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, SE-100 44 Stockholm, Sweden
`tony@nada.kth.se`

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.

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 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 ac-

cessed (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 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 do-

mains (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). Concerning non-linear scale-space concepts, (Guichard 1998) has proposed a morphological scale-space model model that commutes with Galilean transformations, and (Weickert 1998) has studied non-linear scale-spaces that comprise spatial shape adaptation.

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

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

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

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

then L and R are related according to $L(x;\ \Sigma_L, v_L) = R(y;\ \Sigma_R, v_R)$ where the covariance matrices Σ_L and Σ_R satisfy $\Sigma_R = A\Sigma_L A^T$ 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 $x_R - v_R = A(x_L - v_L) + b$ 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

$$\begin{cases} x' = x + v_x t \\ y' = y + v_y t \\ t' = t \end{cases} \quad \text{i.e.} \quad \begin{pmatrix} x' \\ y' \\ t' \end{pmatrix} = \begin{pmatrix} 1 & 0 & v_x \\ 0 & 1 & v_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ t \end{pmatrix} \qquad (3)$$

the spatio-temporal covariance matrix will transform as

$$\begin{pmatrix} C'_{xx} & C'_{xt} & C'_{xt} \\ C'_{xy} & C'_{yy} & C'_{yt} \\ C'_{xt} & C'_{yt} & C'_{tt} \end{pmatrix} = \begin{pmatrix} 1 & 0 & v_x \\ 0 & 1 & v_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} C_{xx} & C_{xt} & C_{xt} \\ C_{xy} & C_{yy} & C_{yt} \\ C_{xt} & C_{yt} & C_{tt} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ v_x & v_y & 1 \end{pmatrix} \qquad (4)$$

while for the mean vector we have

$$\begin{pmatrix} C'_x \\ C'_y \\ C'_t \end{pmatrix} = \begin{pmatrix} 1 & 0 & v_x \\ 0 & 1 & v_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} C_x \\ C_y \\ C_t \end{pmatrix} = \begin{pmatrix} C_x + v_x\, C_t \\ C_y + v_y\, C_t \\ C_t \end{pmatrix} \tag{5}$$

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.

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

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

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

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

$$f_L(x_L) \quad\;-\;\qquad x_R = A x_L + b \qquad\;\to\;\quad f_R(x_R)$$

Fig. 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 smoothing.

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

$$f_{out}(t) = \frac{\mu}{1+\mu} f_{out}(t-1) + \frac{1}{1+\mu} f_{in}(t) \tag{6}$$

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 $M^{(k)} = \sum_{i=1}^{k} \mu_i$ and variance $V^{(k)} = \sum_{i=1}^{k} \mu_i^2 + \mu_i$.

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

$$\mu_k = \tfrac{1}{2} \left(\sqrt{1 + 4V^{(1)} (\gamma - 1) \gamma^{k-1}} - 1 \right). \tag{7}$$

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

$$(\partial_s L)(x; \ s) = (\mathcal{A}L)(x; \ s) = \sum_{\xi \in \mathbb{Z}^D} a_\xi L(x + \xi; \ s), \tag{8}$$

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 (6) 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

$$f_{out}(x,t) = \frac{1}{1 + \mu_t} \begin{pmatrix} a \\ b \\ c \end{pmatrix} f_{out}(x, t - 1) + \frac{1}{1 + \mu_t} \begin{pmatrix} d \\ e \\ f \end{pmatrix} f_{in}(x,t) \tag{9}$$

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:

$$f_{out}(x,t) = \frac{1}{1 + \mu_t}(a \, f_{out}(x + 1, t - 1) + b \, f_{out}(x, t - 1) + c \, f_{out}(x - 1, t - 1)$$
$$+ \, d \, f_{in}(x + 1, t) + e \, f_{in}(x,t) + f \, f_{in}(x - 1, t). \tag{10}$$

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

$$\varphi(w, z) = \sum_{x = -\infty}^{\infty} \sum_{t = -\infty}^{\infty} T(x, t; \ s) \, w^x z^t = \frac{dw^{-1} + e + fw}{1 + \mu_t - (aw^{-1} + b + cw)z} \tag{11}$$

---

[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).

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

$$M = \left.\begin{pmatrix} \varphi_w \\ \varphi_z \end{pmatrix}\right|_{(w,z)=(1,1)} = \begin{pmatrix} \mu_x \\ \mu_t \end{pmatrix}, \tag{12}$$

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

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. While 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, however, one complication 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 it follows that

$$a = \frac{\mu_{xx} + \mu_x^2 + e - 1}{2} - \frac{\mu_{xt} + 2\mu_x\,\mu_{xt}}{2(1 + \mu_t)} \tag{13}$$

$$b = -\mu_{xx} - \mu_x^2 + 1 - e + \mu_t + \frac{2\,\mu_x\,\mu_{xt}}{(1 + \mu_t)}, \tag{14}$$

$$c = \frac{\mu_{xx} + \mu_x^2 + e - 1}{2} + \frac{\mu_{xt} - 2\mu_x\,\mu_{xt}}{2(1 + \mu_t)} \tag{15}$$

$$d = \frac{-\mu_x + 1 - e}{2} + \frac{\mu_{xt}}{2\,(1 + \mu_t)}, \tag{16}$$

$$f = \frac{+\mu_x + 1 - e}{2} - \frac{\mu_{xt}}{2\,(1 + \mu_t)}, \tag{17}$$

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

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

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

and by introducing mixed spatio-temporal derivatives according to

$$\delta_{xt}(f_{in}, f_{out})(x, t) = \begin{pmatrix} 1/2 \\ 0 \\ -1/2 \end{pmatrix} f_{in}(x, t) - \begin{pmatrix} 1/2 \\ 0 \\ -1/2 \end{pmatrix} f_{out}(x, t - 1) \qquad (18)$$

$$\delta_{xxt}(f_{in}, f_{out})(x, t) = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} f_{in}(x, t) - \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} f_{out}(x, t - 1), \qquad (19)$$

we can with $\mu_{xxt} = \nu$ express the spatio-temporal smoothing scheme in (9) as

$$
\begin{aligned}
f_{out}(x, t) - f_{out}(x, t - 1) = \frac{1}{1 + \mu_t}&\big(-\mu_x \delta_x f_{in}(x, t) + \delta_t(f_{in}, f_{out})(x, t) \\
&+ \frac{1}{2}(\mu_{xx} + \mu_x^2 - \frac{2\mu_x \mu_{xt}}{1 + \mu_t}) \, \delta_{xx} f_{out}(x, t - 1) \\
&+ \frac{2}{2} \frac{\mu_{xt}}{1 + \mu_t} \, \delta_{xt}(f_{in}, f_{out})(x, t) \\
&+ \frac{3}{6} \frac{\mu_{xxt}}{1 + \mu_t} \, \delta_{xxt}(f_{in}, f_{out})(x, t)\big) \qquad (20)
\end{aligned}
$$

Alternatively, after introducing the following notation for binomial smoothing with variance ν in the spatial domain

$$\mathrm{Bin}(\nu) f_{in}(x, t) = \begin{pmatrix} \nu/2 \\ 1 - \nu \\ \nu/2 \end{pmatrix} f_{in}(x, t) \qquad (21)$$

the recursive updating scheme can be expressed on the form

$$
\begin{aligned}
f_{out}(x, t) - f_{out}(x, t - 1) = \frac{1}{1 + \mu_t}&\big(-\mu_x \delta_x f_{in}(x, t) + \mathrm{Bin}(\nu) f_{in}(x, t) - f_{out}(x, t - 1) \\
&+ \frac{1}{2}(\mu_{xx} + \mu_x^2 - \frac{2\mu_x \mu_{xt}}{1 + \mu_t} - \nu) \, \delta_{xx} f_{out}(x, t - 1) \\
&+ \frac{\mu_{xt}}{1 + \mu_t} \, \delta_{xt}(f_{in}, f_{out})(x, t)\big) \qquad (22)
\end{aligned}
$$

From this expression it is apparent how the mean and variance parameters μ_x, μ, μ_{xx} and μ_{xt} influence the recursive smoothing scheme. Moreover, we see how the free parameter $\nu = 1 - e$ can be interpreted as 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. Regarding the existence of non-negative values of a, b, c, d, e and f, it can be shown that a necessary and sufficient condition is that

$$
\max\left(\left| \mu_x - \frac{\mu_{xt}}{1 + \mu_t} \right|, \mu_{xx} + \mu_x^2 - \mu_t - \frac{2\mu_x \mu_{xt}}{1 + \mu_t} \right) \le \nu \le
$$
$$
\le \min(1, \mu_{xx} + \mu_x^2 - \frac{2\mu_x \mu_{xt} + |\mu_{xt}|}{1 + \mu_t}) \qquad (23)
$$

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, the effective filter parameters of the composed kernel will be of the form

$$C_x^{(k)} = \sum_{i=1}^{k} \mu_x^{(i)}, \quad C_t^{(k)} = \sum_{i=1}^{k} \mu_t^{(i)}$$

$$C_{xx}^{(k)} = \sum_{i=1}^{k} \mu_{xx}^{(i)}, \quad C_{xt}^{(k)} = \sum_{i=1}^{k} \mu_{xt}^{(i)}, \quad C_{tt}^{(k)} = \sum_{i=1}^{k} (\mu_t^{(i)})^2 + \mu_t^{(i)}. \quad (24)$$

To parameterize these filter shapes, let us start from the transformation properties (4) and (5) 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

$$\begin{pmatrix} C_{xx} & C_{xt} \\ C_{xt} & C_{tt} \end{pmatrix} = \begin{pmatrix} \lambda_{xx} + v^2\lambda_t & v\lambda_t \\ v\lambda_t & \lambda_t \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} C_x \\ C_t \end{pmatrix} = \begin{pmatrix} v\,C_t \\ C_t \end{pmatrix} \quad (25)$$

where C_t and C_{tt} are coupled according to (24).

Permissable combinations of filter parameters. 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 (24) and (25) it follows that

$$\mu_{xx} = \lambda_{xx} + v^2\lambda_t \quad (26)$$

$$\mu_{xt} = v\lambda_t \quad (27)$$

$$\mu_t = \frac{1}{2}\left(\sqrt{1 + 4\lambda_t} - 1\right) \quad (28)$$

$$\mu_x = v\,\mu_t = \frac{v}{2}\left(\sqrt{1 + 4\lambda_t} - 1\right) \quad (29)$$

and for these specific values of the parameters, it holds that $\mu_x = \mu_{xt}/(1 + \mu_t)$. After some calculations it can be shown that 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

$$\frac{(|v| - v^2)}{2}\left(\sqrt{1 + 4\lambda_t} - 1\right) \leq \lambda_{xx} \leq \frac{1 + v^2}{2} + \frac{1 - v^2}{2}\sqrt{1 + 4\lambda_t} \quad (30)$$

If a larger amount of spatial smoothing is required than allowed by this inequality, a straightforward approach is to divide the smoothing step into several layers of the recursive filters 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 $\lambda'_t = \lambda_t/k$. To capture motions with $|v| > 1$, one possible approach is to complement this recursive filtering framework by an additional warping mechanism. Alternatively, we may 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 by a factor h, we finally obtain
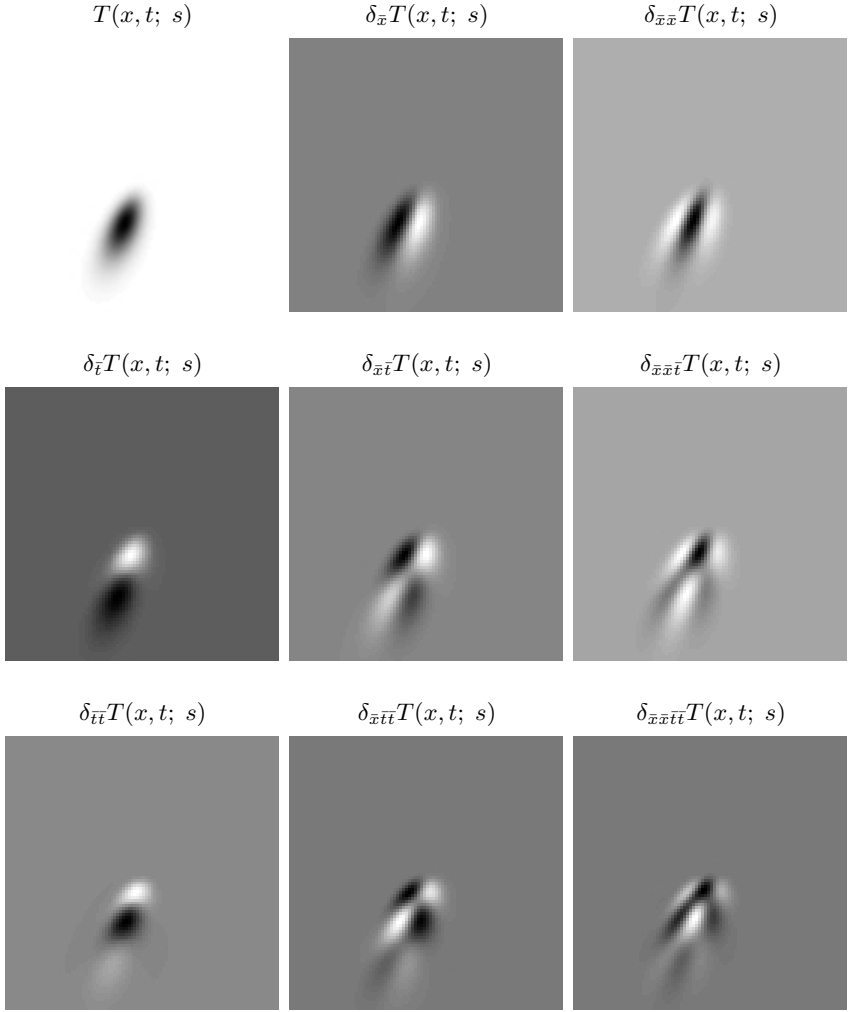
Fig. 2. Equivalent filter kernels corresponding to the result of using a discrete delta function as input and applying repeated non-separable recursive filtering according to (22) with a Galilean transformation parametrization of the filter shape (25) followed by subsequent Galilean based computation of velocity-adapted spatio-temporal derivatives according to (32) 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 $k = 10$ and the image size 150×150 pixels. Horisontal axis: space x. Vertical axis: time t.)

$$\frac{k(h|v| - v^2)}{2} \left(\sqrt{1 + \frac{4\lambda_t}{k}} - 1 \right) \leq \lambda_{xx} \leq k \left(\frac{h^2 + v^2}{2} + \frac{h^2 - v^2}{2} \sqrt{1 + \frac{4\lambda_t}{k}} \right) \tag{31}$$

where the upper bound on the image velocity is now given by $|v| \leq h$. Thus, we will be able to capture higher image velocites at coarser spatial scales using spatio-temporal recursive filtering only. However, if we want to capture rapid motions at fine spatial scales, then an additional warping mechanism is necessary.

Figure 2 shows a few examples of kernels computed in this way, followed by the computation of velocity-adapted spatio-temporal derivatives according to

$$\partial_{\bar{x}} = \partial_x \qquad \partial_{\bar{t}} = v\,\partial_x + \partial_t \qquad (32)$$

2.4   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

$$f_{out}(x,y,t) = \frac{1}{1+\mu_t} \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} f_{out}(x,t-1) + \frac{1}{1+\mu_t} \begin{pmatrix} j & k & l \\ m & n & p \\ q & r & s \end{pmatrix} f_{in}(x,t) \quad (33)$$

where the matrices within parentheses denote computational symbols in the spatial domain (which relate the filter coefficients $a, b, \ldots, s$ to spatial positions). From the generating function of the corresponding filter

$$\varphi(u,v,z) =$$
$$= \frac{pu^{-1} + mu + kv^{-1} + rv + lu^{-1}v^{-1} + su^{-1}v + juv^{-1} + quv}{1 + \mu_t - (fu + du^{-1} + bv + hv^{-1} + cu^{-1}v^{-1} + iu^{-1}v + auv^{-1} + guv)z} \qquad (34)$$

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

$$M = \begin{pmatrix} \varphi_u \\ \varphi_v \\ \varphi_z \end{pmatrix}\Bigg|_{(u,v,z)=(1,1,1)} = \begin{pmatrix} \mu_x \\ \mu_y \\ \mu_t \end{pmatrix}, \qquad (35)$$

$$V = \begin{pmatrix} \varphi_{uu} + \varphi_u - \varphi_u^2 & \varphi_{uv} - \varphi_u\varphi_v & \varphi_{uz} - \varphi_u\varphi_z \\ \varphi_{uv} - \varphi_u\varphi_v & \varphi_{vv} - \varphi_v - \varphi_v^2 & \varphi_{vz} - \varphi_v\varphi_z \\ \varphi_{uz} - \varphi_u\varphi_z & \varphi_{vz} - \varphi_v\varphi_z & \varphi_{zz} + \varphi_z - \varphi_z^2 \end{pmatrix}\Bigg|_{(u,v,z)=(1,1,1)} \qquad (36)$$

$$= \begin{pmatrix} \mu_{xx} & \mu_{xy} & \mu_{xt} \\ \mu_{xy} & \mu_{yy} & \mu_{yt} \\ \mu_{xt} & \mu_{yt} & \mu_t^2 + \mu_t \end{pmatrix}. \qquad (37)$$

Our next aim is to solve for the parameters a, b, ... s in the recursive filter in terms of the parameters $\mu_x, \mu_y, \mu_t, \mu_{xx}, \mu_{xy}, \mu_{yy}, \mu_{xt}$ 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, ... 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 the temporal direction. Thus, we expect to problem to have $18 - 1 - 3 - (6 - 1) = 9$ degrees of freedom, After some calculations, it can be shown that in terms of the difference operators in figure 3, this non-separable recursive smoothing scheme can on incremental form be expressed as

$$
f_{out}(x, y, t) - f_{out}(x, y, t - 1) =
$$
$$
= \frac{1}{1 + \mu_t} (D - \mu_x \, \delta_x \, f_{in}(x, y, t) - \mu_y \, \delta_y \, f_{in}(x, y, t)
$$
$$
+ \frac{1}{2} \left(\mu_{xx} + \mu_x^2 - \frac{2\mu_x \mu_{xt}}{1 + \mu_t} \right) \delta_{xx} f_{out}(x, y, t - 1)
$$
$$
+ \left(\mu_{xy} + \mu_x \mu_y - \frac{\mu_y \mu_{xt} + \mu_x \mu_{yt}}{1 + \mu_t} \right) \delta_{xy} f_{out}(x, y, t - 1)
$$
$$
+ \frac{1}{2} \left(\mu_{yy} + \mu_y^2 - \frac{2\mu_y \mu_{yt}}{1 + \mu_t} \right) \delta_{xy} f_{out}(x, y, t - 1)
$$
$$
+ \frac{\mu_{xt}}{1 + \mu_t} \delta_{xt}(f_{in}, f_{out})(x, y, t)
$$
$$
+ \frac{\mu_{yt}}{1 + \mu_t} \delta_{yt}(f_{in}, f_{out})(x, y, t)) \tag{38}
$$

where the nine degrees of freedom in the solution are contained in

$$
D = f_{in}(x, y, t) - f_{out}(x, y, t - 1)
$$
$$
+ \tfrac{3}{6} \mu_{xxt} \, \delta_{xxt}(f_{in}, f_{out})(x, y, t) + \tfrac{6}{6} \mu_{xyt} \, \delta_{xyt}(f_{in}, f_{out})(x, y, t)
$$
$$
+ \tfrac{3}{6} \mu_{yyt} \, \delta_{yyt}(f_{in}, f_{out})(x, y, t)
$$
$$
+ \tfrac{3}{6} \mu_{xxy} \, \delta_{xxy}(f_{out})(x, y, t) + \tfrac{3}{6} \mu_{xyy} \, \delta_{xyy}(f_{out})(x, y, t)
$$
$$
+ \tfrac{12}{24} \mu_{xxyt} \, \delta_{xxyt}(f_{out}, f_{out})(x, y, t) + \tfrac{12}{24} \mu_{xyyt} \, \delta_{xyyt}(f_{out}, f_{out})(x, y, t)
$$
$$
+ \tfrac{6}{24} \mu_{xxyy} \, \delta_{xxyy}(f_{out})(x, y, t) + \tfrac{30}{120} \mu_{xxyyt} \, \delta_{xxyyt}(f_{out}, f_{out})(x, y, t). \tag{39}
$$

Parameterization of filter shapes based on Galilei transformations. 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

$$
\Sigma = \begin{pmatrix} C_{xx} & C_{xt} & C_{xt} \\ C_{xy} & C_{yy} & C_{yt} \\ C_{xt} & C_{yt} & C_t \end{pmatrix} = \begin{pmatrix} \lambda_1 \cos^2 \alpha + \lambda_2 \sin^2 \alpha & (\lambda_2 - \lambda_1) \cos \alpha \sin \alpha & 0 \\ (\lambda_2 - \lambda_1) \cos \alpha \sin \alpha & \lambda_1 \sin^2 \alpha + \lambda_2 \cos^2 \alpha & 0 \\ 0 & 0 & \lambda_t \end{pmatrix} \tag{40}
$$

Then, by subjecting this matrix to a Galilei transformation (4), we obtain

$$
\Sigma' = \begin{pmatrix} \lambda_1 \cos^2 \alpha + \lambda_2 \sin^2 \alpha + v_x^2 \lambda_t & (\lambda_2 - \lambda_1) \cos \alpha \sin \alpha + v_x v_y \lambda_t & v_x \lambda_t \\ (\lambda_2 - \lambda_1) \cos \alpha \sin \alpha + v_x v_y \lambda_t & \lambda_1 \sin^2 \alpha + \lambda_2 \cos^2 \alpha + v_y^2 \lambda_t & v_y \lambda_t \\ v_x \lambda_t & v_y \lambda_t & \lambda_t \end{pmatrix} \tag{41}
$$

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

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

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

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

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

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

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

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

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

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

Fig. 3. Difference operators used in the construction of the discrete recursive spatio-temporal scale-space representation in 2+1-D space-time (38). 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_{xxt} have been suppressed.

Velocity-adapted spatio-temporal derivatives are finally given by

$$\partial_{\bar{x}} = \partial_x \qquad \partial_{\bar{y}} = \partial_y \qquad \partial_{\bar{t}} = v_x \, \partial_x + v_y \, \partial_y + \partial_t \qquad (42)$$

Figures 4–5 show a few examples of spatio-temporal scale-space kernels generated in this way. Figure 4 shows filters corresponding to $v = 0$, while figure 5 shows corresponding velocity adapted (and non-separable) recursive filters for a non-
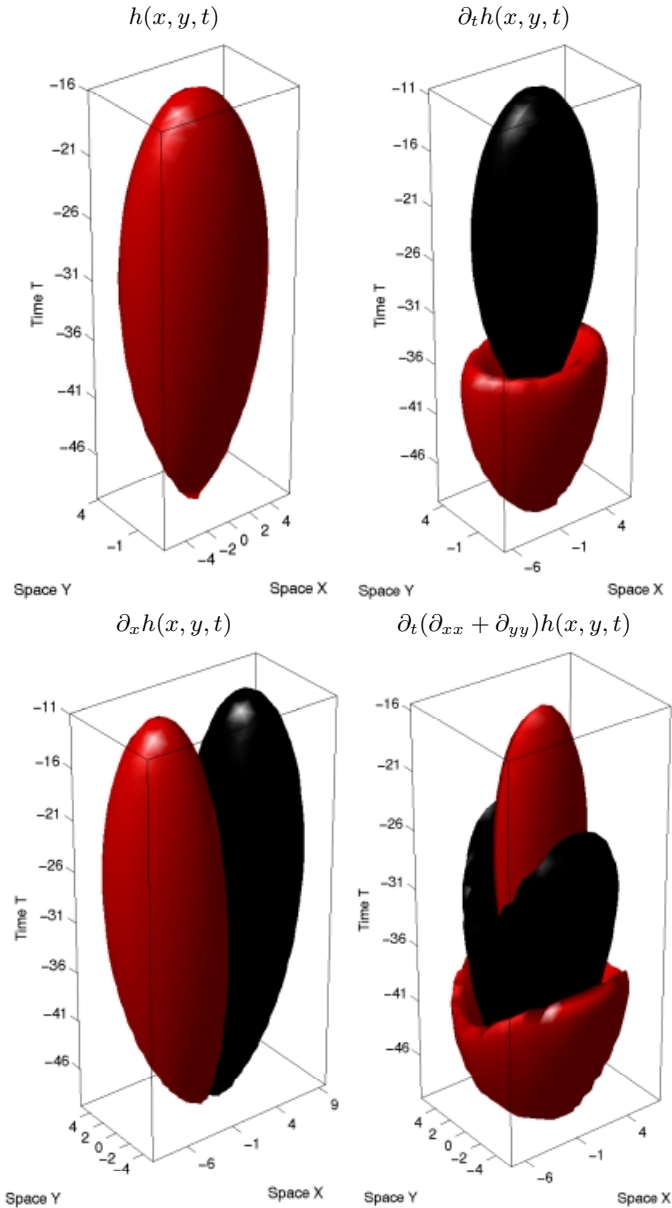
$h(x, y, t)$ $\partial_t h(x, y, t)$

$\partial_x h(x, y, t)$ $\partial_t (\partial_{xx} + \partial_{yy}) h(x, y, t)$

Fig. 4. 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 have been coupled in cascade.)
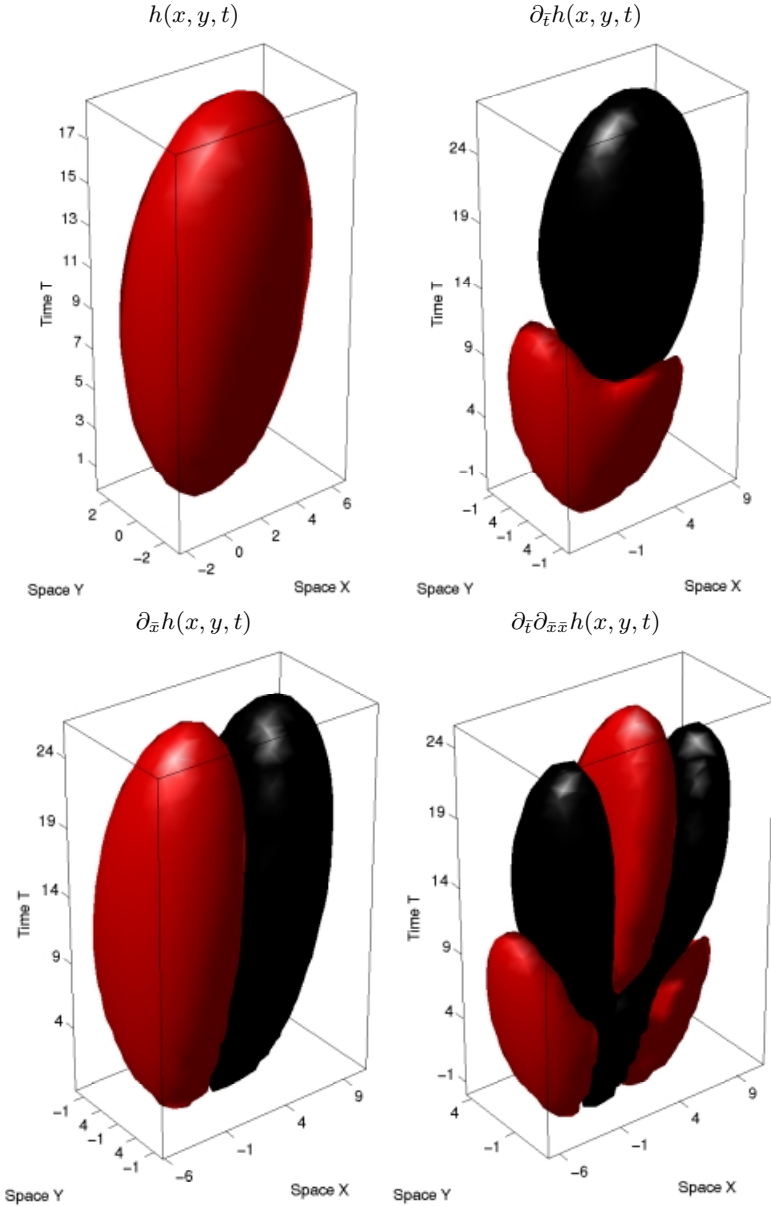
Fig. 5. Level surfaces of spatio-temporal receptive fields for the 2+1-D velocity-adapted spatio-temporal scale-space. (a) The raw smoothing kernel $h(x, y, t)$. (b) First-order temporal derivative $\partial_{\bar{t}} h(x, y, t)$. (c) First-order spatial derivative $\partial_{\bar{x}} h(x, y, t)$. (d) 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 have been coupled in cascade.)

zero velocity $v \neq 0$.[2] 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 (38), and then computing derivative approximations from difference operators in space and time, which are combined according to (42) at every image point.

3   Summary and Outlook

We have presented a theory for non-separable spatio-temporal scale-space kernels, based on time-causal recursive filtering. This theory provides a well-founded framework for 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. We propose that the 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 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 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).

References

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

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

Ballester, C. & Gonzalez, M. (1998), 'Affine invariant texture segmentation and shape from texture by variational methods', *J. Math. Im. Vis.* **9**, 141–171.

Burt, P. J. (1981), 'Fast filter transforms for image processing', *CVGIP* **16**, 20–51.

---

[2] 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.

Crowley, J. L. (1981), A Representation for Visual Information, PhD thesis, Carnegie-Mellon University, 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', *IJCV* **1**, 167–187.

Fleet, D. J. & Langley, K. (1995), 'Recursive filters for optical flow', *IEEE-PAMI* **17**(1), 61–67.

Florack, L. M. J. (1997), *Image Structure*, Kluwer, Netherlands.

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

Freeman, W. T. & Adelson, E. H. (1991), 'The design and use of steerable filters', *IEEE-PAMI* **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', *IJCV* **1**, 279–302.

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

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

Laptev, I. & Lindeberg, T. (2002), Velocity-adaption for direct recognition of activities using multi-scale spatio-temporal receptive field responses. Technical report ISRN KTH/NA/P–02/04–SE, KTH, Stockholm, Sweden.

Lindeberg, T. (1994), *Scale-Space Theory in Computer Vision*, Kluwer, Netherlands.

Lindeberg, T. (1997), Linear spatio-temporal scale-space, *in* 'Proc. Scale-Space'97', Utrecht, Netherlands, Vol. 1252 of Springer LNCS, pp. 113–127.

Lindeberg, T. (2001), Linear spatio-temporal scale-space, report, ISRN KTH/NA/P--01/22--SE, KTH, Stockholm, Sweden.

Lindeberg, T. & Fagerström, D. (1996), Scale-space with causal time direction, *in* 'ECCV'96', Vol. 1064, 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', *IVC* **15**, 415–434.

Nagel, H. & Gehrke, A. (1998), Spatiotemporal adaptive filtering for estimation and segmentation of optical flow fields, *in ECCV'98*, Freiburg, Germany, pp. 86–102.

Schaffalitzky, F. & Zisserman, A. (2001), Viewpoint invariant texture matching and wide baseline stereo, *in Proc. ICCV'01*, Vancouver, Canada.

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

ter Haar Romeny, B., Florack, L. & Nielsen, M. (2001), Scale-time kernels and models, *in Scale-Space'01*, 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', *Vis. Res.* **40**(2), 3685–3702.

Watson, A. & Ahumada, A. (1985), 'Model of human visual-motion sensing', *JOSA* **2**(2), 322–341.

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

Witkin, A. P. (1983), Scale-space filtering, *in* '8th IJCAI', Karlsruhe, Germany, pp. 1019–1022.

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