Time-recursive velocity-adapted spatio-temporal scale-space filters*

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

Technical report ISRN KTH/NA/P–01/23–SE, November 2001.
Revised May 2002.

Abstract

This paper presents a theory for constructing and computing velocity-adapted scale-space filters for spatio-temporal image data. Starting from basic criteria in terms of time-causality, time-recursivity, locality and adaptivity with respect to motion estimates, a family of spatio-temporal recursive filters is proposed and analysed. An important property of the proposed family of smoothing kernels is that the spatio-temporal covariance matrices of the discrete kernels obey similar transformation properties under Galilean transformations as for continuous smoothing kernels on continuous domains. Moreover, the proposed theory provides an efficient way to compute and generate non-separable scale-space representations without need for explicit external warping mechanisms or keeping extended temporal buffers of the past. The approach can thus be seen as a natural extension of recursive scale-space filters from pure temporal data to spatio-temporal domains.

*The support from the Swedish Research Council for Engineering Sciences, TFR, the Swedish Research Council, VR, and from the Royal Swedish Academy of Sciences as well as the Knut and Alice Wallenberg Foundation is gratefully acknowledged.

Contents

1 Introduction

A basic property of real-world image data is that we may perceive and interpret them in different ways depending on the scale of observation. On spatial domains, the understanding of such multi-scale processes has grown substantially during the last two decades, and lead to multi-scale representations such as pyramids (Burt 1981, Crowley 1981) and scale-space representation (Witkin 1983, Koenderink 1984, Lindeberg 1994, Florack 1997). In particular, the linear scale-space theory developed from these premises has close relations to biological vision (Young 1987), and has established itself as a canonical model for early visual processing. Output from linear multi-scale receptive fields can serve as input to a large set of visual modules, including feature detection, shape estimation, grouping, matching, optic flow and recognition.

The world around us, however, consists of spatio-temporal data, in which the temporal dimension plays a special role, and the future cannot be accessed (Koenderink 1988, Lindeberg & Fagerström 1996). Moreover, the spatio-temporal image data arising from a vision system that observes a coherent world will be special in the respect that spatial structures tend to be stable over time.

For analysing spatio-temporal image data with this preferred structure, mechanisms such as velocity adaptation are benefitial (Lindeberg 1997, Nagel & Gehrke 1998). For example, if we compute a separable spatio-temporal scale-space representation of a moving object, then the amount of motion blur will increase with the temporal scale. This issue can be partly dealt with by stabilizing the retinal image by tracking. In the case of imperfect stabilization, however, or for static cameras without tracking ability, alternatively a single camera that observes multiple independently moving objects, a complementary approach for reducing this effect is by adapting the shapes of the scale-space filters to the direction of motion. Moreover, as will be shown in section 2, velocity-adaptation is a necessary pre-requisite for defining spatio-temporal receptive field responses that are invariant under motion.

For image data defined on spatial domains, the related notion of shape adaption has proved to be highly useful for improving the accuracy in surface orientation estimates (Lindeberg & Gårding 1997), for handling image deformations in optic flow computations (Florack et al. 1998), for increasing the robustness when computing image features (Almansa & Lindeberg 2000) and for performing affine invariant segmentation (Ballester & Gonzalez 1998) and matching (Schaffalitzky & Zisserman 2001).

The purpose of this article is to develop a theory for formulating such velocity-adapted time-causal spatio-temporal filters. Specifically, it will be shown how temporal recursive filters can be extended into spatio-temporal recursive filters in such a way that we can control the orientation of the filter in space-time and allow for efficient implementation of non-separable scale-space filtering. It should be emphasized, however, that this paper is mainly concerned with the analysis of such spatio-temporal recursive filters. In a companion paper (Laptev & Lindeberg 2002), it is shown how velocity-adapted spatio-temporal filters can be used for improving the performance of spatio-temporal recognition schemes.

2 Velocity-adapted spatio-temporal scale-space

To model a spatio-temporal scale-space representation, there are several possible approaches. In his pioneering work, (Koenderink 1988) proposed to transform the time axis by a logarithmic transformation that maps the present moment to the unreachable future and applied Gaussian filtering on the transformed domain. Based on a classification of scale-space kernel that guarantee non-creation of local extrema on a one-dimensional domain (Lindeberg 1994), (Lindeberg & Fagerström 1996) formulated time-causal separable spatio-temporal scale-space representations, from which temporal derivatives could be computed without need for any other temporal buffering than the temporal multi-scale representation, with close relations to an earlier approach for estimating optical flow by (Fleet & Langley 1995) and the use of recursive filters on spatial domains (Deriche 1987). With regard to non-separable spatio-temporal scale-space, (Lindeberg 1997) formulated a scale-space theory for non-separable receptive fields, including velocity adaptation for discrete space-time. Other follow-up works based on Koenderinks separable scale-time model have been presented by (Florack 1997, ter Haar Romeny et al. 2001). In relation to non-linear scale-space concepts, (Black 1994) has studied the closely related notion of warping image data in connection with a robust recursive scheme for optic flow estimation, (Guichard 1998) has proposed a morphological scale-space model that commutes with Galilean transformations, and (Weickert 1998) has studied non-linear scale-spaces that comprise spatial shape adaptation mechanisms.

2.1 Transformation properties of spatio-temporal scale-space

For continuous data, a simplified spatio-temporal receptive field model in terms of Gaussian filtering can be used for illustrating the algebraic structure of a spatio-temporal scale-space, if we disregard temporal causality (Lindeberg 1994, Florack 1997). Consider the following shape- (or velocity-) adapted Gaussian kernels

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

where the covariance matrix Σ describes the shape of the kernel and the mean vector m represents the position. This scale-space 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, $f_L(x_L) = f_R(x_R)$ where $x_R = Ax_L + b$, and if linear scale-space representations of these images are defined by

$$L(\cdot;\ \Sigma_L, v_L) = g(\cdot;\ \Sigma_L, v_L) * f_L(\cdot) \quad R(\cdot;\ \Sigma_R, v_R) = g(\cdot;\ \Sigma_R, v_R) * f_R(\cdot) \tag{2}$$

then L and R are related according to

$$L(x;\ \Sigma_L, v_L) = R(y;\ \Sigma_R, v_R) \tag{3}$$

where the covariance matrices Σ_L and Σ_R satisfy

$$\Sigma_R = A \Sigma_L A^T \tag{4}$$

and the velocity terms v_L and v_R in the Gaussian kernels can be traded against coordinate shifts in x_L and x_R as long as the relation

$$x_R - v_R = A(x_L - v_L) + b \tag{5}$$

is satisfied. This closedness property is highly useful whenever we consider visual tasks involving affine image deformations (see figure 1), and has been explored in various respects by (Lindeberg 1994, Lindeberg & Gårding 1997, Florack 1997, Ballester & Gonzalez 1998, Nagel & Gehrke 1998, Schaffalitzky & Zisserman 2001). Specifically, with regard to Galilean motion in the image plane

$$\left\{ \begin{array}{l} x' = x + v_x t \\ y' = y + v_y t \\ t' = t \end{array} \right. \quad \text{i.e.} \quad \begin{pmatrix} x' \\ y' \\ t' \end{pmatrix} = \begin{pmatrix} 1 & 0 & v_x \\ 0 & 1 & v_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ t \end{pmatrix} \tag{6}$$

the spatio-temporal covariance matrix will transform as

$$\begin{pmatrix} C'_{xx} & C'_{xt} & C'_{xt} \\ C'_{xy} & C'_{yy} & C'_{yt} \\ C'_{xt} & C'_{yt} & C'_{tt} \end{pmatrix} = \begin{pmatrix} 1 & 0 & v_x \\ 0 & 1 & v_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} C_{xx} & C_{xt} & C_{xt} \\ C_{xy} & C_{yy} & C_{yt} \\ C_{xt} & C_{yt} & C_{tt} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ v_x & v_y & 1 \end{pmatrix} \tag{7}$$

while for the mean vector we have

$$\begin{pmatrix} C'_x \\ C'_y \\ C'_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & v_x \\ 0 & 1 & v_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} C_x \\ C_y \\ C_t \end{pmatrix} = \begin{pmatrix} C_x + v_x\, C_t \\ C_y + v_y\, C_t \\ C_t \end{pmatrix} \tag{8}$$

It should be noted, however, that these transformation properties are not restricted to Gaussian smoothing kernels only. Rather, they hold for a rather wide family of rapidly decreasing non-negative smoothing functions. One idea we shall follow in this work is to define a family of discrete smoothing kernels such that a similar algebraic structure holds for their covariance matrices and mean vectors.

$$\begin{array}{ccccc} L(x;\ \Sigma_L, v_L) & - & \left\{ \begin{array}{c} x_R = A x_L + b \\ \Sigma_R = A \Sigma_L A^T \\ x_R - v_R = A(x_L - v_L) + b \end{array} \right\} & \rightarrow & R(y;\ \Sigma_R, v_R) \\ \uparrow & & & & \uparrow \\ *g(\cdot;\ \Sigma_L, v_L) & & & & *g(\cdot;\ \Sigma_R, v_R) \\ | & & & & | \\ f_L(x_L) & - & x_R = A x_L + b & \rightarrow & f_R(x_R) \end{array}$$

Figure 1: Commutative diagram of the Gaussian scale-space under linear transformations of the space-time coordinates, implying that the scale-space representations of two affinely deformed image patches can be aligned, either by adapting the shapes of the Gaussian kernels or equivalently by deforming the image data prior to the smoothing operation.

2.2 Time-recursive temporal scale-space for discrete time in 0+1-D

In (Lindeberg 1994, Lindeberg & Fagerström 1996) it was shown that a natural and computationally efficient temporal scale-space concept for a one-dimensional temporal signal (without spatial extent) can be constructed by coupling first-order recursive filters in cascade

$$f_{out}(t) = \frac{\mu}{1+\mu} f_{out}(t-1) + \frac{1}{1+\mu}\, f_{in}(t) \tag{9}$$

The mean of such a filter is μ and the variance μ^2. Thus, by coupling k such filters in cascade, we obtain a filter with mean $M^{(k)}$ and variance $V^{(k)}$ according to

$$M^{(k)} = \sum_{i=1}^{k} \mu_i \quad \text{and} \quad V^{(k)} = \sum_{i=1}^{k} \mu_i^2 + \mu_i. \tag{10}$$

Figure 2 shows examples of coupling such recursive filters in cascade, and figure 3 shows first-order temporal differences (derivative approximations) of these kernels

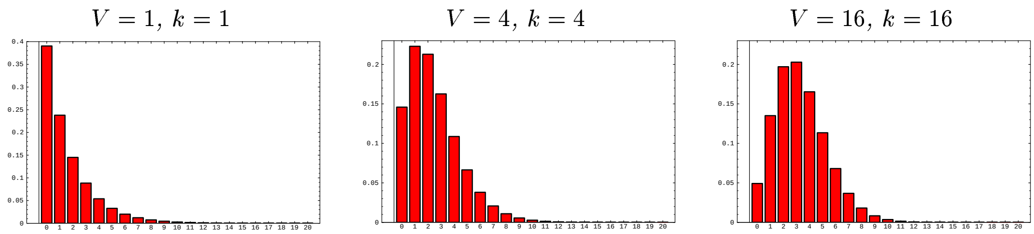

Figure 2: Equivalent temporal filters corresponding to k cascade-coupled recursive filters for which all the primitive time constants μ are equal.

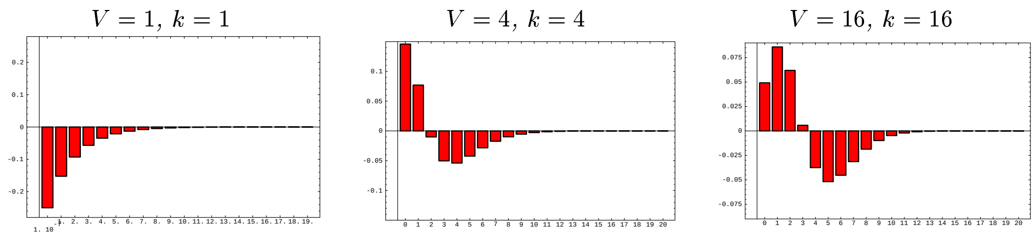

Figure 3: First-order temporal differences of the temporal filters in figure 2.

It can be shown that if we for a given total variance τ^2 in the temporal domain let the time constants become successively smaller $\mu_i = \tau^2/K$ while increasing the number of filtering steps K, then with increasing K these kernels approach the Poisson kernel (Lindeberg 1997), which corresponds to the canonical temporal scale-space concept having a continuous scale parameter on a discrete temporal domain. In practice, however, we should of course rather choose the time constants μ_i such that the variances $V^{(k)}$ are distributed according to a geometric series, which means that the individual filter parameters should with $\gamma = (V^{(K)}/V^{(1)})^{1/K}$ in the minimal case be given by

$$\mu_k = \frac{1}{2}\left(\sqrt{1 + 4V^{(1)}\,(\gamma - 1)\,\gamma^{k-1}} - 1\right). \tag{11}$$

3 Time-recursive non-separable scale-space in 1+1-D

For spatial and spatio-temporal scale-space concepts in higher dimensions, it was shown in (Lindeberg 1997) that for a discrete scale-space representation with a continuous scale parameter, the requirement of non-enhancement of local extrema[1] im-

[1]Non-enhancement of local extrema implies that the intensity value at a local maximum (minimum) must not increase (decrease) when the scale parameter s increases, i.e., $\partial_s L \leq 0$ (≥ 0) must hold at all local maxima (minima).

plies that under variations of a scale parameter s the scale-space family must satisfy a semi-discrete differential equation of the form

$$(\partial_s L)(x;\ s) = (\mathcal{A}L)(x;\ s) = \sum_{\xi \in \mathbb{Z}^D} a_\xi L(x+\xi;\ s), \tag{12}$$

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 $\sum_{\xi \in \mathbb{Z}^D} a_\xi = 0$.

When extending the temporal smoothing scheme (9) from a pure temporal domain to spatio-temporal image data, we propose to use the locality property obtained in this way to include nearest-neighbour computations in space. Thus, for a 1+1-D spatio-temporal signal with one spatial dimension and one temporal dimension, we propose to consider a smoothing scheme of the form

$$f_{out}(x,t) = \frac{1}{1+\mu_t} \begin{pmatrix} a \\ b \\ c \end{pmatrix} f_{out}(x,t-1) + \frac{1}{1+\mu_t} \begin{pmatrix} d \\ e \\ f \end{pmatrix} f_{in}(x,t) \tag{13}$$

where $a, b, c, d, e, f \geq 0$ and the vectors within parentheses denote computational symbols in the spatial domain, corresponding to the following explicit form of the smoothing scheme:

$$\begin{aligned} f_{out}(x,t) = \frac{1}{1+\mu_t} (a\, f_{out}(x+1,t-1) + b\, f_{out}(x,t-1) + c\, f_{out}(x-1,t-1) \\ + d\, f_{in}(x+1,t) + e\, f_{in}(x,t) + f\, f_{in}(x-1,t). \end{aligned} \tag{14}$$

From the generating function $\varphi(w,z)$ of the corresponding filter $T(x,t;\ s)$

$$\varphi(w,z) = \sum_{x=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} T(x,t;\ s)\, w^x z^t = \frac{dw^{-1} + e + fw}{1 + \mu_t - (aw^{-1} + b + cw)z} \tag{15}$$

where w denotes the transformation variable in the spatial domain and z the corresponding transformation variable in the temporal domain, we get the mean vector M and the covariance matrix V of the smoothing kernel as

$$M = \left. \begin{pmatrix} \varphi_w \\ \varphi_z \end{pmatrix} \right|_{(w,z)=(1,1)} = \begin{pmatrix} \mu_x \\ \mu_t \end{pmatrix}, \tag{16}$$

$$V = \left. \begin{pmatrix} \varphi_{ww} + \varphi_w - \varphi_w^2 & \varphi_{wz} - \varphi_w \varphi_z \\ \varphi_{wz} - \varphi_w \varphi_z & \varphi_{zz} + \varphi_z - \varphi_z^2 \end{pmatrix} \right|_{(w,z)=(1,1)} = \begin{pmatrix} \mu_{xx} & \mu_{xt} \\ \mu_{xt} & \mu_t^2 + \mu_t \end{pmatrix}.$$

where the explicit expressions for $\varphi(1,1)$, $\varphi_z(1,1)$, $\varphi_w(1,1)$ and $\varphi_{zz}(1,1)$ are given by

$$\varphi(1,1) = \frac{d+e+f}{1+\mu_t-(a+b+c)} \tag{17}$$

$$\varphi_z(1,1) = \frac{(a+b+c)(d+e+f)}{(1+\mu_t-(a+b+c))^2} \tag{18}$$

$$\varphi_w(1,1) = -\frac{d-bd-2cd+\mu_t d+ae-ce-f+2af+bf-\mu_t f}{(-1+a+b+c-\mu_t)^2} \tag{19}$$

$$\varphi_{zz}(1,1) = \frac{-2(a+b+c)^2\,(d+e+f)}{(-1+a+b+c-\mu_t)^3} \tag{20}$$

while the explicit expressions for $\varphi_{zw}(1,1)$ and $\varphi_{ww}(1,1)$ have been suppressed to save space. Our next aim is to solve for the parameters a, b, c, d, e and f in the recursive filter in terms of the parameters μ_x, μ_t, μ_{xx} and μ_{xt}. One additional constraint, $\varphi(1,1) = 1$, originates from the requirement that the filter should correspond to a normalized filter. Formally, this problem formally has six degrees of freedom in $a \dots f$ and six constraints in terms of the mass, mean and covariance of the filter. One complication, however, originates from the fact that the mean and the variance of the kernel in the temporal domain are coupled. Thus, we expect the problem to have $6-1-2-(3-1)=1$ degree of freedom, and solve for a, b, c, d and f in terms of e and $\mu_x \dots \mu_{xt}$. After some calculations (performed in Mathematica), it follows that

$$a = \frac{\mu_{xx}+\mu_x^2+e-1}{2} - \frac{\mu_{xt}+2\mu_x\,\mu_{xt}}{2(1+\mu_t)} \tag{21}$$

$$b = -\mu_{xx}-\mu_x^2+1-e+\mu_t+\frac{2\,\mu_x\,\mu_{xt}}{(1+\mu_t)}, \tag{22}$$

$$c = \frac{\mu_{xx}+\mu_x^2+e-1}{2} + \frac{\mu_{xt}-2\mu_x\,\mu_{xt}}{2(1+\mu_t)} \tag{23}$$

$$d = \frac{-\mu_x+1-e}{2} + \frac{\mu_{xt}}{2\,(1+\mu_t)}, \tag{24}$$

$$f = \frac{+\mu_x+1-e}{2} - \frac{\mu_{xt}}{2\,(1+\mu_t)}, \tag{25}$$

To interpret these relations, let us first parameterize the single degree of freedom in the solution in terms of $\nu = 1-e$. Then, by rewriting the spatial computational molecules in terms of the spatial difference operators δ_x and δ_{xx} as

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \frac{1}{2}\left(\mu_{xx}+\mu_x^2-\frac{2\mu_x\mu_{xt}}{1+\mu_t}-\nu\right)\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} + \mu_t \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} - \frac{\mu_{xt}}{1+\mu_t}\begin{pmatrix} 1/2 \\ 0 \\ -1/2 \end{pmatrix}$$

$$\begin{pmatrix} d \\ e \\ f \end{pmatrix} = -\mu_x \begin{pmatrix} 1/2 \\ 0 \\ -1/2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \frac{\nu}{2}\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} + \frac{\mu_{xt}}{(1+\mu_t)}\begin{pmatrix} 1/2 \\ 0 \\ -1/2 \end{pmatrix}$$

and by introducing temporal as well as mixed spatio-temporal derivatives according to

$$\delta_t(f_{in}, f_{out})(x,t) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} f_{in}(x,t) - \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} f_{out}(x,t-1) \tag{26}$$

$$\delta_{xt}(f_{in}, f_{out})(x,t) = \begin{pmatrix} 1/2 \\ 0 \\ -1/2 \end{pmatrix} f_{in}(x,t) - \begin{pmatrix} 1/2 \\ 0 \\ -1/2 \end{pmatrix} f_{out}(x,t-1) \tag{27}$$

$$\delta_{xxt}(f_{in}, f_{out})(x,t) = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} f_{in}(x,t) - \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} f_{out}(x,t-1), \tag{28}$$

we can with $\mu_{xxt} = \nu$ express the spatio-temporal smoothing scheme in (13) as

$$\begin{aligned} f_{out}(x,t) - f_{out}(x,t-1) = \\ = \frac{1}{1+\mu_t}\Big(& -\mu_x \delta_x f_{in}(x,t) + \delta_t(f_{in}, f_{out})(x,t) \\ & + \frac{1}{2}(\mu_{xx} + \mu_x^2 - \frac{2\mu_x\mu_{xt}}{1+\mu_t})\, \delta_{xx} f_{out}(x,t-1) \\ & + \frac{2}{2}\,\frac{\mu_{xt}}{1+\mu_t}\, \delta_{xt}(f_{in}, f_{out})(x,t) \\ & + \frac{3}{6}\, \mu_{xxt}\, \delta_{xxt}(f_{in}, f_{out})(x,t)\Big) \end{aligned} \tag{29}$$

From this expression it is apparent how the mean and variance parameters μ_x, μ_t, μ_{xx} and μ_{xt} influence the recursive spatio-temporal smoothing scheme. Alternatively, after introducing the following notation for binomial smoothing with variance ν in the spatial domain

$$\begin{aligned} \text{Bin}(\nu) f_{in}(x,t) &= \begin{pmatrix} \nu/2 \\ 1-\nu \\ \nu/2 \end{pmatrix} f_{in}(x,t) \\ &= \frac{\nu}{2} f_{in}(x+1,t) + (1-\nu)\, f_{in}(x+1,t) + \frac{\nu}{2} f_{in}(x-1,t) \end{aligned} \tag{30}$$

the recursive updating scheme can be expressed on either of the forms

$$\begin{aligned} f_{out}(x,t) - f_{out}(x,t-1) = \\ = \frac{1}{1+\mu_t}\Big(& -\mu_x \delta_x f_{in}(x,t) + \text{Bin}(\nu) f_{in}(x,t) - f_{out}(x,t-1) \\ & + \frac{1}{2}(\mu_{xx} + \mu_x^2 - \frac{2\mu_x\mu_{xt}}{1+\mu_t} - \nu)\, \delta_{xx} f_{out}(x,t-1) \\ & + \frac{\mu_{xt}}{1+\mu_t}\, \delta_{xt}(f_{in}, f_{out})(x,t)\Big) \end{aligned} \tag{31}$$

or

$$
\begin{aligned}
f_{out}(x,t) - f_{out}(x,t-1) = \\
= \frac{1}{1+\mu_t}(& -\mu_x \delta_x f_{in}(x,t) + \text{Bin}(\nu)\delta_t(f_{in}, f_{out})(x,t) \\
& + \frac{1}{2}(\mu_{xx} + \mu_x^2 - \frac{2\mu_x\mu_{xt}}{1+\mu_t})\, \delta_{xx} f_{out}(x,t-1) \\
& + \frac{\mu_{xt}}{1+\mu_t}\, \delta_{xt}(f_{in}, f_{out})(x,t)) \qquad (32)
\end{aligned}
$$

Hence, we can see that the free parameter $\nu = 1 - e$ can be interpreted as either a trade-off parameter between the amount of binomical pre-smoothing of the input signal and the amount of complementary spatial smoothing $\mu_{xx} - \nu$ in the recursive updating scheme (equation (31)), or as a parameter that smoothens the influence of the pointwise error signal $\delta_t(f_{in}, f_{out})(x,t) = f_{in}(x,t) - f_{out}(x,t-1)$ on the recursive smoothing scheme (equation (32)).

3.1 Conditions on ν

To analyse which values of ν give rise to non-negative discretizations, let us combine the positivity requirements on a and b in (21) and (22), which give the following necessary condition on ν:

$$\nu \leq \mu_{xx} + \mu_x^2 - \frac{2\mu_x\mu_{xt} + |\mu_{xt}|}{1+\mu_t} \qquad (33)$$

Positivity of b implies that

$$\nu \geq \mu_{xx} + \mu_x^2 - \mu_t - \frac{2\mu_x\mu_{xt}}{1+\mu_t} \qquad (34)$$

Similarly, from (24) and (25) it is apparent that positivity of d and f implies a lower bound

$$\nu \geq \left| \mu_x - \frac{\mu_{xt}}{1+\mu_t} \right| \qquad (35)$$

while positivity of e implies $\nu \leq 1$. Thus, we obtain the following necessary and sufficient condition on ν for permitting non-negative filter coefficients in the recursive filter:

$$
\begin{aligned}
\max(\left| \mu_x - \frac{\mu_{xt}}{1+\mu_t} \right|, \mu_{xx} + \mu_x^2 - \mu_t - \frac{2\mu_x\mu_{xt}}{1+\mu_t}) \leq \nu \leq \\
\leq \min(1, \mu_{xx} + \mu_x^2 - \frac{2\mu_x\mu_{xt} + |\mu_{xt}|}{1+\mu_t}) \qquad (36)
\end{aligned}
$$

More restrictive bounds on μ_x, μ_t, μ_{xx} and μ_{xt} are obtained if we require the spatial filters $(a,b,c)^T \colon f_{out}(\cdot, t-1) \to f_{out}(\cdot, t)$ and $(d,e,f)^T \colon f_{in}(\cdot, t) \to f_{out}(\cdot, t)$ to be unimodal in the spatial domain, or more strongly that each spatial filter $(a,b,c)^T \colon f_{out}(\cdot, t-1) \to f_{out}(\cdot, t)$ and $(d,e,f)^T \colon f_{in}(\cdot, t) \to f_{out}(\cdot, t)$ should be a discrete scale-space kernel, implying that $b^2 \geq 4ac$ and $e^2 \geq 4df$ (Lindeberg 1994).

3.2 Parametrization of filter shapes.

Given that K such spatio-temporal smoothing filters with parameters $\mu_x^{(i)}$, $\mu_t^{(i)}$, $\mu_{xx}^{(i)}$ and $\mu_{xt}^{(i)}$ are coupled in cascade, then from the additive properties of mean values and variances under convolution it follows that the the effective filter parameters of the composed kernel will be of the form

$$C_x^{(k)} = \sum_{i=1}^{k} \mu_x^{(i)}, \quad C_t^{(k)} = \sum_{i=1}^{k} \mu_t^{(i)}$$
$$C_{xx}^{(k)} = \sum_{i=1}^{k} \mu_{xx}^{(i)}, \quad C_{xt}^{(k)} = \sum_{i=1}^{k} \mu_{xt}^{(i)}, \quad C_{tt}^{(k)} = \sum_{i=1}^{k} (\mu_t^{(i)})^2 + \mu_t^{(i)}. \quad (37)$$

To parameterize these filter shapes, let us start from the transformation properties (7) and (8) of covariance matrices and mean vectors under Galilean motion. Then, with λ_{xx} and λ_t denoting the amount of spatial and temporal smoothing in a frame attached to the object (with separable smoothing), we have

$$\begin{pmatrix} C_{xx} & C_{xt} \\ C_{xt} & C_{tt} \end{pmatrix} = \begin{pmatrix} \lambda_{xx} + v^2 \lambda_t & v\lambda_t \\ v\lambda_t & \lambda_t \end{pmatrix} \quad (38)$$

and

$$\begin{pmatrix} C_x \\ C_t \end{pmatrix} = \begin{pmatrix} v\, C_t \\ C_t \end{pmatrix} \quad (39)$$

where C_t and C_{tt} are coupled according to (37).

3.3 Permissable filter parameters for a single smoothing step

To analyse for which combinations of λ_{xx}, λ_t and v non-negative discretization are possible, let us for simplicity consider one level of a recursive filter. From (37) and (38) it follows that

$$\mu_{xx} = \lambda_{xx} + v^2 \lambda_t \quad (40)$$
$$\mu_{xt} = v\lambda_t \quad (41)$$
$$\mu_t = \frac{1}{2}\left(\sqrt{1 + 4\lambda_t} - 1\right) \quad (42)$$
$$\mu_x = v\,\mu_t = \frac{v}{2}\left(\sqrt{1 + 4\lambda_t} - 1\right) \quad (43)$$

and for these specific values of the parameters, it holds that

$$\mu_x = \frac{\mu_{xt}}{(1 + \mu_t)}. \quad (44)$$

Hence, the lower bound on ν in (35) reduces to the trivial lower bound $\nu \geq 0$. With regard to the necessary and sufficient condition (36) for non-negative filter coeffients, we can then distinguish two cases:

- For large values of λ_{xx} and λ_t, the lower bound in (34) combined with the upper bound $\nu \leq 1$ constitute the main constraints

$$\mu_{xx} + \mu_x^2 - \mu_t - \frac{2\mu_x \mu_{xt}}{1 + \mu_t} \leq 1 \quad (45)$$

With μ_{xx}, μ_{xt}, μ_t and μ_x according to (40)–(43) it is straightforward to show that this requirement reduces to

$$\lambda_{xx} \leq \frac{1+v^2}{2} + \frac{1-v^2}{2}\sqrt{1+4\lambda_t} \tag{46}$$

Thus, given specific values of the temporal smoothing parameter λ_t and the image velocity v, positivity of the filter coeffients implies an upper bound on how much spatial smoothing can be performed by the recursive filter. Moreover, positivity for the filter coefficients implies that we have to require that $|v| \leq 1$ relative to the grid spacing.

- For small values of λ_{xx} and λ_t, the lower bound in (34) combined with the upper bound (33) constitute the main constraint:

$$\mu_{xx} + \mu_x^2 - \mu_t - \frac{2\mu_x\mu_{xt}}{1+\mu_t} \leq \mu_{xx} + \mu_x^2 - \frac{2\mu_x\mu_{xt} + |\mu_{xt}|}{1+\mu_t} \tag{47}$$

  and simply reduce to an upper bound $v \leq 1$ on the image velocity.

- In addition, we have to require that the upper bound (33) on ν is positive

$$\mu_{xx} + \mu_x^2 - \frac{2\mu_x\mu_{xt} + |\mu_{xt}|}{1+\mu_t} \geq 0 \tag{48}$$

  which reduces to a lower bound on the amount of spatial smoothing

$$\lambda_{xx} \geq \frac{|v| - v^2}{2}\left(\sqrt{1+4\lambda_t} - 1\right) \tag{49}$$

To conclude, a necessary prerequisite for the existence of a non-negative filter coefficients is that $|v| \leq 1$. In addition, the range of permissable spatial smoothing parameters λ_{xx} is delimited by

$$\frac{(|v| - v^2)}{2}\left(\sqrt{1+4\lambda_t} - 1\right) \leq \lambda_{xx} \leq \frac{1+v^2}{2} + \frac{1-v^2}{2}\sqrt{1+4\lambda_t} \tag{50}$$

Figures 4(a)–(c) show graphical illustrations of these bounds on λ_{xx} for $\lambda_t = 1$, $\lambda_t = 10$ and $\lambda_t = 100$, while table 1 shows numerical values of the lower and the upper bounds computed in this way.

λ_t	$v = 0$	$v = 1/2$	$v = 1$
1	$[0, 1.62]$	$[0.15, 1.46]$	$[0, 1]$
10	$[0, 3.70]$	$[0.68, 3.03]$	$[0, 1]$
100	$[0, 10.51]$	$[2.38, 8.13]$	$[0, 1]$

Table 1: Ranges of permissable values of the amount of spatial smoothing λ_{xx} for *one layer* of a spatio-temporal recursive filter, given specific values of the amount of temporal smoothing λ_t and the image velocity v.

If a larger amount of smoothing is required than allowed by this inequality, then either an external warping mechanism, a denser temporal sampling or a connectivity pattern that extends outside the range of nearest spatial neighbours is necessary.

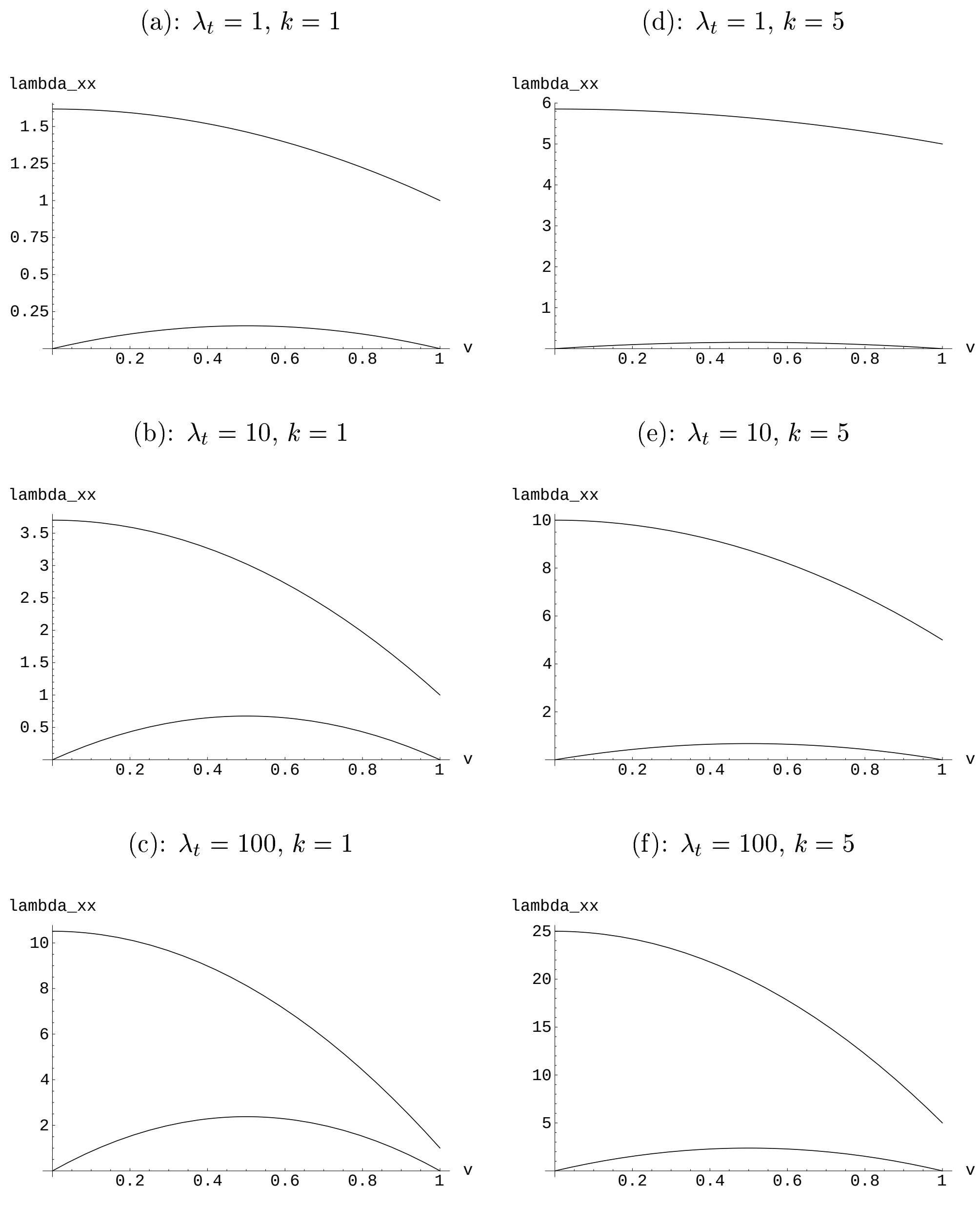

Figure 4: (left column) Graphs of the upper and the lower bounds on the spatial smoothing parameter λ_{xx} according to (50), given a value of the temporal smoothing parameter λ_t, the image velocity v as well as the requirement of non-negative filter parameters in the spatio-temporal smoothing scheme. (right column) Corresponding bounds on λ_{xx} according to (51) when decomposing the smoothing operation into $k = 5$ smoothing steps coupled in cascade.

3.4 Permissable filter parameters for cascade-coupled filters

Another way of increasing the range of image velocities that can be captured by a spatio-temporal recursive filter with non-negative filter parameters is by decomposing the smoothing step into several layers of recursive filters that are coupled in cascade. Using k such filters with identical parameters, the filter parameters for one layer will be related to the filter parameters of the composed filter according to $\lambda'_{xx} = \lambda_{xx}/k$ and $\lambda'_t = \lambda_t/k$. Thus, a decomposition into k layers implies that the bounds on the amount of spatial smoothing assume the form

$$\frac{k(|v| - v^2)}{2} \left(\sqrt{1 + \frac{4\lambda_t}{k}} - 1 \right) \leq \lambda_{xx} \leq k \left(\frac{1 + v^2}{2} + \frac{1 - v^2}{2} \sqrt{1 + \frac{4\lambda_t}{k}} \right) \tag{51}$$

Figures 4(d)–(f) show graphical illustrations of these bounds on λ_{xx} for $k = 5$ recursive filters, while table 1 shows corresponding numerical values for a few image velocities.

λ_t	$v = 0$	$v = 1/2$	$v = 1$
1	$[0, 5.85]$	$[0.21, 5.64]$	$[0, 5]$
10	$[0, 10.00]$	$[1.25, 8.75]$	$[0, 5]$
100	$[0, 25.00]$	$[5.00, 20.00]$	$[0, 5]$

Table 2: Ranges of permissable values of the amount of spatial smoothing λ_{xx} for $k = 5$ *cascade coupled layers* of a spatio-temporal recursive filter, given specific values of the total amount of temporal smoothing λ_t and the image velocity v.

3.5 Permissable filter parameters when adding spatial subsampling

If the number of layers k that is required to satisfy this inequality is too large with regard to computational efficiency, a third alternative is to carry out the computations at a coarser spatial resolution. Using the fact that the filter parameters transform as $\lambda'_{xx} = \lambda_{xx}/h^2$, $\lambda'_t = \lambda_t$ and $v' = v/h$ under a change of spatial resolution[2] by a factor h, we obtain

$$\frac{k(h\,|v| - v^2)}{2} \left(\sqrt{1 + \frac{4\lambda_t}{k}} - 1 \right) \leq \lambda_{xx} \leq k \left(\frac{h^2 + v^2}{2} + \frac{h^2 - v^2}{2} \sqrt{1 + \frac{4\lambda_t}{k}} \right) \tag{52}$$

where the upper bound on the image velocity is now given by $|v| \leq h$. Figures 5(a)–(c) and table 3 show numerical values of these bounds at resolution $h = 4$, while figures 5(d)–(f) and table 3 show corresponding results using $k = 5$ cascade coupled layers of recursive filters at the same level of resolution. From these data, it is apparent how we, without any need for an external warping mechanism, can cover wider ranges of spatial filter parameters λ_{xx} and image velocities v by appropriate choices of the spatial resolution h and the number of layers of recursive filters k. A disadvantage compared to approaches based on warping, however, is that we will not be able to

[2]When changing the spatial resolution in the image representation, the spatial smoothing parameters in the spatio-temporal smoothing scheme are also affected by the smoothing operation that precedes the subsampling operation in a sound pyramid generation scheme. For simplicity, we disregard this effect in this theoretical analysis. In a practical implementation, however, it is straightforward to take this effect into explicit account.

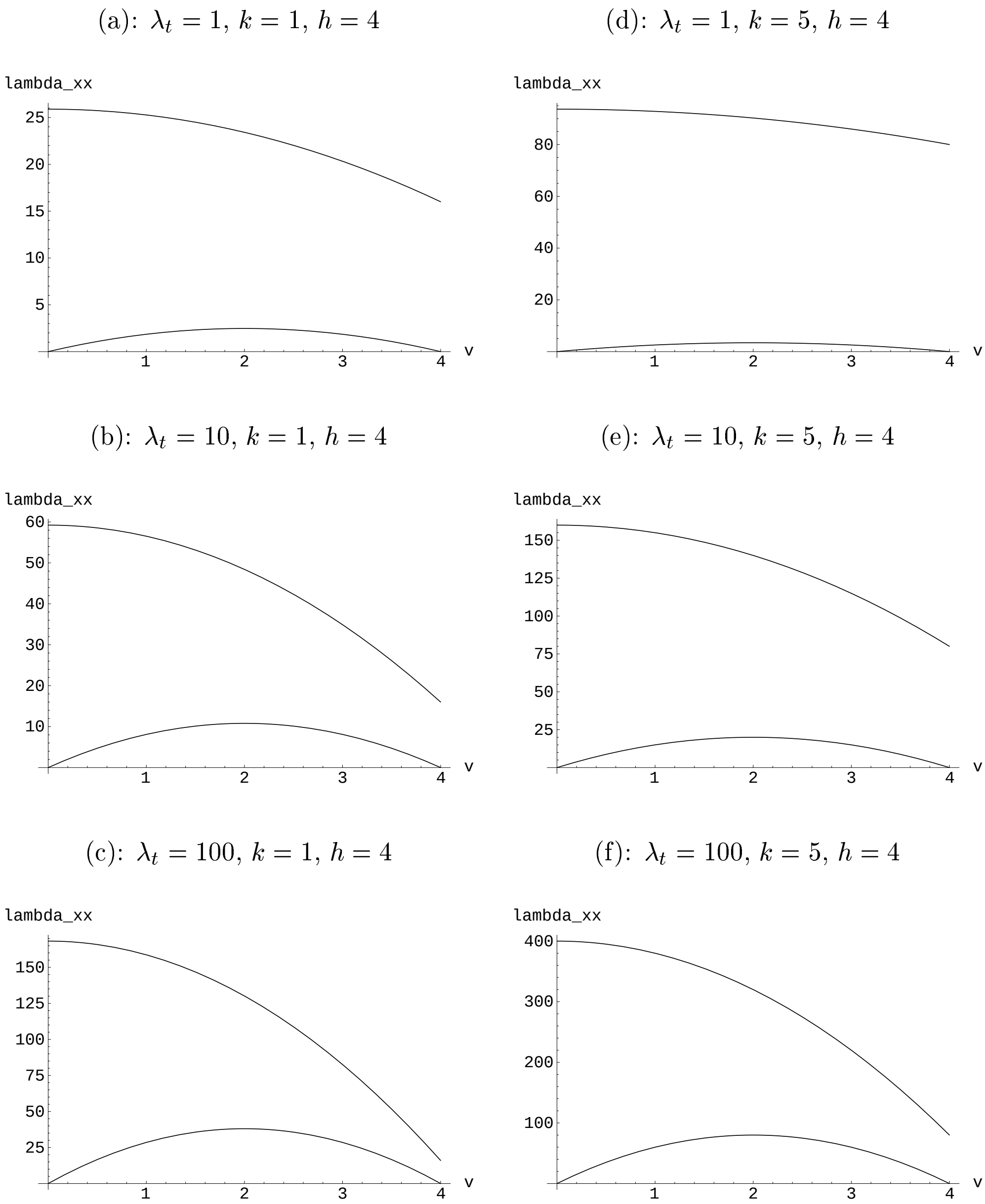

Figure 5: (left column) Graphs of the upper and the lower bounds on the spatial smoothing parameter λ_{xx} according to (52), in the case of spatial subsampling by a factor h and given a value of the temporal smoothing parameter λ_t, the image velocity v as well as the requirement of non-negative filter parameters in the spatio-temporal smoothing scheme. (right column) Corresponding bounds on λ_{xx} according to (52), when decomposing the smoothing operation into $k = 5$ smoothing steps coupled in cascade.

capture fine-scale objects at high velocities, unless the objects are being tracked and fixated by the camera(s) in an active vision system.

λ_t	$v = 0$	$v = 1/2$	$v = 1$	$v = 2$
1	$[0, 25.89]$	$[1.08, 25.75]$	$[1.85, 25.27]$	$[2.47, 23.42]$
10	$[0, 59.22]$	$[4.73, 58.55]$	$[8.10, 56.52]$	$[10.80, 48.42]$
100	$[0, 168.20]$	$[16.65, 165.82]$	$[28.53, 158.69]$	$[38.05, 130.15]$

Table 3: Ranges of permissable values of the amount of spatial smoothing λ_{xx} for *one layer* of a spatio-temporal recursive filter with the computations carried out *at resolution* $h = 4$, given specific values of the total amount of temporal smoothing λ_t relative to a unit grid spacing and the image velocity v.

λ_t	$v = 0$	$v = 1/2$	$v = 1$	$v = 2$
1	$[0, 93.67]$	$[1.49, 93.45]$	$[2.56, 92.81]$	$[3.41, 90.25]$
10	$[0, 160.00]$	$[8.75, 158.75]$	$[15.00, 155.00]$	$[20.00, 140.00]$
100	$[0, 400.00]$	$[35.00, 395.00]$	$[60.00, 380.00]$	$[80.00, 320.00]$

Table 4: Ranges of permissable values of the amount of spatial smoothing λ_{xx} for $k = 5$ *layers* of spatio-temporal recursive filters with the computations carried out *at resolution* $h = 4$, given specific values of the total amount of temporal smoothing λ_t relative to a unit grid spacing and the image velocity v.

3.6 Kernel graphs

Figure 6 shows a few examples of kernels computed in this way, by subjecting a discrete delta function to repeated non-separable recursive filtering and followed by the computation of velocity-adapted spatio-temporal derivatives according to

$$\partial_{\bar{x}} = \partial_x \qquad \partial_{\bar{t}} = v\,\partial_x + \partial_t \tag{53}$$

As can be seen, this spatio-temporal filtering scheme provides a way to express space-time oriented filter kernels of different orders and different orientations. Important properties are that (i) the spatio-temporal kernels are computed in a completely time recursive way, and (ii) without extended temporal buffering of the past. Such properties are required for any real-time system for processing spatio-temporal data.

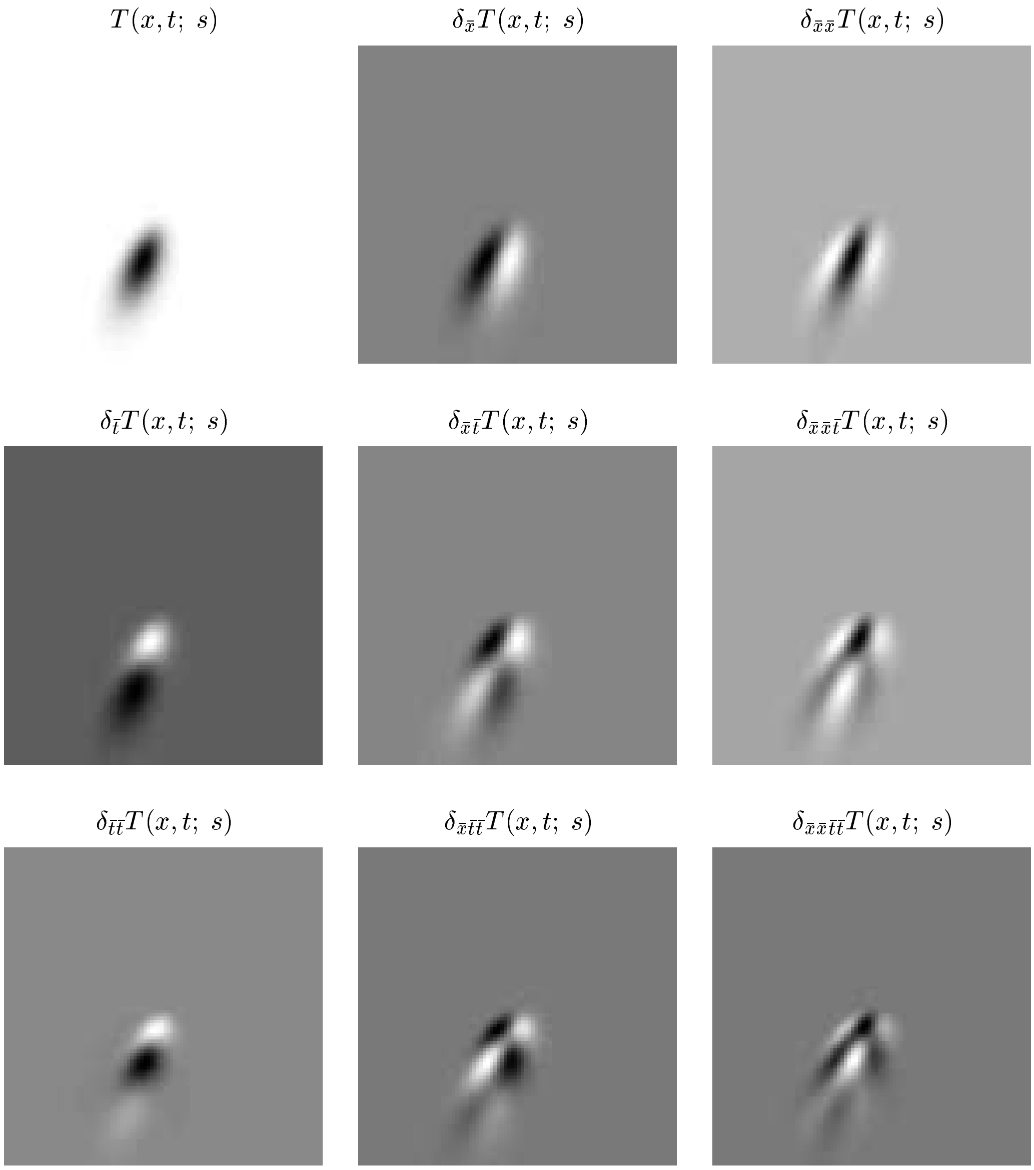

Figure 6: Equivalent filter kernels corresponding to the result of using a discrete delta function as input and applying repeated non-separable recursive filtering according to (31) with a Galilean transformation parametrization of the filter shape (38) followed by subsequent Galilean based computation of velocity-adapted spatio-temporal derivatives according to (53) for different orders of spatial and temporal differentiation. (In all cases, the spatio-temporal smoothing parameters are the same ($\lambda_{xx} = 16, \lambda_t = 64, v = 1/3$) as well as the number of layers $k = 10$. (Image size 150×150 pixels. Horisontal axis: space x. Vertical axis: time t.)

4 Time-recursive non-separable scale-space in 2+1-D

To extend the previously developed recursive velocity adaptation scheme from one to two spatial dimensions, let us consider a smoothing scheme of the form

$$f_{out}(x,y,t) = \frac{1}{1+\mu_t} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} f_{out}(x,t-1) + \frac{1}{1+\mu_t} \begin{pmatrix} j & k & l \\ m & n & p \\ q & r & s \end{pmatrix} f_{in}(x,t) \tag{54}$$

where the matrices within parentheses denote computational symbols in the spatial domain, corresponding to the following operations, written out explicitly

$$\begin{aligned} f_{out}(x,t) &= \\ &= \frac{1}{1+\mu_t}(a\, f_{out}(x-1,y+1,t-1) + b\, f_{out}(x,y+1,t-1) + c\, f_{out}(x+1,y+1,t-1) \\ &\qquad + d\, f_{out}(x-1,y,t-1) + e\, f_{out}(x,y,t-1) + f\, f_{out}(x+1,y,t-1) \\ &\qquad + g\, f_{out}(x-1,y-1,t-1) + h\, f_{out}(x,y-1,t-1) + i\, f_{out}(x+1,y-1,t-1) \\ &\qquad + j\, f_{in}(x-1,y+1,t) + k\, f_{in}(x,y+1,t) + l\, f_{in}(x+1,y+1,t) \\ &\qquad + m\, f_{in}(x-1,y,t) + n\, f_{in}(x,y,t) + p\, f_{in}(x+1,y,t) \\ &\qquad + q\, f_{in}(x-1,y-1,t) + r\, f_{in}(x,y-1,t) + s\, f_{in}(x+1,y-1,t)). \end{aligned}$$

From the generating function of the corresponding filter

$$\begin{aligned} \varphi(u,v,z) &= \\ &= \frac{pu^{-1} + mu + kv^{-1} + rv + lu^{-1}v^{-1} + su^{-1}v + juv^{-1} + quv}{1 + \mu_t - (fu + du^{-1} + bv + hv^{-1} + cu^{-1}v^{-1} + iu^{-1}v + auv^{-1} + guv)z} \end{aligned} \tag{55}$$

where u denotes the transformation variable in the spatial x-direction, v denotes the transformation variable in the spatial y-direction and z the corresponding transformation variable in the temporal domain, we obtain the mean vector M and the covariance matrix V as

$$M = \left. \begin{pmatrix} \varphi_u \\ \varphi_v \\ \varphi_z \end{pmatrix} \right|_{(u,v,z)=(1,1,1)} = \begin{pmatrix} \mu_x \\ \mu_y \\ \mu_t \end{pmatrix}, \tag{56}$$

$$V = \left. \begin{pmatrix} \varphi_{uu} + \varphi_u - \varphi_u^2 & \varphi_{uv} - \varphi_u\varphi_v & \varphi_{uz} - \varphi_u\varphi_z \\ \varphi_{uv} - \varphi_u\varphi_v & \varphi_{vv} - \varphi_v - \varphi_v^2 & \varphi_{vz} - \varphi_v\varphi_z \\ \varphi_{uz} - \varphi_u\varphi_z & \varphi_{vz} - \varphi_v\varphi_z & \varphi_{zz} + \varphi_z - \varphi_z^2 \end{pmatrix} \right|_{(u,v,z)=(1,1,1)} \tag{57}$$

$$= \begin{pmatrix} \mu_{xx} & \mu_{xy} & \mu_{xt} \\ \mu_{xy} & \mu_{yy} & \mu_{yt} \\ \mu_{xt} & \mu_{yt} & \mu_t^2 + \mu_t \end{pmatrix}. \tag{58}$$

Our next aim is to solve for the parameters a, b, $\dots s$ in the recursive filter in terms of the parameters μ_x, μ_y, μ_t, μ_{xx}, μ_{xy}, μ_{yy}, μ_{xt} and μ_{yt}, with the additional constraint $\varphi(1,1,1) = 1$. Structurally, this problem has 18 degrees of freedom in the parameters a, b, $\dots s$, one constraint due to normalization, three constraints in terms of the mean values and six constraints in terms of the covariance matrix, out of which one constraint is redundant due to the coupling between the mean and the variance in

$$\delta_x \, f_{in}(x,y,t) = \begin{pmatrix} 0 & 0 & 0 \\ -1/2 & 0 & +1/2 \\ 0 & 0 & 0 \end{pmatrix} f_{in}(x,y,t)$$

$$\delta_{xx} f_{out}(x,y,t-1) = \begin{pmatrix} 0 & 0 & 0 \\ +1 & -2 & +1 \\ 0 & 0 & 0 \end{pmatrix} f_{out}(x,y,t-1)$$

$$\delta_{xy} f_{out}(x,y,t-1) = \begin{pmatrix} -1/4 & 0 & +1/4 \\ 0 & 0 & 0 \\ +1/4 & 0 & -1/4 \end{pmatrix} f_{out}(x,y,t-1)$$

$$\delta_{xt}(f_{in}, f_{out})(x,y,t) = \begin{pmatrix} 0 & 0 & 0 \\ -1/2 & 0 & +1/2 \\ 0 & 0 & 0 \end{pmatrix} (f_{in}(x,y,t) - f_{out}(x,y,t-1))$$

$$\delta_{xxt}(f_{in}, f_{out})(x,y,t) = \begin{pmatrix} 0 & 0 & 0 \\ +1 & -2 & +1 \\ 0 & 0 & 0 \end{pmatrix} (f_{in}(x,y,t) - f_{out}(x,y,t-1))$$

$$\delta_{xyt}(f_{in}, f_{out})(x,y,t) = \begin{pmatrix} -1/4 & 0 & +1/4 \\ 0 & 0 & 0 \\ +1/4 & 0 & -1/4 \end{pmatrix} (f_{in}(x,y,t) - f_{out}(x,y,t-1))$$

$$\delta_{xyy}(f_{out})(x,y,t) = \begin{pmatrix} -1/2 & 0 & +1/2 \\ +1 & 0 & -1 \\ -1/2 & 0 & +1/2 \end{pmatrix} f_{out}(x,y,t)$$

$$\delta_{xyyt}(f_{in}, f_{out})(x,y,t) = \begin{pmatrix} -1/2 & 0 & +1/2 \\ +1 & 0 & -1 \\ -1/2 & 0 & +1/2 \end{pmatrix} (f_{in}(x,y,t) - f_{out}(x,y,t-1))$$

$$\delta_{xxyy}(f_{out})(x,y,t) = \begin{pmatrix} 1 & -2 & 1 \\ -2 & +4 & -2 \\ 1 & -2 & 1 \end{pmatrix} f_{out}(x,y,t)$$

$$\delta_{xxyyt}(f_{in}, f_{out})(x,y,t) = \begin{pmatrix} +1 & -2 & +1 \\ -2 & +4 & -2 \\ +1 & -2 & +1 \end{pmatrix} (f_{in}(x,y,t) - f_{out}(x,y,t-1))$$

Figure 7: Difference operators used in the construction of the discrete recursive spatio-temporal scale-space representation in 2+1-D space-time (65). To reduce redundancy, operators that are mere rotations of corresponding operators in other directions are not shown. Thus, δ_y, δ_{yy}, δ_{yt}, δ_{yyt}, δ_{xxy} and δ_{xxyt} have been suppressed.

the temporal direction. Thus, we expect to problem to have $18 - 1 - 3 - (6 - 1) = 9$ degrees of freedom, Due to the dimensionality of the problem, however, these degrees of freedom may require a certain effort to handle by a pure bottom-up approach. Therefore, let us instead tackle the problem by the following *ansatz* (which is inspired by the form of the solution (29) in the 1+1-dimensional case as well as from the fact that the higher order difference operators δ_{xxt}, δ_{yyt} $\ldots \delta_{xxyyt}$ do not affect the mean and the variance of the smoothing filter):

$$\begin{aligned} f_{out}(x,y,t) = \frac{1}{1+\mu_t}\Big(& - \mu_x\, \delta_x\, f_{in}(x,y,t) - \mu_y\, \delta_y\, f_{in}(x,y,t) \\ & + f_{in}(x,y,t) + \mu_t\, f_{out}(x,y,t-1) \\ & + \tfrac{1}{2}\, D_{xx}\, \delta_{xx} f_{out}(x,y,t-1) \\ & + \tfrac{2}{2}\, D_{xy}\, \delta_{xy} f_{out}(x,y,t-1) \\ & + \tfrac{1}{2}\, D_{yy}\, \delta_{yy} f_{out}(x,y,t-1) \\ & + \tfrac{2}{2}\, D_{xt} \delta_{xt}(f_{in}, f_{out})(x,y,t) \\ & + \tfrac{2}{2}\, D_{yt} \delta_{yt}(f_{in}, f_{out})(x,y,t) \\ & + \tfrac{3}{6}\, \mu_{xxt}\, \delta_{xxt}(f_{in}, f_{out})(x,y,t) \\ & + \tfrac{6}{6}\, \mu_{xyt}\, \delta_{xyt}(f_{in}, f_{out})(x,y,t) \\ & + \tfrac{3}{6}\, \mu_{yyt}\, \delta_{yyt}(f_{in}, f_{out})(x,y,t) \\ & + \tfrac{3}{6}\, \mu_{xxy}\, \delta_{xxy}(f_{out})(x,y,t) \\ & + \tfrac{3}{6}\, \mu_{xyy}\, \delta_{xyy}(f_{out})(x,y,t) \\ & + \tfrac{12}{24}\, \mu_{xxyt}\, \delta_{xxyt}(f_{out}, f_{out})(x,y,t) \\ & + \tfrac{12}{24}\, \mu_{xyyt}\, \delta_{xyyt}(f_{out}, f_{out})(x,y,t) \\ & + \tfrac{6}{24}\, \mu_{xxyy}\, \delta_{xxyy}(f_{out})(x,y,t) \\ & + \tfrac{30}{120}\, \mu_{xxyyt}\, \delta_{xxyyt}(f_{out}, f_{out})(x,y,t)\Big) \end{aligned} \tag{59}$$

with the explicit expressions for the difference operators given in figure 7. By solving for the essential parameters D_{xx}, D_{xy}, D_{yy}, D_{xt} and D_{yt} in terms of μ_x, μ_y, μ_{xx}, μ_{xy}, μ_{yy}, μ_{xt} and μ_{yt}, it can after some calculations (performed in Mathematica) be shown that

$$D_{xx} = \mu_{xx} + \mu_x^2 - \frac{2\mu_x\mu_{xt}}{1+\mu_t} \tag{60}$$

$$D_{xy} = \mu_{xy} + \mu_x\mu_y - \frac{\mu_y\mu_{xt} + \mu_x\mu_{yt}}{1+\mu_t} \tag{61}$$

$$D_{yy} = \mu_{yy} + \mu_y^2 - \frac{2\mu_y\mu_{yt}}{1+\mu_t} \tag{62}$$

$$D_{xt} = \frac{\mu_{xt}}{1+\mu_t} \tag{63}$$

$$D_{yt} = \frac{\mu_{yt}}{1+\mu_t} \tag{64}$$

while the values of μ_{xxt}, μ_{xyt}, μ_{yyt}, μ_{xxy}, μ_{xyy}, μ_{xxyt}, μ_{xyyt}, μ_{xxyy} and μ_{xxyyt} are formally not contrained as long as they correspond to non-negative values of a, b, $\ldots$,

s. On incremental form, this non-separable recursive smoothing scheme can written

$$
\begin{aligned}
f_{out}(x,y,t) - f_{out}(x,y,t-1) = \\
= \frac{1}{1+\mu_t} (D - \mu_x \, \delta_x \, f_{in}(x,y,t) - \mu_y \, \delta_y \, f_{in}(x,y,t) \\
+ \frac{1}{2} \left(\mu_{xx} + \mu_x^2 - \frac{2\mu_x \mu_{xt}}{1+\mu_t} \right) \delta_{xx} f_{out}(x,y,t-1) \\
+ \left(\mu_{xy} + \mu_x \mu_y - \frac{\mu_y \mu_{xt} + \mu_x \mu_{yt}}{1+\mu_t} \right) \delta_{xy} f_{out}(x,y,t-1) \\
+ \frac{1}{2} \left(\mu_{yy} + \mu_y^2 - \frac{2\mu_y \mu_{yt}}{1+\mu_t} \right) \delta_{xy} f_{out}(x,y,t-1) \\
+ \frac{\mu_{xt}}{1+\mu_t} \, \delta_{xt}(f_{in}, f_{out})(x,y,t) \\
+ \frac{\mu_{yt}}{1+\mu_t} \delta_{yt}(f_{in}, f_{out})(x,y,t))
\end{aligned}
\tag{65}
$$

where

$$
\begin{aligned}
D = & f_{in}(x,y,t) - f_{out}(x,y,t-1) \\
& + \tfrac{3}{6} \, \mu_{xxt} \, \delta_{xxt}(f_{in}, f_{out})(x,y,t) \\
& + \tfrac{6}{6} \, \mu_{xyt} \, \delta_{xyt}(f_{in}, f_{out})(x,y,t) \\
& + \tfrac{3}{6} \, \mu_{yyt} \, \delta_{yyt}(f_{in}, f_{out})(x,y,t) \\
& + \tfrac{3}{6} \, \mu_{xxy} \, \delta_{xxy}(f_{out})(x,y,t) + \tfrac{3}{6} \, \mu_{xyy} \, \delta_{xyy}(f_{out})(x,y,t) \\
& + \tfrac{12}{24} \, \mu_{xxyt} \, \delta_{xxyt}(f_{out}, f_{out})(x,y,t) + \tfrac{12}{24} \, \mu_{xyyt} \, \delta_{xyyt}(f_{out}, f_{out})(x,y,t) \\
& + \tfrac{6}{24} \, \mu_{xxyy} \, \delta_{xxyy}(f_{out})(x,y,t) + \tfrac{30}{120} \, \mu_{xxyyt} \, \delta_{xxyyt}(f_{out}, f_{out})(x,y,t)
\end{aligned}
\tag{66}
$$

In a similar way as for the recursive spatio-temporal scale-space for one spatial dimension, the degrees of freedom in this expression can for sufficiently small values of μ_{xxt}, μ_{xyt}, μ_{yyt} μ_{xxy}, μ_{xyy}, μ_{xxyt}, μ_{xyyt}, μ_{xxyy} and μ_{xxyyt} be interpreted as a mixed smoothing effect on the input signal $f_{in}(\cdot,\cdot,t)$ in combination with the output signal $f_{out}(\cdot,\cdot,t-1)$ at the previous time moment. An important constraint, however, is that the filter parameters should correspond to non-negative values of a, b, $\ldots$, s in (54). A more detailed analysis of this problem is given in appendix A.1.

4.1 Parameterization of filter shapes

If we couple K such spatio-temporal recursive filters in cascade, each filter with filter parameters $\mu_{xx}^{(i)}$, $\mu_{xy}^{(i)}$, $\mu_{yy}^{(i)}$, $\mu_t^{(i)}$, $\mu_{xt}^{(i)}$, $\mu_{yt}^{(i)}$, $\mu_x^{(i)}$ and $\mu_y^{(i)}$, then it follows from the additive property of covariance matrices and mean vectors under convolution that the covariance matrix Σ' and the mean vector M' of the composed filter will be given by

$$
\Sigma' = \begin{pmatrix} C'_{xx} & C'_{xy} & C'_{xt} \\ C'_{xy} & C'_{yy} & C'_{yt} \\ C'_{xt} & C'_{yt} & C'_{tt} \end{pmatrix} = \sum_{i=1}^{k} \begin{pmatrix} \mu_{xx}^{(i)} & \mu_{xy}^{(i)} & \mu_{xt}^{(i)} \\ \mu_{xy}^{(i)} & \mu_{yy}^{(i)} & \mu_{yt}^{(i)} \\ \mu_{xt}^{(i)} & \mu_{yt}^{(i)} & (\mu_t^{(i)})^2 + \mu_t^{(i)} \end{pmatrix}
\tag{67}
$$

and

$$
M' = \begin{pmatrix} C'_x \\ C'_y \\ C'_t \end{pmatrix} = \sum_{i=1}^{k} \begin{pmatrix} \mu_x^{(i)} \\ \mu_y^{(i)} \\ \mu_t^{(i)} \end{pmatrix}
\tag{68}
$$

To parameterize these filter shapes, let us first express the spatial part of the covariance matrix in terms of two eigenvalues (λ_1, λ_2) and one orientation α

$$\Sigma = \begin{pmatrix} C_{xx} & C_{xt} & C_{xt} \\ C_{xy} & C_{yy} & C_{yt} \\ C_{xt} & C_{yt} & C_t \end{pmatrix} = \begin{pmatrix} \lambda_1 \cos^2\alpha + \lambda_2 \sin^2\alpha & (\lambda_2 - \lambda_1)\cos\alpha\,\sin\alpha & 0 \\ (\lambda_2 - \lambda_1)\cos\alpha\,\sin\alpha & \lambda_1 \sin^2\alpha + \lambda_2 \cos^2\alpha & 0 \\ 0 & 0 & \lambda_t \end{pmatrix} \tag{69}$$

Then, by subjecting this matrix to a Galilei transformation (7), we obtain

$$\Sigma' = \begin{pmatrix} C_{xx} & C_{xt} & C_{xt} \\ C_{xy} & C_{yy} & C_{yt} \\ C_{xt} & C_{yt} & C_t \end{pmatrix}$$
$$= \begin{pmatrix} \lambda_1 \cos^2\alpha + \lambda_2 \sin^2\alpha + v_x^2\lambda_t & (\lambda_2 - \lambda_1)\cos\alpha\,\sin\alpha + v_x v_y \lambda_t & v_x\lambda_t \\ (\lambda_2 - \lambda_1)\cos\alpha\,\sin\alpha + v_x v_y\lambda_t & \lambda_1 \sin^2\alpha + \lambda_2 \cos^2\alpha + v_y^2\lambda_t & v_y\lambda_t \\ v_x\lambda_t & v_y\lambda_t & \lambda_t \end{pmatrix} \tag{70}$$

with the mean vector given by

$$M' = \begin{pmatrix} C'_x \\ C'_y \\ C'_t \end{pmatrix} = C_t \begin{pmatrix} v_x \\ v_y \\ 1 \end{pmatrix} \tag{71}$$

These expressions thus provide a natural way to express filters with different shapes and orientations in space-time. Velocity-adapted spatio-temporal derivatives can finally be obtained from the output of such filters according to

$$\partial_{\bar{x}} = \partial_x \qquad \partial_{\bar{y}} = \partial_y \qquad \partial_{\bar{t}} = v_x\,\partial_x + v_y\,\partial_y + \partial_t \tag{72}$$

4.2 Kernel graphs

Figures 8–10 show a few examples of spatio-temporal scale-space kernels generated in this way. Figure 8 shows filters corresponding to $v = 0$, while figures 9–10 show corresponding velocity adapted (and non-separable) recursive filters for the non-zero velocities $v = 0.2$ and $v = 0.8$.[3] While we have here generated explicit filter shapes for the purpose of graphical illustration, the practical approach of implementing this scale-space concept is, of course, by first performing the spatio-temporal smoothing operation once and for all using (54), alternatively (65) and (66), and then computing derivative approximations from difference operators in space and time, which are combined according to (72) at every image point.

[3]For these filters, the choice of higher order filter parameters has been determined numerically, based on the procedure outlined in appendix A.1. For graphical illustration, the results are shown as level surfaces, where the colour of the level surface indicates the polarity.

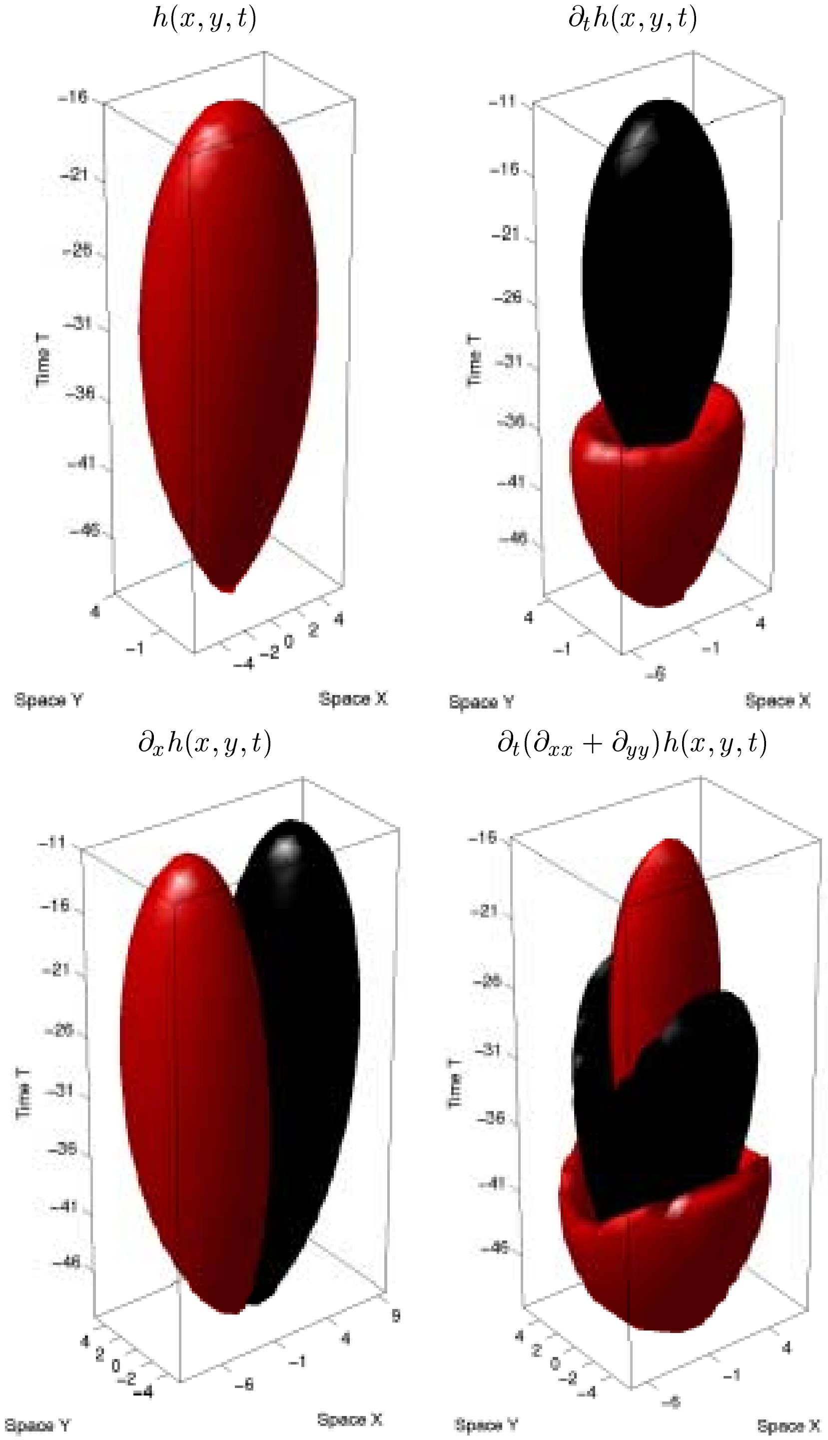

Figure 8: Level surfaces of spatio-temporal receptive fields for the 2+1-D separable recursive spatio-temporal scale-space. (a) The raw smoothing kernel $h(x, y, t)$. (b) First-order temporal derivative $\partial_t h(x, y, t)$. (c) First-order spatial derivative $\partial_x h(x, y, t)$. (d) First-order temporal derivative of Laplacian response $\partial_t(\partial_{xx}+\partial_{yy})h(x, y, t)$. (In all cases, the smoothing parameters have been $\lambda_{xx} = 2$, $\lambda_{yy} = 1$, $\lambda_t = 4$, $v = 0$ and five identical recursive filters with these filter parameters have been coupled in cascade.)

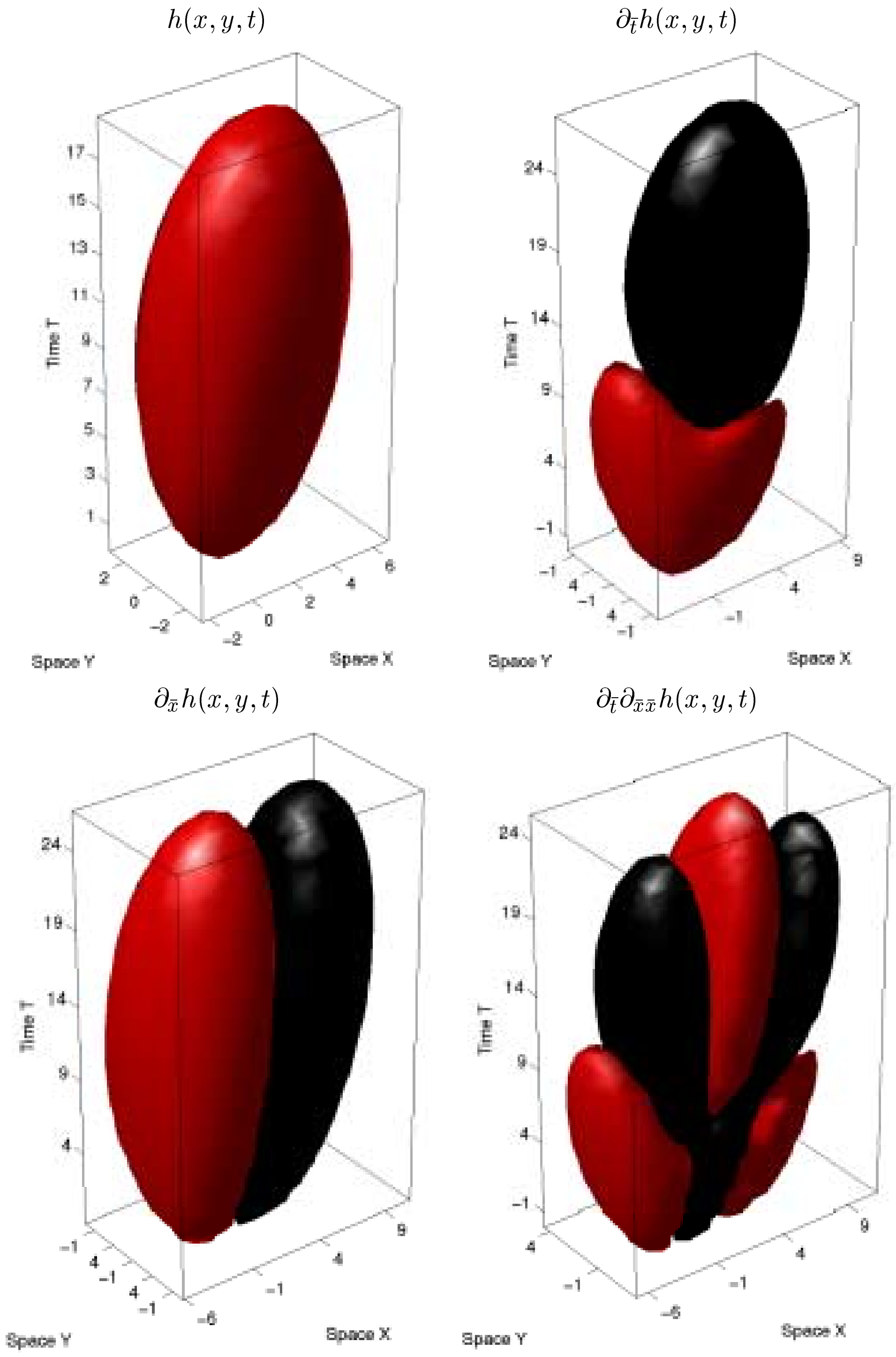

Figure 9: Level surfaces of spatio-temporal receptive fields for the 2+1-D velocity-adapted spatio-temporal scale-space. (a) The raw smoothing kernel $h(x, y, t)$. (b) First-order temporal derivative $\partial_{\bar{t}} h(x, y, t)$. (c) First-order spatial derivative $\partial_{\bar{x}} h(x, y, t)$. (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 $\lambda_{xx} = 2$, $\lambda_{yy} = 1$, $\lambda_t = 4$, $(v_x, v_y) = (0.2, 0.0)$ and five identical filters with these filter parameters have been coupled in cascade.)

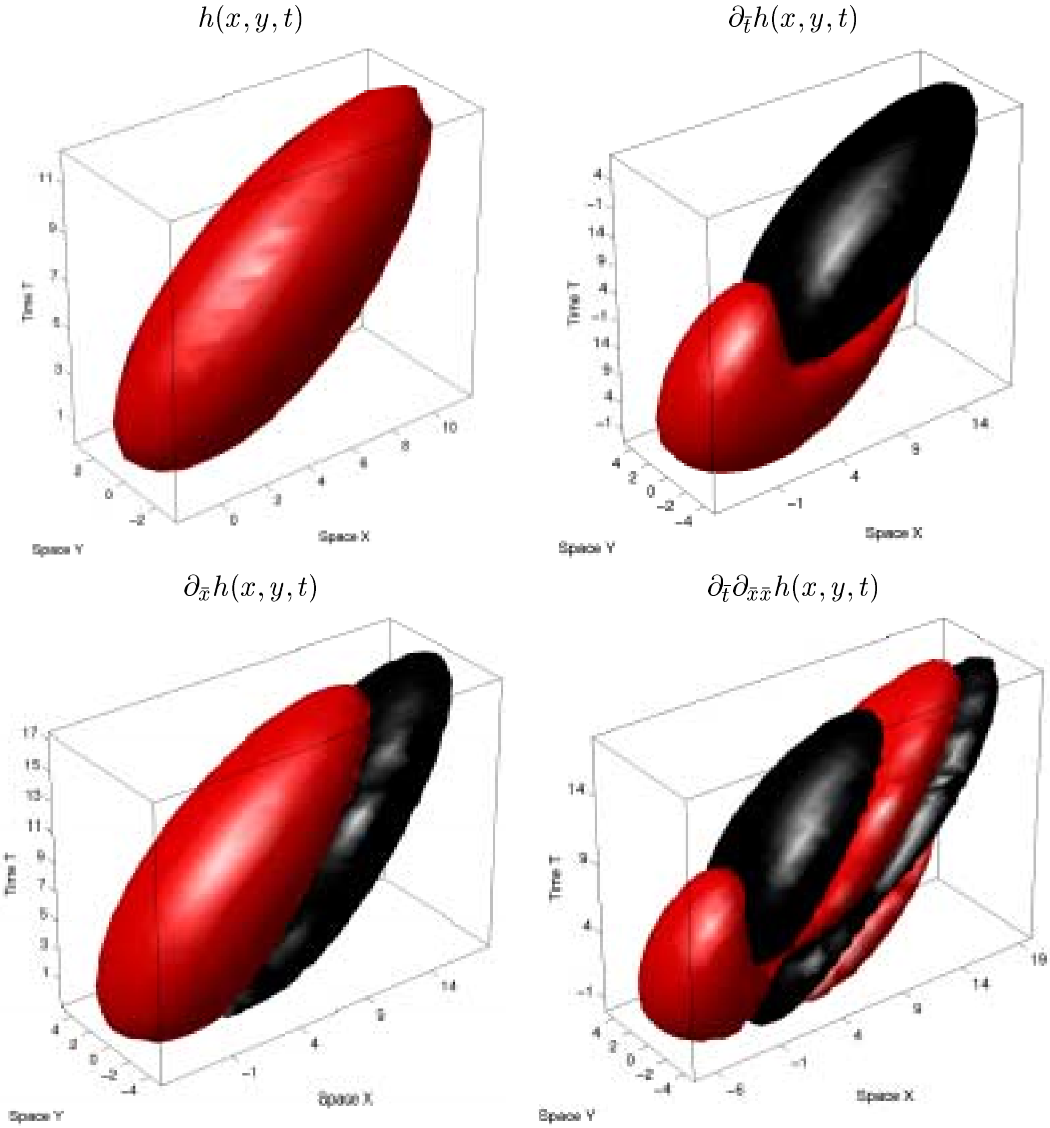

Figure 10: Level surfaces of spatio-temporal receptive fields for the 2+1-D velocity-adapted spatio-temporal scale-space. (a) The raw smoothing kernel $h(x, y, t)$. (b) First-order temporal derivative $\partial_{\bar{t}} h(x, y, t)$. (c) First-order spatial derivative $\partial_{\bar{x}} h(x, y, t)$. (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 $\lambda_{xx} = 5.0$, $\lambda_{yy} = 2.5$, $\lambda_t = 10$, $(v_x, v_y) = (0.8, 0.0)$ and this filtering operation has been divided into ten identical filters coupled in cascade.)

5 Summary and discussion

We have presented a theory for and an analysis of non-separable spatio-temporal scale-space kernels, based on time-causal recursive filtering. These kernels provide a way to perform spatio-temporal filtering with continuous scale and velocity parameters, in such a way that the spatio-temporal covariance matrices of the discrete filters transform in an algebraically similar way as for continuous filters under Galilean motion.

From the requirement of positivity of the filter coefficients, bounds on the permissable image velocities have been derived in terms of the spatial resolution. Basically, these bounds imply that larger image velocities can be can be captured at coarser spatial scales. If a larger amount of smoothing is required at fine spatial scales, however, then either an external warping mechanism, a denser temporal sampling or a connectivity pattern that extends outside the range of nearest spatial neighbours is necessary.

We propose that the general notion of velocity-adapted filtering should be considered as an important mechanism whenever a computer vision system aims at computing spatio-temporal receptive field responses of time-dependent image data under the constraint that the image descriptors are to be invariant under Galilean motion in the image plane. One example of the benefit of such a mechanism is presented in (Laptev & Lindeberg 2002), where it is shown how the incorporation of velocity adaptation improves the performance of spatio-temporal recognition schemes, compared to a more traditional approach of using separable spatio-temporal filters only.

Notably, these receptive field profiles have interesting qualitative similarities 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 1987); see (Lindeberg 2001) for a comparison. There are also interesting relations to methods for optic flow estimation from spatio-temporal filters (Adelson & Bergen 1985, Heeger 1988), steerable filters (Freeman & Adelson 1991) and models of biological receptive fields (Watson & Ahumada 1985, Simoncelli & Heeger 1998).

A Appendix

A.1 Exploring the positivity constraints in 2+1-D

An important constraint on the time-recursive spatio-temporal scale-space for 2+1-dimensional signals according to (65) and (66) is that given specific values of the primary filter parameters μ_x, μ_y, μ_t, μ_{xx}, μ_{xy}, μ_{yy}, μ_{xt} and μ_{yt} for one layer of a set of cascade coupled filters, as set from (67), (68), (70) and (71), the secondary filter parameters μ_{xxt}, μ_{xyt}, μ_{yyt}, μ_{xxy}, μ_{xyy}, μ_{xxyt}, μ_{xyyt}, μ_{xxyy} and μ_{xxyyt} must be determined such that they correspond to non-negative values of a, b, ..., s in the recursive smoothing scheme (54). To construct a numerical procedure for computing suitable values of the secondary parameters μ_{xxt}, μ_{xyt}, μ_{yyt}, μ_{xxy}, μ_{xyy}, μ_{xxyt}, μ_{xyyt}, μ_{xxyy} and μ_{xxyyt}, let us first state the explicit relationships between the filter parameters. By combining (54) with (59) and the computational molecules in figure 7, we obtain the relations between the filter parameters a, b, $\dots s$ and the filter parameters μ_x, μ_y, $\dots \mu_{xxyyt}$ shown in figure 11, where we for simplicity have also introduced the descriptors D_{xx}, D_{xy}, D_{yy}, D_{xt} and D_{yt}, which are related to μ_x, μ_y, μ_t, μ_{xx}, μ_{xy}, μ_{yy}, μ_{xt} and μ_{yt} according to (60)–(64).

By inspection of the expressions in figure 11, it is clear that the problem of finding one combination of parameters μ_{xxt}, μ_{xyt}, μ_{yyt}, μ_{xxy}, μ_{xyy}, μ_{xxyt}, μ_{xyyt}, μ_{xxyy} and μ_{xxyyt} that correspond to non-negative values of a, b, $\dots s$ given specific values of D_{xx}, D_{xyt} $\dots D_{yt}$ is of a similar type as the problem of finding a feasible solution to a linear programming problem given a set of linear inequalities. Hence, if we are just interested in finding any set of values μ_{xxt}, μ_{xyt}, μ_{yyt}, μ_{xxy}, μ_{xyy}, μ_{xxyt}, μ_{xyyt}, μ_{xxyy} and μ_{xxyyt} that lead to non-negative a, b, $\dots s$, we can use the same methodology as in the first phase of a linear programming problem (Press et al. 1992). Next, if we want to define a unique central point among all the possible feasible points, we can make use of the fact that the inequalities in figure 11 delimit a convex polyhedron in the nine-dimensional space spanned by μ_{xxt}, μ_{xyt}, μ_{yyt}, μ_{xxy}, μ_{xyy}, μ_{xxyt}, μ_{xyyt}, μ_{xxyy} and μ_{xxyyt}, and minimimize and maximize each one of the free secondary parameters given the sign constraints on a, b, $\dots s$:[4]

$$\begin{aligned} p_{i_{min}} &= \text{argmin}_{a\geq 0, b\geq 0, \dots, s\geq 0}\, e_i^T (\mu_{xxt}, \mu_{xyt}, \mu_{yyt}, \mu_{xxy}, \mu_{xyy}, \mu_{xxyt}, \mu_{xyyt}, \mu_{xxyy}, \mu_{xxyyt}) \\ p_{i_{max}} &= \text{argmax}_{a\geq 0, b\geq 0, \dots, s\geq 0}\, e_i^T (\mu_{xxt}, \mu_{xyt}, \mu_{yyt}, \mu_{xxy}, \mu_{xyy}, \mu_{xxyt}, \mu_{xyyt}, \mu_{xxyy}, \mu_{xxyyt}) \end{aligned}$$

(see figure 12 for more explicit expressions of these linear programming problems) then choosing the center of gravity of all the feasible solutions obtained in this way

$$p = \frac{1}{9} \sum_{i=1}^{9} \frac{p_{i_{min}} + p_{i_{max}}}{2} \tag{73}$$

Alternatively, we can set the filter parameters from the center of gravity, the analytical center (Sonnevend 1986) or the volumetric center (Vaidya 1989) of the interior of this convex nine-dimensional polyhedron, or optimize the higher-order filter moments.

Of course, it may not always be a combination of μ_{xxt}, μ_{xyt}, μ_{yyt}, μ_{xxy}, μ_{xyy}, μ_{xxyt}, μ_{xyyt}, μ_{xxyy} and μ_{xxyyt} that corresponds to non-negative a, b $\dots s$. If the reason for this is that the amount of spatial smoothing is much larger than the amount of temporal smoothing, the easiest solution is of course to add an explicit spatial

[4] Here, e_i is a unit vector with the ith component equal to one and all the other entries are zero.

$$
\begin{aligned}
a &= \frac{1}{4}\,(-D_{xy} + \mu_{xxy} + \mu_{xxyt} + \mu_{xxyy} + \mu_{xxyyt} + \mu_{xyt} - \mu_{xyy} - \mu_{xyyt}) \\
b &= \frac{1}{2}\,(D_{yy} - D_{yt} - \mu_{xxy} - \mu_{xxyt} - \mu_{xxyy} - \mu_{xxyyt} - \mu_{yyt}) \\
c &= \frac{1}{4}\,(+D_{xy} + \mu_{xxy} + \mu_{xxyt} + \mu_{xxyy} + \mu_{xxyyt} - \mu_{xyt} + \mu_{xyy} + \mu_{xyyt}) \\
d &= \frac{1}{2}\,(D_{xx} + D_{xt} - \mu_{xxt} - \mu_{xxyy} - \mu_{xxyyt} + \mu_{xyy} + \mu_{xyyt}) \\
e &= \mu_t - D_{xx} - D_{yy} + \mu_{xxt} + \mu_{xxyy} + \mu_{xxyyt} + \mu_{yyt} \\
f &= \frac{1}{2}\,(D_{xx} - D_{xt} - \mu_{xxt} - \mu_{xxyy} - \mu_{xxyyt} - \mu_{xyy} - \mu_{xyyt}) \\
g &= \frac{1}{4}\,(D_{xy} - \mu_{xxy} - \mu_{xxyt} + \mu_{xxyy} + \mu_{xxyyt} - \mu_{xyt} - \mu_{xyy} - \mu_{xyyt}) \\
h &= \frac{1}{2}\,(D_{yy} + D_{yt} + \mu_{xxy} + \mu_{xxyt} - \mu_{xxyy} - \mu_{xxyyt} - \mu_{yyt}) \\
i &= \frac{1}{4}\,(-D_{xy} - \mu_{xxy} - \mu_{xxyt} + \mu_{xxyy} + \mu_{xxyyt} + \mu_{xyt} + \mu_{xyy} + \mu_{xyyt})
\end{aligned}
$$

$$
\begin{aligned}
j &= \frac{1}{4}\,(-\mu_{xxyt} - \mu_{xxyyt} - \mu_{xyt} + \mu_{xyyt}) \\
k &= \frac{1}{2}\,(-\mu_y + D_{yt} + \mu_{xxyt} + \mu_{xxyyt} + \mu_{yyt}) \\
l &= \frac{1}{4}\,(-\mu_{xxyt} - \mu_{xxyyt} + \mu_{xyt} - \mu_{xyyt}) \\
m &= \frac{1}{2}\,(\mu_x - D_{xt} + \mu_{xxt} + \mu_{xxyyt} - \mu_{xyyt}) \\
n &= 1 - \mu_{xxt} - \mu_{xxyyt} - \mu_{yyt} \\
p &= \frac{1}{2}\,(-\mu_x + D_{xt} + \mu_{xxt} + \mu_{xxyyt} + \mu_{xyyt}) \\
q &= \frac{1}{4}\,(\mu_{xxyt} - \mu_{xxyyt} + \mu_{xyt} + \mu_{xyyt}) \\
r &= \frac{1}{2}\,(\mu_y - D_{yt} - \mu_{xxyt} + \mu_{xxyyt} + \mu_{yyt}) \\
s &= \frac{1}{4}\,(\mu_{xxyt} - \mu_{xxyyt} - \mu_{xyt} - \mu_{xyyt})
\end{aligned}
$$

Figure 11: Relations between the filter parameters a, b, $\ldots s$ in the explicit form of the recursive spatio-temporal scale-space (54) and the filter parameters μ_x, μ_y, $\ldots \mu_{xxyyt}$ with their associated descriptors D_{xx}, D_{xyt}, $\ldots D_{yt}$ according to (60)–(64) in the incremental diffusion type formulation (65) and (66) of the time-recursive spatio-temporal scale-space.

$$\min / \max\{\mu_{xxt}, \mu_{xyt}, \mu_{yyt}, \mu_{xxy}, \mu_{xyy}, \mu_{xxyt}, \mu_{xyyt}, \mu_{xxyy}, \mu_{xxyyt}\}$$

subject to

$$\begin{pmatrix} 0 & -1 & 0 & -1 & +1 & -1 & +1 & -1 & -1 \\ 0 & 0 & +1 & +1 & 0 & +1 & 0 & +1 & +1 \\ 0 & +1 & 0 & -1 & -1 & -1 & -1 & -1 & -1 \\ +1 & 0 & 0 & 0 & -1 & 0 & -1 & +1 & +1 \\ -1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & -1 \\ +1 & 0 & 0 & 0 & +1 & 0 & +1 & +1 & +1 \\ 0 & +1 & 0 & +1 & +1 & +1 & +1 & -1 & -1 \\ 0 & 0 & +1 & -1 & 0 & -1 & 0 & +1 & +1 \\ 0 & -1 & 0 & +1 & -1 & +1 & -1 & -1 & -1 \\ 0 & +1 & 0 & 0 & 0 & +1 & -1 & 0 & +1 \\ 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & -1 \\ 0 & -1 & 0 & 0 & 0 & +1 & +1 & 0 & +1 \\ -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & -1 \\ +1 & 0 & +1 & 0 & 0 & 0 & 0 & 0 & +1 \\ -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 \\ 0 & -1 & 0 & 0 & 0 & -1 & -1 & 0 & +1 \\ 0 & 0 & -1 & 0 & 0 & +1 & 0 & 0 & -1 \\ 0 & +1 & 0 & 0 & 0 & -1 & +1 & 0 & +1 \end{pmatrix} \begin{pmatrix} \mu_{xxt} \\ \mu_{xyt} \\ \mu_{yyt} \\ \mu_{xxy} \\ \mu_{xyy} \\ \mu_{xxyt} \\ \mu_{xyyt} \\ \mu_{xxyy} \\ \mu_{xxyyt} \end{pmatrix} \leq \begin{pmatrix} -D_{xy} \\ D_{yy} - D_{yt} \\ D_{xy} \\ D_{xx} + D_{xt} \\ \mu_t - D_{xx} - D_{yy} \\ D_{xx} - D_{xt} \\ D_{xy} \\ D_{yy} + D_{yt} \\ -D_{xy} \\ 0 \\ -\mu_y + D_{yt} \\ 0 \\ \mu_x - D_{xt} \\ 1 \\ -\mu_x + D_{xt} \\ 0 \\ \mu_y - D_{yt} \\ 0 \end{pmatrix}$$

Figure 12: Explicit expressions for the 18 linear programming problems used for setting the secondary filter parameters μ_{xxt}, μ_{xyt}, μ_{yyt}, μ_{xxy}, μ_{xyy}, μ_{xxyt}, μ_{xyyt}, μ_{xxyy} and μ_{xxyyt} according to (73) given specific values of the primary filter parameters μ_x, μ_y, μ_t, μ_{xx}, μ_{xy}, μ_{yy}, μ_{xt} and μ_{yt} (here expressed in terms of D_{xx}, D_{xy}, D_{yy}, D_{xt} and D_{yt} according to (60)–(64)). Numerically, these solutions are easy to compute with standard linear programming software.

smoothing step. Otherwise, the main alternatives are basically to: (i) decompose the smoothing operation into a larger number of layers, (ii) complement the smoothing operation by an external warping mechanism to handle the dominant component of the motion, or (iii) aim at spatio-temporal smoothing that corresponds to a lower image velocity.

References

Adelson, E. & Bergen, J. (1985), 'Spatiotemporal energy models for the perception of motion', *J. of the Optical Society of America* **A 2**, 284–299.

Almansa, A. & Lindeberg, T. (2000), 'Fingerprint enhancement by shape adaptation of scale-space operators with automatic scale-selection', *IEEE Transactions on Image Processing* **9**(12), 2027–2042.

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.

Black, M. (1994), Recursive non-linear estimation of discontinuous flow fields, *in* J.-O. Eklundh, ed., 'Proc. 3rd European Conf. on Computer Vision', Vol. 800 of *Lecture Notes in Computer Science*, Springer-Verlag, Stockholm, Sweden, pp. A:138–145.

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., 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.

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

Heeger, D. (1988), 'Optical flow using spatiotemporal filters', *Int. J. of Computer Vision* **1**, 279–302.

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

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

Laptev, I. & Lindeberg, T. (2002), Velocity-adaptation of spatio-temporal receptive fields for direct recognition of activities: An experimental study, report ISRN KTH/NA/P--02/04--SE, Dept. of Numerical Analysis and Computing Science, KTH, Stockholm, Sweden. Also in Proc. ECCV'02 Workshop on Statistical Methods in Video Processing, (to appear).

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. (1997), 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. (2001), Linear spatio-temporal scale-space, report, ISRN KTH/NA/P--01/22--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.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992), *Numerical Recipes in C*, Cambridge University Press, Cambridge, U.K.

Schaffalitzky, F. & Zisserman, A. (2001), Viewpoint invariant texture matching and wide baseline stereo, *in* 'Proc. 8th Int. Conf. on Computer Vision', Vancouver, Canada, pp. II:636–643.

Simoncelli, E. & Heeger, D. (1998), 'A model of neuronal responses in visual area MT', *Vision Research* **38**(5).

Sonnevend, G. (1986), An analytical centre for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming, *in* 'System Modeling and Optimization', Vol. 84 of *Springer Lecture Notes in Control and Sciences*, Budapest, Hungary, pp. 866–875.

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.

Vaidya, P. M. (1989), A new algorithm for minimizing convex functions over convex sets, *in* 'Proc 30th Annual Symposium on Foundations of Computer Sciencee', Research Triangle Park, North Carolina, pp. 338–343.

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.

Watson, A. & Ahumada, A. (1985), 'Model of human visual-motion sensing', *J. of the Optical Society of America* **2**(2), 322–341.

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, Germany, pp. 1019–1022.

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