# Scale-Space with Casual Time Direction

*Tony Lindeberg and Daniel Fagerström*

Computational Vision and Active Perception Laboratory (CVAP)*
KTH (Royal Institute of Technology), Stockholm, Sweden.

**Abstract**

This article presents a theory for multi-scale representation of *temporal data.* Assuming that a real-time vision system should represent the incoming data at different time scales, an additional causality constraint arises compared to traditional scale-space theory—we can only use what has occurred in the past for computing representations at coarser time scales. Based on a previously developed scale-space theory in terms of *non-creation of local maxima with increasing scale*, a complete classification is given of the scale-space kernels that satisfy this property of non-creation of structure and *respect the time direction as causal.* It is shown that the cases of continuous and discrete time are inherently different.

For continuous time, there is no non-trivial time-causal semi-group structure. Hence, the time-scale parameter *must* be discretized, and the only way to construct a linear multi-time-scale representation is by (cascade) convolution with truncated exponential functions having (possibly) different time constants. For discrete time, there is a canonical semi-group structure allowing for a continuous temporal scale parameter. It gives rise to a *Poisson-type temporal scale-space.* In addition, geometric moving average kernels and time-delayed generalized binomial kernels satisfy temporal causality and allow for highly efficient implementations.

It is shown that *temporal derivatives* and derivative approximations can be obtained directly as *linear combinations* of the temporal channels in the multi-time-scale representation. Hence, to maintain a representation of temporal derivatives at multiple time scales, there is no need for other time buffers than the temporal channels in the multi-time-scale representation.

The framework presented constitutes a useful basis for expressing a large class of algorithms for computer vision, image processing and coding.

## 1 Introduction

The notion of multi-scale representation is essential when dealing with measured data, such as images. Philosophically, this need arises from the fact that we perceive real-world structures as meaningful entities only over certain ranges of scale. Traditionally, multi-scale concepts 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 1994)

*The support from the Swedish Research Council for Engineering Sciences, TFR, and the Esprit-NSF collaboration Diffusion is gratefully acknowledged. *Address:* KTH, NADA, S-100 44 Stockholm, Sweden. *Email:* tony@bion.kth.se, danielf@bion.kth.se. http://www.bion.kth.se.

have been developed over a spatial domain, in which data are available in all directions. Most works have avoided the constraints arising from the fact that time runs in a special direction and a genuine real-time vision cannot access the future—only what has occurred in the past can be used for generating representations at different time scales. An early suggestion for how to treat time in a multi-scale context was given by (Koenderink 1988), who proposed to transform the time axis so as to map the present moment to the unreachable infinity. In the transformed domain, he then applied the traditional scale-space concept by Gaussian convolution. The subject of this article is to reconsider the problem of constructing a multi-time-scale representation from an axiomatic viewpoint.

## 2 Continuous and discrete scale-space kernels: Review

A fundamental requirement when constructing a multi-scale representation is that the transformation from a fine scale to a coarser scale should constitute a simplification in the sense that fine-scale image structures should be successively suppressed. In the literature on traditional (spatial) scale-space representation, this property has been formalized in different ways. A noteworthy coincidence is that several different ways of choosing *scale-space axioms* lead to the Gaussian kernel as the unique choice.

In this article, we shall follow the scale-space formulation in (Lindeberg 1990, 1994) based on non-creation of local extrema (zero-crossings) with increasing scale. As shown in the abovementioned references, the class of convolution operators satisfying this requirement can be completely classified based on classical results by (Schoenberg 1953) (see also (Karlin 1968)). Besides translation and rescaling, there are two primitive types of linear and shift-invariant smoothing transformations in the continuous case:

- convolution with *Gaussian kernels*,

$$h(\xi) = e^{-\gamma\xi^2}, \tag{1}$$

- convolution with *truncated exponential functions*,

$$h(\xi) = \begin{cases} e^{-\xi/|\mu|} & \xi \geq 0, \\ 0 & \xi < 0, \end{cases} \qquad h(\xi) = \begin{cases} e^{\xi/|\mu|} & \xi \leq 0, \\ 0 & \xi > 0, \end{cases} \tag{2}$$

Correspondingly, in the discrete case, there are besides rescaling and translation, three primitive types of smoothing transformations (where $f_{out} = h * f_{in}$):

- two-point weighted averaging or *generalized binomial smoothing*,

$$\begin{aligned} f_{out}(x) &= f_{in}(x) + \alpha_i\, f_{in}(x-1) \qquad (\alpha_i \geq 0), \\ f_{out}(x) &= f_{in}(x) + \delta_i\, f_{in}(x+1) \qquad (\delta_i \geq 0), \end{aligned} \tag{3}$$

- moving average or *first-order recursive filtering*,

$$\begin{aligned} f_{out}(x) &= f_{in}(x) + \beta_i\, f_{out}(x-1) \qquad (0 \leq \beta_i < 1), \\ f_{out}(x) &= f_{in}(x) + \gamma_i\, f_{out}(x+1) \qquad (0 \leq \gamma_i < 1), \end{aligned} \tag{4}$$

- *infinitesimal smoothing* described by the generating function

$$H_{semi-group}(z) = e^{t(az^{-1}+bz)}. \qquad (5)$$

In the symmetric case, $a = b = \alpha/2$, this transformation corresponds to convolution with the *discrete analogue of the Gaussian kernel*,

$$T(n;\ \sigma^2) = e^{-\alpha\sigma^2} I_n(\alpha\sigma^2), \qquad (6)$$

where $I_n$ are the modified Bessel functions of integer order.

Among these *scale-space kernels*, we recognize the continuous Gaussian kernel $g(x;\ \sigma^2)$ and its discrete analogue $T(n;\ \sigma^2)$, which arise as unique symmetric choices if the scale parameter is required to be continuous and a semi-group structure is imposed (Lindeberg 1990, 1994). The generalized binomial kernels provide a natural basis for constructing pyramid representations (Burt 1981; Crowley 1981), whereas recursive filters can be used for efficient implementations of smoothing operations (Deriche 1987).

## 3 Time-causal scale-space kernels

The review in the previous section is general and does not take the specific nature of the time direction into account. For scale-space kernels treating the time direction as causal, an obvious requirement is that only function values in the past can be accessed. Hence, the kernels must satisfy $h(t) = 0$ when $t < 0$. Here, we shall analyse the implications of imposing this constraint on scale-space kernels in the continuous and discrete domains.

*Continuous time.* An immediate consequence of the classification of semi-groups of continuous scale-space kernels (the Gaussian kernel is unique) is that we cannot preserve a continuous semi-group structure with respect to the time-scale parameter if the time direction is to be treated as causal. Hence, the *only* choice is to discretize the time-scale parameter. The only primitive scale-space kernels with one-sided support are the truncated exponential functions. After normalization to unit $L_1$-norm they can be written

$$h_{exp}(t;\ \mu) = \frac{1}{\mu} e^{-t/\mu} \qquad (t > 0). \qquad (7)$$

By varying $\mu$, we obtain first-order filters having different time constants. The classification of continuous scale-space kernels implies that a kernel is a time-causal scale-space kernel if and only if it can be decomposed into a sequence of convolutions with such filters. Hence, the architecture on a time-scale representation imposed by this construction is a set of *first-order recursive filters in cascade*, each having a (possibly) different time constants $\mu_i$. Such a filter has mean value $M(h_{composed}(\cdot;\ \mu)) = \sum_{i=1}^{\infty} \mu_i$, variance $\lambda = V(h_{composed}(\cdot;\ \mu)) = \sum_{i=1}^{\infty} \mu_i$, and a (bilateral) Laplace transform of the form

$$H_{composed}(s;\ \mu) = \int_{t=-\infty}^{\infty} (*_{i=1}^{\infty} h_{exp}(t;\ \mu_i))\ e^{-st}\, dt = \prod_{i=1}^{\infty} \frac{1}{1 + \mu_i s}. \qquad (8)$$

If we in analogy with a semi-group requirement, require the transformation from any fine-scale representation to any coarser-scale representation to be a scale-space transformation, then the only possibility is that all the (discrete) scale levels in the multi-time scale representation can be generated by a cascade of such truncated exponential filters.

*Discrete time.* For discrete time sampling, the discrete analogue of the truncated exponential filters are the first-order recursive filters (4). With normalization to unit $l_1$-norm, and $\mu = \beta/(1-\beta)$, their generating functions can be written

$$H_{geom}(z) = \frac{1}{1 - \mu\,(z-1)}, \tag{9}$$

Computationally, these filters are highly efficient, since only few arithmetic operations and no additional time buffering are required to compute the output at time $t+1$ given the output at time $t$. In normalized form, the recursive smoothing operation is

$$f_{out}(t) - f_{out}(t-1) = \frac{1}{1+\mu}\,(f_{in}(t) - f_{out}(t-1)). \tag{10}$$

In analogy with the case of continuous time, a natural way to combine these filters into a discrete multi-time-scale representation is by cascade coupling. The mean and variance of such a composed filter are $M(h_{geom}(\cdot;\ \mu)) = \sum_{i=1}^{\infty} \mu_i$ and $\lambda = V(h_{geom}(\cdot;\ \mu)) = \sum_{i=1}^{\infty} \mu_i^2 + \mu_i$. In the case of discrete time, we can also observe that the generalized binomial kernels (3) indeed satisfy temporal causality, if combined with a suitable time delay. In this respect, there are more degrees of freedom in the case of discrete time sampling.

*Time-causal semi-group structure exists only for discrete time.* The case of discrete time it also special in the sense that a semi-group structure is, indeed, compatible with temporal causality. If we let $q_{-1} = 0$ and $q_1 = \lambda$ in (5) and multiply by the normalization factor $\exp(-\lambda)$, we obtain a generating function of the form $P(z;\ \lambda) = e^{\lambda(z-1)}$ (Lindeberg 1996) with associated filter coefficients

$$p(n;\ \lambda) = e^{-\lambda}\,\frac{\lambda^n}{n!}. \tag{11}$$

This filter corresponds to a Poisson distribution and the kernel $p$ will be referred to as the *Poisson kernel*. Intuitively, it can be interpreted as the limit case of repeated convolution of kernels of the form (9) with time constants $\mu = \lambda/m$:

$$\lim_{m\to\infty} \left(H_{geom}(z;\ \frac{\lambda}{m})\right)^m = \lim_{m\to\infty} \frac{1}{(1 - \frac{\lambda}{m}(z-1))^m} = P(z;\ \lambda). \tag{12}$$

Such a kernel has mean $M(p(\cdot;\ \lambda)) = \lambda$, and variance $V(p(\cdot;\ \lambda)) = \lambda$. From the ratio $\frac{p(n+1;\ \lambda)}{p(n;\ \lambda)} = \frac{\lambda}{n+1}$, it can be seen for $\lambda < 1$ the filter coefficients decrease monotonically for $n \geq 0$, while for $\lambda > 1$ there is a local maximum at the smallest integer less than $\lambda$: $n = [\lambda] > 0$. Similarly, there are two inflexion points at $n \approx \lambda + \frac{1}{2} \pm (\lambda + \frac{1}{4})^{1/2}$. Concerning the qualitative behaviour, it also well-known from statistics that the Poisson distribution approaches the normal distribution with increasing standard deviation (see figure 1).

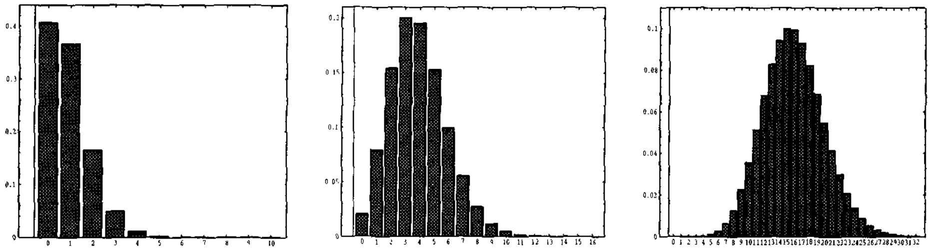

**Figure 1:** Graphs of the Poisson kernels for $\lambda = 0.9$, $3.9$ and $15.9$.

Under variations of $\lambda$, the Poisson kernel satisfies $\partial_\lambda p(n;\ \lambda) = -(p(n;\ \lambda) - p(n-1;\ \lambda))$. Thus, if we define a multi-time-scale representation $L\colon \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ of a discrete signal $f\colon \mathbb{R} \to \mathbb{R}$, having a *continuous time-scale parameter*, by

$$L(t;\ \lambda) = \sum_{n=-\infty}^{\infty} p(n;\ \lambda)\, f(t-n), \tag{13}$$

this representation satisfies the first-order semi-differential equation $\partial_\lambda L = -\delta_- L$, where $\delta_-$ denotes the backward difference operator $\delta_- L(t;\ \lambda) = L(t;\ \lambda) - L(t-1;\ \lambda)$. Hence, in contrast to multi-scale representations of the spatial domain, for which derivatives with respect to scale are related to second-order derivatives/differences in the spatial domain, temporal scale derivatives are here related to *first-order* temporal differences.

Note that a corresponding time-causal structure does not exist for continuous signals. If we apply the same way of reasoning and compute the limit case of primitive kernels of the form (8) for which all $\lambda_i$ are equal, we obtain the trivial semi-group corresponding to translations of the time axis by a time delay $\lambda$.

# 4 Temporal scale-space and temporal derivatives

So far, we have shown how general constraints concerning non-creation of local extrema with increasing scale combined with temporal causality restrict the class of operations that can be used for generating multi-scale representations corresponding to temporal integration over different time scales. When to use these results in practice, an obvious issue concerns how to distribute a (finite) set of discrete scale levels over scales and how to compute temporal derivatives (or derivative approximations) at different time scales.

*Distribution of scale levels.* A useful property of the Poisson-type scale-space (13) is that there is no need for selecting scale levels *in advance*. If we have access to all data in the past, we can compute the temporal scale-space representation at any scale. Assuming that a vision system is to operate at a set of $K$ temporal scales, a natural *a priori* distribution of these scale levels $\lambda_k$ between some minimum scale $\lambda_{min}$ and some maximum scale $\lambda_{max}$ is according to a geometric series $\lambda_k = \gamma^k \lambda_{min}$ where $\gamma^K = \lambda_{max}/\lambda_{min}$.

Concerning the multi-time scale representations having a discrete time-scale parameter, let us assume that a *minimal design* is chosen, in the sense that the transformation between adjacent scales is always of the form (7) or (9). Since variances are additive under convolution, it follows that the time constants between adjacent scales should satisfy $\lambda_k = \lambda_{k-1} + \mu_k$ for continuous signals and $\lambda_k = \lambda_{k-1} + \mu_k + \mu_k^2$ for discrete signals.

*Temporal scale-space derivatives in the continuous case.* Given a continuous signal $f$, assume that a level $k$ in a time-scale representation

$$L(\cdot;\ \lambda_k) = (*_{i=1}^{k} h_{exp}(t;\ \mu_i)) * f \tag{14}$$

has been computed at some temporal scale $\lambda_k$ by cascade filtering with a set of $k$ truncated exponential filters with time constants $\mu_i$. From this representation, a *temporal scale-space derivative* of order $r$ at scale $\lambda_k$ is defined by

$$L_{t^r}(\cdot;\ \lambda_k) = \partial_{t^r} L(\cdot;\ \lambda_k) = \left(\partial_{t^r}(*_{i=1}^{k} h_{exp}(t;\ \mu_i))\right) * f, \tag{15}$$

and the Laplace transform of the composed (equivalent) derivative kernel is

$$H^{(r)}_{composed}(s;\ \lambda_k) = s^r \prod_{i=1}^{k} \frac{1}{1 + \mu_i\, s}. \tag{16}$$

For this kernel to have a net integration effect (well-posed derivative operators), an obvious requirement is that the total order of differentiation should not exceed the total order of integration. Thereby, $r < k$ is a necessary requirement. As a consequence, the transfer function must have finite $L_2$-norm.

A useful observation in this context is that these *temporal scale-space derivatives can be equivalently computed from differences between the temporal channels*. Assume, for simplicity, that all $\mu_i$ are different in (16). Then, a decomposition of $H^{(r)}_{composed}$ into a sum of $r$ such transfer functions at finer scales

$$H^{(r)}_{composed}(s;\ \lambda_k) = \sum_{i=k-r}^{k} B_i\, H_{composed}(s;\ \lambda_i) \tag{17}$$

shows that the weights $B_i$ are given as the solution of a triangular system of equations provided that the necessary condition $r < k$ is satisfied

$$\frac{(-1)^r}{\mu_i^r} \prod_{j=i+1}^{k} \frac{1}{(1 - \mu_j/\mu_i)} = B_i + \sum_{\nu=i+1}^{k} B_\nu \prod_{j=i+1}^{\nu} \frac{1}{(1 - \mu_j/\mu_i)} \qquad (k - r \le i \le k).$$

Hence, each temporal derivative can be computed as a linear combination of the representations at finer time scales. Moreover, the Laplace transforms of the equivalent derivative computation kernels satisfy the recurrence relation

$$H^{(r)}_{composed}(s;\ \lambda_k) = -\frac{1}{\mu_k}\left(H^{(r-1)}_{composed}(s;\ \lambda_k) - H^{(r-1)}_{composed}(s;\ \lambda_{k-1})\right). \quad (18)$$

In other words, higher-order temporal derivatives can be computed as finite differences of lower-order derivatives (analogous to finite difference operators in the spatial domain). Derivative computations will thus be highly efficient.

*Temporal derivative approximations in the discrete case.* In (Lindeberg and Fagerström 1996) it is shown that a corresponding structure holds in the discrete case, for multi-scale temporal derivative approximations obtained by applying (either symmetric or non-symmetric) central difference operators to the discrete

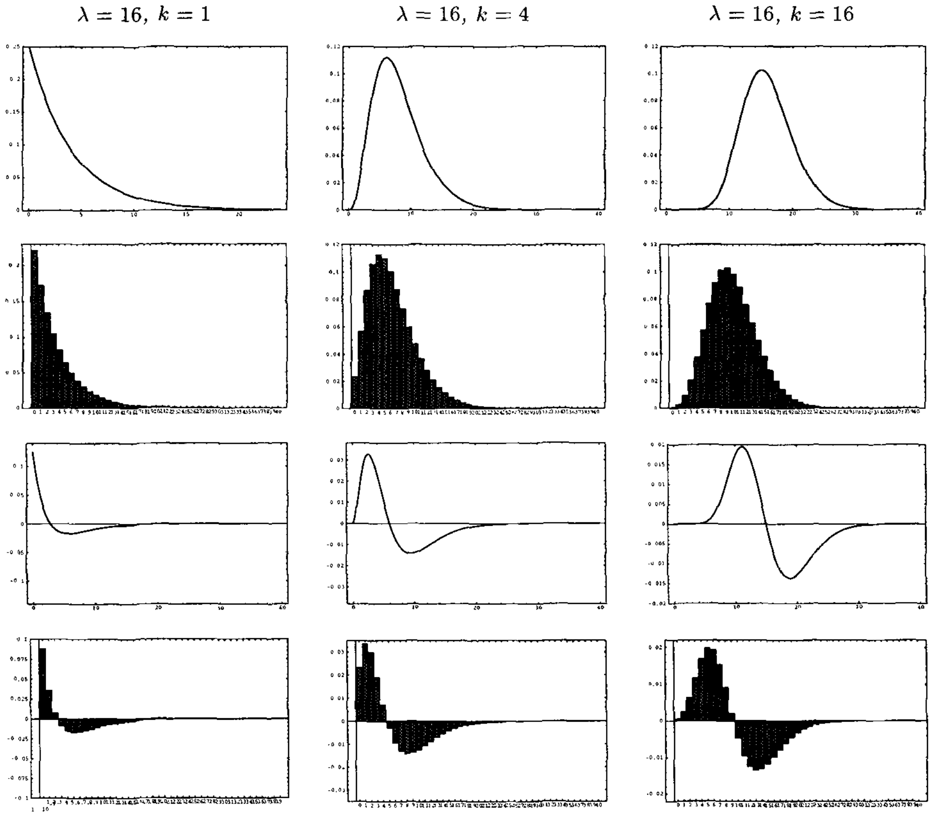


**Figure 2:** Graphs of equivalent smoothing kernels and first-order derivative (approximation) kernels in the continuous and discrete cases, respectively, for $k$ cascade coupled smoothing steps in which all the primitive time constants $\mu_i$ are equal. (Here, $\mu_i$ have been determined from $k$ such that the variance $\lambda$ is the same for all smoothing kernels.)

multi-time-scale representation constructed by cascade convolution with first-order recursive filters of the form (10).

*Temporal derivatives from linear combinations of temporal channels* This special structure is highly useful for practical purposes, since it makes explicit construction of temporal derivative kernels unnecessary. In other words, no other time buffers are necessary for computing temporal scale-space derivatives than the actual channels in the multi-time-scale representation. An intuitive explanation of why this is possible is that the different temporal channels, which represent the incoming data at different time-scales, have different effective temporal delays. Besides the primary effect of producing an integrated representation over a certain time-scale, each such channel serves as a temporal buffer.

*Kernel graphs and trade-off issues.* Figure 2 shows graphs of equivalent (continuous and discrete) convolution kernels for a few combinations of the number of recursive filters in cascade, $k$, and the individual time constants, $\mu_i$.

The parameter values have been chosen such that the variance $\lambda$ of the smoothing kernel is the same for all filters. Hence, they represent different ways of computing the representation at a certain scale.

As can be seen, the kernels are discontinuous if $r \geq k - 1$, whereas the degree of smoothness increases with $k$. To guarantee a certain minimum degree of temporal smoothness at the finest temporal scale, it can therefore (depending on the external sampling conditions) be useful to precede the recursive temporal multi-scale representations by a common pre-smoothing step (such as a few steps of recursive filtering or time-delayed binomial smoothing). For a more detailed analysis, including frequency properties, see (Lindeberg and Fagerström 1996).

## 5 Spatio-temporal scale-space

When to combine these multi-time-scale representations with a spatial representation for dealing with time-varying images, let us first treat space and time as separable dimensions. This is a natural assumption in the absence of further information (such as velocity information). The spatio-temporal scale-space representation we then obtain is the Cartesian product of the spatial and temporal scale-space representations, and is parameterized by a spatial scale parameter $\sigma^2$ and a temporal scale parameter $\lambda$.

Depending on whether the spatial domain $\mathbb{S}$ is continuous or discrete, and correspondingly for the temporal domain $\mathbb{T}$ as well as the domains $\Sigma$ and $\Lambda$ of the spatial and temporal scale parameters, we then obtain one out of twelve possible types of spatio-temporal scale-space representations (see figure 3).

Denote the transfer function of the spatial smoothing kernel by $H_{\mathbb{S}}(u;\ \sigma^2)$ and the transfer function of the temporal smoothing kernel by $H_{\mathbb{T}}(v;\ \lambda)$. Then, the transfer function for mapping a spatio-temporal signal $f\colon \mathbb{S}^N \times \mathbb{T} \to \mathbb{R}$ to its *spatio-temporal scale-space representation* $L\colon \mathbb{S}^N \times \mathbb{T} \times \Sigma \times \Lambda \to \mathbb{R}$ is given by

$$H(u, v;\ \sigma^2, \lambda) = H_{\mathbb{S}}(u;\ \sigma^2)\, H_{\mathbb{T}}(v;\ \lambda). \tag{19}$$

| | | *Spatial domain* $\mathbb{S}$ | |
|---|---|---|---|
| | | Continuous | Discrete |
| *Spatial scale* $\Sigma$ | Continuous | Continuous Gaussian | Discrete Gaussian |
| | Discrete | + Trunc. exp. | + Binom. and geom. averaging |

| | | *Temporal domain* $\mathbb{T}$ | |
|---|---|---|---|
| | | Continuous | Discrete |
| *Temporal scale* $\Lambda$ | Continuous | —— | Poisson kernel |
| | Discrete | Trunc. exp. | + Binom. and geom. averaging |

**Figure 3:** Scale-space kernels satisfying non-creation of local extrema with increasing scale in the cases of a continuous/discrete domain, a continuous/discrete scale parameter, and a spatial/temporal domain without or with preferred direction.

When implementing this operation in practice, the linearity implies that the spatial and temporal smoothing operators commute. For time-recursive temporal smoothing, it will therefore be more efficient to compute the spatial scale-space representation at the finest temporal scale, and then apply subsequent temporal smoothing to each spatial scale layer in this representation. If there is a common temporal smoothing component for all temporal scales (such as time-delayed binomial smoothing to reduce temporal aliasing due to poor temporal sampling), it will be computationally more efficient to apply such filters before constructing the spatial scale-space representation. Concerning temporal derivatives, it was shown that these can be computed by linear combinations of the temporal channels at each spatial scale. Before or after this step, finite difference operators can be applied to compute spatial derivative approximations (see figure 4).

In summary, this spatio-temporal scale-space concept leads to a visual front-end model, which at every time moment outputs a set of spatio-temporal derivatives at different spatio-temporal scales. Concerning time buffering, there is essentially no need for the visual front-end to represent the past in any other ways than as the temporal channels in the multi-time-scale representation. Hence, for two-dimensional image data, we obtain a visual front-end, which over time maintains a four-dimensional representation of the current (delayed) moment. This data set constitutes one time slice of the five-dimensional spatio-temporal representation of the complete history of the visual observer.

Figure 5 shows an example of multi-scale spatio-temporal image descriptors computed in this way. It shows second-order temporal derivatives computed from an image sequence for a number of different values of the spatial and temporal scale parameters. Observe how qualitatively different types of responses are obtained at the different spatio-temporal scales.

A more extensive treatment of this subject is presented in (Lindeberg 1996), including scale-space properties, necessity results and the non-separable case.

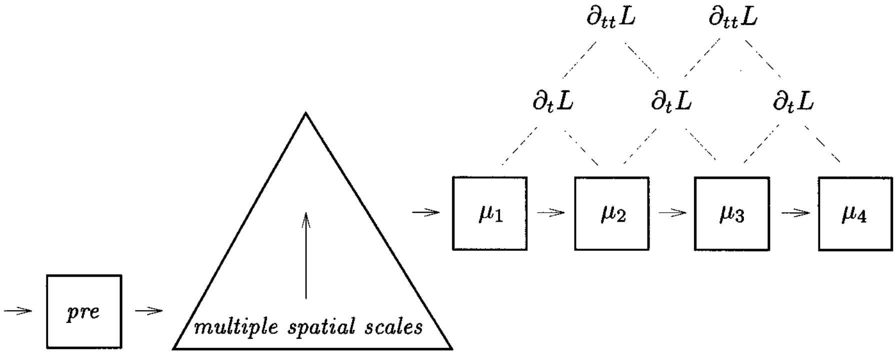


**Figure 4:** The composed architecture of the resulting spatio-temporal visual front-end consists of the following types of processing steps. (i) Optional temporal preprocessing. (ii) Spatial multi-scale representation, *e.g.* a pyramid or a scale-space representation. (iii) One set of recursive temporal smoothing stages associated with each spatial scale. (iv) Temporal derivatives from linear combinations of temporal channels. (v) Spatial derivative approximations from finite spatial differences (not shown in this figure).

# 6 Summary and discussion

We have presented a theory for how the linear scale-space concept can be extended to the temporal domain. The theory is complete in the sense that it provides a complete catalogue of all linear scale-space concepts that satisfy temporal causality in the cases of continuous *vs.* discrete time as well as continuous *vs.* discrete scale. Essentially, there are three main categories.

The construction started from similar scale-space axioms as have been used for deriving the uniqueness of the Gaussian kernel in the spatial domain, namely linearity, shift invariance, symmetry and non-creation of maxima (zero-crossings) with increasing scale. In the case of a continuous scale parameter, the latter assumption corresponds to a semi-group structure. Then, we replaced the symmetry condition by the essential requirement that the time direction should be treated as causal, and only what has occurred in the past may be used as input for computing representations at coarser time scales.

A kernel satisfying these properties was termed a time-causal scale-space kernel, and a complete classification was given for continuous and discrete time domains. For continuous time, the only primitive time-causal kernels are truncated exponential kernels corresponding to first-order integration over time. The discrete correspondences to these are geometric moving average kernels. In the discrete domain, also time-shifted binomial kernels satisfy temporal causality.

In the case of discrete time, and only in this case, there is a non-trivial time-causal semi-group structure. It corresponds to convolution with Poisson kernels, and can be regarded as the canonical model of a temporal scale-space, since it is the only time-causal scale-space having a continuous scale parameter.

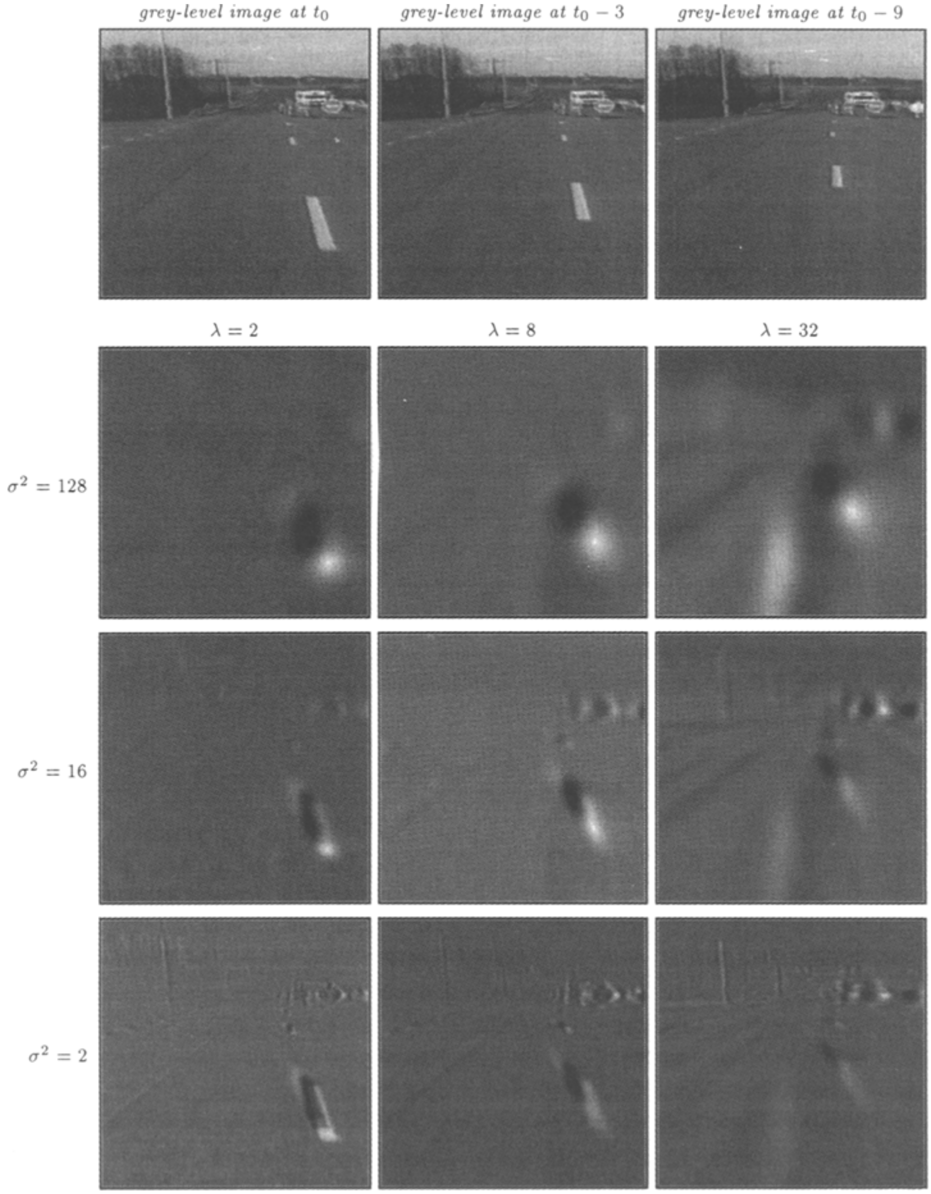


**Figure 5:** Second-order temporal scale-space derivatives computed for a few combinations of spatial scales and temporal scales. The top row shows three frames from the image sequence, whereas the following rows show spatio-temporal data for $\sigma^2 = 2$, 16 and 128 (from the bottom to the top) and $\lambda = 2$, 8 and 32 (from the left to the right).

We analysed derivative operators and derivative approximations with respect to their scale-space properties. Specifically, we made the important observation that the temporal channels themselves contain sufficient information for computing temporal derivatives at the current moment. Hence, there is no need for additional time buffering as would be needed if computing temporal derivative approximations by explicit finite differences.

More generally, the time recursive properties of the smoothing kernels corresponding to a discrete scale parameter imply that it is sufficient for the visual front-end to maintain a representation over time that corresponds to a four-dimensional slice of the entire five-dimensional spatio-temporal scale-space. This dimensionality reduction is of crucial importance, since it substantially reduces computational and hardware requirements.

An attractive property of the presented theory is that it leads to a conceptually very simple architecture (illustrated in figure 4) and allows for computationally highly efficient implementations. To update the temporal information to the next time moment (according to equation (10)) it is sufficient to perform one multiplication and two additions per pixel and spatio-temporal channel. Whereas recursive filters are common in signal processing and constitute a natural choice on an *ad hoc* basis, an important result of this treatment is that this design can be derived by necessity from first principles.

## References


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

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

R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *IJCV*, 1:167–187, 1987.

L. M. J. Florack. *The Syntactical Structure of Scalar Images.* PhD thesis. Dept. Med. Phys. Physics, Univ. Utrecht, NL-3508 Utrecht, Netherlands, 1993.

N. L. Johnson and S. Kotz. *Discrete Distributions.* Houghton Mifflin, Boston, 1969.

S. Karlin. *Total Positivity.* Stanford Univ. Press, 1968.

J. J. Koenderink and A. J. van Doorn. Generic neighborhood operators. *IEEE-PAMI*, 14(6):597–605, 1992.

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

J. J. Koenderink. Scale-time. *Biol. Cyb.*, 58:159–162, 1988.

T. Lindeberg. Scale-space for discrete signals. *IEEE-PAMI*, 12(3):234–254, 1990.

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

T. Lindeberg. Linear spatio-temporal scale-space. in preparation, 1996.

T. Lindeberg and D. Fagerström. Scale-space with causal time direction. Tech. rep. ISRN KTH/NA/P–96/04–SE, NADA, KTH, Stockholm, Sweden, jan 1996.

I. J. Schoenberg. On smoothing operations and their generating functions. *Bull. Amer. Math. Soc.*, 59:199–230, 1953.

A. P. Witkin. Scale-space filtering. In *8th IJCAI*, pages 1019–1022, 1983.

A. L. Yuille and T. A. Poggio. Scaling theorems for zero-crossings. *IEEE-PAMI*, 8:15–25, 1986.