

Linear Spatio-Temporal Scale-Space ^{*}

Tony Lindeberg

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

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

1 Introduction

When analysing sensory data such as images, a fundamental constraint arises from the fact that real-world objects may appear in different ways depending upon the scale of observation. This insight is a major motivation for the development of multi-scale representations such as *pyramids* (Burt 1981; Crowley 1981) and *scale-space representation* (Witkin 1983; Koenderink 1984; Yuille and Poggio 1986; Koenderink and van Doorn 1992; Florack 1993; Lindeberg 1994b).

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

The subject of this article is to develop a scale-space theory for such spatio-temporal image domains, to enable a visual agent to register spatio-temporal events at multiple spatial and temporal scales in a well-founded manner.

Technically, we shall start from a previously stated scale-space formulation in terms of *non-enhancement of local extrema*, and relax the symmetry requirements that have been used in previous applications of this scale-space axiom.

^{*} Earlier version presented in Proc. First Int. Conf. Scale-Space'97 (B. M. ter Haar Romeny, L. M. J. Florack, J. J. Koenderink, and M. A. Viergever, eds.), vol. 1252 of Lecture Notes in Computer Science, (Utrecht, The Netherlands), pp. 113–127, Springer Verlag, Berlin, July 1997.

Examples of phenomena which will be captured by the resulting framework include: (i) spatial shifting mechanisms without need for external warping, (ii) time-causal scale-space filters not extending into the future, and (iii) non-separable and elongated filter kernels over space and time, which allow for shape adaptation in space and velocity adaptation in the direction of motion.

2 Non-enhancement of local extrema: Continuous signals

When (Koenderink 1984) derived the uniqueness of the Gaussian kernel for generating a scale-space representation on a continuous spatial domain, he expressed a *causality* requirement, stating that new level curves must not be created with increasing scales. Then, when (Lindeberg 1990) derived a scale-space theory for discrete image domains, he reformulated this condition as follows:

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

(Lindeberg 1990) also imposed a semi-group structure on the family of convolution kernels, and used a result from functional analysis (Hille and Phillips 1957), stating that a semi-group structure, combined with strong continuity with respect to scale, implies that the scale-space family must have an *infinitesimal generator*. In other words, if a transformation operator \mathcal{T}_s from the input signal f to the scale-space representation $L(\cdot; s)$ at any scale s is defined by

$$L(\cdot; s) = \mathcal{T}_s f(\cdot), \quad (1)$$

then under reasonable regularity requirements there exists a limit case $\mathcal{A}f = \lim_{h \downarrow 0} \frac{\mathcal{T}_h f - f}{h}$ of this operator (the infinitesimal generator) such that the scale-space family satisfies the differential equation

$$\partial_s L(\cdot; s) = \lim_{h \downarrow 0} \frac{L(\cdot; t+h) - L(\cdot; t)}{h} = \mathcal{A}(\mathcal{T}_s f(\cdot)) = \mathcal{A}L(\cdot; s). \quad (2)$$

It was shown that this structure, combined with non-enhancement of local extrema, strongly restricts permissible classes of isotropic scale-space operators.

The main subject of this article is to analyse the consequences of imposing similar constraints on spatial, temporal and spatio-temporal domains under weaker symmetry requirements.

2.1 Continuous domain

On a continuous domain, it is natural to formalize the non-enhancement requirement of local extrema in terms of a sign condition on the derivative of the scale-space family with respect to the scale parameter. Hence, at any non-degenerate

extremum point (extremum point where the determinant of the Hessian matrix is non-zero) we require the following conditions to hold:

$$\partial_s L < 0 \quad \text{at a non-degenerate local maximum,} \quad (3)$$

$$\partial_s L > 0 \quad \text{at a non-degenerate local minimum.} \quad (4)$$

In (Lindeberg 1994a) it is shown that a linear and shift-invariant infinitesimal generator \mathcal{A} satisfying these requirements must be a linear combination of first- and second-order derivatives, where the second-order terms correspond to an elliptic differential operator and the first-order terms are arbitrary.

Rotationally symmetric linear scale-space. If rotational symmetry is added to this structure, it follows that \mathcal{A} has to be proportional to the Laplacian operator, and the scale-space family corresponds to convolution with Gaussian kernels.

Affine Gaussian scale-space. If we relax the condition about rotational symmetry, while keeping a requirement that the corresponding Green's function should be mirror symmetric in every line through the origin (*i.e.*, to avoid spatial shifts, its Fourier transform should be real), we obtain the *affine Gaussian scale-space representation* generated by convolution with non-uniform Gaussian kernels

$$g(x; \Sigma_s) = \frac{1}{(2\pi)^{D/2} \sqrt{\det \Sigma_s}} e^{-x^T \Sigma_s^{-1} x/2}, \quad (5)$$

where Σ_s is a symmetric positive definite (covariance) matrix. If the covariance matrix is written $\Sigma_s = s \Sigma_0$ for some (constant) matrix Σ_0 , then the *shape-adapted affine scale-space representation* satisfies the diffusion equation

$$\partial_s = \frac{1}{2} \nabla^T (\Sigma_0 \nabla L). \quad (6)$$

This scale-space concept has the attractive property that it is *closed* under affine transformations, which is a highly useful property in tasks such as first-order shape estimation, stereo matching and flow estimation (Lindeberg 1994b).

Temporal and spatio-temporal domains. If we, relax the condition that Fourier transform of the Green's function should be real, then we obtain a scale-space generated by *velocity-adapted Gaussian kernels* (see also (Florack *et al.* 1992))

$$g(x; \Sigma_s, v) = \frac{1}{(2\pi)^{D/2} \sqrt{\det \Sigma_s}} e^{-(x-v_s)^T \Sigma_s^{-1} (x-v_s)/2}, \quad (7)$$

which for a given $\Sigma_s = s \Sigma_0$ and a given $v_s = s v_0$ satisfy the diffusion equation

$$\partial_s = \frac{1}{2} \nabla^T (\Sigma_0 \nabla L) - v_0^T \nabla L. \quad (8)$$

This structure can be interpreted as a simple model for scale-spaces on temporal and spatio-temporal domains. Specifically, this model allows the following phenomena to be captured: (i) receptive fields may be adapted to velocity estimates

and thus follow image structures as they move over time, (ii) receptive fields at coarse temporal scales can be associated with a certain time delay.

Unless $\det \Sigma_0 = 0$, however, the support regions of all filters cover the entire domain, and there are no non-trivial semi-groups of time-causal scale-space kernels in this filter class. As we shall see in next section, however, corresponding arguments in the discrete case do indeed give rise to a richer structure.

3 Non-enhancement of local extrema: Discrete signals

For signals defined on a discrete domain, it can be shown that non-enhancement of local extrema implies that the scale-space family $L: \mathbf{Z}^D \times \mathbb{R}^+ \rightarrow \mathbb{R}$ of any discrete signal $f: \mathbf{Z}^D \rightarrow \mathbb{R}$ must satisfy the semi-discrete differential equation

$$(\partial_s L)(x; s) = (\mathcal{A}L)(x; t) = \sum_{\xi \in \mathbf{Z}^D} a_\xi L(x - \xi; s), \quad (9)$$

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

This result follows from similar arguments as the proof of Theorem 4.10 in (Lindeberg 1994b), if the spatial symmetry requirements are relaxed.

In compact operator notation, the solution to (9) at can be written $L = e^{s \mathcal{A}} f$. To give a more explicit parametrization of the filter class spanned by this construction, let us introduce the following basic difference operators, and use them as a basis for expressing the degrees of freedom in the coefficients a_ξ :

$$(\delta_x f)(x) = (f(x+1) - f(x-1)) / 2, \quad (10)$$

$$(\delta_- f)(x) = f(x) - f(x-1), \quad (11)$$

$$(\delta_{xx} f)(x) = f(x+1) - 2f(x) + f(x-1). \quad (12)$$

3.1 1-D discrete spatial domain

On a one-dimensional spatial domain, the class of permissible filter kernels (with coefficients c_n ($n \in \mathbf{Z}$)) generated by (9) can then be written as the solution to

$$\partial_s L = -C_x \delta_x L + \frac{C_{xx}}{2} \delta_{xx} L, \quad (13)$$

where $C_{xx} > 0$ and $|C_x| \leq C_{xx}$ are necessary conditions for the non-central filter coefficients to be non-negative. From the associated generating function

$$\varphi(z) = \sum_{n \in \mathbf{Z}} c_n z^n = e^{s \sum_{\xi \in \mathbf{Z}} a_\xi z^\xi} = e^{s(C_x(z-z^{-1})/2 + C_{xx}(z-2+z^{-1})/2)},$$

we get mean $M = \varphi_z(1) = -s C_x$ and variance $V = \varphi_{zz}(1) + \varphi_z(1) - \varphi_z^2(1) = s C_{xx}$ (measured in units of the grid spacing). Hence, this scale-space concept allows signals to be translated during the smoothing process (without any need for explicit warping), and the amount of diffusion (given by C_{xx}) imposes a bound on the amount of translation (given by C_t) that can be performed, given the positivity constraint on the infinitesimal scale-space generator \mathcal{A} .

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

$$L(x; s) = \sum_{n=-\infty}^{\infty} T(n; s) f(x - n) \quad \text{where} \quad T(n; s) = e^{-s} I_n(s). \quad (14)$$

where I_n are the modified Bessel function of integer order. (Here, we have without loss of generality set $C_{xx} = 1$.)

3.2 1-D discrete temporal domain

On a one-dimensional temporal domain, temporal causality implies that we must require that $a_1 = 0$. The remaining degree of freedom can be parametrized by

$$\partial_s L = -C_t \delta_- L, \quad (15)$$

(where we without loss of generality can set to $C_t = 1$). This gives rise to a *Poisson-type temporal scale-space*, generated by convolution with Poisson kernels

$$L(x; s) = \sum_{n=-\infty}^{\infty} p(n; s) f(x - n) \quad \text{where} \quad p(n; s) = e^{-s} \frac{s^n}{n!} \quad (16)$$

having mean $M = -s$ and variance $V = s$. Notably, this family has no corresponding semi-group structure on continuous time respecting temporal causality.

3.3 2-D discrete spatial domain

On a two-dimensional spatial domain, additional degrees of freedom arise in terms of the eccentricity of the filter and its orientation. The discrete counterpart of the affine Gaussian scale-space in (5) and (6) is obtained if we require the filter kernels in (9) to be mirror symmetric through the origin, *i.e.*, $a_{i,j} = a_{-i,-j}$.

With the difference operators $\delta_{xy} = \delta_x \delta_y$ and $\delta_{xxyy} = \delta_{xx} \delta_{yy}$, the resulting *discrete affine Gaussian scale-space* (allowing for discrete shape adaptation) can be parametrized as the solution to the semi-discrete differential equation

$$\partial_s L = \frac{1}{2} (C_{xx} \delta_{xx} L + 2C_{xy} \delta_{xy} L + C_{yy} \delta_{yy} L) + \frac{C_{xxyy}}{4} \delta_{xxyy} L, \quad (17)$$

where the infinitesimal generator has a computational molecule of the form

$$\mathcal{A} = \frac{1}{2} \begin{pmatrix} -C_{xy}/2 & C_{yy} & C_{xy}/2 \\ C_{xx} & -2(C_{xx} + C_{yy}) & C_{xx} \\ C_{xy}/2 & C_{yy} & -C_{xy}/2 \end{pmatrix} + \frac{C_{xxyy}}{4} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}. \quad (18)$$

This representation can be interpreted as a second-order discretization of the diffusion equation (6) associated with the continuous affine Gaussian scale-space

$$\partial_s L = \frac{1}{2}(C_{xx} L_{xx} + 2C_{xy} L_{xy} + C_{yy} L_{yy}), \quad (19)$$

where $C_{xx} > 0$ and $C_{xx}C_{yy} - C_{xy}^2 > 0$ are necessary requirements for the operator to be elliptic. The free parameter C_{xxyy} (which controls the addition of a discretization of the mixed fourth-order derivative L_{xxyy}) must satisfy

$$|C_{xy}| \leq C_{xxyy} \leq \min(C_{xx}, C_{yy}) \quad (20)$$

to ensure that all non-central coefficients are non-negative. (In practice, the feasibility condition $|C_{xy}| \leq \min(C_{xx}, C_{yy})$ arising from this positivity constraint is always more restrictive than the condition $|C_{xy}| \leq \sqrt{C_{xx}C_{yy}}$ for the operator to be elliptic. This implies that highly eccentric affine Gaussian kernels cannot be represented by non-negative discrete scale-space kernels on a square discrete grid, unless the filters are approximately aligned to the coordinate directions. A more detailed analysis shows that this effect starts to occur when the ratio between the eigenvalues of the covariance matrix Σ exceeds $\epsilon = 3 + 2\sqrt{2} \approx 5.8$.)

To appreciate the importance of C_{xxyy} also in the isotropic case ($C_{xx} = C_{yy} = 1$, $C_{xy} = 0$), introduce discrete approximations ∇_5^2 and $\nabla_{\times 2}^2$ to the 2-D Laplacian operator ∇^2 by $(\nabla_5^2 f)_{0,0} = f_{-1,0} + f_{+1,0} + f_{0,-1} + f_{0,+1} - 4f_{0,0}$ and $(\nabla_{\times 2}^2 f)_{0,0} = 1/2(f_{-1,-1} + f_{-1,+1} + f_{+1,-1} + f_{+1,+1} - 4f_{0,0})$, and reparametrize (17) as $\partial_s L = \frac{1}{2}((1 - \gamma)\nabla_5^2 L + \gamma\nabla_{\times 2}^2 L)$. Then, it can be shown that $\gamma = 1/3$ gives the lowest degree of rotational asymmetry in the Fourier domain.

3.4 1+1-D discrete spatio-temporal domain

To obtain an intuitive understanding for spatio-temporal data, let us first study the case with one spatial dimension (represented by x) and one temporal dimension (represented by t). The general conditions in (9) then give that the computational molecule of \mathcal{A} has to be of the form

$$A = \begin{pmatrix} 0 & 0 & 0 \\ A & -F & E \\ B & C & D \end{pmatrix} \quad (21)$$

for some constants $A, B, C, D, E > 0$ and $F = A + B + C + D + E$. With δ_t denoting the (backward difference) δ_- -operator in the t -direction, $\delta_{xt} = \delta_x \delta_t$ and $\delta_{xxt} = \delta_{xx} \delta_t$, it can be shown that this family can be parametrized as

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

where $C_{xx}, C_t > 0$ and non-negativity of non-central coefficients implies that

$$-C_{xx} \leq C_x \leq C_{xx}, \quad (22)$$

$$-C_t \leq C_{xt} \leq C_t, \quad (23)$$

$$|C_{xt}| \leq C_{xxt} \leq C_t. \quad (24)$$

From the generating function of the corresponding filter, we then get the mean vector M and the covariance matrix V as

$$M = -s \begin{pmatrix} C_x \\ C_t \end{pmatrix}, \quad V = s \begin{pmatrix} C_{xx} & C_{xt} \\ C_{xt} & C_t \end{pmatrix}. \quad (25)$$

Thus, by varying C_x , C_t , C_{xx} within the constraints (22), (23) and (24), we can reach a four-dimensional subset of the five-dimensional manifold spanned by all variations of mean vectors and covariance matrices in two dimensions.

Parametrization of filter shapes. To give a more explicit parametrization of this filter class, let us introduce descriptors P , C , S and Q by

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

Then, the eigenvalues of V are given by $\lambda_{1,2} = \frac{1}{2}(P \pm Q)$ and an eigenvector normalized to unit length in the t -direction can be written $\Delta = (\dot{x}, 1) = (\tan \phi, 1)$,

$$\tan \phi = \text{sign } S \arctan \sqrt{\frac{Q - C}{Q + C}}, \quad (26)$$

corresponding to the following relation between C and S : $(\dot{x}^2 - 1)S = 2\dot{x}C$.

Velocity adaptation. When studying image structures that move over time, it is natural to adapt the shape of the spatio-temporal filter kernel to the direction of motion. A basic motivation for such an approach is to avoid excessive motion blur when observing moving objects at coarse time scales.

Given a velocity estimate $\dot{x} = \tan \phi$ (in units of the grid spacing) and a specification of the amount of smoothing in terms of the eigenvalues of V (now for convenience denoted λ_{xx} and λ_t), we can parametrize C_{xx} , C_{xt} and C_t by

$$\begin{pmatrix} C_{xx} & C_{xt} \\ C_{xt} & C_t \end{pmatrix} = \frac{1}{1 + \dot{x}^2} \begin{pmatrix} \lambda_{xx} + \dot{x}^2 \lambda_t & -\dot{x}(\lambda_{xx} - \lambda_t) \\ -\dot{x}(\lambda_{xx} - \lambda_t) & \lambda_t + \dot{x}^2 \lambda_{xx} \end{pmatrix}. \quad (27)$$

To illustrate how the positivity constraints (22), (23) and (24) restrict the permissible values of λ_{xx} and λ_t , let us consider the limit cases when $|\dot{x}| \ll 1$ and $|\dot{x}| \gg 1$. Then, straightforward calculations lead to the following conditions:

$$|\dot{x}| \lambda_{xx} \leq \lambda_t \leq \frac{1}{|\dot{x}|} \lambda_{xx} \quad (|\dot{x}| \ll 1), \quad (28)$$

$$\frac{1}{|\dot{x}|} \lambda_t \leq \lambda_{xx} \leq |\dot{x}| \lambda_t \quad (|\dot{x}| \gg 1). \quad (29)$$

In other words, at low image velocities on the image domain, the spatial scale imposes a constraint on which temporal scales are meaningful, whereas at high velocities, the temporal scale restricts which spatial scales are meaningful.

In certain situations, it is natural to couple also the mean values of the filter (the spatial shift and the temporal delay) in an analogous way:

$$\begin{pmatrix} C_x \\ C_t \end{pmatrix} = C_t \begin{pmatrix} \dot{x} \\ 1 \end{pmatrix} = \frac{\lambda_t + \dot{x}^2 \lambda_{xx}}{1 + \dot{x}^2} \begin{pmatrix} \dot{x} \\ 1 \end{pmatrix}. \quad (30)$$

Then, however, the positivity condition $|C_x| < C_{xx}$ combined with $|C_x| = |\dot{x}| C_t$ implies that non-negative velocity-adapted filters only exist when $|\dot{x}| < 1$.

To allow for such *complete velocity adaptation* at high velocities, we should thus rather send control signals to the image acquisition system, to allow the brightness pattern to be stabilized in the image domain (tracking/fixation).

Kernel graphs. Figure 1 shows a few examples of filters from this family, for different values of λ_{xx} , λ_t and \dot{x} . Observe how we in this way can generate discrete receptive field profiles with desirable spatio-temporal characteristics.

When $\dot{x} = 0$, this scale-space is separable, corresponding to the tensor product between the spatial and temporal scale-space representations (14) and (16).

3.5 2+1-D discrete spatio-temporal domain

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

If we consider a general three-dimensional infinitesimal generator of the form (9) and require all coefficients $a_\xi = a_{x,y,t}$ to be zero when time t is positive, we then obtain a computational molecule with 17 degrees of freedom. Out of these, 8 degrees of freedom will influence the mean values and the covariance matrices of the filters, while the only restriction on the remaining 9 degrees of freedom will be in terms of intervals (in analogy with C_{xxyy} in the 2-D discrete affine Gaussian scale-space and C_{xxt} in the 1+1-D discrete spatio-temporal scale-space).

With difference operators defined in an analogous way as in previous sections (implying that all differences in the time direction are based on the backward difference operator $\delta_t L(x, y, t) = L(x, y, t) - L(x, y, t - 1)$), this filter class can be parametrized as the solution to a differential equation of the form

$$\begin{aligned} \partial_s L = & -C_x \delta_x L - C_y \delta_y L - C_t \delta_t L \\ & + \frac{1}{2} (C_{xx} \delta_{xx} L + 2C_{xy} \delta_{xy} L + C_{yy} \delta_{yy} L + 2C_{xt} \delta_{xt} L + 2C_{yt} \delta_{yt} L) \\ & + \text{combinations of 1st- and 2nd-order differences w.r.t. } x, y \text{ and } t. \end{aligned} \quad (31)$$

Hence, this filter family allows for both shape adaptation in the spatial domain (variations of C_{xx} , C_{yy} and C_{xy}) and velocity adaptation between the temporal and spatial dimensions (variations of C_x , C_y , C_t , C_{xx} , C_{yy} , C_{xt} and C_{yt}).

A detailed analysis of the structure and properties of this scale-space is given in (Lindeberg 1996). Similar constructions can be done in higher dimensions.

3.6 Connection equations for multi-parameter scale-spaces

So far, we have mainly considered differential equations depending on a single evolution parameter (termed s). The shapes of the filters, however, are determined by a multi-dimensional set of filter parameters (C_x, C_{xx}, C_{xy}, C_t , etc.).

To analyse how filters with different shapes are related differentially, consider any of the semi-discrete differential equations in (13), (15), (17), (22) and (31). Introduce multi-index notation by $x^\omega = x_1^{\omega_1} x_2^{\omega_2} \dots x_D^{\omega_D}$, where $x = (x_1, x_2, \dots, x_D)$ are coordinates on the D -dimensional discrete domain, and let δ_{x^ω} represent a difference operator of order ω_i along the i th dimension. Moreover, let γ_{x^ω} denote the (Taylor) coefficient used for normalizing each coefficient C_{x^ω} for $\omega \in \Omega$. Then, any such semi-differential equation can be written:

$$\partial_s L = \sum_{\omega \in \Omega} \gamma_{x^\omega} C_{x^\omega} \delta_{x^\omega}. \quad (32)$$

Furthermore, let $\hat{\delta}_{x^\omega}(z)$ denote the generating function of the difference operator δ_{x^ω} (e.g., $\hat{\delta}_{x_1}(z) = (z_1 - z_1^{-1})/2$). Then, the generating function $\hat{h}(z)$ of the Green's function of (32) at scale $s = 1$ assumes the form

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

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

$$\partial_{C_{x^\omega}} h = \gamma_{x^\omega} \delta_{x^\omega} h. \quad (33)$$

In other words, if we define an $|\Omega|$ -parameter scale-space representation L of any signal f by considering the solution of (32) at $s = 1$ for all combinations of C_{x^ω} such that $\sum_{\omega \in \Omega} C_{x^\omega} \delta_{x^\omega}$ satisfies the positivity condition in (9), then for any C_{x^ω} this scale-space representation satisfies

$$\partial_{C_{x^\omega}} L = \gamma_{x^\omega} \delta_{x^\omega} L. \quad (34)$$

In other words, this structure shows how representations at coarser scales can be constructed from representations at finer scales in an *incremental* fashion:

$$L(\cdot; C + \Delta s \Delta C) = L(\cdot; C + \Delta s \sum_{\omega \in \Omega} \beta_\omega C_{x^\omega}) \approx L(\cdot; C) + \Delta s \partial_C L(\cdot; C). \quad (35)$$

Of course, we should always choose the directional derivative operator ∂_C

$$\partial_C = \sum_{\omega \in \Omega} \beta_\omega \partial_{C_{x^\omega}} \quad (\beta_\omega \in \mathbb{R}) \quad (36)$$

such that the corresponding spatial derivative operator (obtained from (34))

$$B = \sum_{\omega \in \Omega} \beta_\omega \gamma_{x^\omega} \delta_{x^\omega} \quad (37)$$

satisfies the positivity conditions of the infinitesimal scale-space generator in (9). (By construction, the locality and zero sum conditions are already satisfied.)

Moreover, stability requirements restrict the magnitude of Δs . Observe, however, that these operations only depend on nearest-neighbour relations. Thus, they are highly suitable for parallel implementation on fine-grained architectures.

3.7 Relations between the continuous and discrete models

From the central limit theorem, it follows that the discrete filters generated by (9) will approach Gaussian kernels when the evolution parameter s tends to infinity. Moreover, the algebraic structure of the mean values and the covariance matrices of the scale-space filters is similar in the continuous and discrete domains. In this respect, the continuous theory in section 2 can be regarded as an idealized model of the discrete theory in section 3 when grid effects are negligible.

The advantage of the discrete model is that it allows for a well-founded transition to coarse scales, such that all filters are scale-space kernels also in a discrete sense. Moreover, temporal causality holds exactly for the discrete filter class.

4 Relations to previous work

In his pioneering scale-space formulation for temporal data, (Koenderink 1988) transformed the time axis by mapping the present moment to the unreachable infinity. He introduced a time delay t_0 describing how long it takes before changes in the input can be perceived, and applied Gaussian convolution in transformed domain. In the original time domain, this corresponds to convolution by kernels of the form (where λ_t approximates the variance when $\lambda_t \ll t_0$):

$$h(t; t_0, \sigma_t) = \frac{1}{\sqrt{2\pi} \lambda_t} \left(\frac{t}{t_0} \right)^{-\frac{t_0}{2\lambda_t^2} \log\left(\frac{t}{t_0}\right) - 1} \quad (38)$$

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

The spatio-temporal scale-space model presented here has several similarities to these models, when projected to a pure temporal domain (section 3.2). The shapes of the resulting filter kernels and their temporal derivatives are qualitatively similar to those in (38), if the time delays and the variances are appropriately adapted. Moreover, from (15) it follows that derivatives with respect to time are equal to derivatives with respect to temporal scale.

In addition, if we implement this scale-space model in terms of strictly time-recursive operations (such as a cascade of first-order recursive filters), then the

Poisson-type scale-space at scale λ corresponds to the limit case when the number N of filters tends to infinity, while their individual variances λ/N tend to zero.

For a more general review of filter-based approaches to motion estimation, see *e.g.* the paper by (Barron *et al.* 1994) and the references therein.

5 Relations to biological vision

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

For *LGN neurons*, which have approximately circular center-surround organization, most RFs are space-time separable, and there are two main classes of temporal responses. In a “non-lagged cell”, the first temporal lobe is usually the largest one, whereas for a “lagged cell”, the second lobe typically dominates.

Such temporal response properties are typical for first- and second-order temporal derivatives of a temporal scale-space representation. The spatial response, on the other hand, shows a high similarity to a Laplacian of a Gaussian.

For *simple cells* in the striate cortex, which have spatially oriented RFs, the response properties range from separable to strongly inseparable, where a majority exhibit marked space-time inseparability. The temporal profile is typically biphasic, although some cells have monophasic or triphasic responses.

In terms of temporal derivatives, such behaviour arises from first-, zero- and second-order derivatives, respectively. Concerning the spatial response characteristics, there is a high similarity with derivatives of Gaussians (Young 1987). In fact, for all these linear RFs, spatio-temporal filters with similar response characteristics can be generated by applying difference operators of low orders to the spatio-temporal filters obtained from (27).

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

This structure is nicely compatible with velocity adaptation, as described in section 3.4. Moreover, with regard to the dominance of biphasic temporal response profiles for simple cells, it is interesting to note that spatio-temporal derivatives corresponding to first-order derivatives in the time direction, (more precisely δ_{xt} and δ_t) constitutes a key source of information for controlling the filter shapes for shape adaptation according to (27).

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

From Laplacian-type receptive fields at the lowest level of a computational hierarchy, it is always possible to reconstruct (the rotationally symmetric) scale-space representation by integrating Laplacian responses over scales. Then, a zero-order temporal scale-space representation can always be constructed in a similar way by integrating first- (or second-order) temporal derivatives. Alternatively, or both, temporal scale-space smoothing can be implemented as a cascade of first-order integrators (or recursive filters) over time.

Then, from such a separable spatio-temporal scale-space representation, non-separable spatio-temporal receptive fields as well as elongated receptive fields in the spatial domain can be constructed from the connection equations (34), (35).

Finally, spatial derivative approximations can be computed from spatial differences of this representation, and temporal derivative approximations be obtained from differences between temporal channels. Since all these operations are linear, it follows that scale-space properties transfer from the spatio-temporal scale-space representation to its spatio-temporal derivatives (Lindeberg 1994b).

6 Summary and discussion

We have presented a theory for scale-space representation of spatio-temporal data. Starting from a general condition about non-creation of structure with increasing scales in terms of non-enhancement of local extrema, a complete characterization has been given of the semi-groups of convolution transformations that obey this requirement on different types of image domains. The resulting theory comprises several of the existing continuous and discrete scale-space theories on symmetric spatial domains. In addition, it provides the following extensions:

- An explicit discrete mechanisms for shifting data during a discrete smoothing process, without explicit need for external warping.
- A discrete theory for elongated scale-space filters in the spatial domain, allowing for discrete shape adaptation to local image structures.
- A consistent scale-space representation over discrete time, which respects the time direction as strictly causal and preserves the semi-group structure.
- Simultaneous, *i.e.* non-separable, treatment of spatial and temporal domains.
- Velocity adaptation of spatio-temporal filters in the direction of motion.
- Connection equations allowing filters of different shapes and at different scales in a multi-parameter scale-space to be constructed incrementally.

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

Moreover, the computational structure shows interesting qualitative similarities to recent findings on spatio-temporal receptive fields in biological vision.

Further extensions. Throughout this treatment, we have considered square grids having the same uniform spatial sampling at all scales. More generally, similar arguments based on non-enhancement of local extrema can be carried out on other discrete topologies, such as hexagonal lattices or irregular grids.

The only difference in such situations will be that the spatial summation over nearest-neighbours in (9) should be replaced by a summation over nearest-neighbours defined from the appropriate connectivity concept. The locality, positivity and zero sum conditions transfer in a straightforward manner.

Similar extensions can be performed by decreasing the spatial and temporal sampling densities at coarser spatial and temporal scales. Besides the substantial data reduction by such a subsampling operation, it has the attractive property of increasing the range of velocities that can be captured by velocity adaptation. (Recall that for complete velocity adaptation, positivity of non-central filter coefficients implies a bound on the velocity relative to the grid spacing.)

Such an extension is also nicely compatible with foveated sensors, since large motions can be captured in the periphery, and guide fixation processes which enable fine-scale structures to be resolved at higher resolution in the fovea.

Further relaxation of scale-space axioms. More generally, one could also conceive of relaxing the requirement about translational invariance in the temporal domain, such that the filter family is not required to be a semi-group of convolution kernels. Such extensions are studied in more detail in (Lindeberg 1996).

Acknowledgements

An early version of this work was presented at the PhD School on Scale-Space Theory in Copenhagen, Denmark, May 1996. The support from the Swedish Research Council for Engineering Sciences, TFR, is gratefully acknowledged.

References

- [Barron *et al.*, 1994] J. J. Barron; D. J. Fleet, and S. S. Beachemin. “Performance of Optical Flow Techniques”. *IJCV*, 12(1), 1994.
- [Burt, 1981] P. J. Burt. “Fast Filter Transforms for Image Processing”. *CVGIP*, 16:20–51, 1981.
- [Crowley, 1981] J. L. Crowley. *A Representation for Visual Information*. PhD thesis, Carnegie-Mellon University, Robotics Institute, Pittsburgh, Pennsylvania, 1981.
- [DeAngelis *et al.*, 1995] G. C. DeAngelis; I. Ohzawa, and R. D. Freeman. “Receptive Field Dynamics in the Central Visual Pathways”. *Trends in Neuroscience*, 18(10):451–457, 1995.
- [Fleet and Langley, 1995] D. J. Fleet and K. Langley. “Recursive Filters for Optical Flow”. *IEEE-PAMI*, 17(1):61–67, 1995.
- [Florack, 1993] L. M. J. Florack. *The Syntactical Structure of Scalar Images*. PhD thesis, Dept. Med. Phys. Physics, Univ. Utrecht, NL-3508 Utrecht, Netherlands, 1993.
- [Florack *et al.*, 1992] L. M. J. Florack; B. M. ter Haar Romeny; J. J. Koenderink, and M. A. Viergever. “Families of Tuned Scale-Space Kernels”. In *ECCV'92*, pages 19–23, 1992.
- [Hille and Phillips, 1957] E. Hille and R. S. Phillips. *Functional Analysis and Semi-Groups*, volume XXXI. American Mathematical Society Colloquium Publications, 1957.
- [Koenderink and van Doorn, 1992] J. J. Koenderink and A. J. van Doorn. “Generic neighborhood operators”. *IEEE-PAMI*, 14(6):597–605, 1992.
- [Koenderink, 1984] J. J. Koenderink. “The structure of images”. *Biol. Cyb.*, 50:363–370, 1984.
- [Koenderink, 1988] J. J. Koenderink. “Scale-Time”. *Biol. Cyb.*, 58:159–162, 1988.
- [Lindeberg and Fagerström, 1996] T. Lindeberg and D. Fagerström. “Scale-Space with causal time direction”. In *ECCV'96*, volume 1064, pages 229–240, 1996.
- [Lindeberg, 1990] T. Lindeberg. “Scale-Space for Discrete Signals”. *IEEE-PAMI*, 12(3):234–254, 1990.
- [Lindeberg, 1994a] T. Lindeberg. “On the Axiomatic Foundations of Linear Scale-Space: Combining Semi-Group Structure with Causality vs. Scale Invariance”. Technical Report ISRN KTH/NA/P--94/20--SE, KTH, Stockholm, Sweden, 1994. Extended version to appear in J. Sporring *et al.* (eds.) *Gaussian Scale-Space Theory*, Kluwer, May 1996.
- [Lindeberg, 1994b] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer, Netherlands, 1994.
- [Lindeberg, 1996] T. Lindeberg. “Linear spatio-temporal scale-space”. Technical report, KTH, 1996.
- [Witkin, 1983] A. P. Witkin. “Scale-space filtering”. In *8th IJCAI*, pages 1019–1022, 1983.

- [Young, 1987] R. A. Young. "The Gaussian derivative model for spatial vision: I. Retinal mechanisms". *Spatial Vision*, 2:273-293, 1987.
- [Yuille and Poggio, 1986] A. L. Yuille and T. A. Poggio. "Scaling Theorems for Zero-Crossings". *IEEE-PAMI*, 8:15-25, 1986.

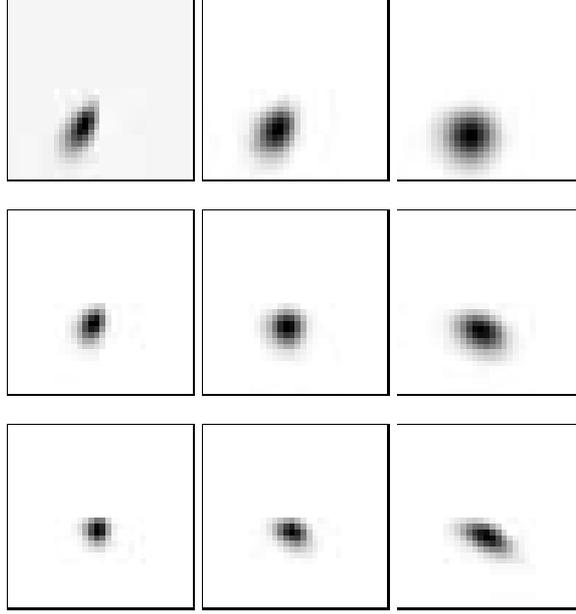


Fig. 1. Discrete velocity-adapted spatio-temporal filters in the 1+1-D spatio-temporal scale-space obtained by varying the eigenvalues λ_{xx} and λ_t in the covariance matrix (27) for a given velocity \dot{x} . For each receptive field, the horizontal dimension represents space and the vertical dimension time (with the origin at the image center). Here, it should be noted that the filter in the upper left corner violates the positivity conditions, which results in a negative lobe. (For this filter, the spatial translation $C_x = \dot{x} C_t$ is too large relative to the spatial smoothing C_{xx} .) The parameter settings were $\lambda_{xx} = 4, 8, 16$ from left to right, $\lambda_t = 4, 8, 16$ from bottom to top.

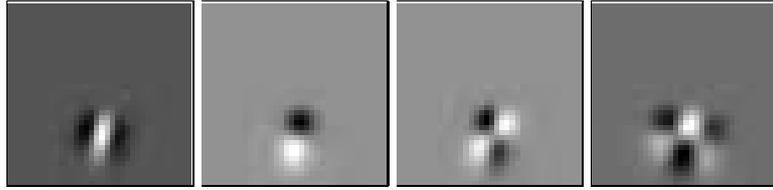
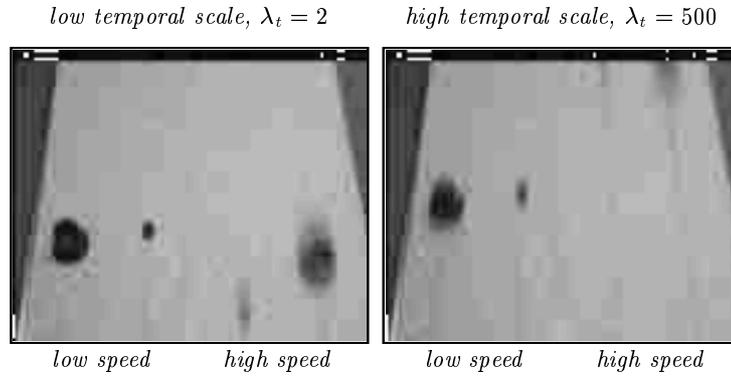


Fig. 2. Examples of discrete spatio-temporal derivative approximations obtained by applying (discrete approximations to) directional derivative operators in xt -space to a zero-order velocity-adapted spatio-temporal receptive field g obtained from (27). With $\delta_{\tilde{x}}$, $\delta_{\tilde{x}\tilde{x}}$ and $\delta_{\tilde{t}}$ denoting first- and second-order directional derivative operators in *tilted* xt -space (with orientation given by $\phi = \arctan \dot{x}$), the figures show from left to right: $\delta_{\tilde{x}\tilde{x}}g$, $\delta_{\tilde{t}}g$, $\delta_{\tilde{x}\tilde{t}}g$ and $\delta_{\tilde{x}\tilde{x}\tilde{t}}g$. (Parameter settings: $\lambda_{xx} = 8$, $\lambda_t = 16$, $\dot{x} = 0.25$.)

Spatio-temporal scale-space representation (at fine spatial scales)



Spatio-temporal derivatives at different spatio-temporal scales

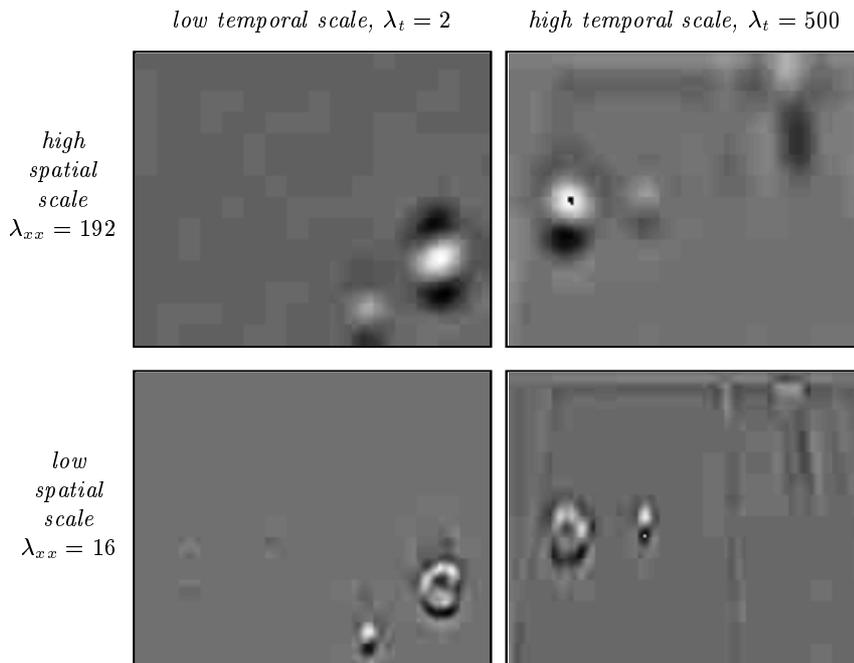


Fig. 3. Illustration of how spatio-temporal events can be perceived in different ways depending on the scale of observation. This figure shows four combinations of spatial and temporal scales in a separable spatio-temporal scale-space computed from a scene where two small size objects and two large size objects move at low and high speeds. Observe how we for different combinations of speeds and object sizes obtain qualitatively different types of responses at different spatio-temporal scales. (These results have been obtained from a time-recursive approximation (Lindeberg and Fagerström 1996) of the scale-space concept presented in this article.)