Scale-Space for Discrete Signals

TONY LINDEBERG

Abstract—We address the formulation of a scale-space theory for *discrete* signals. In one dimension it is possible to characterize the smoothing transformations completely and an exhaustive treatment is given, answering the following two main questions: 1) Which linear transformations remove structure in the sense that the number of local extrema (or zero-crossings) in the output signal does not exceed the number of local extrema (or zero-crossings) in the original signal? 2) How should one create a multiresolution family of representations with the property that a signal at a coarser level of scale never contains more structure than a signal at a finer level of scale?

We propose that there is only one reasonable way to define a scale-space for 1-D discrete signals comprising a *continuous* scale parameter, namely by (discrete) convolution with the family of kernels $T(n; t) = e^{-t}I_n(t)$, where I_n are the modified Bessel functions of integer order. Similar arguments applied in the continuous case *uniquely* lead to the Gaussian kernel.

Some obvious discretizations of the continuous scale-space theory are discussed in view of the results presented. We show that the kernel $T(n; t)$ arises naturally in the solution of a discretized version of the diffusion equation. The commonly adapted technique with a sampled Gaussian can lead to undesirable effects since scale-space violations might occur in the corresponding representation. The result exemplifies the fact that properties derived in the continuous case might be violated after discretization.

A two-dimensional theory, showing how the scale-space should be constructed for images, is given based on the requirement that local extrema must not be enhanced, when the scale parameter is increased continuously. In the separable case the resulting scale-space representation can be calculated by separated convolution with the one-dimensional kernel $T(n; t)$.

The presented discrete theory has computational advantages compared to a scale-space implementation based on the sampled Gaussian, for instance concerning the Laplacian of the Gaussian. The main reason is that the discrete nature of the implementation has been taken into account already in the theoretical formulation of the scale-space representation.

Index Terms—Computer vision, diffusion, Gaussian filtering, multiresolution representation, scale-space discrete smoothing transformations, signal processing.

I. Introduction

It is well-known that objects in the world and details in an image exist only over a limited range of resolution. A classical example is the concept of a branch of a tree which makes sense only on the scale say from 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 levels of scale it is more relevant to talk about the molecules, which form the leaves of the tree, or the forest, in which the tree grows. If one aims at describing the structure of an image, a multiresolution representation is of crucial importance. Then a mechanism, which systematically removes finer details or high-frequency information from an image, is required. This smoothing must be available at any level of scale.

A method proposed by Witkin [23] and Koenderink and van Doorn [11] is to embed the original image in a one-parameter family of derived images, the scale-space, where the parameter t describes the current level of scale resolution. Let us briefly develop the procedure as it is formulated for one-dimensional continuous signals: given a signal $f: R \to R$ a function[1] $L: R \times R_+ \to R$ is defined by $L(x; 0) = f(x)$ and convolution with the Gaussian kernel $g: R \times R_+ \setminus \{0\} \to R$

$$L(x; t) = \int_{\xi=-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\xi^2/2t} f(x - \xi)\, d\xi \qquad (1)$$

if $t > 0$. Equivalently the family can be regarded as defined by the diffusion equation

$$\frac{\partial L}{\partial t} = \frac{1}{2} \frac{\partial^2 L}{\partial x^2} \qquad (2)$$

with initial condition $L(x; 0) = f(x)$. This family possesses some attractive properties.

- As the scale parameter t is increased additional local extrema or additional zero crossings cannot appear.
- Causality in the sense that $L(x; t_2)$ depends exclusively on $L(x; t_1)$ if $t_2 > t_1$ $(t_1, t_2 \geq 0)$.
- The blurring is shift invariant and does not depend upon the image values.

It has been shown by Babaud *et al.* [3] that the Gaussian function is the only kernel in a broad class of functions which satisfies adequate scale-space conditions.

The scale-space theory has been developed and well-established for continuous signals and images. However, it does not tell us at all about how the implementation should be performed computationally for real-life, i.e., discrete signals and images. In principle, we feel that there are two approaches possible.

- Apply the results obtained from the continuous scale-space theory by discretizing the occurring equations. For instance the convolution integral (1) can be approximated by a sum using customary numerical methods. Or, the diffusion equation (2) can be discretized in space with the

Manuscript received November 15, 1988; revised July 31, 1989. This work was supported by the National Swedish Board for Technical Development.

The author is with the Computer Vision and Associative Pattern Processing Laboratory (CVAP), Royal Institute of Technology, S-100 44 Stockholm, Sweden.

IEEE Log Number 8932046.

[1] R_+ denotes the set of real nonnegative numbers.

0162-8828/90/0300-0234$01.00 © 1990 IEEE

ordinary five-point Laplace operator forming a set of coupled ordinary differential equations, which can be further discretized in scale. If the numerical methods are chosen with care we will certainly get reasonable approximations to the continuous numerical values. But we are not guaranteed that the original scale-space conditions, however formulated, will be preserved.

- Define a genuinely discrete theory by postulating suitable axioms.

The goal of this paper is to develop the second item and to address the formulation of a scale-space theory for discrete images. We will start with a one-dimensional signal analysis. In this case it is possible to characterize exactly which kernels can be regarded as smoothing kernels and a complete and exhaustive treatment will be given. One among many questions which are answered is the following: if one performs repeated averaging, does one then get scale-space behavior? We will also present a family of kernels, which are the discrete analog of the Gaussian family of kernels. The set of arguments, which in the discrete case uniquely leads to this family of kernels do in the continuous case uniquely lead to the Gaussian family of kernels.

The structure of the two-dimensional problem is more complex, since it is difficult to formulate what should be meant by preservation of structure in this case. However, arguing that local extrema must not be enhanced when the scale parameter is increased continuously, we will give an answer to how the scale-space for two-dimensional discrete images should be calculated. In the separable case it reduces to separated convolution with the presented one-dimensional discrete analog of the Gaussian kernel. The representation obtained in this way has computational advantages compared to the commonly adapted approach, where the scale-space is based on different versions of the sampled Gaussian kernel. One of many spin-off products which come up naturally is a well-conditioned and efficient method to calculate (a discrete analog of) the Laplacian of the Gaussian. It is well-known that the implementation of the Laplacian of the Gaussian has lead to computational problems [8].

The theory developed in this paper also has the attractive property that it is linked to the continuous theory through a discretized version of the diffusion equation. This means that continuous results may be transferred to the discrete implementation provided that the discretization is done correctly. However, the important point of the scale-space concept outlined here is that the properties we want from a scale-space hold not only in the ideal theory but also in the discretization,[2] since the discrete nature of the problem has been taken into account already in the theoretical formulation of the scale-space representation. Therefore, we believe that the suggested way to implement the scale-space theory really describes the proper way to do it.

The presentation is organized as follows. In Section II we define what we mean by a scale-space representation and a one-dimensional discrete scale-space kernel. Then in a straightforward and constructive manner Section III illustrates some qualitative properties that must be possessed by scale-space kernels. A complete characterization as well as an explicit expression for the generating function of all discrete scale-space kernels are given in Section IV. Section V develops the concept of a discrete scale-space with a *continuous* scale parameter. The formulation is equivalent to the previous scale-space formulation, which in the continuous case leads to the Gaussian kernel. The numerical implementation of this scale-space is treated in Section VI. Section VII discusses discrete scale-space properties of some obvious discretizations of the convolution integral and the diffusion equation. Section VIII describes some problems which occur due to the more complicated topology in two dimensions. In Section IX we develop the scale-space for two-dimensional discrete images. Here we also compare the discrete scale-space representation with the commonly used approach, where the scale-space implementation is based on various versions of the sampled Gaussian kernel. Finally, Section X gives a brief summary of the main results.

The results presented should have implications for image analysis as well as other disciplines of digital signal processing.

II. Scale-Space Axioms

By a scale-space we mean a family of derived signals meant to represent the original signal at various levels of scale. Each member of the family should be associated with a value of a scale parameter intended to somehow describe the current level of scale. This scale parameter, here denoted by t, may be either discrete ($t \in Z_+$) or continuous ($t \in R_+$) and we obtain two different types of discrete scale-spaces—discrete signals with a discrete scale parameter and discrete signals with a continuous scale parameter. However, in both cases we start from the following basic assumptions:

- All representations should be generated by convolution of the original image with a kernel (linear shift-invariant smoothing).
- An increasing value of the scale parameter t should correspond to coarser levels of scale and signals with less structure. Particularly, $t = 0$ should represent to the original signal.
- All signals should be real-valued functions: $Z \to R$ defined on the same infinite grid; in other words no pyramid representations will be used.

The essential requirement is that a signal at a finer level of scale should contain less structure than a signal at a

[2] In a practical implementation we are of course faced with rounding and truncation errors due to finite precision. But the idea with this approach is that we hope to improve our algorithms by including at least the discretization effects already in the theory. In ordinary numerical analysis for simulation of physical phenomena it is almost always possible to reduce these effects by increasing the density of mesh points, in case the current grid is not fine enough to give a prescribed accuracy in the result. However, in computer vision we are often locked to some fixed maximal resolution, beyond which additional image data are not available.

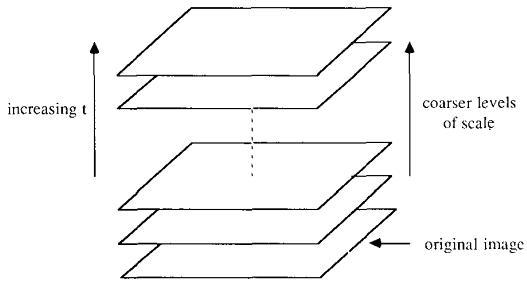


Fig. 1. A scale-space is an ordered set of derived signals/images intended to represent the original signal/image at various levels of scale.

coarser level of scale. If one regards the number of local extrema as one measure of smoothness it is thus necessary that the number of local extrema in space does not increase as we go from a finer to a coarser level of scale. It can be shown that the family of functions generated by convolution with the Gaussian kernel possesses this property in the continuous case. We state it as the basic axiom for our one-dimensional analysis and define the following.

Definition 1: A one-dimensional discrete kernel $K: Z \to R$ is denoted a scale-space kernel if for all signals $f_{\text{in}}: Z \to R$ the number of local extrema in the convolved signal $f_{\text{out}} = K * f_{\text{in}}$ does not exceed the number of local extrema in the original signal.

An important observation to note is that this definition equivalently can be expressed in terms of zero-crossings just by replacing the string "local extrema" with "zero-crossings." The result follows from the facts that a local extremum in a discrete function f is equivalent to a zero-crossing in its first difference Δf, defined by $(\Delta f)(x) = f(x+1) - f(x)$, and that the difference operator commutes with the convolution operator.

However, the stated definition has further consequences. It means that the number of local extrema (zero-crossings) in any nth order difference of the convolved image cannot be larger than the number of local extrema (zero-crossings) in the nth order difference of the original image. Actually, the result can be generalized to arbitrary linear operators.

Proposition 1: Let $K: Z \to R$ be a discrete scale-space kernel and $\mathfrak{L}$ be a linear operator (from the space of real-valued discrete functions to itself), which commutes with K. Then for any $f: Z \to R$ (such that the involved quantities exist) the number of local extrema in $\mathfrak{L}(K * f)$ cannot exceed the number of local extrema in $\mathfrak{L}(f)$.

Proof: Let $g = \mathfrak{L}(f)$. As K is a scale-space kernel the number of local extrema in $K * g$ cannot be larger than the number of local extrema in g. Since K and $\mathfrak{L}$ commute $K * g = K * \mathfrak{L}(f) = \mathfrak{L}(K * f)$ and the result follows. □

This shows that not only the function, but also all its "derivatives" will become smoother. Accordingly, convolution with a discrete scale-space kernel can really be regarded as a smoothing operation.

To realize that the number of local extrema or zero-crossings can increase even in a rather uncomplicated situation consider the input signal

$$f_{\text{in}}(x) = \begin{cases} -3 & \text{if } n = 0 \\ 2 & \text{if } n = \pm 1 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

and convolve it with the kernels $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $(\frac{1}{2}, \frac{1}{2})$, and $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. The results are shown in Fig. 2(b), (c), and (d), respectively. As we see, both the number of local extrema and the number of zero-crossings have increased for the first kernel, but not for the two latter ones. Thus, an operator which naively can be apprehended as a smoothing operator, might actually give a less smooth result. Further, it can really matter if one averages over three instead of two points and how the averaging is performed.

In order to get familiar with the consequences of the definition we will illustrate what this scale-space property means. We start by pointing out a few general qualitative requirements of a scale-space kernel that are necessarily induced by the given axiom. We will also show that the two latter kernels indeed are discrete scale-space kernels.

III. Scale-Space Kernels

A. Positivity and Unimodality in the Spatial Domain

By considering the impulse response it is possible to draw some qualitative conclusions about the properties of a discrete scale-space kernel. Let

$$f_{\text{in}}(x) = \delta(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Then

$$f_{\text{out}}(x) = (K * \delta)(x) = K(x). \tag{5}$$

$\delta(x)$ has exactly one local maximum and no zero-crossings. Therefore in order to be a scale-space kernel K must not have more than one extremum and no zero-crossings. Thus we can state the following.

Proposition 2: All coefficients of a scale-space kernel must have the same sign.

Proposition 3: The coefficient sequence of a scale-space kernel $\{K(n)\}_{n=-\infty}^{\infty}$ must be unimodal.[3]

Without loss of generality we can therefore restrict the rest of the treatment to positive sequences where all $K(n) \geq 0$.

It seems reasonable to require[4] that $K \in l_1$, i.e., that $\sum_{n=-\infty}^{\infty} |K(n)|$ is finite. If f_{in} is bounded and $K \in l_1$ then the convolution is well-defined and the Fourier transform of the filter coefficient sequence exists. It also allows us

[3] A real sequence is called unimodal if it is first ascending (descending) and then descending (ascending).

[4] Some regularity requirement must be imposed on the input signal as well. Throughout our following considerations we will stick to one general convention. If nothing else is explicitly mentioned we assume that f_{in} is sufficiently regular such that the involved quantities exist and are well-defined.

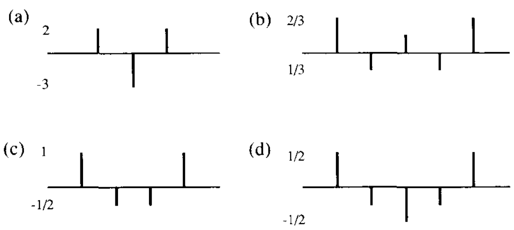


Fig. 2. (a) Input signal. (b) Convolved with $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. (c) Convolved with $(\frac{1}{2}, \frac{1}{2})$. (d) Convolved with $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$.

to normalize the coefficients such that $\Sigma_{n=-\infty}^{\infty} K(n) = 1$. Particularly, the filter coefficients $K(n)$ must then tend to zero as n goes to infinity.

B. Generalized Binomial Kernels

Consider a two-kernel with only two nonzero filter coefficients:

$$K^{(2)}(n) = \begin{cases} p & \text{if } n = 0 \\ q & \text{if } n = -1 \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Assume that $p \geq 0$, $q \geq 0$, and $p + q = 1$.

It is easy to verify that the number of zero-crossings (local extrema) in $f_{\text{out}} = K^{(2)} * f_{\text{in}}$ cannot exceed the number of zero crossings (local extrema) in f_{in}. This result follows from the fact that convolution of f_{in} with $K^{(2)}$ is equivalent to the formation a weighted average of the sequence $\{f_{\text{in}}(x)\}_{x=-\infty}^{\infty}$; see Fig. 3. The values of the output signal can be constructed geometrically and will fall on straight lines connecting the values of the input signal. The offset along the x-axis is determined by the ratio $q/(p + q)$. One realizes that no additional zero-crossings can be introduced by this transformation. Thus, a kernel on the form (6) is a discrete scale-space kernel.

Directly from the definition of a scale-space kernel it follows that if two kernels K_a and K_b are scale-space kernels then $K_a * K_b$ is also a scale-space kernel. Repeated application of this result yields the following.

Proposition 4: All kernels K on the form $*_{i=1}^{n} K_i^{(2)}$, with $K_i^{(2)}$ according to (6), are discrete scale-space kernels.

The filter coefficients generated in this way can be regarded as a kind of generalized binomial coefficient. The ordinary binomial coefficients are obtained, except for a scaling-factor, as a special case if all p_i and q_i are equal. Another formulation of Proposition 4 in terms of generating functions is also possible.

Proposition 5: All kernels with the generating function $\varphi_K(z) = \Sigma_{n=-\infty}^{\infty} K(n) z^n$ on the form

$$\varphi_K(z) = Cz^k \prod_{i=1}^{N} (p_i + q_i z) \tag{7}$$

where $p_i > 0$ and $q_i > 0$ are discrete scale-space kernels.

Proof: The generating function of a kernel on the form (6) is $\varphi_{K_i^{(2)}}(x) = p_i + q_i z$. As convolution in the spatial domain corresponds to multiplication of generating functions Proposition 4 gives us that

$$\varphi_K(z) = \varphi_{K_1^{(2)}}(z)\, \varphi_{K_2^{(2)}}(z) \cdots \varphi_{K_N^{(2)}}(z) \tag{8}$$

is the generating function of a scale-space kernel. A constant scaling-factor C or a translation $\varphi_{\text{transl}}(z) = z^k$ cannot affect the number of local extrema. Therefore these factors can be multiplied onto $\varphi_K(z)$ without changing the scale-space properties. □

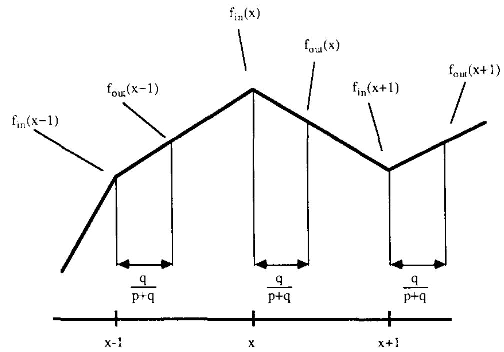


Fig. 3. To convolve a signal f_{in} with a two-kernel $K^{(2)}(n)$ is equivalent to form a weighted average of the sequence $\{f_{in}(x)\}_{x=-\infty}^{\infty}$.

C. No Real Negative Eigenvalues of the Convolution Matrix

If the convolution transformation $f_{\text{out}} = K * f_{\text{in}}$ is represented on matrix form $f_{\text{out}} = Cf_{\text{in}}$ a matrix with constant values along the diagonals $C_{i,j} = K(i - j)$ appears. Such a matrix is called a Toeplitz matrix.

Proposition 6: Let $K: Z \to R$ be a discrete kernel with finite support and filter coefficients $c_n = K(n)$. If for some dimension N the $N \times N$ convolution matrix

$$C^{(N)} = \begin{pmatrix} c_0 & c_{-1} & \cdots & c_{2-N} & c_{1-N} \\ c_1 & c_0 & c_{-1} & \cdots & c_{2-N} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c_{N-2} & \cdots & c_1 & c_0 & c_{-1} \\ c_{N-1} & c_{N-2} & \cdots & c_1 & c_0 \end{pmatrix} \tag{9}$$

has a negative eigenvalue with a corresponding real eigenvector then K cannot be a scale-space kernel.

Proof: See Appendix A.1. □

D. Positivity in the Frequency Domain

The eigenvalues of a Toeplitz matrix are closely related to the Fourier transform of the corresponding sequence of coefficients [7], [6]. This property allows us to derive an interesting Corollary from Proposition 6.

Proposition 7: The Fourier transform $\psi_K(\theta) = \Sigma_{n=-\infty}^{\infty} K(n) e^{-in\theta}$ of a symmetric discrete scale-space kernel K with finite support is nonnegative.

Proof: Let $\lambda_1^{(N)}$ denote the smallest eigenvalue of the convolution matrix of dimension N and let m denote the minimum value[5] the Fourier transform ψ_K assumes on $[-\pi, \pi]$. As a consequence of a theorem by Grenander [7, Section 5.2, p. 65] about the asymptotic distribution of eigenvalues of a finite Toeplitz matrix it follows that

$$\lim_{N\to\infty} \lambda_1^{(N)} = m \qquad \lambda_1^{(N)} \geq m. \tag{10}$$

If m is strictly negative then as $\lim_{N\to\infty} \lambda_1^{(N)} = m$ it follows that $\lambda_1^{(N)}$ will be negative for some sufficiently large N. According to Proposition 6 the kernel cannot be a scale-space kernel. □

E. Unimodality in the Frequency Domain

If a linear transformation is to be regarded as a smoothing transformation it turns out to be necessary that the low frequency components are not suppressed more than the high frequency components. This means that the Fourier transform must not increase when the absolute value of the frequency increases. The occurring unimodality property is easiest to establish for circular convolution. In this case the convolution matrix becomes circulant,[6] which means that its eigenvalues and eigenvectors can be determined analytically.

Proposition 8: Let $\{c_n\}_{n=-\infty}^{\infty}$ be the filter coefficients of a symmetric discrete kernel with $c_n = 0$ if $|n| > N$. For all integers $M \geq N$ it is required that the transformation given by multiplication with the $(2M + 1) \times (2M + 1)$ symmetric circulant matrix $C_C^{(M)}$ (11), defined by $(C_C^{(M)})_{i,j} = c_{i-j}$ $(i, j = 0..M)$ and circulant extension, should be a scale-space transformation. Then, necessarily the Fourier transform $\psi(\theta) = \Sigma_{n=-\infty}^{\infty} c_n e^{-in\theta}$ must be unimodal on $[-\pi, \pi]$.

$$C_C^{(M)} = \begin{bmatrix} c_0 & c_1 & \cdots & c_N & & c_N & \cdots & c_1 \\ c_1 & c_0 & c_1 & & c_N & & \ddots & \vdots \\ \vdots & & \ddots & & & \ddots & & c_N \\ c_N & & & \cdot & & & \ddots & \\ & c_N & & & \cdot & & c_N & \\ & & \ddots & & & \cdot & & c_N \\ c_N & & & \ddots & & & \ddots & \vdots \\ \vdots & \ddots & & & c_N & & c_0 & c_1 \\ c_1 & \cdots & c_N & & & c_N \cdots & c_1 & c_0 \end{bmatrix} \tag{11}$$

Proof: See Appendix A.2. □

The result can be extended to comprise noncircular convolution as well. The idea behind the proof is to construct an input signal consisting of several periods of the signal leading to a scale-space violation in the proof of Proposition 8. Then, the convolution effect on the "interior" periods will be identical to effect on one period by circular convolution. If the signal consists of a sufficient number of periods the boundary effects will be negligible compared to the large number of scale-space violations occurring in the inner parts. The formal details are somewhat technical and can be found in [15, Section 2.6].

Proposition 9: The Fourier transform $\psi_K(\theta) = \Sigma_{n=-\infty}^{\infty} K(n) e^{-in\theta}$ of a symmetric discrete scale-space kernel K with finite support is unimodal on the interval $[-\pi, \pi]$ (with the maximum value at $\theta = 0$).

F. Kernels with Three Nonzero Elements

For a three-kernel $K^{(3)}$ with exactly three nonzero consecutive elements $c_{-1} > 0$, $c_0 > 0$, and $c_1 > 0$ it is possible to determine the eigenvalues of the convolution matrix and the roots of the characteristic equation analytically. It is easy to verify that the eigenvalues λ_μ of the convolution matrix

$$C^{(N)}((c_{-1}, c_0, c_1)) = \begin{bmatrix} c_0 & c_{-1} & & & & \\ c_1 & c_0 & c_{-1} & & & \\ & c_1 & c_0 & c_{-1} & & \\ & & \ddots & \ddots & \ddots & \\ & & & c_1 & c_0 & c_{-1} \\ & & & & c_1 & c_0 \end{bmatrix} \tag{12}$$

are all real and equal to

$$\lambda_\mu = c_0 - 2\sqrt{c_{-1}c_1} \cos\left(\frac{\mu\pi}{N+1}\right) \qquad \mu = 1..N \tag{13}$$

and that the roots of generating function $\varphi_{K^{(3)}}(z) = c_{-1}z^{-1} + c_0 + c_1 z$ are

$$z_{1,2} = \frac{-c_0 \pm \sqrt{c_0^2 - 4c_{-1}c_1}}{2c_1}. \tag{14}$$

From (13) we deduce that if $c_0 < 2\sqrt{c_{-1}c_1}$ then for some sufficiently large N at least one eigenvalue of $C^{(N)}$ will be negative. Thus, according to Proposition 6 the kernel cannot be a scale-space kernel. However, if $c_0^2 \geq 4c_{-1}c_1$ then both the roots of $\varphi_{K^{(3)}}$ will be real and negative. This means that the generating function can be written on the form (7) and the kernel is a scale-space kernel. Consequently, we obtain a complete classification for all possible values of c_{-1}, c_0, and c_1. We conclude the following.

Proposition 10: A three-kernel with positive elements c_{-1}, c_0, and c_1 is a scale-space kernel if and only if $c_0^2 \geq 4c_{-1}c_1$, i.e., if and only if it can be written as the convolution of two two-kernels with positive elements.

At this moment one could ask oneself if the result can be generalized to hold for kernels with arbitrary numbers

[5]Due to symmetry of the kernel, $\psi_K(\theta)$ assumes only real values. The minimum value does certainly exist since $\psi(\theta)$ is a continuous function and the interval $[-\pi, \pi]$ is compact.

[6]In a circulant matrix each row is a circular shift of the previous row, except for the first row which is a circular shift of the last row.

of nonzero filter coefficients, i.e., if all discrete scale-space kernels with finite support have a generating function on the form (7). This question will be answered in the next section.

IV. Kernel Classification

Until now we have postulated an axiom in terms of local extrema or equivalently zero-crossings and investigated some of its consequences for signal transformations expressed as linear convolution with a shift-invariant kernel. We have seen that the sequence of filter coefficients must be positive and unimodal and that its sum should be convergent. For symmetric kernels the Fourier spectrum must be positive and unimodal on $[-\pi, \pi]$.

In this section we will perform a complete characterization of the scale-space kernels. We have studied the literature and seen that several interesting results can be derived from the theory of total positivity. The proofs of the important theorems are sometimes of a rather complicated nature for a reader with an engineering background. We will not burden the presentation with them but give a summarizing result without proof. The reader interested in further details is referred to [15] or the other references mentioned.

The pioneer in the subject of variation-diminishing transforms was I. J. Schoenberg. He studied the subject in a series of papers from 1930 to 1953 [20]–[22]. Later the theory of total positivity was covered in a monumental monograph by Karlin [13]. A recent paper by Ando [2] reviews the field using skew-symmetric vector products and Schur complements of matrices as major tools. The questions issued in this paper constitute a new application of these not too well-known but very powerful results.

Theorem 1: A discrete kernel $K: Z \to R$ is a scale-space kernel if and only if the corresponding sequence of filter coefficients $\{K(n)\}_{n=-\infty}^{\infty}$ is a normalized Pòlya frequency sequence, i.e., if all minors of the infinite matrix

$$\begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & K(0) & K(-1) & K(-2) & \cdot & \cdot \\ \cdot & \cdot & K(1) & K(0) & K(-1) & \cdot & \cdot \\ \cdot & \cdot & K(2) & K(1) & K(0) & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \tag{15}$$

are nonnegative.

There exists a remarkably explicit characterization theorem for the generating function of a normalized Pòlya frequency sequence. It has been proved in several steps by Edrei and Schoenberg; see [22] or [13].

Theorem 2: An infinite sequence $\{K(n)\}_{n=-\infty}^{\infty}$ is a normalized Pòlya frequency sequence if and only if its generating function $\varphi_K(z) = \sum_{n=-\infty}^{\infty} K(n) z^n$ is of the form

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

$$c > 0; \quad k \in Z; \quad q_{-1}, q_1, \alpha_i, \beta_i, \gamma_i, \delta_i \geq 0$$

$$\beta_i, \gamma_i < 1;$$

$$\sum_{i=1}^{\infty} (\alpha_i + \beta_i + \gamma_i + \delta_i) < \infty. \tag{16}$$

The product structure of this expression corresponds to the previously mentioned property that if K_a and K_b are scale-space kernels then also $K_a * K_b$ is a scale-space kernel. The meanings of the leading factors C and z^k are just rescaling and translation. In $(1 + \alpha_i z)$ and $(1 + \delta_i z^{-1})$ we recognize rewritten versions of the generating functions of two-kernels. The factors in the denominator are Taylor expansions of geometric series, which correspond to recursive average processes of the forms $f_{\text{out}}(x) = f_{\text{in}}(x) + \beta_i f_{\text{out}}(x - 1)$ and $f_{\text{out}}(x) = f_{\text{in}}(x) + \gamma_i f_{\text{out}}(x + 1)$. The exponential factor describes infinitesimal smoothing. Its interpretation will become clearer in the next section, when we derive the discrete scale-space with a continuous scale parameter.

For kernels with finite support the generating function will be reduced to $\varphi_K(z) = c z^k \prod_{i=1}^{\infty} (1 + \alpha_i z)(1 + \delta_i z^{-1})$, which except for rescaling and translation is the generating function of the class of generalized binomial kernels in Propositions 4 and 5. Hence, we have the following.

Theorem 3: The kernels on the form $*_{i=1}^{n} K_i^{(2)}$, with $K_i^{(2)}$ according to (6), are (except for rescaling and translation) the only discrete scale-space kernels with finite support.

An immediate consequence of this result is that convolution with a finite scale-space kernel can be decomposed into convolution with kernels having two strictly positive consecutive filter coefficients.

The representation (16), which gives a catalog of all one-dimensional discrete smoothing kernels, can sometimes be very convenient for further analysis. For example, starting from (16) it is almost trivial to show that the Fourier transform of a symmetric discrete scale-space kernel is unimodal and nonnegative on the interval $[-\pi, \pi]$. Due to the symmetry we have $q_{-1} = q_1$, $\alpha_\nu = \delta_\nu$, and $\beta_\nu = \gamma_\nu$. As a first step one replaces z with $e^{-i\theta}$ (which gives the Fourier transform) and shows that each one of the factors $e^{(q_{-1} z^{-1} + q_1 z)}$, $(1 + \alpha_\nu z)(1 + \delta_\nu z^{-1})$, and $((1 - \beta_\nu z)(1 - \gamma_\nu z^{-1}))^{-1}$ is a nonnegative and unimodal function of θ on $[-\pi, \pi]$. The remaining details are left to the reader.

V. Axiomatic Scale-Space Construction

A. Discrete Scale-Space with Discrete Scale Parameter

With the classification result from the previous section in mind an apparent way to get a multiresolution representation of a discrete signal f is to define a set of discrete functions L_i ($i = 0..n$) where $L_0 = f$ and each coarser level is calculated by convolutions from the previous one $L_i = K_{i \leftarrow i-1} * L_{i-1}$ ($i = 1..n$). The kernels $K_{i \leftarrow i-1}$ should be appropriately selected scale-space kernels corresponding to suitable amounts of blurring. The scale-space condition for each kernel guarantees that signals at coarser levels of scale (larger value of i) do not contain more structure than signals at finer levels of scale. This leads to a so-called sampled scale-space with a *discrete* scale parameter. Combined with a subsampling operator it provides a theoretical basis for the pyramid representations. However, one problem arises. How should one select the kernels/scale-levels *a priori* in order to achieve a sufficiently dense sampling in scale?

B. Discrete Scale-Space with Continuous Scale Parameter

The goal in this section is to tie together scale-space kernels corresponding to different degrees of smoothing in a systematic manner such that a *continuous* resolution parameter can be introduced. The concept of a continuous scale parameter is of considerable importance, since we will no longer be locked to fixed predetermined discrete levels of scale. It allows us to defocus signals with an arbitrary amount of blurring, which will certainly make it easier to locate and trace events in scale-space. Of course, it is impracticable to generate the representations at all levels of scale in a real implementation. However, the important idea is that, in contrast to the pyramid representations where the scale levels are fixed in advance, with a continuous scale parameter the scale-space representation at *any* level of scale can be calculated if desired.

We will not consider the question about how to choose a suitable set of scale levels in a practical case. Imagine for instance that we want to trace events, like local extrema, zero-crossings, edges [4] or convex and concave regions, as the blurring proceeds in scale-space. In order to analyze the behavior in scale-space, the continuum of multiresolution representations must be sampled at some levels of scale. It is certainly a nontrivial problem to make an appropriate selection of these levels. The point of a scale-space having a continuous scale parameter is that it provides a theoretical framework in which the scale steps can be varied arbitrarily. We do not need to select any set of scale levels *in advance*, but can leave the decision open to the actual situation. In other words, the continuous scale parameter allows for data-driven determination of scale levels.

We start from the axioms given in Section II and postulate that the scale-space should be generated by convolution with a one-parameter family of kernels, i.e., $L(x; 0) = f(x)$ and

$$L(x; t) = \sum_{n=-\infty}^{\infty} T(n; t) f(x - n) \qquad t > 0. \quad (17)$$

This form on the smoothing formula reflects the requirements about linear shift-invariant smoothing and a continuous scale parameter. The amount of structure in a signal must not increase with scale. This means that for any $t_2 > t_1$ the number of local extrema in $L(x; t_2)$ must not exceed the number of local extrema in $L(x; t_1)$. Particularly, by setting t_1 to zero we realize that each $T(\cdot\,; t)$ must be a scale-space kernel.

In order to simplify the analysis a semigroup requirement $T(\cdot\,; s) * T(\cdot\,; t) = T(\cdot\,; s + t)$ is imposed on the family of kernels. This property makes it possible to calculate the representation $L(\cdot\,; t_2)$ at a coarser level t_2 from the representation $L(\cdot\,; t_1)$ at a finer level t_1 ($t_2 > t_1$) by convolution with a kernel *from the* one-parameter *family*. In summary,

$$\begin{aligned} L(\cdot\,; t_2) &= \{\text{definition}\} = T(\cdot\,; t_2) * f = \{\text{semigroup}\} \\ &= \big(T(\cdot\,; t_2 - t_1) * T(\cdot\,; t_1)\big) * f \\ &= \{\text{associativity}\} \\ &= T(\cdot\,; t_2 - t_1) * \big(T(\cdot\,; t_1) * f\big) \\ &= \{\text{definition}\} = T(\cdot\,; t_2 - t_1) * L(\cdot\,; t_1). \end{aligned} \quad (18)$$

As each $T(\cdot\,; t)$ is a scale-space kernel the semigroup property ensures that the scale-space property holds between any two levels of scale. It also means that all scale levels will be treated in a similar manner, i.e., the transformation from a fine scale level to a coarser scale level always follows the same law.

We will show below that the conditions mentioned, combined with a normalization criterion $\Sigma_{n=-\infty}^{\infty} T(n; t) = 1$ and a symmetry constraint $T(-n; t) = T(n; t)$, determine the family of kernels up to a positive scaling parameter[7] α. One obtains,

$$T(n; t) = e^{-\alpha t} I_n(\alpha t) \quad (19)$$

where I_n are the modified Bessel functions of integer order. These functions with real arguments are except for an alternating sign sequence equal to the ordinary Bessel functions J_n of integer order with purely imaginary arguments.

$$I_n(t) = I_{-n}(t) = (-1)^n J_n(it) \qquad n \geq 0, t > 0. \quad (20)$$

Theorem 4: Given any one-dimensional signal $f: Z \to R$ let $L: Z \times R_+ \to R$ be a one-parameter family of functions defined by $L(x; 0) = f(x)$ ($x \in Z$) and $L(x; t) = \Sigma_{n=-\infty}^{\infty} T(n; t) f(x - n)$ ($x \in Z$, $t > 0$), where $T: Z \times R_+ \to R$ is a one-parameter family of symmetric functions satisfying the semigroup property $T(\cdot\,; s) * T(\cdot\,; t) = T(\cdot\,; s + t)$ and the normalization criterion $\Sigma_{n=-\infty}^{\infty} T(n; t) = 1$. For all signals f it is required that if $t_2 > t_1$ then the number of local extrema (zero-crossings) in $L(x; t_2)$ must not exceed the number of local extrema (zero-crossings) in $L(x; t_1)$. Then necessarily (and sufficiently), $T(n;$

[7]For simplicity, the parameter α, which only affects the scaling of the scale parameter, will be set to 1 after the end of this section.

$t) = e^{-\alpha t} I_n(\alpha t)$ for some nonnegative real α, where I_n are the modified Bessel functions of integer order.

Proof: As mentioned above each kernel $T(n; t)$ must be a scale-space kernel. A theorem by Karlin [13, p. 354] states that the only semigroup of normalized Pólya frequency sequences has a generating function on the form $\varphi(z) = e^{t(az^{-1}+bz)}$ where $t > 0$ and $a, b \geq 0$. This result, which forms the basis of the proof, can be easily understood from Theorem 2. If a family $h(\cdot; t)$ possesses the semigroup property $h(\cdot; s) * h(\cdot; t) = h(\cdot; s + t)$ then its generating function must obey the relation $\varphi_{h(\cdot;s)} \cdot \varphi_{h(\cdot;t)} = \varphi_{h(\cdot;s+t)}$ for all nonnegative s and t. This excludes the factors z^k, $(1 + \alpha_i z)$, $(1 + \delta_i z^{-1})$, $(1 - \beta_i z)$, and $(1 - \gamma_i z^{-1})$ from (16). What remains are the constant and the exponential factors. The argument of the exponential factor must also be linear in t in order to fulfill the adding property of the scale parameters of the kernels under convolution.

Due to the symmetry the generating function must satisfy $\varphi_h(z^{-1}) = \varphi_h(z)$, which in our case leads to $a = b$. For simplicity, let $a = b = \alpha/2$, and we get the generating function for the modified Bessel functions of integer order, see [1, (9.6.33)].

$$\varphi_t(z) = e^{(\alpha t/2)(z^{-1}+z)} = \sum_{n=-\infty}^{\infty} I_n(\alpha t) z^n \quad (21)$$

We obtain a normalized kernel if we let $T: Z \times R_+ \to R$ be defined by $T(n; t) = e^{-\alpha t} I_n(\alpha t)$. Set z to 1 in the generating function $e^{(\alpha t/2)(z^{-1}+z)} = \Sigma_{n=-\infty}^{\infty} I_n(\alpha t) z^n$. Then it follows that $\Sigma_{n=-\infty}^{\infty} I_n(\alpha t) = e^{\alpha t}$, which means that $\Sigma_{n=-\infty}^{\infty} T(n; t) = 1$. The semigroup property is trivially preserved after normalization. □

Consequently, this result provides us with an explicit controlled method to preserve structure in the spatial domain as we let the original signal erode by blurring it to a coarser level of scales.

If the relation (21) is multiplied by the factor $e^{-\alpha t}$ and z is replaced with $e^{-i\theta}$ one gets the analytical expression for the Fourier transform of $T(n; t)$.

Proposition 11: The Fourier transform of the kernel $T(n; t) = e^{-\alpha t} I_n(\alpha t)$ is

$$\psi_T(\theta) = \sum_{n=-\infty}^{\infty} T(n; t) e^{-in\theta} = e^{\alpha t(\cos\theta - 1)}. \quad (22)$$

For completeness, it should be mentioned that the variance of the kernel $T(n; t)$ is $\Sigma_{n=-\infty}^{\infty} n^2 T(n; t) = t$. This can be shown easily from a recurrence relation (26) for modified Bessel functions and the normalization condition.

In Appendix A.4 we give graphs of the discrete analog of the Gaussian kernel $T(n; t)$ and the continuous Gaussian kernel $g(\xi; t)$.

C. Equivalent Formulation for Continuous Signals

If similar arguments are applied in the continuous case we obtain the Gaussian kernel. The following is a summary.

Theorem 5: Given any one-dimensional continuous signal $f: R \to R$, let $L: R \times R_+ \to R$ be a one-parameter family of functions defined by $L(x; 0) = f(x)$ $(x \in R)$ and $L(x; t) = \int_{\xi=-\infty}^{\infty} g(\xi; t) f(x - \xi)\, d\xi$ $(x \in R, t > 0)$, where $g: R \times R_+ \to R$ is a one-parameter family of symmetric functions satisfying the semigroup property $g(\cdot; s) * g(\cdot; t) = g(\cdot; s + t)$ and the normalization criterion $\int_{\xi=-\infty}^{\infty} g(\xi; t)\, d\xi = 1$. For all signals f it is required that if $t_2 > t_1$ then the number of local extrema (zero-crossings) in $L(x; t_2)$ must not exceed the number of local extrema (zero-crossings) in $L(x; t_1)$. Suppose also that $g(\xi; t)$ is Borel-measurable as a function of t. Then necessarily (and sufficiently), $g(\xi; t) = (2\pi\alpha t)^{-1/2} \exp(-\xi^2/2\alpha t)$ for some nonnegative real α.

This result, which is proved in [15, Section 4.1], gives further support for the firm belief that Theorem 4 states the canonical way to define a scale-space for discrete signals. It also provides an alternative formulation of the uniqueness of the Gaussian kernel for scale-space filtering of continuous signals [3]. The assumption of Borel-measurability means no important restriction. It is well-known that all continuous functions are Borel-measurable.

VI. Numerical Implementation

According to the definition of the scale-space for discrete signals the representation of a signal f at a scale-level t is given by,

$$L(x; t) = \sum_{n=-\infty}^{\infty} T(n; t) f(x - n) \qquad x \in Z, t > 0 \quad (23)$$

where $T(n; t) = e^{-t} I_n(t)$. When this transformation is to be implemented computationally a few numerical problems must be considered.

- The infinite convolution sum must be replaced with a finite one.
- Normally, the modified Bessel functions are not available as standard library routines. Therefore, we must design an algorithm to generate the required filter coefficients $T(n; t)$ for a given value of t.
- A realistic signal is finite, but a finite approximation of (23) might need additional values.

In this section we will discuss the first two items. We will not go into the complications, which arise from finite signals. Instead we assume that f is defined for all those integers, where signal values are required for our algorithms.

A. Truncation and Filter Coefficient Generation

A reasonable approach to approximate (23) is to truncate the infinite sum for some sufficiently large value of N,

$$L(x; t) \approx \sum_{n=-N}^{N} T(n; t) f(x - n) \qquad x \in Z, t > 0 \quad (24)$$

chosen such that the absolute error in L due to truncation does not exceed a given error limit ϵ_{trunc}. If we assume that f is bounded ($|f(x)| \leq M$) we get the sufficient condition

$$2M \sum_{n=N+1}^{\infty} T(n; t) \leq \epsilon_{\text{trunc}}. \tag{25}$$

An easy way to generate the filter coefficients is to use the recurrence relation, see [1, (9.6.26)],

$$I_{n-1}(t) - I_{n+1}(t) = \frac{2n}{t} I_n(t) \tag{26}$$

which is always stable for backward iteration. One can use Miller's algorithm [19, p. 142] and start the recurrence with an arbitrary seed $I_{N_{\text{start}}} = 1$ and $I_{N_{\text{start}}+1} = 0$ for a sufficiently large start index N_{start}. As n decreases the iterates obtained from (26) will successively approach the correct solution. The sequence of iterates can be normalized if $I_0(t)$ is calculated by a separate routine. Once a sufficient number of filter coefficients has been calculated, it is easy to determine how many that are actually needed from the condition $\Sigma_{n=-N}^{N} T(n; t) \geq 1 - \epsilon_{\text{trunc}}/M$. A more detailed investigation as well as an algorithm generating the filter coefficients $T(n; t)$ can be found in [15, Sections 5 and A.3].

Another possibility is of course to start from (22) and perform the convolutions in the frequency domain instead.

VII. Numerical Approximations

A. Sampled Gaussian Kernel

A commonly adapted technique to implement the scale-space theory for discrete signals has been to discretize the convolution integral (1) using the rectangle rule of integration. This leads to the approximation formula

$$\tilde{L}(x; t) = \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-n^2/2t} f_{\text{in}}(x - n), \tag{27}$$

i.e., discrete convolution with the sampled Gaussian kernel. However, this representation might lead to undesirable effects. One can show (see [15, Section B.1]) that the transformation from the zero level $\tilde{L}(x; 0)$ to a higher level always preserves the number of local extrema (zero-crossings), but that the transformation from an arbitrary low level $\tilde{L}(x; t_1)$ to an arbitrary higher level $\tilde{L}(x; t_2)$ is in general *not* a scale-space transformation. Thus, we are not guaranteed that the amount of structure will decrease with scale. More precisely we state the following.

Proposition 12: The transformation from a low level $t_1 \geq 0$ to an arbitrary higher level $t_2 > t_1$ in the representation (27) generated by discrete convolution with the sampled Gaussian kernel is a scale-space transformation if and only if t_1 is zero or the ratio t_2/t_1 is an odd integer.

Proof: See Appendix A.5.

The result constitutes an example of the fact that properties derived in the continuous case might be violated after discretization. The main reason why the scale-space property fails to hold between arbitrary levels is because the semigroup property of the Gaussian kernel is not preserved after discretization.

B. Discretized Diffusion Equation

The scale-space family generated by (17) and (19) can be interpreted in terms of a discretized version of the diffusion equation. It is not difficult to verify the following (see below). A more constructive proof is given in [15, Section 6.4].

Theorem 6: Given a discrete signal $f: Z \to R$ in l_1, the discrete scale-space representation $L(x; t) = \Sigma_{n=-\infty}^{\infty} T(n; t) f(x - n)$ is the solution of the system of ordinary differential equations

$$\frac{\partial L(x; t)}{\partial t} = \frac{1}{2}\left(L(x+1; t) - 2L(x; t) + L(x-1; t)\right)$$
$$x \in Z \tag{28}$$

with initial conditions $L(x; 0) = f(x)$, i.e., the system of differential equations obtained if the diffusion equation (2) is discretized in space but solved analytically in time.

Proof: From the relation $2I_n'(t) = I_{n-1}(t) + I_{n+1}(t)$ for modified Bessel functions [1, (9.6.26)] one easily shows that the kernel $T(n; t) = e^{-t} I_n(t)$ satisfies:

$$\frac{\partial T(n; t)}{\partial t} = \frac{1}{2}\left(T(n+1; t) - 2T(n; t) + T(n-1; t)\right). \tag{29}$$

The rest follow from straightforward calculations. The regularity condition on f justifies a change of order between differentiation and infinite summation. □

This provides another motivation for the selection of T as the canonical discrete scale-space kernel. If (28) is further discretized in scale using Euler's method, we obtain the iteration formula:

$$L_{i,k+1} = \frac{\Delta t}{2} L_{i+1,k} + (1 - \Delta t) L_{i,k} + \frac{\Delta t}{2} L_{i-1,k}. \tag{30}$$

Proposition 10 states that the corresponding kernel is a scale-space kernel if and only if

$$\Delta t \leq \tfrac{1}{2}. \tag{31}$$

From Theorem 3, one can easily show that all symmetric scale-space kernels with finite support can be derived from kernels of this latter form. Hence, they provide a possible set of primitive kernels for the scale-space with a discrete scale parameter mentioned in Section V-A.

Proposition 13: All symmetric discrete scale-space kernels with finite support arise from repeated application of the discretization of the diffusion equation (30), using, if necessary, different $\Delta t_k \in [0, \frac{1}{2}]$.

These results all seem to point in the same direction. The natural way to apply the scale-space theory to discrete signals is apparently by discretization of the diffusion equation, not the convolution integral.

VIII. From One to Two Dimensions

We have discussed discrete aspects of the one-dimensional scale-space theory. Positivity and unimodality properties have been shown to hold for scale-space kernels as well as their Fourier transforms. We saw that the interesting kernels could be completely characterized in terms of Pólya frequency sequences, which possess an explicit expression for their generating functions.

Then we introduced a continuous scale parameter and showed that the only reasonable way to define a scale-space for discrete signals is by convolution with the one-parameter family of kernels $T(n; t) = e^{-t}I_n(t)$, where I_n are the modified Bessel functions of integer order. Similar arguments applied in the continuous case uniquely lead to the Gaussian kernel. The kernel T does also have the attractive property that it is equivalent to the analytical solution of a certain discretization of the diffusion equation. The idea of a continuous scale parameter even for discrete signals is of considerable importance, since it permits arbitrary degrees of smoothing, i.e., we are no longer restricted to specific predetermined levels of scale. We saw that scale-space violations might occur in the family of representations generated by discrete convolution with the sampled Gaussian kernel.

The extension to two dimensions is not obvious, since it is possible to show that there does not exist any nontrivial kernel on R^2 or Z^2 with the property that it never introduces new local extrema. Lifshitz and Pizer [14] present an illuminating counterexample:

Imagine a two-dimensional image function consisting of two hills, one of them somewhat higher than the other one; see Fig. 4. 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. The the top of the lower hill is no longer a local maximum. Let this configuration be the input image. When 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.

The same argument can be carried out in the discrete case. Of course, we have to consider connectivity when we define what we mean by local extrema. But this question is only of formal nature. Given an arbitrary nontrivial convolution kernel, it is always possible to create a counterexample. Therefore, it is not clear what we should mean with a scale-property in two space dimensions. We cannot generalize the formulation in terms of zero-crossings either. From the counterexample it is apparent that a level curve might split into two during erosion. Consequently, we cannot expect to find a nontrivial kernel never increasing the number of zero-crossing curves either.

Anyway, we should not be too disappointed. In some sense the decomposition of the scene is not intuitively preposterous despite its consequences. The narrow ridge is a fine-scale phenomenon and should subsequently disappear before the coarse-scale peaks. In this case it is rather the measure on structure than the smoothing method which is wrong.

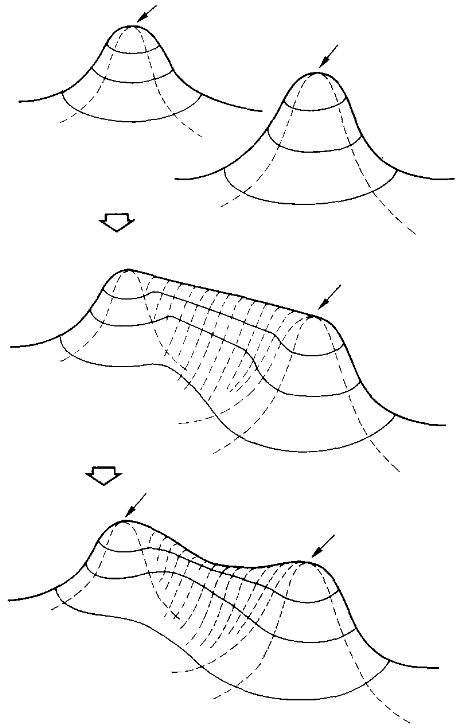


Fig. 4. New local extrema can be created by the diffusion equation in the two-dimensional case.

Therefore, when extending the theory to higher dimensions, we should not be too locked to the previously given definition of a discrete scale-space kernel. In one dimension the number of local extrema is a good measure of structure on which a theory can be founded—in two dimensions obviously not. Instead the previously given treatment should be understood in a wider sense as a characterization of which one-dimensional convolution transformations can be regarded as smoothing transformations.

Is it true that the discrete analog of the Gaussian kernel used as a separated kernel is the natural discrete kernel in two dimensions? If one, due to computational considerations, wants to use separable discrete kernels, one could, of course, heuristically argue that the kernel should at least have a good performance in one dimension. Another indication in that direction is obtained if one studies a discretized version of the two-dimensional diffusion equation. In [15, Section B.3] it is shown that separated convolution with the one-dimensional discrete analog of the Gaussian kernel describes the solution of the system of ordinary differential equations, which appears if the diffusion equation is discretized in space but not in time (scale).

In the next section we will develop a two-dimensional theory based on somewhat modified axioms, which however in one dimension turns out to give the same result as the previous formulation. In a special case the resulting scale-space representation will be reduced to separated convolution with the discrete analog of the Gaussian kernel.

IX. 2-D Discrete Scale-Space Formulation

From the discussion in the previous section is clear that the one-dimensional treatment cannot be generalized directly to higher dimensions. However, an important point with the study we have performed, is that we have acquired a deep understanding about which one-dimensional linear transformations can be regarded as smoothing transformations. We have also shown that the only reasonable way to convert the one-dimensional scale-space theory from continuous images to discrete images is by discretization of the diffusion equation.

Koenderink and van Doorn [11] derive the two-dimensional scale-space for continuous images from three assumptions—causality, homogeneity, and isotropy. The leading idea is that every gray-level at a coarse level of scale should be possible to trace to the same gray-level at a finer level of scale. In other words, no new gray-level surfaces[8] should be created in the scale-space representation when the scale parameter increases (see Fig. 5). Using differential geometry they show that these requirements uniquely lead to the diffusion equation, or equivalently to convolution with the Gaussian kernel.

It is of course impossible to apply these ideas directly, since there does not exist any direct correspondences to level curves and differential geometry in the discrete case. However, an alternative way to express the previous ideas is to require that *if for some scale level t_0 a point (x_0, y_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. In other words, *local extrema should not be enhanced when the scale parameter is increased continuously*.

It is clear that this formulation is equivalent to the formulation in terms of gray-levels for continuous images, since if the gray-level value at a local maximum (minimum) would increase (decrease) a new gray-level would be created. Inversely, if a new gray-level is created then some local maximum must have increased or some local minimum must have decreased.

An intuitive description of this requirement is that it prevents local extrema from being enhanced and from "popping up from nowhere" when the scale parameter increases. As we have seen earlier, we cannot prevent the number of local extrema from ever being increased. However, the idea is that those creation events should be "few."

[8]With a gray-level surface we mean an isosurface in scale-space, i.e., a connected set of points $(x, y; t) \in R^2 \times R$ such that $L(x, y; t) = z_0$ for some gray-level value z_0.

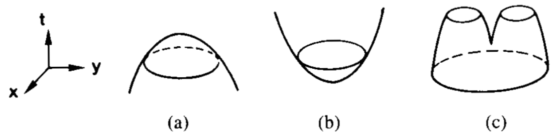


Fig. 5. Gray-level surfaces $L(x, y; t) = z_0$. (a) Causal (and generic) gray-level surface. (b) Noncausal (and impossible) gray-level surface. (c) Gray-level surface corresponding to the example in Fig. 4 where one gray-level curve splits into two.

Below we will show that this condition combined with a continuous scale parameter means a strong restriction on the smoothing method also in the discrete case, and we will again obtain a discretized version of the diffusion equation. In a special case the resulting scale-space representation will be reduced to the family of functions generated by separated convolution with the discrete analog of the Gaussian kernel $T(n; t)$.

A. Definitions

Before getting into the detailed scale-space formulation we will need to make a few definitions. The eight-neighbors of a point $(x, y) \in Z^2$ will be denoted $N_8(x, y)$. If the central point is included as well we will use the notation $N_8^+(x, y)$. The notion of extremum points will be as follows.

Definition 2: A point (x, y) is said to be a local maximum point for a function $g: Z^2 \to R$ if $g(x, y) \geq g(\xi, \eta)$ for all $(\xi, \eta) \in N_8(x, y)$.

Definition 3: A point (x, y) is said to be a local minimum point for a function $g: Z^2 \to R$ if $g(x, y) \leq g(\zeta, \eta)$ for all $(\xi, \eta) \in N_8(x, y)$.

The final result will be expressed in terms of two common discrete operators approximating the two-dimensional Laplace operator $\partial^2/\partial x^2 + \partial^2/\partial y^2$ namely the five-point operator ∇_5^2 and the cross operator $\nabla_\times^2$, defined by[9]:

$$(\nabla_5^2 f)(x, y) = (f(x-1, y) + f(x+1, y) + f(x, y-1) + f(x, y+1) - 4f(x, y)) \quad (32)$$

$$(\nabla_\times^2 f)(x, y) = \tfrac{1}{2}(f(x-1, y-1) + f(x-1, y+1) + f(x+1, y-1) + f(x+1, y+1) - 4f(x, y)). \quad (33)$$

B. Axiomatic 2-D Discrete Scale-Space Construction

When we construct the scale-space for two-dimensional discrete images we follow the ideas from the one-dimensional case, see Section V-B. We start by postulating that the scale-space should be generated by convolution with

[9]In our considerations the step length h is set to 1.

a one-parameter family of kernels, i.e., $L(x, y; 0) = f(x, y)$ and $L(x, y; t) = \Sigma_{m=-\infty}^{\infty} \Sigma_{n=-\infty}^{\infty} T(m, n; t) f(x - m, y - n)$ if $t > 0$. As mentioned earlier, this form on the smoothing formula corresponds to the requirements about linear shift-invariant smoothing and a continuous scale parameter. We want both coordinate directions to be processed identically. Therefore all kernels should be symmetric. We will also impose a semigroup condition on the family T. This means that all scale levels will be treated similarly and that the transformation from a lower scale level to a higher scale level will always be given by convolution with a kernel from the family, compare to (18).

The smoothing criterion will be the requirement about local extrema given in the previous section. It is convenient to express it as a condition on the derivative of the scale-space family with respect to the scale parameter. (In the proof below it will be shown that the requirements on T, combined with a continuity condition, mean that L is differentiable.)

Theorem 7 (Necessity): Given any two-dimensional image $f: Z^2 \to R$ let $L: Z^2 \times R_+ \to R$ be a one-parameter family of functions defined by $L(x, y; 0) = f(x, y)$ and

$$L(x, y; t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} T(m, n; t) f(x - m, y - n) \qquad t > 0 \tag{34}$$

where $T: Z^2 \times R_+ \to R$ is a one-parameter family of kernels in l_1 satisfying

- the semigroup property $T(\cdot, \cdot; s) * T(\cdot, \cdot; t) = T(\cdot, \cdot; s + t)$
- the symmetry constraints $T(-x, y; t) = T(x, y; t)$, $T(y, x; t) = T(x, y; t)$
- the continuity[10] requirement $\| T(\cdot, \cdot; t) \delta(\cdot, \cdot) \|_1 \to 0$ when $t \downarrow 0$.

For all images f it is required that if for some scale level t_0 a point (x_0, y_0) is a local extremum point for the mapping $(x, y) \mapsto L(x; y; t_0)$ then the derivative of L with respect to t must satisfy

$$\frac{\partial L}{\partial t}(x_0, y_0; t_0) \le 0 \quad \text{if } (x_0, y_0) \text{ is a local maximum point} \tag{35}$$

$$\frac{\partial L}{\partial t}(x_0, y_0; t_0) \ge 0 \quad \text{if } (x_0, y_0) \text{ is a local minimum point.} \tag{36}$$

Then necessarily, the scale-space family L must obey the differential equation

$$\frac{\partial L}{\partial t} = \alpha \nabla_5^2 L + \beta \nabla_\times^2 L \tag{37}$$

for some $\alpha \ge 0$ and $\beta \ge 0$.

Proof: See Appendix A.3. □

From the proof it is apparent that if similar arguments are applied in the one-dimensional case, we are uniquely led to the one-dimensional scale-space concept developed earlier in Theorem 4 and Theorem 6. This shows that, combined with the requirements about a continuous scale parameter and semigroup structure, the condition about suppression of local extrema is in one dimension equivalent to the condition about decreasing number of local extrema.

Consequently, also this formulation in terms of local extrema has lead to a discretized version of the diffusion equation. But here in the two-dimensional case there is apparently another degree of freedom in the class of possible smoothing operators, since a linear combination of the two discrete Laplacian operators ∇_5^2 and $\nabla_\times^2$ is admitted on the right hand side of the differential equation. The effects of these parameters will be illuminated in Section IX-C. However, first we will show the sufficiency, which is much easier to establish.

Theorem 8 (Sufficiency): Given a discrete image $f: Z^2 \to R$ let $L: Z^2 \times R_+ \to R$ be the representation generated by $L(x, y; 0) = f(x, y)$ and

$$\frac{\partial L}{\partial t} = \alpha \nabla_5^2 L + \beta \nabla_\times^2 L \tag{38}$$

where $\alpha \ge 0$ and $\beta \ge 0$. If for some scale level t_0 a point (x_0, y_0) is a local extremum point for the mapping $(x, y) \mapsto L(x, y; t_0)$ then the derivative of L with respect to t satisfies

$$\frac{\partial L}{\partial t}(x_0, y_0; t_0) \le 0 \quad \text{if } (x_0, y_0) \text{ is a local maximum point} \tag{39}$$

$$\frac{\partial L}{\partial t}(x_0, y_0; t_0) \ge 0 \quad \text{if } (x_0, y_0) \text{ is a local minimum point.} \tag{40}$$

Proof: The result follows almost trivially if we rewrite the differential equation on the form

$$\begin{aligned}
\frac{\partial L}{\partial t}(x, y; t) = {} & \alpha[L(x, y - 1; t) - L(x, y; t)] \\
& + \alpha[L(x, y + 1; t) - L(x, y; t)] \\
& + \alpha[L(x - 1, y; t) - L(x, y; t)] \\
& + \alpha[L(x + 1, y; t) - L(x, y; t)] \\
& + \tfrac{1}{2}\beta[L(x - 1, y - 1; t) - L(x, y; t)] \\
& + \tfrac{1}{2}\beta[L(x + 1, y - 1; t) - L(x, y; t)] \\
& + \tfrac{1}{2}\beta[L(x - 1, y + 1; t) - L(x, y; t)] \\
& + \tfrac{1}{2}\beta[L(x + 1, y + 1; t) - L(x, y; t)].
\end{aligned} \tag{41}$$

[10] δ denotes the two-dimensional discrete delta function, which assumes the value 1 at (0, 0) and is zero elsewhere. In fact, this condition about continuity in norm can be weakened; see [16].

If for some scale level t, a point (x, y) is a local maximum point then all differences (within brackets) become nonpositive, which means that $\partial L/\partial t\,(x, y; t) \le 0$ provided that $\alpha \ge 0$ and $\beta \ge 0$. Similarly, if a point is a local minimum point the differences are all nonnegative and $\partial L/\partial t(x, y; t) \ge 0$. □

C. Parameter Determination

If (37) is rewritten on the form

$$\frac{\partial L}{\partial t} = C((1-\gamma)\nabla_5^2 L + \gamma \nabla_\times^2 L) \tag{42}$$

one realizes that the interpretation of the parameter C is just a trivial rescaling of the scale parameter. Thus, without loss of generality[11] we may set C to $\frac{1}{2}$ in order to get the same scaling constant as in the one-dimensional case (28). What is left to investigate is how the remaining degree of freedom in the parameter $\gamma \in [0, 1]$ affects the scale-space representation.

If $\gamma = 1$ then an undesired situation appears. Since the cross-operator only links diagonal points, the system of ordinary differential equations given by (42) can be split into two *uncoupled* systems, one operating on the points with even coordinate sum $x + y$ and the other operating on the points with odd coordinate sum. It is clear that this is really an unwanted behavior, since even after a substantial amount of "blurring" the gray-level landscape may still have a rather saw-toothed shape.

Further arguments showing that γ must not be too large can be obtained if one studies the kernel, which describes the transformation from a fine level to a coarse level in the scale-space representation (42). It is not difficult to show, see [16], that its Fourier transform is

$$\begin{aligned}\psi_T(u, v) &= \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} T(m, n; t)\, e^{-i(mu+nv)} \\ &= \exp\big[-(2-\gamma)\,t + (1-\gamma)(\cos(u) \\ &\quad + \cos(v))\,t + \gamma\cos(u)\cos(v)\,t\big].\end{aligned} \tag{43}$$

It is an easy exercise to verify (see [16]) that this function is unimodal if and only if $\gamma \le \frac{1}{2}$.

The transformation kernel is separable if and only if its Fourier transform is separable, i.e., if and only if $\psi_T(u, v)$ can be written on the form $U_T(u)\, V_T(v)$ for some functions U_T and V_T. From (43) we realize that this separation is possible if and only if $\gamma = 0$. Hence, we have the following.

Proposition 14: The convolution kernel associated with the scale-space representation defined by $L(x, y; t) = f(x, y)$ and

$$\frac{\partial L}{\partial t} = \frac{1}{2}((1-\gamma)\nabla_5^2 L + \gamma\nabla_\times^2 L) \tag{44}$$

is separable if and only if $\gamma = 0$. Then L is given by

$$L(x, y; t) = \sum_{m=-\infty}^{\infty} T(m; t) \sum_{n=-\infty}^{\infty} T(n; t) \cdot f(x-m, y-n) \qquad t > 0 \tag{45}$$

where $T(n; t) = e^{-t} I_n(t)$ and I_n are the modified Bessel functions of integer order.

If (42) is discretized further is scale using Euler's explicit method with scale step Δt (see [16]), we get an iteration kernel with the coefficients

$$\begin{bmatrix} \frac{\gamma\Delta t}{4} & \frac{(1-\gamma)\Delta t}{2} & \frac{\gamma\Delta t}{4} \\ \frac{(1-\gamma)\Delta t}{2} & 1-(2-\gamma)\Delta t & \frac{(1-\gamma)\Delta t}{2} \\ \frac{\gamma\Delta t}{4} & \frac{(1-\gamma)\Delta t}{2} & \frac{\gamma\Delta t}{4} \end{bmatrix}. \tag{46}$$

Clearly, this kernel is unimodal if and only if $\gamma \le \frac{2}{3}$. One can show, see [16], that it is separable if and only if $\gamma = \Delta t$. In that case the corresponding one-dimensional kernel is a discrete scale-space kernel in the sense given in Definition 1 if and only if $\Delta t \le \frac{1}{2}$; see (31). This gives a further indication that γ should not exceed $\frac{1}{2}$.

It is worth mentioning, that if the extremum definitions, Definition 2 and Definition 3, would have been based on four-neighbors instead of eight-neighbors, then $\gamma = 0$ would have appeared as a necessary condition in Theorem 7; see [16].

If $\gamma = \frac{1}{3}$ we get the nine-point operator ∇_9^2; see [5, Section 7.7.2]. If is not difficult to show that (see [16]) for large spatial scales, this value of γ gives the "most" isotropic second order approximation of the continuous Laplacian operator. It is not clear that rotational invariance is a primary quality to be aimed at in the discrete case, since we are anyway locked to a fixed square grid. But if we use a nonzero value of γ, the discrete scale-space representation can always be calculated efficiently in the Fourier domain, using (43).

We leave the question about definite selection of γ open. However, from a computational point of view it seems very plausible that $\gamma = 0$ should not be a too bad choice. As we will see in the next section the analytical expressions for some derived quantities will also become simple in this case. A possible disadvantage with that approach is that it emphasizes the x- and y-directions as being special directions.

D. Implementational Effects

The scale-space representation obtained from the discrete theory has some implementational advantages compared to the commonly adapted approach, where the scale-space implementation is based on different versions of the sampled Gaussian kernel. Consider for instance the computation of the Laplacian of the Gaussian $\nabla^2 G$ of an image f. It is well-known that $\nabla^2 G$ is not a separable ker-

[11] The case when $C = 0$ is obviously not interesting since then all scale-space representations would be equal.

nel—a clear disadvantage in terms of computational efficiency. It is also known that the straightforward implementation consisting of smoothing with the sampled Gaussian kernel followed by application of a discrete Laplacian gives unsatisfactory results, since the values obtained in this way deviate too much from the sampled values of $\nabla^2 G$. A common approach to circumvent this problem has been by calculation of difference of Gaussians (DOG) instead [17]. However, this method only gives an approximate result, and there is a tradeoff between cancellation of digits and accuracy. It also requires computation of two smoothed representations instead of one.

One interpretation of the Laplacian of the Gaussian in the continuous case is as the derivative of the scale-space representation with respect to the scale parameter.[12] This connection gives us a natural way to define *the discrete analog of the Laplacian of the Gaussian*, namely as *the derivative of the discrete scale-space family with respect to the scale parameter*, or equivalently as the result of application of the discrete Laplacian operator Δ_d on the scale-space representation. From the diffusion equation (42) we get

$$\Delta_d L = \frac{\partial L}{\partial t} = \frac{1}{2}\left((1-\gamma)\nabla_5^2 L + \gamma \nabla_\times^2 L\right) = \Delta_d (T * f)$$
$$= T * (\Delta_d f) = (\Delta_d T) * f. \tag{47}$$

In this discrete case Δ_d commutes with the smoothing kernel and we can compute the discrete analog of the Laplacian of the Gaussian in two sweeps—a smoothing step followed by application of the discrete Laplacian or a discrete Laplacian step followed by smoothing. We could of course also calculate the Laplacian of the smoothing kernel as a first step and then convolve the result with the image, but then we lose the separability.[13] However, note that with the discrete formulation *all methods give the same result since the (discrete) smoothing operator commutes with the (discrete) Laplacian*. Preferably, we should use the same value γ in all discrete Laplacian operators. The computations required to calculate the discrete analog of the Laplacian of the Gaussian of an image result are, if $\gamma = 0$, just one separable two-dimensional smoothing step and an efficient application of the discrete Laplacian.

The discrete scale-space does also provide a convenient formulation of gradient calculations if $\gamma = 0$. Let δ_x denote the central difference operator in the x-direction defined by $(\delta_x f)(x, y) = \frac{1}{2}(f(x+1, y) - f(x-1, y))$. Then, similarly to the previous case $\delta_x L$ can be computed either by application of δ_x on the smoothed image, the original image or on the smoothing kernel. The effect of the gradient calculation is given by the effect δ_x has on the one-dimensional kernel applied in the x-direction. From the recurrence relation for the modified Bessel functions (26) we get an explicit analytical expression for $\delta_x T(x; t)$, namely

$$\delta_x T(x; t) = \frac{1}{2} e^{-t}\left(I_{x+1}(t) - I_{x-1}(t)\right)$$
$$= \frac{1}{2} e^{-t}\left(-\frac{2x}{t} I_x(t)\right) = -\frac{x}{t} T(x; t). \tag{48}$$

Note the similarity with the derivative of the Gaussian kernel $(\partial/\partial x)\, G(x; t) = -(x/t)\, G(x; t)$. If one instead would use the approach with a sampled Gaussian it is clear that convolution with the sampled x-gradient of the Gaussian would not have given the same result as application of δ_x on the scale-space representation.

Another minor problem concerns the behavior of the sampled Gaussian kernel for small values of t. It is well-known that under these circumstances the central coefficient of the sampled Gaussian can become very large and the sum of the corresponding filter coefficients will exceed[14] one, sometimes substantially. However, with the discrete approach the kernels are inherently bounded for small values of t since they approach the discrete delta function (instead of the continuous one) when t tends to zero.

The effects mentioned in this section are all due to the difference between continuous theory and discrete implementation. *The main reason why they arise is because the involved operators, which commute in the continuous case, do not commute when the discretization operator is involved* (compare with violated semigroup property discussed in Section VII-A). With the discrete scale-space theory presented in this paper we feel that we have accomplished a structured way to eliminate this kind of problem.

E. 2-D Summary and Discussion

The proper way to apply the scale-space theory to two-dimensional discrete images is apparently by discretization of the diffusion equation. Starting from a requirement that local extrema must not be enhanced when the scale parameter is increased continuously, we have shown that a necessary and sufficient condition for a family of derived representations to be a scale-space family is that it satisfies the differential equation

$$\frac{\partial L}{\partial t} = \frac{1}{2}\left((1-\gamma)\nabla_5^2 L + \gamma \nabla_\times^2 L\right) \tag{49}$$

where γ is a real constant in $[0, 1]$. Our recommendation is that γ should not exceed $\frac{1}{2}$. $\gamma = 0$ gives a separable convolution kernel, while $\gamma = \frac{1}{3}$ leads to a spatially more isotropic smoothing effect on coarse scale objects. In the

[12]With this terminology the zero-crossings of the Laplacian of the Gaussian of an image are those points in scale-space, which are locally stationary in t.

[13]As mentioned earlier, the convolution kernel is separable only if $\gamma = 0$.

[14]It has been suggested that this effect should be compensated for by renormalization of the filter coefficient sequence. But this operation does not solve the major problem since the mutual relation between the coefficients remains unchanged anyway. It only leads to a rescaling of the output image. The problem with the sampled Gaussian kernel for small values of t is rather that it appears as having a smaller t-value than it should.

separable case the scale-space representation can be calculated by separated convolution with the presented one-dimensional discrete analog of the Gaussian kernel $T(n; t)$.

We have seen that the discrete scale-space representation given by discretization of the diffusion equation has computational advantages compared to the commonly used approach, where the scale-space implementation is based on various versions of the sampled Gaussian kernel. It can be expected that the difference is largest for small values of the scale parameter, when the sampled Gaussian kernel and the discrete analog of the Gaussian kernel deviate as most. When the scale parameter increases these two kernels approach each other (see [15, Section 5.3] and Fig. 9 in Appendix A.4); and we might expect that the difference becomes smaller. This effect can also be understood from another point of view. At coarse levels of scale the large scale phenomena dominate in the scale-space representation, which means that the grid effects become smaller, since a characteristic length in the smoothed image will be large compared to the distance between adjacent grid points. It is difficult to say generally how large the numerical effects are in an actual implementation and how seriously they affect the output result, since this is very much determined by the algorithms working on the scale-space representation and the goal of the analysis in which the scale-space part is just one of the modules. However, in Fig. 6 we have tried to visualize how some measures on the difference between the sampled Gaussian kernel and the discrete analog of the Gaussian kernel behave as a function of the scale parameter. The graphs verify that the difference is largest for small values of t and show that it increases with higher order differences. Also note the large difference between the sampled second derivative of the Gaussian kernel and the second difference of the sampled Gaussian kernel.

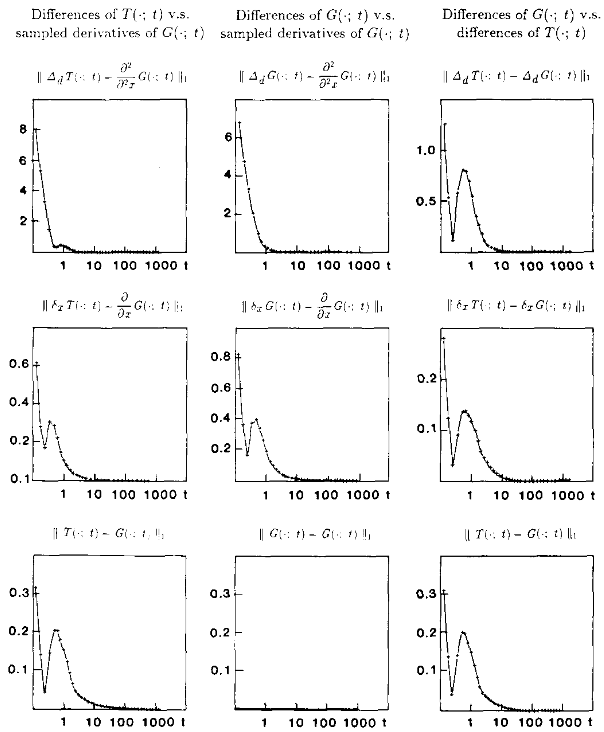


Fig. 6. l_1 norms of some differences between the sampled Gaussian $G(n; t)$ and the discrete analog of the Gaussian kernel $T(n; t)$ in the one-dimensional case. Here, Δ_d denotes the second difference operator defined by $(\Delta_d f)(x) = f(x + 1) - 2f(x - 1)$.

The scale-space concept developed here in two dimensions can be generalized directly to higher dimensions as well. Then the right hand side becomes a convex combination of the possible discrete second-order approximations of the n-dimensional Laplacian operator. In the separable case the scale-space representation can be calculated by separated convolution with the one-dimensional discrete analog of the Gaussian kernel.

Finally, it should be explicitly stressed that the presented discrete scale-space theory is closely linked to the continuous scale-space theory through the discretization of the diffusion equation. This means that continuous results can be transferred to discrete implementation provided that the discretization is performed correctly. The discussion in the previous section is intended to exemplify the technique.

X. Conclusions

This paper gives a basic and extensive treatment of discrete aspects of the scale-space theory. A genuinely discrete scale-space theory is developed and its connection to the continuous scale-space theory is explained. Special attention is given to discretization effects, which occur when results from the continuous scale-space theory are to be implemented computationally. The one-dimensional problem is solved completely in an axiomatic manner. The two-dimensional problem is more complex, but we answer the question about how the two-dimensional discrete scale-space should be constructed. The main results can be summarized as follows (references to central theorems and appropriate parts of this paper are given within parentheses):

- The proper way to apply the scale-space theory to discrete signals and discrete images is by discretization of the diffusion equation, *not* the convolution integral (Theorems 4, 6, 7, Proposition 12, and Sections IX-D, VII).
- The discrete scale-space obtained in this way can be described by convolution with the kernel $T(n; t)$, which is the discrete analog of the Gaussian kernel (Theorem 4, Proposition 14, and Section IX-C).
- A scale-space implementation based on the sampled Gaussian kernel might lead to undesirable effects and computational problems, especially at fine levels of scale (Proposition 12 and Section IX-D).
- The one-dimensional discrete smoothing transformations can be characterized exactly and a complete catalog is given (Theorems 1, 2).
- All finite support one-dimensional discrete smoothing transformations arise from repeated averaging over

two adjacent elements (Theorem 3 and Propositions 4, 5). The kernel $T(n;\, t)$ describes the limit case of such an averaging process ([15, Section 6.4]).

- The symmetric one-dimensional discrete smoothing kernels are nonnegative and unimodal, both in the spatial and the frequency domain (Propositions 2, 3, 7, 9 and Section IV).

The important idea of the scale-space concept suggested in this paper is that the discrete nature of the implementation has been taken into account already in the theoretical formulation of the scale-space representation.

XI. Philosophy

The formulation in terms of the diffusion equation appears to be a natural unification of the existing scale-space theory for continuous signals and the presented scale-space theory for discrete signals. One could say that the *primary* formulation of the scale-space theory is by the *diffusion equation*. Then,

- The Gaussian kernel appears as the fundamental solution of the *continuous* diffusion equation.
- The discrete analog of the Gaussian kernel is the fundamental solution of the *discrete* diffusion equation.

During recent years "Gaussian smoothing" has become a widespread concept in the computer vision society. However, in view of these results we should rather say "diffusion smoothing."

Appendix
Proofs

A.1 Proof of Proposition 6

Because of Proposition 2, it is sufficient to study kernels having only nonnegative filter coefficients. Assume that the convolution matrix $C^{(N)}$ has a real negative eigenvalue for some dimension N and a corresponding real eigenvector v. Let I_N be the index set $1..N$. Create an input signal f_{in}, which is equal to the components of v for $x \in I_N$ and zero otherwise. Convolve this signal with the kernel. Then for $x \in I_N$ the values of $K * f_{\text{in}}$ will be equal to the corresponding components of $C^{(N)}v$ (see Fig. 7). As v is an eigenvector with a negative eigenvalue the components of $C^{(N)}v$ and v have opposite signs. This means that v, $C^{(N)}v$, and $K * f_{\text{in}}$ all have the same number of internal zero-crossings provided that we observe only the components in I_N.

The reversal of these components and the positivity of the filter coefficients guarantee that at least one additional zero-crossing will occur in the output signal. Let α denote the index of the first nonzero component of f_{in}. If $f_{\text{in}}(\alpha)$ is positive (negative) then due to the negative eigenvalue $K * f_{\text{in}}(\alpha)$ will be negative (positive). Since the filter coefficients are nonnegative the first nonzero component of $K * f_{\text{in}}$ (at position β) will have the same sign as $f_{\text{in}}(\alpha)$, i.e., positive (negative). Consequently, we have found at least one additional zero-crossing in $K * f_{\text{in}}$ between these two positions (α and β). The same argument can be carried out at other end point producing another scale-space violation. This shows that K cannot be a scale-space kernel. □

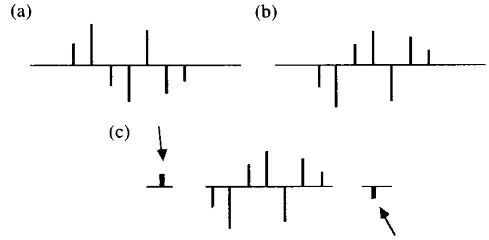


Fig. 7. (a) The eigenvector v. (b) The components of $C^{(N)}v$ having indexes $1..N$. (c) The components of $K * f_{\text{in}}$.

A.2 Proof of Proposition 8

We will introduce a temporary definition. If $\boldsymbol{x}$ is a vector of length L let $V(\boldsymbol{x})$ denote the number of zero-crossings in the sequence of components $x_1, x_2, \cdots, x_L, x_1$. By verification one shows that the eigenvalues λ_m and eigenvectors v_m of $C_C^{(M)}$ are

$$\lambda_m = \sum_{n=-M}^{M} c_n e^{-(2\pi i m n/(2M+1))} = \sum_{m=-N}^{N} c_n e^{-(2\pi i m n/(2M+1))}$$

$$m = -M..0..M \qquad (50)$$

$$(\boldsymbol{v}_m)_k = \sin\left(\frac{2\pi mk}{2M+1}\right) \qquad m = -M..-1,\; k = -M..0..M$$

$$(\boldsymbol{v}_m)_k = \cos\left(\frac{2\pi mk}{2M+1}\right) \qquad m = 0..M,\; k = -M..0..M. \qquad (51)$$

We note that $V(v_m)$ increases as $|m|$ increases. Further, the eigenvalues $\lambda_m = \psi(2\pi m/(2M+1))$ of $C_C^{(M)}$ are uniformly sampled values of the Fourier transform and a larger value of $|m|$ corresponds to a larger absolute value of the argument of ψ.

Now, assume that the Fourier spectrum is not unimodal. (Without loss of generality we can presuppose that ψ is nonnegative on $[-\pi, \pi]$, because otherwise, according to Proposition 7, the kernel cannot be a scale-space kernel.) Then, as ψ is a continuous function of θ it is possible to find some sufficiently large integer $\tilde{M}$ such that there exists $\theta_\alpha = (2\pi\alpha/(2\tilde{M}+1))$ and $\theta_\beta = (2\pi\beta/(2\tilde{M}+1))$ satisfying $\psi(\theta_\beta) > \psi(\theta_\alpha)$ for some integers $\beta > \alpha$ in $[0, \tilde{M}]$.

To summarize, $C_C^{(\tilde{M})}$ has eigenvalues $\lambda_\beta > \lambda_\alpha$ and corresponding eigenvectors with $V(v_\beta) > V(v_\alpha)$. We will show that this situation leads to a scale-space violation. The scale-space properties are not affected by a scaling factor. Therefore, we can equivalently study $B = (1/\lambda_\beta)\, C_C^{(\tilde{M})}$. For both eigenvectors we define the smallest and largest absolute values $v^{(\text{absmin})}$ and $v^{(\text{absmax})}$ by

$$v^{(\text{absmin})} = \min_{k=1..N} |v_k|; \qquad v^{(\text{absmax})} = \max_{k=1..N} |v_k|. \qquad (52)$$

Let $x = cv_\alpha + v_\beta$ where c is chosen large enough such that $V(x) = V(v_\alpha)$. This can always be achieved if $|c| v_\alpha^{(\text{absmin})} > v_\beta^{(\text{absmax})}$, since then the components of x and v_α will have pairwise same signs. ($v_\alpha^{(\text{absmin})}$ will be strictly positive as all components of v_α are nonzero.) Then consider $Bx = (1/\lambda_\beta)(c\lambda_\alpha v_\alpha + \lambda_\beta v_\beta)$ and study

$$B^k x = c\left(\frac{\lambda_\alpha}{\lambda_\beta}\right)^k v_\alpha + v_\beta. \tag{53}$$

For a fixed value of c we can always find a sufficiently large value of k such that $V(B^k x) = V(v_\beta)$. In a manner similar to above, one verifies that the condition $|c| \, |\lambda_\alpha/\lambda_\beta|^k v_\alpha^{(\text{absmax})} < v_\beta^{(\text{absmin})}$ suffices. Subsequently, $V(B^k x) > V(x)$ which shows that the transformation induced by B^k is not a scale-space transformation. Therefore, B cannot be a scale-space kernel since at least one scale-space violation must have occurred in the series of k successive transformations. □

A.3 Proof of Theorem 7

The proof consists of two parts. In the first step we show that the requirements on T imply that the family L obeys a linear differential equation. In the second step we construct counterexamples from various simple test functions in order to delimit the class of possible operators.

(I): Assume that f is sufficiently regular, i.e., $f \in l_1$, and define a family of operators $\{\mathfrak{I}_t, t > 0\}$ from l_1 to l_1 by $\mathfrak{I}_t f = T(\cdot, \cdot; t) * f$. Due to the conditions imposed on the kernels it will satisfy the relation

$$\lim_{t \to t_0} \left\| (\mathfrak{I}_t - \mathfrak{I}_{t_0}) f \right\|_1 = \lim_{t \to t_0} \left\| (\mathfrak{I}_{t-t_0} - \mathfrak{J})(\mathfrak{I}_{t_0} f) \right\|_1 = 0 \tag{54}$$

where $\mathfrak{J}$ is the identity operator. Such a family is called a strongly continuous semigroup of operators; see [9, pp. 58–59].

A semigroup is often characterized by its infinitesimal generator $\mathfrak{A}$ defined by, see [9, p. 307],

$$\mathfrak{A} f = \lim_{h \downarrow 0} \frac{\mathfrak{I}_h f - f}{h}. \tag{55}$$

The set of elements f for which $\mathfrak{A}$ exists is denoted $\mathfrak{D}(\mathfrak{A})$. (This set is not empty and it never reduces to the zero element. Actually, it is even dense in l_1 [9, p. 307].) If this operator exists, we obtain

$$\lim_{h \downarrow 0} \frac{L(\cdot, \cdot; t+h) - L(\cdot, \cdot; t)}{h} = \lim_{h \downarrow 0} \frac{\mathfrak{I}_{t+h} f - \mathfrak{I}_t f}{h}$$
$$= \lim_{h \downarrow 0} \frac{\mathfrak{I}_h(\mathfrak{I}_t f) - (\mathfrak{I}_t f)}{h} = \mathfrak{A}(\mathfrak{I}_t f) = \mathfrak{A} L(\cdot, \cdot; t). \tag{56}$$

According to a Theorem by Hille and Phillips [9, p. 308] strong continuity implies that $(\partial/\partial t)(\mathfrak{I}_t f) = \mathfrak{A}\mathfrak{I}_t f = \mathfrak{I}_t \mathfrak{A} f$ for all $f \in \mathfrak{D}(\mathfrak{A})$. Hence, the scale-space family L must obey the differential equation

$$\frac{\partial L}{\partial t} = \mathfrak{A} L \tag{57}$$

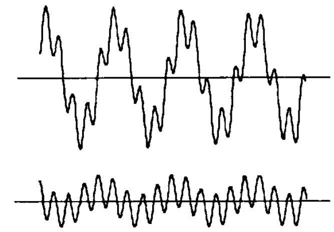


Fig. 8. (a) Input signal consisting of a low frequency component of high amplitude and a high frequency component of low amplitude. (b) In the output signal the low frequency component has been suppressed while the high frequency component remains unchanged. As we see, additional zero-crossings have been introduced.

where $\mathfrak{A}$ is a linear operator. Because of the shift invariance, $\mathfrak{A} L$ can be written as

$$(\mathfrak{A} L)(x, y; t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a_{m,n} L(x - m, y - n; t). \tag{58}$$

(II): The extremum point conditions (35) and (36) mean that $\mathfrak{A}$ must be local, i.e., that $a_{m,n} = 0$ if $|m| > 1$ or $|n| > 1$. This is easily understood from the following counterexample.

First, assume that $a_{\tilde{m},\tilde{n}} > 0$ where either $|\tilde{m}| > 1$ or $|\tilde{n}| > 1$, and define a function $f_1 : Z^2 \to R$ by

$$f_1(x, y) = \begin{cases} \epsilon > 0 & \text{if } (x, y) = (0, 0) \\ 0 & \text{if } (x, y) \in N_8(0, 0) \\ 1 & \text{if } (x, y) = (\tilde{m}, \tilde{n}) \\ 0 & \text{otherwise.} \end{cases} \tag{59}$$

Obviously $(0, 0)$ is a local maximum point for f_1. From (57) and (58), it follows that $(\partial L/\partial t)(0, 0; 0) = \epsilon a_{0,0} + a_{\tilde{m},\tilde{m}}$. It is clear that this value can be positive provided that ϵ has been chosen small enough. Hence, L cannot satisfy (35). In a similar manner one shows that also $a_{\tilde{m},\tilde{n}} < 0$ leads (let $\epsilon < 0$) to a violation against the extremum point condition (36).

Consequently, $a_{\tilde{m},\tilde{n}}$ must be zero if either of $|\tilde{m}|$ or $|\tilde{n}|$ is larger than one. Thereby, (57) will be reduced to

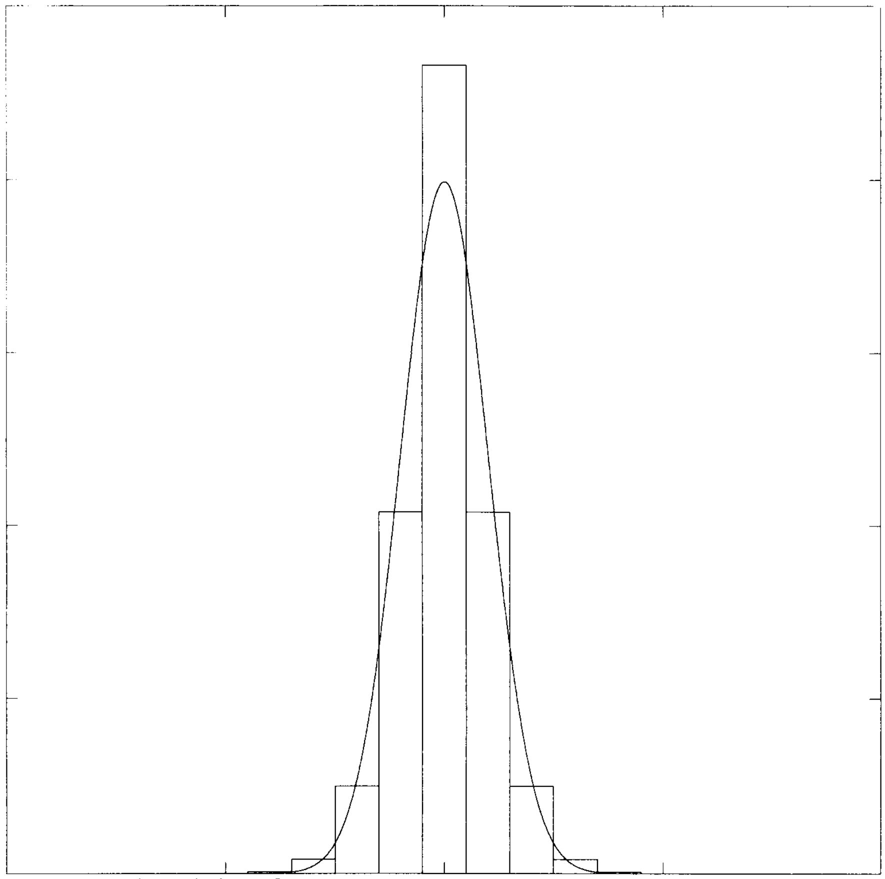
$$\frac{\partial L}{\partial t}(x, y; t) = \sum_{(m,n) \in N_8^+(0,0)} a_{m,n} L(x - m, y - n; t). \tag{60}$$

Due to the symmetry conditions, opposite coefficients must be equal, i.e., $a_{-m,n} = a_{m,n}$ and $a_{n,m} = a_{m,n}$. Thus, (60) can be written

$$\frac{\partial L}{\partial t} = \begin{pmatrix} a & b & a \\ b & c & b \\ a & b & a \end{pmatrix} L. \tag{61}$$

Then, consider the function

$$f_2(x, y) = \begin{cases} 1 & \text{if } (x, y) \in N_8^+(0, 0) \\ 0 & \text{otherwise.} \end{cases} \tag{62}$$



(a)

Fig. 9. The discrete analog of the Gaussian kernel (block diagrams) and the continuous Gaussian kernel (smooth curves) at (a) $t = 1.0$, (b) $t = 10.0$, and (c) $t = 100.0$.

With the given definitions of extremum points it is clear that $(0, 0)$ is both a local maximum point and a local minimum point. Hence $(\partial L/\partial t)(0, 0; 0)$ must be zero and we obtain the relation $4a + 4b + c = 0$. This means that (61) can be split into two components:

$$\frac{\partial L}{\partial t} = \begin{pmatrix} a & b & a \\ b & c & b \\ a & b & a \end{pmatrix} L = \alpha \begin{pmatrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{pmatrix} L + \beta \begin{pmatrix} 1/2 & & 1/2 \\ & -2 & \\ 1/2 & & 1/2 \end{pmatrix} L \qquad (63)$$

provided that $\alpha = b$ and $\beta = 2a$. The condition $4a + 4b + c = 0$ is trivially satisfied.

Finally, by considering the test function

$$f_3(x, y) = \begin{cases} \epsilon > 0 & \text{if } (x, y) = (0, 0) \\ 1 & \text{if } (x, y) = (\tilde{m}, \tilde{n}) \\ 0 & \text{otherwise} \end{cases} \qquad (64)$$

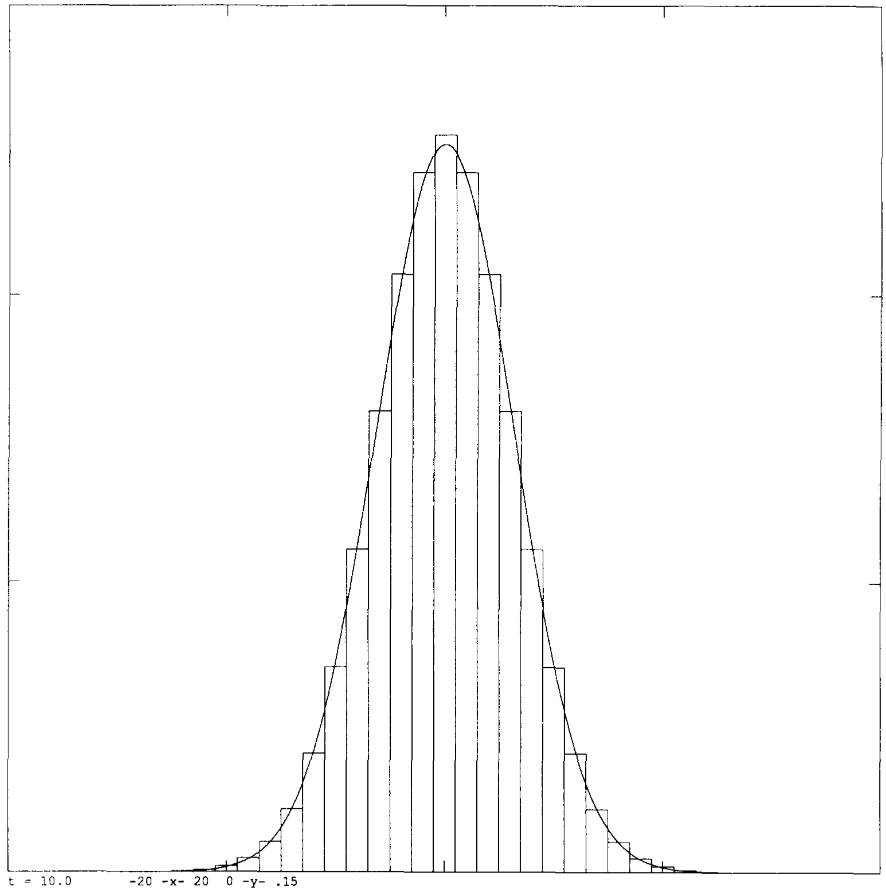
for each $(\tilde{m}, \tilde{n})$ in $N_8(0, 0)$ one easily realizes that $a_{m,n}$ must be nonnegative if $(m, n) \in N_8(0, 0)$. This shows that $\alpha \geq 0$ and $\beta \geq 0$. □

A.4. Kernel Graphs

To illustrate the difference between the discrete analog of the Gaussian kernel and the continuous Gaussian kernel we have drawn their graphs at a few levels of scale. In Fig. 9 the block diagrams represent the values of the discrete analog of the Gaussian kernel and the smooth curves represent the continuous Gaussian kernel. As we can see, the difference between the two kernels is largest at fine levels of scale and becomes smaller as the kernels approach each other at coarser levels of scale.

A.5. Proof of Proposition 12

Assume that we construct the "scale-space" for a discrete signal by convolution with the sampled Gaussian kernel, i.e., given a discrete signal $f: Z \to R$ we define the family of functions $\tilde{L}: Z \times R_+ \to R$ by $\tilde{L}(x; 0) =$



(b)

Fig. 9. (*Continued.*)

$f(x)$ $(x \in Z)$ and

$$\tilde{L}(x; t) = \sum_{n=-\infty}^{\infty} g(n; t) f(x - n) \qquad (x \in Z, t > 0) \tag{65}$$

where

$$g(n; t) = \frac{1}{\sqrt{2\pi t}} e^{-(n^2/2t)} \qquad (n \in Z, t > 0). \tag{66}$$

We will need an expression for the generating function for the discrete kernel corresponding to the sampled Gaussian. For simplicity we let $q_t = e^{-1/2t}$. One can show [21] that

$$\varphi_t(z) = \sum_{n=-\infty}^{\infty} g(n; t) z^n = \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} q_t^{n^2}$$

$$z^n = C_t \prod_{n=0}^{\infty} (1 + q_t^{2n+1} z)(1 + q_t^{2n+1} z^{-1}) \tag{67}$$

where

$$C_t = \frac{1}{\sqrt{2\pi t}} \prod_{n=1}^{\infty} (1 - q_t^{2n}). \tag{68}$$

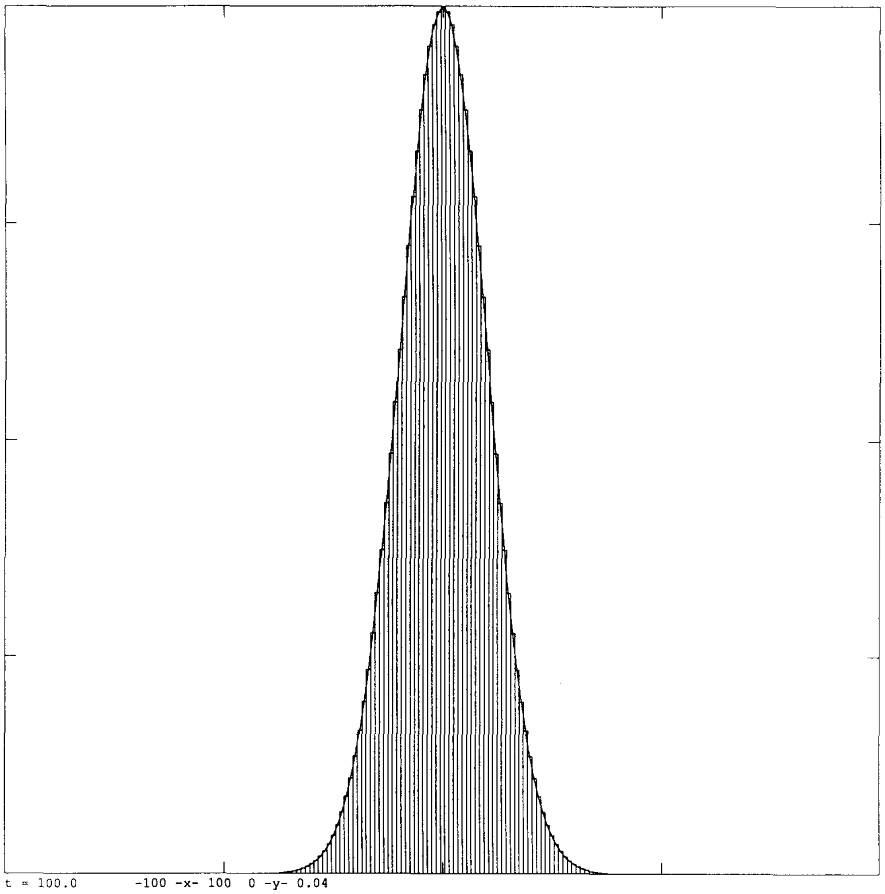
Comparison to the complete characterization of the generating function of a discrete scale-space kernel (16) in Theorem 2 shows that the sampled Gaussian kernel is a discrete scale-space kernel. This means that for any signal f the number of local extrema in $\tilde{L}(x; t)$ $(t > 0)$ does not exceed the number of local extrema in f. However, we will show that this scale-space property does *not* hold between two *arbitrary* levels.

Let t_1 and t_2 be two levels $(t_2 > t_1 > 0)$ of the representation (65) and let φ_{in} be the generating function of the input signal. Then the generating functions of $\tilde{L}(x; t_1)$ and $\tilde{L}(x; t_2)$ are

$$\varphi_{\tilde{L}_1}(z) = \varphi_{t_1}(z)\, \varphi_{\text{in}}(z) \quad \varphi_{\tilde{L}_2}(z) = \varphi_{t_2}(z)\, \varphi_{\text{in}}(z). \tag{69}$$

Let φ_{diff} describe the transformation from $\tilde{L}(x; t_1)$ to $\tilde{L}(x; t_2)$. Thus,

$$\varphi_{\tilde{L}_2}(z) = \varphi_{\text{diff}}(z)\, \varphi_{\tilde{L}_1}(z). \tag{70}$$



(c)

Fig. 9. (*Continued.*)

Combination of (69), (70), and (68) gives

$$\varphi_{\text{diff}}(z) = \frac{\varphi_{\tilde{L}_2}(z)}{\varphi_{\tilde{L}_1}(z)} = \frac{C_{t_2}}{C_{t_1}} \cdot \frac{\prod_{m=0}^{\infty} (1 + q_{t_2}^{2m+1} z)(1 + q_{t_2}^{2m+1} z^{-1})}{\prod_{n=0}^{\infty} (1 + q_{t_1}^{2n+1} z)(1 + q_{t_1}^{2n+1} z^{-1})}. \quad (71)$$

According to the complete characterization of scale-space kernels it follows that the corresponding kernel is a scale-space kernel if and only if (71) can be written on the form (16). Then, for every factor $(1 + q_{t_1}^{2n+1} z^{\pm 1})$ in the denominator there must exist a corresponding factor in the numerator $(1 + q_{t_2}^{2m+1} z^{\pm 1})$, i.e., for every n there must exist an m such that

$$q_{t_1}^{2n+1} = q_{t_2}^{2m+1}. \quad (72)$$

Insertion of $q_{t_i} = e^{-(1/2t_i)}$ and reduction gives the necessary and sufficient requirement

$$2m = \frac{t_2}{t_1}(2n + 1) - 1. \quad (73)$$

It is clear that this relation cannot hold for all $n \in Z$ if t_1 and t_2 are chosen arbitrarily. The transformation from $\tilde{L}(x;\, t_1)$ to $\tilde{L}(x;\, t_2)$ $(t_2 > t_1)$ is a scale-space transformation if and only if the ratio t_2/t_1 is an odd integer.

Acknowledgment

I would like to thank Prof. J.-O. Eklundh for valuable advice as well as for reading and suggesting improvements to several versions of this manuscript. His guidance and support and his way of always finding time for discussions is very much appreciated. I also wish to thank Dr. S. Carlsson for stimulating discussions, which provided a large inspiration to this work, and Dr. T. Kolsrud for useful help with rudiments on semigroups.

References

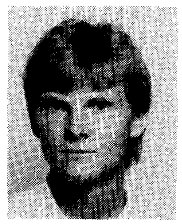
[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Applied Mathematics Series, 55), Nat. Bureau Standards, 1964.

[2] T. Ando, "Totally positive matrices," *Linear Algebra Applicat.*, vol. 90, pp. 165–219, 1987.

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

[4] F. Bergholm, "Edge focusing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 6, pp. 726–741, 1987.

[5] G. Dahlquist, Å. Björk, and N. Anderson, *Numerical Methods*. London: Prentice-Hall, 1974.
[6] R. M. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 6, pp. 725–730, 1972.
[7] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. Los Angeles, CA: University of California Press, 1958.
[8] W. E. L. Grimson and E. C. Hildreth, "Comments on digital step edges from zero crossings of second directional derivatives," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, no. 1, pp. 121–127, 1985.
[9] E. Hille, and R. S. Phillips, *Functional Analysis and Semi-Groups*, vol. XXXI. Providence, RI: Amer. Math. Soc. Colloquium Publ., 1957.
[10] I. I. Hirschmann and D. V. Widder, *The Convolution Transform*. Princeton, NJ: Princeton University Press, 1955.
[11] J. J. Koenderink and A. J. van Doorn, "The structure of images," *Biol. Cybern.*, vol. 50, pp. 363–370, 1984.
[12] —, "Dynamic shape," *Biol. Cybern.*, vol. 53, pp. 383-396, 1986.
[13] S. Karlin, *Total Positivity*, vol. I. Stanford, CA: Stanford University Press, 1968.
[14] L. M. Lifshitz and S. M. Pizer, "A multiresolution hierarchical approach to image segmentation based on intensity extrema," Dep. Comput. Sci. and Radiol., Univ. North Carolina, Chapel Hill, Internal Rep., 1987.
[15] T. P. Lindeberg, "On the construction of a scale-space for discrete images," Dep. Numer. Anal. Comput. Sci., Royal Inst. Technol., Stockholm, Sweden, Internal Rep. TRITA-NA-P8808, 1988.
[16] —, "Two-dimensional discrete scale-space formulation," Dep. Numer. Anal. Comput. Sci., Royal Inst. Technol. Stockholm, Sweden, Internal Rep., 1989.
[17] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc. London*, ser. B, vol. 207, pp. 187–217, 1980.
[18] E. Norman, "A discrete analogue of the Weierstrass transform," *Proc. Amer. Math. Soc.*, vol. 11, pp. 596–604, 1960.
[19] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*. London: Cambridge University Press, 1986.
[20] I. J. Schoenberg, "Über variationsvermindernde lineare Transformationen," *Mathematische Zeitschrift*, vol. 32, pp. 321–328, 1930.
[21] —, "Some analytical aspects of the problem of smoothing," in *Studies and Essays*. New York: Courant Anniversary Volume, 1948, pp. 351–370.
[22] —, "On smoothing operations and their generating functions," *Bull. Amer. Math. Soc.*, vol. 59, pp. 199–230, 1953.
[23] A. P. Witkin, "Scale-space filtering," in *Proc. 7th Int. Joint Conf. Artificial Intelligence*, 1983, pp. 1019–1022.



Tony Lindeberg was born in Stockholm, Sweden, in 1964. He received the M.Sc. degree in engineering physics and applied mathematics from the Royal Institute of Technology, Stockholm, Sweden, in 1987.

He is currently with the Computer Vision Group (CVAP) under Prof. Jan-Olof Eklundh, at the Royal Institute of Technology, where he is working on his Ph.D. thesis. His primary research interests in computer vision are scale-space methods and reconstruction of shape.