Linear Spatio-Temporal Scale-Space[*]

Tony Lindeberg

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

http://www.nada.kth.se/~tony
Email: tony@nada.kth.se

Abstract

This article shows how a linear scale-space formulation previously expressed for spatial domains extends to spatio-temporal data. Starting from the main assumptions that: (i) the scale-space should be generated by convolution with a semi-group of filter kernels and that (ii) local extrema must not be enhanced when the scale parameter increases, a complete taxonomy is given of the linear scale-space concepts that satisfy these conditions on spatial, temporal and spatio-temporal domains, including the cases with continuous as well as discrete data.

Key aspects captured by this theory include that: (i) time-causal scale-space kernels must not extend into the future, (ii) filter shapes can be tuned from specific context information, permitting mechanisms such local shifting, shape adaptation and velocity adaptation, all expressed in terms of local diffusion operations.

Receptive field profiles generated by the proposed theory show high qualitative similarities to receptive field profiles recorded from biological vision.

Contents

1 Introduction

When analysing sensory data, such as images, a fundamental constraint arises from the fact that real-world objects may appear in different ways depending upon the scale of observation. This insight is a major motivation for the development of multi-scale representations such as *pyramids* (Burt 1981, Crowley 1981) and *scale-space representation* (Iijima 1962, Witkin 1983, Koenderink 1984, Babaud et al. 1986, Yuille & Poggio 1986, Koenderink & van Doorn 1992, Lindeberg 1994, Pauwels et al. 1995, Florack 1997); see also (Alvarez et al. 1993, ter Haar Romeny 1994, Sporring et al. 1996, ter Haar Romeny et al. 1997, Weickert 1998, Nielsen et al. 1999, Kerckhove 2001).

A key reason why a multi-scale representation is essential is that the type of information that can be extracted from an image measurement is largely determined by the relationship between the size of the image operator (the probe size) and the size of the image structures. In the lack of any *a priori* knowledge about what image structures are relevant in a given situation, there is no way for the visual agent to know what probe size is appropriate for analysing a given data set. Hence, if one aims at a vision system with the ability to handle large variabilities in its input data, the only reasonable approach is to consider image measurements performed for a large range of probe sizes. During the last two decades, there has been a fruitful wave of research on this topic. A main stream of research has been concerned with the choice of image operators, and several results have been presented showing that within the class of linear operations, the Gaussian kernel and its derivatives arise as a canonical family of image operators (see the abovementioned references as well as the references therein). Complementary works have demonstrated that these filters can serve as a basis for expressing a large number of visual operations, including feature detection, stereo matching, computation of optic flow, estimation of shape cues and recognition.

Traditionally, however, most works on multi-scale representations have been concerned with image data defined on *spatial domains*, characterized by the fact that image data are accessible in all directions. (Notable exceptions are presented in (Koenderink 1988, Lindeberg & Fagerström 1996) and will be discussed later.) The world around us, on the other hand, gives rise to *spatio-temporal data*, in which time plays a very special role and the future cannot be accessed. Moreover, the structure of spatio-temporal data is usually special, in the sense that most real-world objects tend to persist over time in a coherent manner.

The subject of this article is to develop a scale-space theory for such spatio-temporal image domains, to enable a visual agent to register spatio-temporal events at multiple spatial and temporal scales in a well-founded manner. Technically, we shall start from a previously stated scale-space formulation in terms of *non-enhancement of local extrema*, (which has the attractive property that it applies both continuous and discrete data) and relax the symmetry requirements that have been used in previous applications of this scale-space axiom.

It will be shown that a parameterized family of linear scale-spaces is defined by this construction, including (i) the traditional linear scale-space representation for rotationally symmetric spatial domains, (ii) the affine Gaussian scale-space for spatial image domains subject to affine transformations, (iii) temporal linear scale-spaces for time dependent image data, and (iv) spatio-temporal linear scale-space for spatio-temporal image domains subject to motion.

Examples of properties that will be captured by the proposed framework include: (i) mechanisms for spatial shifting expressed in terms of diffusion equations and operating without need for external warping modules, (ii) time-causal scale-space filters not extending into the

future, and (iii) non-separable and elongated filter kernels over space and time, which allow for shape adaptation in space and velocity adaptation along the direction of motion. Common for all these scale-spaces is that they are generated from a parameterized family of continuous, discrete or semi-discrete diffusion processes, implying that they can be computed by local connections in a distributed computational architecture. Notably the receptive field profiles generated by this scale-space concept have high qualitative similarity to receptive fields profiles recorded from biological vision (DeAngelis et al. 1995, Valois et al. 2000) in analogy with previously established relations between spatial receptive fields and Gaussian derivative operators (Young 1985, Young 1987).

2 Choice of scale-space axioms

When constructing a multi-scale representation, one may ask: Is any choice of smoothing operation feasible? Of main importance when constructing a scale-space representation is that new (spurious) image structures are not introduced with increasing scale. This condition has been formalized in different ways by different authors, and a large number of different choices of scale-space axioms have been considered. Besides linearity and shift invariance, which form the basis when considering *linear* scale-spaces, some of the most commonly used requirements include causality, non-creation of local maxima or zero-crossings, scale invariance and semi-group structure complemented by more specific requirements, such as isotropy, homogeneity, symmetry, rotational invariance, separability, normalization, positivity, regularity, continuity and differentiability.[1]

Out of this large family of possibilities, we shall here make the following choice: To formalize the requirement that new structures are not created with increasing scale, we will use a reformulation of Koenderinks causality requirement, expressed as non-enhancement of local extrema (see also figure 1):

> *Non-enhancement of local extrema:* If for some scale level s_0 a point x_0 is a non-degenerate local maximum for the scale-space representation at that level (regarded as a function of the space coordinates only) then its value must not increase when the scale parameter increases. Analogously, if a point is a non-degenerate local minimum then its value must not decrease when the scale parameter increases.

Formulations closely related to this have been used when proving the uniqueness of Gaussian filtering for generating scale-space representations for continuous or discrete signals (Koenderink 1984, Babaud et al. 1986, Yuille & Poggio 1986, Lindeberg 1990, Lindeberg 1994, Lindeberg 1996) with extension to non-linear diffusion schemes by (Weickert 1998). While scale invariance may at first also be regarded as a very desirable and thus a primary property of a multi-scale representation, this criterion has the limitation that it can only be formulated for signals defined on a continuous domain. The notion of non-enhancement of local extrema has the conceptual advantage that it applies to both continuous and discrete data. Moreover, it gives a more rigorous statement of the requirement that new structures should not be created (or enhanced) when the scale parameter is increased. On the other hand, the formulation of the non-enhancement criterion requires the scale-space representation to

---

[1] Some of these listed requirements can be deduced from others, and we shall not aim at any extensive literature survey in this paper. The reader is referred to the sources mentioned in the introduction and the references therein.

satisfy certain regularity properties with respect to variations of the scale parameter. These requirements are, however, automatically fulfilled if we require the family of scale-space kernels to form a *semi-group*, a complementary requirement in scale-space formulations based on scale invariance.
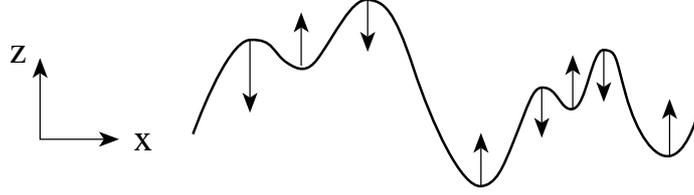
Figure 1: The non-enhancement condition of local extrema means that the grey-level value of a local maximum must not increase with scale and that the grey-level of a local minimum must not decrease.

2.1   Structural scale-space axioms

We would like to model an uncommitted visual front-end, that performs linear and shift-invariant operations. Hence, we assume that each scale level $L(\cdot;\ s)$ is generated by convolving the original image f with a convolution kernel $T(\cdot;\ s)$,

$$L(\cdot;\ s) = T(\cdot;\ s) * f(\cdot), \tag{1}$$

in operator form written

$$L(\cdot;\ s) = \mathcal{T}_s f(\cdot). \tag{2}$$

Then, to ensure regularity with respect to scale, we assume that the family of scale-space kernels $T(\cdot;\ s)$ forms a *semi-group*

$$T(\cdot;\ s_1) * T(\cdot;\ s_2) = T(\cdot;\ s_1 + s_2). \tag{3}$$

This condition means that all scale levels are computed from conceptually similar operations, and that, in addition, the transformation from any fine scale level to any coarser scale level is of the same form as the transformation from the original image

$$L(\cdot;\ s_2) = T(\cdot;\ s_2 - s_1) * L(\cdot;\ s_1). \tag{4}$$

Another important consequence of imposing the semi-group requirement on the family of convolution kernels is that if we assume reasonable continuity requirements of T with respect to variations of the scale parameter s, then it follows (from a general result in functional analysis (Hille & Phillips 1957)) that there exists a limit case operator, the *infinitesimal generator*

$$\mathcal{A}f = \lim_{h \downarrow 0} \frac{\mathcal{T}_h f - f}{h}, \tag{5}$$

such that the scale-space family satisfies the differential equation

$$\partial_s L(\cdot;\ s) = \lim_{h \downarrow 0} \frac{L(\cdot;\ t + h) - L(\cdot;\ s)}{h} = \mathcal{A}(\mathcal{T}_s f(\cdot)) = \mathcal{A} L(\cdot;\ s). \tag{6}$$

3

In previous works based on this form of scale-space formulation, additional (spatial) symmetry requirements have been imposed on the scale-space kernels (typically rotational symmetry), leading to the Gaussian kernel and its discrete counterpart. The purpose of this paper is to explore the richer structure that can be obtained if these symmetry requirements are relaxed.

3   Continuous signals

On a continuous domain, it is natural to formalize the non-enhancement requirement of local extrema in terms of a sign condition on the derivative of the scale-space family with respect to the scale parameter. Hence, at any non-degenerate extremum point (extremum where the determinant of the Hessian matrix is non-zero) we require the following conditions to hold:

$$\partial_s L \leq 0 \qquad \text{at a non-degenerate local maximum,} \tag{7}$$

$$\partial_s L \geq 0 \qquad \text{at a non-degenerate local minimum.} \tag{8}$$

In Appendix A it is shown that a linear and shift-invariant infinitesimal generator $\mathcal{A}$ satisfying this requirement of non-enhancement of local extrema corresponds to a linear combination of first- and second-order derivatives, where the second-order terms correspond to an elliptic differential operator and the coefficients for the first-order terms are arbitrary. Thereby, by necessity and sufficiency the scale-space family satisfies the differential equation

$$\partial_s L = \frac{1}{2} \nabla^T \left(\Sigma_0 \nabla L \right) - v_0^T \nabla L. \tag{9}$$

for some positive semi-definite covariance matrix Σ_0 and some translation vector v_0. If we take a discrete delta function as input, the interpretation of this evolution equation is that at any time moment the solution corresponds to a Gaussian kernel with covariance matrix $\Sigma_s = s\Sigma_0$ centered at $v_s = sv_0$. Thus, we can interpret the impulse response as a gradually growing elongated Gaussian kernel, which moves with velocity v_0 with respect to the evolution parameter s. In terms of filtering operations, this scale-space can equivalently be constructed by convolution with *velocity-adapted Gaussian kernels*

$$g(x;\ \Sigma_s, v_s) = \frac{1}{(2\pi)^{D/2}\sqrt{\det \Sigma_s}}\, e^{-(x-v_s)^T \Sigma_s^{-1}(x-v_s)/2}, \tag{10}$$

which for a given $\Sigma_s = s\Sigma_0$ and a given $v_s = sv_0$ satisfy the diffusion equation (9). The Fourier transform of this shifted Gaussian kernel is

$$\hat{g}(\omega;\ \Sigma_s, v_s) = \int_{x \in \mathbb{R}^N} g(x;\ \Sigma_s, v_s)\, e^{-i\omega^T x}\, dx = e^{i\omega^T x - \omega^T \Sigma_s \omega/2}. \tag{11}$$

Transformation property under linear transformations.   This scale-space concept has the attractive property that it is *closed* under affine transformations: If two image patterns f_L and f_R are related by an affine transformation

$$f_L(\xi) = f_R(\eta), \tag{12}$$

where

$$\eta = A\xi + b, \tag{13}$$

4

and if linear scale-space representations of these images are defined by

$$L(\cdot;\ \Sigma_L, v_L) = g(\cdot;\ \Sigma_L, v_L) * f_L(\cdot), \quad R(\cdot;\ \Sigma_R, v_R) = g(\cdot;\ \Sigma_R, v_R) * f_R(\cdot), \qquad (14)$$

then L and R are related by (Lindeberg 1994)

$$L(x;\ \Sigma_L, v_L) = R(y;\ \Sigma_R, v_R), \qquad (15)$$

where the covariance matrices Σ_L and Σ_R satisfy

$$\Sigma_R = A\Sigma_L A^T, \qquad (16)$$

and the velocity terms v_L and v_R in the Gaussian kernels can be traded against coordinate shifts in x and y as long as the following relation is satisfied:

$$y - v_R = A(x - v_L) + b. \qquad (17)$$

This property is highly useful in connection with visual tasks involving image deformations, such as stereo matching, flow estimation and shape estimation. The closedness under affine transformation allows for perfect modelling and matching of image data under first-order approximations of image deformations due to motion or the perspective mapping, and has been explored by e.g. (Lindeberg & Gårding 1997, Ballester & Gonzalez 1998, Nagel & Gehrke 1998, Schaffalitzky & Zisserman 2001).

In practice, there are two principally different ways of computing scale-space representations under affine alignment — either by deforming the filter shapes or by deforming the image data before the smoothing operation. This equivalence is made explicit in the commutative diagram in figure 2, and the two approaches may have their respective advantages when expressing algorithms and computational mechanisms.

3.1   Special cases

The abovementioned relations provide a general structure for linear scale-space concepts on shift-invariant continuous domains. Specifically, it includes the following special cases:

Rotationally symmetric linear scale-space.   If we require the scale-space representation generated by (9), or equivalently (10), to be rotationally symmetric, then it follows by necessity that v_s and v_0 must be zero and that the covariance matrices Σ_s and Σ_0 must be proportional to the unit matrix. Thus, the diffusion operator $\mathcal{A}$ will be proportional to the Laplacian operator, and the filter kernels will be rotationally symmetric Gaussians. With regard to image deformations, the closedness properties are then restricted to translations, rotations and rescalings. On the other hand, this scale-space concept is separable, corresponding to the convolution with one-dimensional Gaussian kernels along each dimension. This is the original scale-space concept as proposed by (Witkin 1983) and (Koenderink 1984).

Affine Gaussian scale-space.   If we relax the condition about rotational symmetry, while keeping a requirement that the corresponding Green's function should be mirror symmetric on every line through the origin *i.e.*, to avoid spatial shifts, the Fourier transform should be

$$L(x;\ \Sigma_L, v_L) \quad - \quad \left\{ \begin{array}{c} \eta = A\xi + b \\ \Sigma_R = A\Sigma_L A^T \\ y - v_R = A(x - v_L) + b \end{array} \right\} \quad \rightarrow \quad R(y;\ \Sigma_R, v_R)$$

$$\uparrow \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \uparrow$$

$$*g(\cdot;\ \Sigma_L, v_L) \qquad\qquad\qquad\qquad\qquad\qquad *g(\cdot;\ \Sigma_R, v_R)$$

$$| \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad |$$

$$f_L(\xi) \qquad - \qquad\qquad \eta = A\xi + b \qquad\qquad \rightarrow \qquad f_R(\eta)$$

Figure 2: Commutative diagram of the Gaussian scale-space representation under linear transformations of the spatial coordinates in the original image. The commutative property implies that the scale-space representations of two affinely deformed image patches can be affinely aligned, either by adapting the shapes of the Gaussian kernels or by deforming the image data prior to smoothing.
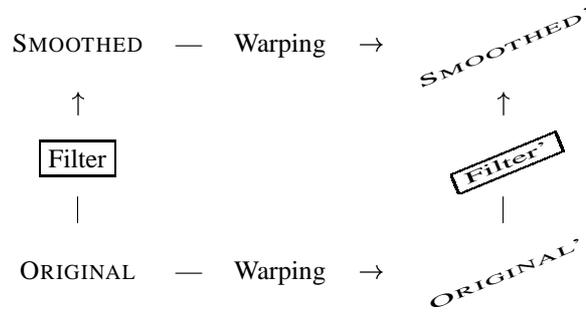
Figure 3: In practice, the commutative property implies that smoothed image data in scale-space representations can be matched by a similar warping operation as the original image data, provided that the shape of the smoothing kernel is transformed accordingly.
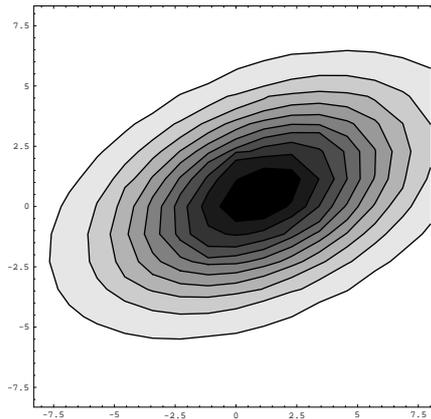
Figure 4: In terms of linear filters, the shift-invariant linear spatio-temporal scale-space concept for continuous signals corresponds to convolution with Gaussian kernels in space-time, with the center of the kernel determined by the offset vector v_s and the shape and orientation of the kernel determined by the covariance matrix Σ_0.

real, we obtain the *affine Gaussian scale-space representation*, generated by convolution with non-uniform Gaussian kernels

$$g(x; \ \Sigma_s) = \frac{1}{(2\pi)^{D/2}\sqrt{\det \Sigma_s}} e^{-x^T \Sigma_s^{-1} x/2}, \tag{18}$$

where Σ_s is a symmetric positive definite (covariance) matrix. This scale-space concept has the same closedness property under affine transformations as the scale-space generated by (10), besides the fact that the additional degree in freedom in the translational parameters in (17) has been eliminated by a choice of zero offset in the filter kernels. This scale-space concept has been studied by (Lindeberg 1994, Griffin 1996) and is highly useful when computing surface shape under local affine distortion and (Lindeberg & Gårding 1997) and performing affine invariant segmentation (Ballester & Gonzalez 1998) and matching (Schaffalitzky & Zisserman 2001). Combined with derivative operations, it can also serve as a natural idealized model for filter banks (Freeman & Adelson 1991, Simoncelli et al. 1992) consisting of elongated directional filters (Perona 1992).

**Temporal and spatio-temporal domains.** While the affine Gaussian scale-space generated by (18) has essentially the same invariance properties as the spatio-temporal scale-space representation generated by (10), one motivation for keeping the velocity term v in (10) arises when studying time dependent data. On a temporal domain, the non-zero offset in the Gaussian kernel can be used as a simplified model of the fact that all computations require non-zero computation time, and time averages can only be computed from data that have occurred in the past. This requirement of *time causality* implies that any temporal receptive field has to be associated with a non-zero time delay. Moreover, on a spatio-temporal domain, we may want the receptive fields to follow the direction of motion, in such a way that the centers and the shapes of the receptive fields are adapted to the motion estimates; see figure 5 for an illustration. Such velocity adaptation (Lindeberg 1997*a*) is useful for reducing the temporal blur induced by observing objects that move relative to the camera, and is a natural mechanism to include in modules for multi-scale motion estimation (Nagel & Gehrke 1998, Florack et al. 1998) and tracking at coarse temporal scales (Bretzner & Lindeberg 1997); compare
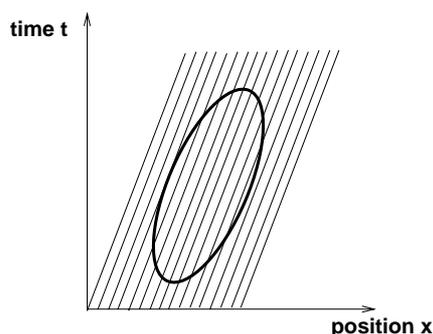

Figure 5: By adapting the shape and the position of a spatio-temporal smoothing kernel to the direction of motion, we can compute image descriptors that are invariant to e.g. constant velocity motion. This property can for example be used for reducing the effect of motion blur, when computing image descriptors of moving objects at coarse temporal scales.

7

also with the earlier work by (Florack et al. 1992) based on arguments from dimensional analysis.

With respect to temporal image processing, however, the filters in this Gaussian filter class do not respect time causality in a strict sense. Although the total mass of the filter coefficients that imply access to the future can be made arbitrarily small, by a suitable choice of time delay associated with the scale parameter in the scale direction, all filters in this filter class have support regions that cover the entire time axis, and are not suitable for real-time processing of temporal image data. To handle this problem (Koenderink 1988) proposed to transform the time axis by a logarithmic transformation that maps the present moment to infinity. Then, he applied Gaussian smoothing in the transformed domain, thus avoiding violations of temporal causality. Based on a classification of continuous and discrete scale-space kernels in (Lindeberg 1994), (Lindeberg & Fagerström 1996) presented another set of time-causal scale-space kernels based on cascaded recursive filters for discrete time and first-order integrators in the case of continuous time, with close relations to an earlier approach for estimating optical flow (Fleet & Langley 1995). Follow-up works based on Koenderinks logarithmic mapping of the time axis have then been presented by (Florack 1997, ter Haar Romeny et al. 2001) and (Lindeberg 2001).

We shall return to this subject in section 7. Before that, however, let us first develop a corresponding theory for discrete image data, which turns out to give rise to a richer structure with regard to semi-groups respecting temporal causality.

4 Discrete signals

If the general structure of a linear scale-space representation described in section 2.1 is combined with the requirement of non-enhancement of local extrema on a discrete domain, then it can be shown that the scale-space family $L: \mathbb{Z}^D \times \mathbb{R}^+ \to \mathbb{R}$ of the discrete signal $f: \mathbb{Z}^D \to \mathbb{R}$ must satisfy a semi-discrete differential equation of the form

$$(\partial_s L)(x; \ s) = (\mathcal{A}L)(x; \ s) = \sum_{\xi \in \mathbb{Z}^D} a_\xi L(x + \xi; \ s), \tag{19}$$

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

- the *locality* condition $a_\xi = 0$ if $|\xi|_\infty > 1$,

- the *positivity* constraint $a_\xi \geq 0$ if $\xi \neq 0$, and

- the *zero sum* condition $\sum_{\xi \in \mathbb{Z}^D} a_\xi = 0$.

A proof of this result is given in Appendix B. To investigate the solutions of this equation, a general approach that we shall follow in this paper consists of computing the generating function

$$\varphi(z; \ s) = \sum_{n \in \mathbb{Z}^D} c_n z^n \tag{20}$$

of the set of filter coefficients c_n in the filter for computing the scale-space representation L from the input signal f. By formally transforming (19) into generating functions

$$\partial_s \left(\varphi_C(z; \ s) \right) = \left(\sum_{\xi \in \mathbb{Z}^D} a_\xi z^\xi \right) \varphi_C(z; \ s) \tag{21}$$

8

where $a_\xi z^\xi$ should be interpreted as multi-index notation of $a_{\xi_1,\ldots,\xi_D} z_1^{\xi_1} \ldots z_D^{\xi_D}$, we obtain

$$\varphi(z;\ s) = e^{s \sum_{\xi \in \mathbb{Z}^D} a_\xi z^\xi} \tag{22}$$

In compact operator notion, the solution of (19) at scale s can equivalently be written

$$L = e^{s\,\mathcal{A}} f \tag{23}$$

and this expression can be regarded as a general parameterization of scale-space kernels on discrete image domains. Alternatively, we obtain the Fourier transform of the kernel by substituting $z = e^{-i\omega}$ into the generating function, which gives[2]

$$\psi(\omega;\ s) = \sum_{n \in \mathbb{Z}^D} c_n e^{in\omega} = e^{s \sum_{\xi \in \mathbb{Z}^D} a_\xi e^{i\xi\omega}}. \tag{24}$$

The subject of this section is to interpret this result on various types of image domains involving spatial, temporal and spatio-temporal image data, respectively, in various numbers of dimensions. The methodology we shall follow is to reparameterize the filter class in terms of the following basic difference operators, which will be used as a basis for expressing the degrees of freedom in the coefficients a_ξ:

$$(\delta_x f)(x) = \left(f(x+1) - f(x-1)\right)/2, \tag{25}$$

$$(\delta_- f)(x) = f(x) - f(x-1), \tag{26}$$

$$(\delta_{xx} f)(x) = f(x+1) - 2f(x) + f(x-1). \tag{27}$$

4.1   1–D discrete spatial domain

On a one-dimensional spatial domain, the semi-differential equation in (19) assumes the form

$$\partial_s L(x;\ s) = a_{-1} L(x-1;\ s) + a_0 L(x;\ s) + a_1 L(x+1;\ s) \tag{28}$$

where the positivity and zero sum conditions imply that the diffusion coefficients a_ξ must satisfy

$$a_{-1}, a_1 \geq 0, \quad \text{and} \quad a_1 + a_0 + a_{-1} = 0. \tag{29}$$

In terms of the difference operators δ_x and δ_{xx}, we can reparameterize (28) according to

$$\partial_s L = -C_x\, \delta_x L + \frac{C_{xx}}{2}\, \delta_{xx} L \tag{30}$$

where C_x and C_{xx} are related to a_{-1}, a_0 and a_{-1} by

$$\frac{C_x}{2} + \frac{C_{xx}}{2} = a_{-1}, \tag{31}$$

$$-C_{xx} = a_0, \tag{32}$$

$$-\frac{C_x}{2} + \frac{C_{xx}}{2} = a_1. \tag{33}$$

---

[2]The expression (24) provides a straightforward way of generating the filter coefficient of scale-space kernels, by computing the inverse FFT of the Fourier transform expressed in closed form. When computing multiple levels in a scale-space representations, however, it may be computationally more efficient to start from the diffusion equation (19) and/or to make use of recursive filters (Deriche 1987), in particular concerning scale-spaces on temporal and spatio-temporal domains (Lindeberg & Fagerström 1996, Lindeberg 2001).

Since $a_{-1}, a_1 \geq 0$ and $a_0 \leq 0$, it follows that

$$|C_x| \leq C_{xx} \qquad (34)$$

is a necessary and sufficient condition for the filter coefficients to satisfy the positivity and zero sum requirements. In terms of the difference operators δ_x and δ_{xx}, with generating functions $(z - z^{-1})/2$ and $z - 2 + z^{-1}$, respectively, we can express the generating function corresponding to the solution of (30)

$$\varphi(z) = \sum_{n \in \mathbb{Z}} c_n z^n = e^{s \sum_{\xi \in \mathbb{Z}} a_\xi z^\xi} = e^{s\left(-C_x\,(z-z^{-1})/2 + C_{xx}\,(z-2+z^{-1})/2\right)}, \qquad (35)$$
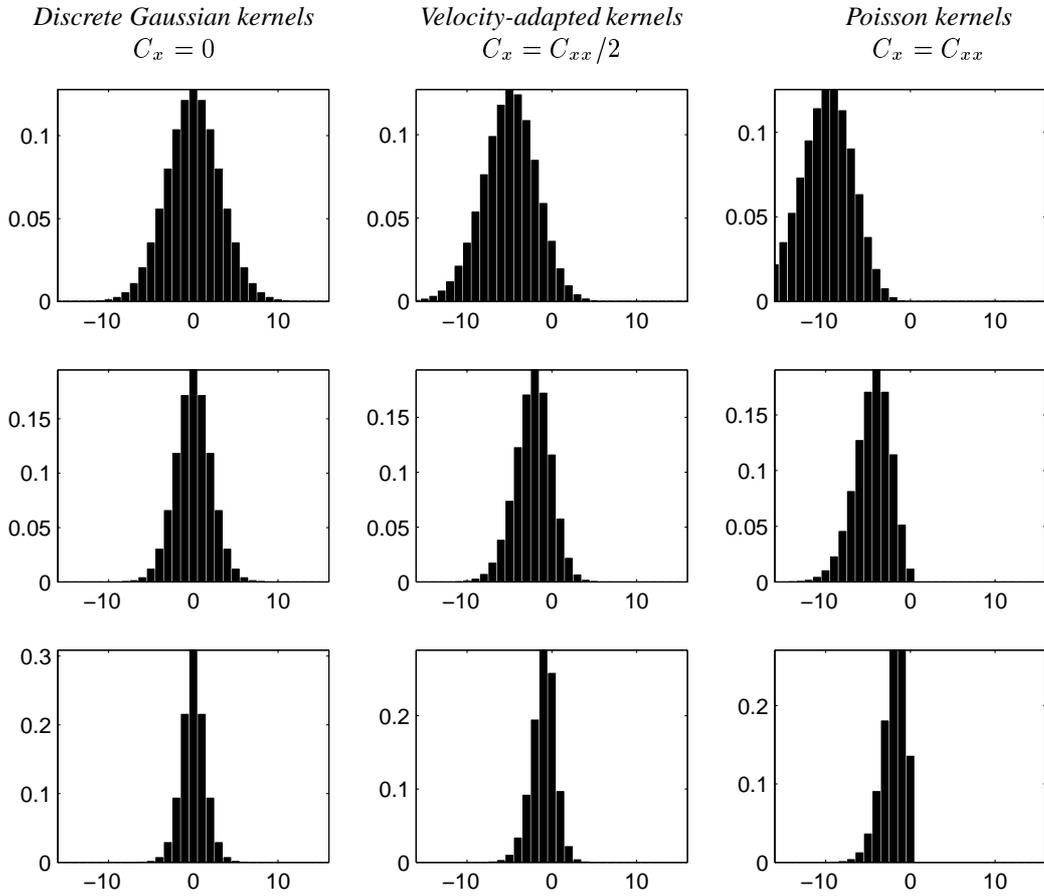
Figure 6: Graphs of kernels generated from the family of one-dimensional diffusion equations (30) for different values of the parameters C_x and C_{xx}. The graphs in the left column shows the discrete analogue of the Gaussian kernel (39) corresponding to $C_x = 0$, the graphs in the middle column shows velocity adapted discrete Gaussians for the (arbitrary) selection $C_x = C_{xx}/2$, while the graphs in the right column are Poisson kernels, which respect temporal causality and correspond to the extreme case $C_x = C_{xx}$. The variance ranges from $C_{xx} = 4$ in the bottom row to $C_{xx} = 20$ in the top row.

which by a change of variables $z = e^{-i\theta}$ gives rise to the corresponding Fourier transform

$$\psi(\theta) = \varphi(e^{-i\theta}) = \sum_{n \in \mathbb{Z}} c_n e^{-in\theta} = e^{s \sum_{\xi \in \mathbb{Z}} a_\xi e^{-i\xi\theta}} = e^{s\,(iC_x \sin\theta - C_{xx}\,(1-\cos\theta))} \qquad (36)$$

it can be seen that the mean value of the kernel (the center of gravity) is

$$M = \varphi_z(1) = -s\,C_x \qquad (37)$$

while its variance (the spatial extent) is

$$V = \varphi_{zz}(1) + \varphi_z(1) - \varphi_z^2(1) = s\,C_{xx}. \qquad (38)$$

(Both of these entities are expressed in units of the grid spacing).[3] The interpretation of this scale-space concept is thus that it allows signals to be translated a certain amount during the smoothing process, without any need for explicit warping. The coefficient C_x corresponds to the velocity term v_0 in the continuous diffusion equation (9), while the parameter C_{xx} combined with the evolution parameter s determines the amount of diffusion.

A notable difference compared to the continuous case, however, is that the positivity constraints on infinitesimal generator impose a bound on the amount on translation that can be performed given a certain amount of diffusion. Implicitly, this bound is expressed in units of the grid spacing. If we are interested in capturing rapid motions within this framework, and if we want to avoid the inclusion of additional explicit warping mechanisms, it is thus central to include a multi-resolution representation such that rapid motions can be first captured using a coarse grid spacing and using larger amounts of diffusion. Then, refinements or adjustments can be performed at finer scales using a denser grid, whenever more accurate localization properties are needed.

In the symmetric case, $C_x = 0$, it can be shown that this scale-space family reduces to the discrete scale-space generated by convolution with the *discrete analogue of the Gaussian kernel*

$$L(x;\ s) = \sum_{n=-\infty}^{\infty} T(n;\ s)\,f(x-n) \quad \text{where} \quad T(n;\ s) = e^{-s} I_n(s) \qquad (39)$$

and I_n are the modified Bessel functions of integer order (Abramowitz & Stegun 1964). (Here, we have without loss of generality set $C_{xx} = 1$.) As has been shown in (Lindeberg 1990), this family is also the unique semi-group structure on a symmetric discrete domain that guarantees non-creation of local extrema with increasing scales.

The left column in figure 6 shows a set of centered scale-space kernels (with $C_x = 0$) for different values of C_{xx} (ranging from $C_{xx} = 4$ to $C_{xx} = 20$). The middle column shows corresponding kernels with $C_x = C_{xx}/2$, while the right column shows corresponding results in the extreme case $C_x = C_{xx}$ (which is described in more detail in next section). Figures 7–8 show first- and second-order discrete derivative approximation kernels obtained by applying the first- and second-order difference operators δ_x and δ_{xx} to these scale-space kernels.

4.2 1–D discrete temporal domain

On a one-dimensional temporal domain, temporal causality implies that we must require that $a_{-1} = 0$. Relative to the analysis in previous section, this corresponds to the extreme case

---

[3]Note that that these expressions for the mean value and the standard deviation are only valid for filter kernels with non-negative values.

$C_x = C_{xx}$. The constraint that remains on the coefficients a_ξ is the zero sum constraint, $a_0 + a_1 = 0$. For notational convenience, let us parameterize this degree of freedom in terms of the δ_- operator and a free parameter $C_t > 0$,

$$\partial_s L = -C_t \, \delta_- L, \tag{40}$$

where we without loss of generality can set $C_t = 1$. It can be shown that this gives rise to a *Poisson-type temporal scale-space*, generated by convolution with Poisson kernels

$$L(x; \; s) = \sum_{n=-\infty}^{\infty} p(n; \; s) \, f(x - n) \quad \text{where} \quad p(n; \; s) = e^{-s} \frac{s^n}{n!} \quad (n \geq 0) \tag{41}$$

Figure 7: First-order discrete derivative approximation kernels generated by applying the first-order central difference operator δ_x to kernels generated from the family of one-dimensional diffusion equations (30) for different values of the parameters C_x and C_{xx}. The kernels in the left column originate from the discrete analogue of the Gaussian kernel (39) corresponding to $C_x = 0$, the kernels in the middle column are computed from velocity adapted discrete Gaussians for the (arbitrary) selection $C_x = C_{xx}/2$, while the kernels have been obtained from Poisson kernels, which respect temporal causality and correspond to the extreme case $C_x = C_{xx}$. The variance ranges from $C_{xx} = 4$ in the bottom row to $C_{xx} = 20$ in the top row.
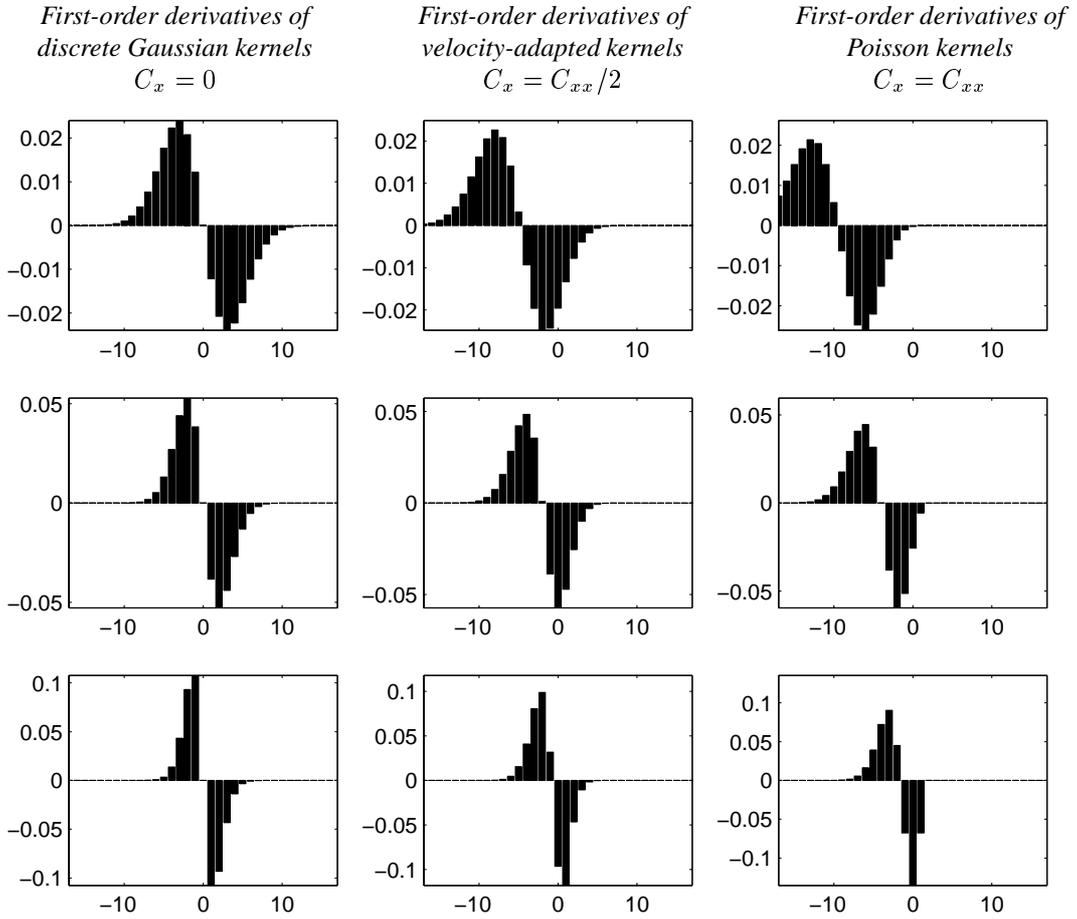
12

Figure 8: Second-order discrete derivative approximation kernels generated by applying the first-order central difference operator δ_x to kernels generated from the family of one-dimensional diffusion equations (30) for different values of the parameters C_x and C_{xx}. The kernels in the left column originate from the discrete analogue of the Gaussian kernel (39) corresponding to $C_x = 0$, the kernels in the middle column are computed from velocity adapted discrete Gaussians for the (arbitrary) selection $C_x = C_{xx}/2$, while the kernels have been obtained from Poisson kernels, which respect temporal causality and correspond to the extreme case $C_x = C_{xx}$. The variance ranges from $C_{xx} = 4$ in the bottom row to $C_{xx} = 20$ in the top row.

having mean $M = -s$ and variance $V = s$. Note that the evolution equation for this temporal scale-space (40) is expressed in terms of first-order differences only, while the corresponding evolution equation (30) on the spatial domain involves second-order differences.

The right column in figure 6 shows graphs of Poisson kernels for C_t ranging from $C_t = 4$ to $C_t = 20$, while the right columns in figures 7–8 show first-order derivatives of these kernels.

13

4.3 2–D discrete spatial domain

When we go from a one-dimensional to a two-dimensional image domain, additional degrees of freedom in the kernel design arise in terms of the eccentricity and the orientation of the filter. For simplicity, let us require the filter kernels in (19) to be symmetric[4] $a_{i,j} = a_{-i,-j}$. Then, the computational molecule of the infinitesimal operator $\mathcal{A}$ can be written as

$$\mathcal{A} = \begin{pmatrix} a_{-1,1} & a_{0,1} & a_{1,1} \\ a_{-1,0} & a_{0,0} & a_{1,0} \\ a_{-1,-1} & a_{0,-1} & a_{1,-1} \end{pmatrix} = \begin{pmatrix} D & C & B \\ A & -E & A \\ B & C & D \end{pmatrix} \tag{42}$$

for some $A, B, C, D \geq 0$ and $E = A + B + C + D$. In terms of the previously mentioned difference operators, and with $\delta_{xy} = \delta_x \, \delta_y$ and $\delta_{xxyy} = \delta_{xx} \, \delta_{yy}$, this computational molecule can with

$$A = \frac{C_{xx}}{2} - \frac{C_{xxyy}}{2}, \tag{43}$$

$$B = \frac{C_{xy}}{4} + \frac{C_{xxyy}}{4}, \tag{44}$$

$$C = \frac{C_{yy}}{2} - \frac{C_{xxyy}}{2}, \tag{45}$$

$$D = -\frac{C_{xy}}{4} + \frac{C_{xxyy}}{4}, \tag{46}$$

be reparameterized as

$$\mathcal{A} = \frac{1}{2} \begin{pmatrix} -C_{xy}/2 & C_{yy} & C_{xy}/2 \\ C_{xx} & -2(C_{xx} + C_{yy}) & C_{xx} \\ C_{xy}/2 & C_{yy} & -C_{xy}/2 \end{pmatrix} + \frac{C_{xxyy}}{4} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}. \tag{47}$$

Diffusion equation. With the parameterization in terms of C_{xx}, C_{xy}, C_{yy} and C_{xxyy}, the corresponding *discrete affine Gaussian scale-space* is given as the solution of the semi-discrete differential equation

$$\partial_s L = \frac{1}{2} \left(C_{xx} \, \delta_{xx} L + 2 C_{xy} \, \delta_{xy} L + C_{yy} \, \delta_{yy} L \right) + \frac{C_{xxyy}}{4} \, \delta_{xxyy} L, \tag{48}$$

which in turn can be interpreted as a second-order discretization of the diffusion equation associated with the continuous affine Gaussian scale-space

$$\partial_s L = \tfrac{1}{2}(C_{xx} \, L_{xx} + 2 C_{xy} \, L_{xy} + C_{yy} \, L_{yy}), \tag{49}$$

where $C_{xx}, C_{yy} > 0$ and $C_{xx} C_{yy} - C_{xy}^2 > 0$ are necessary conditions for the infinitesimal operator $\mathcal{A}$ to be positive definite.

Free parameter and positivity constraint. There is one free parameter $C_{xxyy} \geq 0$ in (48), which controls the addition of a discretization of the mixed fourth-order derivative L_{xxyy}. From (43)–(46) and $A, B, C, D \geq 0$ it follows that this parameter must satisfy

$$|C_{xy}| \leq C_{xxyy} \leq \min(C_{xx}, C_{yy}) \tag{50}$$

---

[4]In the non-symmetric case, we obtain additional degrees of freedom corresponding to translations in the spatial domain, in analogy with the velocity term v_s in (10).

to ensure that the positivity constraint is satisfied. In practice, the feasibility condition $|C_{xy}| \leq \min(C_{xx}, C_{yy})$ arising from this positivity constraint is always more restrictive than the condition $|C_{xy}| \leq \sqrt{C_{xx} C_{yy}}$ for the infinitesimal operator to be positive semi-definite. This implies that highly eccentric affine Gaussian kernels cannot be represented by non-negative discrete scale-space kernels on a square discrete grid, unless the orientation of the filter is approximately aligned to the coordinate directions.

To analyse for which positive definite covariance matrices the positivity condition may be violated, let us reparameterize the covariance matrix in terms of its eigenvalues $\lambda_{1,2} > 0$ as well as the orientation α of the eigenvectors eigenvectors

$$C_{xx} = \lambda_1 \cos^2 \alpha + \lambda_2 \sin^2 \alpha, \tag{51}$$

$$C_{xy} = (\lambda_2 - \lambda_1) \cos \alpha \, \sin \alpha, \tag{52}$$

$$C_{yy} = \lambda_1 \sin^2 \alpha + \lambda_2 \cos^2 \alpha. \tag{53}$$

where $\lambda_{1,2}$ and α are computed from the elements in C according to

$$\lambda_{1,2} = \frac{1}{2}(P \pm Q) \tag{54}$$

$$\alpha = \frac{1}{2} \arg(C, S) \tag{55}$$

with

$$P = C_{xx} + C_{yy} = \lambda_1 + \lambda_2 \tag{56}$$

$$C = C_{xx} - C_{yy} = (\lambda_1 - \lambda_2) \cos 2\alpha \tag{57}$$

$$S = 2C_{xy} = (\lambda_1 - \lambda_2) \sin 2\alpha \tag{58}$$

and $Q = \sqrt{C^2 + S^2}$. Then, the positivity requirement assumes the form

$$|\lambda_1 - \lambda_2| \, |\cos \alpha \, \sin \alpha| \leq \min(\lambda_1 \cos^2 \alpha + \lambda_2 \sin^2 \alpha, \lambda_1 \sin^2 \alpha + \lambda_2 \cos^2 \alpha), \tag{59}$$

which can be rewritten as

$$|\lambda_1 - \lambda_2| \, |\sin 2\alpha| \leq \min(\lambda_1 + \lambda_2 + (\lambda_1 - \lambda_2) \cos 2\alpha, \lambda_1 + \lambda_2 - (\lambda_1 - \lambda_2) \cos 2\alpha) \tag{60}$$

or equivalently

$$|\lambda_1 - \lambda_2| \, (|\cos 2\alpha| + |\sin 2\alpha|) \leq \lambda_1 + \lambda_2 \tag{61}$$

The "worst case orientation" is given by $\alpha = \frac{\pi}{8}$, and violations of the positivity requirement start to occur when λ_1 and λ_2 obey the relation

$$\sqrt{2} \, |\lambda_1 - \lambda_2| = \lambda_1 + \lambda_2, \tag{62}$$

corresponding to a ratio of the eigenvalues equal to $\epsilon = 3 + 2\sqrt{2} \approx 5.8$. A similar bound on the positivity requirement has been derived by (Weickert 1998) in the context of non-negative discretizations of non-linear diffusion equations.

Filter shapes. From the explicit expression for the generating function (22), here given by

$$\varphi(z) = \exp(C_{xx}\,(z - 2 + z^{-1})/2 + C_{yy}\,(w - 2 + w^{-1})/2+ \tag{63}$$
$$C_{xy}\,(z - z^{-1})(w - w^{-1})/4 + C_{xxyy}\,(z - 2 + z^{-1})\,(w - 2 + w^{-1})/4),$$

it follows that the filters in this family have mean vectors and covariance matrices given by

$$M = \begin{pmatrix} \varphi_z \\ \varphi_w \end{pmatrix} \Bigg|_{(z,w)=(1,1)} = -s \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{64}$$

$$V = \begin{pmatrix} \varphi_{zz} + \varphi_z - \varphi_z^2 & \varphi_{zw} - \varphi_z \varphi_w \\ \varphi_{zw} - \varphi_z \varphi_w & \varphi_{ww} + \varphi_w - \varphi_w^2 \end{pmatrix} \Bigg|_{(z,w)=(1,1)} = s \begin{pmatrix} C_{xx} & C_{xy} \\ C_{xy} & C_{yy} \end{pmatrix}. \tag{65}$$

provided that the positivity requirement is satisfied. This expression provides a formal justification for using the eigenvalue and eigenvector parameterization of Σ according to (51)–(53).

Behaviour for low frequencies. Let us transform the generating function (63) into the corresponding Fourier transform by $(z, w) = (e^{iu}, e^{iv})$, which gives

$$\psi(u, v) = \varphi(e^{iu}, e^{iv})$$
$$= \exp(-C_{xx}\,(1 - \cos u) - C_{yy}\,(1 - \sin u) + C_{xy} \sin u \sin v$$
$$+ C_{xxyy}\,(1 - \cos u)\,(1 - \sin u)) \tag{66}$$

By transforming this expression to polar coordinates $(u, v) = \omega(\cos \beta, \sin \beta)$, and performing a Taylor expansion for low angular frequencies ω:

$$\psi(\omega \cos \beta, \omega \sin \beta) = -\frac{1}{2}\left(C_{xx} \cos^2 \beta + 2C_{xy} \cos \beta \sin \beta + C_{yy} \sin^2 \beta\right) \omega^2$$
$$+ \frac{1}{24}\left(C_{xx} \cos^4 \beta + 4C_{xy} \cos^3 \beta \sin \beta + 6C_{xxyy} \cos^2 \beta \sin^2 \beta \right.$$
$$\left. + 4C_{xy} \cos \beta \sin^3 \beta + C_{yy} \sin^4 \beta\right) \omega^4 + \mathcal{O}(\omega^6), \tag{67}$$

which with $C_{xxyy} = (C_{xx} + C_{yy})/6 + \rho/3$ can be rewritten as

$$\psi(\omega \cos \beta, \omega \sin \beta) = -\frac{1}{2}\left(C_{xx} \cos^2 \beta + 2C_{xy} \cos \beta \sin \beta + C_{yy} \sin^2 \beta\right) \omega^2$$
$$+ \frac{1}{24}\left(C_{xx} \cos^2 \beta + 2C_{xy} \cos \beta \sin \beta + C_{yy} \sin^2 \beta\right) \omega^4$$
$$+ \frac{1}{12}\left(C_{xy} \cos \beta \sin \beta + \rho \cos^2 \beta \sin^2 \beta\right) \omega^4 + \mathcal{O}(\omega^6), \tag{68}$$

it can be seen that when $C_{xy} = 0$ the choice $\rho = 0$, implying that

$$C_{xxyy} = \frac{C_{xx} + C_{yy}}{6}, \tag{69}$$

gives a Fourier transform in which the fourth-order terms have a similar angular dependency, here an elliptic shape, as the second-order terms. Indeed, this elliptic shape coincides with the shape of the corresponding Fourier transform of the Gaussian kernel on a continuous spatial domain.

16

Rotationally symmetric scale-space. Specifically, for an isotropic scale-space, with $C_{xx} = C_{yy} = 1$ and $C_{xy} = 0$, insertion of (69) into (68) gives a Taylor expansion where both the second- and the fourth-order terms are independent of the angle β. In this respect, $C_{xxyy} = (C_{xx} + C_{yy})/3$ gives the discrete operator having the lowest degree of rotational asymmetry for low frequencies. In terms of the following two discrete approximations of the two-dimensional Laplacian operator:

$$(\nabla_5^2 f)_{0,0} = f_{-1,0} + f_{+1,0} + f_{0,-1} + f_{0,+1} - 4f_{0,0} \tag{70}$$

$$(\nabla_{\times^2}^2 f)_{0,0} = 1/2(f_{-1,-1} + f_{-1,+1} + f_{+1,-1} + f_{+1,+1} - 4f_{0,0}), \tag{71}$$

and the following parameterization of the isotropic diffusion equation

$$\partial_s L = \frac{1}{2}\left((1-\gamma)\,\nabla_5^2 L + \gamma\,\nabla_{\times^2}^2 L\right), \tag{72}$$

this choice corresponds to $\gamma = \frac{1}{3}$ with the computational molecule

$$\frac{2}{3}\,\nabla_5^2 + \frac{1}{3}\,\nabla_{\times^2}^2 = \frac{1}{6}\begin{pmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{pmatrix} \tag{73}$$

On the other hand, $C_{xxyy} = 0$, corresponding to $\gamma = 0$, gives a separable scale-space, corresponding to the application of the one-dimensional scale-space concept (39) along each dimension.

Choice of C_{xxyy}. In view of the general constraint from the positivity requirement (50) as well as the specific result (69) when $C_{xy} = 0$, we have in our implementation chosen C_{xxyy} as function of C_{xx}, C_{xy} and C_{yy} by blending these requirements as follows:

$$C_{xxyy} = \frac{C_{xx} + C_{yy}}{6}\left(1 - \frac{|C_{xy}|}{\min(C_{xx}, C_{yy})}\right) + |C_{xy}|. \tag{74}$$

When $C_{xy} = 0$, this choice gives C_{xxyy} according to (69), while in the extreme case $|C_{xy}| = \min(C_{xx}, C_{yy})$, it gives the only permissible choice $C_{xxyy} = |C_{xy}|$.

Another alternative could be to derive the value of C_{xxyy} that minimizes the deviation between the shapes of the level curves in the fourth-order term from the desired elliptic shape of the level curves in the second-order term.

Kernel graphs. Figure 9 shows a few examples of filter kernels generated in this way, with C_{xx}, C_{xy} and C_{yy} specified from the eigenvalues $\lambda_{1,2}$ as well as the orientation of the covariance matrix, according to (51)–(53).

Directional derivative approximations. Figure 10 shows first- and second-order directional derivatives of such kernels, computed from the general expression for the n-th order directional derivative in the direction θ:

$$\partial_\theta^n L = (\cos\theta\,\partial_x + \sin\theta\,\partial_y)^n L, \tag{75}$$

for orders up to two, explicitly written out as

$$\partial_{\bar{\theta}} L = \cos\theta\, L_x + \sin\theta\, L_y, \tag{76}$$

$$\partial_\theta^2 L = \cos^2\theta\, L_{xx} + 2\cos\theta\sin\theta\, L_{xy} + \sin^2\theta\, L_{yy}. \tag{77}$$

17
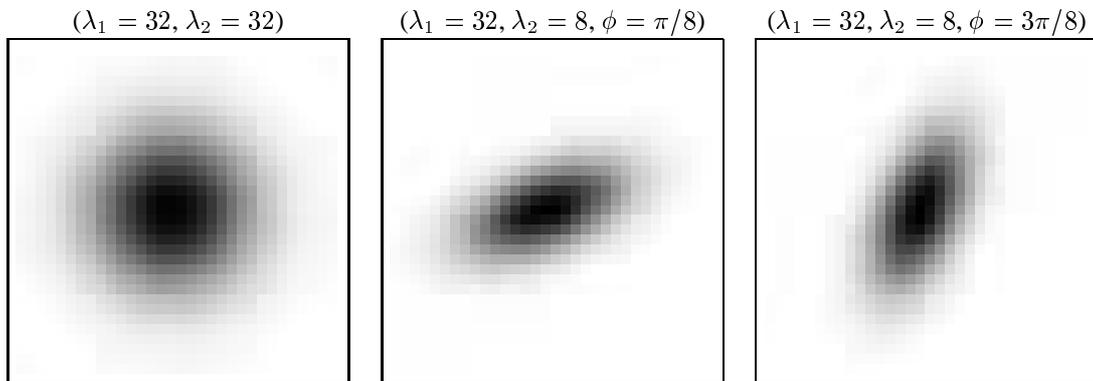
Figure 9: Discrete affine Gaussian kernels with varying eccentricity and orientation. (Image size: 32×32 pixels.)
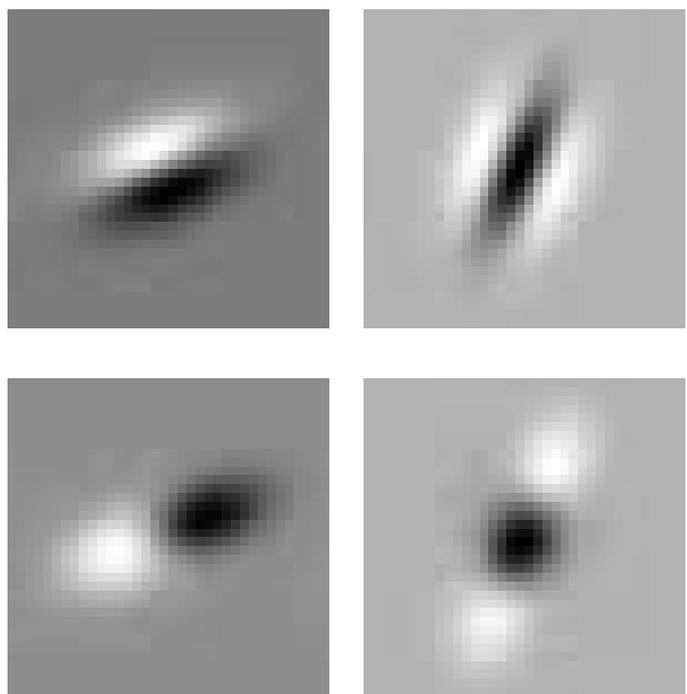

Figure 10: Discrete derivative approximations obtained by applying first- and second-order directional derivative operators to elongated discrete affine Gaussian kernels with eigenvalues $(\lambda_1, \lambda_2) = (32, 8)$ and orientation $\phi = \pi/8$ and $3\pi/8$, respectively. (Image size: 32×32 pixels.)

4.4   1+1–D discrete spatio-temporal domain

To obtain an intuitive understanding of how a discrete scale-space concept applies to spatio-temporal data, let us first study the case with one spatial dimension (represented by x) and one temporal dimension (represented by t). The general conditions in (19), combined with temporal causality, then give that the computational molecule of $\mathcal{A}$ has to be of the form

$$\mathcal{A} = \begin{pmatrix} 0 & 0 & 0 \\ A & -F & E \\ B & C & D \end{pmatrix} \tag{78}$$

for some constants $A, B, C, D, E > 0$ and $F = A + B + C + D + E$. (The zero entries in this operator indicate forbidden access to the future.)

Diffusion equation.   Let δ_t denote the backward difference operator δ_- in the t-direction, $\delta_{xt} = \delta_x\,\delta_t$ and $\delta_{xxt} = \delta_{xx}\,\delta_t$. Then, with

$$A = +\frac{C_x}{2} + \frac{C_{xx}}{2} - \frac{C_{xt}}{2} - \frac{C_{xxt}}{2}, \tag{79}$$

$$B = +\frac{C_{xt}}{2} + \frac{C_{xxt}}{2}, \tag{80}$$

$$C = C_t - C_{xxt}, \tag{81}$$

$$D = -\frac{C_{xt}}{2} + \frac{C_{xxt}}{2}, \tag{82}$$

$$E = -\frac{C_x}{2} + \frac{C_{xx}}{2} + \frac{C_{xt}}{2} - \frac{C_{xxt}}{2}, \tag{83}$$

this family of filter kernels can be parameterized as

$$\partial_s L = -C_x\,\delta_x L - C_t\,\delta_t L + \frac{1}{2}\left(C_{xx}\,\delta_{xx}L + 2C_{xt}\,\delta_{xt}L\right) - \frac{3}{6}\,C_{xxt}\,\delta_{xxt}L, \tag{84}$$

where $C_{xx}, C_t > 0$ and non-negativity of non-central coefficients implies that the following conditions must hold (see (81) and compare (79) with (83) and (80) with (82))

$$|C_{xt}| \le C_{xxt} \le \min(C_{xx}, C_t), \tag{85}$$

$$|C_x - C_{xt}| \le C_{xx} - C_{xxt}, \tag{86}$$

It can be shown that the conditions (85) and (86) together constitute both necessary and sufficient conditions for the positivity requirement to be satisfied. Moreover, these conditions can be summarized into a single necessary and sufficient condition as follows

$$|C_{xt}| \le C_{xxt} \le \min(C_t, C_{xx} - |C_x - C_{xt}|). \tag{87}$$

With reference to the positivity condition (34) on the filter coefficients on a one-dimensional discrete spatial domain, it can be shown that (87) implies that

$$|C_x| \le C_{xx}. \tag{88}$$

Hence, the positivity requirement on a 1+1-D discrete spatio-temporal domain is more restrictive than for a 1-D discrete spatial domain. Moreover, the requirement on C_{xxt} in (87) is more restrictive than the corresponding requirement (50) on the coefficient C_{xxyy} in the discrete scale-space concept for a two-dimensional discrete domain.

Filter shapes. From the generating function of the corresponding filter

$$
\begin{aligned}
\varphi(w, z) = \exp(&- C_x \, (w - w^{-1})/2 - C_t \, (1 - z^{-1}) \\
&+ C_{xx} \, (w - 2 + w^{-1})/2 + C_{xt} \, (w - w^{-1}) \, (1 - z^{-1})/2 \\
&- C_{xxt} \, (w - 2 + w^{-1}) \, (1 - z^{-1})/2),
\end{aligned}
\tag{89}
$$

we then get the mean vector M and the covariance matrix V as

$$
M = \left(\begin{array}{c} \varphi_w \\ \varphi_z \end{array} \right) \Bigg|_{(w,z)=(1,1)} = -s \left(\begin{array}{c} C_x \\ C_t \end{array} \right),
\tag{90}
$$

$$
V = \left(\begin{array}{cc} \varphi_{ww} + \varphi_w - \varphi_w^2 & \varphi_{wz} - \varphi_w \varphi_z \\ \varphi_{wz} - \varphi_w \varphi_z & \varphi_{zz} + \varphi_z - \varphi_z^2 \end{array} \right) \Bigg|_{(w,z)=(1,1)} = s \left(\begin{array}{cc} C_{xx} & C_{xt} \\ C_{xt} & C_t \end{array} \right).
\tag{91}
$$

provided that the positivity requirements are satisfied.

Note that the temporal component of the mean vector is coupled to the temporal component of the covariance matrix. Thus, by varying C_x, C_t, C_{xx} and C_{xt} within the constraints (85), (86) and (88), we can reach a four-dimensional subset of the five-dimensional manifold spanned by all variations of mean vectors and covariance matrices in two dimensions.

Parameterization of filter shapes based on eigenvalues To give a more explicit parameterization of the filters in this filter class, let us first in analogy with section 4.3 introduce descriptors P, C, S and Q according to

$$
P = C_{xx} + C_t, \qquad C = C_{xx} - C_t, \qquad S = 2 \, C_{xt} \qquad Q = \sqrt{C^2 + S^2}.
\tag{92}
$$

Then, the eigenvalues of V are given by

$$
\lambda_{1,2} = \frac{1}{2}(P \pm Q)
\tag{93}
$$

and the eigenvectors of V can (for example) be written

$$
e_1 = \left(\begin{array}{c} Q + C \\ S \end{array} \right), \qquad e_2 = \left(\begin{array}{c} -S \\ Q + C \end{array} \right)
\tag{94}
$$

where the orientations have been chosen such that the first eigenvector is aligned with the x-direction when the coefficient C_{xt} is zero.

Alternatively, we can parameterize the orientation of V in terms of an eigenvector normalized to unit length in the t-direction

$$
e_v = \Delta = \left(\begin{array}{c} \dot{x} \\ 1 \end{array} \right) = \left(\begin{array}{c} \tan \phi \\ 1 \end{array} \right).
\tag{95}
$$

This choice corresponds to

$$
\cos \phi = \frac{1}{\sqrt{1 + \dot{x}^2}}, \qquad \sin \phi = \frac{\dot{x}}{\sqrt{1 + \dot{x}^2}},
\tag{96}
$$

and insertion into the parameterization (51)–(53) of the covariance matrix gives

$$
\left(\begin{array}{cc} C_{xx} & C_{xt} \\ C_{xt} & C_t \end{array} \right) = \frac{1}{1 + \dot{x}^2} \left(\begin{array}{cc} \lambda_{xx} + \dot{x}^2 \, \lambda_t & -\dot{x} \, (\lambda_{xx} - \lambda_t) \\ -\dot{x} \, (\lambda_{xx} - \lambda_t) & \lambda_t + \dot{x}^2 \, \lambda_{xx} \end{array} \right).
\tag{97}
$$

where λ_{xx} and λ_t are the eigenvalues of V, while the orientation of V is given by the angle $\phi = \arctan \dot{x}$. Specifically, when $\dot{x} = 0$ we have $C_{xx} = \lambda_{xx}$, $C_{xt} = 0$ and $C_t = \lambda_t$.

We obtain the explicit relationship between these two parameterizations of V by inserting C_{xx}, C_{xt} and C_t from (97) into (92), which gives

$$P = \lambda_{xx} + \lambda_t, \tag{98}$$

$$C = \frac{1 - \dot{x}^2}{1 + \dot{x}^2} (\lambda_{xx} - \lambda_t), \tag{99}$$

$$S = \frac{-2\dot{x}}{1 + \dot{x}^2} (\lambda_{xx} - \lambda_t), \tag{100}$$

$$Q = |\lambda_{xx} - \lambda_t|. \tag{101}$$

Specifically, by comparing e_v in (95) with e_2 in (94) we get

$$\dot{x} = \tan \phi = \frac{-S}{Q + C} \tag{102}$$

and from a comparison of (99) with (100) it follows that

$$(\dot{x}^2 - 1) S + 2\dot{x} C = 0. \tag{103}$$

Parameterization of filter shapes based on Galilei transformations   With regard to motion analysis, another natural way of parameterizing the smoothing kernel is in terms of the transformation property (16) of covariance matrices under linear transformations. For an object moving with constant velocity v

$$\left\{ \begin{array}{l} x' = x + vt \\ t' = t \end{array} \right. \quad \text{i.e.} \quad \left(\begin{array}{c} x' \\ t' \end{array} \right) = \left(\begin{array}{cc} 1 & v \\ 0 & 1 \end{array} \right) \left(\begin{array}{c} x \\ t \end{array} \right) \tag{104}$$

it follows that the covariance matrices will be related according to

$$\left(\begin{array}{cc} C'_{xx} & C'_{xt} \\ C'_{xt} & C'_t \end{array} \right) = \left(\begin{array}{cc} 1 & v \\ 0 & 1 \end{array} \right) \left(\begin{array}{cc} C_{xx} & C_{xt} \\ C_{xt} & C_t \end{array} \right) \left(\begin{array}{cc} 1 & 0 \\ v & 1 \end{array} \right) \tag{105}$$

and for the mean vectors

$$\left(\begin{array}{c} C'_x \\ C'_t \end{array} \right) = \left(\begin{array}{cc} 1 & v \\ 0 & 1 \end{array} \right) \left(\begin{array}{c} C_x \\ C_t \end{array} \right) = \left(\begin{array}{c} C_x + v\,C_t \\ C_t \end{array} \right) \tag{106}$$

If we want to express a smoothing operation on a moving object that is equivalent to separable smoothing in a corresponding static coordinate frame, $C_{xx} = \lambda_{xx}$, $C_{xt} = 0$, $C_t = \lambda_t$ and $C_x = 0$, this corresponds to the relations

$$\left(\begin{array}{cc} C'_{xx} & C'_{xt} \\ C'_{xt} & C'_t \end{array} \right) = \left(\begin{array}{cc} \lambda_{xx} + v^2\,\lambda_t & v\lambda_t \\ v\lambda_t & \lambda_t \end{array} \right) \tag{107}$$

and

$$\left(\begin{array}{c} C'_x \\ C'_t \end{array} \right) = \left(\begin{array}{c} C_x + v\,C_t \\ C_t \end{array} \right) \tag{108}$$

which differ from the relations in (97). Note, however, that the orientation of the eigenvector of this covariance matrix will not coincide with the direction of motion. The reason for this is similar as the reason why the principal axes of an ellipse are not preserved under a skew transformation (see figure 11 for an illustration).
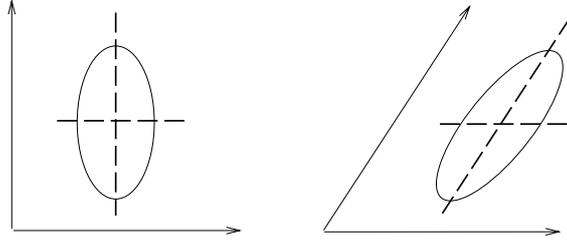
Figure 11: (a) A drawing of an ellipse with its principal axes indicated. (b) The result of subjecting the drawing in (a) to a skew transformation. Note that the principal axes of the ellipse are not preserved under the skewing transformation.

Velocity adaptation. When studying image structures that move coherently over time, it is natural to adapt the shape of the spatio-temporal filter kernel to the direction of motion. There are several motivations for using such an approach: One practically important reasons are to avoid excessive motion blur when observing moving objects at coarse time scales. A conceptually important reason is to make it possible to define image descriptors that are invariant under velocity changes. A third reason is that the result of a velocity adaptation procedure, once it has converged, can serve as an indicator of the direction of motion. Furthermore, compared to motion estimates computed using an arbitrary selection of filter shapes, we can expect motion estimates to be more accurate if the motion estimates are computed from spatio-temporal filters with their shapes adapted to the direction of motion, since then the trajectories of objects in space-time will be enhanced. Here, we shall consider the following forms of velocity adaptation:

Partial velocity adaptation: Partial velocity adaptation means that the shape of the covariance matrix is adapted to the direction of motion, while the mean vector is not. There are two special cases:

- *Eigenvector based partial velocity adaptation* means that the main eigenvector of the covariance matrix (91) is adapted to the direction of motion, such that the relations (95) and (103) are satisfied. A main motivation for adapting the shapes of smoothing kernels in this way, is that we can expect that the traces of objects that move coherently in space-time will be enhanced. Given a velocity estimate $\dot{x}$, eigenvector based partial velocity adaptation thus means that the parameters C_{xx}, C_{xt} and C_t should be coupled according to:

$$(\dot{x}^2 - 1) C_{xt} + \dot{x} (C_{xx} - C_t) = 0. \tag{109}$$

- *Galilean transformation based partial velocity adaptation* means that the shape of the covariance matrix is determined as (107) given a velocity estimate v and a set of smoothing parameters λ_{xx} and λ_t in an imagined moving frame attached to the motion. A main motivation for this form of velocity adaptation is that the filter parameters will then transform in a similar way as the motion of an object, and we can then derive image descriptors that are invariant under motion.

Formally, the mean vector is not specified by partial velocity adaptation. There are, however, two natural choices. If we let $C_x = C_{xt}$, then the positivity constraint (86)

22

becomes minimally restrictive (see also below), while if we let $C_x = 0$ then we obtain filters without spatial offset.

Complete velocity adaptation: Complete velocity adaptation means that in addition to a requirement on the shape of the covariance matrix, we also constrain it mean value. The following special cases will be considered:

- *Eigenvector based complete velocity adaptation* means that in addition to the requirement (109) of eigenvector based partial velocity adaptation, we also adapt the mean values of the filters to the direction of motion such that

$$\begin{pmatrix} C_x \\ C_t \end{pmatrix} = C_t \begin{pmatrix} \dot{x} \\ 1 \end{pmatrix}. \tag{110}$$

- *Galilean transformation based complete velocity adaptation* means that in addition to the requirement (107) of Galilean transformation based partial velocity adaptation, we also restrict the mean vector of the smoothing kernel according to (108).

In the following, we will analyse for which combinations of velocity estimates and smoothing parameters partial velocity adaptation and complete velocity adaptation, respectively, are possible, given the positivity requirements of the non-central coefficients in the infinitesimal generator.

Free parameter and positivity constraint. In the semi-discrete diffusion equation (84) that generates this spatio-temporal scale-space, there is one free parameter $C_{xxt} \geq 0$, which controls the addition of the mixed third-order derivative L_{xxt}. In equation (87), we expressed a necessary and sufficient condition for how the positivity of non-central filter coefficients restricts the possible values of this parameter, given a selection of C_{xx}, C_t, C_x and C_{xt}:

$$|C_{xt}| \leq C_{xxt} \leq \min(C_t, C_{xx} - |C_x - C_{xt}|). \tag{111}$$

From the form of this relation, it is evident that this positivity condition is always more restrictive than the condition $|C_{xt}| \leq \sqrt{C_{xx}C_t}$ for the covariance matrix to be positive semi-definite. This implies that spatio-temporal scale-space filters with highly eccentric filter shapes cannot be represented by non-negative discrete scale-space kernels on a square grid, unless the orientation of the filter is approximately aligned to the coordinate directions or the sampling densities in space and scale are adjusted.

When is eigenvector based partial velocity adaptation possible? To make explicit how the positivity constraints restrict the permissible values of λ_{xx} and λ_t in the case of eigenvector based partial velocity adaptation, let us insert the parameterization (97) of the covariance matrix into the positivity requirement (111). Then, for partial velocity adaptation with $C_x = 0$, the positivity requirement assumes the form

$$|\dot{x}| |\lambda_{xx} - \lambda_t| \leq \min(\lambda_t + \dot{x}^2 \lambda_{xx}, \lambda_{xx} + \dot{x}^2 \lambda_t - |\dot{x}| |\lambda_{xx} - \lambda_t|), \tag{112}$$

which after a few calculations can be reduced to the following requirements:

- If $\lambda_{xx} \geq \lambda_t$:

$$\lambda_{xx} \leq \frac{(1 + |\dot{x}|)}{|\dot{x}|(1 - |\dot{x}|)} \lambda_t \quad \text{for small } |\dot{x}| < 1, \tag{113}$$

$$\lambda_{xx} \leq \frac{|\dot{x}| (2 + |\dot{x}|)}{(2|\dot{x}| - 1)} \lambda_t \quad \text{for large } |\dot{x}| > \tfrac{1}{2}, \tag{114}$$

- If $\lambda_{xx} \leq \lambda_t$:

$$\lambda_t \leq \frac{|\dot{x}| (1 + |\dot{x}|)}{(|\dot{x}| - 1)} \lambda_{xx} \quad \text{for large } |\dot{x}| > 1, \tag{115}$$

$$\lambda_t \leq \frac{(1 + 2|\dot{x}|)}{|\dot{x}| (2 - |\dot{x}|)} \lambda_{xx} \quad \text{for small } |\dot{x}| < 2, \tag{116}$$

Here, the condition (113) is maximally restrictive when $|\dot{x}| = \sqrt{2} - 1 \approx 0.414$, while the condition (114) is maximally restrictive when $|\dot{x}| = (\sqrt{5} + 1)/2 \approx 1.618$. With respect to the relation between λ_{xx} and λ_t, these conditions correspond to the ratios $\lambda_{xx}/\lambda_t = 3 + 2\sqrt{2} \approx 5.828$ and $\lambda_{xx}/\lambda_t = (3 + \sqrt{5})/2 \approx 2.618$, respectively. In a similar way, the condition (115) is maximally restrictive when $|\dot{x}| = 1 + \sqrt{2} \approx 2.414$, while the condition (116) is maximally restrictive when $|\dot{x}| = (\sqrt{5} + 1)/2 \approx 1.618$, and these values correspond to the ratios $\lambda_t/\lambda_{xx} = 3 + 2\sqrt{2} \approx 5.828$ and $\lambda_t/\lambda_{xx} = (3 + \sqrt{5})/2 \approx 2.618$, respectively.

If we disregard the specific velocity dependency in these relations, we can thus conclude that partial velocity adaptation with $C_x = 0$ can always be achieved if the ratio between the eigenvalues λ_{xx} and λ_t is below $\kappa = (3 + \sqrt{5})/2 \approx 2.618$. If we in addition (for reasons that will be apparent later) restrict the image velocities to always satisfy $|\dot{x}| \leq 1$, then this range increases to ratios up to $\kappa = 3 + 2\sqrt{2} \approx 5.828$, that is the same bound as for the discrete affine scale-space concept considered in section 4.3). This bound is also obtained for partial velocity adaptation with $C_x = C_{xt}$.

When is eigenvector based complete velocity adaptation possible? For complete velocity adaptation, the positivity requirement can be written

$$|\dot{x}| |\lambda_{xx} - \lambda_t| \leq \min(\lambda_t + \dot{x}^2 \lambda_{xx}, (1 - |\dot{x}| - |\dot{x}|^3) \lambda_{xx} + \dot{x}^2 \lambda_t) \tag{117}$$

and reduces to the following requirements

- If $\lambda_{xx} \geq \lambda_t$: A similar condition as in (113) with the additional condition

$$\lambda_{xx} \leq \frac{(1 + |\dot{x}|)}{(|\dot{x}|^3 + 2|\dot{x}| - 1)} \lambda_t \quad \text{for large } |\dot{x}| < |\dot{x}|_{crit} \tag{118}$$

where $|\dot{x}|_{crit} \approx 0.453$ is the real solution of the equation $|\dot{x}|^3 + 2|\dot{x}| - 1 = 0$.

- If $\lambda_{xx} \leq \lambda_t$: A similar condition as in (115) with the additional condition

$$\lambda_t \leq \left(\frac{1}{|\dot{x}|} + 1 + |\dot{x}| \right) \lambda_{xx} \quad \text{for small } |\dot{x}| < 1 \tag{119}$$

Here, the first condition (118) does not impose any constraints on $\lambda_{xx} \geq \lambda_t$ as long as the velocity is below the threshold $|\dot{x}| < |\dot{x}|_{crit}$. Thus, the bound $\lambda_{xx}/\lambda_t \leq 3 + 2\sqrt{2}$ according to (113) is the dominant factor. In the second case when $\lambda_{xx} \leq \lambda_t$, the constraint (119) dominates and is more restrictive than (115). If we disregard the explicit velocity dependency, it leads to the bound $\lambda_t \leq 3\lambda_{xx}$, where $|\dot{x}| < 1$ is a necessary prerequisite.

When is Galilean transformation based partial velocity adaptation possible? For Galilean transformation based partial velocity adaptation, we can in the necessary and sufficient condition (87) on the filter coefficients, first observe that the first condition $|C_{xt}| \leq C_t$ reduces to $|v|\lambda_t \leq \lambda_t$, thus corresponding to the necessary requirement $|v| \leq 1$ relative to the grid spacing. Concerning the second requirement $|C_{xt}| \leq C_{xx} - |C_x - C_{xt}|$, let us study the two cases $C_x = 0$ and $C_x = C_{xt}$:

- With $C_x = C_{xt}$, the requirement reduces to $|C_{xt}| \leq C_{xx}$ corresponding to $|v|\lambda_t \leq \lambda_{xx} + v^2\lambda_t$, with the necessary requirement, $\lambda_{xx} \geq (|v| - v^2)\lambda_t$. Thus, given a fixed value of the temporal scale λ_t, we obtain a minimum permissible value on the spatial scale λ_{xx}.

- With $C_x = 0$, the requirement reduces to $|C_{xt}| \leq C_{xx}/2$, and we obtain a corresponding more restrictive bound $\lambda_{xx} \geq (2|v| - v^2)\lambda_t$ with $|v| \leq 1$.

When is Galilean transformation based complete velocity adaptation possible? In the case of complete velocity adaptation, where the mean values in addition have to be coupled according to $C'_x = v\,C'_t$, the second positivity requirement can be written $|v|\lambda_t \leq \lambda_{xx} + v^2\lambda_t - |v\lambda_t - v\lambda_t|$, corresponding to previously mentioned requirement in the case when $C_x = C_{xt}$, i.e., $\lambda_{xx} \geq (|v| - v^2)\lambda_t$ with $|v| \leq 1$.

Velocity adaptation in relation to the vision system as a whole. Besides the explicit expressions derived above, a main conclusion from this analysis is that it is not always possible to achieve velocity adaptation when the image velocity is high, and the diffusion equation is the only tool available for adapting the filter shapes to the direction of motion.[5] If we are interested in defining image operations that are invariant under image motion, then this is a strong motivation for adding other stabilization mechanisms to the vision system, such as explicit image warping or by sending control signals to the image acquisition system, in order to stabilize the brightness pattern in the image domain.

In practice, such functionalities can be implemented by tracking, fixation and shifting mechanisms, to which a velocity adaptation mechanisms in turn can serve as an important complement, to handle deviations between actual and predicted motion.

Choice of C_{xxt}. Motivated by the expression (74) for choosing the free parameter C_{xxyy} in the affine scale-space concept for two-dimensional signals, we have in our implementation chosen C_{xxt} as function of C_{xx}, C_{xt}, C_t and C_x according to

$$C_{xxt} = \frac{C_{xx} + C_t}{6}\left(1 - \frac{|C_{xt}|}{\min(C_t, C_{xx} - |C_x - C_{xt}|)}\right) + |C_{xt}|. \qquad (120)$$

When $C_{xy} = 0$, this choice gives C_{xxt} according to (69), while in the extreme case $|C_{xy}| = \min(C_t, C_{xx} - |C_x - C_{xt}|)$, it gives the only permissible choice $C_{xxt} = |C_{xy}|$.

---

[5]In practice, however, the problem of defining a velocity adaptation mechanism with positive discretization can, however, be reduced by either using a higher density of temporal sampling, performing the filtering at a coarser spatial resolution, or by making use of a complementary warping or stabilization module.

Velocity-adapted spatio-temporal derivatives. When computing derivatives in space-time of velocity-adapted spatio-temporal filters, or from a velocity-adapted spatio-temporal scale-space representation, it is natural to express these derivatives in a velocity-adapted $(\bar{x}, \bar{t})$-frame. Based on the eigenvectors of the covariance matrix, we can thus express eigenvector based velocity adapted spatio-temporal derivatives according to

$$\partial_{\bar{x}} = \cos\phi\,\partial_x + \sin\phi\,\partial_t \qquad \partial_{\bar{t}} = \cos\phi\,\partial_t - \sin\phi\,\partial_x \tag{121}$$

where $\phi = \arctan\dot{x}$ and $\dot{x}$ is the motion estimate. Alternatively, in terms of the transformation property under Galilean transformations, we can formulate Galilean transformation based velocity adapted spatio-temporal derivatives as

$$\partial_{\bar{x}} = \partial_x \qquad \partial_{\bar{t}} = -v\,\partial_x + \partial_t \tag{122}$$

Spatio-temporal smoothing kernels. Figure 12 shows a few examples of spatio-temporal filters from this filter family,[6] obtained by varying the eigenvalues λ_{xx} and λ_t in the case of complete velocity adaptation based on (109) and (110), given a fixed velocity estimate $\dot{x}$. Figure 13 shows the effect of varying the image velocity, while keeping the temporal scale parameter C_t constant. Alternatively, one could keep the eigenvalues $(\lambda_{xx}, \lambda_t)$ constant. Observe from these examples how we by varying the spatial smoothing parameter λ_{xx} and the temporal smoothing parameter λ_t can generate discrete receptive field profiles having different spatio-temporal characteristics. Specifically, due to the time-causality, the amount of temporal delay is bound to increase with the amount of temporal smoothing.

Figures 14–16 show such velocity-adapted spatio-temporal derivatives of spatio-temporal scale-space kernels for derivative orders up to two with regard to either space or scale. Figure 14 shows filter kernels corresponding to zero velocity, figure 15 shows kernels constructed by eigenvector based complete velocity adaptation, while figure 16 shows corresponding results with Galilean transformation based complete velocity adaptation.

The spatio-temporal kernels without velocity adaptation (corresponding to $\dot{x} = 0$) are separable in space-time, corresponding to the tensor product between the spatial and temporal scale-space representations (39) and (41). The velocity-adapted kernels corresponding to velocity adaptation with $\dot{x} \neq 0$ are, however, not separable.

---

[6]Throughout this section, we use the convention that the horizontal axis represents the spatial dimension x, the vertical axis represents time t, and the origin is at the center of the image.
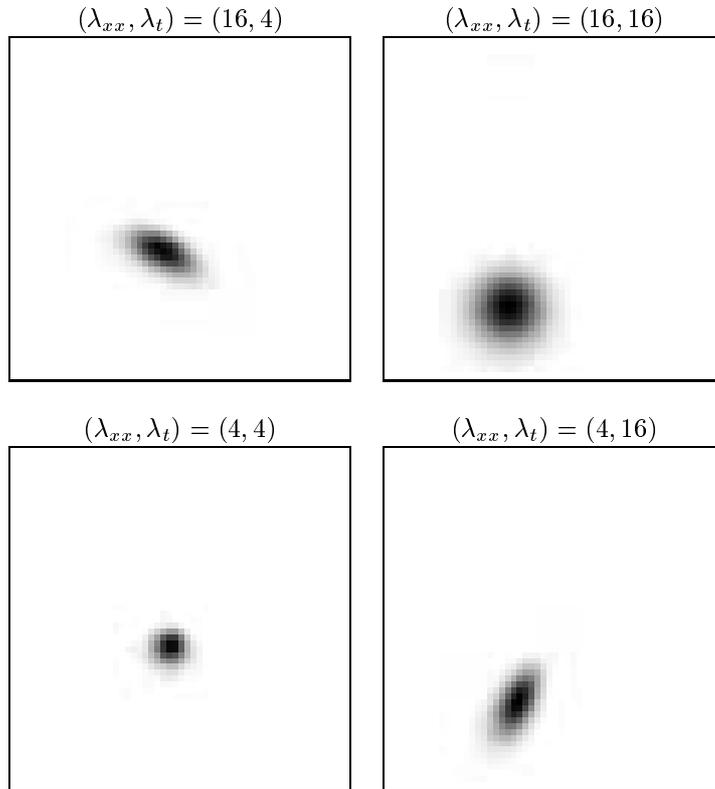
Figure 12: Velocity-adapted spatio-temporal kernels for a few values of the spatial smoothing parameter λ_{xx} and the temporal smoothing parameter λ_t. The horizontal axis represents the spatial dimension x, the vertical axis represents time t, and the origin is at the center. Here, complete velocity adaptation has been performed, resulting in a spatial offset in addition to the temporal delay that is always induced by time-causal temporal smoothing. (In all cases here, the velocity is $\dot{x} = 1/2$ in units of the grid spacing, and the image size is 50×50 pixels.)
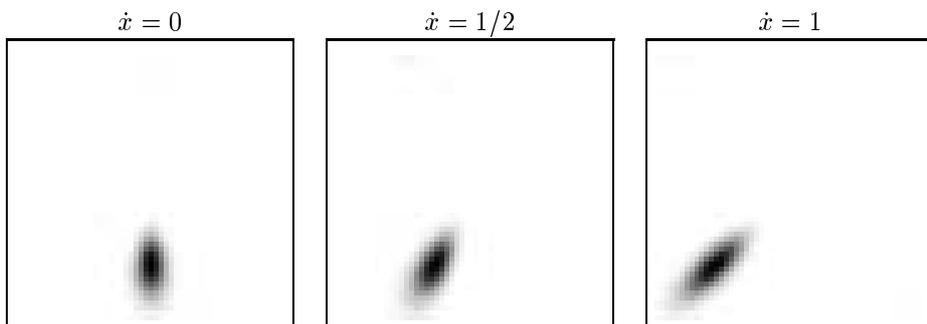
Figure 13: Spatio-temporal receptive fields for different values of the velocity parameter $\dot{x}$ using complete velocity adaptation. Observe how the spatial offset increases with the image velocity. Here, we have enforced the temporal delay to be the same in all cases, by keeping the temporal scale parameter $C_t = 16$ constant and determining the other components of the covariance matrix from the constraint that the spatial eigenvalue should satisfy $\lambda_{xx} = 4$. (Image size: 50×50 pixels.)
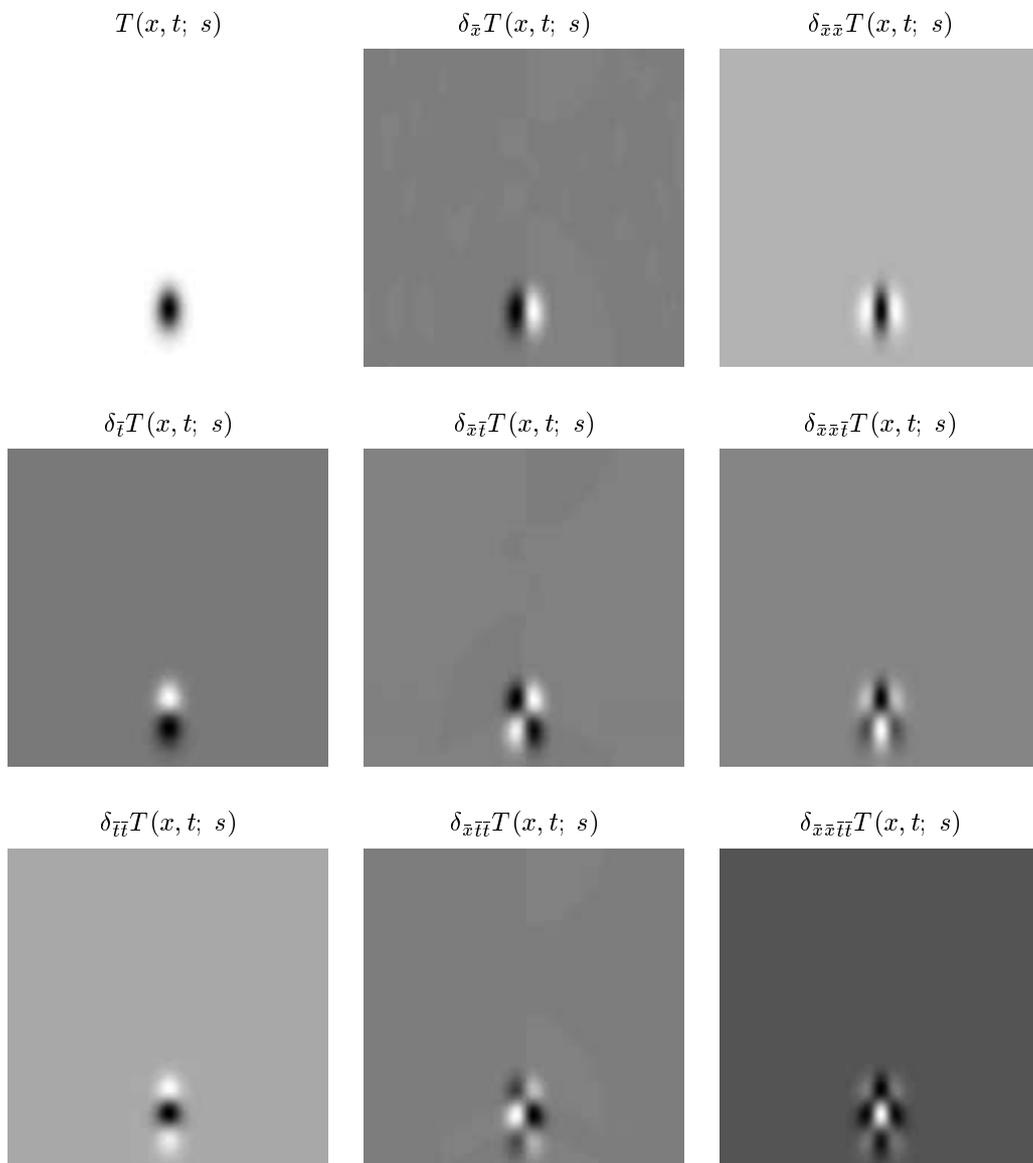
Figure 14: Regular (separable) spatio-temporal derivatives of spatio-temporal kernels for different orders of spatial and temporal differentiation. (In all cases the spatio-temporal smoothing parameters are fixed ($\lambda_{xx} = 16$, $\lambda_t = 48$, $\dot{x} = 0$) as well as the image size 150×150 pixels.)
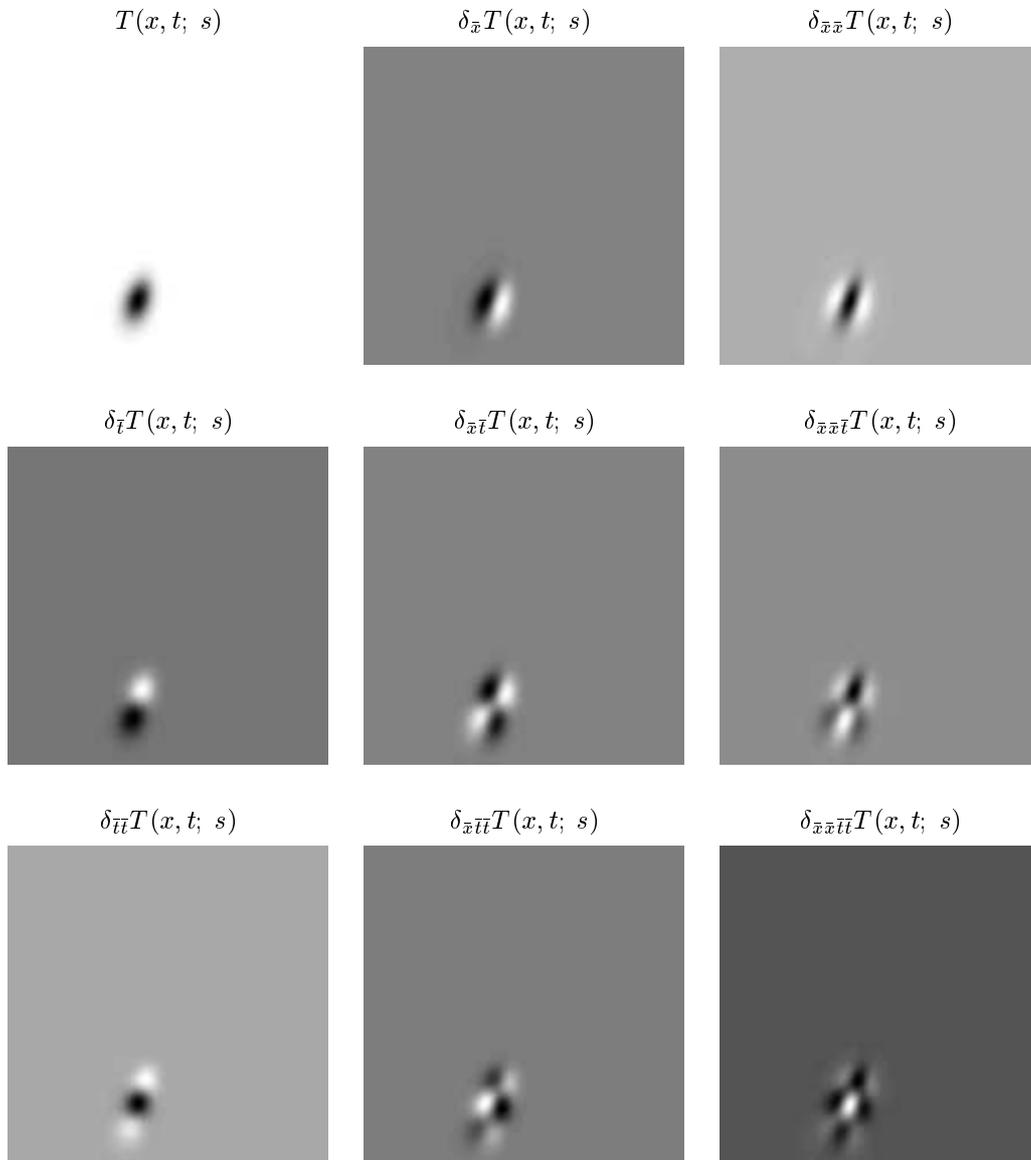
Figure 15: Eigenvector based velocity-adapted spatio-temporal derivatives of eigenvector based velocity-adapted spatio-temporal kernels for different orders of spatial and temporal differentiation. (In all cases the spatio-temporal smoothing parameters are fixed ($\lambda_{xx} = 16, \lambda_t = 48, \dot{x} = 1/3$) as well as the image size 150×150 pixels.)

$T(x, t; \; s)$　　　　$\delta_{\bar{x}} T(x, t; \; s)$　　　　$\delta_{\bar{x}\bar{x}} T(x, t; \; s)$

$\delta_{\bar{t}} T(x, t; \; s)$　　　　$\delta_{\bar{x}\bar{t}} T(x, t; \; s)$　　　　$\delta_{\bar{x}\bar{x}\bar{t}} T(x, t; \; s)$

$\delta_{\bar{t}\bar{t}} T(x, t; \; s)$　　　　$\delta_{\bar{x}\bar{t}\bar{t}} T(x, t; \; s)$　　　　$\delta_{\bar{x}\bar{x}\bar{t}\bar{t}} T(x, t; \; s)$

Figure 16: Galilean transformation based velocity-adapted spatio-temporal derivatives of Galilean transformation based velocity-adapted spatio-temporal kernels for different orders of spatial and temporal differentiation. (In all cases the spatio-temporal smoothing parameters are fixed ($\lambda_{xx} = 16, \lambda_t = 48, v = 1/3$) as well as the image size 150×150 pixels.)
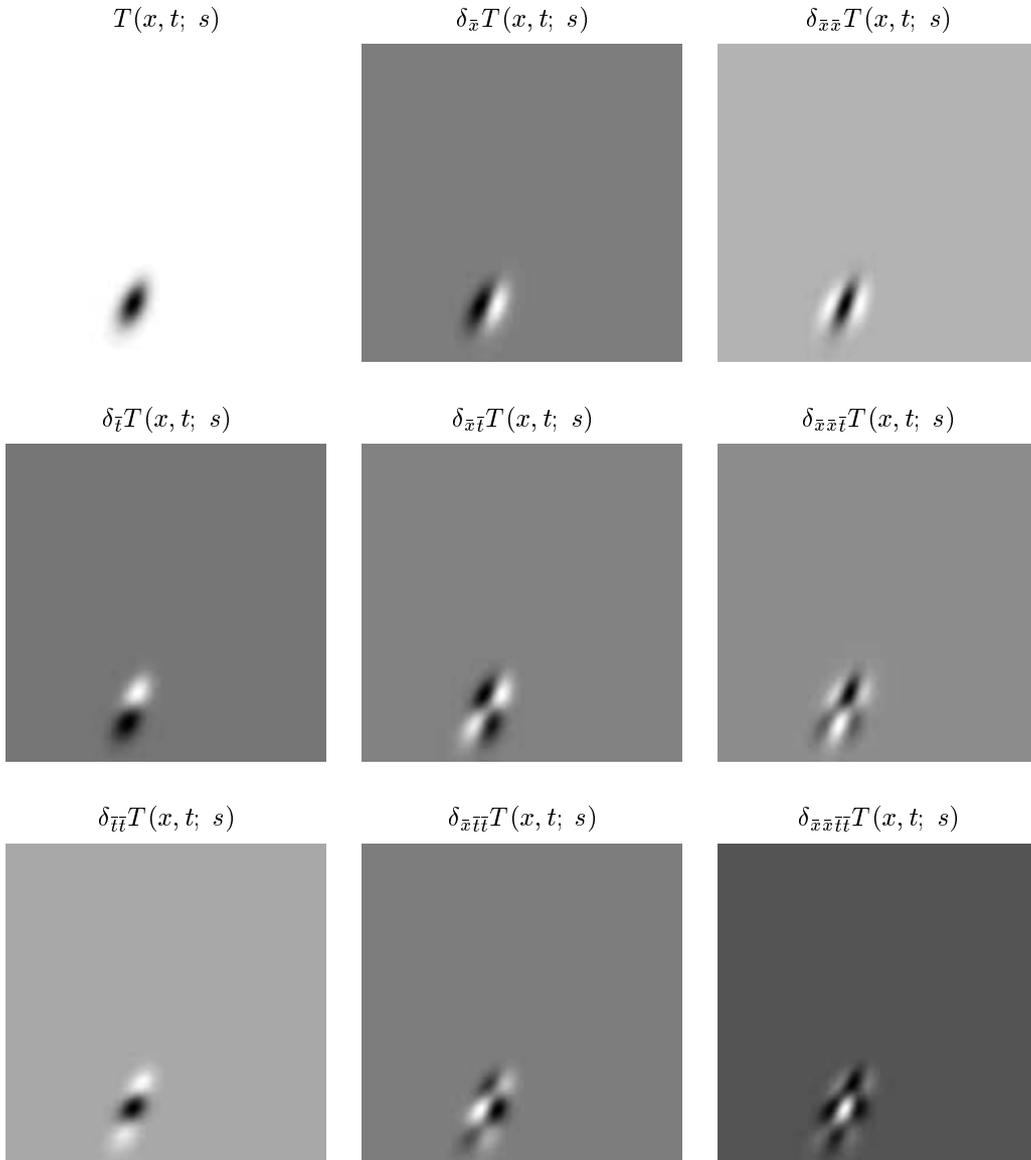
30

4.5 2+1–D discrete spatio-temporal domain

The most interesting situation for a visual observer is of course a spatio-temporal image domain with two spatial dimensions and one temporal dimension.

If we consider a general three-dimensional infinitesimal generator of the form (19) and require all coefficients $a_\xi = a_{x,y,t}$ to be zero when time t is positive, we obtain a computational molecule with 17 degrees of freedom. With horisontal and vertical indices representing the two spatial dimensions, and with time represented in terms of layers (choice of matrices), we can write the general computational molecule for such nearest-neighbour processing in $2 + 1 - D$ space-time as:

$$\mathcal{A} = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} A & B & C \\ D & -S & E \\ F & G & H \end{pmatrix}, \begin{pmatrix} I & J & K \\ L & M & N \\ P & Q & R \end{pmatrix} \right) \tag{123}$$

where the zero entries indicate forbidden access to the future, while the zero sum condition implies $A + B + C + D + E + F + G + H + I + J + K + L + M + N + P + Q + R = S$. Out of these parameters, 8 degrees of freedom will influence the mean values and the covariance matrices of the filters, while the only restriction on the remaining 9 degrees of freedom will be in terms of intervals on the corresponding coefficients (in analogy with the coefficient C_{xxyy} in the 2–D discrete affine Gaussian scale-space and the coefficient C_{xxt} in the 1+1–D discrete spatio-temporal scale-space).

Diffusion equation. To study this filter family, let us first introduce a set of difference operators δ_x, δ_y, δ_t, δ_{xx}, δ_{xy}, δ_{yy}, δ_{xt}, δ_{yt}, δ_{xxy}, δ_{xyy}, δ_{xxt}, δ_{xyt}, δ_{yyt}, δ_{xxyy}, δ_{xxyt}, δ_{xyyt} and δ_{xxyyt} as illustrated in figure 17. Then, with relations between the parameters $A \ldots R$ in (123) with corresponding filter parameters C_x, C_y, C_t, C_{xx}, C_{xy}, C_{yy}, C_{xt}, C_{yt}, C_{xxy}, C_{xyy}, C_{xxt}, C_{xyt}, C_{yyt}, C_{xxyy}, C_{xxyt}, C_{xyyt} and C_{xxyyt}, this filter family can be expressed as the solution of a differential equation of the form[7]

$$
\begin{aligned}
\partial_s L = &- C_x \, \delta_x L - C_y \, \delta_x L - C_t \, \delta_t L \\
&+ \frac{1}{2} \left(C_{xx} \, \delta_{xx} L + 2 \, C_{xy} \, \delta_{xy} L + C_{yy} \, \delta_{yy} L + 2 \, C_{xt} \, \delta_{xt} L + 2 \, C_{yt} \, \delta_{yt} L \right) \\
&- \frac{1}{6} \left(3 \, C_{xxy} \, \delta_{xxy} L + 3 \, C_{xyy} \, \delta_{xyy} L + 3 \, C_{xxt} \, \delta_{xxt} L + 6 \, C_{xyt} \, \delta_{xyt} L + 3 \, C_{yyt} \, \delta_{yyt} L \right) \\
&+ \frac{1}{24} \left(6 \, C_{xxyy} \, \delta_{xxyy} L + 12 \, C_{xxyt} \, \delta_{xxyt} L + 12 \, C_{xyyt} \, \delta_{xyyt} L \right) \\
&- \frac{1}{120} \left(30 \, C_{xxyyt} \, \delta_{xxyyt} L \right)
\end{aligned}
\tag{124}
$$

The interpretation of the filter parameters $C_x \ldots C_{xxyyt}$ is that

- spatial shape adaptation can be performed by variations of C_{xx}, C_{yy} and C_{xy},

- spatial shifts can be accomplished by varying C_x and C_y,

- velocity adaptation can be achieved by appropriate combinations of C_x, C_y, C_t, C_{xx}, C_{xy}, C_{yy}, C_{xt} and C_{yt}, and

---

[7]In this expression, the coefficient related to a specific filter parameter $C_{x^a y^b t^c}$ is $(-1)^{(a+b+c)}/(a! \, b! \, c!)$.

$$
\delta_x = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ -1/2 & 0 & +1/2 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right)
$$

$$
\delta_t = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ 0 & +1 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right)
$$

$$
\delta_{xx} = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ +1 & -2 & +1 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right)
$$

$$
\delta_{xy} = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} -1/4 & 0 & +1/4 \\ 0 & 0 & 0 \\ +1/4 & 0 & -1/4 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right)
$$

$$
\delta_{xt} = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ -1/2 & 0 & +1/2 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ +1/2 & 0 & -1/2 \\ 0 & 0 & 0 \end{pmatrix} \right)
$$

$$
\delta_{xyy} = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} -1/2 & 0 & +1/2 \\ +1 & 0 & -1 \\ -1/2 & 0 & +1/2 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right)
$$

$$
\delta_{xxt} = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ +1 & -2 & +1 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ -1 & +2 & -1 \\ 0 & 0 & 0 \end{pmatrix} \right)
$$

$$
\delta_{xyt} = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} -1/4 & 0 & +1/4 \\ 0 & 0 & 0 \\ +1/4 & 0 & -1/4 \end{pmatrix} , \begin{pmatrix} +1/4 & 0 & -1/4 \\ 0 & 0 & 0 \\ -1/4 & 0 & +1/4 \end{pmatrix} \right)
$$

$$
\delta_{xxyy} = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} 1 & -2 & 1 \\ -2 & +4 & -2 \\ 1 & -2 & 1 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right)
$$

$$
\delta_{xyyt} = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} -1/2 & 0 & +1/2 \\ +1 & 0 & -1 \\ -1/2 & 0 & +1/2 \end{pmatrix} , \begin{pmatrix} +1/2 & 0 & -1/2 \\ -1 & 0 & +1 \\ +1/2 & 0 & -1/2 \end{pmatrix} \right)
$$

$$
\delta_{xxyyt} = \left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} +1 & -2 & +1 \\ -2 & +4 & -2 \\ +1 & -2 & +1 \end{pmatrix} , \begin{pmatrix} -1 & +2 & -1 \\ +2 & -4 & +2 \\ -1 & +2 & -1 \end{pmatrix} \right)
$$

Figure 17: Computational molecules for the difference operators used in the construction of the discrete spatio-temporal scale-space representation in 2+1-D space-time. To reduce redundancy, operators that are mere rotations of corresponding operators in the horisontal direction are not shown, i.e., δ_y, δ_{yy}, δ_{yt}, δ_{yyt}, δ_{xxy} and δ_{xxyt} have been suppressed.

$$A = -\frac{C_{xy}}{4} + \frac{C_{xxyy}}{4} + \frac{C_{xxyyt}}{4} + \frac{C_{xyt}}{4} - \frac{C_{xxy}}{4} + \frac{C_{xyy}}{4} + \frac{C_{xxyt}}{4} - \frac{C_{xyyt}}{4}$$

$$B = -\frac{C_y}{2} + \frac{C_{yy}}{2} + \frac{C_{yt}}{2} - \frac{C_{xxyy}}{2} - \frac{C_{xxyyt}}{2} - \frac{C_{yyt}}{2} + \frac{C_{xxy}}{2} - \frac{C_{xxyt}}{2}$$

$$C = +\frac{C_{xy}}{4} + \frac{C_{xxyy}}{4} + \frac{C_{xxyyt}}{4} - \frac{C_{xyt}}{4} - \frac{C_{xxy}}{4} - \frac{C_{xyy}}{4} + \frac{C_{xxyt}}{4} + \frac{C_{xyyt}}{4}$$

$$D = +\frac{C_x}{2} + \frac{C_{xx}}{2} - \frac{C_{xt}}{2} - \frac{C_{xxyy}}{2} - \frac{C_{xxyyt}}{2} - \frac{C_{xxt}}{2} - \frac{C_{xyy}}{2} + \frac{C_{xyyt}}{2}$$

$$E = -\frac{C_x}{2} + \frac{C_{xx}}{2} + \frac{C_{xt}}{2} - \frac{C_{xxyy}}{2} - \frac{C_{xxyyt}}{2} - \frac{C_{xxt}}{2} + \frac{C_{xyy}}{2} - \frac{C_{xyyt}}{2}$$

$$F = +\frac{C_{xy}}{4} + \frac{C_{xxyy}}{4} + \frac{C_{xxyyt}}{4} - \frac{C_{xyt}}{4} + \frac{C_{xxy}}{4} + \frac{C_{xyy}}{4} - \frac{C_{xxyt}}{4} - \frac{C_{xyyt}}{4}$$

$$G = +\frac{C_y}{2} + \frac{C_{yy}}{2} - \frac{C_{yt}}{2} - \frac{C_{xxyy}}{2} - \frac{C_{xxyyt}}{2} - \frac{C_{yyt}}{2} - \frac{C_{xxy}}{2} + \frac{C_{xxyt}}{2}$$

$$H = -\frac{C_{xy}}{4} + \frac{C_{xxyy}}{4} + \frac{C_{xxyyt}}{4} + \frac{C_{xyt}}{4} + \frac{C_{xxy}}{4} - \frac{C_{xyy}}{4} - \frac{C_{xxyt}}{4} + \frac{C_{xyyt}}{4}$$

$$I = -\frac{C_{xxyyt}}{4} - \frac{C_{xyt}}{4} - \frac{C_{xxyt}}{4} + \frac{C_{xyyt}}{4}$$

$$J = -\frac{C_{yt}}{2} + \frac{C_{xxyyt}}{2} + \frac{C_{yyt}}{2} + \frac{C_{xxyt}}{2}$$

$$K = -\frac{C_{xxyyt}}{4} + \frac{C_{xyt}}{4} - \frac{C_{xxyt}}{4} - \frac{C_{xyyt}}{4}$$

$$L = +\frac{C_{xt}}{2} + \frac{C_{xxyyt}}{2} + \frac{C_{xxt}}{2} - \frac{C_{xyyt}}{2}$$

$$M = +C_t - C_{xxyyt} - C_{xxt} - C_{yyt}$$

$$N = -\frac{C_{xt}}{2} + \frac{C_{xxyyt}}{2} + \frac{C_{xxt}}{2} + \frac{C_{xyyt}}{2}$$

$$P = -\frac{C_{xxyyt}}{4} + \frac{C_{xyt}}{4} + \frac{C_{xxyt}}{4} + \frac{C_{xyyt}}{4}$$

$$Q = +\frac{C_{yt}}{2} + \frac{C_{xxyyt}}{2} + \frac{C_{yyt}}{2} - \frac{C_{xxyt}}{2}$$

$$R = -\frac{C_{xxyyt}}{4} - \frac{C_{xyt}}{4} + \frac{C_{xxyt}}{4} - \frac{C_{xyyt}}{4}$$

Figure 18: Relation between parameterizations of filter parameters for the 2+1-D spatio-temporal scale-space in terms of either (i) coefficients $A \ldots R$ in the general computational molecule (123) for nearest-neighbour processing in 2+1-D space-time and (ii) coefficients $C_x \ldots C_{xxyyt}$ of discrete derivative approximations as shown in figure 17.

- the other filter parameters can chosen in a complementary way, in order to achieve a positive discretization and to optimize filter performance in different ways.

From the relations in figure 18, more explicit conditions can be derived on the filter coefficients, in a similar way as the analysis of the sign conditions for the 1-D spatial, 1-D temporal, 2-D spatial and 1+1-D spatio-temporal scale-spaces treated previously. The analysis is, however, somewhat more cumbersome and these calculations are left to the reader.

Concerning the settings of the essential filter parameters, i.e., C_x, C_y, C_t, C_{xx}, C_{xy}, C_{yy}, C_{xt} and C_{yt}, the methodology we shall follow is to set the parameters C_{xx}, C_{xy}, C_{yy} that determine the shape of the kernel in the spatial domain based on the results from the two-dimensional discrete scale-space outlined in section 4.3. Then, concerning the orientation of the filter in space-time we shall choose C_x, C_y, C_t, C_{xt} and C_{yt} based on an extension from one to two dimensions of the results for the 1+1-dimensional spatio-temporal scale-space described in section 4.4. Initially, we shall here restrict ourselves to a parameterization based on Galilean transformations.

Parameterization of filter shapes based on Galilei transformations To parameterize these filter shapes with respect to motion analysis, let us initially consider an elongated spatial smoothing filter generated according to the spatial affine scale-space concept in section 4.3 and parameterized as

$$\left(\begin{array}{cc} C_{xx} & C_{xy} \\ C_{xy} & C_{yy} \end{array} \right) = \left(\begin{array}{cc} \lambda_1 \cos^2 \alpha + \lambda_2 \sin^2 \alpha & (\lambda_2 - \lambda_1) \cos \alpha \sin \alpha \\ (\lambda_2 - \lambda_1) \cos \alpha \sin \alpha & \lambda_1 \sin^2 \alpha + \lambda_2 \cos^2 \alpha \end{array} \right) \quad (125)$$

where $\lambda_{1,2}$ are the eigenvalues, and α describes the orientation of the eigenvector corresponding to α_1. Next, let us extend (125) to a space-time separable spatio-temporal smoothing kernel with variance λ_t along the temporal dimension

$$\left(\begin{array}{ccc} C_{xx} & C_{xt} & C_{xt} \\ C_{xy} & C_{yy} & C_{yt} \\ C_{xt} & C_{yt} & C_t \end{array} \right) = \left(\begin{array}{ccc} \lambda_1 \cos^2 \alpha + \lambda_2 \sin^2 \alpha & (\lambda_2 - \lambda_1) \cos \alpha \sin \alpha & 0 \\ (\lambda_2 - \lambda_1) \cos \alpha \sin \alpha & \lambda_1 \sin^2 \alpha + \lambda_2 \cos^2 \alpha & 0 \\ 0 & 0 & \lambda_t \end{array} \right) \quad (126)$$

and consider the transformation property of this smoothing kernel under a Galilean transformation with two spatial dimensions

$$\left\{ \begin{array}{l} x' = x + v_x t \\ y' = x + v_y t \\ t' = t \end{array} \right. \quad \text{i.e.} \quad \left(\begin{array}{c} x' \\ y' \\ t' \end{array} \right) = \left(\begin{array}{ccc} 1 & 0 & v_x \\ 0 & 1 & v_y \\ 0 & 0 & 1 \end{array} \right) \left(\begin{array}{c} x \\ y \\ t \end{array} \right) \quad (127)$$

From the transformation property of a covariance matrix under a Galilean transformation

$$\left(\begin{array}{ccc} C'_{xx} & C'_{xt} & C'_{xt} \\ C'_{xy} & C'_{yy} & C'_{yt} \\ C'_{xt} & C'_{yt} & C'_t \end{array} \right) = \left(\begin{array}{ccc} 1 & 0 & v_x \\ 0 & 1 & v_y \\ 0 & 0 & 1 \end{array} \right) \left(\begin{array}{ccc} C_{xx} & C_{xt} & C_{xt} \\ C_{xy} & C_{yy} & C_{yt} \\ C_{xt} & C_{yt} & C_t \end{array} \right) \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ v_x & v_y & 1 \end{array} \right) \quad (128)$$

as well as the corresponding transformation property of its mean vector

$$\left(\begin{array}{c} C'_x \\ C'_y \\ C'_t \end{array} \right) = \left(\begin{array}{ccc} 1 & 0 & v_x \\ 0 & 1 & v_y \\ 0 & 0 & 1 \end{array} \right) \left(\begin{array}{c} C_x \\ C_y \\ C_t \end{array} \right) = \left(\begin{array}{c} C_x + v_x C_t \\ C_y + v_y C_t \\ C_t \end{array} \right) \quad (129)$$

34

it then follows that the explicit expression for the simultaneously velocity-adapted and shape-adapted covariance matrix will be of the form

$$
\begin{pmatrix} C'_{xx} & C'_{xt} & C'_{xt} \\ C'_{xy} & C'_{yy} & C'_{yt} \\ C'_{xt} & C'_{yt} & C'_{t} \end{pmatrix} =
$$
$$
\begin{pmatrix} \lambda_1 \cos^2 \alpha + \lambda_2 \sin^2 \alpha + v_x^2 \lambda_t & (\lambda_2 - \lambda_1) \cos \alpha \, \sin \alpha + v_x v_y \lambda_t & v_x \lambda_t \\ (\lambda_2 - \lambda_1) \cos \alpha \, \sin \alpha + v_x v_y \lambda_t & \lambda_1 \sin^2 \alpha + \lambda_2 \cos^2 \alpha + v_y^2 \lambda_t & v_y \lambda_t \\ v_x \lambda_t & v_y \lambda_t & \lambda_t \end{pmatrix} \quad (130)
$$

Alternatively, with the velocity vector parameterized as $(v_x, v_y)^T = v(\cos \beta, \sin \beta)^T$ we have

$$
\begin{pmatrix} C'_{xx} & C'_{xt} & C'_{xt} \\ C'_{xy} & C'_{yy} & C'_{yt} \\ C'_{xt} & C'_{yt} & C'_{t} \end{pmatrix} =
$$
$$
\begin{pmatrix} \lambda_1 \cos^2 \alpha + \lambda_2 \sin^2 \alpha + \lambda_t \, v^2 \cos^2 \beta & (\lambda_2 - \lambda_1) \cos \alpha \, \sin \alpha + \lambda_t \, v^2 \cos \beta \, \sin \beta & \lambda_t \, v \, \cos \beta \\ (\lambda_2 - \lambda_1) \cos \alpha \, \sin \alpha + \lambda_t \, v^2 \cos \beta \, \sin \beta & \lambda_1 \sin^2 \alpha + \lambda_2 \cos^2 \alpha + \lambda_t \, v^2 \sin^2 \beta & \lambda_t \, v \, \sin \beta \\ \lambda_t \, v \, \cos \beta & \lambda_t \, v \, \sin \beta & \lambda_t \end{pmatrix}
$$
$$
(131)
$$

With $C_x = C_y = 0$ in the original domain, corresponding to centered kernels, the mean value of the kernel will in the case of complete velocity adaptation be given by

$$
\begin{pmatrix} C'_x \\ C'_y \\ C'_t \end{pmatrix} = C_t \begin{pmatrix} v_x \\ v_y \\ 1 \end{pmatrix} \quad (132)
$$

Velocity-adapted spatio-temporal derivatives. Velocity-adapted spatio-temporal derivatives in the velocity-adapted spatio-temporal scale-space are then obtained as

$$
\partial_{\bar{x}} = \partial_x \qquad \partial_{\bar{y}} = \partial_x \qquad \partial_{\bar{t}} = -v_x \, \partial_x - v_y \, \partial_y + \partial_t \quad (133)
$$

where $(v_x, v_y)^T$ denotes the velocity estimate.

Kernel graphs. Figure 19 shows a few examples of spatio-temporal scale-space kernels generated in this way for the separable spatio-temporal scale-space, while figure 20 shows corresponding results for a non-separable velocity-adapted scale-space expressed in a Galilean frame. (For these filters, the choice of higher order filter parameters has been determined numerically. For graphical illustrations, the results are shown as level surfaces, where the colour of the level surface indicates the polarity.)
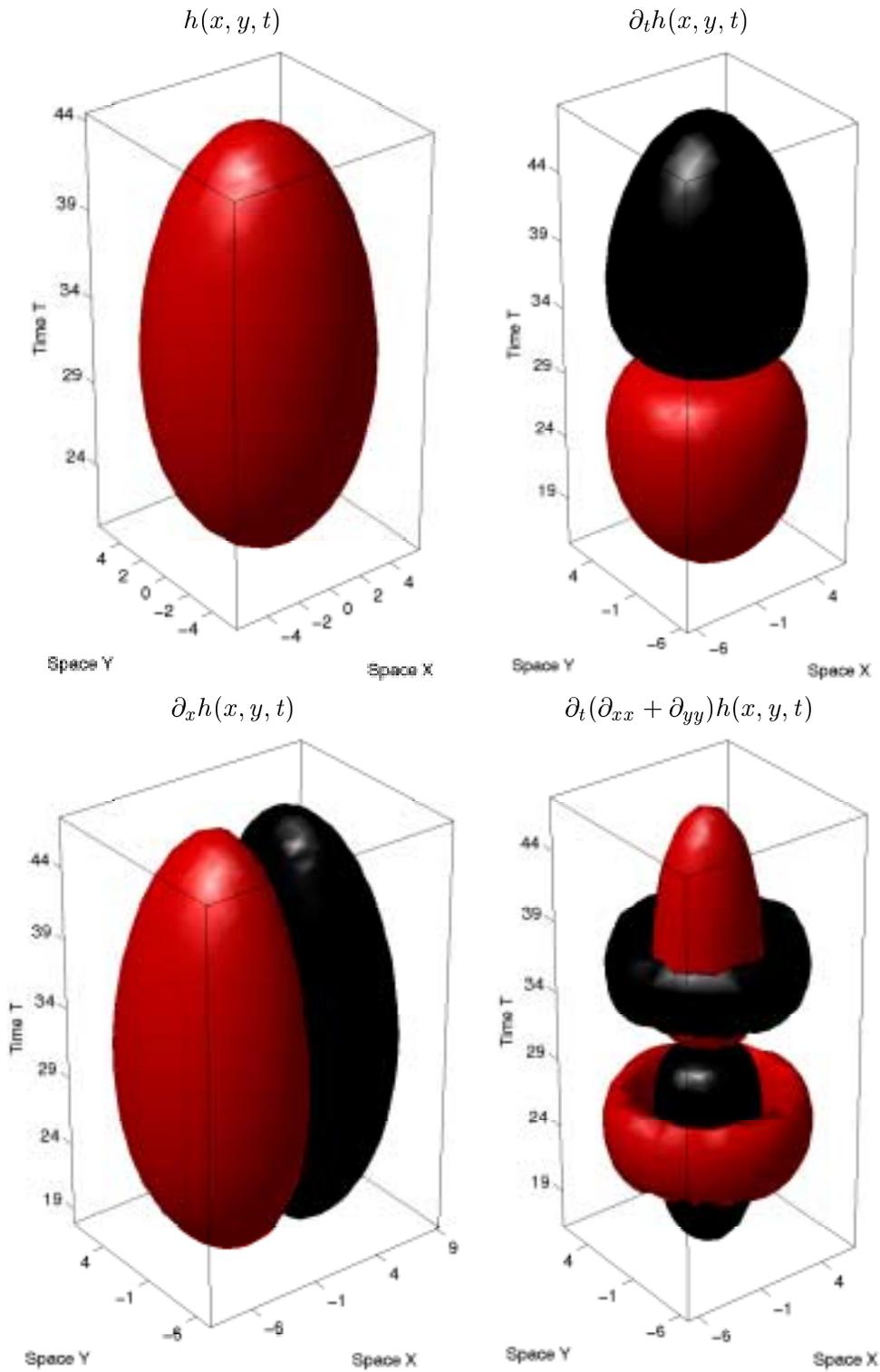
Figure 19: Level surfaces of spatio-temporal receptive fields for the 2+1-D separable spatio-temporal scale-space. (a) The raw smoothing kernel $h(x, y, t)$. (b) First-order temporal derivative $\partial_t h(x, y, t)$. (c) First-order spatial derivative $\partial_x h(x, y, t)$. (d) First-order temporal derivative of Laplacian response $\partial_t(\partial_{xx} + \partial_{yy})h(x, y, t)$. (In all cases, the smoothing parameters have been $\lambda_{xx} = 8$, $\lambda_{yy} = 8$, $\lambda_t = 8$ and $v = 0$.)

36

$$h(x, y, t) \qquad \partial_{\bar{t}} h(x, y, t)$$

$$\partial_{\bar{x}} h(x, y, t) \qquad \partial_{\bar{t}} \partial_{\bar{x}\bar{x}} h(x, y, t)$$
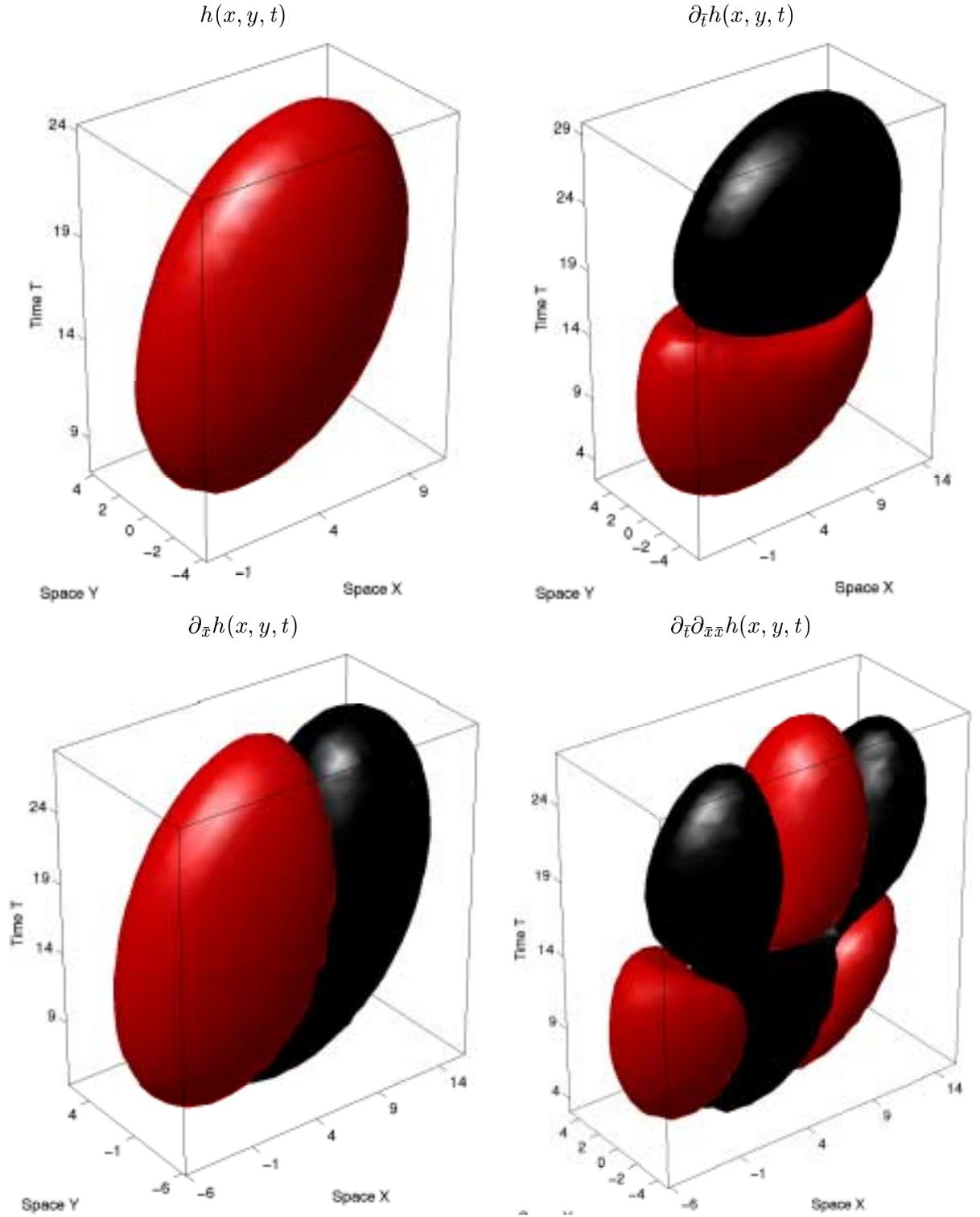
Figure 20: Level surfaces of spatio-temporal receptive fields for the 2+1-D velocity-adapted spatio-temporal scale-space. (a) The raw smoothing kernel $h(x, y, t)$. (b) First-order temporal derivative $\partial_{\bar{t}} h(x, y, t)$. (c) First-order spatial derivative $\partial_{\bar{x}} h(x, y, t)$. (d) First-order temporal derivative of second-order derivative in the velocity direction $\partial_{\bar{t}} \partial_{\bar{x}\bar{x}} h(x, y, t)$. (In all cases, the smoothing parameters have been $\lambda_{xx} = 4$, $\lambda_{yy} = 2$, $\lambda_t = 16$ and $(v_x, v_y) = (0.3, 0.0)$.)

37

5 Connection relations for discrete multi-parameter scale-spaces

In the analysis above, we have formulated differential equations that generate different forms of scale-spaces in terms of diffusion equations that are expressed in terms of one evolution parameter (s), as well as a set of filter parameters $(C_{x^a y^b t^c})$. The role of the evolution parameter in this context is to describe how the size of the filter grows as the differential equation evolves over s, while the shape of the filter was described implicitly in terms of the filter parameters $C_{x^a y^b t^c}$.

An alternative way of viewing the set of filters that can be generated in this way, or equivalently the set of scale-spaces, is in terms of a *multi-parameter filter family*, or as a *multi-parameter scale-space*. If we consider the solution of (19) for a fixed value of s (where we for simplicity make take $s = 1$) and for a discrete delta function as input, we obtain a family of scale-space filters parameterized by a set of filter parameters $C_{x^a y^b t^c}$.

The subject of this section is to describe how inter-relations can be expressed between such filters, either (i) differentially, corresponding to partial differential equations, or (ii) finite relations, corresponding to a semi-group structure. This analysis basically corresponds to an analysis in terms of Lie groups or Lie algebras (Olver 1986).

5.1 Infinitesimal relations in terms of connection equations

Let us consider any of the semi-discrete differential equations in (30), (40), (48), (84) and (124) or any other semi-discrete diffusion equation of the form (19). Introduce multi-index notation by $x^\omega = x_1^{\omega_1} x_2^{\omega_2} \ldots x_D^{\omega_D}$, where $x = (x_1, x_2, \ldots, x_D)$ are coordinates on the D-dimensional discrete domain, and let δ_{x^ω} represent a difference operator of order ω_i along the ith dimension. Moreover, let $\gamma_{x^\omega} = (-1)^{(\omega_1 + \omega_2 + \cdots + \omega_D)} / (\omega_1! \, \omega_2 \ldots \omega_D!)$ denote the Taylor coefficient used for normalizing each coefficient C_{x^ω} for $\omega \in \Omega$. Then, any such semi-differential equation can be written on the form:

$$\partial_s L = \sum_{\omega \in \Omega} \gamma_{x^\omega} \, C_{x^\omega} \, \delta_{x^\omega} \,. \tag{134}$$

Furthermore, let $\hat{\delta}_{x^\omega}(z)$ denote the generating function of the difference operator δ_{x^ω} (*e.g.*, $\hat{\delta}_{x_1}(z) = (z_1 - z_1^{-1})/2$). Then, the generating function $\hat{h}(z)$ of the Green's function of (134) at scale $s = 1$ assumes the form

$$\hat{h}(z) = e^{\hat{A}(z)} = \exp\left(\sum_{\omega \in \Omega} \gamma_{x^\omega} \, C_{x^\omega} \, \hat{\delta}_{x^\omega}(z) \right) = \prod_{\omega \in \Omega} \exp\left(\gamma_{x^\omega} \, C_{x^\omega} \, \hat{\delta}_{x^\omega}(z) \right),$$

and under small variations of any C_{x^ω}, the Green's function $h(x)$ satisfies

$$\partial_{C_{x^\omega}} h = \gamma_{x^\omega} \, \delta_{x^\omega} h. \tag{135}$$

In other words, if we define an $|\Omega|$-parameter scale-space representation L of any signal f by considering the solution of (134) at $s = 1$ for all combinations of C_{x^ω} such that $\sum_{\omega \in \Omega} C_{x^\omega} \delta_{x^\omega}$ satisfies the positivity condition in (19), then for any C_{x^ω} this scale-space representation satisfies

$$\partial_{C_{x^\omega}} L = \gamma_{x^\omega} \, \delta_{x^\omega} L. \tag{136}$$

This structure shows how representations at coarser scales can be constructed from representations at finer scales in an *incremental* fashion, implying that form small Δs we have:

$$L(\cdot;\ C + \Delta s\,\Delta C) = L(\cdot;\ C + \Delta s \sum_{\omega \in \Omega} \beta_\omega\,C_{x^\omega}) \approx L(\cdot;\ C) + \Delta s\,\partial_C L(\cdot;\ C). \quad (137)$$

Of course, we have to restrict the magnitude of Δs, relative to the units of grid spacing. Moreover, we should always choose the directional derivative operator ∂_C

$$\partial_C = \sum_{\omega \in \Omega} \beta_\omega\,\partial_{C_{x^\omega}} \qquad (\beta_\omega \in \mathbb{R}) \quad (138)$$

such that the corresponding spatial derivative operator (obtained from (136))

$$\mathcal{B} = \sum_{\omega \in \Omega} \beta_\omega\,\gamma_{x^\omega}\,\delta_{x^\omega} \quad (139)$$

satisfies the positivity conditions of the infinitesimal scale-space generator in (19). By construction, the locality and zero sum conditions are already satisfied; see figure 21 for a schematic illustration.

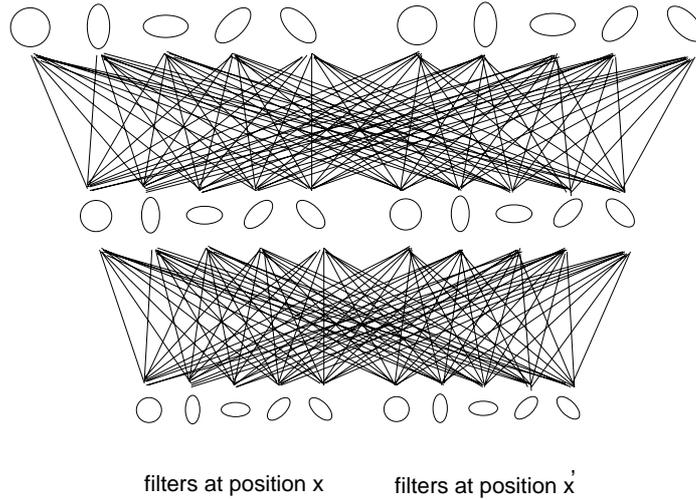filters at position x   filters at position x$'$

Figure 21: Schematic illustration of how the infinitesimal connection relations (136) and (138), or their discretized counterpart (137), describe how filters of different shapes are related differentially.

Note that when these operations are discretized in the domain of the filter parameters C_{x^ω}, they will in the spatial domain only depend on nearest-neighbour relations. Thus, they could be highly suitable for computing coarser-scale receptive fields from finer-scale receptive fields in a fine-grained computational architecture consisting of a large number of parallel processors with restricted connectivity.

5.2   Finite relations in terms of multi-parameter semi-group structure

An alternative way of constructing coarser-scale receptive fields from finer-scale receptive fields is by exploiting the $|\Omega|$-dimensional equivalence to the one-dimensional semi-group.

Let $C = (C_{x^{\omega_1}}, \ldots, C_{x^{\omega_{|\Omega|}}}$ represent an $|\Omega|$-dimensional vector containing all the filter parameters C_{x^ω}. Moreover, let $\hat{\mathcal{A}}(z; \; C)$ denote the generating function of the corresponding infinitesimal generator

$$\mathcal{A}(\cdot; \; C) = \sum \gamma_{x^\omega} C_{x^\omega} \delta_{x^\omega}. \tag{140}$$

Then, from the additive property of variances under convolution (or alternatively the multiplicative property of generating functions), it follows that if we convolve two filter kernels $h(\cdot; \; C_1)$ and $h(\cdot; \; C_2)$ with filter parameter vectors C_1 and C_2, respectively, then the filter parameters will be added in a similar fashion

$$h(\cdot; \; C_1) * h(\cdot; \; C_2) = h(\cdot; \; C_1 + C_2). \tag{141}$$

Computationally, this means that if we allow for longer range connectivity in the computational architecture than is possible in terms of nearest-neighbour-connections, we can construct coarse-scale receptive fields from finer-scale receptive fields that do not necessarily have the same shape; see figure 22 for an illustration.
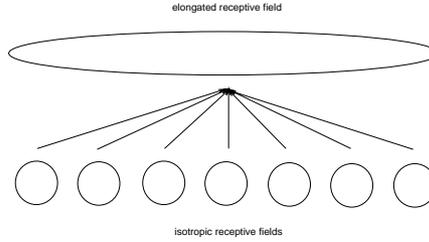

Figure 22: The multi-parameter semi-group property implies that filters at coarse scales can be constructed from linear combinations of filters at finer scales; for example an elongated receptive field at a coarse scale can be constructed from a set of isotropic receptive fields at finer scales, by summation (convolution) of the filter outputs in space.

6 Relations to biological vision

In a recent review, (DeAngelis et al. 1995) present an overview of recent results on temporal response properties of receptive fields in the central visual pathways. Foremost, the authors point out the limitations of defining receptive fields in the spatial domain only, and emphasize the need to characterize receptive fields in the *joint* space-time domain, in order to describe how a neuron processes the visual image. Then, for basic cell types in the LGN and the striate cortex, they essentially describe the spatio-temporal response characteristics as follows:

LGN neurons: The neurons in the LGN have approximately circular center-surround organization in the spatial domain (see figure 23(a)), and most receptive fields are separable in space-time. There are two main classes of temporal responses for such cells In a "non-lagged cell" the first temporal lobe is the largest one (figure 25(a)), whereas for a "lagged cell" the second lobe dominates (figure 25(b)). Such temporal response properties are typical for first- and second-order temporal derivatives of a temporal scale-space representation (compare e.g. with the graphs of the first- and second-order differences of the Poisson kernels in figure 7
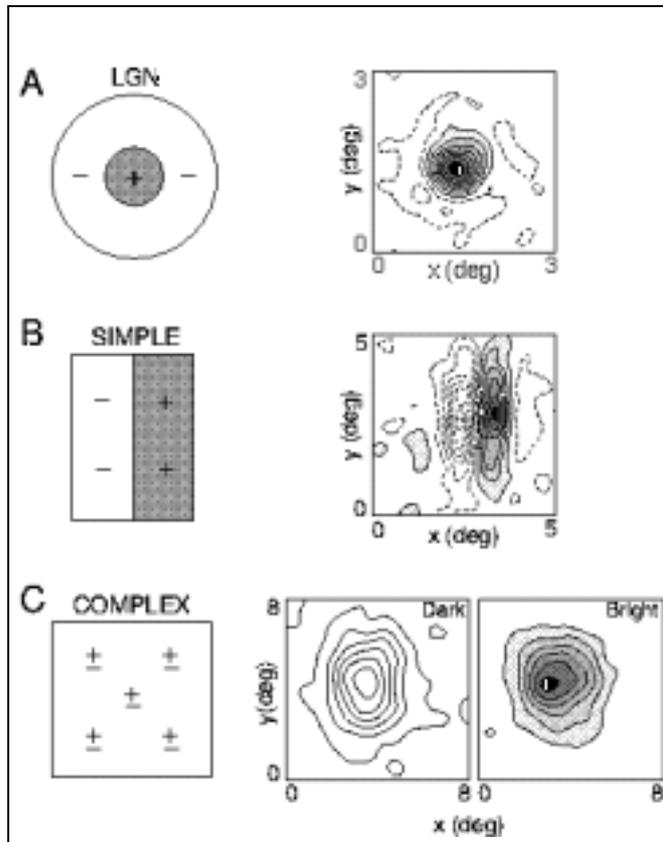
Figure 23: Examples of receptive field profiles in the spatial domain as reported by (DeAngelis et al. 1995). (a) Receptive fields in the LGN have approximately circular center-surround responses in the spatial domain. In terms of Gaussian derivatives, this spatial response profile can be modelled by the Laplacian of the Gaussian $\nabla^2 g(x; t)$. (b) Simple cells in the cerebral cortex do usually have strong directional preference in the spatial domain. In terms of Gaussian derivatives, this spatial response can be modelled as a directional derivative of a possibly elongated Gaussian kernel. (c) Complex cells are non-linear and do not obey the superposition principle.
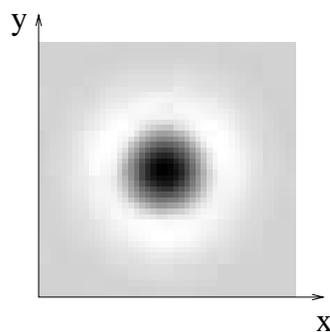

Figure 24: The discrete Laplacian of an isotropic two-dimensional spatial smoothing kernel generated according to the discrete framework in section 4.3.

and figure 8, respectively). The spatial response, on the other hand, shows a high similarity to a Laplacian of a Gaussian.

Within the abovementioned spatio-temporal scale-space theory, we could thus model the qualitative shape of these circular center-surround receptive fields in the LGN as:

$$h_{LGN}(x, y, t; \ s, C_t) = \pm \delta_{t^n} (\delta_{xx} + \delta_{yy}) \, h(x, y; \ s) \, p(t; \ C_t) \tag{142}$$

where the sign determines whether the cell is of type "on-center-off-surround" or "off-center-on-surround", the parameter $n = 1, 2$ describes the order of differentiation with respect to time, $h(x, y; \ s)$ is an isotropic smoothing kernel in the spatial domain generated by the scale-space concept in section 4.3 with spatial scale parameters $C_{xx} = C_{yy} = s$ and $C_{xy} = 0$, and $p(t; \ C_t)$ is a Poisson kernel in the temporal domain with temporal scale parameter C_t as generated from the discrete temporal scale-space concept outlined in section 4.2. Figure 24 shows an illustration of the response properties of such a receptive field.

Simple cells: For simple cells in the striate cortex, the receptive fields are oriented in the spatial domain (see figure 23(b)). In the joint space-time domain, the spatio-temporal response properties range from separable (figure 28) to strongly inseparable (figure 30), where a majority exhibit marked space-time inseparability. The temporal profile is reported to be typically biphasic, although some cells are reported to have monophasic or triphasic responses.

In terms of temporal derivatives, a biphasic behaviour arises from first-order derivatives, a monophasic behaviour from zero-order derivatives and a triphasic behaviour from second-order derivatives. Concerning the oriented spatial response characteristics, there is a high similarity with directional derivatives of Gaussian kernels (Young 1987) or directional derivatives of smoothing kernels within the corresponding discrete framework in the spatial domain (as outlined in section 4.3 and illustrated in figure 10 regarding directional responses).

In fact, for all these linear receptive fields, spatio-temporal filters with qualitatively similar response characteristics can be generated by applying difference operators of low orders to the spatio-temporal filters obtained from the spatio-temporal scale-space framework outlined in section 4.4. Figure 29 and figure 31 show a few examples of separable and inseparable kernels obtained in this way from the presented discrete framework

$$\partial_{\bar{x}}^m \partial_{\bar{t}}^n T(x, t; \ C_x, C_{xx}, C_t, C_{xt}) \tag{143}$$

with tilted directional derivative operators $\partial_{\bar{x}}$ and $\partial_{\bar{t}}$ according to (122) and the filter coefficients C_x, C_{xx}, C_t, C_{xt} determined from (107) and (108).

Motion selectivity. Concerning motion selectivity, (DeAngelis et al. 1995) report that most cortical neurons are quite sensitive to stimulus velocity, and the speed tuning is more narrow than for LGN cells. Simple cells with inseparable receptive fields have directional preference, while cells with space-time separable receptive fields do not. Moreover, the preferred direction of motion corresponds to the orientation of the filter in space-time.

This structure is nicely compatible with velocity adaptation, as described in section 4.4. Within the abovementioned terminology, separable receptive fields correspond to spatio-temporal scale-space kernels without velocity adaptation, while inseparable receptive fields correspond to kernels that are explicitly adapted to non-zero velocities.

The directional preference of the cells in the spatial domain can, in turn, be controlled by the parameters C_{xx}, C_{xy} and C_{yy} in the two-dimensional scale-space concept outlined in
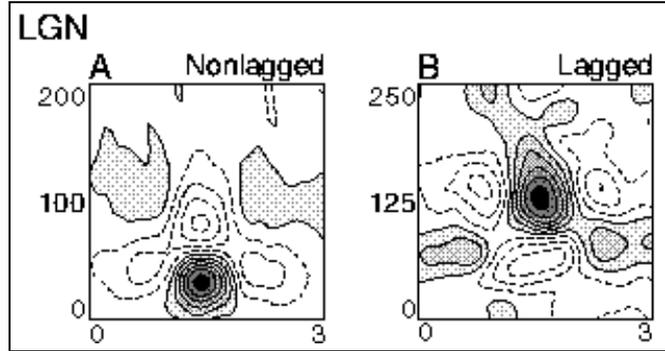
Figure 25: Examples of separable receptive field profiles in the LGN as reported by (DeAngelis et al. 1995). There are two main categories of such cells; (a) for a non-lagged cell, the first temporal lobe dominates, while (b) for a lagged cell the second temporal lobe is strongest. In terms of the spatio-temporal receptive field model presented in this paper, non-lagged cells can be modelled by first-order temporal derivatives, while the shape of lagged cells resembles of second-order temporal derivatives (see figure 26 and figure 27).

Figure 26: Separable spatio-temporal receptive fields obtained by applying first- and second-order temporal differences to second-order spatial differences of spatio-temporal smoothing kernels generated by the spatio-temporal scale-space concept outlined in section 4.4. Compare the qualitative shapes of these kernels with the kernels in figure 25.
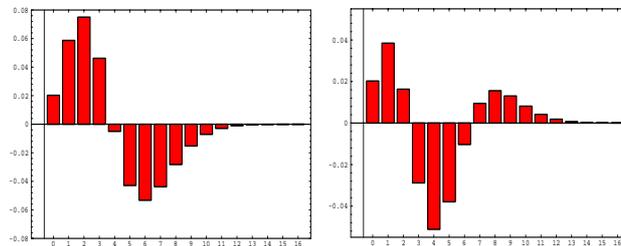
Figure 27: First- and second order differences of Poisson kernels generated by the temporal scale-space concept outlined in section 4.2. Observe that for a first-order difference, the first peak has the highest absolute value, while for a second-order difference the second peak is strongest.

Figure 28: Examples of separable receptive field profiles in striate cortex as reported by (DeAngelis et al. 1995): (a) a non-lagged cell reminiscent of a first-order temporal derivative in time and a second-order derivative in space (compare with figure 29(a)) (b) a lagged cell reminiscent of a second-order temporal derivative in time and a second-order derivative in space (compare with figure 29(b)).
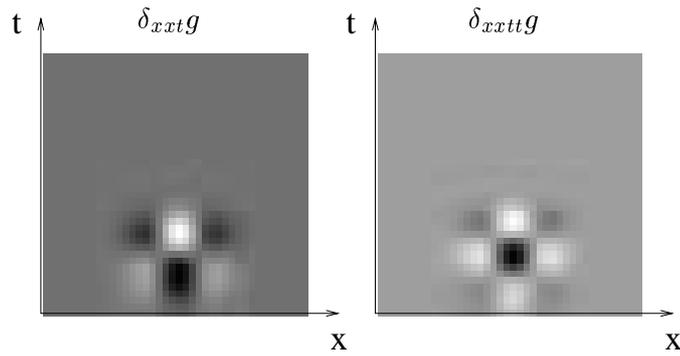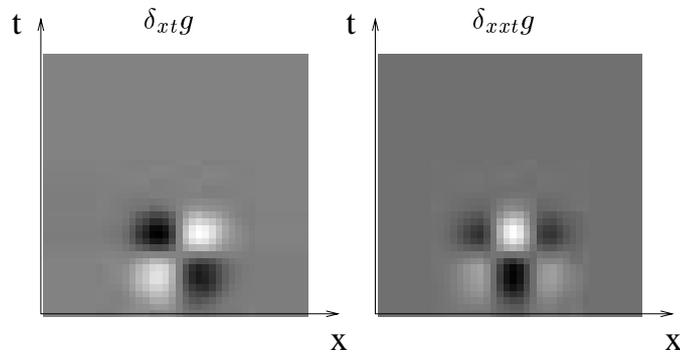

Figure 29: Separable spatio-temporal receptive fields obtained by applying first- and second-order temporal differences to second-order spatial differences of spatio-temporal smoothing kernels generated by the spatio-temporal scale-space concept outlined in section 4.4. Compare the qualitative shapes of these kernels with the kernels in figure 28.
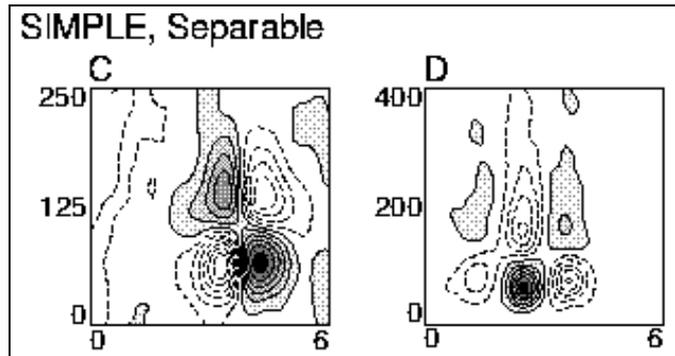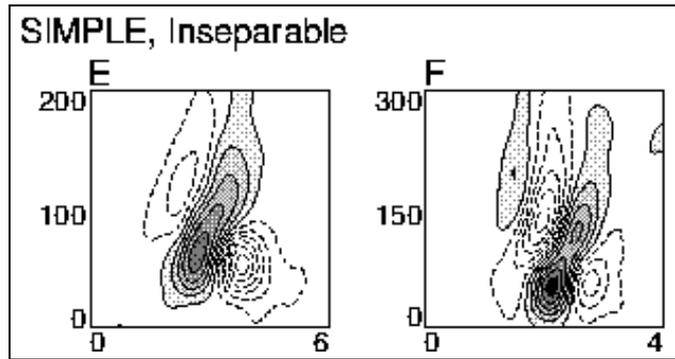
Figure 30: Examples of separable receptive field profiles in striate cortex as reported by (DeAngelis et al. 1995): (a) a receptive reminiscent of a second-order derivative in tilted space-time (compare with figure 31(a)) (b) a receptive reminiscent of a third-order derivative in tilted space-time (compare with figure 31(b)).

Figure 31: Non-separable spatio-temporal receptive fields obtained by applying tilted second- and third-order derivative operations in space-time to spatio-temporal smoothing kernels generated by the spatio-temporal scale-space concept outlined in section 4.4. Compare the qualitative shapes of these kernels with the kernels in figure 30.
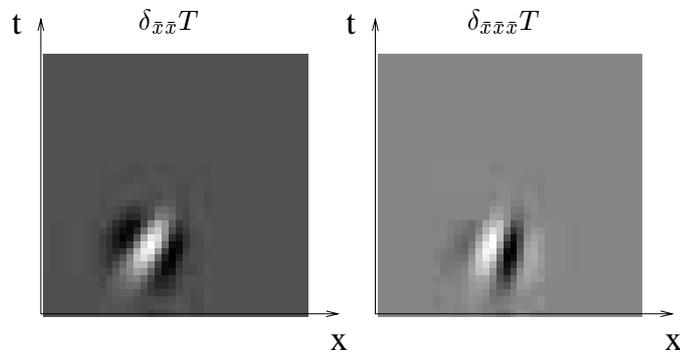
section 4.3. We obtain receptive fields without directional preference in the spatial domain if we set $C_{xx} = C_{yy}$ and $C_{xy} = 0$, and separable receptive fields if we in addition choose $C_{xt} = C_{yt} = 0$. Assuming that C_x and C_y can be neglected (e.g. by setting these parameters to zero), the filter shape will then be determined solely by the spatial scale $s = C_{xx} = C_{yy}$ and the temporal scale C_t. Conversely, we can construct inseparable kernels with strong directional preference by appropriate combinations of non-zero values of C_{xy}, C_{xt} and C_{yt}.

The abovementioned fact that a majority of the cells are inseparable in space-time is indeed nicely compatible with a description in terms of a multi-parameter scale-space as outlined in section 5. If the vision system is to give a reasonable coverage of a set of filter parameters $C_{x^a y^b t^c}$, then the set of filters corresponding to separable receptive fields (corresponding to the filter parameters $C_{x^a y^b t^c} = 0$ when either $a+b$ or c is non-zero) will be much smaller than the set of filters allowing for non-zero values of the mixed parameters over space and time.

Complex cells. Besides the abovementioned linear receptive fields, there is a large number of non-linear receptive fields, which do not obey the superposition principle, and which are referred to as complex cells. The response profile of such a cell in the spatial domain is typically of the form illustrated in figure 23(c).

In their study of spatio-temporal receptive field properties, (DeAngelis et al. 1995) also report a large number of complex cells with non-linear response profiles in the joint space-time domain; see figure 32 for an example. Within the framework of the presented spatio-temporal scale-space concept, it is interesting to note that non-linear receptive fields with qualitatively similar properties can be constructed by squaring first- and second-order derivative responses and summing up these components (Koenderink & van Doorn 1990). Provided that the filters are appropriately normalized, we can then construct a quasi quadrature measure (Lindeberg 1997b) e.g. as

$$QL = L_\xi^2 + L_\tau^2 + C\left(L_{\xi\xi}^2 + 2L_\xi L_\tau + L_{\tau\tau}^2\right) \tag{144}$$

where $L_\xi = \sqrt{s}L_x$ and $L_\tau = \sqrt{C_t}L_t$ denote normalized derivatives with respect to normalized coordinates $\xi = x/\sqrt{s}$ and $\tau = t/\sqrt{C_t}$, respectively. Figure 33 shows the result of computing the response of this quasi quadrature measure to a discrete delta function. Such a computational structure is nicely compatible with a recent results by (Valois et al. 2000), who show that first- and second-order receptive fields typically occur in pairs that can be modelled as approximate Hilbert pairs.

Successive construction of receptive fields. (DeAngelis et al. 1995) also discuss how spatio-temporal receptive fields transform along the geniculostriate pathway to increasingly specialized characteristics. Notably, such a construction is straightforward to carry out within the presented spatio-temporal scale-space framework.

From Laplacian-type receptive fields at the lowest level of a computational hierarchy, it is always possible to reconstruct (the rotationally symmetric) scale-space representation by integrating Laplacian responses over scales. With ∇_D^2 denoting the D-dimensional discrete approximation to the Laplacian operator used for generating the scale-space, we have

$$L(x;\ s) = -\int_{s'=s}^{\infty} \partial_s L(x;\ s')\,ds' = \{\text{eq. (9) isotropic case}\} = -\frac{1}{2}\int_{s'=s}^{\infty} \nabla_D^2 L(x;\ s')\,ds'. \tag{145}$$
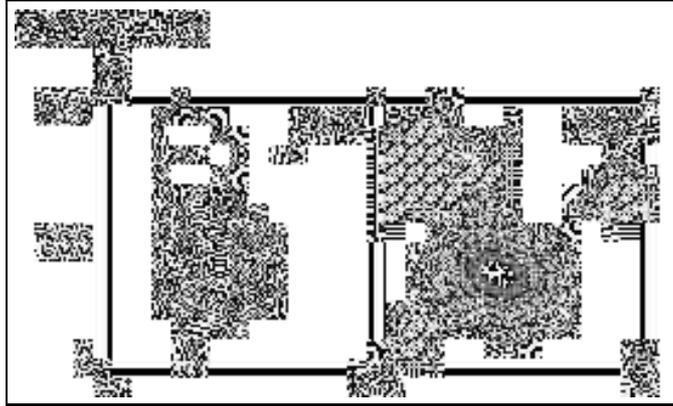
Figure 32: Response profile of a complex cell in the joint space-time domain as reported by (DeAngelis et al. 1995). Within the framework of the spatio-temporal scale-space framework presented in this paper, such a response property can be obtained by a quasi-quadrature combination of first- and second-order receptive fields; see figure 33.
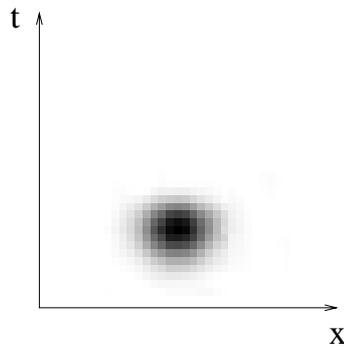
Figure 33: The response of a spatio-temporal quasi quadrature measure $\mathcal{Q}L = L_\xi^2 + L_\tau^2 + C\,(L_{\xi\xi}^2 + 2L_\xi L_\tau + L_{\tau\tau}^2)$ to a discrete delta function.

Then, a zero-order temporal scale-space representation can always be constructed in a similar way by integrating first-order temporal derivatives across temporal scales

$$L(t;\; C_t) = -\int_{C_t'=C_t}^{\infty} \partial_{C_t} L(t;\; C_t')\, dC_t' = \{\text{eq. (40)}\} = \int_{C_t'=C_t}^{\infty} \delta_t L(t;\; C_t')\, dC_t' \qquad (146)$$

Alternatively, temporal scale-space smoothing can be implemented as a cascade of first-order integrators (or recursive filters) over time. Then, from such a separable spatio-temporal scale-space representation, non-separable spatio-temporal receptive fields as well as elongated receptive fields in the spatial domain can be constructed from the connection relations described in section 5.

Finally, spatial derivative approximations can be computed from spatial differences of this representation, and temporal derivative approximations be obtained from differences between temporal channels. Since all these operations are linear, it follows that scale-space properties transfer from the spatio-temporal scale-space representation to its spatio-temporal derivatives. Moreover, since all these operations can be expressed in terms of local connections, either between neighbouring spatial positions, or over scales at a given point, they are straightforward to implement on a fine-grained massively parallel computational architecture.

7 Relations to previous works

In his pioneering scale-space formulation for temporal data, (Koenderink 1988) transformed the time axis by mapping the present moment to the unreachable infinity. He introduced a time delay t_0 that describes how long it takes before changes in the input can be perceived, and applied Gaussian smoothing on a logarithmically transformed time axis. Given the current time moment t and previous events occurring in the past $u < t$, we can thus define a logarithmically warped time axis τ_t in relation to the current moment t as

$$\tau_{(t)}(u) = -\log\left(\frac{t-u}{t_0}\right) \qquad (147)$$

where the interpretation of t_0 as a temporal delay of the system is apparent from the fact that $t - u = t_0$ corresponds to $\tau_{(t)}(u) = 0$. Then, given a signal $f(t)$, we define a time-warped signal $F(\tau_t)$ as $F(\tau) = f(t)$ and compute a multi-scale representation λ as

$$\lambda(\sigma^2) = \int_{\tau=-\infty}^{\infty} F(\tau)\, \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\tau^2/2\sigma^2}\, d\tau \qquad (148)$$

where σ is a scale parameter on the temporal axis. In the original time domain, this corresponds to computing $L(\sigma^2) = \lambda(\sigma^2)$ according to

$$L(\sigma^2) = \int_{u=-\infty}^{t} f(u)\, \frac{1}{\sqrt{2\pi}\sigma(t-u)}\, e^{-\frac{1}{2\sigma^2}\log^2\left(\frac{t-u}{t_0}\right)}\, du \qquad (149)$$

i.e., to temporal smoothing with a filter of the form

$$h_{log}(u, t;\; \sigma^2, t_0) = \frac{1}{\sqrt{2\pi}\sigma(t-u)}\, e^{-\frac{1}{2\sigma^2}\log^2\left(\frac{t-u}{t_0}\right)} \qquad (150)$$

Figure 34(a) shows an example of such a temporal smoothing filter, while figure 34(b–c) shows first- and second-order derivatives with respect to time. Figure 34(b–c) shows

corresponding temporal derivatives with respect to logarithmically transformed time $\partial_\tau = t\,\partial_t$. This form of temporal smoothing filter has been considered in follow-up works by (Florack 1997) and (ter Haar Romeny et al. 2001).

$$h_{log}(t;\ \sigma^2, t_0) \qquad\qquad \partial_t h_{log}(t;\ \sigma^2, t_0) \qquad\qquad \partial_\tau h_{log}(t;\ \sigma^2, t_0)$$
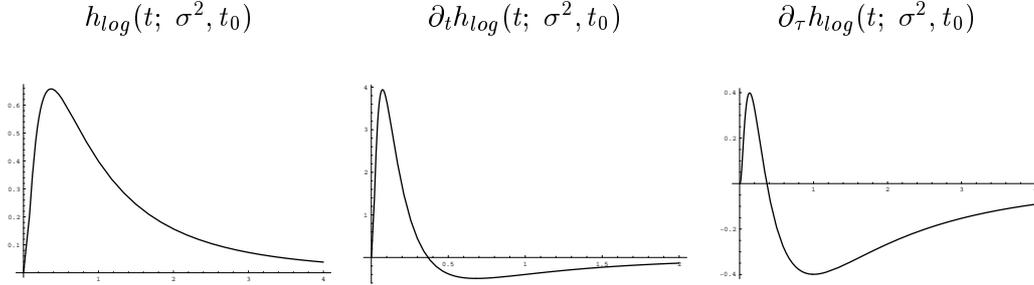

Figure 34: (a) Temporal smoothing functions corresponding to Gaussian convolution on a logarithmically transformed time axis ($\sigma = 1$, $t_0 = 0$, $\alpha = 1$). (b) First-order derivative with respect to time t. (c) First-order derivative with respect to logarithmically transformed time τ.

Related works have been presented by several other authors. Based on one-dimensional scale-space kernels, which guarantee non-creation of local extrema with increasing scales and respect the time direction as causal (Lindeberg 1990), (Lindeberg & Fagerström 1996) expressed a strictly time-recursive temporal scale-space model, in which temporal derivatives are computed from differences between temporal channels at different scales. A similar computation of temporal derivatives has been proposed by (Fleet & Langley 1995).

The spatio-temporal scale-space model presented here has several qualitative similarities to these previously presented models when projected onto a pure temporal domain. From a comparison between the kernel graphs in figures 6–8, figure 27 and figure 34, we can see that all these temporal scale-space concepts give rise to receptive fields with qualitatively similar skewed filter shapes in the temporal domain. Quantitatively, however, the kernels differ, in particular regarding the shapes of the higher-order derivative kernels. An interesting open question concerns to investigate which of these temporal scale-space models gives rise to the highest degree of agreement with receptive fields as registered from biological vision.

With regard to the temporal scale-space concepts with a discrete scale parameter there is also a large qualitative similarity with Koenderinks scale-space concept involving a logarithmically transformed time axis. If we distribute the temporal scale levels according to a geometric series, as suggested in (Lindeberg 2001), then the effective smoothing effect for the temporal smoothing filters corresponds to remembering the past in terms of a set of logarithmically distributed recursive filters coupled in cascade.

On the other hand, if we implement the temporal scale-space models presented in this paper in terms of a set of strictly time-recursive operations (such as a cascade of first-order recursive filters), then the Poisson-type scale-space at scale λ corresponds to the limit case when the number N of filters tends to infinity, while their individual variances λ/N tend to zero. In a similar way, the non-separable spatio-temporal scale-space concept presented in section 4.4 can be seen as the limit case of non-separable recursive filters according to (Lindeberg 2001) that are coupled in cascade. In this respect, these scale-space models with continuous scale parameters can be seen as idealized models on discretized domains.

Concerning comparisons between the continuous and discrete models, if follows from the central limit theorem that the discrete filters generated by (19) will approach Gaussian

kernels when the evolution parameter s tends to infinity. Moreover, the algebraic structure of the mean values and the covariance matrices of the scale-space filters is similar in the continuous and discrete domains. In this respect, the continuous theory in section 3 can be regarded as an idealized model of the discrete theory in in section 4 when grid effects are negligible. An advantage of the discrete model is that allows for a well-founded transition to coarse scales, such that all filters are scale-space kernels also in a discrete sense. Moreover, temporal causality holds exactly for the discrete filter class, and positive discretizations are guaranteed. Related works on non-negative discretizations of diffusion equations on spatial domains have been presented by (Weickert 1998).

There are also interesting connections to morphological scale-space concepts that commute with Galilean transformations (Guichard 1998), non-linear scale-space that are isomorphic to linear scale-spaces on transformed domains (Florack 2001) as well as scale-spaces that are invariant under arbitrary gauge groups (Salden et al. 2001).

8   Summary and discussion

We have presented a theory for linear scale-space representation of spatio-temporal data. Starting from a general condition about non-creation of structure with increasing scales in terms of non-enhancement of local extrema, a complete characterization has been given of the semi-groups of convolution transformations that obey this requirement on different types of image domains. The resulting theory comprises several of the existing continuous and discrete scale-space theories on symmetric spatial domains, with extensions to non-symmetric spatial domains as well as temporal and spatio-temporal domains. A main underlying message is that a much richer structure of spatio-temporal filters can be obtained if we start from a reformulation of Koenderinks causality requirement into non-enhancement of local extrema, and relax the requirement of spatial symmetry that was prevalent in the earliest scale-space formulations and most follow-up works. Indeed, by considering general covariance matrices and nearest-neighbour processing for discrete sampled data, a large family of filter shapes can be generated, with high qualitative similarities to receptive field profiles registered in biological vision. Specifically, the following types of mechanisms have been considered, which constitute extensions to previous scale-space formulations.

- An explicit discrete mechanisms for shifting data during a discrete smoothing process, without explicit need for external warping.

- A discrete theory for elongated scale-space filters in the spatial domain, allowing for discrete shape adaptation to local image structures.

- A consistent scale-space representation over discrete time, which respects the time direction as strictly causal and preserves the semi-group structure.

- Simultaneous, *i.e.* non-separable, treatment of spatial and temporal domains.

- Velocity adaptation of spatio-temporal filters along the direction of motion.

- Connection equations allowing filters of different shapes and at different scales in a multi-parameter scale-space to be constructed incrementally.

Whereas such mechanisms could also be constructed on an intuitive basis, it has been shown that they arise as *consequences* of a small set of scale-space axioms.

Compared to other possible discretizations of scale-space theory, this scale-space concept has the advantage that scale-space properties are transferred completely to the discrete domain, and the mean values and the covariance matrices of the kernels are transformed in the same way as under a corresponding linear transformation of a continuous domain. On a discrete domain it is not possible to perfectly warp kernels without introducing discretization artifacts; the presented theory, however, makes it possible to express kernels of different shapes over a continuum of parameters expressed in terms of covariance matrices and mean values. Extensions to time-recursive computations for efficient non-separable spatio-temporal filters without explicit temporal buffering are presented in (Lindeberg 2001).

Further extensions.   Throughout this treatment, we have considered square grids having the same uniform spatial sampling at all scales. More generally, similar arguments based on non-enhancement of local extrema can be carried out on other discrete topologies, such as hexagonal lattices or irregular grids. Besides minor complications (concerning book-keeping), the only difference in such situations will be that the spatial summation over nearest-neighbours in (19) should be replaced by a summation over nearest-neighbours defined from the appropriate connectivity concept. The locality, positivity and zero sum conditions transfer in a straightforward manner.

Similar extensions can be performed by decreasing the spatial and temporal sampling densities at coarser spatial and temporal scales. Besides the substantial data reduction obtained by such a subsampling operation, it has the attractive property of increasing the range of velocities that can be captured by velocity adaptation. (Recall from section 4.4 that for velocity adaptation, the positivity requirement on non-central filter coefficients implies a bound on the velocity relative to the grid spacing.) Such an extension is also nicely compatible with foveated sensors, since large motions can be captured in the periphery, and guide fixation processes, which enable fine-scale structures to be resolved at higher resolution in the fovea.

Further relaxation of scale-space axioms.   More generally, one could also conceive of relaxing the requirement about translational invariance in the temporal domain, such that the filter family is not required to be a semi-group of convolution kernels in the original temporal domain. Such extensions enable closer connections to the work by (Koenderink 1988) and will be considered in further work.

A Necessity and sufficiency: Continuous signals

This appendix gives a formal statement of the fact that non-enhancement of local extrema combined with semi-group structure and regularity properties with respect to the scale parameter imply that the scale-space family must satisfy a second-order differential equation, in which the second-order terms correspond to a positive semi-definite differential operator. Earlier versions of this scale-space formulation have been presented in (Lindeberg 1996) and (Lindeberg 1997a).[8]

A.1 Definitions

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

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

- $T(\cdot;\ 0) = \delta(\cdot)$,

- *the semi-group property $T(\cdot;\ s_1) * T(\cdot;\ s_2) = T(\cdot;\ s_1 + s_2)$,*

- *the continuity requirement $\parallel T(\cdot;\ s) - \delta(\cdot) \parallel_1 \to 0$ when $s \downarrow 0$.*

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

$$L(x;\ s) = \int_{\xi \in \mathbb{R}^N} T(\xi;\ s)\, f(x - \xi). \tag{151}$$

is said to be the continuous pre-scale-space representation of f generated by T.

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

Lemma 3 *(A continuous pre-scale-space representation is differentiable)*
Let $L \colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ be the continuous pre-scale-space representation of a signal $f \colon \mathbb{R}^N \to \mathbb{R}$ in L_1. Then, L satisfies the differential equation

$$\partial_s L = \mathcal{A} L \tag{152}$$

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

---

[8]In earlier presentation of this material, the proofs were carried out in the rotationally symmetric case, while it was mentioned that relaxation of rotational symmetry implies that a linear combination of first- and second-order derivatives. This is the first time the proofs are presented in full detail.

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

$$\lim_{s \to s_0} \parallel (\mathcal{T}_s - \mathcal{T}_{s_0})f \parallel_1 = \lim_{s \to s_0} \parallel (\mathcal{T}_{s-s_0} - \mathcal{I})(\mathcal{T}_{s_0} f) \parallel_1 = 0, \tag{153}$$

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

$$\mathcal{A}f = \lim_{h \downarrow 0} \frac{\mathcal{T}_h f - f}{h}. \tag{154}$$

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

$$\begin{aligned}
\lim_{h \downarrow 0} \frac{L(\cdot, \cdot; \ s + h) - L(\cdot, \cdot; \ s)}{h} &= \lim_{h \downarrow 0} \frac{\mathcal{T}_{s+h}f - \mathcal{T}_s f}{h} \\
&= \lim_{h \downarrow 0} \frac{\mathcal{T}_h(\mathcal{T}_s f) - (\mathcal{T}_s f)}{h} = \mathcal{A}(\mathcal{T}_s f) = \mathcal{A}L(\cdot; \ s). \tag{155}
\end{aligned}$$

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

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

Definition 4 *(Pre-scale-space property: Non-enhancement of local extrema)*
A continuous pre-scale-space representation $L \colon \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ of a smooth (infinitely continuously differentiable) signal is said to possess continuous pre-scale-space properties, or equivalently not to enhance local extrema, if for every value of the scale parameter $s_0 \in \mathbb{R}_+$ it holds that if $x_0 \in \mathbb{R}^N$ is a critical point for the mapping $x \mapsto L(x_0; \ s_0)$ and if the Hessian matrix at this point is non-degenerate, then the derivative of L with respect to s at this point has the same sign as the Hessian matrix, i.e.

$$\operatorname{sign} \partial_s L = \operatorname{sign} \operatorname{trace} \mathcal{H}L. \tag{156}$$

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

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

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

A.2 Necessity

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

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

$$\partial_s L = \frac{1}{2} \nabla^T \left(\Sigma_0 \nabla L \right) - v_0^T \nabla L. \tag{157}$$

with initial condition $L(\cdot;\ 0) = f(\cdot)$ for some positive semi-definite covariance matrix Σ_0 and some vector v_0.

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

C.1. The extremum point condition (7) combined with definitions 5-6 mean that $\mathcal{A}$ must be a pure differential operator. This can be easily understood by studying the following class of counterexamples: Consider a smooth (infinitely continuously differentiable) function $f \colon \mathbb{R}^N \to \mathbb{R}$ such that f has a maximum point at the origin at which the Hessian matrix is negative definite and for some $\varepsilon > 0$ $f(x) = 0$ when $|x| \in [\frac{\varepsilon}{2}, \varepsilon]$. Split this function into two components

$$f = f_I + f_E \tag{158}$$

where

$$\nabla f_I(0) = 0, \tag{159}$$

$$\operatorname{sign} \mathcal{H} f_I(0) < 0, \tag{160}$$

$$f_I = 0 \quad \text{when} \quad |x| \geq \varepsilon/2, \tag{161}$$

$$f_E = 0 \quad \text{when} \quad |x| \leq \varepsilon. \tag{162}$$

Assume first that $f_E = 0$. Then, $\partial_s f = \mathcal{A}(f_I + f_E) = \mathcal{A}f_I$, and we must have $\mathcal{A}f_I = C_1 < 0$. Fixate these $\mathcal{A}$ and f_I and consider then any f_E for which $\mathcal{A}f_E = C_2 \neq 0$. Then, for $f = f_I + \beta_1 f_E$ it holds that

$$\partial_s L = \mathcal{A}f_I + \beta_1 \mathcal{A}F_E = C_1 + \beta_1 C_2. \tag{163}$$

Obviously, the sign of this expression can be made positive and (7) be violated by a suitable choice of β_1. Hence, for any $\varepsilon > 0$ we have that $\mathcal{A}f_E$ must be identically zero for all functions that assume non-zero values outside the region $|x| < \varepsilon$. In other words, $\mathcal{A}$ must be

a *local operator* and $\mathcal{A}f$ can only exploit information from f at the central point. This means that for any smooth function $\mathcal{A}f$ must be of the form

$$\mathcal{A}f = \sum_{\xi \in \mathbb{Z}_+^N} a_\xi L_{x^\xi} \tag{164}$$

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

C.2. The extremum point condition (7) also means that $\mathcal{A}L$ must not contain a term proportional to L or derivatives of order higher than two. This can be seen by considering a test signal of the form

$$f(x) = x_1^2 + x_2^2 + x_N^2 + \beta_2 x^\eta \tag{165}$$

for some $\eta = (\eta_1, \eta_2, \ldots, \eta_N) \in \mathbb{Z}^N$ with $|\eta| = |\eta_1| + |\eta_2| + \cdots + |\eta_N| > 2$. If $a_\xi \neq 0$ for some $\xi \in \mathbb{Z}^N$ it is clear that we can choose $\eta = \xi$ and by a suitable choice of β_2 we can make the sign of $\mathcal{A}f$ arbitrary and hence violate (7). Similarly, by considering a test signal of the form

$$f(x) = x_1^2 + x_2^2 + x_N^2 + \beta_3 \tag{166}$$

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

C.3. Concerning the remaining first- and second-order terms, let us first note that the non-enhancement condition of local extrema does not impose any constraints on the first-order terms, since the influence of the first-order terms vanishes at critical points.

Regarding the second-order terms, we can next observe that without loss of generality, the influence of the second-order terms can be written

$$\partial_s L = \frac{1}{2} \nabla^T (\Sigma_0 \nabla L) \tag{167}$$

for some symmetric matrix Σ_0. Let us next show by counter-example that Σ_0 must be positive semi-definite, i.e. if Σ_0 would have a strictly negative eigenvalue, then the non-enhancement property of local extrema would be violated. Such a violation occurs if there exists a positive definite matrix H such that $\partial_s L(0; 0) < 0$ for $L(x; 0) = x^T H x$, i.e. if for a given Σ_0 with both positive and negative eigenvalues, there exists some positive definite H such that

$$\partial_s L = \frac{1}{2} \nabla^T (\Sigma_0 \nabla L) = \nabla^T (\Sigma_0 H x) = \text{trace}(\Sigma_0 H) < 0. \tag{168}$$

To construct such a H, let us assume that Σ_0 has eigenvectors $e_1 \ldots e_N$ with associated eigenvalues $\lambda_1 \ldots \lambda_N$, and choose H to have the same set of eigenvectors while having different eigenvalues $\mu_1 \ldots \mu_N$. Then, from

$$\partial_s L = \text{trace}(\Sigma_0 H) = \sum_{i=1}^N \mu_i \lambda_i \tag{169}$$

where all $\mu_i > 0$ and at least one $\lambda_i < 0$, it is clear that we can find a suitable combination of $\mu_i > 0$ such that $\partial_s L < 0$. Hence, the requirement of non-enhancement of local extrema implies that the second-order terms must correspond to a positive definite quadratic form, while the first-order terms are unconstrained. $\qquad \square$

A.3 Sufficiency

To prove sufficiency, i.e., the reverse statement of theorem 7 is straightforward.

Theorem 8 *(Scale-space for continuous signals: Sufficiency)*
Given a semi-definite covariance matrix Σ_0, an arbitrary vector v_0, and any smooth function $f \in L_1$, the solution of the diffusion equation

$$\partial_s L = \frac{1}{2} \nabla^T \left(\Sigma_0 \nabla L \right) - v_0^T \nabla L. \tag{170}$$

with initial condition $L(\cdot;\ 0) = f$ constitutes a continuous scale-space representation. Specifically, L obeys

$$\partial_s L \leq 0 \qquad \text{at a non-degenerate local maximum}, \tag{171}$$

$$\partial_s L \geq 0 \qquad \text{at a non-degenerate local minimum}. \tag{172}$$

Proof: The regularity properties of the solution are apparent from the regularity properties of solutions of the diffusion equation. To verify non-enhancement of local extrema, consider any non-degenerate local extremum x_0 at scale s_0 with Hessian matrix H_0. Using the fact that $\nabla^T (\Sigma_0 \nabla L) = \text{trace}(\Sigma_0 H_0)$ for a signal L with Hessian matrix H_0 it follows that at this critical point the derivative of the intensity with respect to the scale parameter is given by

$$\partial_s L = \text{trace}(\Sigma_0 H_0). \tag{173}$$

If x_0 is a local minimum then H_0 is positive semidefinite. Without loss of generality we can assume that the coordinate system is oriented such that $H_0 = diag(H_{11}, H_{22}, \ldots H_{NN})$. Then, it follows that

$$\text{trace}(\Sigma_0 H_0) = \Sigma_{11} H_{11} + \Sigma_{22} H_{22} + \ldots \Sigma_{NN} H_{NN} \geq 0. \tag{174}$$

If x_0 is a local minimum, we can apply the same way of reasoning to $-L$. $\square$

B   Necessity and sufficiency: Discrete signals

This section presents a general result concerning for linear shift-invariant scale-space representations of discrete signals; the result is a generalization of earlier results presented in (Lindeberg 1990) and (Lindeberg 1994).

B.1   Definitions

For discrete signals, let us first analogue statements to Definitions 1–2, Lemma 3 and Definitions 4–6:

Definition 9 *(Discrete pre-scale-space family of kernels)*
A one-parameter family of kernels $T \colon \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ is said to be a discrete pre-scale-space family of kernels if it satisfies

- $T(\cdot;\ 0) = \delta(\cdot)$,

- *the semi-group property $T(\cdot; \ s_1) * T(\cdot; \ s_2) = T(\cdot; \ s_1 + s_2)$,*

- *the continuity requirement $\| \ T(\cdot; \ s) - \delta(\cdot) \ \|_1 \to 0$ when $s \downarrow 0$.*

Definition 10 *(Discrete pre-scale-space representation) Let $f \colon \mathbb{Z}^N \to \mathbb{R}$ be a discrete signal and let $T \colon \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ be a discrete pre-scale-space family of kernels. Then, the one-parameter family of signals $L \colon \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ given by*

$$L(x; \ s) = \sum_{\xi \in \mathbb{Z}^N} T(\xi; \ s) \, f(x - \xi). \tag{175}$$

is said to be the discrete pre-scale-space representation of f generated by T.

Lemma 11 *(A discrete pre-scale-space representation is differentiable)*
Let $L \colon \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ be the discrete pre-scale-space representation of a signal $f \colon \mathbb{Z}^N \to \mathbb{R}$ in l_1. Then, L satisfies the differential equation

$$\partial_s L = \mathcal{A} L \tag{176}$$

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

Proof: Similar too the proof of Lemma 3 $\qquad \qquad \square$

Definition 12 *(Pre-scale-space property: Non-enhancement of local extrema)*
A discrete pre-scale-space representation $L \colon \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ of a smooth (infinitely continuously differentiable) signal is said to possess pre-scale-space properties, or equivalently not to enhance local extrema, if for every value of the scale parameter $s_0 \in \mathbb{R}_+$ it holds that if $x_0 \in \mathbb{Z}^N$ is a critical point for the mapping $x \mapsto L(x; \ s_0)$ and if the Hessian matrix at this point is non-degenerate, then the derivative of L with respect to s satisfies

$$\partial_s L \leq 0 \qquad \text{at a non-degenerate local maximum,} \tag{177}$$

$$\partial_s L \geq 0 \qquad \text{at a non-degenerate local minimum.} \tag{178}$$

Definition 13 *(Discrete scale-space family of kernels) A one-parameter family of discrete pre-scale-space kernels $T \colon \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ is said to be a discrete scale-space family of kernels if for any discrete signal $f \colon \mathbb{Z}^N \to \mathbb{R} \in l_1$ the discrete pre-scale-space representation of f generated by T possesses pre-scale-space properties,* i.e., *if for any signal local extrema are never enhanced.*

Definition 14 *(Discrete scale-space representation) A discrete pre-scale-space representation $L \colon \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ of a signal $f \colon \mathbb{Z}^N \to \mathbb{R}$ generated by a family of discrete kernels $T \colon \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$, which are discrete scale-space kernels, is said to be a discrete scale-space representation of f.*

B.2 Necessity

Theorem 15 *(Scale-space for discrete signals: Necessity)*
A discrete scale-space representation $L \colon \mathbb{Z}^N \times \mathbb{R}_+ \to \mathbb{R}$ of a discrete signal $f \colon \mathbb{Z}^N \to \mathbb{R}$ satisfies the differential equation

$$\partial_s L = \mathcal{A} L \tag{179}$$

with initial condition $L(\cdot; \ 0) = f(\cdot)$ for some infinitesimal scale-space generator $\mathcal{A}$ characterized by:

- *the* locality *condition* $a_\xi = 0$ *if* $|\xi|_\infty > 1$,

- *the* positivity *constraint* $a_\xi \geq 0$ *if* $\xi \neq 0$, *and*

- *the* zero sum *condition* $\sum_{\xi \in \mathbb{Z}^D} a_\xi = 0$.

Proof: The proof consists of two parts. The first part has already been presented in lemma 11, where it was shown that the requirements on the kernels imply that the family L obeys a linear differential equation. Because of the shift invariance $\mathcal{A}L$ can be written in the form

$$(\partial_s L)(x;\ s) = (\mathcal{A}L)(x;\ s) = \sum_{\xi \in \mathbb{Z}^D} a_\xi L(x + \xi;\ s), \tag{180}$$

In the second part counterexamples will be constructed from various simple test functions in order to delimit the class of possible operators.

D.1. The extremum point conditions (177), (178) combined with definitions 5-6 mean that $\mathcal{A}$ must be *local*, i.e., that $a_\xi = 0$ if $\xi \notin N_+(0)$. This is easily understood by studying the following counterexample: First, assume that $a_{\xi_0} > 0$ for some $\xi_0 \notin N_+(0)$ and define a function $f_1 \colon \mathbb{Z}^N \to \mathbb{R}$ by

$$f_1(x) = \begin{cases} \varepsilon > 0 & \text{if } x = 0, \\ 0 & \text{if } x \in N(0), \\ 1 & \text{if } x = \xi_0, \text{ and} \\ 0 & \text{otherwise.} \end{cases} \tag{181}$$

Obviously, 0 is a local maximum point for f_1. From (152) and (19) one obtains $\partial_s L(0;\ 0) = \epsilon a_0 + a_{\xi_0}$. It is clear that this value can be positive provided that ε is chosen small enough. Hence, L cannot satisfy (177). Similarly, it can also be shown that $a_{\xi_0} < 0$ leads to a violation of the non-enhancement property (178) (let $\varepsilon < 0$). Consequently, a_ξ must be zero if $\xi \notin N_+(0)$.

D.2. With the given (weak) definitions of local extremum points it is clear that 0 is both a local maximum point and a local minimum point. Hence $\partial_s L(0;\ 0)$ must be zero, and the *coefficients sum to zero*

$$\sum_{\xi \in \mathbb{Z}^N} a_\xi = 0, \tag{182}$$

D.3. Finally, by observing that due to the zero sum condition, (19) can be written

$$\partial_s L = (\mathcal{A}L)(x;\ s) = \sum_{\xi \in N(0)} a_\xi (L(x - \xi;\ s) - L(x;\ s)), \tag{183}$$

and by considering the test function

$$f_3(x, y) = \begin{cases} \epsilon > 0 & \text{if } x = 0, \\ -1 & \text{if } x = \tilde{\xi}, \text{ and} \\ 0 & \text{otherwise,} \end{cases} \tag{184}$$

for some $\tilde{\xi}$ in $N(0)$ one easily realizes that a_ξ must be *non-negative* if $\xi \in N(0)$. The initial condition $L(\cdot;\ 0) = f$ is a direct consequence of the definition of pre-scale-space kernel. $\quad\square$

B.3 Sufficiency

Theorem 16 *(Scale-space for discrete signals: Sufficiency)*
Given an infinitesimal generator $\mathcal{A}$ corresponding to

$$(\mathcal{A}L)(x;\ s) = \sum_{\xi \in \mathbb{Z}^D} a_\xi L(x + \xi;\ s), \tag{185}$$

where the coefficients a_ξ satisfy

- *the* locality *condition $a_\xi = 0$ if $|\xi|_\infty > 1$,*

- *the* positivity *constraint $a_\xi \geq 0$ if $\xi \neq 0$, and*

- *the* zero sum *condition $\sum_{\xi \in \mathbb{Z}^D} a_\xi = 0$,*

the solution of the diffusion equation

$$\partial_s L = \mathcal{A}L \tag{186}$$

with initial condition $L(\cdot;\ 0) = f$ constitutes a discrete scale-space representation. Specifically, L obeys

$$\partial_s L \leq 0 \qquad \text{at a non-degenerate local maximum,} \tag{187}$$
$$\partial_s L \geq 0 \qquad \text{at a non-degenerate local minimum.} \tag{188}$$

Proof: From (183) it follows that the influence of L at any extremum point $(x_0;\ s_0)$ in scale-space can be written

$$\partial_s L(x_0;\ s_0) = (\mathcal{A}L)(x_0;\ s_0) = \sum_{\xi \in N(0)} a_\xi (L(x_0 - \xi;\ s_0) - L(x_0;\ s_0)), \tag{189}$$

If x_0 is a local maximum, then all differences $L(x_0 - \xi;\ s_0) - L(x_0;\ s_0)$ are negative. Combined with the positivity of a_ξ it follows that $\partial_s L(x_0;\ s_0) < 0$. If x_0 is a local minimum, we apply the same way of reasoning to $-L$. $\qquad\square$

C Difference operators

Figure 35 shows computational molecules of difference operators that are commonly used in this paper for analysing spatio-temporal scale-space concepts for discrete signals:[9]

---

[9]Extensions and combinations of these operators to 2+1-D space-time are shown in figure 17.

1–D spatial domain: x

$$\delta_x = \tfrac{1}{2} \begin{pmatrix} -1 & 0 & 1 \end{pmatrix} \qquad \delta_{xx} = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$$

1–D temporal domain: t

$$\delta_t = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

2–D spatial domain: (x, y)

$$\delta_x = \tfrac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \qquad \delta_y = \tfrac{1}{2} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

$$\delta_{xx} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{pmatrix} \qquad \delta_{xy} = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix} \qquad \delta_{yy} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\delta_{xxyy} = \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}$$

1+1–D spatio-temporal domain: (x, t)

$$\delta_x = \tfrac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \qquad \delta_t = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

$$\delta_{xx} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{pmatrix} \qquad \delta_{xt} = \tfrac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & -1 \end{pmatrix}$$

$$\delta_{xxt} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ -1 & 2 & -1 \end{pmatrix}$$

Figure 35: The difference operators used as basis for parameterizing the semi-discrete diffusion equations (30), (40), (48) and (84) for generating discrete linear scale-spaces on spatial, temporal and spatio-temporal domains, respectively.

References

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

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

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

Ballester, C. & Gonzalez, M. (1998), 'Affine invariant texture segmentation and shape from texture by variational methods', *J. of Mathematical Imaging and Vision* **9**, 141–171.

Bretzner, L. & Lindeberg, T. (1997), On the handling of spatial and temporal scales in feature tracking, *in* B. M. ter Haar Romeny, L. M. J. Florack, J. J. Koenderink & M. A. Viergever, eds, 'Scale-Space Theory in Computer Vision: Proc. First Int. Conf. Scale-Space'97', Vol. 1252 of *Lecture Notes in Computer Science*, Springer Verlag, New York, Utrecht, The Netherlands, pp. 128–139.

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

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

DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. (1995), 'Receptive field dynamics in the central visual pathways', *Trends in Neuroscience* **18**(10), 451–457.

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

Fleet, D. J. & Langley, K. (1995), 'Recursive filters for optical flow', *IEEE Trans. Pattern Analysis and Machine Intell.* **17**(1), 61–67.

Florack, L. M. J. (1997), *Image Structure*, Series in Mathematical Imaging and Vision, Kluwer Academic Publishers, Dordrecht, Netherlands.

Florack, L. M. J. (2001), 'Non-linear scale-space isomorphic to the linear case with applications to scalar, vector and multispectral images', *J. of Mathematical Imaging and Vision* **15**(1), 39–53.

Florack, L. M. J., ter Haar Romeny, B. M., Koenderink, J. J. & Viergever, M. A. (1992), Families of tuned scale-space kernels, *in* G. Sandini, ed., 'Proc. 2nd European Conf. on Computer Vision', Vol. 588 of *Lecture Notes in Computer Science*, Springer-Verlag, Santa Margherita Ligure, Italy, pp. 19–23.

Florack, L., Niessen, W. & Nielsen, M. (1998), 'The intrinsic structure of optic flow incorporating measurement duality', *Int. J. of Computer Vision* **27**(3), 263–286.

Freeman, W. T. & Adelson, E. H. (1991), 'The design and use of steerable filters', *IEEE Trans. Pattern Analysis and Machine Intell.* **13**(9), 891–906.

Griffin, L. (1996), Critical point events in affine scale space, *in* J. Sporring, M. Nielsen, L. Florack & P. Johansen, eds, 'Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory', Kluwer Academic Publishers, Copenhagen, Denmark, pp. 165–180.

Guichard, F. (1998), 'A morphological, affine, and galilean invariant scale-space for movies', *IEEE Trans. Image Processing* **7**(3), 444–456.

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

Iijima, T. (1962), Observation theory of two-dimensional visual patterns, Technical report, Papers of Technical Group on Automata and Automatic Control, IECE, Japan.

Kerckhove, M., ed. (2001), *Scale-Space and Morphology: Proc. Scale-Space'01*, Lecture Notes in Computer Science, Springer Verlag, New York, Vancouver, Canada.

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

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

Koenderink, J. J. & van Doorn, A. J. (1990), 'Receptive field families', *Biological Cybernetics* **63**, 291–298.

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

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

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

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

Lindeberg, T. (1997*a*), Linear spatio-temporal scale-space, *in* B. M. ter Haar Romeny, L. M. J. Florack, J. J. Koenderink & M. A. Viergever, eds, 'Scale-Space Theory in Computer Vision: Proc. First Int. Conf. Scale-Space'97', Vol. 1252 of *Lecture Notes in Computer Science*, Springer Verlag, New York, Utrecht, The Netherlands, pp. 113–127.

Lindeberg, T. (1997*b*), On automatic selection of temporal scales in time-casual scale-space, *in* G. Sommer & J. J. Koenderink, eds, 'Proc. AFPAC'97: Algebraic Frames for the Perception-Action Cycle', Vol. 1315 of *Lecture Notes in Computer Science*, Springer Verlag, Berlin, Kiel, Germany, pp. 94–113.

Lindeberg, T. (2001), Time-recursive velocity-adapted spatio-temporal scale-space filters, report, ISRN KTH/NA/P--01/23--SE, Dept. of Numerical Analysis and Computing Science, KTH, Stockholm, Sweden.

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

Lindeberg, T. & Gårding, J. (1997), 'Shape-adapted smoothing in estimation of 3-D depth cues from affine distortions of local 2-D structure', *Image and Vision Computing* **15**, 415–434.

Nagel, H. & Gehrke, A. (1998), Spatiotemporal adaptive filtering for estimation and segmentation of optical flow fields, *in* 'Proc. 5th European Conf. on Computer Vision', Springer-Verlag, Freiburg, Germany, pp. 86–102.

Nielsen, M., Johansen, P., Olsen, O. F. & Weickert, J., eds (1999), *Scale-Space Theories in Computer Vision: Proc. Second Int. Conf. Scale-Space'99*, Lecture Notes in Computer Science, Springer Verlag, New York, Corfu, Greece.

Olver, P. J. (1986), *Applications of Lie Groups to Differential Equations*, Vol. 107 of *Graduate Texts in Mathematics*, Springer Verlag, New York. Second Edition 1993.

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

Perona, P. (1992), 'Steerable-scalable kernels for edge detection and junction analysis', *Image and Vision Computing* **10**, 663–672.

Salden, A. H., ter Haar Romeny, B. & Viergever, M. A. (2001), 'A dynamic scale-space paradigm', *J. of Mathematical Imaging and Vision* **15**, 127–168.

Schaffalitzky, F. & Zisserman, A. (2001), Viewpoint invariant texture matching and wide baseline stereo, *in* 'Proc. 8th Int. Conf. on Computer Vision', Vancouver, Canada.

Simoncelli, E. P., Freeman, W. T., Adelson, E. H. & Heeger, D. J. (1992), 'Shiftable multi-scale transforms', *IEEE Trans. Information Theory* **38**(2).

Sporring, J., Nielsen, M., Florack, L. & Johansen, P., eds (1996), *Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory*, Series in Mathematical Imaging and Vision, Kluwer Academic Publishers, Copenhagen, Denmark.

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

ter Haar Romeny, B., Florack, L., Koenderink, J. J. & Viergever, M., eds (1997), *Scale-Space Theory in Computer Vision: Proc. First Int. Conf. Scale-Space'97*, Lecture Notes in Computer Science, Springer Verlag, New York, Utrecht, Netherlands.

ter Haar Romeny, B., Florack, L. & Nielsen, M. (2001), Scale-time kernels and models, *in* 'Scale-Space and Morphology: Proc. Scale-Space'01', Lecture Notes in Computer Science, Springer Verlag, New York, Vancouver, Canada.

Valois, R. L. D., Cottaris, N. P., Mahon, L. E., Elfer, S. D. & Wilson, J. A. (2000), 'Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity', *Vision Research* **40**(2), 3685–3702.

Weickert, J. (1998), *Anisotropic Diffusion in Image Processing*, Teubner-Verlag, Stuttgart, Germany.

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

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

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

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