal{D}}\,\mathcal{T}\,\mathcal{G}\,f$	$= \mathcal{T}\,\mathcal{S}_{\mathcal{D}}\,\mathcal{G}\,f$
rotation	$(\mathcal{R}L)(x;\, t)$	$=$	$L(Rx;\, t)$	$\mathcal{S}_{\mathcal{D}}\,\mathcal{G}\,\mathcal{R}\,f =$	$\mathcal{S}_{\mathcal{D}}\,\mathcal{R}\,\mathcal{G}\,f$	$= \mathcal{R}\,\mathcal{S}_{\mathcal{D}}\,\mathcal{G}\,f$
uniform scaling	$(\mathcal{U}L)(x;\, t)$	$=$	$L(sx;\, t)$	$\mathcal{S}_{\mathcal{D}}\,\mathcal{G}\,\mathcal{U}\,f =$	$\mathcal{S}_{\mathcal{D}}\,\mathcal{U}\,\mathcal{G}\,f$	$= \mathcal{U}\,\mathcal{S}_{\mathcal{D}}\,\mathcal{G}\,f$
affine intensity	$(\mathcal{A}L)(x;\, t)$	$=$	$aL(x;\, t) + b$	$\mathcal{S}_{\mathcal{D}}\,\mathcal{G}\,\mathcal{A}\,f =$	$\mathcal{S}_{\mathcal{D}}\,\mathcal{A}\,\mathcal{G}\,f$	$= \mathcal{S}_{\mathcal{D}}\,\mathcal{G}\,f$

In other words, feature detectors formulated in terms of differential singularities by definition commute with a number of elementary transformations of the spatial and intensity domains, and it does not matter whether the transformation is performed before or after the smoothing step. Some simple examples of feature detectors that can be expressed in this way are listed below.

---

[7] Here, R is a rotation matrix, Δx is a vector ($\in \mathbb{R}^2$), while a, b and s are scalar constants. The definitions of the transformed singularity sets are as follows; $\mathcal{T}\mathcal{S}_{\mathcal{D}}L = \{(x;\, t) : \mathcal{D}L(x + \Delta x;\, t) = 0\}$, $\mathcal{R}\mathcal{S}_{\mathcal{D}}L = \{(x;\, t) : \mathcal{D}L(Rx;\, t) = 0\}$, and $\mathcal{U}\mathcal{S}_{\mathcal{D}}L = \{(x;\, t) : \mathcal{D}L(sx;\, \underline{s}^2 t) = 0\}$.

6.2.1 Examples of feature detectors

A natural way to define *edges* from a continuous grey-level image $L : \mathbb{R}^2 \to \mathbb{R}$ is as the union of the points for which the gradient magnitude assumes a maximum in the gradient direction. This method is usually referred to as "non-maximum suppression", (see e.g. Canny (1986), or Korn (1988)). Assuming that the second and third order directional derivatives of L in the v-direction are not simultaneously zero, a necessary and sufficient condition for P_0 to be a gradient maximum in the gradient direction may be stated as:

$$\begin{cases} L_{\bar{v}\bar{v}} = 0, \\ L_{\bar{v}\bar{v}\bar{v}} < 0. \end{cases} \tag{70}$$

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

$$\begin{cases} \tilde{L}_{\bar{v}\bar{v}} = & L_{\bar{v}}^2 L_{\bar{v}\bar{v}} = & L_x^2 L_{xx} + 2 L_x L_y L_{xy} + L_y^2 L_{yy} = 0, \\ \tilde{L}_{\bar{v}\bar{v}\bar{v}} = & L_{\bar{v}}^3 L_{\bar{v}\bar{v}\bar{v}} = & L_x^3 L_{xxx} + 3 L_x^2 L_y L_{xxy} + 3 L_x L_y^2 L_{xyy} + L_y^3 L_{yyy} < 0. \end{cases} \tag{71}$$

Interpolating for zero-crossings of $\tilde{L}_{\bar{v}\bar{v}}$ within the sign-constraints of $\tilde{L}_{\bar{v}\bar{v}\bar{v}}$ gives a straightforward method for sub-pixel edge detection (Lindeberg 1993).

An entity commonly used for *junction detection* is the curvature of level curves in intensity data, see e.g. Kitchen (1982) or Koenderink and Richards (1988). In terms of directional derivatives it can be expressed as

$$\kappa = \frac{L_{\bar{u}\bar{u}}}{L_{\bar{v}}}. \tag{72}$$

In order to to give a stronger response near edges, the level curve curvature is usually multiplied by the gradient magnitude $L_{\bar{v}}$ raised to some power k. A natural choice is $k = 3$. This leads to a polynomial expression, see e.g. Brunnström et al. (1992),

$$|\tilde{\kappa}| = |L_{\bar{v}}^2 L_{\bar{u}\bar{u}}| = |L_y^2 L_{xx} - 2 L_x L_y L_{xy} + L_x^2 L_{yy}|. \tag{73}$$

Since the sum of the order of differentiation with respect to x and y is the same for all terms in this sum, it follows that junction candidates given by extrema in $\tilde{\kappa}$ also are skew invariant (Blom 1992). Assuming that the first- and second-order differentials of $\tilde{\kappa}$ are not simultaneously degenerate, a necessary and sufficient condition for a point P_0 to be a maximum in this rescaled level curve curvature is that:

$$\begin{cases} \partial_{\bar{u}}(\tilde{\kappa}) = 0, \\ \partial_{\bar{v}}(\tilde{\kappa}) = 0, \\ \mathcal{H}(\tilde{\kappa}) = \tilde{\kappa}_{\mathcal{H}} = \tilde{\kappa}_{\bar{u}\bar{u}} \tilde{\kappa}_{\bar{v}\bar{v}} - \tilde{\kappa}_{\bar{u}\bar{v}}^2 > 0, \\ \mathrm{sign}(\tilde{\kappa}) \tilde{\kappa}_{\bar{u}\bar{u}} < 0. \end{cases} \tag{74}$$

Interpolating for simultaneous zero-crossings in $\partial_{\bar{u}}(\tilde{\kappa})$ and $\partial_{\bar{u}}(\tilde{\kappa})$ gives a sub-pixel junction detector. Zero-crossings of the Laplacian

$$\nabla^T \nabla L = L_{\bar{u}\bar{u}} + L_{\bar{v}\bar{v}} = L_{xx} + L_{yy} = 0 \tag{75}$$

have been used for *stereo matching* (see, e.g., Marr (1982)) and *blob detection* (see, e.g., Blostein and Ahuja (1987)). Blob detection methods can also be formulated in terms of local extrema (see, e.g., Lindeberg and Eklundh (1990, 1991)).

Zero-crossings of the Laplacian have been used also for edge detection, although the localization becomes quite poor at curved edges. This can be understood from the relation between the Laplace operator and the second derivative in the gradient direction

$$\nabla^T \nabla L = L_{\bar{u}\bar{u}} + L_{\bar{v}\bar{v}} = L_{\bar{v}\bar{v}} + \kappa L_{\bar{v}}. \tag{76}$$

This example constitutes a simple indication of how theoretical analysis of feature detectors becomes tractable when expressed in terms of the suggested differential geometric framework.

There are several ways to define ridges from intensity data. One way is to start from the (p, q)-system considered in section 6.1.1 and to state the requirement that for a point to be regarded as belonging to a bright ridge it must satisfy

$$\left\{ \begin{array}{rcl} L_{\bar{p}} & = & 0, \\ L_{\bar{p}\bar{p}} & < & 0, \end{array} \right. \quad \text{or} \quad \left\{ \begin{array}{rcl} L_{\bar{q}} & = & 0, \\ L_{\bar{q}\bar{q}} & < & 0, \end{array} \right.$$

depending on whether the $\bar{p}$- or the $\bar{q}$-direction corresponds to the maximum absolute value of the principal curvature. It is straightforward to show that in terms of the (u, v)-system, this differential geometric ridge detector can equivalently be expressed as the zero-crossings of $L_{\bar{u}\bar{v}}$ that satisfy $L_{\bar{u}\bar{u}}^2 - L_{\bar{v}\bar{v}}^2 > 0$. The sign of $L_{\bar{u}\bar{u}}$ determines the polarity; $L_{\bar{u}\bar{u}} < 0$ corresponds to bright ridges, and $L_{\bar{u}\bar{u}} > 0$ to dark ridges.

(To derive these relationships, replace (x, y) by (u, v) in the expressions for $L_{\bar{p}}$, $L_{\bar{q}}$, $L_{\bar{p}\bar{p}}$, and $L_{\bar{q}\bar{q}}$ obtained by combining equations (61) and (62) (and use $L_{\bar{u}} = 0$). This gives,

$$L_{\bar{p}} = \sin \beta' \, L_{\bar{v}}, \quad L_{\bar{q}} = -\cos \beta' \, L_{\bar{v}},$$

where $\cos \beta'$ and $\sin \beta'$ represent the results of replacing (x, y) by (u, v) in (62). Then, the condition $L_{\bar{p}} L_{\bar{q}} = 0$ can be rewritten as $L_{\bar{u}\bar{v}} = 0$. Moreover, since in general $L_{\bar{v}} \neq 0$, it follows that $L_{\bar{p}} = 0$ corresponds to $\sin \beta' = 0$. Then, we have $L_{\bar{p}\bar{p}} = \cos^2 \beta' L_{\bar{u}\bar{u}}$ and $L_{\bar{q}\bar{q}} = \cos^2 \beta' L_{\bar{v}\bar{v}}$, which means that $L_{\bar{p}\bar{p}} < 0$ corresponds to $L_{\bar{u}\bar{u}} < 0$. The requirement for the p-direction to correspond to the maximum absolute value of the principal curvatures can be written $|L_{\bar{u}\bar{u}}| > |L_{\bar{v}\bar{v}}|$. Similar results are obtained for the case $L_{\bar{q}} = 0$.)

Examples of edge detection, junction detection and blob detection using this technique at a number of different scales in scale-space are shown in Figure 12. Observe how different structures in the image manifest themselves at different scales in scale-space, and how the complexity in the feature data decreases with scale. Figure 13 shows corresponding results with ridge detection. Examples of shape cue computation using the Gaussian derivative framework are given by Jones and Malik (1992), and Lindeberg and Gårding (1993).

7 Behaviour across scales: Deep structure

The treatment so far has been concerned with the formal definition of the scale-space representation, and the definition of image descriptors at any single scale. A very important problem concerns how to relate structures at different scales. This subject has been termed *deep structure* by Koenderink (1984). When a pattern is subjected to scale-space smoothing, its shape changes. This gives rise to the notion of "dynamic shape", which as argued by Koenderink and van Doorn (1986), is an essential component of any shape description of natural objects.

| scale-space representation | edges | junction candidates | dark grey-level blobs |

$t = 1$

$t = 4$
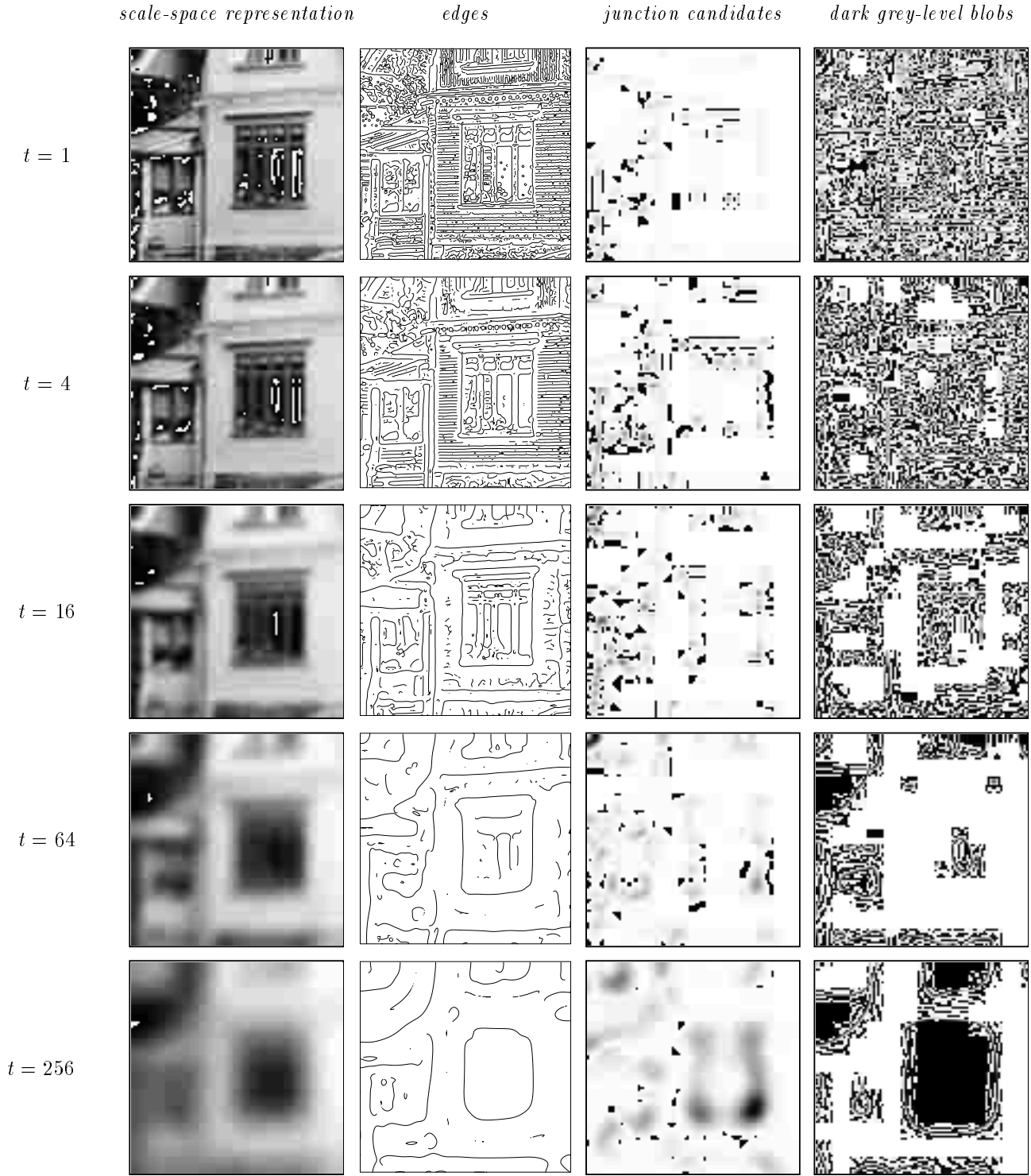
$t = 16$

$t = 64$

$t = 256$

Figure 12: Example of multi-scale feature detection in scale-space using singularities of differential invariants. (left) Smoothed grey-level images. (middle left) Edges defined by $L_{\bar{v}\bar{v}} = 0$ and $L_{\bar{v}\bar{v}\bar{v}} < 0$. (middle right) Magnitude of $\tilde{\kappa}$. (right) Dark grey-level blobs (every blob corresponds to a local minimum in grey-level). Scale levels from top to bottom: $t = 2$, 8, 32, and 128.

Figure 13: Examples of differential geometric ridge detection (without thresholding): (a)–(b) dark ridges from a detail from a telephone image at scale levels $t = 16$ and 64, (c)–(d) bright ridges from an aerial image at scale levels $t = 16$ and 64.

7.1 Iso-intensity linking

An early suggestion by Koenderink (1984) to relate structures at different scales was to identify points across scales that have the same grey-level and correspond to paths of steepest ascent along level surfaces in scale-space.

Since the tangent vectors of such paths must be in the tangent plane to the level surface, and the spatial component must be parallel to (L_x, L_y), these *iso-intensity paths* are the integral paths of the vector field

$$v = (L_x L_t, L_y L_t, -(L_x^2 + L_y^2)). \tag{77}$$

Lifshitz and Pizer (1990) considered such paths in scale-space, and constructed a multi-scale "stack" representation, in which the grey-level at which an extremum disappeared was used for defining a region in the original image by local thresholding on that grey-level.

Although the representation was demonstrated to be applicable for certain segmentation problems in medical image analysis, Lifshitz and Pizer observed the serious problem of *non-containment*, which basically means that a point, which at one scale has been classified as belonging to a certain region (associated with a local maximum), can escape from that region when the scale parameter increases. Moreover, such paths can be intertwined in a rather complicated way.

7.2 Feature based linking (differential singularities)

The main cause to problem in the iso-intensity linking is that grey-levels corresponding to a features tracked over scales change under scale-space smoothing. For example, concerning a local extremum it is a necessary consequence of the diffusion equation that the grey-level at the maximum point must decrease with scale. For this reason, it may be more natural to identify *features* across scales rather than grey-levels. A type of representation defined in this way is the *scale-space primal sketch* of blob-like image structures (extrema with extent) defined at all scales in scale-space and linked into a tree-like data structure (Lindeberg 1991, 1992).

More generally, consider a feature, which at any level of scale can be defined by

$$h(x; \ t) = 0 \quad (x \in \mathbb{R}^N, t \in \mathbb{R}_+) \tag{78}$$

38

for some function $h : \mathbb{R}^N \times \mathbb{R}_+ \rightarrow \mathbb{R}^M$. For example, the differential singularities treated in Section 6.2 are of this form. Using the implicit function theorem it is then formally easy to analyze the dependence of x on t in the solution to (78). Here, some simple examples will be presented of how such analysis can be performed; see Lindeberg (1992) for a more extensive treatment. Consider, for simplicity, data given as two-dimensional images. Then, it is sufficient to study the cases when M is either 1 or 2.

7.2.1 Pointwise entities

If $M = 2$ then then the features will in general be isolated points. The implicit function theorem states that these points form smooth paths across scales (one-dimensional curves in three-dimensional scale-space) provided that the Jacobian $\partial_x h$ is non-degenerate. The drift velocity along such a path can be written

$$\partial_t x = -(\partial_x^T h)^{-1} \partial_t h. \tag{79}$$

Concerning *critical points* in the grey-level landscape, we have $h = (L_x, L_y)^T$, and the explicit expression for the drift velocity can be written

$$\partial_t x = -\frac{1}{2}(\mathcal{H}L)^{-1} \nabla^T \nabla(\nabla L), \tag{80}$$

where $\mathcal{H}L$ denotes the Hessian matrix of L, and the fact that the spatial derivatives satisfy the diffusion equation has been used for expressing derivatives of L with respect to t in terms of derivatives with respect to x. A similar analysis can be performed concerning, e.g. junctions given as maxima in $\tilde{\kappa}$, although the expressions then contain derivatives up to order five.

This result gives an estimate of the drift velocity of the features due to scale-space smoothing, and provides a theoretical basis for relating and, hence, linking corresponding features across scales in a well-defined manner.

7.2.2 Curve entities

If $M = 1$, then the set of feature points will in general be curves when treated at a single scale and surfaces when treated at all scales. Hence, there is no longer any unique correspondence between points at adjacent scales. This ambiguity is similar to the so-called "aperture problem" in motion analysis. Nevertheless, the normal component of the drift can be determined. If s represents a coordinate along the normal direction, then the drift velocity can be expressed as

$$\partial_t s = - \tilde{h}_s^{-1} \tilde{h}_t = - \frac{h_t}{|\nabla h|} . \tag{81}$$

For example, concerning an *edge* given by non-maximum suppression ($\alpha = \tilde{L}_{\bar{v}\bar{v}} = 0$), the drift velocity in the normal direction assumes the form

$$(\partial_t u, \partial_t v) = -\frac{L_{\bar{v}}(L_{\bar{u}\bar{u}\bar{v}\bar{v}} + L_{\bar{v}\bar{v}\bar{v}\bar{v}}) + 2L_{\bar{u}\bar{v}}(L_{\bar{u}\bar{u}\bar{u}} + L_{\bar{u}\bar{v}\bar{v}})}{2((L_{\bar{v}}L_{\bar{u}\bar{v}\bar{v}} + 2L_{\bar{u}\bar{v}}L_{\bar{u}\bar{u}})^2 + (L_{\bar{v}}L_{\bar{v}\bar{v}\bar{v}} + 2L_{\bar{u}\bar{v}}^2)^2)}(\frac{\alpha_{\bar{u}}}{L_{\bar{v}}}, \frac{\alpha_{\bar{v}}}{L_{\bar{v}}}), \tag{82}$$

where

$$\begin{aligned} \alpha_{\bar{u}} &= L_{\bar{v}}^2 L_{\bar{u}\bar{v}\bar{v}} + 2L_{\bar{v}}L_{\bar{u}\bar{v}}L_{\bar{u}\bar{u}}, \\ \alpha_{\bar{v}} &= L_{\bar{v}}^2 L_{\bar{v}\bar{v}\bar{v}} + 2L_{\bar{v}}L_{\bar{u}\bar{v}}^2, \end{aligned} \tag{83}$$

represent the components of the normal vector $(\alpha_{\bar{u}}, \alpha_{\bar{v}})$ to the edge expressed in the (u, v) coordinate system. Unfortunately, this expression cannot be further simplified unless additional constraints are posed on L. For a *straight edge*, however, where all partial derivatives with respect to u are zero, it reduces to

$$(\partial_t u, \partial_t v) = -\frac{1}{2} \frac{L_{\bar{v}\bar{v}\bar{v}}}{L_{\bar{v}\bar{v}}}(0, 1). \tag{84}$$

This analysis can be used for stating a formal description of the *edge focusing* method developed by Bergholm (1987), in which edges are detected at a coarse scale and then tracked to finer scales; see also Clark (1988) and Lu and Jain (1989) concerning the behaviour of edges in scale-space.

Note the qualitative difference between linking across scales in the scale-space representation of a signal and the corresponding problem in a pyramid. In the first case, the linking process can be expressed in terms of differential equations, while in the second case it corresponds to a combinatorial matching problem. It is well-known that it is a hard algorithmic problem to obtain stable links across scales in a pyramid.

7.3 Bifurcations in scale-space

Previous section states that the scale linking is well-defined whenever the appropriate submatrix of the Jacobian of h is non-degenerate, When the Jacobian degenerates, *bifurcations* may occur.

Concerning critical points in the grey-level landscape, the situation is simple. In the one-dimensional case, the generic bifurcation event is the annihilation of a pair consisting of a local maximum and a minimum point, while in the two-dimensional case a pair consisting of a saddle point and an extremum can be both annihilated and created[8] with increasing scale. A natural model of this so-called *fold singularity* is the polynomial

$$L(x; \, t) = x_1^3 + 3x(t - t_0) + \sum_{i=1}^{N} \pm (x_i^2 + t - t_0), \tag{85}$$

which also satisfies the diffusion equation; see also Poston and Stewart (1978), Koenderink and van Doorn (1986), Lifshitz and Pizer (1990), and Lindeberg (1992). The positions of the critical points are given by

$$x_1(t) = \pm\sqrt{t_0 - t} \quad (x_i = 0 \; i > 1) \tag{86}$$

i.e. the critical points merge along a parabola, and the drift velocity tends to infinity at the bifurcation point.

Johansen (1993) gives a more detailed differential geometric study of such bifurcations, covering also a number of cases, which generically are unstable when treated in a single image. Under more general parameter variations, however, like in image sequences, such singularities can be expected to be stable in the sense that a small disturbance of the original signal causes the singular point to appear at a slightly different time moment.

---

[8] An example of a creation event is given in the end of Section 4.4.

8 Scale sampling

Although the scale-space concept comprises a continuous scale parameter, it is necessary to actually compute the smoothed representations at some discrete set of sampled scale levels. The fact that drift velocities may (momentarily) tend to infinity indicates that in general some mechanism for adaptive scale must be used. Such an approach also simplifies algorithms for tracking structures across scales. Under the assumption of generic signals a sufficiently fine scale sampling should make the scale matching problem trivial. This property substantially simplifies the algorithm for building the scale-space primal sketch (Lindeberg 1991).

8.1 *Effective scale*

Nevertheless, it is necessary to have some default strategy for scale selection when no information is available. Introduce a transformed scale parameter, *effective scale* τ, such that default scale sampling corresponds to uniform scale steps $\Delta\tau$ measured in this unit. Then, a straightforward scaling argument shows that for continuous signals, the transformation function expressing τ as function of t should be given by

$$\tau(t) = A + B \log t, \tag{87}$$

and that the ratio between successive scale values should be constant. This result can also be obtained directly from scale invariance and the Pi-theorem. If a dimensionless parameter τ is to be defined from t, then necessarily

$$d\tau = B \frac{dt}{t} \tag{88}$$

for some B, which gives the logarithmic transformation. Some more care must be taken if the lifetime of a structure in scale-space is to be used for measuring significance in discrete signals, since otherwise a structure existing at scale zero would be assigned an infinite lifetime. An analysis in (Lindeberg 1992) shows that a natural way to introduce such a scale parameter for discrete signals is by

$$\tau(t) = A + B \log p(t), \tag{89}$$

where $p(t)$ constitutes a measure of the "amount of structure" in a signal at scale t, for practical use the number of local extrema in a reference signal. It can be shown that under rather general conditions on a one-dimensional signal, $\tau(t)$ given by (89) reduces to (87), while for a discrete signal $\tau(t)$ is approximately linear at fine scales, and asymptotically approaches the logarithmic behaviour when t increases. Hence, the latter approach provides a well-defined way to model the transition from the genuine discrete behaviour at fine scales to coarser scales where the continuous approximation is valid.

9 Information content

Originating from the computational vision model by Marr and Hildreth (1980), in which zero-crossings of the Laplacian play a dominant role, substantial efforts have been spent on analysing the information content of those features in scale-space. Problems that have been treated concern whether the original signal can be reconstructed, solely using the evolution properties over scales of these zero-crossing curves.

The main result is that such reconstruction is possible (up to certain constants), although unstable unless regularized in some sense, see e.g. Yuille and Poggio (1988), and Hummel and Moniot (1989).

An interesting result is presented by Johansen et al (1986), who show that if also negative scales are considered, then a band-limited one-dimensional signal is up to a a multiplicative constant determined by its "top points", that is, the points in scale-space where bifurcations between critical points occur.

When considering the entire scale-space representation, it is, of course, obvious that the original signal can be "reconstructed" from the scale-space representation. The philosophy that underlies this presentation is to us the scale-space representation for making explicit certain aspects of the information content, rather than for deriving any "minimal and complete" representation, from which the signal can be reconstructed.

10   Non-uniform scale-space

Although the linear scale-space representation generated by smoothing with the rotationally symmetric Gaussian kernel provides a theoretically well-founded framework for handling image structures at different scale, the scale-space smoothing has the negative property that it leads to shape distortions. For example, smoothing across "object boundaries" can affect both the shape and the localization of edges in edge detection. Similarly, surface orientation estimates computed by shape from texture algorithms are affected, since the anisotropy of a surface pattern may decrease when smoothed using a rotationally symmetric Gaussian

To reduce these problems, Perona and Malik (1990) proposed the use of *anisotropic diffusion*. The basic idea is to modify the conductivity $c(x; t)$ in a non-linear version of the diffusion equation

$$\partial_t L = \frac{1}{2} \nabla^T (c(x; t)\nabla L) \tag{90}$$

such as to favour intra-region smoothing to inter-region smoothing. In principle, they solved the diffusion equation

$$\partial_t L = \frac{1}{2} \nabla^T (h(|\nabla L(x; t)|)\nabla L) \tag{91}$$

for some monotonic decreasing function $h: \mathbb{R}_+ \to \mathbb{R}_+$. The intuitive effect of this evolution is that the conductivity will be low where the gradient magnitude is high and vice versa.

This idea has been further developed by several authors. Nordström (1990) showed that by adding a bias term to the diffusion equation, it was possible to relate this method to earlier considered regularization approaches by Terzopoulos (1983) and Mumford and Shah (1985). Alternative modifications in terms of adaptive smoothing schemes have been presented by Saint-Marc *et al.* (1991), Nitzberg and Shiota (1992), and Whittaker and Pizer (1993). .

By adopting an axiomatic approach, Alvarez *et al.* have shown that given certain constraints on a visual front-end, a natural choice of non-linear diffusion equation is the equation

$$\partial_t L = \frac{1}{2} |\nabla L| \nabla^T (\nabla L/|\nabla L|) = \frac{1}{2} L_{\bar{u}\bar{u}}, \tag{92}$$

where $L_{\bar{u}\bar{u}}$ represents the second order derivative in the tangent direction to a level curve. This evolution means that level curves move in the normal direction with a velocity proportional to the curvature of the level curves. For a slightly modified version of (92),

$$\partial_t L = \frac{1}{2} \left(|\nabla L|^2 \, L_{\bar{u}\bar{u}} \right)^{1/3}, \tag{93}$$

the solutions are relative invariant under affine transformations of the spatial coordinates. This property has been used by Sapiro and Tannenbaum (1993) for defining an affine invariant curve evolution scheme.

An interesting approach to describing non-linear diffusion more generally has been pursued by Florack (1993) who considers general non-linear coordinate transformations of the spatial coordinates as a common framework for expressing such operations. Interestingly, this approach covers several of the above-mentioned methods.

Trivially, it holds (and follows from the maximum principle) that if $h > 0$ then any non-linear scale-space representation of the form (90) satisfies the causality requirement, or equivalently the non-enhancement property of local extrema: $\partial_t L < 0$ at local maxima, and $\partial_t L > 0$ at local minima.

Improvements relative to the rotationally symmetric scale-space representation can also be obtained using linear theory. As has been argued by several authors, it can be advantageous to use filters that correspond to different scale values along different directions; for example a large scale value along the direction of an edge, and a smaller scale value in the perpendicular direction. At junctions, where several directions meet, the converse behaviour can be advantageous. In shape from texture, Lindeberg (1994) and Lindeberg and Gårding (1994) demonstrate that linear shape adaption of the smoothing kernels can be used for improving the accuracy in surface orientation estimates, and in fact making a shape from texture method invariant with respect to the locally linearized perspective mapping.

Acknowledgements

References

M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions*. Applied Mathematics Series, National Bureau of Standards, 55 ed., 1964.

L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel, "Axioms and fundamental equations of image processing", tech. rep., Ceremade, Universit'e Paris-Dauphine, Paris, France, 1992.

J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, "Uniqueness of the Gaussian kernel for scale-space filtering", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 8, no. 1, pp. 26–33, 1986.

F. Bergholm, "Edge focusing", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 9, pp. 726–741, 1987.

P. Bijl, *Aspects of Visual Contrast Detection*. PhD thesis, University of Utrecht, University of Utrecht, Dept. of Med. Phys., Princetonplein 5, Utrecht, the Netherlands, May 1991.

J. Blom, *Topological and Geometrical Aspects of Image Structure*. PhD thesis, Dept. Med. Phys. Physics, Univ. Utrecht, NL-3508 Utrecht, Netherlands, 1992.

D. Blostein and N. Ahuja, "Representation and three-dimensional interpretation of image texture: An integrated approach", in *Proc. 1st Int. Conf. on Computer Vision*, (London), pp. 444–449, IEEE Computer Society Press, 1987.

C. de Boor, *A Practical Guide to Splines*, vol. 27 of *Applied Mathematical Sciences*. New York: Springer-Verlag, 1978.

J. W. Bruce and P. J. Giblin, *Curves and Singularities*. Cambridge: Cambridge University Press, 1984.

K. Brunnström, T. Lindeberg, and J.-O. Eklundh, "Active detection and classification of junctions by foveation with a head-eye system guided by the scale-space primal sketch", in *Proc. 2nd European Conf. on Computer Vision* (G. Sandini, ed.), vol. 588 of *Lecture Notes in Computer Science*, (Santa Margherita Ligure, Italy), pp. 701–709, Springer-Verlag, May. 1992.

P. J. Burt, "Fast filter transforms for image processing", *Computer Vision, Graphics, and Image Processing*, vol. 16, pp. 20–51, 1981.

P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code", *IEEE Trans. Communications*, vol. 9:4, pp. 532–540, 1983.

J. Canny, "A computational approach to edge detection", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 8, no. 6, pp. 679–698, 1986.

V. Cantoni and S. Levialdi, eds., *Pyramidal Systems for Computer Vision*. Berlin: Springer-Verlag, 1986.

A. Chehikian and J. L. Crowley, "Fast computation of optimal semi-octave pyramids", in *Proc. 7th Scandinavian Conf. on Image Analysis*, (Aalborg, Denmark), pp. 18–27, Aug. 1991.

J. J. Clark, "Singularity theory and phantom edges in scale-space", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 10, no. 5, pp. 720–727, 1988.

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

J. L. Crowley and A. C. Parker, "A representation for shape based on peaks and ridges in the Difference of Low-Pass Transform", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 6, no. 2, pp. 156–170, 1984.

J. L. Crowley and A. C. Sanderson, "Multiple resolution representation and probabilistic matching of 2-D gray-scale shape", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 9, no. 1, pp. 113–121, 1987.

J. L. Crowley and R. M. Stern, "Fast computation of the Difference of Low Pass Transform", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 6, pp. 212–222, 1984.

D. Donoho and I. Johnston, "Wavelet decomposition of image data: Theory and experience", in *Statistics and images* (K. Mardia, ed.), (Abington, Oxford), Carfax Publishing Co., 1993.

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

L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, "General intensity transformations", in *Proc. 7th Scandinavian Conf. on Image Analysis* (P. Johansen and S. Olsen, eds.), (Aalborg, Denmark), pp. 338–345, Aug. 1991.

L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, "Images: Regular tempered distributions", in *Proceedings NATO workshop 'Shape in Picture* (Y.O. Ying, A. Toet, and H. Heijmanns, eds.), NATO ASI Series F, (Driebergen, Netherlands), Springer Verlag, New York, September 1992.

L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, "Scale and the differential structure of images", *Image and Vision Computing*, vol. 10, pp. 376–388, Jul. 1992.

L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, "Cartesian differential invariants in scale-space", *J. of Mathematical Imaging and Vision*, vol. 3, no. 4, pp. 327–348, 1993.

L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, "General intensity transformations and differential invariants", *J. of Mathematical Imaging and Vision*, 1994. In press.

J. Gårding and T. Lindeberg, "Direct estimation of local surface shape in a fixating binocular vision system", in *Proc. 3rd European Conference on Computer Vision* (J.-O. Eklundh, ed.), vol. 800 of *Lecture Notes in Computer Science*, (Stockholm, Sweden), pp. 365–376, Springer-Verlag, May. 1994. (To appear).

J. H. Grace and A. Young, *Algebra of Invariants.* Bronx, New York: Chelsea Publishing Company, 1965.

W. A. van de Grind, J. J. Koenderink, and A. J. van Doorn, "The distribution of human motion detector properties in the monocular visual field", *Vision Research*, vol. 26, no. 5, pp. 797–810, 1986.

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

A. R. Hanson and E. M. Riseman, "Processing cones: A parallel computational structure for scene analysis", tech. rep., Computer and Information Science, Univ. of Massachusetts, Amherst, Massachusetts, 1974.

D. Hilbert, "Über die vollen Invariantensystemen", *Mathematische Annalen*, vol. 42, pp. 313–373, 1893.

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

I. I. Hirschmann and D. V. Widder, *The Convolution Transform.* Princeton, New Jersey: Princeton University Press, 1955.

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

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

R. A. Hummel and R. Moniot, "Reconstructions from zero-crossings in scale-space", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 37, no. 12, pp. 2111–2130, 1989.

P. Johansen, "On the classification of toppoints in scale space", *J. of Mathematical Imaging and Vision*, 1993. To appear.

P. Johansen, S. Skelboe, K. Grue, and J. D. Andersen, "Representing signals by their top points in scale-space", in *Proc. 8:th Int. Conf. on Pattern Recognition*, (Paris, France), pp. 215–217, Oct. 1986.

D. G. Jones and J. Malik, "A computational framework for determining stereo correspondences from a set of linear spatial filters", in *Proc. 2nd European Conf. on Computer Vision* (G. Sandini, ed.), vol. 588 of *Lecture Notes in Computer Science*, (Santa Margherita Ligure, Italy), pp. 395–410, Springer-Verlag, May. 1992.

K. Kanatani, *Group-Theoretical Methods in Image Understanding*, vol. 20 of *Series in Information Sciences*. Spatial Vision, 1990.

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

J. Kay, "Wavelets", in *Statistics and images* (K. Mardia, ed.), (Abington, Oxford), Carfax Publishing Co., 1993.

L. Kitchen and A. Rosenfeld, "Gray-level corner detection", *Pattern Recognition Letters*, vol. 1, no. 2, pp. 95–102, 1982.

A. Klinger, "Pattern and search statistics", in *Optimizing Methods in Statistics* (J.S. Rustagi, ed.), (New York), Academic Press, 1971.

C. B. Knudsen and H. I. Christensen, "On methods for efficient pyramid generation", in *Proc. 7th Scandinavian Conf. on Image Analysis*, (Aalborg, Denmark), pp. 28–39, Aug. 1991.

J. J. Koenderink, "The structure of images", *Biological Cybernetics*, vol. 50, pp. 363–370, 1984.

J. J. Koenderink, *Solid Shape*. Cambridge, Massachusetts: MIT Press, 1990.

J. J. Koenderink, "Generic neighborhood operators", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 14, pp. 597–605, Jun. 1992.

J. J. Koenderink and W. Richards, "Two-dimensional curvature operators", *J. of the Optical Society of America*, vol. 5:7, pp. 1136–1141, 1988.

J. J. Koenderink and A. J. van Doorn, "Visual detection of spatial contrast; influence of location in the visual field, target extent and illuminance level", *Biological Cybernetics*, vol. 30, pp. 157–167, 1978.

J. J. Koenderink and A. J. van Doorn, "Dynamic shape", *Biological Cybernetics*, vol. 53, pp. 383–396, 1986.

J. J. Koenderink and A. J. van Doorn, "Representation of local geometry in the visual system", *Biological Cybernetics*, vol. 55, pp. 367–375, 1987.

J. J. Koenderink and A. J. van Doorn, "Generic neighborhood operators", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 14, pp. 597–605, Jun. 1992.

A. F. Korn, "Toward a symbolic representation of intensity changes in images", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 10, no. 5, pp. 610–625, 1988.

L.M. Lifshitz and S.M. Pizer, "A multiresolution hierarchical approach to image segmentation based on intensity extrema", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 12, no. 6, pp. 529–541, 1990.

T. Lindeberg, "Scale-space for discrete signals", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 12, pp. 234–254, 1990.

T. Lindeberg, *Discrete scale space theory and the scale space primal sketch*. PhD thesis, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, May. 1991. A revised and extended version to appear in Kluwer Int. Series in Engineering and Computer Science.

T. Lindeberg, "Scale-space behaviour and invariance properties of differential singularities", in *Shape in Picture: Proc. NATO workshop on Shape in Picture* (Y. O. Ying, A. Toet, and H. Heijmanns, eds.), NATO ASI Series F, (Driebergen Netherlands), Springer Verlag, New York, Sep. 1992. (Also available in Tech. Rep. ISRN KTH/NA/P-92/26-SE from Royal Inst of Technology) (In press).

T. Lindeberg, "Scale-space behaviour of local extrema and blobs", *J. of Mathematical Imaging and Vision*, vol. 1, pp. 65–99, Mar. 1992.

T. Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention", *Int. J. of Computer Vision*, vol. 11, no. 3, pp. 283–318, 1993.

T. Lindeberg, "Discrete derivative approximations with scale-space properties: A basis for low-level feature extraction", *J. of Mathematical Imaging and Vision*, vol. 3, no. 4, pp. 349–376, 1993.

T. Lindeberg, "Effective scale: A natural unit for measuring scale-space lifetime", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 15, pp. 1068–1074, Oct. 1993.

T. Lindeberg, "On scale selection for differential operators", in *Proc. 8th Scandinavian Conf. on Image Analysis* (K. Heia K. A. Høgdra, B. Braathen, ed.), (Tromsø, Norway), pp. 857–866, Norwegian Society for Image Processing and Pattern Recognition, May. 1993.

T. Lindeberg, "Scale selection for differential operators", Tech. Rep. TRITA-NA-P9403, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Jan. 1994.

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

T. Lindeberg and J.-O. Eklundh, "Scale detection and region extraction from a scale-space primal sketch", in *Proc. 3rd Int. Conf. on Computer Vision*, (Osaka, Japan), pp. 416–426, Dec. 1990.

T. Lindeberg and J.-O. Eklundh, "The scale-space primal sketch: Construction and experiments", *Image and Vision Computing*, vol. 10, pp. 3–18, Jan. 1992.

T. Lindeberg and L. Florack, "On the decrease of resolution as a function of eccentricity for a foveal vision system", Tech. Rep. TRITA-NA-P9229, Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology, Oct. 1992. (Submitted).

T. Lindeberg and J. Gårding, "Shape from texture from a multi-scale perspective", in *Proc. 4th Int. Conf. on Computer Vision* (H.-H. Nagel et. al., ed.), (Berlin, Germany), pp. 683–691, IEEE Computer Society Press, May. 1993.

T. Lindeberg and J. Gårding, "Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D structure", in *Proc. 3rd European Conference on Computer Vision* (J.-O. Eklundh, ed.), vol. 800 of *Lecture Notes in Computer Science*, (Stockholm, Sweden), pp. 389–400, Springer-Verlag, May. 1994. (To appear).

Y. Lu and R. C. Jain, "Behaviour of edges in scale space", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 11, no. 4, pp. 337–356, 1989.

S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 11, no. 7, pp. 674–694, 1989.

S. G. Mallat and S. Zhong, "Characterization of signals from multi-scale edges", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 14, no. 7, pp. 710–723, 1992.

S.G. Mallat, "Multifrequency channel decompositions of images and wavelet models", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 37, pp. 2091–2110, 1989.

D. Marr, *Vision*. W.H. Freeman, New York, 1982.

D. C. Marr and E. C. Hildreth, "Theory of edge detection", *Proc. Royal Society London* B, vol. 207, pp. 187–217, 1980.

P. Meer, E. S. Baugher, and A. Rosenfeld, "Frequency domain analysis and synthesis of image pyramid generating kernels", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 9, pp. 512–522, 1987.

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

M. Nitzberg and T. Shiota, "Non-linear image filtering with edge and corner enhancement", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 14, no. 8, pp. 826–833, 1992.

E. J. Pauwels, P. Fiddelaers, T. Moons, and L. J. van Gool, "An extended class of scale-invariant and recursive scale-space filters", tech. rep., Dept. Elektrotechniek, Katholieke Universiteit Leuven, B-3001 Heverlee, Belgium, 1994.

T. Poston and I. Stewart, *Catastrophe Theory and its Applications*. London: Pitman, 1978.

A. Rosenfeld, *Multiresolution Image Processing and Analysis*, vol. 12 of *Springer Series in Information Sciences*. Springer-Verlag, 1984.

W. B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer, and L. Raphael, eds., *Wavelets and Their Applications*. Boston, Massachusetts: Jones and Barlett Publishers, 1992.

P. Saint-Marc, J.-S. Chen, and G. Medioni, "Adaptive smoothing: A general tool for early vision", *IEEE Trans. Pattern Analysis and Machine Intell.*, pp. 514–529, 1991.

A. H. Salden, L. M. J. Florack, and B. M. ter Haar Romeny, "Differential geometric description of 3D scalar images". Internal Report 3DCV # 91-10.

A. H. Salden, B. M. ter Haar Romeny, and L. M. J. Florack, "A complete and irreducible set of local orthogonally invariant features of 2-dimensional images". Internal Report 3DCV # 91-05.

G. Sapiro and A. Tannenbaum, "Affine invariant scale-space", *Int. J. of Computer Vision*, vol. 11, no. 1, pp. 25–44, 1993.

I. J. Schoenberg, "Contributions to the problem of approximation of equidistant data by analytic functions", *Quarterly of Applied Mathematics*, vol. 4, pp. 45–99, 1946.

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

I. J. Schoenberg, "On smoothing operations and their generating functions", *Bull. Amer. Math. Soc.*, vol. 59, pp. 199–230, 1953.

L. Schwartz, *Théorie des Distributions*, vol. I, II of *Actualités scientifiques et industrielles; 1091,1122*. Paris: Publications de l'Institut de Mathématique de l'Université de Strasbourg, 1950–1951.

M. Spivak, *Differential Geometry*, vol. 1–5. Berkeley, California, USA: Publish or Perish, Inc., 1975.

G. Strang, *Introduction to Applied Mathematics*. Massachusetts: Wellesley-Cambridge Press, 1986.

J. O. Strömberg, "A modified Franklin system and higher order splines as unconditional basis for Hardy spaces", in *Proc. Conf. in Harmonic Analysis in Honor of Antoni Zygmund* (Beckner W. et al., ed.), vol. II, Wadworth Mathematical Series, 1983.

S. Tanimoto, ed., *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 11:7. 1989.

S. Tanimoto and A. Klinger, eds., *Structured Computer Vision*. New York: Academic Press, 1980.

S. Tanimoto and T. Pavlidis, "A hierarchical structure for picture processing", *Computer Vision, Graphics, and Image Processing*, vol. 4, pp. 104–119, 1975.

D. Terzopoulos, "Multilevel computational processes for visual surface reconstruction", *Computer Vision, Graphics, and Image Processing*, vol. 24, pp. 52–95, 1983.

A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*. Washington DC: Winston and Wiley, 1977.

V. Torre and Th. A. Poggio, "On edge detection", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 8, no. 2, pp. 147–163, 1986.

L. Uhr, "Layered 'recognition cone' networks that preprocess, classify and describe", *IEEE Trans. Computers*, pp. 759–768, 1972.

H. Weyl, *The Classical Groups, Their Invariants and Representations*. Princeton, NJ: Princeton University Press, 1946.

R. T. Whitaker and S. M. Pizer, "A multi-scale approach to nonuniform diffusion", *Computer Vision, Graphics, and Image Processing*, vol. 57, no. 1, 1993.

D. V. Widder, *The Heat Equation*. New York: Academic Press, 1975.

R. Wilson and A. H. Bhalerao, "Kernel design for efficient multiresolution edge detection and orientation estimation", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 14, no. 3, pp. 384–390, 1992.

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

R. A. Young, "The Gaussian derivative theory of spatial vision: Analysis of cortical cell receptive field line-weighting profiles", Tech. Rep. GMR-4920, Computer Science Department, General Motors Research Lab., Warren, Michigan, 1985.

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

A. L. Yuille, "The creation of structure in dynamic shape", in *IEEE Second Conf. on Computer Vision*, (Tampa), pp. 685–689, 1988.

A. L. Yuille and T. A. Poggio, "Scaling theorems for zero-crossings", *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 8, pp. 15–25, 1986.

A. L. Yuille and T. A. Poggio, "Scaling and fingerprint theorems for zero-crossings", in *Advances in Computer Vision* (C. Brown, ed.), pp. 47–78, Lawrence Erlbaum, 1988.