Journal of Mathematical Imaging and Vision (2026) 68:19
https://doi.org/10.1007/s10851-026-01296-8

Hybrid Lie Semi-Group and Cascade Structures for the Generalized Gaussian Derivative Model for Visual Receptive Fields

Tony Lindeberg[1]

Received: 19 September 2025 / Accepted: 3 April 2026
© The Author(s) 2026

Abstract

Because of the variabilities of real-world image structures under the natural image transformations, that arise when observing similar objects or spatio-temporal events under different viewing conditions, the receptive field responses computed in the earliest layers of the visual hierarchy may be strongly influenced by such geometric image transformations. One way of handling this variability is by basing the vision system on covariant receptive field families, which expand the receptive field shapes over the degrees of freedom in the image transformations. This paper addresses the problem of deriving relationships between spatial and spatio-temporal receptive field responses obtained for different values of the shape parameters in the resulting multi-parameter families of receptive fields. For this purpose, we derive both (i) infinitesimal relationships, roughly corresponding to a combination of notions from semi-groups and Lie groups, as well as (ii) macroscopic cascade smoothing properties, which describe how receptive field responses at coarser spatial and temporal scales can be computed by applying smaller support incremental filters to the output from corresponding receptive fields at finer spatial and temporal scales, structurally related to the notion of Lie algebras, although with directional preferences. For the receptive field models based on spatio-temporal smoothing using a sole combination of affine Gaussian smoothing kernels over image space with non-causal temporal Gaussian kernels over the temporal domain, we derive reasonably complete results in this respect, by exploiting the specific structure of the joint 2+1-D Gaussian spatio-temporal model. This permits characterizations in terms of higher-dimensional generalizations of the notion of Hermite polynomials, as well as characterizations in terms of joint spatio-temporal covariance matrices. For the time-causal spatio-temporal receptive field model, where the temporal smoothing is instead performed using the time-causal limit kernel, we derive less complete results, however, still highly useful for the special case when the velocity parameters assume equal values in both the incremental convolution kernel as well as in the corresponding layers of the spatio-temporal scale-space representation. The presented results provide (i) a deeper understanding of the relationships between spatial and spatio-temporal receptive field responses for different values of the filter parameters, which can be used for both (ii) designing more efficient schemes for computing receptive field responses over populations of multi-parameter families of receptive fields, as well as (iii) formulating idealized theoretical models of the computations of simple cells in biological vision.

Keywords Receptive field · Filter bank · Filter parameters · Semi-group · Lie group · Lie algebra · Gaussian derivative · Image transformations · Simple cells · Scale space · Vision

1 Introduction

When a visual observer views objects or spatial–temporal events in the environment, the image data reaching the visual sensor may be subject to substantial variabilities caused by variabilities in the viewing conditions, as resulting from varying the distance, the viewing direction and the relative motion between the objects in the world and the observer. The influence of the resulting geometric image transformations will,

The support from the Swedish Research Council (contract 2022-02969) is gratefully acknowledged.

✉ Tony Lindeberg
tony@kth.se

[1] Computational Brain Science Lab, Department of Computational Science and Technology, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden

in turn, strongly affect the responses of the receptive fields in the early layers of the visual hierarchy.

To handle such variabilities in image structures, caused by variations in the viewing conditions, the notion of multi-parameter scale spaces has been proposed regarding the families of receptive fields (Lindeberg [47, 52]). Specifically, from the desirable property of provable covariance properties of the families of spatial and/or spatio-temporal receptive fields under the appropriate families of geometric image transformations, the shapes of the spatial or spatio-temporal receptive fields ought to be expanded over the degrees of freedom of the corresponding image transformations (Lindeberg [59]). Notably, such a notion has been proposed as a tentative model for variabilities of simple cells in the primary cortex of higher mammals (Lindeberg [60]).

If one in an idealized vision system is to take the view that a rich set of receptive field responses is to be computed for a rich variety of shape parameters of the receptive fields, one can raise the question if this could be done in a more efficient manner than computing each receptive field response separately. Similarly, one may ask how receptive field responses computed for different values of the filter parameters could be related. For example, for the pure spatial scale-space representation, defined from convolution with either isotropic or anisotropic affine Gaussian kernels, the underlying spatial smoothing filters obey cascade smoothing properties over spatial scales, implying that any representation at a coarser spatial scale can be computed by applying a filter or a set of filters to the representations at any finer scale.

For purposes of computer implementation, making use of that cascade smoothing property can substantially reduce the computational work, by making it possible to build the implementation based on a set of filtering operations of smaller spatial support compared to a naive implementation of each receptive field response, which thereby reduces the amount of computations. Similarly, by the special property of the spatial smoothing operation, the cascade smoothing property resulting from the semi-group property of the underlying either isotropic or anisotropic affine Gaussian smoothing kernels guarantees a simplifying property from finer to coarser level of scales, crucial for the formal definition of a simplifying property from finer to coarser scales in the corresponding spatial scale-space representation (Iijima [32], Koenderink [37], Koenderink and van Doorn [39], Lindeberg [43], [47], Weickert *et al.* [89]).

The subject of this paper is to generalize the above-mentioned ideas for the specific generalized Gaussian derivative model for visual receptive fields, initially derived in Lindeberg [47, 48] and then refined in Lindeberg [50, 52, 55, 59, 60]. Using the algebraic properties of the idealized models for visual receptive fields according to this framework, we will derive closed-form expressions for derivatives with respect to the filter parameters in this model as well as explicit macroscopic cascade relations between receptive field responses for different parameter settings, roughly corresponding to a combination of Lie group structures with multi-parameter semi-group structures, comprising 3 effective parameters in the purely spatial case and 4 or 6 effective parameters in the joint spatio-temporal case, although for special reasons of separating the degree of freedom corresponding to purely spatial scaling transformations, overparameterized over 4 or 7 parameters when using affine Gaussian kernels for the spatial smoothing operations.

This will result in (i) sets of infinitesimal generators and (ii) sets of macroscopic cascade smoothing properties over subsets of the parameter spaces, for which the evolution properties can only performed in the positive parameter directions. For other subsets in the parameter space, the corresponding relationships between receptive field responses for different parameter settings will, however, be bidirectional.

The intention behind these theoretical results to be presented is that they could be used (i) in network structures that combine responses computed for different values of the filter parameters in an actual either computer or biophysical implementation of receptive field responses, and for (ii) understanding the theoretical relationships between receptive field responses that are to be computed from different sets of filter parameters for receptive field models based on multi-parameter scale spaces.

Specifically, in relation to the recently proposed tentative model for variabilities of simple cells in the primary cortex of higher mammals in Lindeberg [60], the presented results provide a detailed theoretical foundation for how to realize the alternative conceptual model in that paper, where the receptive field responses corresponding to simple cells are not computed for all spatial and/or temporal scales, but instead because of the semi-group properties with respect to the spatial and the temporal scale parameters only computed at the finest spatial and/or temporal scales at each image position and temporal moment. Thereby, as considered as a possible design option for an idealized vision system, the image representations at higher layers in the visual hierarchy would not make use of explicit receptive field responses computed for simple cells for each spatial and/or temporal scale, but could instead compute equivalent receptive field responses at higher levels in the visual hierarchy, from the responses of the explicitly computed responses of simple cells as only implemented at the finest spatial and/or temporal scales.

Additionally, if an idealized vision system would be based on computing receptive field responses as corresponding to the output from simple cells, then the relationships between receptive field responses for different values of the filter parameters constitute a theoretical foundation for how such responses could be computed in a hierarchical network structure, where receptive field responses are computed in a way that constitutes an extension to the cascade smoothing properties over the spatial and/or the temporal scale parameters, to enable a computationally more efficient implementation.

Although we will in the following theoretical analysis not make use of any explicit Lie group or Lie algebra formalisms in the derivations of the relationships to be presented between receptive field responses for different parameter settings, the derived relations will reveal relationships very closely related to both the notions of differential Lie group structure and macroscopic Lie algebra structures of the corresponding representations for receptive field responses computed for different values of the filter parameters. An important distinction, however, is that the evolution properties over some of the parameters will correspond to unidirectional semi-groups instead of bidirectional groups, because the evolution properties of those parameters can only be performed in one direction.

1.1 Structure of this Article

The presentation is organized as follows: After describing relations to previous work in Sect. 2, Sect. 3 starts by reviewing the generalized Gaussian derivative model for visual receptive fields, including its covariance properties under important classes of geometric image transformations in terms of (i) spatial scaling transformations, (ii) spatial affine transformations, (iii) Galilean transformations and (iv) temporal scaling transformations. From the closedness properties of the receptive field responses under these classes of geometric image transformation, we specifically explain why the aim of being able to match the receptive field responses computed under different viewing conditions leads to the notion of multi-parameter filter banks of receptive fields, which motivates this study of theoretically showing how receptive field responses computed for different values of the filter parameters can be related. This section also defines the notation and the terminology in the way as it will be used in the forthcoming more technical sections.

Section 4 then begins the new theoretical work by deriving infinitesimal relationships between receptive field responses for the subclass of receptive field models that are based on spatial or spatio-temporal smoothing solely based on Gaussian kernels over image space or joint space-time. Section 5 then continues by deriving macroscopic relations between filter responses computed for different sets of filter parameters, based on either purely spatial or joint spatio-temporal cascade smoothing properties of the corresponding receptive field representations, defined from spatial or spatio-temporal smoothing based on solely using Gaussian smoothing kernels over image space or joint space-time.

Section 6 then complements with a set of less complete results for the time-causal theory of visual receptive fields, where the previous use of non-causal smoothing over the temporal domain is replaced by smoothing with the time-causal limit kernel aimed at real-world image data, for which the future cannot be accessed. Finally, Sect. 7 concludes with a summary and discussion.

2 Relations to Previous Work

Concerning theoretical modeling of visual receptive fields, the regular Gaussian derivative model was initially proposed by Koenderink and van Doorn [37–39] and used for modeling biological receptive fields by Young [94] and his co-workers in Young *et al.* [95, 96]. This regular Gaussian derivative model has also been used as a component in more developed models of biological vision by Lowe [62], May and Georgeson [64], Hesse and Georgeson [30], Georgeson *et al.* [23], Wallis and Georgeson [87], Hansen and Neumann [29], Wang and Spratling [88], Pei *et al.* [69], Ghodrati *et al.* [25], Kristensen and Sandberg [40], Abballe and Asari [1], Ruslim *et al.* [77] and Wendt and Faul [90].

The generalization into receptive field families based on multi-parameter scale spaces goes back to the formulation of affine Gaussian scale-space representation in Lindeberg and Gårding [61] and early formulations of discrete scale-space representations over either purely spatial or joint spatio-temporal domains in Lindeberg [44, 46]. These notions were then extended into the generalized Gaussian derivative model for either purely spatial or joint spatio-temporal domains in Lindeberg [47, 48, 52] and with additional extensions concerning provable covariance properties of receptive field families in Lindeberg [55, 59, 60].

The variabilities in receptive field shapes generated by the affine Gaussian scale space have been used for computing more accurate estimates of local surface orientation from monocular and binocular cues by Lindeberg and Gårding [61] and Rodríguez *et al.* [74], for computing affine invariant image features for image matching under wide baselines by Baumberg [10], Mikolajczyk and Schmid [65], Mikolajczyk *et al.* [66], Tuytelaars and van Gool [86], Lazebnik *et al.* [41], Rothganger *et al.* [75, 76], Lia *et al.* [42], Eichhardt and Chetverikov [21] and Dai *et al.* [18], for performing affine invariant segmentation by Ballester and González [3], for constructing affine covariant SIFT descriptors by Morel and Yu [67], Yu and Morel [97] and Sadek *et al.* [79], for modeling receptive fields in biological vision by Lindeberg [48, 52], for affine invariant tracking by Giannarou *et al.* [26], and for formulating affine covariant metrics by Fedorov *et al.* [22]. This rich variety of application domains clearly demonstrates how the use of variabilities in the shapes of the receptive fields, as implied by covariant receptive field representations (here with regard to spatial affine transformations), can be used for improving the performance of visual operations.

For modeling spatial receptive fields in vision, also Gabor models have been used by Marcelja [63], Jones and Palmer

[35, 36], Ringach [72, 73], Serre *et al.* [83], Sabatini *et al.* [78], Baspinar *et al.* [4, 8, 9], De and Horwitz [19] and others. The use of Lie symmetries for deriving receptive field shapes based on Gabor filters under the SE(2) group has been studied by Citti and Sarti *et al.* [16] and under the 2-D affine group by Sarti *et al.* [81]. Modeling of spatio-temporal receptive fields using Gabor functions has been proposed by Cocci *et al.* [17] and Barbieri *et al.* [5].

Both the complex-valued Gabor functions used in the Gabor model and the real-valued Gaussian functions underlying the Gaussian derivative model minimize the uncertainty relation (Barbieri *et al.* [4], Lindeberg [49]).

For handling the influence of geometric image transformations in terms of spatial scaling transformations in deep networks, provably scale covariant (also referred to as scale equivariant) network architectures have been developed by Worrall and Welling [92], Bekkers [11], Sosnovik *et al.* [84, 85], Lindeberg [51, 53], Jansson and Lindeberg [33, 34], Zhu *et al.* [99], Penaud *et al.* [70], Sangalli *et al.* [80], Zhan *et al.* [98], Yang *et al.* [93], Wimmer *et al.* [91], Barisin *et al.* [6, 7] and Perzanowski and Lindeberg [71]. These classes of deep networks have specifically been demonstrated to handle image data subject to scaling variations in a much more robust way compared to non-covariant deep networks.

With the combined treatment of the 4 main classes of geometric image transformations considered in this paper, involving (i) spatial scaling transformations, (ii) spatial affine transformations, (iii) Galilean transformations and (iv) temporal scaling transformations, the underlying aims of this work bear close similarities to approaches in the area of geometric deep learning (Bronstein *et al.* [15], Gerken *et al.* [24]), which also consider visual processing operations that are covariant under wider classes of geometric image transformations.

For an introductory overview of Lie groups and Lie algebras, to which the presented relations between receptive field responses for different values of the filter parameters will be closely related to, see *e.g.,* Hall [28]. For an extensive treatment of one-parameter semi-groups and their infinitesimal generators, see Hille and Phillips [31]. For further relations between semi-groups and partial differential equations as well as extensions of the theory to two-parameter semi-groups, to which the presented methodology will also bear very close relations, see Pazy [68], Goldstein [27] and Al-Sharif and Khalil [2].

3 The Generalized Gaussian Derivative Model for Visual Receptive Fields

3.1 Spatial and Spatio-Temporal Smoothing Kernels

The generalized Gaussian derivative model for visual receptive fields (Lindeberg [52]) is based on:

- smoothing any purely spatial image $f \colon \mathbb{R}^2 \to \mathbb{R}$ with affine Gaussian kernels [1] $T \colon \mathbb{R}^2 \times \mathbb{R}_+ \times \mathbb{S}^2_+ \to \mathbb{R}$ of the form

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

  where

  - $x = (x_1, x_2)^T \in \mathbb{R}^2$ denotes the image coordinates,
  - $s \in \mathbb{R}_+$ is the spatial scale parameter, and
  - $\Sigma \in \mathbb{S}^2_+$ is a spatial covariance matrix that specifies the shape of the affine Gaussian kernel, or

- smoothing any video sequence or video stream $f \colon \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ with joint spatio-temporal smoothing kernels $T \colon \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{S}^2_+ \times \mathbb{R}_+ \times \mathbb{R}^2 \to \mathbb{R}$ of the form

$$T(x, t;\ s, \Sigma, \tau, v) = g(x - v\, t;\ s, \Sigma)\, h(t;\ \tau), \tag{2}$$

  where

  - $t \in \mathbb{R}$ is the time variable,
  - $v = (v_1, v_2)^T \in \mathbb{R}^2$ is an image velocity,
  - $\tau \in \mathbb{R}_+$ is the temporal scale parameter, and
  - $h \colon \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ is a temporal smoothing kernel.

In the case of a non-causal temporal domain, *i.e.,* for pre-recorded video, for which the future can be accessed, the temporal smoothing kernel $h(t;\ \tau)$ in the spatio-temporal smoothing kernel $T(x, t;\ s, \Sigma, \tau, v)$ can be chosen as a Gaussian kernel

$$h(t;\ \tau) = g(t;\ \tau) = \frac{1}{\sqrt{2\pi}\sqrt{\tau}}\, e^{-t^2/2\tau}. \tag{3}$$

In the case of a time-causal domain, *i.e.*, for real-time situations when the future cannot be accessed, the temporal smoothing kernel $h(t;\ \tau)$ in the spatio-temporal smoothing kernel $T(x, t;\ s, \Sigma, \tau, v)$ can be chosen as the time-causal limit kernel (Lindeberg [50] Sect. 5; Lindeberg [54] Sect. 3)

$$h(t;\ \tau) = \psi(t;\ \tau, c), \tag{4}$$

[1] In this treatment, $\mathbb{S}^2_+$ denotes the set of symmetric positive definite 2×2 matrices.

characterized by having a Fourier transform of the form

$$\hat{\Psi}(\omega;\ \tau, c) = \prod_{k=1}^{\infty} \frac{1}{1 + i\, c^{-k} \sqrt{c^2 - 1}\, \sqrt{\tau}\, \omega}, \tag{5}$$

where $c > 1$ is a distribution parameter describing the ratio between adjacent discrete temporal scale levels

$$\tau_k = \tau_0\, c^{2k} \tag{6}$$

for some initial temporal scale level $\tau_0 \in \mathbb{R}_+$. For common purposes of implementation, the distribution parameter c is often chosen as $c = \sqrt{2}$ or $c = 2$.

In fact, it is shown in an axiomatic way in Lindeberg [47] that the forms of the spatial smoothing kernel (1) and the spatio-temporal smoothing kernel (2) are uniquely determined, given natural symmetry requirements on a visual front-end, with conceptual extensions of those ideas to a time-causal temporal domain in Lindeberg [50, 52]. The underlying symmetry requirements used for the axiomatic derivations are based on symmetry properties of the environment, in combination with internal consistency requirements, to guarantee theoretically well-founded treatment of image structures over multiple spatial and temporal scales, see Lindeberg [47, 52] for further details.

3.2 Receptive Field Families in Terms of Spatial and Spatio-Temporal Derivatives

For defining either purely spatial or joint spatio-temporal receptive fields, to be used for processing visual image data, the either purely spatial or joint spatio-temporal smoothing operations are, in turn, combined with either purely spatial differential operators $\mathcal{D}_x$ or joint spatio-temporal differentiation operators $\mathcal{D}_{x,t}$, to give rise to spatial receptive fields applied to spatial image data $f : \mathbb{R}^2 \to \mathbb{R}$ of the form

$$\mathcal{R}\, f(\cdot) = \mathcal{D}_x (T(\cdot;\ s, \Sigma) * f(\cdot)), \tag{7}$$

or joint spatio-temporal receptive fields applied to video sequences or video streams $f : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ of the form

$$\mathcal{R}\, f(\cdot, \cdot) = \mathcal{D}_{x,t} (T(\cdot, \cdot;\ s, \Sigma, \tau, v) * f(\cdot, \cdot)). \tag{8}$$

Regarding the choice of what spatial differentiation operators $\mathcal{D}_x$ to use, common choices include directional derivative operators in the orientation $\varphi \in [-\pi, \pi]$ of the form

$$\mathcal{D}_x = \partial_\varphi^m = (\cos\varphi\, \partial_{x_1} + \sin\varphi\, \partial_{x_2})^m, \tag{9}$$

the spatial gradient operator

$$\mathcal{D}_x = \nabla_x = (\partial_{x_1}, \partial_{x_2})^T, \tag{10}$$

or the Hessian matrix operator

$$\mathcal{D}_x = \mathcal{H}_x = \nabla_x \nabla_x^T = \begin{pmatrix} \partial_{x_1 x_1} & \partial_{x_1 x_2} \\ \partial_{x_1 x_2} & \partial_{x_2 x_2} \end{pmatrix}. \tag{11}$$

Concerning joint spatio-temporal derivative operators $\mathcal{D}_{x,t}$, a natural choice is to combine the above purely spatial derivative operators $\mathcal{D}_x$ with velocity-adapted temporal derivative operators $\partial_{\bar{t}}^n$ according to

$$\partial_{\bar{t}}^n = (\partial_t + v_1\, \partial_{x_1} + v_2\, \partial_{x_2}), \tag{12}$$

leading to joint spatio-temporal derivative operators $\mathcal{D}_{x,t}$ of the form

$$\mathcal{D}_{x,t} = \mathcal{D}_x\, \partial_{\bar{t}}^n, \tag{13}$$

or alternatively using space-time separable (not velocity-adapted) spatio-temporal derivative operators of the form

$$\mathcal{D}_{x,t} = \mathcal{D}_x\, \partial_t^n. \tag{14}$$

3.3 Visualizations of Spatial and Spatio-Temporal Receptive Fields

Figures 1, 2 and 3 show sample illustrations of receptive fields according to the generalized Gaussian derivative model for visual receptive fields. Figure 1 shows purely spatial receptive fields of the form

$$T_{\varphi^m}(x;\ s, \Sigma) = \partial_{\varphi^m}\, g(x;\ s, \Sigma), \tag{15}$$

obtained by applying directional derivative operators of the form (9) of orders $m = 1$ and $m = 2$ to affine Gaussian kernels according to (1), based on the specific parameterization of the elements in the spatial covariance matrix Σ according to

$$\Sigma_{11} = \lambda_1 \cos^2\varphi + \lambda_2 \sin^2\varphi, \tag{16}$$

$$\Sigma_{12} = (\lambda_1 - \lambda_2) \cos\varphi\, \sin\varphi, \tag{17}$$

$$\Sigma_{22} = \lambda_1 \sin^2\varphi + \lambda_2 \cos^2\varphi, \tag{18}$$

with the eigenvalues λ_1 and λ_2 of the spatial covariance matrix Σ parameterized in terms of the corresponding spatial standard deviations σ_1 and σ_2 according to

$$\lambda_1 = \sigma_1^2, \tag{19}$$

$$\lambda_2 = \sigma_2^2. \tag{20}$$

This then leads to the following explicit expression for the affine Gaussian derivative kernel (Lindeberg [56] Eqs (163)

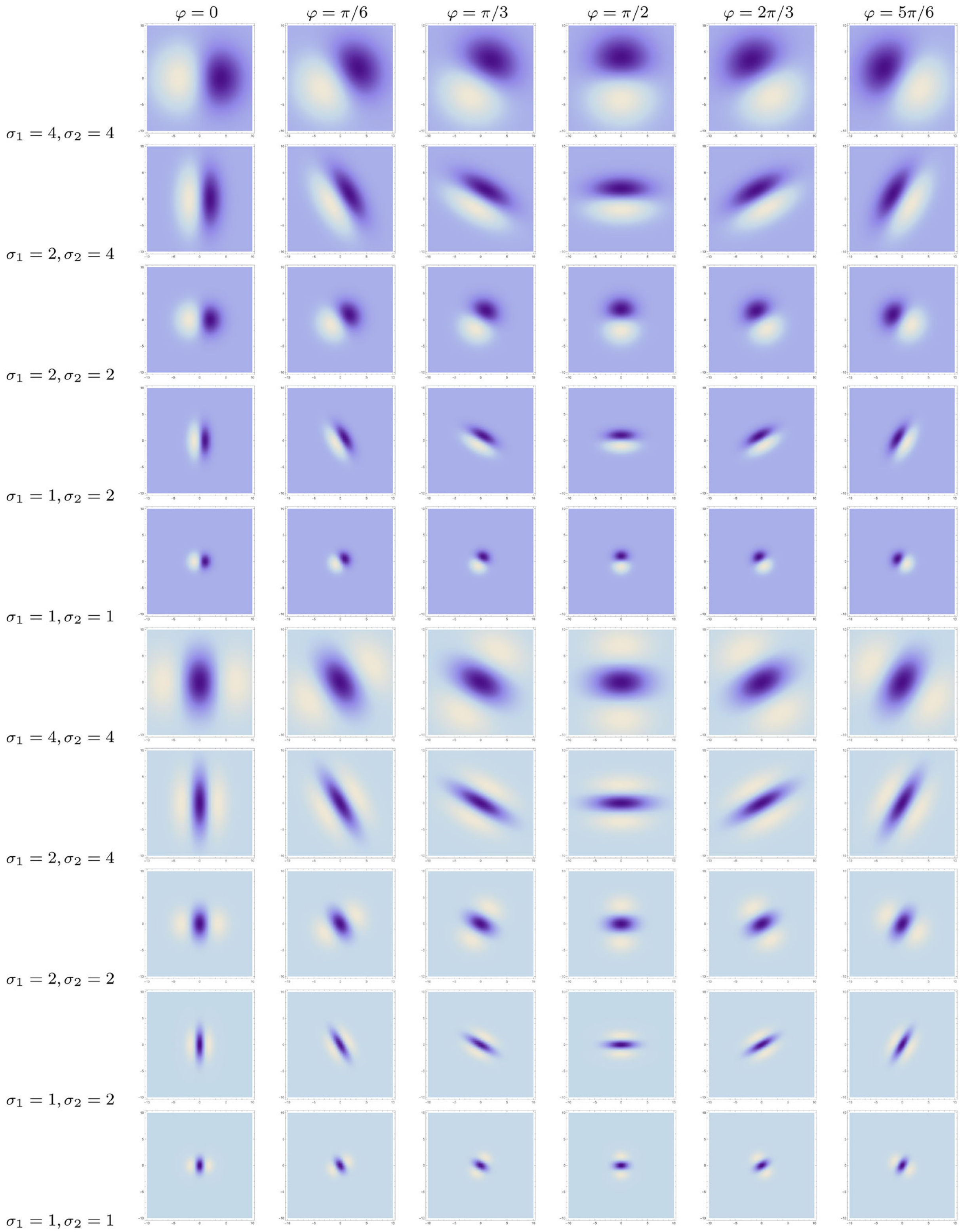

Fig. 1 Purely spatial receptive fields in terms of directional derivatives ∂_φ^m of affine Gaussian kernels $g(x;\ s, \Sigma)$ of the form (1) for orders $m = 1$ and $m = 2$, shown for different combinations of the spatial scale parameters σ_1 and σ_2, corresponding to two different eccentricities $\epsilon = \sigma_2/\sigma_1 \in \{1, 2\}$ of the receptive fields, according to the explicit parameterization of the affine Gaussian kernels according to (16)–(18) and (21)–(22). (Horizontal axes: Horizontal image coordinate $x_1 \in [-10, 10]$. Vertical axes: Vertical image coordinate $x_2 \in [-10, 10]$.)

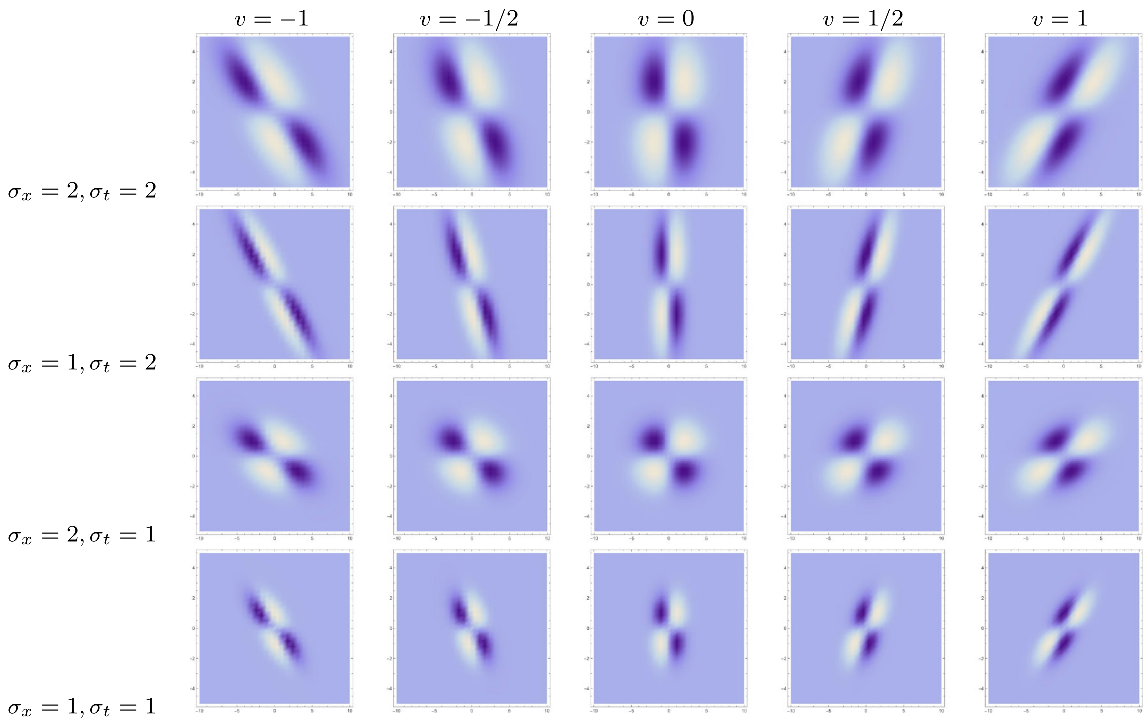

Fig. 2 Non-causal joint spatio-temporal receptive fields over a 1+1-D spatio-temporal domain in terms of the mixed first-order spatial derivative and the first-order velocity-adapted temporal derivative of the form $T_{x\bar{t}}(x, t;\; s, \tau, v)$ according to (24) as the product of a velocity-adapted 1-D Gaussian kernel $g_{1D}(x - v\,t;\; s)$ over the spatial domain and the non-causal Gaussian kernel $h(t;\; \tau) = g(t;\; \tau)$ over the temporal domain according to (3). The spatio-temporal receptive fields are shown for different values of the spatial scale parameter $\sigma_x = \sqrt{s}$ and the temporal scale parameter $\sigma_t = \sqrt{\tau}$ in dimensions of [length] and [time]. (Horizontal axes: Spatial image coordinate $x \in [-10, 10]$. Vertical axes: Temporal variable $t \in [-4, 4]$.)

and (164))

$$g_{\text{aff}}(x, y;\; \sigma_1, \sigma_2, \varphi) = \frac{1}{2\pi\sigma_1\sigma_2}\, e^{-A/2\sigma_1^2\sigma_2^2}, \tag{21}$$

where

$$A = (\sigma_2^2\, x_1^2 + \sigma_1^2\, x_2^2)\cos^2\varphi + (\sigma_1^2\, x_1^2 + \sigma_2^2\, x_2^2)\sin^2\varphi \\ -2\,(\sigma_1^2 - \sigma_2^2)\, x_1\, x_2\, \cos\varphi\, \sin\varphi. \tag{22}$$

The eccentricity of such a kernel, which represents its degree of elongation, is defined as the ratio

$$\epsilon = \frac{\sigma_2}{\sigma_1}. \tag{23}$$

Figures 2 and 3 show spatio-temporal receptive fields corresponding to the mixed spatio-temporal derivative

$$T_{x\bar{t}}(x, t;\; s, \tau, v) = \partial_x \partial_{\bar{t}}\, (g(x - v\,t;\; s)\, h(t;\; \tau)) \tag{24}$$

over a 1+1-D spatio-temporal domain, where the 2-D affine Gaussian kernel (1) reduces to a 1-D Gaussian kernel

$$g(x;\; s) = \frac{1}{\sqrt{2\pi\, s}}\, e^{-x^2/2s}. \tag{25}$$

Figure 2 shows such receptive fields in the case of a non-causal temporal domain, where the temporal smoothing is performed with the non-causal temporal Gaussian kernel according to (3), whereas Fig. 3 shows such receptive fields in the case of a time-causal temporal domain, where the temporal smoothing is performed with the time-causal limit kernel according to (4),

3.4 Relations to Scale-Normalized Derivatives

In Lindeberg [52, 59, 60], so-called unnormalized spatio-temporal derivative expressions of the above forms are additionally combined with different forms of scale normalization, as depending on the actual values of the spatial scale parameter s, the spatial covariance matrix Σ and the temporal scale parameter τ, to define scale-normalized spatial

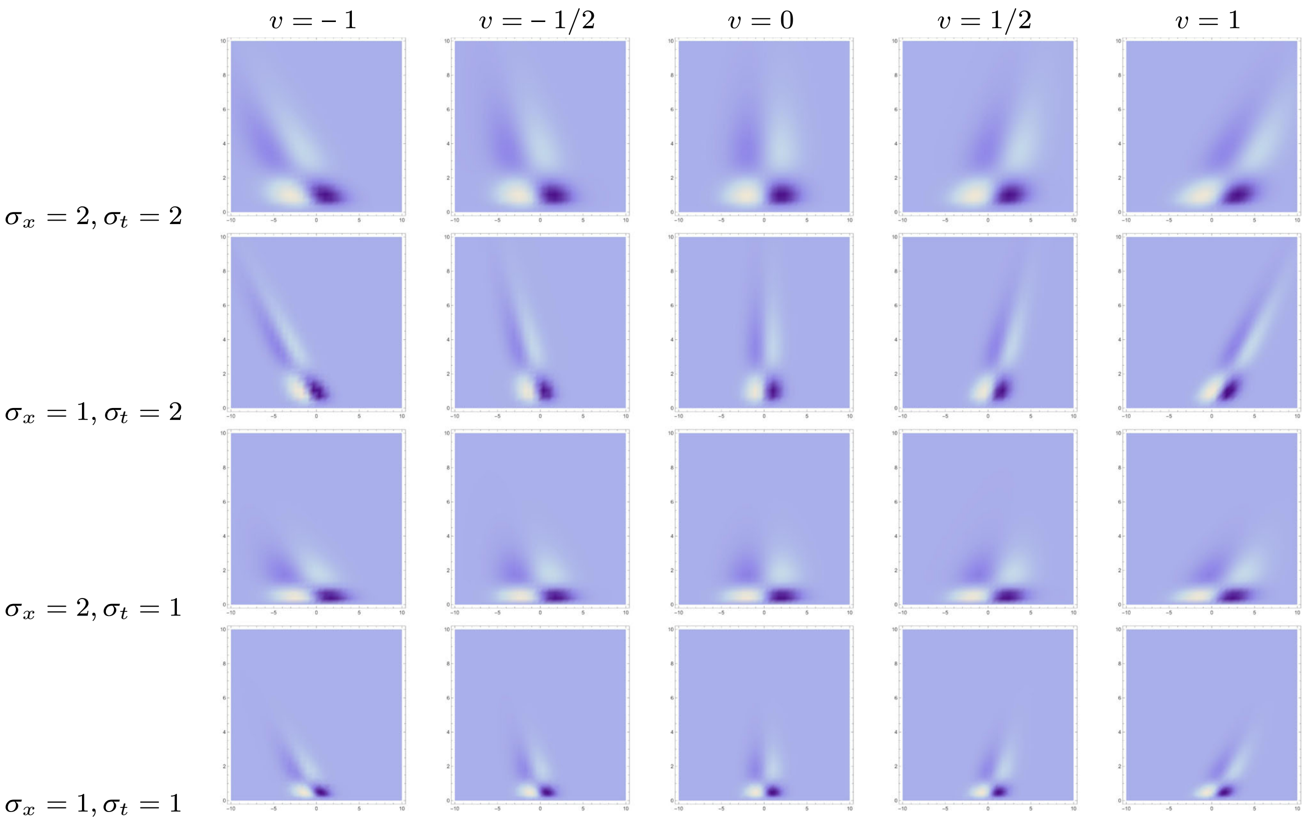

Fig. 3 Time-causal joint spatio-temporal receptive fields over a 1+1-D spatio-temporal domain in terms of the mixed first-order spatial derivative and the first-order velocity-adapted temporal derivative of the form $T_{x\bar{t}}(x, t;\; s, \tau, v)$ according to (24) as the product of a velocity-adapted 1-D Gaussian kernel $g_{1D}(x - v\,t;\; s)$ over the spatial domain and the time-causal limit kernel $h(t;\; \tau) = \psi(t;\; \tau, c)$ over the temporal domain according to (4) with the distribution parameter set to $c = 2$. The spatio-temporal receptive fields are shown for different values of the spatial scale parameter $\sigma_x = \sqrt{s}$ and the temporal scale parameter $\sigma_t = \sqrt{\tau}$ in dimensions of [length] and [time]. (Horizontal axes: Spatial image coordinate $x \in [-10, 10]$. Vertical axes: Temporal variable $t \in [0, 8]$)

and spatio-temporal derivative operators [2] [3] that are provably scale covariant, in the sense that the magnitude of the

[2] For in-depth treatments about how to extend the regular unnormalized spatial and spatio-temporal derivative operators $\mathcal{D}_x$ and $\mathcal{D}_{x,t}$ to corresponding scale-normalized spatial and spatio-temporal derivative operators that make the full derivative-based receptive field responses in (7) and (8) provably covariant under the composed geometric image transformations in (26), (33), (34) and (35), see Lindeberg [59, 60]. Basically, to achieve covariance under with respect to uniform spatial scaling transformations of the form $x' = S_x\, x$, where $S_x \in \mathbb{R}_+$ is a spatial scaling factor, the partial derivative operators ∂_{x_i} in the spatial gradient operator ∇_x according to (10) and the spatial Hessian operator $\mathcal{H}_x$ according to (11) should be replaced by regular scale-normalized spatial derivatives according to $\partial_{x_i,\text{norm}} = s^{1/2}\,\partial_{x_i}$ according to Lindeberg [45] Eq. (6) for the scale normalization power $\gamma = 1$, while the regular directional derivative operator ∂_φ^m according to (9) should be replaced by the scale-normalized directional derivative operator $\partial_{\varphi,\text{norm}}^m$ according to Lindeberg [59] Eq. (33).
To achieve covariance with respect to non-isotropic spatial affine transformations of the form $x' = A\,x$, where A is a positive definite 2×2 affine transformation matrix, the regular gradient operator ∇_x according to (10) should be replaced by the scale-normalized affine gradient operator $\nabla_{x,\text{affnorm}}$ according to Lindeberg [59] Eq. (111), while the regular Hessian operator $\mathcal{H}_x$ according to (11) should be replaced by the scale-normalized affine Hessian operator $\mathcal{H}_{x,\text{affnorm}}$ according to Lindeberg [59] Eq. (140).
To achieve both covariance under Galilean transformations of the form $x' = x + u\,t$, where $u = (u_1, u_2)^T \in \mathbb{R}^2$ is a 2-D motion vector, and covariance under temporal scaling transformations of the form $t' = S_t\,t$, where $S_t \in \mathbb{R}_+$ is a temporal scaling factor, the temporal differentiation of the receptive fields should be performed in terms of scale-normalized velocity-adapted temporal derivative operators obtained by replacing the regular velocity-adapted temporal derivative operator $\partial_{\bar{t}}^n$ in (12) by its scale-normalized correspondence $\partial_{\bar{t},\text{norm}}^n$ according to Lindeberg [59] Eq. (168).

[3] If using scale-normalized derivative operators in the models for the visual receptive fields, it should, however, be observed that the introduction of scale-normalized derivatives will substantially influence the evolution properties over scales of the receptive field responses. This follows, since the scale-dependent scale normalization factors used in the formulations of scale-normalized derivative operators, as extensively described in Lindeberg [59], would then also need to be differentiated with respect to the spatial and the temporal scale parameters. For this reason, we do in this paper isolate the evolution properties over the filter parameters to only receptive field models expressed in terms of regular, not scale-normalized, spatial and/or spatio-temporal derivative expressions. If the use of scale-normalized derivatives is warranted, as it can be because of the requirement of making it possible to match the output of derivative-based receptive field operators on the image data, this can be easily accomplished by performing the scale normalization in

filter responses can be perfectly matched under specific families of geometric image transformations. In this treatment, we will, however, not consider the effects of such scale normalization and will instead focus solely on the effects of the either purely spatial or joint spatio-temporal smoothing operations, obtained by convolution with kernels of the forms (1) or (2) in combination with either purely spatial or joint spatio-temporal derivative computations, in turn, leading to receptive field responses of the forms (7) or (8). For the task of relating the responses of linear receptive fields of the forms (7) or (8), such an analysis is sufficient, provided that the spatial differentiation operators $\mathcal{D}_x$ and the spatio-temporal differential operators $\mathcal{D}_{x,t}$ in these receptive field models are purely linear operators.

3.5 Covariance Properties of the Spatial and the Spatio-Temporal Smoothing Kernels

With regard to variabilities of the shapes of the spatial or the spatio-temporal receptive fields in relation to the variabilities in spatial or spatio-temporal image transformations involved in the image formation process that determines the structures of the either purely spatial or joint spatio-temporal image data f, an important consequence of the formulation of the spatial smoothing kernel $T(x;\; s, \Sigma)$ according to (1) and the formulation of the joint spatio-temporal smoothing kernel $T(x, t;\; s, \Sigma, \tau, v)$ according to (2) is that these forms of convolution kernels are covariant under important basic classes of geometric image transformations.

3.5.1 Purely Spatial Covariance Properties

For the case of a purely spatial domain, let us consider an image transformation of the form

$$f'(x') = f(x) \quad \text{for} \quad x' = S_x\, A, \tag{26}$$

where

- $S_x \in \mathbb{R}_+$ is a spatial scaling factor,
- A is a non-singular 2×2 matrix that we could assume to be normalized in a way as corresponding to orthographic projection of a smooth local surface patch.

The geometric motivation for using this model is that (see Fig. 4 for an illustration):

- the spatial scaling factor S_x does to first order of approximation reflect the size variations caused by varying the distance between the object and the observer, whereas
- the spatial affine transformation matrix A does to first order of approximation reflect the image deformations caused by varying the viewing direction for a visual observer that views a smooth local surface patch.

Then, if we define purely spatial scale-space representations $L \colon \mathbb{R}^2 \times \mathbb{R}_+ \times \mathbb{S}^2_+ \to \mathbb{R}$ and $L' \colon \mathbb{R}^2 \times \mathbb{R}_+ \times \mathbb{S}^2_+ \to \mathbb{R}$ of the purely spatial images f and f', respectively, according to

$$L(\cdot;\; s, \Sigma) = T(\cdot;\; s, \Sigma) * f(\cdot), \tag{27}$$

$$L'(\cdot;\; s', \Sigma') = T(\cdot;\; s', \Sigma') * f'(\cdot), \tag{28}$$

it follows that the scale-space representations can be perfectly matched under the geometric image transformation (26)

$$L'(x';\; s', \Sigma') = L(x;\; s, \Sigma), \tag{29}$$

provided that the parameters of the receptive fields are matched according to Lindeberg [59] Equations (252)–(253):

$$s' = S_x^2\, s, \tag{30}$$

$$\Sigma' = A\, \Sigma\, A^T. \tag{31}$$

Due to the overparameterization of the degrees of freedom spanned by the spatial scale parameter s and the spatial covariance matrix Σ, however, a minimal necessity requirement can also be stated in terms of the following combined matching criterion (Lindeberg [59] Eq. (118)):

$$s'\, \Sigma' = s\, (S_x\, A)\, \Sigma\, (S_x A)^T = s\, S_x^2\, A\, \Sigma\, A^T. \tag{32}$$

3.5.2 Joint Spatio-Temporal Covariance Properties

For the case of a joint spatio-temporal domain, let us consider a joint spatio-temporal image transformation of the form

$$f'(x', t') = f(x, t) \tag{33}$$

for

$$x' = S_x\, (A\, x + u\, t), \tag{34}$$

$$t' = S_t\, t, \tag{35}$$

where for the entities not already defined in connection with Equation (26), we have that

- $u = (u_1, u_2)^T \in \mathbb{R}^2$ is a 2-D velocity vector that may correspond to the orthographic projection of the 3-D motion field U of the local surface patch observed by the vision system, and
- $S_t \in \mathbb{R}_+$ is a temporal scaling factor.

a second stage, after the computation of regular, not scale-normalized, derivative operators in a first processing stage.

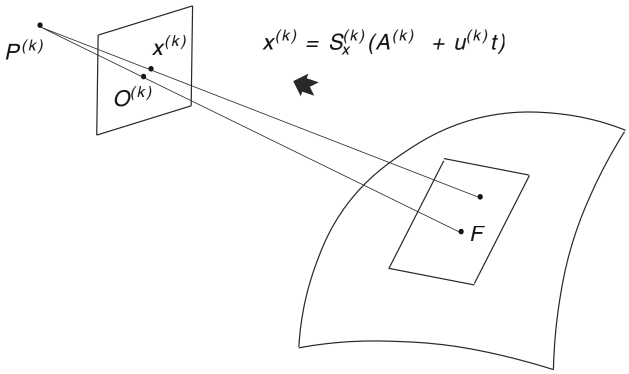

Fig. 4 Illustration of the geometry underlying the composed locally linearized projection models in Equation (26) and Equations (33)–(35), when extended to a multi-view imaging situation, with each view indexed by k. We consider a local, in the spatio-temporal case possibly moving, surface patch, which is projected to an arbitrary view in a multi-view locally linearized projection model, with the fixation point F on the surface mapped to the origin $O^{(k)} = 0$ in the image plane for the observer with the optic center $P^{(k)}$. Then, any point in the tangent plane to the surface at the fixation point, as parameterized by the local coordinates ξ in a coordinate frame attached to the tangent plane of the surface with $\xi = 0$ at the fixation point F, is by the local linearization mapped to the image point $x^{(k)}$. (Figure reproduced from Lindeberg [59] with permission (OpenAccess))

The geometric motivation for using this model is beyond the reasons for using a combination of uniform spatial scaling transformations and spatial affine transformations according to the previously stated motivation in Sect. 3.5.1 that (see Fig. 4 for an illustration):

- the Galilean transformation of the form $x' = x + u\, t$ models the effect of relative motions between the observed objects or spatial–temporal events and the viewing direction, whereas
- the temporal scaling factor S_t models the effect of observing similar spatio-temporal events that may occur either faster or slower relative to a previously observed reference view.

Then, if we define the joint spatio-temporal scale-space representations $L \colon \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{S}_+^2 \times \mathbb{R}_+ \times \mathbb{R}^2 \to \mathbb{R}$ and $L' \colon \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{S}_+^2 \times \mathbb{R}_+ \times \mathbb{R}^2 \to \mathbb{R}$ of the video sequences or video streams f and f', respectively, according to

$$L(\cdot, \cdot;\ s, \Sigma, \tau, v) = T(\cdot, \cdot;\ s, \Sigma, \tau, v) * f(\cdot, \cdot), \tag{36}$$

$$L'(\cdot, \cdot;\ s', \Sigma', \tau', v') = T(\cdot, \cdot;\ s', \Sigma', \tau', v') * f'(\cdot, \cdot), \tag{37}$$

it follows that the scale-space representations can be perfectly matched under the composed geometric image transformation given by (33), (34) and (35)

$$L'(x', t';\ s', \Sigma', \tau', v') = L(x, t;\ s, \Sigma, \tau, v), \tag{38}$$

provided that the purely spatial parameters s, s', A and A' are matched according to (30) and (31) and additionally the temporal scale parameters τ and τ' as well as the velocity parameters v and v' are matched according to Lindeberg [59] Equations (254)–(255):

$$\tau' = S_t^2\, \tau, \tag{39}$$

$$v' = \frac{S_x}{S_t}(A\, v + u). \tag{40}$$

3.6 Motivation for Filter Banks of Receptive Field Responses

A main consequence of the above covariance properties is therefore that from a requirement of the vision system to be covariant under the basic classes of geometric image transformations, if we want to have the ability to match the effects of the either purely spatial or joint spatio-temporal smoothing operations that occur in the models for the receptive fields, then we have to have the ability to make use of receptive field responses computed for matching sets of filter parameters, either (s, Σ) or (s', Σ') in the purely spatial case or (s, Σ, τ, v) and (s', Σ', τ', v') in the joint spatio-temporal case.

Thus, the requirement of establishing an identity between receptive field responses under the influence of *a priori* unknown geometric image transformations, caused by different viewing conditions between different observations of the same scene or a similar type of spatio-temporal event, leads to the strategy of basing the filtering operations in an idealized vision system on filter banks with receptive field responses computed for a large variability of the filter parameters. This theoretical background forms the conceptual foundation for the main topic in this paper of relating receptive field responses computed for different values of the filter parameters of the receptive fields.

4 Infinitesimal Hybrid Lie Semi-Group Structures for the Generalized Gaussian Derivative Model

Given the above formulations of models for either purely spatial or joint spatio-temporal receptive fields in terms of either purely spatial or joint spatio-temporal derivative expressions applied to either purely spatial or joint spatio-temporal smoothing kernels, as obtained from multi-parameter scale-space representations, one may ask how such receptive field responses are related between different values of the filter parameters. Specifically, one may ask if and then how a receptive field response at a coarser level of spatial and temporal scales may be computed from receptive field responses at

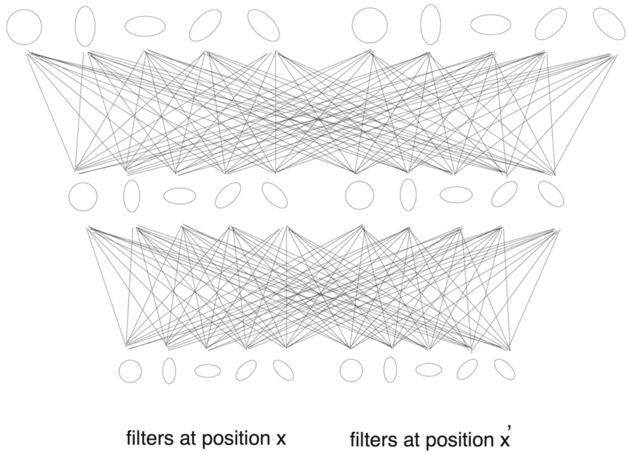

Fig. 5 Schematic illustration of how receptive fields for different parameter values and at different image positions can be interrelated to each other, as constituting the main subject of study in this paper. Here, we derive such relationships between receptive field responses obtained for different values of the shape parameters of the receptive fields, either in terms of (i) infinitesimal relationships closely related to the notions of Lie groups and infinitesimal generators for semi-groups, and (ii) macroscopic relationships in terms of cascade smoothing properties, with close relations to the notion of Lie algebras, as can be related to Lie groups *via* the exponential map. Compared to regular Lie groups and Lie algebras, some of the evolution relations do, however, have a directional preference, implying that the evolution can only be performed in a single direction, and not in the reverse direction. (In this figure, each line represents a connection between two receptive fields for different values of their parameters and/or their spatial positions)

finer spatial and temporal scales, as schematically illustrated in Fig. 5.

In this section, we will derive a set of differential relationships that show how either spatial receptive field responses

$$\mathcal{R}f(\cdot) = \mathcal{D}_x T(\cdot;\ s, \Sigma) * f(\cdot) \tag{41}$$

or spatio-temporal receptive field responses

$$\mathcal{R}f(\cdot,\cdot) = \mathcal{D}_{x,t} T(\cdot,\cdot;\ s, \Sigma, \tau, v) * f(\cdot,\cdot) \tag{42}$$

computed for different values of the filter parameters (s, Σ) or (s, Σ, τ, v) are related under infinitesimal perturbations of the filter parameters, in the respective cases of either purely spatial receptive fields or joint spatio-temporal receptive fields, initially restricted to the situation when both the spatial and the temporal smoothing kernels are based on Gaussian kernels, and therefore differentiable with respect to all of their respective parameters. [4]

To a major extent, this will be structurally similar to the notion of a Lie group structure over the filter parameters, as expressed in terms of relations between derivatives with respect to the filter parameters (s, Σ) or (s, Σ, τ, v) in relation to derivatives with respect to the either spatial or joint spatio-temporal image coordinates x or (x, t). Since the evolution properties with respect to the spatial scale parameter s and the temporal scale parameter τ can, however, only performed in the direction of increasing values of the scale parameters, the corresponding transformations will therefore not form a full group, but only a semi-group with regard to these two dimensions of variabilities in the filter parameters. For the derived infinitesimal evolution equations with respect to these scale parameters, the resulting notions will therefore be closer to the notion of infinitesimal generators for semi-groups over continuous evolution parameters. For two of the dimensions of the variability, concerning the mixed element Σ_{12} in the spatial covariance matrix Σ as well as the variabilities with respect to the elements v_1 and v_2 in the velocity vector v, the variabilities are, on the other hand, not as constrained. For these reasons, we will refer to the resulting structures, to be derived, as hybrid Lie semi-group structures.

Before starting deriving the resulting infinitesimal relationships over the different filter parameters, let us first describe the methodology that we will use in this section, by exploiting the fact that for the purely spatial receptive field model as well as for the joint spatio-temporal receptive field model, for the case of a non-causal temporal domain, the convolution kernels are all expressed in terms of higher-dimensional Gaussian kernels. This implies that we can make use of similar algebraic properties for performing the proofs as arise in the definition of Hermite functions and Hermite polynomials, although here generalized to a higher-dimensional setting for the specific forms of spatial or spatio-temporal smoothing kernels, as occur in our theoretically principled models for visual receptive fields.

4.1 Methodology to be Used

For the one-dimensional Gaussian kernel, it is well-known that the computation of derivatives of this kernel corresponds to multiplying the Gaussian function with Hermite polynomials. For our way of parameterizing the Gaussian kernel, the probabilistic Hermite polynomials $\mathrm{He}_m(x)$, defined by

$$\mathrm{He}_m(x) = (-1)^m e^{x^2/2}\, \partial_{x^m} \left(e^{-x^2/2}\right), \tag{43}$$

[4] When using the time-causal limit kernel $h(t;\ \tau) = \psi(t;\ \tau, c)$ according to (4) as the temporal smoothing kernel in the spatio-temporal smoothing kernel $T(x, t;\ s, \Sigma, \tau, v)$, an important restriction is that the temporal scale levels are genuinely discrete. Therefore, the corresponding temporal or spatio-temporal smoothing kernels cannot be differentiated with respect to the discrete scale parameter $\tau_k = \tau_0\, c^{2k}$ according to (6).

imply that

$$\partial_{x^m}\left(e^{-x^2/2}\right) = (-1)^m \operatorname{He}_m(x)\, e^{-x^2/2} \tag{44}$$

and

$$\partial_{x^m}\left(e^{-x^2/2s}\right) = (-1)^m \operatorname{He}_m(\frac{x}{\sqrt{s}})\, e^{-x^2/2s}\, \frac{1}{\sqrt{s}^m}. \tag{45}$$

This means that the n:th-order Gaussian derivative kernel in 1-D can be written as

$$\begin{aligned}\partial_{x^m}\left(g(x;\ s)\right) &= \frac{1}{\sqrt{2\pi}\sqrt{s}}\,\partial_{x^m}\left(e^{-x^2/2s}\right)\\ &= \frac{1}{\sqrt{2\pi}\sqrt{s}}\frac{(-1)^m}{\sqrt{s}^m}\operatorname{He}_m(\frac{x}{\sqrt{s}})\, e^{-x^2/2s}\\ &= \frac{(-1)^m}{\sqrt{s}^m}\operatorname{He}_m(\frac{x}{\sqrt{s}})\, g(x;\ s),\end{aligned} \tag{46}$$

in other words as a polynomial multiplied with the original Gaussian function.

In this section, we will derive corresponding generalizations of Hermite polynomials for the 2-D and 2+1-D Gaussian kernels that correspond to the either purely spatial or joint spatio-temporal smoothing operations in the spatial and spatio-temporal receptive field models according to (7) and (8).

Specifically, we will frequently make use of the property that if two different differential operators give rise to the same generalized Hermite polynomials, when applied to the either purely spatial or joint spatio-temporal smoothing kernels, then also the corresponding spatial or spatio-temporal scale-space representations have to be equal, as well as any spatial or temporal derivative operator, or any first-order differentiation operator with respect to a parameter of the spatial or temporal smoothing kernel.

Stated more formally in a compact manner, let

- p denote either 2-D spatial or 2+1-D spatio-temporal image coordinates, $p = (x_1, x_2)^T$ or $p = (x_1, x_2, t)^T$,
- $g(p;\ P)$ denote a corresponding either 2-D spatial or joint 2+1-D spatio-temporal smoothing kernel depending on a set of filter parameters P,
- ∂_{P_i} denote a partial derivative operator with respect to the i:th filter parameter, and
- $\mathcal{D}_{i,p}$ denote a possibly composed differentiation operator with respect to the image coordinates p.

Then, if we for the partial derivative operator ∂_{P_i} with respect to the filter parameter P_i are able to show that this partial derivative operator does for the either purely spatial or joint spatio-temporal smoothing function $g(p;\ P)$ for the corresponding either spatial or joint spatio-temporal scale-space representation

$$L(\cdot;\ P) = g(\cdot;\ P) * f(\cdot) \tag{47}$$

of the either purely spatial or joint spatio-temporal input image $f\colon \mathbb{R}^2 \to \mathbb{R}$ or $f\colon \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ satisfy

$$\partial_{P_i} g(p;\ P) = \mathcal{D}_{i,p} g(p;\ P), \tag{48}$$

then it follows from the linearity of the scale-space representation L that

$$\begin{aligned}&\partial_{P_i} L(\cdot;\ P) = \partial_{P_i} g(\cdot;\ P) * f(\cdot) =\\ &= \mathcal{D}_{i,p} g(\cdot;\ P) * f(\cdot) = \mathcal{D}_{i,p} L(\cdot;\ P).\end{aligned} \tag{49}$$

Similarly, for any linear either purely spatial receptive field operator $\mathcal{R}$ defined from applying a spatial differential operator $\mathcal{D}_x$ to a purely spatial scale-space representation

$$\mathcal{R} f(\cdot) = \mathcal{D}_x L(\cdot;\ P), \tag{50}$$

or for any linear joint spatio-temporal receptive field operator $\mathcal{R}$ defined from applying a spatio-temporal differential operator $\mathcal{D}_{x,t}$ to a the joint spatio-temporal scale-space representation

$$\mathcal{R} f(\cdot) = \mathcal{D}_{x,t} L(\cdot;\ P), \tag{51}$$

it holds that

$$\partial_{P_i}(\mathcal{R} f) = \mathcal{D}_{i,p}(\mathcal{R} f). \tag{52}$$

Specifically, since for all these generalized Gaussian derivative models for visual receptive fields, the result of applying the differential operators to the either 2-D purely spatial or 2+1-D joint spatio-temporal Gaussian kernels will be generalized Hermite functions multiplied by the underlying Gaussian kernels

$$\partial_{P_i} g(p;\ P) = H_{P_i}(p;\ P)\, g(p;\ P), \tag{53}$$

$$\mathcal{D}_{i,p} g(p;\ P) = H_{i,p}(p;\ P)\, g(p;\ P), \tag{54}$$

implying that the respective generalized Hermite polynomials are given by

$$H_{P_i}(p;\ P) = \frac{\partial_{P_i} g(p;\ P)}{g(p;\ P)}, \tag{55}$$

$$H_{i,p}(p;\ P) = \frac{\mathcal{D}_{i,p} g(p;\ P)}{g(p;\ P)}, \tag{56}$$

it will for the purposes of the following proofs to be performed be sufficient to show that the corresponding generalized Hermite polynomials are equal

$$H_{P_i}(p;\ P) = H_{i,p}(p;\ P), \tag{57}$$

in order to conclude that a differentiation operator $\mathcal{D}_{i,p}$ with respect to the image coordinates p for the first-order partial derivative ∂_{P_i} with respect to the filter parameter P_i of the receptive field is given by

$$\partial_{P_i} = \mathcal{D}_{i,p}. \tag{58}$$

Specifically, if we let the operator $\mathcal{D}_p$ denote the either purely spatial differentiation operator $\mathcal{D}_x$ used for computing a spatial-derivative-based purely spatial receptive field response according to (7), or alternatively using the same notation $\mathcal{D}_p$ for denoting a spatio-temporal-derivative-based joint spatio-temporal receptive field response $\mathcal{D}_{x,t}$ according to (8), such that those receptive field responses can be combined on the compact form

$$\mathcal{R}\, f(\cdot) = \mathcal{D}_p(g(;\ P) * f(\cdot)), \tag{59}$$

then it follows from the linearity of the involved differentiation operators that the infinitesimal relationships between differentiation with respect to a filter parameter ∂_{P_i} and the corresponding either purely spatial or joint spatio-temporal differentiation operator $\mathcal{D}_{i,p}$ according to (58) generalize to corresponding receptive field responses according to

$$\partial_{P_i}\, \mathcal{R}\, f(\cdot) = \mathcal{D}_{i,p}\, \mathcal{R}\, f(\cdot). \tag{60}$$

Based on this general property, we will in the following Sects. 4.2–4.4 focus solely on the properties of the generalized Hermite polynomials that arise from differentiating the either purely spatial or joint spatio-temporal smoothing kernels in the idealized receptive field models with respect to the image coordinates and the filter parameters for the respective cases of: (i) purely spatial receptive fields, (ii) joint spatio-temporal receptive fields based on smoothing with spatially isotropic Gaussian kernels, for which the spatial covariance matrix Σ reduces to a unit matrix $\Sigma = I$, and (iii) joint spatio-temporal receptive fields based on smoothing with possibly anisotropic affine Gaussian kernels, for which we assume that the spatial covariance matrix Σ can generally assume values not equal to a unit matrix I.

By dividing the analysis into these three types of receptive field models, we specifically begin with the simpler purely spatial model, in order to then increase the complexity via the isotropic spatio-temporal model toward the fully affine spatio-temporal model, which obeys covariance properties over the largest set of geometric image transformations considered in this study.

$$\frac{\partial_{x_1} g(x;\ s, \Sigma)}{g(x;\ s, \Sigma)} = \frac{\Sigma_{22} x_1 - \Sigma_{12} x_2}{\Sigma_{12}^2 s - \Sigma_{11} \Sigma_{22} s}, \tag{61}$$

$$\frac{\partial_{x_2} g(x;\ s, \Sigma)}{g(x;\ s, \Sigma)} = \frac{\Sigma_{11} x_2 - \Sigma_{12} x_1}{s\,(\Sigma_{12}^2 - \Sigma_{11} \Sigma_{22})}, \tag{62}$$

$$\frac{\partial_{x_1 x_1} g(x;\ s, \Sigma)}{g(x;\ s, \Sigma)} = \frac{\Sigma_{22}^2\,(x_1^2 - \Sigma_{11} s) + \Sigma_{12}^2\,(\Sigma_{22} s + x_2^2) - 2 \Sigma_{12} \Sigma_{22} x_1 x_2}{s^2\,(\Sigma_{12}^2 - \Sigma_{11} \Sigma_{22})^2}, \tag{63}$$

$$\frac{\partial_{x_1 x_2} g(x;\ s, \Sigma)}{g(x;\ s, \Sigma)} = \frac{\Sigma_{12}\,(\Sigma_{11} \Sigma_{22} s - \Sigma_{11} x_2^2 - \Sigma_{22} x_1^2) + \Sigma_{11} \Sigma_{22} x_1 x_2 + \Sigma_{12}^3 (-s) + \Sigma_{12}^2 x_1 x_2}{s^2\,(\Sigma_{12}^2 - \Sigma_{11} \Sigma_{22})^2}, \tag{64}$$

$$\frac{\partial_{x_2 x_2} g(x;\ s, \Sigma)}{g(x;\ s, \Sigma)} = \frac{\Sigma_{11}^2\,(x_2^2 - \Sigma_{22} s) + \Sigma_{11} \Sigma_{12} (\Sigma_{12} s - 2 x_1 x_2) + \Sigma_{12}^2 x_1^2}{s^2\,(\Sigma_{12}^2 - \Sigma_{11} \Sigma_{22})^2}, \tag{65}$$

$$\frac{\partial_{s} g(x;\ s, \Sigma)}{g(x;\ s, \Sigma)} = -\frac{\Sigma_{11}\,(x_2^2 - 2 \Sigma_{22} s) + 2 \Sigma_{12}^2 s - 2 \Sigma_{12} x_1 x_2 + \Sigma_{22} x_1^2}{2 s^2\,(\Sigma_{12}^2 - \Sigma_{11} \Sigma_{22})}, \tag{66}$$

$$\frac{\partial_{\Sigma_{11}} g(x;\ s, \Sigma)}{g(x;\ s, \Sigma)} = \frac{\Sigma_{22}^2\,(x_1^2 - \Sigma_{11} s) + \Sigma_{12}^2\,(\Sigma_{22} s + x_2^2) - 2 \Sigma_{12} \Sigma_{22} x_1 x_2}{2 s\,(\Sigma_{12}^2 - \Sigma_{11} \Sigma_{22})^2}, \tag{67}$$

$$\frac{\partial_{\Sigma_{12}} g(x;\ s, \Sigma)}{g(x;\ s, \Sigma)} = \frac{\Sigma_{12}\,(\Sigma_{11} \Sigma_{22} s - \Sigma_{11} x_2^2 - \Sigma_{22} x_1^2) + \Sigma_{11} \Sigma_{22} x_1 x_2 + \Sigma_{12}^3 (-s) + \Sigma_{12}^2 x_1 x_2}{s\,(\Sigma_{12}^2 - \Sigma_{11} \Sigma_{22})^2}, \tag{68}$$

$$\frac{\partial_{\Sigma_{22}} g(x;\ s, \Sigma)}{g(x;\ s, \Sigma)} = \frac{\Sigma_{11}^2\,(x_2^2 - \Sigma_{22} s) + \Sigma_{11} \Sigma_{12} (\Sigma_{12} s - 2 x_1 x_2) + \Sigma_{12}^2 x_1^2}{2 s\,(\Sigma_{12}^2 - \Sigma_{11} \Sigma_{22})^2}. \tag{69}$$

Fig. 6 Generalized Hermite polynomials as arising from derivatives of the purely spatial affine Gaussian kernel $T(x;\ s, \Sigma) = g(x;\ s, \Sigma)$ according to (1) with respect the image coordinates $x = (x_1, x_2)^T$ up to order 2, as well as with respect to the spatial scale parameter s and the elements Σ_{11}, Σ_{12} and Σ_{22} of the spatial covariance matrix Σ

4.2 Purely Spatial Receptive Fields Based on Affine Gaussian Smoothing

For the case of a purely spatial domain, let us consider the following parameterization of the affine Gaussian kernel

$$g(x;\ s, \Sigma) = \frac{1}{2\pi\, s \sqrt{\det \Sigma}}\, e^{-x^T \Sigma^{-1} x/2s}, \tag{70}$$

which leads to the corresponding affine Gaussian scale space of any 2-D spatial image $f \colon \mathbb{R}^2 \to \mathbb{R}$:

$$L(\cdot;\ s, \Sigma) = g(\cdot;\ s, \Sigma) * f(\cdot). \tag{71}$$

The motivation why we overparameterize the effective variability in the shapes of the receptive fields in this way is for reasons of geometric interpretation with regard to viewing variations for a visual observer that may observe the same local surface patch from different distances and viewing directions. If the affine matrix A in a locally linearized perspective viewing model

$$f'(x') = f(x) \qquad \text{for} \qquad x' = S_x\, A\, x \tag{72}$$

is normalized such that it corresponds to an a local orthonormal projection, then by the requirement of covariance under the group of geometric image transformations, a variability in depth Z directly maps to a variability in the spatial scaling factor S_x, which in terms of the receptive fields will correspond to a variability in the spatial scale parameter s. Similarly, a variability in the affine transformation matrix A will map to a variability in the spatial covariance matrix Σ.

By differentiating the affine Gaussian kernel (1) with respect to the image coordinates $x = (x_1, x_2)^T$ and the parameters s and the elements $\Sigma_{i,j}$ in the 2-D positive symmetric definite covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}, \tag{73}$$

we get the explicit expressions for these derivatives according to Fig. 6. [5] By comparing these expressions, we can then verify the following infinitesimal relationships with respect to variations in the scale parameter s and the elements Σ_{11}, Σ_{12} and Σ_{22} of the spatial covariance matrix Σ:

$$\partial_s = \frac{1}{2} \nabla_x^T\, \Sigma\, \nabla_x = \frac{1}{2} \left(\Sigma_{11}\, \partial_{x_1 x_1} + 2\Sigma_{12}\, \partial_{x_1 x_2} + \Sigma_{22}\, \partial_{x_2 x_2} \right), \tag{74}$$

$$\partial_{\Sigma_{11}} = \frac{1}{2}\, s\, \partial_{x_1 x_1}, \tag{75}$$

$$\partial_{\Sigma_{12}} = \frac{1}{2}\, s\, \partial_{x_1 x_2}, \tag{76}$$

$$\partial_{\Sigma_{22}} = \frac{1}{2}\, s\, \partial_{x_2 x_2}. \tag{77}$$

With regard to the differential relationships over the parameters s, Σ_{11} and Σ_{22}, the corresponding partial differential equations are parabolic and can thereby only be evolved in the corresponding positive directions. Hence, the expressions for ∂_s, $\partial_{\Sigma_{11}}$ and $\partial_{\Sigma_{22}}$ correspond to the notions of infinitesimal generators for multi-parameter semi-groups. The differential relationship over the parameter Σ_{12} does, on the other hand, correspond to a hyperbolic partial differential equation and could therefore be evolved in both directions.

4.3 Spatio-Temporal Receptive Fields Based on Isotropic Spatial Smoothing

Over a joint spatio-temporal image domain, let us first combine an isotropic Gaussian kernel over the spatial domain, for which the spatial covariance matrix Σ is a unit matrix $\Sigma = I$, with a non-causal temporal Gaussian kernel of the form

$$h(t;\ \tau) = \frac{1}{\sqrt{2\pi\tau}}\, e^{-t^2/2\tau}, \tag{78}$$

which for any value of the velocity parameter $v = (v_1, v_2)^T \in \mathbb{R}^2$ gives the following joint spatio-temporal Gaussian kernel

$$T(x, t;\ s, \tau, v) = g(x - v\,t;\ s, I)\, h(t;\ \tau) = \frac{1}{2\sqrt{2}\pi^{3/2}\, s\sqrt{\tau}}\, e^{-(x - v\,t)^T (x - v\,t)/2s}\, e^{-t^2/2\tau}. \tag{79}$$

The corresponding spatio-temporal scale-space representation $L \colon \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}^2 \to \mathbb{R}$ of any 2+1-D video sequence $f \colon \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ is then given by

$$L(\cdot, \cdot;\ s, \tau, v) = T(\cdot, \cdot;\ s, \tau, v) * f(\cdot, \cdot), \tag{80}$$

for which it is natural to introduce a velocity-adapted temporal derivative operator according to

$$\partial_{\bar{t}} = \partial_t + v_1\, \partial_{x_1} + v_2\, \partial_{x_2}. \tag{81}$$

Differentiating the spatio-temporal kernel $T(x, t;\ s, \tau, v)$ according to (79) with respect to the spatial image coordinates $x = (x_1, x_2)^T$ and the time variable t up to order 2, as well as with respect to the spatial scale parameter s, the temporal scale parameter τ and the elements v_1 and v_2 of the velocity vector v, leads to the partial derivatives shown in Fig. 7 in Appendix A.

[5] The computations of these expressions, as well as all the other generalized Hermite polynomials in this paper, have been performed in Wolfram Mathematica.

From these derivative expressions, we can in turn verify the following infinitesimal relationships over variations of the set of filter parameters (s, τ, v) of the receptive fields:

$$\partial_s = \frac{1}{2}\left(\partial_{xx} + \partial_{yy}\right), \tag{82}$$

$$\begin{aligned} \partial_\tau &= \frac{1}{2}\partial_{\bar{t}\bar{t}} \\ &= \frac{1}{2}\Big(\partial_{tt} + 2v_1\,\partial_{xt} + 2v_2\,\partial_{yt} \\ &\qquad + v_1^2\,\partial_{xx} + 2v_1\,v_2\,\partial_{xy} + v_2^2\,\partial_{yy}\Big), \end{aligned} \tag{83}$$

$$\partial_{v_1} = \tau\,\partial_{x_1\bar{t}}, \tag{84}$$

$$\partial_{v_2} = \tau\,\partial_{x_2\bar{t}}. \tag{85}$$

Here, concerning the differential relationships over the parameters s and τ, the corresponding partial differential equations are parabolic and can therefore only be evolved in the corresponding positive directions. Hence, the expressions for ∂_s and ∂_τ correspond to the notions of infinitesimal generators for multi-parameter semi-groups. The differential relationship over the parameters v_1 and v_2 do, on the other hand, correspond to a hyperbolic partial differential equations and could therefore be evolved in both directions.

4.4 Spatio-Temporal Receptive Fields Based on Affine Gaussian Smoothing

To combine the variabilities under spatial affine transformations and Galilean transformations, let us finally consider a joint spatio-temporal kernel obtained by combining the general affine Gaussian kernel $g(x;\ s, \Sigma)$ according to (1) with the non-causal temporal Gaussian kernel $h(t;\ \tau) = g_{1D}(t;\ \tau)$ according to (78), leading to the joint spatio-temporal kernel

$$T(x, t;\ s, \Sigma, \tau, v) = g(x - v\,t;\ s, \Sigma)\,g_{1D}(t;\ \tau) \tag{86}$$

with the corresponding spatio-temporal scale-space representation $L\colon \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{S}^2_+ \times \mathbb{R}_+ \times \mathbb{R}^2 \to \mathbb{R}$ of any 2+1-D video sequence $f\colon \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$: according to

$$L(\cdot, \cdot;\ s, \Sigma, \tau, v) = T(\cdot, \cdot;\ s, \Sigma, \tau, v) * f(\cdot, \cdot). \tag{87}$$

Differentiating the above spatio-temporal kernel $T(x, t;\ s, \Sigma, \tau, v)$ with respect the spatial image coordinates $x = (x_1, x_2)^T$ and the time variable t up to order 2, as well as with respect to the spatial scale parameter s, the elements Σ_{11}, Σ_{12} and Σ_{22} of the spatial covariance matrix Σ, the temporal scale parameter τ and the elements v_1 and v_2 of the velocity vector v, then give the partial derivatives shown in Fig. 8 in Appendix B.

From these derivative expressions, we can in turn verify the following infinitesimal relationships over the set of the filter parameters (s, Σ, τ, v) of the receptive fields:

$$\begin{aligned} \partial_s &= \frac{1}{2}\nabla_x^T\,\Sigma\,\nabla_x = \\ &= \frac{1}{2}\left(\Sigma_{11}\,\partial_{x_1x_1} + 2\Sigma_{12}\,\partial_{x_1x_2} + \Sigma_{22}\,\partial_{x_2x_2}\right), \end{aligned} \tag{88}$$

$$\partial_{\Sigma_{11}} = \frac{1}{2}\,s\,\partial_{x_1x_1}, \tag{89}$$

$$\partial_{\Sigma_{12}} = \frac{1}{2}\,s\,\partial_{x_1x_2}, \tag{90}$$

$$\partial_{\Sigma_{22}} = \frac{1}{2}\,s\,\partial_{x_2x_2}, \tag{91}$$

$$\begin{aligned} \partial_\tau &= \frac{1}{2}\partial_{\bar{t}\bar{t}} \\ &= \frac{1}{2}\Big(\partial_{tt} + 2v_1\,\partial_{xt} + 2v_2\,\partial_{yt} \\ &\qquad + v_1^2\,\partial_{xx} + 2v_1\,v_2\,\partial_{xy} + v_2^2\,\partial_{yy}\Big), \end{aligned} \tag{92}$$

$$\partial_{v_1} = \tau\,\partial_{x_1\bar{t}}, \tag{93}$$

$$\partial_{v_2} = \tau\,\partial_{x_2\bar{t}}. \tag{94}$$

For this joint combination of the computational structures in the two partially simplified receptive field models in Sects. 4.2 and 4.3, for the differential relationships over the parameters s, Σ_{11}, Σ_{22} and τ, it holds that the corresponding partial differential equations are parabolic and can therefore only be evolved in the corresponding positive directions. Hence, the expressions for ∂_s, $\partial_{\Sigma_{11}}$, $\partial_{\Sigma_{22}}$ and ∂_τ correspond to the notions of infinitesimal generators for multi-parameter semi-groups. The differential relationships over the parameters Σ_{12}, v_1 and v_2 do, on the other hand, correspond to hyperbolic partial differential equations and could therefore be evolved in both directions.

5 Macroscopic Cascade Structures for the Generalized Gaussian Derivative Model

To avoid explicit integration of the previous infinitesimal relationships between the receptive field responses for different values of the filter parameters, an alternative way to model such relationships is in terms of macroscopic relations. For simplicity, we again start by deriving such relations for the receptive field models that are purely based on smoothing with Gaussian kernels, thus initially excluding the non-causal spatio-temporal receptive field model based on convolution with the time-causal limit kernel (4) over the temporal domain.

5.1 Cascade Smoothing Properties Based on Spatial or Spatio-Temporal Covariance Matrices

Due to the special properties of the N-dimensional Gaussian kernel, it holds that the convolution of two Gaussian kernels $g \colon \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{S}_+^N \to \mathbb{R}$

$$g(\cdot;\ \mu_1, \Sigma_1) * g(\cdot;\ \mu_2, \Sigma_2) = g(\cdot;\ \mu_1 + \mu_2, \Sigma_1 + \Sigma_2) \tag{95}$$

with mean vectors $m_1, m_2 \in \mathbb{R}^N$ and spatial covariance matrices $\Sigma_1, \Sigma_2 \in \mathbb{S}_+^N$ is a Gaussian kernel with added mean vectors and added covariance matrices.

5.1.1 Purely Spatial Receptive Fields Based on Affine Gaussian Smoothing

Applied to the purely spatial smoothing operation, corresponding to convolution with the affine Gaussian kernel, it therefore follows that the scale-space representation $L(\cdot;\ s_2, \Sigma_2)$ at a coarser spatial scale (s_2, Σ_2) can be computed from the scale-space representation $L(\cdot;\ s_1, \Sigma_1)$ at any finer spatial scale (s_1, Σ_1) according to

$$L(\cdot;\ s_2, \Sigma_2) = g(\cdot;\ \Delta s, \Delta\Sigma) * L(\cdot;\ s_1, \Sigma_1). \tag{96}$$

This result holds provided the filter parameters $(\Delta s, \Delta\Sigma)$ of the incremental smoothing kernel $g(\cdot;\ \Delta s, \Delta\Sigma)$ satisfy

$$s_2\, \Sigma_2 = \Delta s\, \Delta\Sigma + s_1\, \Sigma_1, \tag{97}$$

that is, provided that the incremental effective spatial covariance matrix

$$\Delta s\, \Delta\Sigma = s_2\, \Sigma_2 - s_1\, \Sigma_1 \tag{98}$$

is a positive symmetric definite matrix. In other words, the resulting cascade smoothing property according to (96) can be computed in the forward direction of a cone in the parameter space of the purely spatial kernels.

Due to the linearity of the corresponding operators, this cascade smoothing property extends to the receptive field responses defined by applying spatial derivative operators $\mathcal{D}_x$ to the spatial scale-space representation of the form

$$\mathcal{R}\, f(\cdot) = \mathcal{D}_x L(\cdot;\ s, \Sigma), \tag{99}$$

thus implying an explicit cascade smoothing property for spatial-derivative-based receptive fields of the form

$$\mathcal{D}_x L(\cdot;\ s_2, \Sigma_2) = T(\cdot;\ \Delta s, \Delta\Sigma) * \mathcal{D}_x L(\cdot;\ s_1, \Sigma_1). \tag{100}$$

5.1.2 Joint Spatio-Temporal Receptive Fields Based on Affine Gaussian Smoothing

To express a corresponding cascade smoothing property over the joint spatio-temporal domain, let us make use of the fact that the composed effect of the smoothing operation, with a 2-D either isotropic Gaussian kernel or an anisotropic affine Gaussian kernel over the spatial domain in combination with smoothing with a 1-D non-causal temporal Gaussian kernel over the temporal domain, can equivalently be obtained by smoothing with a joint spatio-temporal Gaussian kernel over the 2+1-D joint spatio-temporal domain. Thus, when using an anisotropic affine spatial Gaussian kernel, leading to a joint spatio-temporal kernel of the form (86)

$$T(x, t;\ s, \Sigma, \tau, v) = g(x - v\,t;\ s, \Sigma)\, h(t;\ \tau)$$
$$= \frac{1}{2\sqrt{2}\pi^{3/2}\, s\sqrt{\tau}}\, e^{-(x-v\,t)^T \Sigma^{-1} (x-v\,t)/2s}\, e^{-t^2/2\tau}, \tag{101}$$

this operation can equivalently be described as a 2+1-D convolution with a joint spatio-temporal kernel $g_{2+1\text{-D}} \colon \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{S}_+^3 \to \mathbb{R}$ of the form

$$g_{2+1\text{-D}}(p;\ m_{2+1\text{-D}}, \Sigma_{2+1\text{-D}})$$
$$= \frac{1}{(2\pi)^{3/2}\sqrt{\det \Sigma_{2+1\text{-D}}}}\, e^{-(p - m_{2+1\text{-D}})^T \Sigma_{2+1\text{-D}}^{-1} (p - m_{2+1\text{-D}})/2} \tag{102}$$

over the spatio-temporal image coordinates $p = (x_1, x_2, t)^T$.

To determine the filter parameters $m_{2+1\text{-D}}$ and $\Sigma_{2+1\text{-D}}$ in the joint spatio-temporal smoothing kernel (102), let us with the reformulation of the spatio-temporal smoothing kernel (79) as

$$T(p;\ s, \Sigma, \tau, v) = T(x, t;\ s, \Sigma, \tau, v) \tag{103}$$

calculate the joint mean vector $m_{2+1\text{-D}}$ and the joint covariance matrix $\Sigma_{2+1\text{-D}}$ according to

$$m_{2+1\text{-D}} = \frac{\int_{p \in \mathbb{R}^{2+1}} p\, T(p;\ s, \Sigma, \tau, v)\, dp}{\int_{p \in \mathbb{R}^{2+1}} T(p;\ s, \Sigma, \tau, v)\, dp}, \tag{104}$$

$$\Sigma_{2+1\text{-D}} = \frac{\int_{p \in \mathbb{R}^{2+1}} p\, p^T\, T(p;\ s, \Sigma, \tau, v)\, dp}{\int_{p \in \mathbb{R}^{2+1}} T(p;\ s, \Sigma, \tau, v)\, dp} - m_{2+1\text{-D}}\, m_{2+1\text{-D}}^T, \tag{105}$$

which gives $m_{2+1\text{-D}} = (0, 0, 0)^T$ and

$$\Sigma_{2+1\text{-D}} = \begin{pmatrix} \Sigma_{11}\, s + \tau\, v_1^2 & \Sigma_{12}\, s + \tau\, v_1\, v_2 & \tau\, v_1 \\ \Sigma_{12}\, s + \tau\, v_1\, v_2 & \Sigma_{22}\, s + \tau\, v_2^2 & \tau\, v_2 \\ \tau\, v_1 & \tau\, v_2 & \tau \end{pmatrix}. \tag{106}$$

Notably, concerning the structure of the expression for the joint spatio-temporal covariance matrix, a general 3×3 spatio-temporal covariance matrix has 6 degrees of freedom, while this expression also has 6 effective degrees of freedom in terms of the filter parameters s, Σ_{11}, Σ_{12}, Σ_{22}, τ, v_1 and v_2, if we disregard the overparameterization induced by the combination of the spatial scale parameter s with the elements Σ_{11}, Σ_{12}, Σ_{22} of the spatial covariance matrix Σ.

Thus, let us assume that we have two configurations of parameter settings that would correspond to a cascade smoothing property over the filter parameters for two joint spatio-temporal smoothing kernels of the form (101), with

- the joint convolution kernel $T(\cdot, \cdot;\ s_i, \Sigma_i, \tau_i, v_i)$ in the lower layer with the velocity vector $v_i = (v_{i,1}, v_{i,2})^T$ and the elements of the spatial covariance matrix Σ_i being $\Sigma_{i,11}$, $\Sigma_{i,12}$ and $\Sigma_{i,22}$, and with
- the joint convolution kernel $T(\cdot, \cdot;\ s_j, \Sigma_j, \tau_j, v_j)$ in the higher layer with the velocity vector $v_j = (v_{j,1}, v_{j,2})^T$ and the elements of the spatial covariance matrix Σ_j being $\Sigma_{j,11}$, $\Sigma_{j,12}$ and $\Sigma_{j,22}$.

Then, from the additive properties of the joint spatio-temporal covariance matrices over 2+1-D space-time, a cascade smoothing property would be possible, if there would exist

- an incremental joint spatio-temporal convolution kernel $T(\cdot, \cdot;\ \Delta s, \Delta\Sigma, \Delta\tau, \Delta v)$ with the velocity vector $\Delta v = (\Delta v_1, \Delta v_2)^T$ and the elements of the spatial covariance matrix $\Delta\Sigma$ being $\Delta\Sigma_{11}$, $\Delta\Sigma_{12}$ and $\Delta\Sigma_{22}$,

such that

$$T(\cdot, \cdot;\ s_j, \Sigma_j, \tau_j, v_j) = T(\cdot, \cdot;\ \Delta s, \Delta\Sigma, \Delta\tau, \Delta v) * T(\cdot, \cdot;\ s_i, \Sigma_i, \tau_i, v_i), \quad (107)$$

thus implying an explicit cascade smoothing property of the form

$$L(\cdot, \cdot;\ s_j, \Sigma_j, \tau_j, v_j) = T(\cdot, \cdot;\ \Delta s, \Delta\Sigma, \Delta\tau, \Delta v) * L(\cdot, \cdot;\ s_i, \Sigma_i, \tau_i, v_i) \quad (108)$$

between the non-causal spatio-temporal scale-space representations $L(\cdot, \cdot;\ s_i, \Sigma_i, \tau_i, v_i)$ and $L(\cdot, \cdot;\ s_j, \Sigma_j, \tau_j, v_j)$.

Thereby, the following relations between the filter parameters would have to be satisfied:

$$\Sigma_{j,11}\, s_j + \tau_j\, v_{j,1}^2 = = \Delta\Sigma_{11}\, \Delta s + \Delta\tau\, \Delta v_1^2 + \Sigma_{i,11}\, s_i + \tau_i\, v_{i,1}^2, \quad (109)$$

$$\Sigma_{j,12}\, s_j + \tau_j\, v_{j,1}\, v_{j,2} = = \Delta\Sigma_{12}\, \Delta s + \Delta\tau\, \Delta v_1\, \Delta v_2 + \Sigma_{i,12}\, s_i + \tau_i\, v_{i,1}\, v_{i,2}, \quad (110)$$

$$\Sigma_{j,22}\, s_j + \tau_j\, v_{j,2}^2 = \Delta\Sigma_{22}\, \Delta s + \Delta\tau\, \Delta v_2^2 + \Sigma_{i,22}\, s_i + \tau_i\, v_{i,2}^2, \quad (111)$$

$$\tau_j\, v_{j,1} = \Delta\tau\, \Delta v_1 + \tau_i\, v_{i,1}, \quad (112)$$

$$\tau_j\, v_{j,2} = \Delta\, \tau\, \Delta\, v_2 + \tau_i\, v_{i,2}, \quad (113)$$

$$\tau_j = \Delta\tau + \tau_i. \quad (114)$$

Solving for the parameters Δs, $\Delta\Sigma_{11}$, $\Delta\Sigma_{12}$, $\Delta\Sigma_{22}$, $\Delta\tau$, Δv_1 and Δv_2 of the incremental spatio-temporal convolution kernel $T(\cdot, \cdot;\ \Delta s, \Delta\Sigma, \Delta\tau, \Delta v)$ then gives

$$\Delta\tau = \tau_j - \tau_i, \quad (115)$$

$$\Delta v_1 = \frac{\tau_j\, v_{j,1} - \tau_i\, v_{i,1}}{\tau_j - \tau_i}, \quad (116)$$

$$\Delta v_2 = \frac{\tau_j\, v_{j,2} - \tau_i\, v_{i,2}}{\tau_j - \tau_i}, \quad (117)$$

$$\Delta s\, \Delta\Sigma_{11} = \Sigma_{j,11}\, s_j - \Sigma_{i,11}\, s_i + \frac{\tau_i\, \tau_j\, (v_{i,1} - v_{j,1})^2}{\tau_i - \tau_j}, \quad (118)$$

$$\Delta s\, \Delta\Sigma_{12} = \Sigma_{j,12}\, s_j - \Sigma_{i,12}\, s_i + \quad (119)$$

$$+ \frac{\tau_i\, \tau_j\, (v_{i,1} - v_{j,1})(v_{i,2} - v_{j,2})}{\tau_i - \tau_j}, \quad (120)$$

$$\Delta s\, \Delta\Sigma_{22} = \Sigma_{j,22}\, s_j - \Sigma_{i,22}\, s_i + \frac{\tau_i\, \tau_j\, (v_{i,2} - v_{j,2})^2}{\tau_i - \tau_j}. \quad (121)$$

Given that the resulting parameters Δs, $\Delta\Sigma_{11}$, $\Delta\Sigma_{12}$, $\Delta\Sigma_{22}$, $\Delta\tau$, Δv_1 and Δv_2 for the incremental spatio-temporal convolution kernel should obey the consistency requirements

$$\Delta\Sigma_{11}\, \Delta s \geq 0, \quad (122)$$

$$\Delta\Sigma_{22}\, \Delta s \geq 0, \quad (123)$$

$$(\Delta\Sigma_{11}\, \Delta s)(\Delta\Sigma_{22}\, \Delta s) - (\Delta\Sigma_{12}\, \Delta s)^2 \geq 0, \quad (124)$$

$$\Delta\tau \geq 0, \quad (125)$$

it thereby follows that it is possible to implement a cascade smoothing property

$$L(\cdot, \cdot;\ s_j, \Sigma_j, \tau_j, v_j) = T(\cdot, \cdot;\ \Delta s, \Delta\Sigma, \Delta\tau, \Delta v) * L(\cdot, \cdot;\ s_i, \Sigma_i, \tau_i, v_i) \quad (126)$$

for the joint spatio-temporal scale-space representations of video sequences f according to

$$L(\cdot, \cdot;\ s, \Sigma, \tau, v) = T(\cdot, \cdot;\ s, \Sigma, \tau, v) * f(\cdot, \cdot). \quad (127)$$

Due to the linearity of the corresponding operators, this cascade smoothing property extends to receptive field responses defined by applying spatio-temporal derivative operators $\mathcal{D}_{x,t}$ to the spatio-temporal scale-space representation of the form

$$\mathcal{R} f(\cdot, \cdot) = \mathcal{D}_{x,t} L(\cdot, \cdot;\ s, \Sigma, \tau, v), \tag{128}$$

thus implying an explicit cascade smoothing property of the spatio-temporal derivative-based spatio-temporal receptive fields of the form

$$\begin{aligned} &\mathcal{D}_{x,t} L(\cdot, \cdot;\ s_j, \Sigma_j, \tau_j, v_j) \\ &\quad = T(\cdot, \cdot;\ \Delta s, \Delta\Sigma, \Delta\tau, \Delta v) * \mathcal{D}_{x,t} L(\cdot, \cdot;\ s_i, \Sigma_i, \tau_i, v_i). \end{aligned} \tag{129}$$

5.1.3 Joint Spatio-Temporal Receptive Fields Based on Isotropic Gaussian Smoothing

In relation to the results from the previous section, it is worth observing that if the spatial smoothing operation would instead only be performed based on an isotropic Gaussian kernel, with the spatial covariance matrix Σ equal to the unit matrix $\Sigma = I$, then the parameters (s, τ, v) in the spatio-temporal receptive fields would only span 4 out of the general 6 degrees of freedom in the joint spatio-temporal covariance matrix $\Sigma_{2+1\text{-D}}$ according to (106). Thereby, the configurations for which it would be possible to implement a corresponding cascade smoothing property

$$\begin{aligned} &L(\cdot, \cdot;\ s_j, \tau_j, v_j) = \\ &= T(\cdot, \cdot;\ \Delta s, \Delta\tau, \Delta v) * L(\cdot, \cdot;\ s_i, \tau_i, v_i) \end{aligned} \tag{130}$$

for the joint spatio-temporal scale-space representations of video sequences f according to

$$L(\cdot, \cdot;\ s, \tau, v) = T(\cdot, \cdot;\ s, \tau, v) * f(\cdot, \cdot) \tag{131}$$

would on the other hand be highly degenerate. Thus, beyond the previously mentioned closedness property under anisotropic spatial affine transformations, the use of an affine Gaussian kernel for spatial smoothing also offers clear advantages in terms of the ability for an idealized vision system, to implement spatio-temporal cascade smoothing properties for the receptive fields, in order to decrease the amount of computational work in filter bank implementations of joint spatio-temporal receptive fields.

5.1.4 The Special Case with Equal Image Velocities

Concerning the relationships derived in Sect. 5.1.2, it is of particular interest to consider the special case when all the image velocities are equal, *i.e.*,

$$v_j = \Delta v = v_i. \tag{132}$$

Then, Eqs. (112) and (113) both reduce to the criterion (114)

$$\tau_j = \Delta\tau + \tau_i. \tag{133}$$

Furthermore, given the assumption that this condition is satisfied, the relationships (109)–(111) then reduce to the following relationships

$$\Sigma_{j,11}\, s_j = \Delta\Sigma_{11}\, \Delta s + \Sigma_{i,11}\, s_i, \tag{134}$$

$$\Sigma_{j,12}\, s_j = \Delta\Sigma_{12}\, \Delta s + \Sigma_{i,12}\, s_i, \tag{135}$$

$$\Sigma_{j,22}\, s_j = \Delta\Sigma_{22}\, \Delta s + \Sigma_{i,22}\, s_i, \tag{136}$$

corresponding to similar relationships as for the elements of the purely spatial receptive field model in Equation (97).

An explanation for this simplicity of the resulting relationships is because of the Galilean covariance property of the spatio-temporal receptive field model, which then up to a Galilean transformation is isomorphic to a purely space-time separable model for the spatio-temporal receptive fields based on spatio-temporal smoothing operations of the form

$$\begin{aligned} &T(x, t;\ s, \Sigma, \tau) = g(x;\ s, \Sigma)\, h(t;\ \tau) = \\ &= \frac{1}{2\sqrt{2}\pi^{3/2}\, s\sqrt{\tau}}\, e^{-x^T\Sigma^{-1}x/2s}\, e^{-t^2/2\tau}. \end{aligned} \tag{137}$$

With regard to the special case with joint spatio-temporal receptive fields based on isotropic Gaussian smoothing with $\Sigma = I$ considered in Sect. 5.1.3, it is, however, interesting to note that those degeneracies do not in any negative way affect the possibility for the receptive fields to span that subgroup of receptive fields covariant to spatial and temporal scaling transformations and Galilean transformations.

6 Relationships Between Receptive Field Responses for Different Parameter Settings for Spatio-Temporal Receptive Fields Based on Temporal Smoothing with The Time-Causal Limit Kernel

In the treatment so far, we have made use of special algebraic properties, as arising from using spatial and spatio-temporal smoothing kernels solely based on Gaussian kernels, for deriving both infinitesimal and macroscopic relationships between the receptive field responses computed for different values of the filter parameters.

When replacing the non-causal temporal Gaussian kernel $h(t;\ \tau) = g(t;\ \tau)$ according to (3) by the time-causal limit kernel $h(t;\ \tau) = \psi(t;\ \tau, c)$ according to (4), where $c > 1$ is a distribution parameter that specifies the ratio between the temporal scale levels in units of the temporal standard deviation of the kernel, we obtain a joint spatio-temporal smoothing kernel of the form

$$T(x, t;\ s, \Sigma, \tau, v, c) = g(x - v\,t;\ s, \Sigma)\, \psi(t;\ \tau, c) = \frac{1}{2\pi\, s}\, e^{-(x-v\,t)^T \Sigma^{-1} (x-v\,t)/2s}\, \psi(t;\ \tau, c). \quad (138)$$

This kernel cannot, however, be differentiated with respect to the here fully discrete temporal scale parameter $\tau_i = \tau_0\, c^{2i}$. Additionally, expressing the derivative of the time-causal limit kernel with respect to the temporal variable t is neither straightforward, because of the relatively complex explicit expression for the time-causal limit kernel over the temporal domain:

$$\Psi(\cdot;\ \tau, c) = *_{k=1}^{\infty} h_{\exp}(\cdot;\ \mu_k), \quad (139)$$

where

$$h_{\exp}(t;\ \mu_k) = \begin{cases} \frac{1}{\mu_k}\, e^{-t/\mu_k} & t \geq 0, \\ 0 & t < 0, \end{cases} \quad (140)$$

and the time constants μ_k of the individual truncated exponential kernels are given by

$$\mu_k = c^{-k} \sqrt{c^2 - 1} \sqrt{\tau}. \quad (141)$$

Specifically, we do not have any straightforward correspondence to generalized Hermite polynomials based on temporal derivatives of the time-causal limit kernel.

Furthermore, in terms of cascade smoothing properties, the time-causal limit kernel does not form a semi-group as the 1-D temporal Gaussian kernel $h(t;\ \tau) = g(t;\ \tau)$ does. Instead, the time-causal limit obeys the following cascade smoothing property between adjacent levels of temporal scales (see Lindeberg [54] Eq. (28)):

$$\Psi(\cdot;\ \tau, c) = h_{\exp}(\cdot;\ \tfrac{\sqrt{c^2-1}}{c} \sqrt{\tau}) * \Psi(\cdot;\ \tfrac{\tau}{c^2}, c). \quad (142)$$

From these structural differences, arising from using the time-causal limit kernel as the temporal smoothing kernel instead of the non-causal temporal Gaussian kernel, we cannot expect to be able to directly apply similar methodologies for deriving infinitesimal or macroscopic evolution properties over the filter parameters for the time-causal spatio-temporal kernel (138), as we used for deriving such relationships for the fully Gaussian smoothing-based receptive field models in Sects. 4 and 5.

As will be described further in Sect. 6.1 below, the infinitesimal relationships do, however, carry over regarding the purely spatial image coordinates x, the spatial scale parameter s, as well as for the elements Σ_{11}, Σ_{12} and Σ_{22} of the spatial covariance matrix Σ. Furthermore, as will be described in the following Sect. 6.2, in the special case when the image velocities are all equal in the cascade model for the receptive field responses, the macroscopic relationships for the spatial components in the spatio-temporal receptive field model do also carry over from the purely spatial case.

6.1 Infinitesimal Relationships Between Receptive Field Responses Computed for Different Values of the Filter Parameters

Regarding infinitesimal relationships, the differential relationships that we derived in Sect. 4.4 regarding derivative operators solely expressed over the spatial domain will also hold if we replace the non-causal Gaussian kernel $h(t;\ \tau) = g_{1D}(t;\ \tau)$ in (86) by the time-causal limit kernel $h(t;\ \tau) = \psi(t;\ \tau, c)$. Thus, if we for any spatio-temporal video stream $f : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ define a time-causal spatio-temporal scale-space representation according to

$$L(\cdot, \cdot;\ s, \Sigma, \tau, v, c) = T(\cdot, \cdot;\ s, \Sigma, \tau, v, c) * f(\cdot, \cdot) \quad (143)$$

with the joint time-causal spatio-temporal smoothing kernel $T(\cdot;\ s, \Sigma, \tau, v, c)$ according to (138), then it follows that the following infinitesimal relationships will hold over the purely spatial filter parameters (s, Σ):

$$\partial_s = \frac{1}{2} \nabla_x^T \Sigma\, \nabla_x = \frac{1}{2} \left(\Sigma_{11}\, \partial_{x_1 x_1} + 2\Sigma_{12}\, \partial_{x_1 x_2} + \Sigma_{22}\, \partial_{x_2 x_2} \right), \quad (144)$$

$$\partial_{\Sigma_{11}} = \frac{1}{2}\, s\, \partial_{x_1 x_1}, \quad (145)$$

$$\partial_{\Sigma_{12}} = \frac{1}{2}\, s\, \partial_{x_1 x_2}, \quad (146)$$

$$\partial_{\Sigma_{22}} = \frac{1}{2}\, s\, \partial_{x_2 x_2}. \quad (147)$$

This can be easily understood by observing that when calculating the ratios between derivatives with respect to either the spatial coordinates x or the spatially based filter parameters (s, Σ), the occurrence of the time-causal limit kernel will cancel in the resulting expressions. Thus, we obtain similar results for these ratios computed from the time-causal spatio-temporal smoothing kernel as shown in Fig. 9 in Appendix C, as we previously obtained for the corresponding non-causal spatio-temporal smoothing kernel shown in Fig. 8 in Appendix B.

By comparing those generalized Hermite polynomials, we can then establish that the infinitesimal relationships (144)–(147) are valid.

Notably, without differentiating the time-causal limit kernel, we can also compute the generalized Hermite polynomials corresponding to differentiation with respect to the elements v_1 and v_2 of the velocity vector v. So far, we have, however, not yet been able to match these partial derivatives to corresponding expressions in terms of derivatives with respect to the spatio-temporal image coordinates (x_1, x_2, t).

6.2 Macroscopic Cascade Structures Between Receptive Field Responses Computed for Different Values of the Filter Parameters

Based on the cascade smoothing property (142) of the time-causal limit kernel (4), let us for spatio-temporal smoothing kernels $T(x, t;\; s, \Sigma, \tau, v, c)$ of the form (138) consider formulating a cascade smoothing property for the time-causal spatio-temporal kernels of the form

$$\begin{aligned} & L(\cdot, \cdot;\; s_j, \Sigma_j, \tau_j, v_j, c) \\ & \quad = \Delta T(\cdot, \cdot;\; \Delta s, \Delta\Sigma, \Delta\tau, \Delta v, c) \\ & \quad * L(\cdot, \cdot;\; s_i, \Sigma_i, \tau_i, v_i, c), \end{aligned} \tag{148}$$

where the incremental kernel $\Delta T(\cdot, \cdot;\; \Delta s, \Delta\Sigma, \Delta\tau, \Delta v, c)$ is given by

$$\begin{aligned} & \Delta T(x, t;\; \Delta s, \Sigma, \Delta\tau, \Delta v, c) = \\ & = g(x - \Delta v\, t;\; \Delta s, \Delta\Sigma)\, h_{\exp}(t;\; \sqrt{\Delta\tau}) \\ & = \frac{1}{2\pi\, \Delta s}\, e^{-(x - \Delta v\, t)^T \Delta\Sigma^{-1} (x - \Delta v\, t)/2\Delta s} \times \\ & \quad h_{\exp}(t;\; \sqrt{\Delta\tau}) \end{aligned} \tag{149}$$

here, for simplicity, restricted to adjacent temporal scales

$$\tau_i = \frac{\tau_j}{c^2} \tag{150}$$

with

$$\Delta\tau = \tau_j - \tau_i = \frac{c^2 - 1}{c^2}\, \tau_j. \tag{151}$$

The motivation for the last relationships is that the temporal scale levels in these types of temporal and spatio-temporal scale-space representations are inherently discrete according to a geometric distribution of the form (6).

6.2.1 The Special Case with Equal Image Velocities

Unfortunately, it appears as a rather complex problem to derive closed-form expressions for the relationships between the parameters $(\Delta s, \Delta\Sigma, \Delta\tau, \Delta v)$ of the incremental kernel from the parameters $(s_i, \Sigma_i, \tau_i, v_i)$ and $(s_j, \Sigma_j, \tau_j, v_j)$ of two involved layers of receptive field responses in the general case, when the image velocities Δv, v_i and v_j may be different. A main reason to technical complications is that the time-causal limit kernel does not have any compact explicit expression, since it is given as the convolution of an infinite number of truncated exponential kernels.

In the special case when the image velocities are all equal

$$v_j = \Delta v = v_i, \tag{152}$$

we can, however, in analogy with the previous treatment in Sect. 5.1.4, make use of the Galilean covariance property to infer that, beyond the relationship between adjacent scale levels according to (150)

$$\tau_i = \frac{\tau_j}{c^2}, \tag{153}$$

the purely spatial parameters of the receptive fields should obey similar relationships as in Equations (134)–(136):

$$\Sigma_{j,11}\, s_j = \Delta\Sigma_{11}\, \Delta s + \Sigma_{i,11}\, s_i, \tag{154}$$

$$\Sigma_{j,12}\, s_j = \Delta\Sigma_{12}\, \Delta s + \Sigma_{i,12}\, s_i, \tag{155}$$

$$\Sigma_{j,22}\, s_j = \Delta\Sigma_{22}\, \Delta s + \Sigma_{i,22}\, s_i. \tag{156}$$

In this way, we can hence determine the parameters of the incremental spatio-temporal kernel according to (149) in the special case of equal velocity parameters v_i, Δv and v_j.

We leave it as an open problem for future work to analyze the potential applicability of the incremental kernel of the form (149) to formulate time-causal cascade model between the scale-space representations $L(\cdot, \cdot;\; s_i, \Sigma_i, \tau_i, v_i, c)$ and $L(\cdot, \cdot;\; s_j, \Sigma_j, \tau_j, v_j, c)$ at adjacent scales τ_i and τ_j of the form (148) in the case of non-equal velocity parameters v_i, Δv and v_j.

7 Summary and Discussion

We have presented an in-depth theoretical analysis about how receptive field responses computed using linear receptive fields corresponding to simple cells in the primary visual cortex, as expressed in terms of the generalized Gaussian derivative model for visual receptive fields, can be related between different values of the filter parameters.

After an overview of the generalized Gaussian derivative model for visual receptive fields in Sect. 3, with emphasis on its relation to handle variabilities in image structures, as generated by natural geometric image transformations, and as modeled by local linearizations in terms of (i) uniform spatial scaling transformations, (ii) spatial affine transformations, (iii) Galilean transformations and (iv) temporal scaling transformations, we have specifically, from the requirement of covariance properties of the receptive fields under such geometric image transformations, motivated the importance of multi-parameter families of visual receptive fields. This motivation originates from the desirable requirement of making it possible to, up to first order of approximation, match the receptive field responses that have been computed from image data acquired under different viewing conditions. The requirement of closedness of the representation of receptive field responses under such geometric image transformations then means that the shapes of the receptive fields ought to be expanded over the degrees of freedom of the corresponding geometric image transformations.

For our considered axiomatically determined normative theory of visual receptive fields, this leads to spatial and spatio-temporal receptive fields that are based on applying spatial and/or spatio-temporal derivative operators to a multi-parameter spatial and/or spatio-temporal scale-space representation of the either purely spatial or joint spatio-temporal image data. Specifically, we have by the covariance properties of the receptive fields in Eqs. (30), (31) and (39)–(40) related the variabilities in the shapes of the receptive fields to the parameters of the general class of composed geometric image transformations of the form (34) and (35), which, as illustrated in Fig. 4, represents the first-order variability in image structures, when observing the surfaces of smooth objects in dynamic scenes.

Then, we have in Sects. 4 and 5 developed two main ways of interrelating the receptive field responses between different parameter settings for this generalized Gaussian derivative model for visual receptive fields, either in terms of (i) infinitesimal differential relations closely related to the notions of Lie groups and the infinitesimal generators of continuous semi-groups, or (ii) macroscopic cascade smoothing relations, closely related to the notion of Lie algebras, although with a directional preference, in the respect that the evolution can only performed in positive directions for those dimensions of the parameter space that correspond to uni-directional semi-group structures, as opposed to bi-directional Lie group structures.

Specifically, in Sect. 4, we studied three variants of models for spatial or spatio-temporal smoothing operations in the receptive fields based on pure Gaussian smoothing over the spatial or spatial–temporal domains, in terms of either (i) a purely spatial model with 3 effective degrees of freedom in the parameter space, (ii) a joint spatio-temporal model with a restriction to isotropic Gaussian smoothing over the spatial domain, resulting in 4 effective degrees of freedom in the parameter space, and (ii) a joint spatio-temporal model with general affine Gaussian smoothing over the spatial domain, leading to 6 effective degrees of freedom in the parameter space. By the special algebraic structures of the resulting receptive field representations, when using solely Gaussian smoothing over both the spatial and the temporal domains, we could introduce the notion of generalized Hermite polynomials of the resulting spatial or spatio-temporal smoothing operations, to substantially simplify the process of deriving and proving the corresponding differential evolution equations over infinitesimal perturbations of the filter parameters, largely corresponding to the notion of infinitesimal generators for more traditional semi-groups.

Then, in Sect. 5 we derived macroscopic evolution properties of the either purely spatial or joint spatio-temporal receptive field representations based on the sole use of Gaussian smoothing operators over both the spatial and the temporal domains, by combining the separate Gaussian smoothing operations over the either spatial or temporal domains into a joint spatio-temporal Gaussian smoothing kernel over the joint spatio-temporal domain. Specifically, by calculating the joint spatio-temporal covariance matrix of that 2+1-D joint spatio-temporal smoothing kernel, we derived a joint spatio-temporal cascade smoothing property, which specifies how a receptive field response at a coarser spatial and temporal scale can be computed by applying an incremental spatio-temporal smoothing kernel to the receptive field responses at finer spatial and temporal scales.

The way that the macroscopic cascade smoothing properties in Sect. 5 are related to the infinitesimal evolution properties in Sect. 4 is structurally similar to the way a Lie algebra is related to a Lie group according to the exponential map, see *e.g.,* Hall [28] Sect. 3.7. A structural difference compared to Lie groups and Lie algebras, however, is that here we have directional preferences in the evolution equations with regard to some of the parameters, thus with closer relationships to the notions of infinitesimal generators and semi-groups, as described by Hille and Phillips [31], Pazy [68] and Goldstein [27]. The evolution equations that we have derived are, however, neither strict infinitesimal generators nor semi-groups with respect to all the parameters. In this respect, the derived relationships can be seen as hybrid relationships between Lie groups and Lie algebras on one side and infinitesimal generators and semi-groups on the other side.

In Sect. 6, we then extended the relationships derived for the non-causal fully Gaussian-based spatio-temporal model to the time-causal spatio-temporal model, where the temporal smoothing operation is performed by convolution with the time-causal limit kernel (4) instead of the non-causal temporal Gaussian kernel (3). Due to the discrete nature of the temporal scale levels τ_k in the corresponding spatio-temporal scale representation, differential evolution properties over the temporal scale parameter cannot be expressed for that time-causal spatio-temporal scale-space representation. Additionally, due to the current lack of sufficiently good tools to derive compact closed-form expressions for the temporal derivatives of the time-causal limit kernel, we restricted the differential evolution properties to derivatives with respect the spatial scale parameter s and the spatial covariance matrix Σ of the spatial affine Gaussian kernel, which are similar as for the non-causal spatio-temporal scale-space representation studied in Sect. 4.3. Similarly, regarding the notion of a cascade property for the case of time-causal spatio-temporal receptive fields, we restricted the treatment to the special case when the image velocity of the incremental kernel is equal to the image velocities used for computing the receptive field responses before and after the cascade smoothing step, which are also assumed to be equal.

From the viewpoint of the underlying normative theory for visual receptive fields, the receptive field shapes are parameterized by over the 4-D (although effectively 3-D) parameters (s, Σ) in the purely spatial case and by the 7-D (although effectively 6-D) parameters (s, Σ, τ, v) in the joint spatio-temporal case. If we regard this theory as an idealized model of the visual processing in the primary visual cortex, then flattening out these higher-dimensional spaces onto the 2-D cortical surface, implies that points p and parameters P that may be relatively nearby in the higher-dimensional spaces $(p;\ P) = (x;\ s, \Sigma)$ or $(p;\ P) = (x;\ s, \Sigma, \tau, v)$ may be further away from each other when flattened out on the 2-D cortical surface.

In this way, if we transfer the cascade smoothing properties derived in this paper onto such a flattened cortical manifold, then the corresponding connections between receptive fields at different positions on the cortical surface will lead to a set of "horizontal connections" between different neurons. Specifically, due to the flattening of the higher-dimensional manifold of parameterized receptive responses onto the 2-D cortical surface, some of these connections will by necessity be "long range" on the cortical surface, although they would have been of shorter range in the original higher-dimensional manifold of receptive field responses. In this respect, the presented theory offers a possible explanation of a certain kind of "horizontal connections" constructed from a different purpose than the more commonly used explanation of integrating receptive field responses at different positions in the visual field into higher-order "gestalt" entities, such as contours (see *e.g.,* Bolz and Gilbert [13], Bosking *et al.* [14], Ben-Shahar and Zucker [12] and Sarti *et al.* [82]). [6]

For modeling visual receptive fields, there is, as previously described in Sect. 2, also an alternative school of using idealized models based on Gabor functions, which consist of the multiplication of Gaussian functions with complex sine waves. For the purely Gaussian smoothing-based component in the Gabor model, it seems likely that structurally related evolution properties over the receptive field parameters could be obtained for the corresponding Gabor models of visual receptive fields. Regarding variations in the modulation frequency or the orientation of the complex sine wave component in the Gabor model, the corresponding relationships could, however, be expected to be of a different nature.

In this treatment, we have therefore focused solely on the topic of deriving evolution properties over the parameters of the generalized Gaussian derivative model, which obeys very special closedness properties that could not be expected to hold for more generic classes of kernels. Thus, we leave the topic of deriving corresponding infinitesimal, or, if possible, macroscopic relationships for the Gabor model to future work. For a complementary in-depth analysis of similarities *vs.* differences between the generalized Gaussian derivative and the Gabor models regarding orientation selectivity properties, see Lindeberg [58].

By the derived evolution properties over variations of the filter parameters of the receptive fields according to the generalized Gaussian derivative model for visual receptive fields, we do foremost obtain a better theoretical understanding about how receptive field responses are related between different values of the filter parameters. These either infinitesimal or macroscopic evolution properties can, in turn, be used in computational schemes to compute banks of receptive field responses suitable for matching the responses of receptive fields applied to image data that have been acquired under substantial variations in the viewing conditions, which according to the theory described in Sect. 3.5 implies that the shapes of the receptive fields would have to be expanded over the degrees of freedom of the corresponding geometric image transformations. Thereby, it becomes possible to, to first order of approximation, match the receptive field responses between different views of the same object or a similar spatio-temporal event and thus establish an identity operation between the receptive field responses computed for different observations of the same object or a similar spatio-temporal event.

These theoretical results do also have possible implications for theories about biological vision, where indeed such a model concerning an expansion over receptive field shapes over the degrees of freedom, corresponding to the degrees of freedom in the here studied generalized Gaussian derivative model for visual receptive fields, has in Lindeberg [60] been recently proposed for predicting and explaining variabilities of simple cells in the primary visual cortex of higher mammals. From statistics of the populations of neurons in the early visual pathway, with about 1 M output channels from the retina mapped to 1 M output channels from the lateral geniculate nucleus (LGN) to the primary visual cortex (V1), to about 190 M neurons in V1 with 37 M output channels (see DiCarlo *et al.* [20] Fig. 3), such a substantial expansion of the number of receptive fields from the LGN to V1 would indeed be consistent with an expansion of the shapes of the receptive fields over shape parameters of the receptive fields.

[6] Let us, however, remark that with this statement we do not in any way intend to argue against the use of "horizontal connections" between different visual neurons for computing higher-order "gestalt" entities, such as contours. Instead, our purpose is to propose a complementary explanation regarding a possible subset of such "horizontal connections".

Specifically, the evolution properties between receptive field responses for different sets of filter parameters derived in this paper provide a theoretical explanation for the alternative design strategy of a representation of the receptive field responses of simple cells proposed in Lindeberg [60], where an idealized vision system could by the use of cascade smoothing properties over the spatial and temporal scale parameters choose to only implement receptive field responses corresponding to the finest spatial and temporal scale levels in the sensorium, corresponding to the simple cells in the primary visual cortex of higher mammals. Then, at higher levels in the visual hierarchy, equivalent receptive field responses, corresponding to coarser levels of spatial and temporal scales, could be computed in an indirect manner, either based on integration of the infinitesimal evolution equations derived in Sect. 4 or based on the macroscopic evolution properties derived in Sect. 5.

See also Lindeberg [57] for a closely related theoretical study of this problem in relation to the degrees of freedom of spatial affine transformations, the degrees of freedom of the spatial component of the generalized Gaussian derivative model, and existing neurophysiological support for such variabilities of the shapes of simple cells in the primary visual cortex.

Appendix

A Generalized Hermite Polynomials for the Isotropic Spatio-Temporal Gaussian Kernel $T(x, t;\; s, \tau, v)$

Figure 7 lists generalized Hermite polynomials for the isotropic spatio-temporal Gaussian kernel $T(x, t;\; s, \tau, v)$ according to (79).

$$\frac{\partial_{x_1} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{t v_1 - x_1}{s}, \tag{157}$$

$$\frac{\partial_{x_2} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{t v_2 - x_2}{s}, \tag{158}$$

$$\frac{\partial_{t} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{-t\left(v_1^2 + v_2^2\right) + v_1 x_1 + v_2 x_2}{s} - \frac{t}{\tau}, \tag{159}$$

$$\frac{\partial_{\bar{t}} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = -\frac{t}{\tau}, \tag{160}$$

$$\frac{\partial_{x_1 x_1} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{(x_1 - t v_1)^2 - s}{s^2}, \tag{161}$$

$$\frac{\partial_{x_1 x_2} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{(t v_1 - x_1)(t v_2 - x_2)}{s^2}, \tag{162}$$

$$\frac{\partial_{x_2 x_2} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{(x_2 - t v_2)^2 - s}{s^2}, \tag{163}$$

$$\frac{\partial_{x_1 t} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{s\left(t^2(-v_1) + t x_1 + \tau v_1\right) - \tau(t v_1 - x_1)\left(t\left(v_1^2 + v_2^2\right) - v_1 x_1 - v_2 x_2\right)}{s^2 \tau}, \tag{164}$$

$$\frac{\partial_{x_2 t} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{s\left(t^2(-v_2) + t x_2 + \tau v_2\right) - \tau(t v_2 - x_2)\left(t\left(v_1^2 + v_2^2\right) - v_1 x_1 - v_2 x_2\right)}{s^2 \tau}, \tag{165}$$

$$\frac{\partial_{x_1 \bar{t}} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{t(x_1 - t v_1)}{s \tau}, \tag{166}$$

$$\frac{\partial_{x_2 \bar{t}} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{t(x_2 - t v_2)}{s \tau}, \tag{167}$$

$$\frac{\partial_{tt} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{\left(-t\left(v_1^2 + v_2^2\right) + v_1 x_1 + v_2 x_2\right)^2}{s^2} - \frac{-2t^2\left(v_1^2 + v_2^2\right) + 2t(v_1 x_1 + v_2 x_2) + \tau\left(v_1^2 + v_2^2\right)}{s \tau} + \frac{t^2 - \tau}{\tau^2}, \tag{168}$$

$$\frac{\partial_{\bar{t}\bar{t}} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{t^2 - \tau}{\tau^2}, \tag{169}$$

$$\frac{\partial_{s} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{-2s + t^2\left(v_1^2 + v_2^2\right) - 2t(v_1 x_1 + v_2 x_2) + x_1^2 + x_2^2}{2s^2}, \tag{170}$$

$$\frac{\partial_{\tau} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{t^2 - \tau}{2\tau^2}, \tag{171}$$

$$\frac{\partial_{v_1} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{t(x_1 - t v_1)}{s}, \tag{172}$$

$$\frac{\partial_{v_2} T(x, t;\ s, \tau, v)}{T(x, t;\ s, \tau, v)} = \frac{t(x_2 - t v_2)}{s}. \tag{173}$$

Fig. 7 Generalized Hermite polynomials as arising from derivatives of the isotropic spatio-temporal Gaussian kernel $T(x, t;\ s, \tau, v)$ according to (79), based on smoothing with a spatially isotropic Gaussian kernel, with respect to the image coordinates $x = (x_1, x_2)^T$ and the time variable t up to order 2, as well as with respect to the spatial scale parameter s, the temporal scale parameter τ and the elements v_1 and v_2 of the velocity vector v

B Generalized Hermite polynomials For The Affine Spatio-Temporal Gaussian Kernel $T(x, t;\ s, \Sigma, \tau, v)$

Figure 8 lists generalized Hermite polynomials for the affine spatio-temporal Gaussian kernel $T(x, t;\ s, \Sigma, \tau, v)$ according to (86).

$$\frac{\partial_{x_1} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{\Sigma_{12}tv_2 - \Sigma_{12}x_2 - \Sigma_{22}tv_1 + \Sigma_{22}x_1}{s\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)}, \tag{174}$$

$$\frac{\partial_{x_2} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{-\Sigma_{11}tv_2 + \Sigma_{11}x_2 + \Sigma_{12}tv_1 - \Sigma_{12}x_1}{\Sigma_{12}^2 s - \Sigma_{11}\Sigma_{22}s}, \tag{175}$$

$$\frac{\partial_t T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{\Sigma_{11}\Sigma_{22}st + \Sigma_{11}\tau v_2(tv_2 - x_2) + \Sigma_{12}^2(-s)t + \Sigma_{12}\tau(-2tv_1v_2 + v_1x_2 + v_2x_1) + \Sigma_{22}\tau v_1(tv_1 - x_1)}{s\tau\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)}, \tag{176}$$

$$\frac{\partial_{\bar{t}} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = -\frac{t}{\tau}, \tag{177}$$

$$\frac{\partial_{x_1x_1} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{\Sigma_{22}^2\left((x_1 - tv_1)^2 - \Sigma_{11}s\right) + \Sigma_{12}^2\left(\Sigma_{22}s + (x_2 - tv_2)^2\right) - 2\Sigma_{12}\Sigma_{22}(tv_1 - x_1)(tv_2 - x_2)}{s^2\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)^2}, \tag{178}$$

$$\frac{\partial_{x_1x_2} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{1}{s^2\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)^2}\left(\Sigma_{11}\Sigma_{12}\Sigma_{22}s - \Sigma_{11}\Sigma_{12}(x_2 - tv_2)^2 + \Sigma_{11}\Sigma_{22}(tv_1 - x_1)(tv_2 - x_2) + \right.$$
$$\left. + \Sigma_{12}^3(-s) + \Sigma_{12}^2(tv_1 - x_1)(tv_2 - x_2) - \Sigma_{12}\Sigma_{22}(x_1 - tv_1)^2\right) \tag{179}$$

$$\frac{\partial_{x_2x_2} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{\Sigma_{11}^2\left((x_2 - tv_2)^2 - \Sigma_{22}s\right) + \Sigma_{11}\Sigma_{12}(\Sigma_{12}s - 2(tv_1 - x_1)(tv_2 - x_2)) + \Sigma_{12}^2(x_1 - tv_1)^2}{s^2\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)^2}, \tag{180}$$

$$\frac{\partial_{x_1\bar{t}} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{t(-\Sigma_{12}tv_2 + \Sigma_{12}x_2 + \Sigma_{22}tv_1 - \Sigma_{22}x_1)}{s\tau\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)}, \tag{181}$$

$$\frac{\partial_{x_2\bar{t}} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{t(\Sigma_{11}tv_2 - \Sigma_{11}x_2 - \Sigma_{12}tv_1 + \Sigma_{12}x_1)}{s\tau\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)}, \tag{182}$$

$$\frac{\partial_{\bar{t}\bar{t}} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{t^2 - \tau}{\tau^2}, \tag{183}$$

$$\frac{\partial_s T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = -\frac{\Sigma_{11}\left((x_2 - tv_2)^2 - 2\Sigma_{22}s\right) + 2\Sigma_{12}^2 s - 2\Sigma_{12}(tv_1 - x_1)(tv_2 - x_2) + \Sigma_{22}(x_1 - tv_1)^2}{2s^2\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)}, \tag{184}$$

$$\frac{\partial_{\Sigma_{11}} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{\Sigma_{22}^2\left((x_1 - tv_1)^2 - \Sigma_{11}s\right) + \Sigma_{12}^2\left(\Sigma_{22}s + (x_2 - tv_2)^2\right) - 2\Sigma_{12}\Sigma_{22}(tv_1 - x_1)(tv_2 - x_2)}{2s\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)^2}, \tag{185}$$

$$\frac{\partial_{\Sigma_{12}} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{1}{s\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)^2}\left(\Sigma_{11}\Sigma_{12}\Sigma_{22}s - \Sigma_{11}\Sigma_{12}(x_2 - tv_2)^2 + \Sigma_{11}\Sigma_{22}(tv_1 - x_1)(tv_2 - x_2) + \right.$$
$$\left. + \Sigma_{12}^3(-s) + \Sigma_{12}^2(tv_1 - x_1)(tv_2 - x_2) - \Sigma_{12}\Sigma_{22}(x_1 - tv_1)^2\right), \tag{186}$$

$$\frac{\partial_{\Sigma_{22}} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{\Sigma_{11}^2\left((x_2 - tv_2)^2 - \Sigma_{22}s\right) + \Sigma_{11}\Sigma_{12}(\Sigma_{12}s - 2(tv_1 - x_1)(tv_2 - x_2)) + \Sigma_{12}^2(x_1 - tv_1)^2}{2s\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)^2}, \tag{187}$$

$$\frac{\partial_{\tau} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{t^2 - \tau}{2\tau^2}, \tag{188}$$

$$\frac{\partial_{v_1} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = \frac{t(-\Sigma_{12}tv_2 + \Sigma_{12}x_2 + \Sigma_{22}tv_1 - \Sigma_{22}x_1)}{s\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)}, \tag{189}$$

$$\frac{\partial_{v_2} T(x,t;\ s,\Sigma,\tau,v)}{T(x,t;\ s,\Sigma,\tau,v)} = -\frac{t(-\Sigma_{11}tv_2 + \Sigma_{11}x_2 + \Sigma_{12}tv_1 - \Sigma_{12}x_1)}{s\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)}. \tag{190}$$

Fig. 8 Generalized Hermite polynomials as arising from derivatives of the affine spatio-temporal Gaussian kernel $T(x, t;\ s, \Sigma, \tau, v)$ according to (86), based on smoothing with a spatially anisotropic affine Gaussian kernel, with respect to the image coordinates $x = (x_1, x_2)^T$ and the time variable t up to order 2, as well as with respect to the spatial scale parameter s, the elements Σ_{11}, Σ_{12} and Σ_{22} of the spatial covariance matrix Σ, the temporal scale parameter τ and the elements v_1 and v_2 of the velocity vector v. (The explicit expressions for $\partial_{x_1t} T(x, t;\ s, \Sigma, \tau, v)$, $\partial_{x_2t} T(x, t;\ s, \Sigma, \tau, v)$ and $\partial_{tt} T(x, t;\ s, \Sigma, \tau, v)$ have been omitted here, because they do not fit within two lines each)

C Generalized Hermite Polynomials for the Time-causal Affine Spatio-Temporal Kernel $T(x, t;\ s, \Sigma, \tau, v, c)$

Figure 9 lists generalized Hermite polynomials for the time-causal affine spatio-temporal Gaussian kernel $T(x, t;\ s, \Sigma, \tau, v, c)$ according to (138).

$$
\begin{aligned}
\frac{\partial_{x_1} T(x,t;\ s,\Sigma,\tau,v,c)}{T(x,t;\ s,\Sigma,\tau,v,c)} &= \frac{\Sigma_{12}tv_2 - \Sigma_{12}x_2 - \Sigma_{22}tv_1 + \Sigma_{22}x_1}{s\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)}, && (191)\\
\frac{\partial_{x_2} T(x,t;\ s,\Sigma,\tau,v,c)}{T(x,t;\ s,\Sigma,\tau,v,c)} &= \frac{-\Sigma_{11}tv_2 + \Sigma_{11}x_2 + \Sigma_{12}tv_1 - \Sigma_{12}x_1}{\Sigma_{12}^2 s - \Sigma_{11}\Sigma_{22}s}, && (192)\\
\frac{\partial_{x_1x_2} T(x,t;\ s,\Sigma,\tau,v,c)}{T(x,t;\ s,\Sigma,\tau,v,c)} &= \frac{1}{s^2\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)^2}\left(\Sigma_{11}\Sigma_{12}\Sigma_{22}s - \Sigma_{11}\Sigma_{12}(x_2 - tv_2)^2 + \Sigma_{11}\Sigma_{22}(tv_1 - x_1)(tv_2 - x_2) + \right.\\
&\quad \left. + \Sigma_{12}^3(-s) + \Sigma_{12}^2(tv_1 - x_1)(tv_2 - x_2) - \Sigma_{12}\Sigma_{22}(x_1 - tv_1)^2\right), && (193)\\
\frac{\partial_{x_2x_2} T(x,t;\ s,\Sigma,\tau,v,c)}{T(x,t;\ s,\Sigma,\tau,v,c)} &= \frac{\Sigma_{11}^2\left((x_2 - tv_2)^2 - \Sigma_{22}s\right) + \Sigma_{11}\Sigma_{12}(\Sigma_{12}s - 2(tv_1 - x_1)(tv_2 - x_2)) + \Sigma_{12}^2(x_1 - tv_1)^2}{s^2\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)^2}, && (194)\\
\frac{\partial_{s} T(x,t;\ s,\Sigma,\tau,v,c)}{T(x,t;\ s,\Sigma,\tau,v,c)} &= -\frac{\Sigma_{11}\left((x_2 - tv_2)^2 - 2\Sigma_{22}s\right) + 2\Sigma_{12}^2 s - 2\Sigma_{12}(tv_1 - x_1)(tv_2 - x_2) + \Sigma_{22}(x_1 - tv_1)^2}{2s^2\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)}, && (195)\\
\frac{\partial_{\Sigma_{11}} T(x,t;\ s,\Sigma,\tau,v,c)}{T(x,t;\ s,\Sigma,\tau,v,c)} &= \frac{\Sigma_{22}^2\left((x_1 - tv_1)^2 - \Sigma_{11}s\right) + \Sigma_{12}^2\left(\Sigma_{22}s + (x_2 - tv_2)^2\right) - 2\Sigma_{12}\Sigma_{22}(tv_1 - x_1)(tv_2 - x_2)}{2s\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)^2}, && (196)\\
\frac{\partial_{\Sigma_{12}} T(x,t;\ s,\Sigma,\tau,v,c)}{T(x,t;\ s,\Sigma,\tau,v,c)} &= \frac{1}{s\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)^2}\left(\Sigma_{11}\Sigma_{12}\Sigma_{22}s - \Sigma_{11}\Sigma_{12}(x_2 - tv_2)^2 + \Sigma_{11}\Sigma_{22}(tv_1 - x_1)(tv_2 - x_2) + \right.\\
&\quad \left. + \Sigma_{12}^3(-s) + \Sigma_{12}^2(tv_1 - x_1)(tv_2 - x_2) - \Sigma_{12}\Sigma_{22}(x_1 - tv_1)^2\right), && (197)\\
\frac{\partial_{\Sigma_{22}} T(x,t;\ s,\Sigma,\tau,v,c)}{T(x,t;\ s,\Sigma,\tau,v,c)} &= \frac{\Sigma_{11}^2\left((x_2 - tv_2)^2 - \Sigma_{22}s\right) + \Sigma_{11}\Sigma_{12}(\Sigma_{12}s - 2(tv_1 - x_1)(tv_2 - x_2)) + \Sigma_{12}^2(x_1 - tv_1)^2}{2s\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)^2}, && (198)\\
\frac{\partial_{v_1} T(x,t;\ s,\Sigma,\tau,v,c)}{T(x,t;\ s,\Sigma,\tau,v,c)} &= \frac{t(-\Sigma_{12}tv_2 + \Sigma_{12}x_2 + \Sigma_{22}tv_1 - \Sigma_{22}x_1)}{s\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)}, && (199)\\
\frac{\partial_{v_2} T(x,t;\ s,\Sigma,\tau,v,c)}{T(x,t;\ s,\Sigma,\tau,v,c)} &= -\frac{t(-\Sigma_{11}tv_2 + \Sigma_{11}x_2 + \Sigma_{12}tv_1 - \Sigma_{12}x_1)}{s\left(\Sigma_{12}^2 - \Sigma_{11}\Sigma_{22}\right)}. && (200)
\end{aligned}
$$

Fig. 9 Generalized Hermite polynomials as arising from derivatives of the time-causal affine spatio-temporal kernel $T(x, t;\ s, \Sigma, \tau, v, c)$ according to (138), based on smoothing with a spatially anisotropic affine Gaussian kernel, with respect to only the image coordinates $x = (x_1, x_2)^T$ up to order 2, as well as with respect to only the spatial scale parameter s, the elements Σ_{11}, Σ_{12} and Σ_{22} of the spatial covariance matrix Σ and the elements v_1 and v_2 of the velocity vector v

Author Contributions TL developed the theory, generated the illustrations and wrote the paper.

Funding Open access funding provided by Royal Institute of Technology.

Data Availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest T.L. is a member of the editorial board of Journal of Mathematical Imaging and Vision.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Abballe, L., Asari, H.: Natural image statistics for mouse vision. PLoS ONE **17**(1), e0262763 (2022)
2. Al-Sharif, S., Khalil, R.: On the generator of two parameter semigroups. Appl. Math. Comput. **156**(2), 403–414 (2004)
3. Ballester, C., Gonzalez, M.: Affine invariant texture segmentation and shape from texture by variational methods. J. Math. Imaging Vision **9**, 141–171 (1998)
4. Barbieri, D., Citti, G., Sanguinetti, G., Sarti, A.: An uncertainty principle underlying the functional architecture of V1. J. Phys. Paris **106**(5–6), 183–193 (2012)
5. Barbieri, D., Citti, G., Cocci, G., Sarti, A.: A cortical-inspired geometry for contour perception and motion integration. J. Math. Imaging Vision **49**(3), 511–529 (2014)
6. Barisin, T., Angulo, J., Schladitz, K., Redenbach, C.: Riesz feature representation: scale equivariant scattering network for classification tasks. SIAM J. Imag. Sci. **17**(2), 1284–1313 (2024)
7. Barisin, T., Schladitz, K., Redenbach, C.: Riesz networks: scale invariant neural networks in a single forward pass. J. Math. Imaging Vision **66**(3), 246–270 (2024)
8. Baspinar, E., Citti, G., Sarti, A.: A geometric model of multi-scale orientation preference maps via Gabor functions. J. Math. Imaging Vision **60**, 900–912 (2018)
9. Baspinar, E., Sarti, A., Citti, G.: A sub-Riemannian model of the visual cortex with frequency and phase. J. Math. Neurosci. **10**(1), 11 (2020)
10. Baumberg, A.: Reliable feature matching across widely separated views. In Proceedings of Computer Vision and Pattern Recognition (CVPR'00), vol. 1, pp. 1774–1781. (2000)

11. Bekkers, E. J.: B-spline CNNs on Lie groups. International Conference on Learning Representations (ICLR 2020). preprint at https://openreview.net/forum?id=H1gBhkBFDH, preprint at arXiv:1909.12057 (2020)
12. Ben-Shahar, O., Zucker, S.: Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. Neural Comput. **16**(3), 445–476 (2004)
13. Bolz, J., Gilbert, C.D.: The role of horizontal connections in generating long receptive fields in the cat visual cortex. Eur. J. Neurosci. **1**(3), 263–268 (1989)
14. Bosking, W.H., Zhang, Y., Schofield, B., Fitzpatrick, D.: Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. J. Neurosci. **17**(6), 2112–2127 (1997)
15. Bronstein, M. M., Bruna, J., Cohen, T., Veličković, P.: Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. Preprint at arXiv preprint arXiv:2104.13478 (2021)
16. Citti, G., Sarti, A.: A cortical based model of perceptual completion in the roto-translation space. J. Math. Imaging Vision **24**(3), 307–326 (2006)
17. Cocci, G., Barbieri, D., Sarti, A.: Spatiotemporal receptive fields of cells in V1 are optimally shaped for stimulus velocity estimation. J. Opt. Soc. Am. A **29**(1), 130–138 (2011)
18. Dai, J., Jin, S., Zhang, J., Nguyen, T.Q.: Boosting feature matching accuracy with pairwise affine estimation. IEEE Trans. Image Process. **29**, 8278–8291 (2020)
19. De, A., Horwitz, G.D.: Spatial receptive field structure of double-opponent cells in macaque V1. J. Neurophysiol. **125**(3), 843–857 (2021)
20. DiCarlo, J.J., Zoccolan, D., Rust, N.C.: How does the brain solve visual object recognition? Neuron **73**(3), 415–434 (2012)
21. Eichhardt, I., Chetverikov, D.: Affine correspondences between central cameras for rapid relative pose estimation. In Proc. European Conference on Computer Vision (ECCV 2018), vol. 11210 of Springer LNCS, pp. 482–497. (2018)
22. Fedorov, V., Arias, P., Sadek, R., Facciolo, G., Ballester, C.: Linear multiscale analysis of similarities between images on Riemannian manifolds: practical formula and affine covariant metrics. SIAM J. Imag. Sci. **8**(3), 2021–2069 (2015)
23. Georgeson, M.A., May, K.A., Freeman, T.C.A., Hesse, G.S.: From filters to features: Scale-space analysis of edge and blur coding in human vision. J. Vis. **7**(13), 7.1 (2007)
24. Gerken, J.E., Aronsson, J., Carlsson, O., Linander, H., Ohlsson, F., Petersson, C., Persson, D.: Geometric deep learning and equivariant neural networks. Artif. Intell. Rev. **56**(12), 14605–14662 (2023)
25. Ghodrati, M., Khaligh-Razavi, S.-M., Lehky, S.R.: Towards building a more complex view of the lateral geniculate nucleus: recent advances in understanding its role. Prog. Neurobiol. **156**, 214–255 (2017)
26. Giannarou, S., Visentini-Scarzanella, M., Yang, G.-G.: Probabilistic tracking of affine-invariant anisotropic regions. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 130–143 (2013)
27. Goldstein, J.A.: Semigroups of Linear Operators and Applications. Oxford Science Publications, Oxford Mathematical Monographs (1985)
28. Hall, B. C.: Lie groups, Lie algebras, and representations, 2nd Edition. Springer, (2015)
29. Hansen, T., Neumann, H.: A recurrent model of contour integration in primary visual cortex. J. Vis. **8**(8), 8.1 (2008)
30. Hesse, G.S., Georgeson, M.A.: Edges and bars: where do people see features in 1-D images? Vision. Res. **45**(4), 507–525 (2005)
31. Hille,E., Phillips,R. S.: Functional Analysis and Semi-Groups, vol. XXXI. American Mathematical Society Colloquium Publications, (1957)
32. Iijima, T.: Basic theory on normalization of pattern (in case of typical one-dimensional pattern). Bulletin of the Electrotechnical Laboratory **26**, 368–388 (1962). (**(in Japanese)**)
33. Jansson,Y., Lindeberg, T.: Exploring the ability of CNNs to generalise to previously unseen scales over wide scale ranges. In Proceedings on International Conference on Pattern Recognition (ICPR 2020), pp. 1181–1188. (2021)
34. Jansson, Y., Lindeberg, T.: Scale-invariant scale-channel networks: deep networks that generalise to previously unseen scales. J. Math. Imaging Vision **64**(5), 506–536 (2022)
35. Jones, J., Palmer, L.: The two-dimensional spatial structure of simple receptive fields in cat striate cortex. J. Neurophysiology **58**, 1187–1211 (1987)
36. Jones, J., Palmer, L.: An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J. Neurophysiology **58**, 1233–1258 (1987)
37. Koenderink, J.J.: The structure of images. Biol. Cybern. **50**(5), 363–370 (1984)
38. Koenderink, J.J., van Doorn, A.J.: Representation of local geometry in the visual system. Biol. Cybern. **55**(6), 367–375 (1987)
39. Koenderink, J.J., van Doorn, A.J.: Generic neighborhood operators. IEEE Trans. Pattern Anal. Mach. Intell. **14**(6), 597–605 (1992)
40. Kristensen, D.G., Sandberg, K.: Population receptive fields of human primary visual cortex organised as DC-balanced bandpass filters. Sci. Rep. **11**(1), 22423 (2021)
41. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using local affine regions. IEEE Trans. Pattern Anal. Mach. Intell. **27**(8), 1265–1278 (2005)
42. Liao, K., Liu, G., Hui, Y.: An improvement to the SIFT descriptor for image representation and matching. Pattern Recogn. Lett. **34**(11), 1211–1220 (2013)
43. Lindeberg, T.: On the axiomatic foundations of linear scale-space. In Sporring, J., Nielsen, M., Florack, L., and Johansen, P. (eds.), Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory, pp. 75–97, Copenhagen, Denmark, Springer (1996)
44. Lindeberg,T.: Linear spatio-temporal scale-space. In Proceedings on International Conference on Scale-Space Theory in Computer Vision (Scale-Space'97), vol. 1252 of Springer LNCS, pp. 113–127. Springer, (1997)
45. Lindeberg, T.: Feature detection with automatic scale selection. Int. J. Comput. Vision **30**(2), 77–116 (1998)
46. Lindeberg, T.: Linear spatio-temporal scale-space. Technical Report ISRN KTH/NA/P–01/22–SE, Dept. of Numerical Analysis and Computer Science, KTH, Nov. (2001). http://www.csc.kth.se/cvap/abstracts/cvap257.html
47. Lindeberg, T.: Generalized Gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space. J. Math. Imaging Vision **40**(1), 36–81 (2011)
48. Lindeberg, T.: A computational theory of visual receptive fields. Biol. Cybern. **107**(6), 589–635 (2013)
49. Lindeberg, T.: Generalized axiomatic scale-space theory. In Hawkes, p. (ed) Advances in Imaging and Electron Physics, vol. 178, pp. 1–96. Elsevier, (2013b)
50. Lindeberg, T.: Time-causal and time-recursive spatio-temporal receptive fields. J. Math. Imaging Vision **55**(1), 50–88 (2016)
51. Lindeberg, T.: Provably scale-covariant continuous hierarchical networks based on scale-normalized differential expressions coupled in cascade. J. Math. Imaging Vision **62**(1), 120–148 (2020)
52. Lindeberg, T.: Normative theory of visual receptive fields. Heliyon **7**(1), e05897 (2021). https://doi.org/10.1016/j.heliyon.2021.e05897
53. Lindeberg, T.: Scale-covariant and scale-invariant Gaussian derivative networks. J. Math. Imaging Vision **64**(3), 223–242 (2022)
54. Lindeberg, T.: A time-causal and time-recursive scale-covariant scale-space representation of temporal signals and past time. Biol. Cybern. **117**(1–2), 21–59 (2023)
55. Lindeberg, T.: Covariance properties under natural image transformations for the generalized Gaussian derivative model for visual

receptive fields. Front. Comput. Neurosci. **17**(1189949), 1–23 (2023)
56. Lindeberg, T.: Discrete approximations of Gaussian smoothing and Gaussian derivatives. J. Math. Imaging Vision **66**(5), 759–800 (2024)
57. Lindeberg, T.: Relationships between the degrees of freedom in the affine Gaussian derivative model for visual receptive fields and 2-D affine image transformations, with application to covariance properties of simple cells in the primary visual cortex. Biol. Cybern. **119**(2), 15 (2025)
58. Lindeberg, T.: Orientation selectivity properties for the affine Gaussian derivative and the affine Gabor models for visual receptive fields. J. Comput. Neurosci. **53**(1), 61–98 (2025)
59. Lindeberg, T.: Unified theory for joint covariance properties under geometric image transformations for spatio-temporal receptive fields according to the generalized Gaussian derivative model for visual receptive fields. J. Math. Imaging Vision **67**(4), 1–49 (2025)
60. Lindeberg, T.: On sources to variabilities of simple cells in the primary visual cortex: A principled theory for the interaction between geometric image transformations and receptive field responses. Preprint at arXiv preprint arXiv:2509.02139, (2025d)
61. Lindeberg, T., Gårding, J.: Shape-adapted smoothing in estimation of 3-D shape cues from affine distortions of local 2-D structure. Image Vis. Comput. **15**(6), 415–434 (1997)
62. Lowe,D. G.: Towards a computational model for object recognition in IT cortex. In Biologically Motivated Computer Vision, vol. 1811 of Springer LNCS, pp. 20–31. Springer, (2000)
63. Marcelja, S.: Mathematical description of the responses of simple cortical cells. J. Optical Soc. of America **70**(11), 1297–1300 (1980)
64. May, K.A., Georgeson, M.A.: Blurred edges look faint, and faint edges look sharp: the effect of a gradient threshold in a multi-scale edge coding model. Vision. Res. **47**(13), 1705–1720 (2007)
65. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. Int. J. Comput. Vision **60**(1), 63–86 (2004)
66. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., van Gool, L.: A comparison of affine region detectors. Int. J. Comput. Vision **65**(1–2), 43–72 (2005)
67. Morel, J.-M., Yu, G.: ASIFT: A new framework for fully affine invariant image comparison. SIAM J. Imag. Sci. **2**(2), 438–469 (2009)
68. Pazy, A.: Semi-groups of Linear Operators and Applications to Partial Differential Equations. Springer-Verlag, Applied Mathematical Sciences (1983)
69. Pei, Z.-J., Gao, G.-X., Hao, B., Qiao, Q.-L., Ai, H.-J.: A cascade model of information processing and encoding for retinal prosthesis. Neural Regen. Res. **11**(4), 646 (2016)
70. Penaud-Polge, V., Velasco-Forero, S., Angulo, J.: Fully trainable Gaussian derivative convolutional layer. In International Conference on Image Processing (ICIP 2022), pp. 2421–2425. (2022)
71. Perzanowski, A., Lindeberg, T.: Scale generalisation properties of extended scale-covariant and scale-invariant Gaussian derivative networks on image datasets with spatial scaling variations. J. Math. Imaging Vision **67**(3), 1–39 (2025)
72. Ringach, D.L.: Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. J. Neurophysiol. **88**, 455–463 (2002)
73. Ringach, D.L.: Mapping receptive fields in primary visual cortex. J. Physiol. **558**(3), 717–728 (2004)
74. Rodríguez, M., Delon, J., Morel, J.-M.: Covering the space of tilts: application to affine invariant image comparison. SIAM J. Imag. Sci. **11**(2), 1230–1267 (2018)
75. Rothganger, F., Lazebnik, S., Schmid, C., Ponce, J.: 3D object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints. Int. J. Comput. Vision **66**(3), 231–259 (2006)
76. Rothganger, F., Lazebnik, S., Schmid, C., Ponce, J.: Segmenting, modeling, and matching video clips containing multiple moving objects. IEEE Trans. Pattern Anal. Mach. Intell. **29**(3), 477–491 (2007)
77. Ruslim, M. A., Burkitt, A. N., Lian,Y.: Learning spatio-temporal V1 cells from diverse LGN inputs. bioRxiv, 2023.11.30.569354 (2023)
78. Sabatini, S.P., Gastaldi, G., Solari, F., van Hulle, K.P.M., Diaz, J., Ros, E., Pugeault, N., Krüger, N.: A compact harmonic code for early vision based on anisotropic frequency channels. Comput. Vis. Image Underst. **114**(6), 681–699 (2010)
79. Sadek, R., Constantinopoulos, C., Meinhardt, E., Ballester, C.C., Caselles, V.: On affine invariant descriptors related to SIFT. SIAM J. Imag. Sci. **5**(2), 652–687 (2012)
80. Sangalli, M., Blusseau, S., Velasco-Forero, S., Angulo, J.: Scale equivariant U-net. In Proceedings British Machine Vision Conference (BMVC 2022), pp. 763. (2022)
81. Sarti, A., Citti, G., Petitot, J.: The symplectic structure of the primary visual cortex. Biol. Cybern. **98**(1), 33–48 (2008)
82. Sarti, A., Citti, G., Petitot, J.: Functional geometry of the horizontal connectivity in the primary visual cortex. J. Phys. Paris **103**(1–2), 37–45 (2009)
83. Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T.: Robust object recognition with cortex-like mechanisms. IEEE Trans. Pattern Anal. Mach. Intell. **29**(3), 411–426 (2007)
84. Sosnovik, I., Szmaja, M., Smeulders, A.: Scale-equivariant steerable networks. International Conference on Learning Representations (ICLR 2020). preprint at arXiv:1910.11093. (2020)
85. Sosnovik, I., Moskalev, A., Smeulders, A.: DISCO: Accurate discrete scale convolutions. British Machine Vision Conference (BMVC 2021). preprint at arXiv:2106.02733. (2021)
86. Tuytelaars, T., van Gool, L.: Matching widely separated views based on affine invariant regions. Int. J. Comput. Vision **59**(1), 61–85 (2004)
87. Wallis, S.A., Georgeson, M.A.: Mach edges: local features predicted by 3rd derivative spatial filtering. Vision. Res. **49**(14), 1886–1893 (2009)
88. Wang, Q., Spratling, M.W.: Contour detection in colour images using a neurophysiologically inspired model. Cogn. Comput. **8**(6), 1027–1035 (2016)
89. Weickert, J., Ishikawa, S., Imiya, A.: Linear scale-space has first been proposed in Japan. J. Math. Imaging Vision **10**(3), 237–252 (1999)
90. Wendt, G., Faul, F.: Binocular luster elicited by isoluminant chromatic stimuli relies on mechanisms similar to those in the achromatic case. J. Vis. **24**(3), 7–7 (2024)
91. Wimmer, T., Golkov, V., Dang, H. N., Zaiss, M., Maier, A., Cremers, D.: Scale-equivariant deep learning for 3D data. arXiv preprint arXiv:2304.05864, (2023)
92. Worrall, D., Welling,M.: Deep scale-spaces: Equivariance over scale. In Advances in Neural Information Processing Systems (NeurIPS 2019), pp. 7366–7378, (2019)
93. Yang, Y., Dasmahapatra, S., Mahmoodi, S.: Scale-equivariant UNet for histopathology image segmentation. Proc. Mach. Learn. Res. **194**, 130–148 (2022)
94. Young, R.A.: The Gaussian derivative model for spatial vision: I. Retinal mechanisms. Spatial Vision **2**(4), 273–293 (1987)
95. Young, R.A., Lesperance, R.M.: The Gaussian derivative model for spatio-temporal vision: II cortical data. Spat. Vis. **14**, 321–389 (2001)
96. Young, R.A., Lesperance, R.M., Meyer, W.W.: The Gaussian derivative model for spatio-temporal vision: I cortical model. Spat. Vis. **14**, 261–319 (2001)
97. Yu,G., Morel,J.-M.: A fully affine invariant image comparison method. In Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009), pp. 1597–1600, (2009)

98. Zhan, W., Sun, G., Li,Y.: Scale-equivariant steerable networks for crowd counting. In Proc. International Conference on Control and Robotics Engineering (ICCRE 2022), pp. 174–179, (2022)
99. Zhu, W., Qiu, Q., Calderbank, R., Sapiro, G., Cheng, X.: Scale-translation-equivariant neural networks with decomposed convolutional filters. J. Mach. Learn. Res. **23**(68), 1–45 (2022)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Tony Lindeberg is a Professor of Computer Science — Computer Vision at KTH Royal Institute of Technology in Stockholm, Sweden. He was born in 1964, received his MSc degree in 1987, his PhD degree in 1991, became docent in 1996, and was appointed professor in 2000. He was a Research Fellow at the Royal Swedish Academy of Sciences between 2000 and 2010. His research interests in computational vision relate to scale-space theory, image features, object recognition, spatio-temporal recognition, video analysis, deep networks and computational modelling of biological vision. He has developed theories and methodologies for continuous and discrete scale-space representation, visual and auditory receptive fields, hierarchical and deep networks, detection of image features and salient image structures, automatic scale selection, scale-covariant and scale-invariant features, affine-covariant and affine-invariant features, affine and Galilean normalization, temporal, spatio-temporal and spectro-temporal scale-space concepts as well as spatial and spatio-temporal image descriptors for image-based recognition. He does also work on computational modelling of hearing and has previously worked on topics in medical image analysis, brain activation and gesture recognition. He is the author of the book Scale-Space Theory in Computer Vision.