Research article

Normative theory of visual receptive fields

Tony Lindeberg *

Computational Brain Science Lab, Division of Computational Science and Technology, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden

ARTICLE INFO

ABSTRACT

This article gives an overview of a normative theory of visual receptive fields. We describe how idealized functional models of early spatial, spatio-chromatic and spatio-temporal receptive fields can be derived in a principled way, based on a set of axioms that reflect structural properties of the environment in combination with assumptions about the internal structure of a vision system to guarantee consistent handling of image representations over multiple spatial and temporal scales. Interestingly, this theory leads to predictions about visual receptive field shapes with qualitatively very good similarities to biological receptive fields measured in the retina, the LGN and the primary visual cortex (V1) of mammals.

1. Introduction

The light distribution that reaches a visual sensor, such as the retina, carries information about the environment to a visual observer. The information necessary to infer properties about the surrounding world from this light distribution is, however, not contained in the measurement of image intensity at any single image point in isolation. Instead, the relevant information is mediated by the *relationships* between image intensities over local neighbourhoods. An underlying reason for this is that the incoming light constitutes an *indirect* source of information that depends on the interaction between geometric and material properties of objects in the surrounding world and on external illumination sources. Another main reason why cues to the environment need to be collected over *regions* in image space as opposed to at single image points is that the measurement process by itself requires the accumulation of energy over non-infinitesimal support regions over space and time. Such a region in the visual field, for which a neuron responds to visual stimuli, is traditionally referred to as a *receptive field* (Hubel and Wiesel [1, 2, 3]) (see Fig. 1).

In this work, we focus on a functional description of receptive fields, regarding how a neuron with a purely spatial receptive field responds

to visual stimuli over image space, and regarding how a neuron with a spatio-temporal receptive field responds to visual stimuli over space and time (DeAngelis et al. [4, 5]).

If we consider the theoretical and computational problem of designing a vision system that is going to derive properties of the surrounding world from light reflected from it, we may ask what types of image operations should be performed on the image data. Would it be possible to perform *any* type of image operation, or are there classes of image operations that are more natural or more effective? Specifically, with regard to the notion of receptive fields, we may ask what shapes of receptive field profiles would be reasonable or desirable. Is it possible to express a theory for how receptive fields "ought to" respond to visual data?

From a first inspection, such a problem could possibly be regarded as intractable, unless the prerequisites of the question could be further specified. It does, however, turn out to be possible to address this question in a systematic manner, based on a framework known as *scale-space theory* (Iijima [6]; Witkin [7]; Koenderink [8]; Koenderink and van Doorn [9, 10]; Lindeberg [11, 12, 13, 14]; Florack [15]; Sporring et al. [16]; Weickert et al. [17]; ter Haar Romeny [18]), which has been developed in the area of computer vision. This field has established a

---
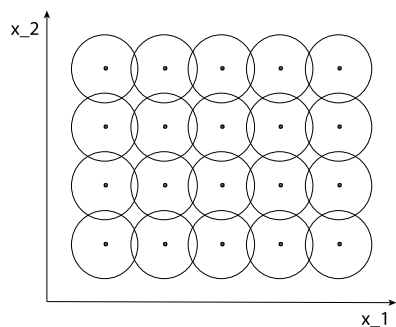
Fig. 1. A traditional definition of the notion of a receptive field is as a region in the visual field for which a visual sensor/neuron/operator responds to visual stimuli. In this figure, we have illustrated a set of receptive fields over the spatial domain that partially overlap, and where all the receptive fields have the same size. More generally, we could consider distributions of receptive fields over space or space-time that have varying sizes, shapes and orientations in image space as well as having different directions in joint space-time. Adjacent receptive fields could also have substantially larger relative overlap than displayed here. In this work, we focus on a functional description of such linear receptive fields, concerning how a neuron responds to visual stimuli over image space regarding spatial receptive fields or over joint space-time regarding spatio-temporal receptive fields.

Fig. 2. Basic factors that influence the formation of images for an eye with a two-dimensional retina that observes objects in the three-dimensional world. In addition to the position, the orientation and the motion of the object in 3-D, the perspective projection onto the retina is affected by the viewing distance, the viewing direction and the relative motion of the eye in relation to the object, the spatial and the temporal sampling characteristics of the neurons in the retina as well the usually unknown external illumination field in relation to the geometry of the scene and the observer.
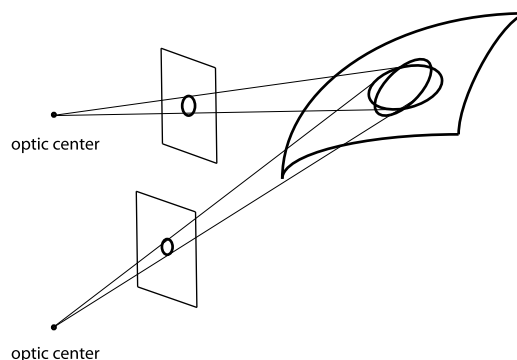
Fig. 3. Illustration of the importance of covariance of the receptive field responses under natural image transformations. Consider a vision system that computes image features from image data based on image operations that are formulated over rotationally symmetric support regions in the spatial image domain. If such image measurements are performed for two different viewing directions relative to the same three-dimensional surface patch, then the back-projections of the image operations onto the tangent plane surface of the object will, in general, correspond to different regions in physical space over which corresponding information will be weighted differently. If such image features are in turn to be used for deriving three-dimensional shape cues of the object from binocular cues, such as surface orientation, then there will be a systematic error caused by the mismatch between the backprojections of the receptive fields from the image domain onto the world. By requiring the family of receptive fields to be covariant under local affine image deformations, it is possible to reduce this amount of mismatch, such that the backprojected receptive fields can be made *equal*, when projected onto the tangent plane of the surface by local linearizations of the perspective mapping. In this way, the source to error caused by mismatch between the two different receptive fields is eliminated. Corresponding effects occur when analyzing spatio-temporal image data based on receptive fields that are restricted to being space-time separable only. If an object is observed over time from two observations having different relative motions between the viewing direction and the observer, then the corresponding receptive fields cannot be matched unless the family of receptive fields possesses sufficient covariance properties under local Galilean transformations.

paradigm of imposing a set of *structural constraints* on the first stages of visual processing that reflect *symmetry properties* of the environment. From an axiomatic treatment based on such assumptions, it turns out to be possible to restrict the class of permissible image operations substantially.
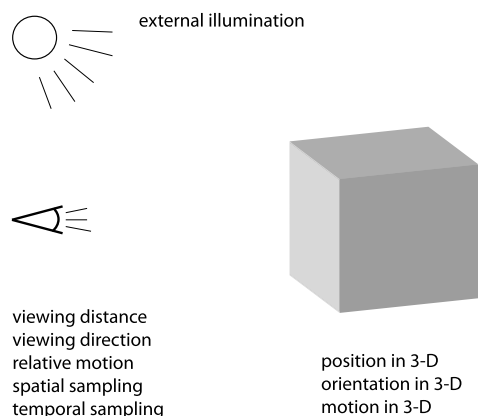
The subject of this article is to present a comprehensive overview of a theory for how structural requirements on the first stages of visual processing as formulated based on scale-space theory can be used for deriving idealized functional models of visual receptive fields, and to develop implications of how these theoretical results can be used when modelling receptive fields in the retina, the lateral geniculate nucleus and the primary visual cortex. A main message is that we derive idealized functional models for linear receptive fields *by necessity*, starting from a small set of symmetry requirements that reflect properties of the visual world that one may naturally require an idealized vision system, or a biological organism subject to strong evolutionary pressure, to be adapted to, to enable a consistent handling of receptive field responses in terms of provable covariance or invariance properties under natural image transformations (see Fig. 2).

If the receptive field responses do not obey covariance under the basic classes of geometric image transformations, then it will not be possible to match the early image representations between *e.g.* different views of the same scene, which implies that the vision system will perform systematic errors *e.g.* when deriving shape cues from a three-dimensional scene, as illustrated in Fig. 3.

In this respect, the treatment has similarities to the way theories are formulated in theoretical physics, where symmetry properties of the environments constitute key components underlying the formulation of physical theories of the world. The treatment that will follow will be general in the sense that it *encompasses spatial, spatio-chromatic and spatio-temporal receptive fields within the same unified theory*.

Early mathematical necessity results underlying this theory were presented in [13], and earlier versions of this theory in a format for a computational biology audience have been presented in [19, 20]. More recently, a very much improved model for the case of a time-causal spatio-temporal domain was presented in [21], where the future cannot be accessed and the receptive fields have to be solely based on information from the present moment and a compact buffer of the past, written for an audience in the area of mathematical imaging.

That material may, however, be less easy to access for vision researchers in biology, medicine or psychophysics. Specifically, the replacement of certain assumptions (axioms) in [19, 20] with new as-

sumptions (axioms) in [21] may require substantial efforts for readers not previously familiar with this type of theoretical modelling. This has motivated the need for an overview article of the improved theory, intended to be more easy to access, and with a more direct focus on biological implications. Thus, this paper presents an improved ax-

iomatic structure on a compact form more easy to access compared to the original publications and updated with respect to the better time-causal model.

It will be shown that the presented framework leads to predictions of *receptive field profiles* in good agreement with receptive field measurements reported in the literature (Hubel and Wiesel [1, 2, 3]; DeAngelis et al. [4, 5]; Conway and Livingstone [22]; Johnson et al. [23]). Specifically, explicit phenomenological models will be given of neurons in the LGN and simple cells in the primary visual cortex, with comparisons to related models in terms of Gabor functions (Marčelja [24]; Jones and Palmer [25, 26]; Ringach [27, 28]), differences of Gaussians (Rodieck [29]) and Gaussian derivatives (Koenderink and van Doorn [9]; Young [30]; Young et al. [31, 32]). An important consequence of the theory is that the evolution properties of the receptive field profiles can be described by diffusion equations. They are therefore suitable for implementation in a biological architecture, since the computations can be expressed in terms of communications between neighbouring computational units, where either a single computational unit or a group of computational units may be interpreted as corresponding to a neuron or a group of neurons.[1] Such computational models based on diffusion equations do also arise in mean field theories that approximate the computations that are performed by populations of neurons (Omurtag et al. [33]; Mattia and Guidic [34]; Faugeras et al. [35]).

1.1. Structure of this article

This paper is organized as follows: Section 2 gives an overview of and motivations to the assumptions that the theory is based on. A set of structural requirements is formulated to capture the effect of natural image transformations onto the illumination field that reaches the retina and to guarantee internal consistency between image representations that are computed from receptive field responses over multiple spatial and temporal scales.

This set of structural requirements partially overlaps with the structural requirements in [19, 20], while the axiomatic structure has been substantially changed regarding a time-causal temporal domain according to the more recent theory in [21]. This is the most practically relevant case for realistic modelling of biological vision, since there is no way to access the future in a real-time situation, but which is not at all handled in the earlier spatio-temporal modeling work by e.g. Young et al. [31, 32].

Section 3 describes linear receptive families that arise as consequences of these assumptions for the cases of either a purely spatial domain or a joint spatio-temporal domain. The issue of how to perform relative normalization between receptive field responses over multiple spatial and temporal scales is treated, so as to enable comparisons between receptive field responses at different spatial and temporal scales. We also show how the influence of illumination transformations and exposure control mechanisms on the receptive field responses can be handled, by describing invariance properties obtained by applying the derived linear receptive fields over a logarithmically transformed intensity domain.

The consequences of these assumptions for spatial and spatio-temporal domains, which are described in Sections 3.1 and 3.2, constitute more explicit reformulations of results in [19, 20]. These reformulations have additionally been made so that they encompass a time-causal spatio-temporal domain based on results in [21]. The material in Section 3.3 is a new statement of normalization results, partly based a theory for scale selection in [36], while extended from spatially isotropic Gaussian kernels to affine Gaussian kernels, and partially also

based on results in [21]. Section 3.4 describes an adaptation of theoretical results in [19] and [37] to this specific domain.

Section 4 shows examples of how spatial, spatio-chromatic and spatio-temporal receptive fields in the retina, the LGN and the primary visual cortex can be well modelled by the derived receptive field families.

Several of these figures are similar to figures in [19, 20] and in [21]. These results have, however, also been cleaned by replacing the previous time-causal models in [19, 20] with the much better time-causal theory in [21], and updating the distribution parameter c from the previous default value $c = \sqrt{2}$ to the value $c = 2$ found more suitable for computer vision algorithms that operate on these time-causal spatio-temporal receptive fields with real-time requirements of shorter temporal delays [38].

Section 5 gives relations to previous work, including conceptual and theoretical comparisons to previous use of Gabor models of receptive fields, approaches for learning receptive fields from image data and previous applications of a logarithmic transformation of the image intensities. Finally, Section 6 summarizes some of the main results.

2. Assumptions underlying the theory: structural requirements

In the following, we shall describe a set of structural requirements that can be stated concerning: (i) spatial geometry, (ii) spatio-temporal geometry, (iii) the image measurement process with its close relationship to the notion of scale, (iv) internal representations of image data that are to be computed by a general purpose vision system and (v) the parameterization of image intensity with regard to the influence of illumination variations.

For modelling the image formation process, we will at any point in the retina approximate the spherical retina by a perspective projection onto the tangent plane of the retinal surface at that image point, below represented as the image plane. Additionally, we will approximate the possibly non-linear geometric transformations regarding spatial and spatio-temporal geometry by local linearizations at every image point, and corresponding to the derivatives of the possibly non-linear transformations. In these ways, the theoretical analysis can be substantially simplified, while still enabling accurate modelling of essential functional properties of receptive fields in relation to the effects of natural image transformations as arising from interactions with the environment.

By necessity, some parts of the presentation in this section will be somewhat technical, if we want to clearly mathematically define the assumptions that the theory rests upon. For the hasty reader, who may be more interested in the implications of the theory, we have made a schematic summary of most of the main assumptions in Fig. 4. After getting an overview of these assumptions from this figure, the hasty reader may then proceed to Section 3, where it is shown how these assumptions lead to idealized families of visual receptive fields, then backtracking to this section again if necessary.

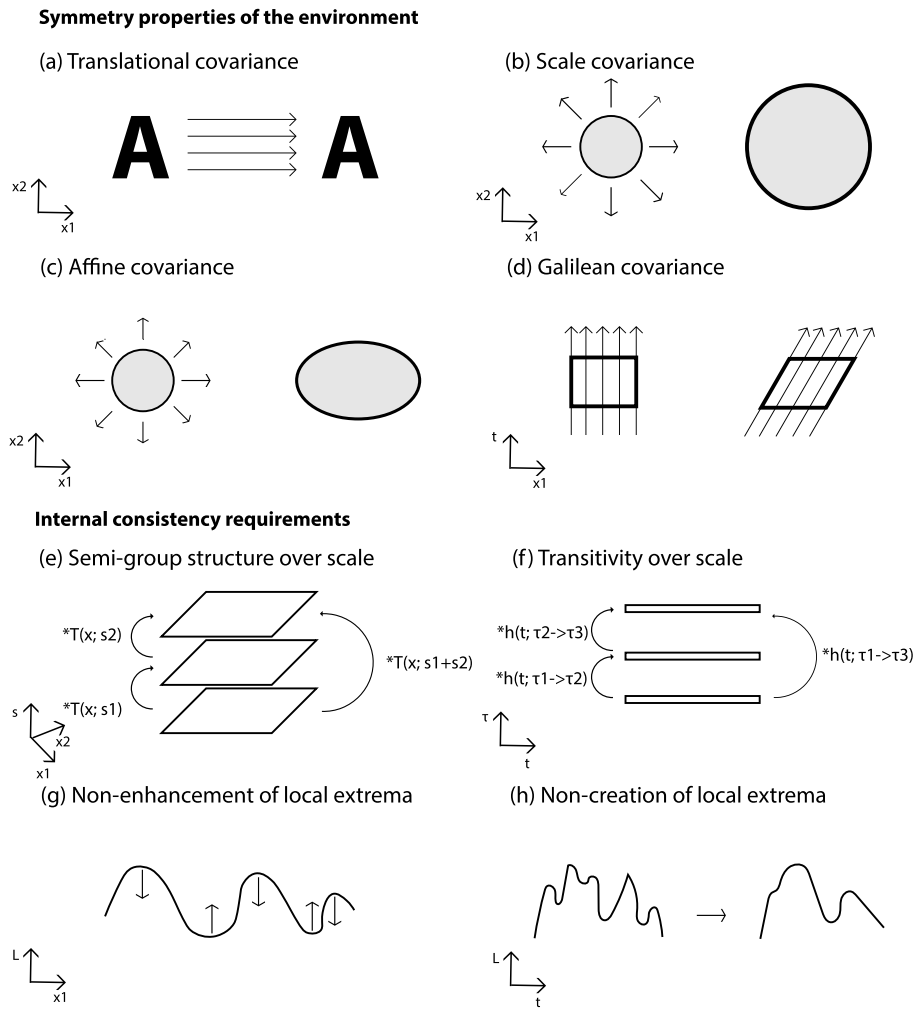### 2.1. Static image data over a spatial domain

In the following, we will describe a theoretical model for the computational function of applying visual receptive fields to local image patterns.

For time-independent data f over a two-dimensional spatial image domain, we would like to define a family of image representations $L(\cdot; s)$ over a possibly multi-dimensional scale parameter s, where the internal image representations $L(\cdot; s)$ are computed by applying some parameterized family of image operators $\mathcal{T}_s$ to the image data f:

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

Specifically, we will assume that the family of image operators $\mathcal{T}_s$ should satisfy:

---

[1] Specifically, both the spatial and temporal smoothing and the computation of spatial and temporal derivatives, that constitute the computational primitives in the receptive field models that we will arrive at, can be performed by local connections between neighbouring neurons in our time-causal model.

Symmetry properties of the environment

(a) Translational covariance

(b) Scale covariance

(c) Affine covariance

(d) Galilean covariance

Internal consistency requirements

(e) Semi-group structure over scale

(f) Transitivity over scale

*T(x; s2)

*T(x; s1+s2)

s

*T(x; s1)

x2

x1

*h(t; τ2->τ3)

*h(t; τ1->τ2)

τ

t

*h(t; τ1->τ3)

(g) Non-enhancement of local extrema

(h) Non-creation of local extrema

L

x1

L

t

Fig. 4. Schematic illustration of main assumptions underlying the proposed normative theory for visual receptive fields, regarding (i) transformation properties of the environment and (ii) internal consistency requirements to guarantee internally consistent image representations over multiple spatial and temporal scales. (a) *Translational covariance* means that visual representations of objects should be processed in a similar manner over the entire visual field. (b) *Scale covariance* means that scaling transformations, as occur in the visual domain because of objects of different size and objects at different distances to the observer, should be processed in a similar manner such that the receptive field responses can be matched. (c) *Affine covariance* is a generalization of scale covariance to non-uniform scaling transformations, as occur when surface structures are foreshortened for surfaces with a non-frontal slant angle relative to the tangent plane of the surface. (d) *Galilean covariance* means that if we observe objects or events that move relative to a fixed viewing direction, then these visual patterns should be processed in a conceptually similar way as if we observe the same patterns with the gaze direction following the same objects or events, and in such a way that the two types of spatio-temporal image representations can be matched. (e) The assumption of a *semi-group structure over spatial scales* implies that with a spatial smoothing operation in terms of convolution operations, which follows from a combination of the assumptions of translational covariance and linearity, the composition of two spatial smoothing operations with scale parameters s_1 and s_2 should be a spatial smoothing operation of a similar form and with added values of scale parameters $s_1 + s_2$. (f) The assumption of a *transitivity structure over temporal scales* implies that the composition of two temporal smoothing operations from temporal scales τ_1 to τ_2 and from temporal scales τ_2 to τ_3 should be a similar type of temporal smoothing operation from temporal scales τ_1 to τ_3 (while without imposing an additive structure of the temporal scale parameters). (g) The assumption of *non-enhancement of local extrema* means that the spatial smoothing operation that determines the shape of the spatial receptive fields should obey the property that the smoothed intensity value L at a spatial maximum must not increase with increasing scale and that the intensity value at a spatial minimum must not decrease with increasing scale. (h) The assumption of *non-creation of local extrema* implies that the temporal smoothing operation that determines the temporal shape of the spatio-temporal receptive fields must not increase the number of local extrema in a purely temporal signal.

a) Linearity. For the earliest processing stages to have as low risk as possible of making irreversible decisions that may affect later processing stages, we assume that the first layers of receptive fields should be linear

$$\mathcal{T}_s(a_1 f_1 + a_2 f_2) = a_1 \mathcal{T}_s f_1 + b_1 \mathcal{T}_s f_2. \qquad (2)$$

Specifically, linearity implies that any particular scale-space properties (to be detailed below) that we derive for the zero-order image representation L will transfer to any spatial derivative $L_{x_1^{\alpha_1} x_2^{\alpha_2}}$ of L, so that

$$L_{x_1^{\alpha_1} x_2^{\alpha_2}}(\cdot\,;\, s) = \partial_{x_1^{\alpha_1} x_2^{\alpha_2}}(\mathcal{T}_s f(\cdot)) = \mathcal{T}_s(\partial_{x_1^{\alpha_1} x_2^{\alpha_2}} f(\cdot)) \qquad (3)$$

where α_1 and α_2 are the derivative orders for the two spatial dimensions x_1 and x_2.

This means that different types of image structures, irrespective of what order of spatial differentiation they respond mostly to, will be treated in a structurally similar manner. In this way, we reduce the risk that the first layers of visual receptive fields could make early decisions dedicated to certain types of image structures that later processing stages could then not later recover from.

In this sense, the assumption of linearity reflects the requirement of a lack of bias to particular types of image structures, with the underlying aim that the processing performed in the first processing stages should be *generic*, to be used as input for a large variety of visual tasks. By

the assumption of linearity, local image structures that are captured by *e.g.* first- or second-order derivatives will be treated in a structurally similar manner, which would not necessarily be the case if the first local neighbourhood processing stage of the first layer of receptive fields would instead be allowed to be genuinely non-linear.[2]

This genericity property is closely related to the basic property of the mammalian vision system, that the computations performed in the retina[3], the LGN[4] and the primary visual cortex provide general purpose output that is used as input to higher-level visual areas.

b) Translational covariance. To ensure that the visual interpretation of an object should be the same irrespective of its position in the image plane, we assume that the first processing stages should be covariant under translations, so that if an object is moved a distance $\Delta x = (\Delta x_1, \Delta x_2)$ in the image plane, the receptive field response should remain on a similar form, while shifted with the same distance. Formally, this requirement can be stated that the family of image operators $\mathcal{T}_s$ should commute with the shift operator defined by $S_{\Delta x}(f)(x) = f(x - \Delta x)$:

$$\mathcal{T}_s \left(S_{\Delta x} f \right) = S_{\Delta x} \left(\mathcal{T}_s f \right). \tag{4}$$

In other words, if we shift the input by a translation and then apply the receptive field operator $\mathcal{T}_s$, the result should be similar as applying the receptive field operator to the original input and then shifting the result.

c) Convolution structure. Together, the assumptions about linearity and translational covariance imply that $\mathcal{T}_s$ will correspond to a convolution operator [54]. This implies that the representation L can be computed from the image data f by convolution with some parameterized family of convolution kernels $T(\cdot; s)$:

---

[2] While it is known that many receptive fields may have both linear and non-linear components, we do here, as a conceptual simplification, focus on the regime where receptive fields in the LGN and the primary visual cortex can be well approximated by a linear model. Note, however, that this assumption about linearity of the first layers of visual receptive fields does not exclude the possibility of defining later stage non-linear receptive fields that operate on the output from the linear receptive fields, such as the computations performed by complex cells in the primary visual cortex [1, 2, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51]. Specifically, in [51, Section 5] a functional energy model of complex cells is proposed, based on simple cell responses modelled by affine Gaussian derivatives as described in this work. Neither does the assumption of linearity exclude the possibility of transforming the raw image intensities by a pointwise non-linear mapping function prior to the application of linear receptive fields over local neighbourhoods. In Section 3.4, it will be specifically shown that a pointwise logarithmic transformation of the image intensities prior to the application of linear receptive fields has theoretical advantages of enabling invariance properties of derivative-based receptive field responses under local multiplicative illumination transformations.

[3] Concerning the use of a linear model for the receptive fields in the retina used in this paper, it should be noted that there are also more sophisticated non-linear computations known to be performed in the retina [52]. Light adaptation and gain control in the retina, we model by an initial logarithmic transformation prior to the receptive fields, which has the desirable property of leading to invariant receptive field responses under multiplicative illumination transformations, as will be described in Section 3.4. Direction sensitive cells that respond to different directions of motion, to handle objects that move rapidly over the visual field and to decrease temporal delays in a time-causal context by the ability of performing motion extrapolation to perform temporal predictions, we capture some of the effects of by studying the influence of Galilean transformations, Probably, the function of such non-linear mechanisms could also be constrained from structural properties of the world, in a similar way as we here constrain the shapes of visual receptive fields from theoretical assumptions. The use of a pure linear model in this treatment, should in this context be seen as a conceptual simplification and a way to simplify the theoretical analysis.

[4] Concerning the use of a pure feed-forward model of the receptive fields used in this paper, it should be noted that there is a also a large number of top-down connections to the LGN from higher visual areas [53], which we do not explicitly model in this treatment.

$$L(\cdot; s) = T(\cdot; s) * f(\cdot). \tag{5}$$

d) Semi-group structure over spatial scales. To ensure that the transformation from any finer scale s_1 to any coarser scale $s_1 + s_2$ should be of the same form for any $s_2 > 0$ (a requirement of algebraic closedness), we assume that the result of convolving two kernels $T(\cdot; s_1)$ and $T(\cdot; s_2)$ from the family with each other should be a kernel within the same family of kernels and with added parameter values $T(\cdot; s_1 + s_2)$:

$$T(\cdot; s_1) * T(\cdot; s_2) = T(\cdot; s_1 + s_2). \tag{6}$$

This assumption specifically implies that the representation $L(\cdot; s_2)$ at a coarse scale s_2 can be computed from the representation $L(\cdot; s_1)$ at a finer scale $s_1 < s_2$ by a convolution operation of the same form (5) as the transformation from the original image data f, while using the difference in scale levels $s_2 - s_1$ as the parameter

$$L(\cdot; s_2) = T(\cdot; s_2 - s_1) * L(\cdot; s_1). \tag{7}$$

This property does in turn imply that if we are able to derive specific properties of the family of transformations $\mathcal{T}_s$ (to be detailed below), then these properties will not only hold for the transformation from the original image data f to the representations $L(\cdot; s)$ at coarser scales, but also between any pair of scale levels $s_2 > s_1$, with the aim that image representations at coarser scales should be regarded as simplifications of corresponding image representations at finer scales.

In terms of mathematical concepts, this form of algebraic structure is referred to as a semi-group structure over spatial scales

$$\mathcal{T}_{s_1} \mathcal{T}_{s_2} = \mathcal{T}_{s_1 + s_2}. \tag{8}$$

e) Scale covariance under spatial scaling transformations. If a visual observer looks at the same object from different distances, we would like the internal scale-space representations derived from the receptive field responses to be sufficiently similar, so that the object can be recognized as the same object, while appearing with a different size on the retina. Specifically, it is thereby natural to require that the receptive field responses should be of a similar form, while resized in the image plane.

This corresponds to a requirement of spatial scale covariance under uniform scaling transformations of the spatial domain $x' = S_s x$:

$$L'(x'; s') = L(x; s) \quad \Leftrightarrow \quad \mathcal{T}_{S_s(s)} S_s f = S_s \mathcal{T}_s f \tag{9}$$

to hold for some transformation $s' = S_s(s)$ of the scale parameter s.

f) Affine covariance under spatial affine transformations. If a visual observer looks at the same local surface patch from two different viewing directions, then the local surface patch may be deformed in different ways onto the different views and with different amounts of perspective foreshortening from the different viewing directions. If we approximate the local deformations caused by the perspective mapping by local affine transformations, then the transformation between the two differently deformed views of the local surface patch can in turn be described by a composed local affine transformation $x' = A x$.

If we are to use receptive field responses as a basis for higher level visual operations, it is natural to require that the receptive field response of an affine deformed image patch should remain on a similar form, while being reshaped by a corresponding affine transformation.

This corresponds to a requirement of affine covariance under general affine transformations $x' = A x$:

$$L'(x'; s') = L(x; s) \quad \Leftrightarrow \quad \mathcal{T}_{A(s)} \mathcal{A} f = \mathcal{A} \mathcal{T}_s f \tag{10}$$
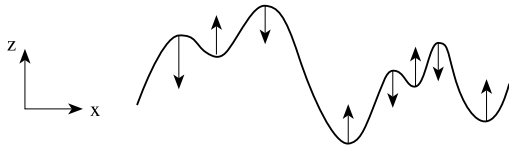
to hold for some transformation $s' = A(s)$ of the scale parameter.

g) Non-creation of new structure with increasing scale. If we apply the family of transformations $\mathcal{T}_s$ for computing representations at coarser scales from representations at finer scales according to (1) and (7), there could be a potential risk that the family of transformations could

Fig. 5. Illustration of the notion of non-enhancement of local extrema, which is a way to restrict the class of possible image operations by preventing new structures from being created from finer to coarser levels of scales. Non-enhancement of local extrema means that the value at a local maximum must not increase and that the value at a local minimum must not decrease with increasing scale s.

amplify spurious structures in the input to produce macroscopic amplifications in the representations at coarser scales that do not directly correspond to simplifications of corresponding structures in the original image data. To prevent such undesirable phenomena from occurring, we require that local spurious structures must not be amplified and express this condition in terms of the evolution properties over scales at local maxima and minima in the image intensities as smoothed by the family of convolution kernels $T(\cdot; s)$: If a point x_0 for some scale s_0 is a local maximum point in the image plane, then the value at this maximum point $L(x_0; s_0)$ must not increase to coarser scales $s > s_0$. Similarly, if a point is a local minimum point in the image plane, then the value at this minimum point $L(x_0; s_0)$ must not decrease to coarser scales $s > s_0$.

Formally, this requirement that new structures should not be created from finer to coarser scales can be formalized into the requirement of *non-enhancement of local extrema*, which implies that if at some scale s_0 a point x_0 is a local maximum (minimum) for the mapping from x to $L(x; s_0)$, then (see Fig. 5):

- $(\partial_s L)(x; s) \leq 0$ at any spatial maximum,
- $(\partial_s L)(x; s) \geq 0$ at any spatial minimum.

This condition implies a strong condition on the class of possible smoothing kernels $T(\cdot; s)$.

2.2. Time-dependent image data over joint space-time

To model the computational function of spatio-temporal receptive fields in time-dependent image patterns, we do for a time-dependent spatio-temporal domain first inherit the structural requirements regarding a spatial domain and complement the spatial scale parameter s by a temporal scale parameter τ. In addition, we assume:

a) Scale covariance under temporal scaling transformations. If a similar type of spatio-temporal event $f(x,t)$ occurs at different speeds, faster or slower, it is natural to require that the receptive field responses should be of a similar form, while occurring correspondingly faster or slower.

This corresponds to a requirement of temporal scale covariance under a requirement of temporal scaling transformation of the temporal domain $t' = S_\tau t$:

$$L'(x',t'; s',\tau') = L(x,t; s,\tau) \quad \Leftrightarrow \quad \mathcal{T}_{S_\tau(s,\tau)} S_\tau f = S_\tau \mathcal{T}_{s,\tau} f \quad (11)$$

to hold for some transformation $(s',\tau') = S_\tau(s,\tau)$ of the spatio-temporal scale parameters (s,τ).

b) Galilean covariance under Galilean transformations. If an observer looks at the same object in the world for different relative motions $v = (v_1, v_2)$ between the object and the observer, it is natural to require that the internal scale-space representations of the object should be sufficiently similar, so as to enable a coherent perception of the object under different relative motions relative to the observer. Specifically, we may require that the receptive field responses under relative motions should remain on the same form, while being transformed in a corresponding way as the relative motion pattern.

If we at any point in space-time locally linearize the possibly non-linear motion pattern $x(t) = (x_1(t), x_2(t))$ by a local Galilean transformation $x' = x + vt$ over space-time

$$f' = \mathcal{G}_v f \quad \Leftrightarrow \quad f'(x',t') = f(x,t) \quad \text{with} \quad x' = x + vt, \quad (12)$$

then the requirement of guaranteeing a consistent visual interpretation under different relative motions between the object and the observer can be stated as a requirement of Galilean covariance:

$$L'(x',t'; s',\tau') = L(x,t; s,\tau) \quad \Leftrightarrow \quad \mathcal{T}_{G_v(s,\tau)} \mathcal{G}_v f = \mathcal{G}_v \mathcal{T}_{s,\tau} f \quad (13)$$

to hold for some transformation $G_v(s,\tau)$ of the spatio-temporal scale parameters (s,τ).

c) Semi-group structure over temporal scales in the case of a non-causal temporal domain. To ensure that the image representations between different spatio-temporal scale levels (s_1,τ_1) and (s_2,τ_2) should be sufficiently well-behaved internally, we will make use of different types of assumptions depending on whether the temporal domain is regarded as time-causal or non-causal. Over a time-causal temporal domain, the future cannot be accessed, which is the basic condition for real-time visual perception by a biological organism. Over a non-causal temporal domain, the temporal kernels may extend to the relative future in relation to any pre-recorded time moment, which is sometimes used as a conceptual simplification when analysing pre-recorded time-dependent data, although not at all realistic in a real-world setting.

For the case of a non-causal temporal domain, we make use of a similar type of semi-group property (8) as formulated over a purely spatial domain, while extending the semi-group property over both the spatial scale parameter s and the temporal scale parameter τ:

$$\mathcal{T}_{s_1,\tau_1} \mathcal{T}_{s_2,\tau_2} = \mathcal{T}_{s_1+s_2,\tau_1+\tau_2}. \quad (14)$$

In analogy with the case of a purely spatial domain, this requirement guarantees that the transformation from any finer spatio-temporal scale level (s_1,τ_1) to any coarser spatio-temporal scale level $(s_2,\tau_2) \geq (s_1,\tau_1)$ will always be of the same form (algebraic closedness)

$$L(\cdot,\cdot; s_2,\tau_2) = \mathcal{T}_{s_2-s_1,\tau_2-\tau_1} L(\cdot,\cdot; s_1,\tau_1). \quad (15)$$

Specifically, this assumption implies that if we are able to establish desirable properties of the family of transformations $\mathcal{T}_{s,\tau}$ (to be detailed below), then these relations hold between any pair of spatio-temporal scale levels (s_1,τ_1) and (s_2,τ_2) with $(s_2,\tau_2) \geq (s_1,\tau_1)$.

d) Cascade structure over temporal scales in the case of a time-causal temporal domain. Since it can be shown that the assumption of a semi-group structure over temporal scales leads to undesirable temporal dynamics in terms of *e.g.* longer temporal delays for a time-causal temporal domain [55, Appendix A], we do for a time-causal temporal domain instead assume a weaker cascade smoothing property over temporal scales for the temporal smoothing kernel over temporal scales

$$L(\cdot; \tau_2) = h(\cdot; \tau_1 \mapsto \tau_2) * L(\cdot; \tau_1), \quad (16)$$

where the temporal kernels $h(t; \tau)$ should for any triplets of temporal scale values and temporal delays τ_1, τ_2 and τ_3 obey the transitive property

$$h(\cdot; \tau_1 \mapsto \tau_2) * h(\cdot; \tau_2 \mapsto \tau_3) = h(\cdot; \tau_1 \mapsto \tau_3). \quad (17)$$

This weaker assumption of a cascade smoothing property (16) still ensures that an image representation at a coarser temporal scale τ_2 should with a corresponding requirement of an accompanying simplifying condition on the family of kernels h (to be detailed below) constitute a simplification of the representation at a finer temporal scale τ_1, while not implying as hard constraints as a semi-group structure.

e) Non-enhancement of local space-time extrema in the case of a non-causal temporal domain. In the case of a non-causal temporal domain,

we again build on the notion of non-enhancement of local extrema to guarantee that the image representations at coarser spatio-temporal scales should constitute true simplifications of corresponding representations at finer scales. Over a spatio-temporal domain, we do, however, state the requirement in terms of local extrema over joint space-time instead of over local extrema over image space. If a point (x_0, t_0) for some scale (s_0, τ_0) is a local maximum point over space-time, then the value at this maximum point $L(x_0, t_0; s_0, \tau_0)$ must not increase to coarser spatio-temporal scales $(s, \tau) \geq (s_0, \tau_0)$. Similarly, if a point is a local minimum point over space-time, then the value at this minimum point $L(x_0, t_0; s_0, \tau_0)$ must not decrease to coarser spatio-temporal scales $(s, \tau) \geq (s_0, \tau_0)$.

Formally, this requirement of non-creation of new structure from finer to coarser spatio-temporal scales can be stated as follows: If at some scale (s_0, τ_0) a point (x_0, t_0) is a local maximum (minimum) for the mapping from (x, t) to $L(x, t; s_0, \tau_0)$, then

- $\alpha\,(\partial_s L)(x, t; \ s, \tau) + \beta\,(\partial_\tau L)(x, t; \ s, \tau) \leq 0$ at any spatio-temporal maximum
- $\alpha\,(\partial_s L)(x, t; \ s, \tau) + \beta\,(\partial_\tau L)(x, t; \ s, \tau) \geq 0$ at any spatio-temporal minimum

should hold in any positive spatio-temporal direction defined from any non-negative linear combinations of α and β. This condition implies a strong condition on the class of possible smoothing kernels $T(\cdot, \cdot; \ s, \tau)$.

f) Non-creation of new local extrema or zero-crossings for a purely temporal signal in the case of a time-causal temporal domain. In the case of a time-causal temporal domain, we do instead state a requirement for purely temporal signals, based on the cascade smoothing property (16). We require that for a purely temporal signal $f(t)$, the transformation from a finer temporal scale τ_1 to a coarser temporal scale τ_2 must not increase the number of local extrema or the number of zero-crossings in the signal.

3. Idealized receptive field families

3.1. Spatial image domain

Based on the above assumptions in Section 2.1, it can be shown [13] that when complemented with certain regularity assumptions in terms of Sobolev norms, they imply[5] that a spatial scale-space representation L as determined by these assumptions must satisfy a diffusion equation of the form[6]

---

[5] The formal derivation of this result concerning spatial receptive fields is based on two main conceptual steps: In a first stage, we establish that from the assumptions of linearity and translational covariance in combination with a semi-group structure over a continuous scale parameter and certain regularity assumptions in terms of Sobolev norms, the internal scale-space representations according to (1) must obey an evolution equation over scale of the form $\partial_s L = \mathcal{A}L$ for some linear and translation covariant operator $\mathcal{A}$, see [13, Lemma 2]. In a second stage, we show that by adding the requirement of non-enhancement of local extrema, the operator $\mathcal{A}$ may only correspond to second- and first-order derivatives at the central point, corresponding to a diffusion equation of the form (18), see [13, Theorem 5].

[6] For the theoretical analysis in this treatment, we consider the affine diffusion equation as defined over the entire infinite image plane, $x = (x_1, x_2)^T \in \mathbb{R}^2$, and normally with initial equation $L(x_1, x_2; s) = f(x_1, x_2)$ for $s = 0$. When using this equation to model the effect of visual operations for simulation experiments on finite images, we solve this equation with adiabatic boundary conditions, corresponding to no heat flow across the image boundaries, and equivalent to reflections of the image data at the image boundaries, when implementing the spatial smoothing operation in terms of explicit Gaussian filtering. This turns out to be a good approximation for moderate fields of view. An alternative approach for handling large fields of view is by reformulating this equation on a sphere, corresponding to computations on the viewsphere.

$$\partial_s L = \frac{1}{2}\,\nabla^T (\Sigma \nabla L) - \delta^T \nabla L \qquad (18)$$

for some symmetric positive definite covariance matrix $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}$ and some translation vector $\delta = (\delta_1, \delta_2)^T$, where ∇ denotes the (vertical) spatial gradient operator and ^{T} its transpose such that $\nabla^T = (\partial_{x_1}, \partial_{x_2})$.

Expanding the matrix and vector notation to elements, this equation can also be written

$$\partial_s L = \frac{1}{2}\,(\Sigma_{11} L_{x_1 x_1} + 2\Sigma_{12} L_{x_1 x_2} + \Sigma_{22} L_{x_2 x_2}) - \delta_1 L_{x_1} - \delta_2 L_{x_2}, \qquad (19)$$

where the subscripts with respect to s, x_1 and x_2 denote derivatives with respect to these variables.

This expression is physically analogous to a diffusion equation that describes how a heat distribution corresponding to the image intensities L evolves as function of time in an inhomogeneous medium with spatial scale s here taking the role of time, with the intensities of the input image f as initial condition $L(x_1, x_2; s) = f(x_1, x_2)$ for $s = 0$. The first term, that depends upon Σ, describes how the image intensity L diffuses as the scale parameter s increases, as function of the anisotropic heat conductivity Σ, which determines how the image intensities may diffuse differently in different spatial directions. The second term, that depends upon δ, describes how the image intensities are translated to other spatial positions as function of a spatial drift velocity δ.

The first effect results in a smoothing effect that may be different in different spatial directions as determined by the anisotropic covariance matrix Σ. With regard to spatial image transformations, variations of the scale parameter s lead to solutions that obey the assumption of scale covariance, to handle objects of different sizes in the world and objects at different distances to the observer.

More general variations of the shape of the covariance matrix Σ do additionally allow for affine covariance, to enable matching of objects that are viewed from different viewing directions relative to the local tangent plane of a smooth surface.

The second translation effect, as determined by the translation vector δ, is relevant for handling image disparities between binocular eyes or, for time-dependent images, image structures that move as function of time.

In terms of convolution kernels, the solution of (18) corresponds to convolution with Gaussian kernels of the form

$$g(x; \ \Sigma_s, \delta_s) = \frac{1}{2\pi \sqrt{\det \Sigma_s}}\, e^{-(x - \delta_s)^T \Sigma_s^{-1} (x - \delta_s)/2}, \qquad (20)$$

which for a given $\Sigma_s = s\,\Sigma$ and a given $\delta_s = s\,\delta$ satisfy (18) (see Appendix A in the supplement for an explicit proof of the property that the family of internal spatial scale-space representations L generated by convolution with kernels of the form (20) satisfies the diffusion equation (18)).

If we additionally require these kernels to be mirror symmetric through the origin, then we obtain *affine Gaussian kernels*

$$g(x; \ \Sigma) = \frac{1}{2\pi \sqrt{\det \Sigma}}\, e^{-x^T \Sigma^{-1} x/2}. \qquad (21)$$

Their spatial derivatives constitute a canonical family for expressing receptive fields over a spatial domain that can be summarized and reparameterized on the form

$$T(x; \ s, \Sigma) = g(x; \ s\,\Sigma) = \frac{1}{2\pi s \sqrt{\det \Sigma}}\, e^{-x^T \Sigma^{-1} x/2s}, \qquad (22)$$

where we have separated the parameters into two components; a scalar scale parameter s that represents the size of the Gaussian, *i.e.*, how large it is in the image domain, and a matrix Σ that determines its shape, *i.e.*, how eccentric it is (the ratio between the sizes in the perpendicular principal directions for an anisotropic Gaussian) and the orientation of the main principal axis in the image domain.
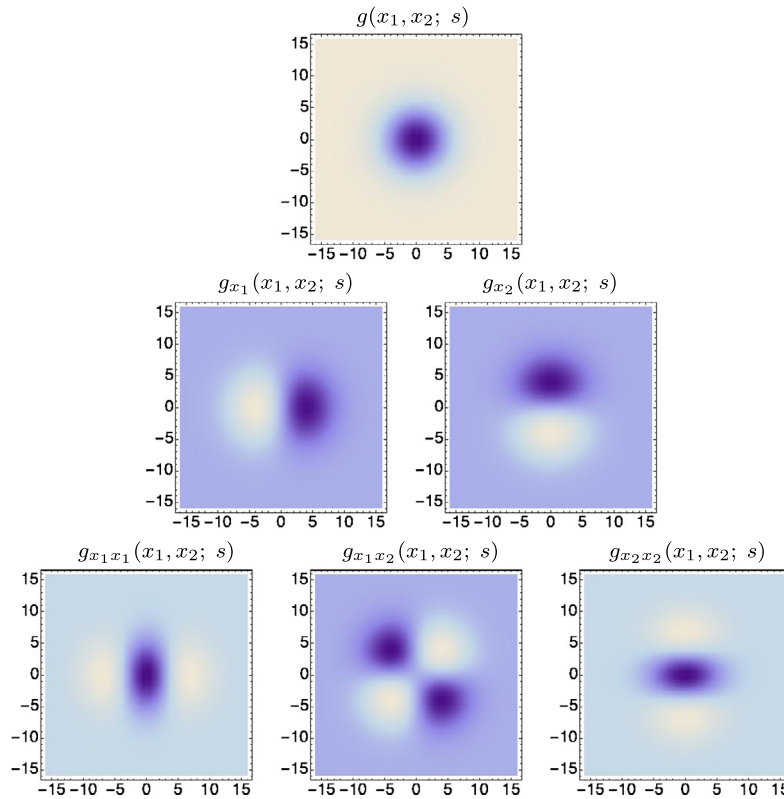
Fig. 6. Illustrations of spatial receptive fields formed by the 2-D rotationally symmetric Gaussian kernel (for $s = 16$) and its partial derivatives up to order two. The resulting receptive fields are closed under translations, rotations and scaling transformations. This means that if an image is transformed in these ways, then it will always be possible to find some possibly other receptive field such that the receptive field responses of the original image and the transformed image can be perfectly matched.

The spatial scale-space representations that are obtained by convolution with kernels of this form obey (i) spatial scale covariance as described in Appendix C in the supplement and illustrated in Fig. 21 in the supplement and (ii) spatial affine covariance as described in Appendix D in the supplement and illustrated in Fig. 22 in the supplement.

Incorporating the fact that spatial derivatives of the kernels (22) are also compatible with the assumptions underlying this theory, this does specifically for the case of a two-dimensional spatial image domain lead to spatial receptive fields that can be compactly summarized on the form

$$T_{\varphi^{m_1} \perp \varphi^{m_2}}(x_1, x_2; \ s, \Sigma) = \partial_\varphi^{m_1} \partial_{\perp\varphi}^{m_2} \left(g(x_1, x_2; \ s\Sigma) \right), \quad (23)$$
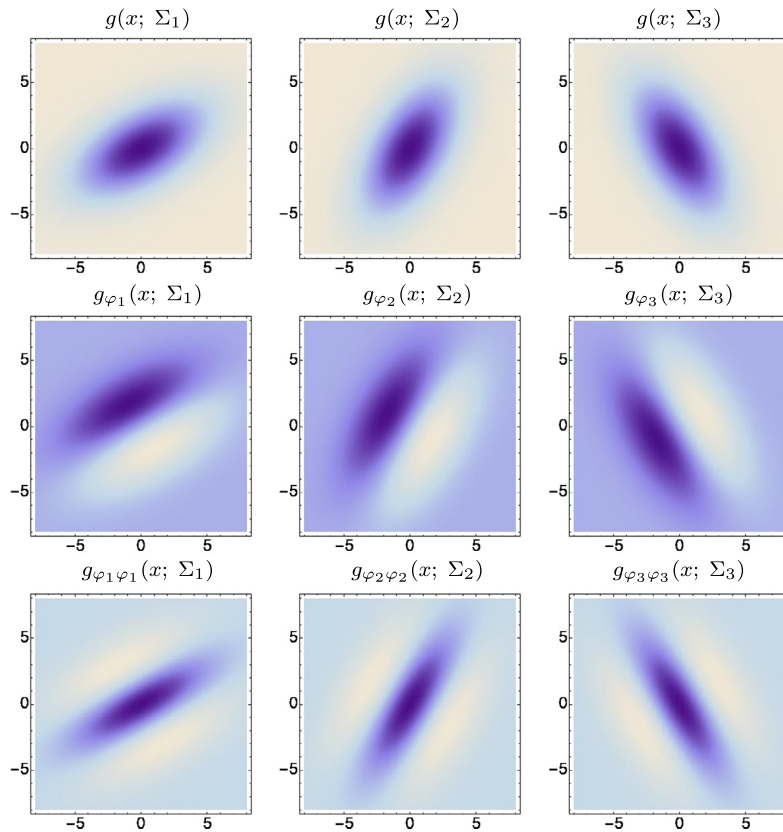
where

- $x = (x_1, x_2)$ denotes the spatial coordinates,
- s denotes the spatial scale[7] in units of $s = \sigma^2$, where σ has dimension [length] and corresponds to the standard deviation of the Gaussian kernel for an isotropic covariance matrix with $\Sigma = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$,

- Σ denotes a spatial covariance matrix determining the shape of a spatial affine Gaussian kernel (this covariance matrix is assumed to be symmetric positive definite such that $x^T \Sigma x = \Sigma_{11} x_1^2 + 2\Sigma_{12} x_1 x_2 + \Sigma_{22} x_2^2 > 0$ for any $x = (x_1, x_2)^T \neq 0$),
- $\partial_\varphi = \cos \varphi \, \partial_{x_1} + \sin \varphi \, \partial_{x_2}$ and $\partial_{\perp\varphi} = \sin \varphi \, \partial_{x_1} - \cos \varphi \, \partial_{x_2}$ denote spatial directional derivative operators in two orthogonal directions φ and $\perp\varphi$ aligned with the eigenvectors of the covariance matrix Σ, where ∂_{x_1} and ∂_{x_2} denote differentiation with respect the spatial coordinates x_1 and x_2,
- m_1 and m_2 denote orders of spatial differentiation in the spatial direction φ and its orthogonal direction $\perp\varphi$, respectively,
- $g(x; \ s\Sigma) = \frac{1}{2\pi s \sqrt{\det \Sigma}} e^{-x^T \Sigma^{-1} x / 2s}$ is an affine Gaussian kernel with its size determined by the spatial scale parameter s and its shape by the spatial covariance matrix Σ.

Fig. 6 and Fig. 7 show examples of spatial receptive fields from this family up to second order of spatial differentiation. Fig. 6 shows partial derivatives of the Gaussian kernel for the specific case when the covariance matrix Σ is restricted to a unit matrix and the Gaussian kernel thereby becomes rotationally symmetric. The resulting family of receptive fields is closed under scaling transformations over the spatial domain, implying that if an object is seen from different distances to the observer, then it will always be possible to find a transformation of the scale parameter s between the two image domains such that the receptive field responses computed from the two image domains can be matched. Fig. 7 shows examples of affine Gaussian receptive fields for covariance matrices Σ that do not correspond to rescaled copies of the unit matrix. The resulting full family of affine Gaussian derivative kernels is closed under general affine transformations, implying that for two different perspective views of a local smooth surface patch, it will always be possible to find a transformation of the covariance matrices Σ between the two domains so that the receptive field responses can

---

[7] Here, we parameterize the spatial scale parameter in units of s, where s after the mathematical derivations turns out to have dimension [length]². With this parameterization, the scale values become additive under the convolution operation (5), because of the semi-group property in Equation (6). The scale values in units of s also become additive under evolution according to the diffusion equation (18). If two Gaussian convolutions with scale parameters s_1 and s_2 are performed after each other, the composed smoothing effect is $s_3 = s_1 + s_2$. If the scale parameters would instead have been measured in units of $\sigma = \sqrt{s}$, with σ having dimension [length], then the corresponding composed smoothing effect would instead be written $\sigma_3 = \sqrt{\sigma_1^2 + \sigma_2^2}$.

Fig. 7. Illustrations of spatial receptive fields formed by affine Gaussian kernels and directional derivatives of these up to order two of, here visualized for three different covariance matrices Σ_1, Σ_2 and Σ_3 that correspond to the major eigendirections $\theta_1 = \pi/6$, $\theta_2 = \pi/3$ and $\theta_3 = 2\pi/3$ of the covariance matrix and with directional derivatives computed in the corresponding orthogonal directions φ_1, φ_2 and φ_3. The resulting family of receptive fields is closed under general affine transformations of the spatial domain, including translations, rotations, scaling transformations and perspective foreshortening. In this figure, however, only variabilities in the orientation of the filter are illustrated, thereby disregarding variabilities in both the size and the degree of elongation. This closedness property implies that receptive field responses computed from different views of a smooth local surface patch can be perfectly matched, if the transformation between the two views can be modelled as a local affine transformation. (Scale parameters $s_1 = 16$ and $s_2 = 4$ in the orthogonal eigendirections of the spatial covariance matrices Σ_i.)
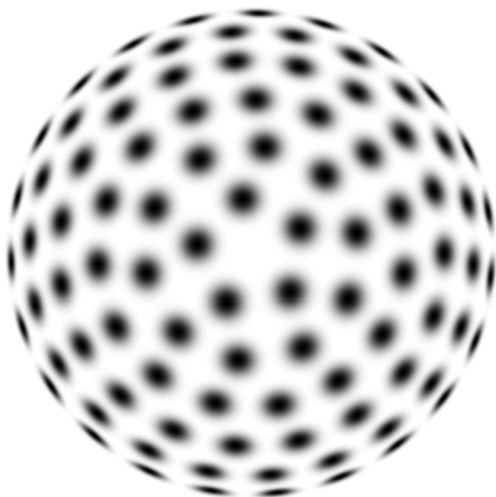
Fig. 8. Illustration of the variability of zero-order affine Gaussian receptive fields for a uniform distribution on a hemisphere. In the most idealized version of the theory, one can think of all affine receptive fields with their directional derivatives in preferred directions aligned to the eigendirections of the covariance matrix Σ as being present at any position in the image domain. This variability makes it possible to perfectly match the first-order variability of receptive field responses under variations of the slant and tilt directions of a smooth surface patch.

be matched, if the transformation between the two image domains is approximated by a local affine transformation.

In the most idealized version of the theory, one should think of receptive fields for all combinations of filter parameters as being present at every image point, as illustrated in Fig. 8 concerning affine Gaussian receptive fields over different orientations in image space and different eccentricities.

3.2. Spatio-temporal image domain

Over a non-causal spatio-temporal domain, corresponding arguments as in Section 3.1 lead to a similar form of diffusion equation as in Equation (18), while expressed over the joint space-time domain $p = (x, t)$. After splitting the composed affine Gaussian spatio-temporal smoothing kernel corresponding to (20), while expressed over the joint space-time domain, into separate smoothing operations over space and time, this leads to zero-order spatio-temporal receptive fields of the form [13, 19] (see Appendix B.1 in the supplement for an overview of the logical steps in the derivation that lead to this result):

$$T(x_1, x_2, t; \ s, \tau; \ v, \Sigma) = g(x_1 - v_1 t, x_2 - v_2 t; \ s\,\Sigma)\, h(t; \ \tau), \qquad (24)$$

where the temporal smoothing kernel $h(t; \ \tau)$ is a one-dimensional Gaussian kernel. After combining that result with the results from corresponding theoretical analysis for a time-causal spatio-temporal domain in [13, 21] (see Appendix B.2 in the supplement for an overview of the logical steps behind this construction), we are lead to a similar form of spatio-temporal smoothing operation, while then using a temporal

smoothing kernel $h(t; \tau)$ that corresponds to a set of truncated exponential kernels coupled in cascade.

The resulting spatio-temporal scale-space representation obeys (i) spatial scale covariance as described in Appendix E in the supplement and illustrated in Fig. 23 in the supplement, (ii) spatial affine covariance as described in Appendix F in the supplement and illustrated in Fig. 24 in the supplement, (iii) Galilean covariance as described in Appendix G in the supplement and illustrated in Fig. 25 in the supplement and (iv) temporal scale covariance as described in Appendix H in the supplement and illustrated in Fig. 26 in the supplement.

After noting that spatial, temporal and spatio-temporal derivatives of the spatio-temporal smoothing kernels (24) are also compatible with the assumptions because of the linearity assumption, the resulting spatio-temporal derivative kernels constituting the spatio-temporal extension of the spatial receptive field model (23) can be reparameterized and summarized on the following form (see [13, 19, 20, 21]):

$$T_{\varphi^{m_1} \perp \varphi^{m_2} \bar{t}^n}(x_1, x_2, t; s, \tau; v, \Sigma) =$$
$$\partial_\varphi^{m_1} \partial_{\perp\varphi}^{m_2} \partial_{\bar{t}}^n \left(g(x_1 - v_1 t, x_2 - v_2 t; s\Sigma) h(t; \tau) \right), \quad (25)$$

where

- $x = (x_1, x_2)$ denotes the spatial coordinates,
- t denotes time,
- s denotes the spatial scale (in dimension of $[\text{length}]^2$),
- τ denotes the temporal scale (in dimension of $[\text{time}]^2$),
- $v = (v_1, v_2)^T$ denotes a local image velocity,
- Σ denotes a spatial covariance matrix determining the shape of a spatial affine Gaussian kernel,
- $\partial_\varphi = \cos\varphi \, \partial_{x_1} + \sin\varphi \, \partial_{x_2}$ and $\partial_{\perp\varphi} = \sin\varphi \, \partial_{x_1} - \cos\varphi \, \partial_{x_2}$ denote spatial directional derivative operators in two orthogonal directions φ and $\perp\varphi$ aligned with the eigenvectors of the covariance matrix Σ,
- $\partial_{\bar{t}} = v_1 \partial_{x_1} + v_2 \partial_{x_2} + \partial_t$ is a velocity-adapted temporal derivative operator aligned to the direction of the local image velocity $v = (v_1, v_2)^T$,
- m_1 and m_2 denote orders of spatial differentiation,
- n denotes the order of temporal differentiation,
- $g(x; s\Sigma) = \frac{1}{2\pi s \sqrt{\det \Sigma}} e^{-x^T \Sigma^{-1} x / 2s}$ is an affine Gaussian kernel with its size determined by the spatial scale parameter s and its shape determined by the spatial covariance matrix Σ,
- $g(x_1 - v_1 t, x_2 - v_2 t; s\Sigma)$ denotes a spatial affine Gaussian kernel that moves with image velocity $v = (v_1, v_2)$ in space-time and
- $h(t; \tau)$ is a temporal smoothing kernel over time corresponding to a Gaussian kernel $h(t; \tau) = g(t; \tau) = 1/\sqrt{2\pi\tau} \exp(-t^2/2\tau)$ in the case of non-causal time or a cascade of first-order integrators or equivalently truncated exponential kernels coupled in cascade $h(t; \tau) = h_{composed}(\cdot; \mu)$ according to (27) over a time-causal temporal domain.

This family of spatio-temporal scale-space kernels can be seen as a canonical family of linear receptive fields over a spatio-temporal domain.

For the case of a time-causal temporal domain, the result states that truncated exponential kernels of the form

$$h_{exp}(t; \mu_k) = \begin{cases} \frac{1}{\mu_k} e^{-t/\mu_k} & t \geq 0 \\ 0 & t > 0 \end{cases} \quad (26)$$

coupled in cascade constitute the natural temporal smoothing kernels. These do in turn lead to a composed temporal smoothing kernel of the form

$$h_{composed}(\cdot; \mu) = *_{k=1}^K h_{exp}(\cdot; \mu_k) \quad (27)$$

and corresponding to a set of first-order integrators coupled in cascade (see Fig. 9).
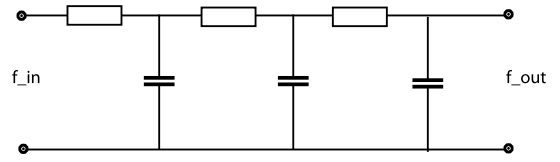

Fig. 9. Illustration of the time-causal receptive field model in terms of an electric wiring diagram composed of a set of resistors and capacitors that emulate a series of first-order integrators coupled in cascade. In this model, the time-varying voltage f_{in} represents the time varying input signal, whereas the time-varying voltage f_{out} represents the time-varying output signal at a coarser temporal scale. From the theory for temporal scale-space kernels for one-dimensional signals (Lindeberg [21, 56]; Lindeberg and Fagerström [57]), it holds that the corresponding equivalent truncated exponential kernels are the only primitive temporal smoothing kernels that guarantee both temporal causality and non-creation of local extrema (or zero-crossings) with increasing temporal scale.

Two natural ways of distributing the discrete time constants μ_k over temporal scales are studied in detail in [21, 55] corresponding to either a uniform or a logarithmic distribution in terms of the composed variance

$$\tau_K = \sum_{k=1}^K \mu_k^2. \quad (28)$$

Specifically, it is shown in [21, Section 5] that in the case of a logarithmic distribution of the discrete temporal scale levels, it is possible to consider an infinite number of temporal scale levels that cluster infinitely dense near zero temporal scale

$$\ldots \frac{\tau_0}{c^6}, \frac{\tau_0}{c^4}, \frac{\tau_0}{c^2}, \tau_0, c^2\tau_0, c^4\tau_0, c^6\tau_0, \ldots \quad (29)$$

so that a *scale-covariant time-causal limit kernel* $\Psi(t; \tau, c)$ can be defined obeying self-similarity and scale covariance over temporal scales and with a Fourier transform of the form

$$\hat{\Psi}(\omega; \tau, c) = \prod_{k=1}^{\infty} \frac{1}{1 + i \, c^{-k} \sqrt{c^2 - 1} \sqrt{\tau} \, \omega}. \quad (30)$$

Fig. 10 and Fig. 11 show spatio-temporal kernels over a $1+1$-dimensional spatio-temporal domain using approximations of the time-causal limit kernel for temporal smoothing over the temporal domain and the Gaussian kernel for spatial smoothing over the spatial domain. Fig. 10 shows space-time separable receptive fields corresponding to image velocity $v = 0$, whereas Fig. 11 shows unseparable velocity-adapted receptive fields corresponding to a non-zero image velocity $v \neq 0$.

The family of space-time separable receptive fields for zero image velocities is closed under spatial scaling transformations for arbitrary spatial scaling factors as well as for temporal scaling transformations with temporal scaling factors that are integer powers of the distribution parameter c of the time-causal limit kernel. The full family of velocity-adapted receptive fields for general non-zero image velocities is additionally closed under Galilean transformations, corresponding to variations in the relative motion between the objects in the world and the observer. Given that the full families of receptive fields are explicitly represented in the vision system, this means that it will be possible to perfectly match receptive field responses computed under the following types of natural image transformations: (i) objects of different size in the image domain as arising from *e.g.* viewing the same object from different distances, (ii) spatio-temporal events that occur with different speed, faster or slower, and (iii) objects and spatio-temporal events that are viewed with different relative motions between the objects/event and the visual observer.

If additionally the spatial smoothing is performed over the full family of spatial covariance matrices Σ, then receptive field responses can also be matched (iv) between different views of the same smooth local surface patch.
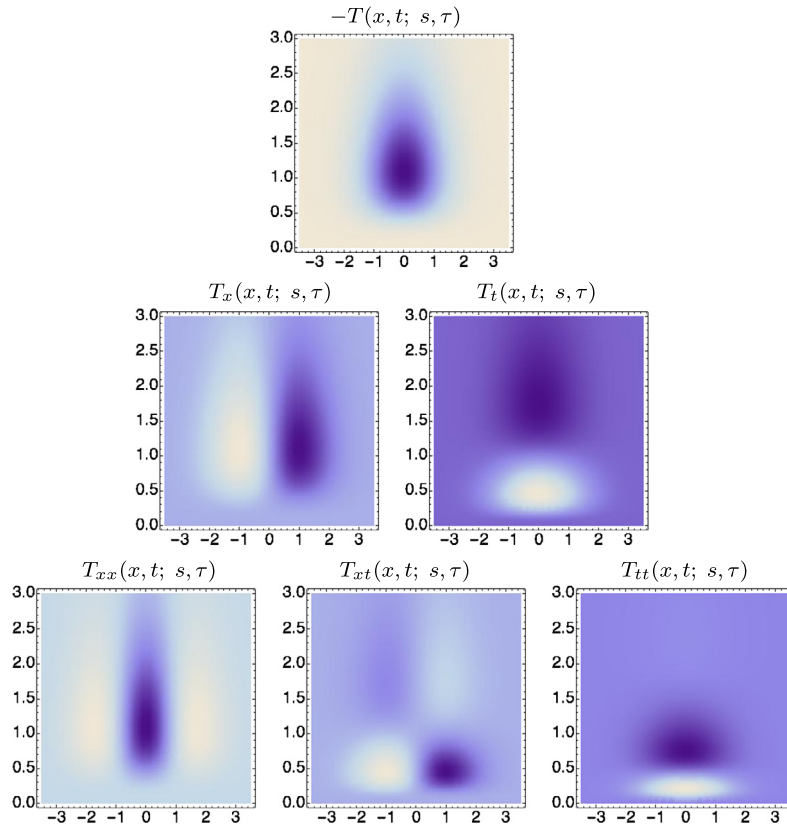
Fig. 10. *Illustrations of space-time separable receptive fields $T_{x^m t^n}(x, t; s, \tau) = \partial_{x^m t^n}(g(x; s) h(t; \tau))$ up to order two, formed from by the composition of Gaussian kernels over the spatial domain x for spatial scale parameter $s = 1$ and a set of truncated exponential kernels coupled in cascade over the temporal domain t according to Equation (27), with a logarithmic distribution of the intermediate temporal scale levels that approximates the time-causal limit kernel in Equation (30) with the following parameters: $\tau = 1$, $K = 7$, $c = 2$, $v = 0$. The corresponding family of spatio-temporal receptive fields is closed under spatial scaling transformations as well as under temporal scaling transformations for temporal scaling factors that are integer powers of the distribution parameter c of the temporal smoothing kernel. (Horizontal axis: space x. Vertical axis: time t.)*

3.3. Scale normalisation of spatial and spatio-temporal receptive fields

When computing receptive field responses over multiple spatial and temporal scales, there is an issue about how the receptive field responses should be normalized so as to enable appropriate comparisons between receptive field responses at different scales. Issues of scale normalisation of the derivative based receptive fields defined from scale-space operations are treated in [36, 58, 59] regarding spatial receptive fields and in [21, 38, 55] regarding spatio-temporal receptive fields.

a) Scale-normalized spatial receptive fields. Let s_φ and $s_{\perp\varphi}$ denote the eigenvalues of the composed affine covariance matrix $s\Sigma$ in the spatial receptive field model (23) and let ∂_φ and $\partial_{\perp\varphi}$ denote directional derivative operators along the corresponding eigendirections. Then, the scale-normalized spatial derivative kernel corresponding to the receptive field model (23) is given by

$$T_{\varphi^{m_1} \perp\varphi^{m_2}, norm}(x_1, x_2; s, \Sigma) = s_\varphi^{m_1\gamma_s/2} s_{\perp\varphi}^{m_2\gamma_s/2} \partial_\varphi^{m_1} \partial_{\perp\varphi}^{m_2} \left(g(x_1, x_2; s\Sigma) \right), \qquad (31)$$

where γ_s denotes the spatial scale normalization parameter of γ-normalized derivatives and specifically the choice $\gamma_s = 1$ leads to maximum scale invariance in the sense that the magnitude response of the spatial receptive field will be preserved under uniform spatial scaling transformations $(x_1', x_2') = (S_s x_1, S_s x_2)$, provided that the spatial scale levels are appropriately matched $(s_\varphi', s_{\perp\varphi}') = (S_s^2 s_\varphi, S_s^2 s_{\perp\varphi})$.

b) Scale-normalized spatial receptive fields in the case of a non-causal spatio-temporal domain. For the case of a non-causal spatio-temporal domain, where the temporal smoothing operation in the spatio-temporal receptive field model is performed by a non-causal Gaussian temporal kernel $h(t; \tau) = g(t; \tau) = 1/\sqrt{2\pi\tau} \exp(-t^2/2\tau)$, the scale-normalized

spatio-temporal derivative kernel corresponding to the spatio-temporal receptive field model (25) is with corresponding notation regarding the spatial domain as in (31) given by

$$\begin{aligned} &T_{\varphi^{m_1} \perp\varphi^{m_2} \bar{t}^n, norm}(x_1, x_2, t; s, \tau; v, \Sigma) \\ &= s_\varphi^{m_1\gamma_s/2} s_{\perp\varphi}^{m_2\gamma_s/2} \tau^{n\gamma_\tau/2} \\ &\quad \partial_\varphi^{m_1} \partial_{\perp\varphi}^{m_2} \partial_{\bar{t}}^n \left(g(x_1 - v_1 t, x_2 - v_2 t; s\Sigma) h(t; \tau) \right), \end{aligned} \qquad (32)$$

where γ_s and γ_τ denote the spatial and temporal scale normalization parameters of γ-normalized derivatives and specifically the choice $\gamma_s = 1$ and $\gamma_\tau = 1$ leads to maximum scale invariance in the sense that the magnitude response of the spatio-temporal receptive field will be preserved under independent scaling transformations of the spatial and the temporal domains $(x_1', x_2', t') = (S_s x_1, S_s x_2, S_\tau t)$, provided that both the spatial and temporal scale levels are appropriately matched $(s_\varphi', s_{\perp\varphi}', \tau') = (S_s^2 s_\varphi, S_s^2 s_{\perp\varphi}, S_\tau^2 \tau)$.

c) Scale-normalized spatial receptive fields in the case of a time-causal spatio-temporal domain. For the case of a time-causal spatio-temporal domain, where the temporal smoothing operation in the spatio-temporal receptive field model is performed by truncated exponential kernels coupled in cascade $h(t; \tau) = h_{composed}(\cdot; \mu)$ (27), the corresponding scale-normalized spatio-temporal derivative kernel corresponding to the spatio-temporal receptive field model (25) is given by

$$\begin{aligned} &T_{\varphi^{m_1} \perp\varphi^{m_2} \bar{t}^n, norm}(x_1, x_2, t; s, \tau; v, \Sigma) \\ &= s_\varphi^{m_1\gamma_s/2} s_{\perp\varphi}^{m_2\gamma_s/2} \alpha_{n,\gamma_\tau}(\tau) \\ &\quad \partial_\varphi^{m_1} \partial_{\perp\varphi}^{m_2} \partial_{\bar{t}}^n \left(g(x_1 - v_1 t, x_2 - v_2 t; s\Sigma) h(t; \tau) \right), \end{aligned} \qquad (33)$$
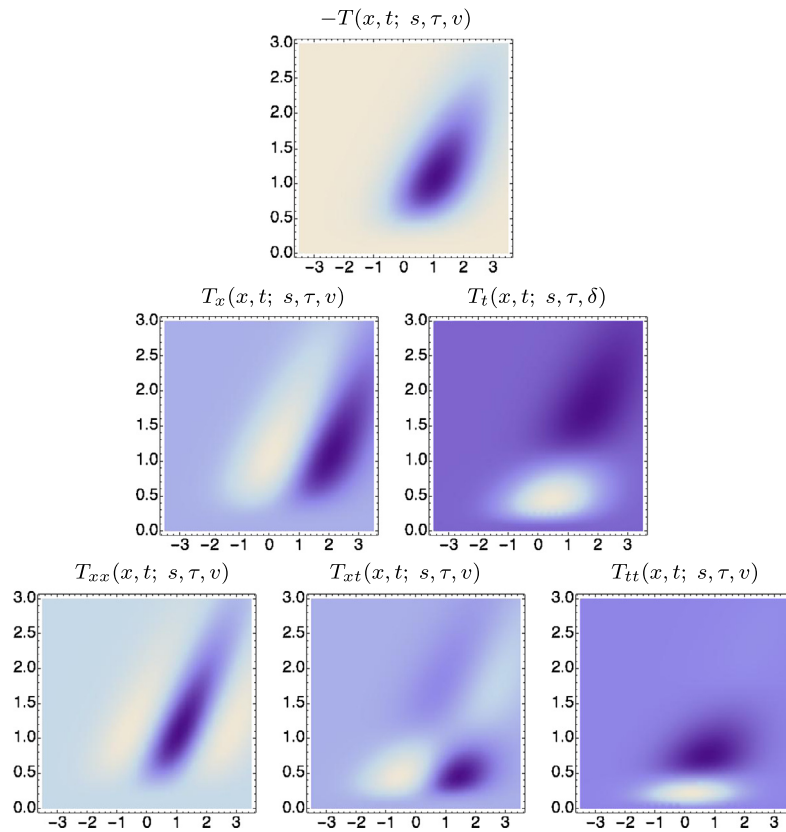
Fig. 11. *Illustrations of velocity-adapted spatio-temporal receptive fields $T_{x^m t^n}(x, t; \; s, \tau, v) = \partial_{x^m t^n}(g(x - vt; \; s) \, h(t; \; \tau))$ up to order two, formed from the composition of Gaussian kernels over the spatial domain x for spatial scale parameter $s = 1$ and a set of truncated exponential kernels coupled in cascade over the temporal domain t according to Equation (27), with a logarithmic distribution of the intermediate temporal scale levels that approximates the time-causal limit kernel in Equation (30) with the following parameters: $\tau = 1$, $K = 7$, $c = 2$, $v = 1$. In addition to spatial and temporal scaling transformations, the corresponding family of receptive fields is also closed under Galilean transformations. (Horizontal axis: space x. Vertical axis: time t.)*

where γ_s and γ_τ denote the spatial and temporal scale normalization parameters of γ-normalized derivatives and $\alpha_{n,\gamma_\tau}(\tau)$ is the temporal scale normalization factor, which for the case of variance-based normalization is given by

$$\alpha_{n,\gamma_\tau}(\tau) = \tau^{n\gamma_\tau/2} \tag{34}$$

in agreement with (32), while for the case of L_p-normalization it is given by [21, Equation (76)]

$$\alpha_{n,\gamma_\tau}(\tau) = \frac{G_{n,\gamma_\tau}}{\|h_{t^n}(\cdot; \; \tau)\|_p}, \tag{35}$$

with G_{n,γ_τ} denoting the L_p-norm of the n:th order scale-normalized derivative of a non-causal Gaussian temporal kernel with scale normalization parameter γ_τ. In the specific case when the temporal smoothing is performed using the scale-invariant limit kernel (30), the magnitude response will for the maximally scale invariant choice of scale normalization parameters $\gamma_s = 1$ and $\gamma_\tau = 1$ be preserved under independent scaling transformations of the spatial and the temporal domains $(x_1', x_2', t') = (S_s x_1, S_s x_2, S_\tau t)$ for general spatial scaling factors S_s and for temporal scaling factors $S_\tau = c^j$ that are integer powers of the distribution parameter c of the scale-invariant limit kernel, provided that both the spatial and temporal scale levels are appropriately matched $(s_\varphi', s_{\perp\varphi}', \tau') = (S_s^2 s_\varphi, S_s^2 s_{\perp\varphi}, S_\tau^2 \tau)$.

3.4. Invariance to local multiplicative illumination variations or variations in exposure parameters

The treatment so far has been concerned with modelling receptive fields under natural geometric image transformations, modelled as local

scaling transformations, local affine transformations and local Galilean transformations representing the essential dimensions in the variability of a local linearization of the perspective mapping from a local surface patch in the world to the tangent plane of the retina. A complementary issue concerns how to model receptive field responses under variations in the external illumination and under variations in the internal exposure mechanisms of the eye that adapts the diameter of the pupil and the sensitivity of the photoreceptors to the external illumination. In this section, we will present a solution for this problem regarding the subset of intensity transformations that can be modelled as local multiplicative intensity transformations.

To handle image data under illumination variations in a theoretically well-founded manner, it is natural to represent the image data on a logarithmic luminosity scale

$$f(x_1, x_2, t) \sim \log I(x_1, x_2, t). \tag{36}$$

Then, it can be shown that receptive field responses computed from such logarithmic luminosities can be *interpreted physically* as a superposition of relative variations of surface structure and relative variations of illumination variations. To demonstrate why this follows, let us assume: (i) a perspective camera model extended with (ii) a thin circular lens that gathers incoming light from different directions and (iii) a Lambertian illumination model that is complemented with (iv) a spatially varying albedo factor for modelling the light that is reflected from surface patterns in the world. Then, it can be shown [19, Section 2.3] that a spatio-temporal receptive field response

$$L_{\varphi^{m_1} \perp \varphi^{m_2} \bar{t}^n}(\cdot, \cdot; \; s, \; \tau) = \partial_{\varphi^{m_1} \perp \varphi^{m_2} \bar{t}^n} \mathcal{T}_{s,\tau} f(\cdot, \cdot) \tag{37}$$

of the image data f, where $\mathcal{T}_{s,\tau}$ represents the spatio-temporal smoothing operator (here corresponding to a spatio-temporal smoothing kernel of the form (24)) can be expressed as

$$L_{\varphi^{m_1}\perp\varphi^{m_2}\tilde{\iota}^n}(x_1,x_2,t;\;s,\tau) =$$
$$= \partial_{\varphi^{m_1}\perp\varphi^{m_2}\tilde{\iota}^n}\mathcal{T}_{s,\tau}\left(\log\rho(x_1,x_2,t)+\log i(x_1,x_2,t)\right.$$
$$\left.+\log C_{cam}(\tilde{f}(t))+V(x_1,x_2)\right), \qquad (38)$$

where

(i) $\rho(x_1,x_2,t)$ is a spatially dependent *albedo factor* that reflects *properties of surfaces of objects* in the environment (note that this entity may in general refer to points on different surfaces in the world depending on the viewing direction),

(ii) $i(x_1,x_2,t)$ represents a spatially dependent *illumination field* (note that the amount of incoming light may be different for different surfaces world as mapped to corresponding image coordinates (x_1,x_2) over time t),

(iii) $C_{cam}(\tilde{f}(t)) = \frac{\pi}{4}\frac{d}{f}$ represents the possibly time-dependent *internal camera parameters* with the ratio $\tilde{f} = f/d$ referred to as the *effective f-number*, where d denotes the diameter of the lens and f the focal distance, and

(iv) $V(x_1,x_2) = -2\log(1+x_1^2+x_2^2)$ represents a geometric *natural vignetting* effect corresponding to the factor $\log\cos^4(\phi)$ for a planar image plane, with ϕ denoting the angle between the viewing direction (x_1,x_2,f) and the surface normal $(0,0,1)$ of the image plane. This vignetting term disappears for a spherical camera model.

From the way Equation (38) is structured, we can observe that if we have a non-zero order of spatial differentiation with at least some of $m_1 > 0$ or $m_2 > 0$, then the influence of the internal camera parameters in $C_{cam}(\tilde{f}(t))$ will vanish because of the spatial differentiation with respect to x_1 or x_2. In a corresponding manner, the effects of any other multiplicative exposure control mechanism will also vanish. Moreover, for any multiplicative transformation of the illumination field $i'(x_1,x_2) = C\,i(x_1,x_2)$, where C is a constant, the logarithmic luminosity will be transformed as $\log i'(x_1,x_2) = \log C + \log i(x_1,x_2)$. This implies that the dependence on C will disappear after any spatial or temporal differentiation.

Thus, given that the image measurements are performed on a logarithmic brightness scale, the spatio-temporal receptive field responses will be automatically invariant under local multiplicative illumination variations as well as under local multiplicative variations in the exposure parameters of the retina and the eye.

4. Modelling the computational function of biological receptive fields using idealized receptive field profiles

An established methodology to characterize the spatial and temporal response properties of receptive fields in the central visual pathways consists of performing neurophysiological cell recordings of the responses of visual neurons to white noise stimuli. DeAngelis et al. [4, 5] have presented comprehensive surveys of advances made in this way. In these works, the authors emphasize that it is necessary to characterize receptive fields over the *joint* space-time domain, and that it is thus not sufficient to study receptive fields over the spatial domain only. Then, the authors describe basic classes of spatial and spatio-temporal receptive fields in the LGN and the primary visual cortex. Conway and Livingstone [22] and Johnson et al. [23] show results of corresponding investigations concerning spatio-chromatic receptive fields.

In the following, we will outline how the above derived theory for idealized functional models of linear receptive fields can be used for modelling such spatial, spatio-chromatic and spatio-temporal response properties of biological neurons. Indeed, we will show that the derived theory for idealized functional models of linear receptive fields leads to
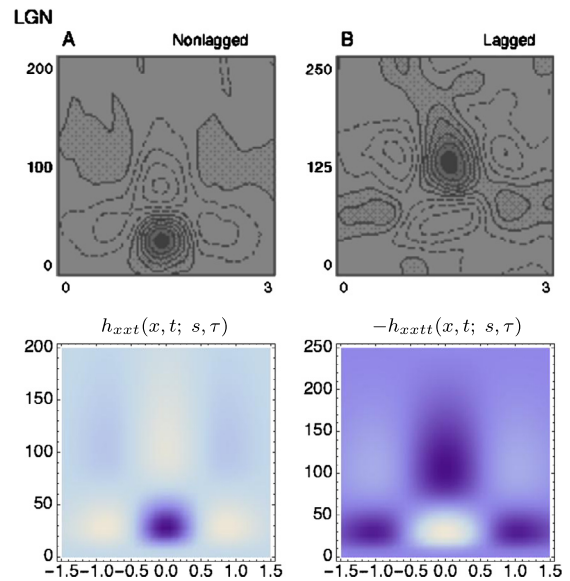
Fig. 12. *Spatio-temporal modelling of LGN neurons.* Regarding space-time separable receptive fields in the lateral geniculate nucleus (LGN), there are two main types: For a "non-lagged cell", the first temporal lobe is strongest, whereas for a "lagged cell", the second temporal lobe is the strongest one. The top row shows examples of such neurons reported by DeAngelis et al. [4]. In the bottom row, we have modelled these receptive fields by idealized spatio-temporal receptive fields of the form $T(x,t;\;s,\tau) = \partial_{x^m}\partial_{t^n}(g(;\;s)\,h(t;\;\tau))$ according to Equation (25), for $m = 2$ corresponding to a Laplacian of Gaussian over the spatial domain, and with the temporal smoothing function $h(t;\;\tau)$ expressed as a cascade of first-order integrators or equivalently truncated exponential kernels of the form (27) and using a logarithmic distribution of the intermediate temporal scale levels. Specifically, in the (left) we model a "non-lagged cell" by first-order temporal derivatives, whereas we model (right) a "lagged cell" using second-order temporal derivatives. Parameter values with $\sigma_x = \sqrt{s}$ and $\sigma_t = \sqrt{\tau}$: (a) h_{xxt}: $\sigma_x = 0.5$ degrees, $\sigma_t = 60$ ms, $c = 2$. (b) h_{xxtt}: $\sigma_x = 0.6$ degrees, $\sigma_t = 140$ ms, $c = 2$. (Horizontal dimension: space x. Vertical dimension: time t.) (The figures in the top row are reprinted with permission.)

predictions of receptive field profiles that are qualitatively very similar to *all* the linear spatial and spatio-temporal receptive field types presented in (DeAngelis et al. [4, 5]) and also to schematic simplifications of most of the spatio-chromatic receptive fields shown in (Conway and Livingstone [22]) and (Johnson et al. [23]).

### 4.1. Spatial and spatio-temporal receptive fields in the LGN

Neurophysiological studies by DeAngelis et al. [4, 5] and others report that most neurons in the lateral geniculate nucleus (LGN), have receptive fields (i) with approximately circular center-surround organization over image space (ii) and that they are separable over space-time. Furthermore, they are two main types of temporal responses: (i) for a "non-lagged cell" the first temporal lobe is the strongest one (Fig. 12(left)), whereas (ii) for "lagged cell" the second temporal lobe is strongest (Fig. 12(right)), see also Ghodrati et al. [53] for a more extensive overview of properties of LGN neurons

When using a time-causal temporal smoothing kernel, the first peak of a first-order temporal derivative will be strongest, whereas the second peak of a second-order temporal derivative will be strongest (see [21, Fig. 2]). Thus, according to this theory, non-lagged LGN cells can be seen as corresponding to first-order time-causal temporal derivatives, whereas lagged LGN cells can be seen as corresponding to second-order time-causal temporal derivatives.

Comparing to the proposed framework for idealized receptive fields, the spatial response of such a neuron is highly similar to a Laplacian of a Gaussian, which leads to a composed idealized receptive field model of the form [19, Equation (108)]
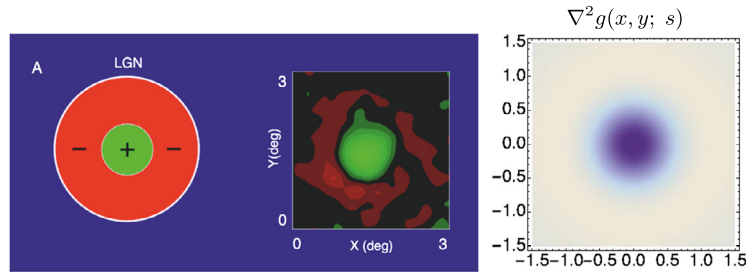
Fig. 13. *Spatial modelling of LGN neurons.* (left) DeAngelis et al. [4] report that LGN neurons have approximately circular center-surround responses over the spatial domain. (right) In terms of our idealized receptive field models, such a spatial dependency can be modelled by the Laplacian of the Gaussian $\nabla^2 g(x, y; s) = (x^2 + y^2 - 2s)/(2\pi s^3) \exp(-(x^2 + y^2)/2s)$, here with $\sigma_s = \sqrt{s} = 0.6$ in units of degrees of visual angle. (Left and middle figures reprinted with permission.)
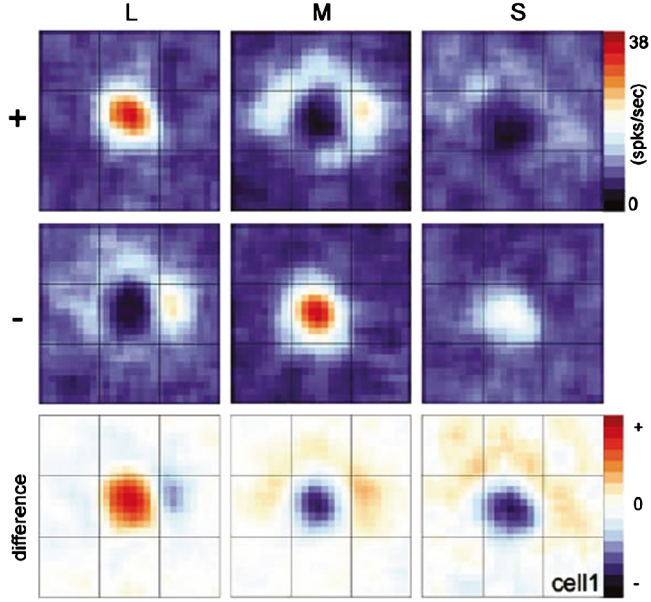

Fig. 14. *Receptive field responses of a spatio-chromatic double-opponent neuron* according to Conway and Livingstone [22, Fig. 2, Page 10831]. Here, the colour channels L, M and S basically correspond to red, green and blue colour channels, respectively, from which corresponding red/green and yellow/blue colour-opponent channels can be computed from the difference between L to M and the difference between L + M to S, respectively. © 2006 Society for Neuroscience with permission

$$h_{LGN}(x, y, t; \; s, \tau) = \pm(\partial_{xx} + \partial_{yy}) \, g(x, y; \; s) \, \partial_{t^n} \, h(t; \; \tau). \tag{39}$$

In Fig. 13, we show the result of modelling the spatial component of a receptive field in the LGN with a Laplacian of the Gaussian. Such a Laplacian of the Gaussian is also applicable for spatial modelling of on-center/off-surround and off-center/on-surround receptive fields in the retina. In Fig. 12, we show results of joint spatio-temporal modelling of space-time separable receptive fields in the LGN, with the temporal smoothing over the temporal domain expressed as a cascade of truncated exponential kernels of the form (27) and complemented by first-or second-order derivatives.

In previous work by (Rodieck [29]), differences of Gaussians have been shown to constitute a very good approximation of the spatial component of receptive fields in the retina and the LGN. The Laplacian of Gaussian model $(\partial_{xx} + \partial_{yy}) \, g(x, y; \; s)$ over the spatial domain is closely related to such differences of Gaussians. This relationship can be shown from fact that the rotationally symmetric Gaussian satisfies the isotropic diffusion equation [37]:

$$\frac{1}{2} \nabla^2 L(x, y; \; s) = \partial_s L(x, y; \; s)$$

$$\approx \frac{L(x, y; \; s + \Delta s) - L(x, y; \; s)}{\Delta s}$$
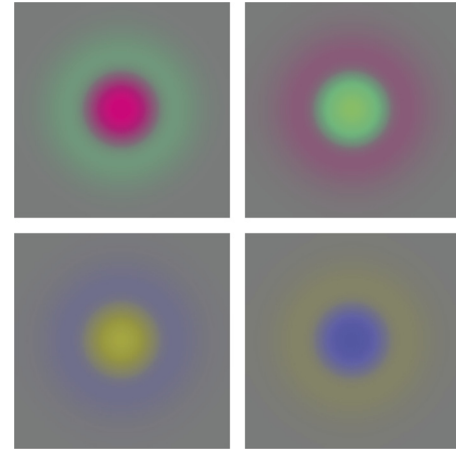
Fig. 15. *Modelling of double-opponent neurons using idealized spatio-chromatic receptive fields over the spatial domain.* Here, we have applied the spatial Laplacian operator to positive and negative red/green and yellow/blue colour opponent channels, respectively. These receptive fields can be seen as idealized models of the spatial component of double-opponent spatio-chromatic receptive fields in the LGN.

$$= \frac{DOG(x, y; \; s, \Delta s)}{\Delta s}. \tag{40}$$

This relationship means that differences of Gaussians approximate derivatives over scale, which in turn correspond to Laplacian responses. Conceptually, this implies very good agreement with the spatial component of the LGN model (39) based on Laplacians of Gaussians. In more recent work, Bonin et al. [60] has also found that LGN responses in cats can be well modelled by differences of Gaussians in combination with temporal smoothing, also complemented by a non-linear contrast gain control mechanism (which we do not model specifically here, although the logarithmic brightness scale considered in this treatment will handle variabilities in illumination that could also be handled by non-linear gain control).

4.2. Double-opponent spatio-chromatic receptive fields in the LGN

Conway and Livingstone [22] have presented a study of spatio-chromatic response properties of V1 neurons in the alert Macaque monkey. They report the finding of *double-opponent cells*, that simultaneously compute both spatial and chromatic opponency. These cells have receptive fields with approximately circular red/green and yellow/blue colour-opponent response properties, see Fig. 14, and which are claimed to constitute the first layer of spatially opponent colour computations.

If we in analogy with the previous modelling of rotationally symmetric on-center/off-surround and off-center/on-surround receptive fields in the LGN by Laplacian of Gaussians (39), apply the Laplacian of the Gaussian operator to red/green and yellow/blue colour-opponent channels,
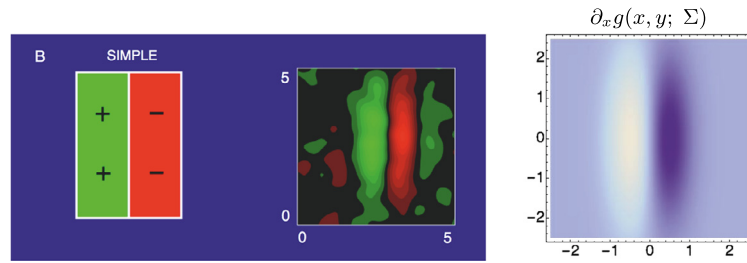
Fig. 16. Computational modelling of a receptive field profile over the spatial domain in the primary visual cortex (V1) as reported by DeAngelis et al. [4, 5] using affine Gaussian derivatives: (middle) Receptive field profile of a simple cell over image intensities as reconstructed from cell recordings, with positive weights represented as green and negative weights by red. (left) Stylized simplification of the receptive field shape. (right) Idealized model of the receptive field from a first-order directional derivative of an affine Gaussian kernel $\partial_x g(x, y; \Sigma) = \partial_x g(x, y; \lambda_x, \lambda_y)$ according to (21), here with $\sigma_x = \sqrt{\lambda_x} = 0.5$ and $\sigma_y = \sqrt{\lambda_y} = 1.5$ in units of degrees of visual angle, and with positive weights with respect to image intensities represented by white and negative values by violet. (Left and middle figures reprinted with permission.)
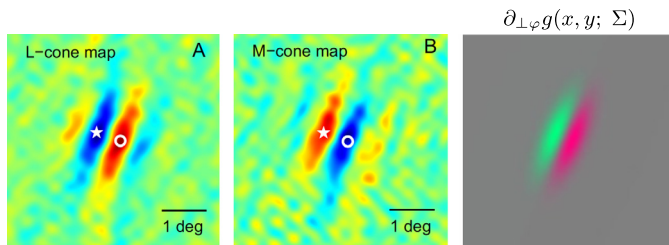
Fig. 17. Modelling of double-opponent simple cells in the primary visual cortex (V1) in terms of affine Gaussian derivatives over colour-opponent channels, based on neurophysiological cell recordings by Johnson et al. [23]: (left) Responses to L-cones corresponding to long wavelength red cones, with positive weights represented by red and negative weights by blue. (middle) Responses to M-cones corresponding to medium wavelength green cones, with positive weights represented by red and negative weights by blue. (right) Idealized model of the receptive field from a first-order directional derivative of an affine Gaussian kernel $\partial_{\perp \varphi} g(x, y; \Sigma)$ according to (21) over a red-green colour-opponent channel for $\sigma_1 = \sqrt{\lambda_1} = 0.6$ and $\sigma_2 = \sqrt{\lambda_2} = 0.2$ in units of degrees of visual angle, $\alpha = 67$ degrees and with positive weights for the red-green colour-opponent channel represented by red and negative values by green. (Left and middle figures: Copyright 2008 of Society for Neuroscience with permission.)

$$\begin{pmatrix} f \\ u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}, \tag{41}$$
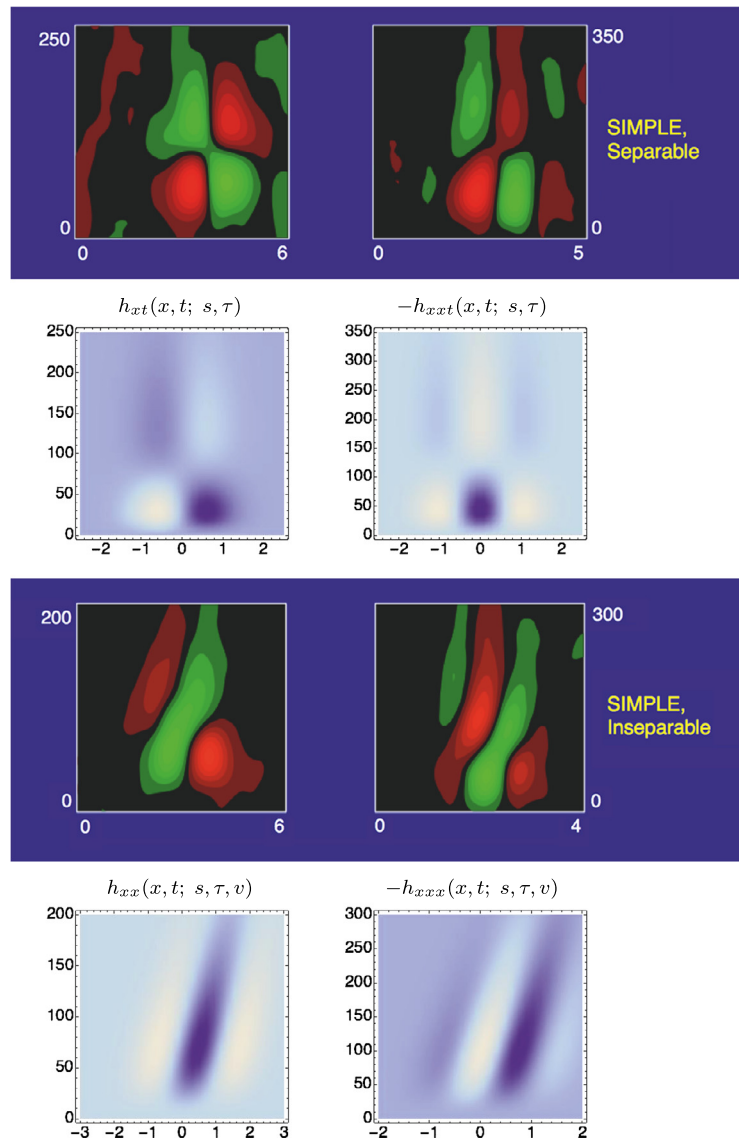
respectively, we get equivalent spatio-chromatic receptive fields that correspond to red-center/green-surround, green-center/red-surround, yellow-center/blue-surround or blue-center/yellow-surround, respectively, see Fig. 15. This corresponds to applying the following spatio-chromatic receptive field model to the RGB channels

$$h_{double-opponent}(x, y; s) = \pm (\partial_{xx} + \partial_{yy}) g(x, y; s) \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 \end{pmatrix}, \tag{42}$$

and which constitutes an idealized model for the spatio-chromatic response properties of double-opponent cells.

4.3. Spatial, spatio-chromatic and spatio-temporal receptive fields in V1

In their study of neurons in the primary visual cortex (V1), DeAngelis et al. [4, 5] report that the receptive fields of V1 neurons, in general, have different response properties compared to LGN neurons in the following ways: (i) they are oriented in the spatial domain and (ii) they are sensitive to specific ranges of stimulus velocities. According to the pioneering work by Hubel and Wiesel [1, 2, 3]), simple cells are additionally characterized by the following properties: (iii) they have precisely localized "on" and "off" subregions, (iv) spatial summation takes place over each subregion, (v) there is spatial antagonism be-

tween on- and off-subregions, and (vi) visual responses to stationary or moving spots can be predicted from the spatial subregions.

Fig. 16 shows an example of the spatial dependency of a simple cell, that can be well modelled by a first-order affine Gaussian derivative over image intensities. Fig. 17 shows corresponding results for a colour-opponent receptive field of a simple cell in V1, that can be modelled as a first-order affine Gaussian spatio-chromatic derivative over an R-G colour-opponent channel.

Biological support for using multiple affine receptive fields, over an expansion of the shapes of the affine covariance matrices Σ, can be obtained from neurophysiological measurements by Goris et al. [50], who show that there is a large variability in the orientation selectivity of simple and complex cells (see Fig. 19). With regard to the presented theoretical model for simple cells in Equation (23), possibly extended with a colour-opponent representation (41) for spatio-chromatic receptive fields, this means that we could think of all affine receptive fields, with their directional derivatives in preferred directions aligned to the eigendirections of the covariance matrix Σ, as being present at any position in the image domain (see Fig. 8). Such a variability makes it possible to perfectly match the first-order variability of receptive field responses under variations of the slant and tilt directions of a smooth surface patch.

In Fig. 18, we show spatio-temporal dependencies of a set of separable and inseparable simple cells in V1 that can be modelled using the general idealized model of spatio-temporal receptive fields in Equation (25), based on Gaussian derivatives over image space and temporal derivatives of a set of truncated exponential kernels coupled in cascade (27). The results in the upper part show space-time separable spatio-temporal receptive fields corresponding to zero image velocity $v = 0$, and corresponding to either first- or second-order spatial derivatives over image space in combination with first-order temporal derivatives over time. The results in the lower part show inseparable spatio-temporal receptive fields corresponding to non-zero image velocities and based on either second- or third-order spatial derivatives over image space.

To conclude, from these figures we can see that the qualitative shape of biological receptive fields, as recorded by neurophysiological measurements, can be quite well modelled by the proposed idealized receptive field models that result from the presented normative theory of visual receptive fields.

5. Relations to previous work

In earlier work, Young [30] has also proposed to model spatial visual receptive fields by Gaussian derivatives and shown that visual receptive fields in cats and monkeys can be well modelled by Gaussian derivatives up to order four. Young et al. [31, 32] have also proposed to model spatio-temporal receptive fields by Gaussian derivatives over the spatio-temporal domain. This corresponds to the non-causal purely

Fig. 18. Modelling of space-time separable and inseparable simple cells in the primary visual cortex (V1) based on neural cell recordings reported by DeAngelis et al. [4]. The idealized spatio-temporal receptive fields are of the form $T(x, t; s, \tau, v) = \partial_{x^m} \partial_{t^n} (g(x - vt; s) h(t; \tau))$ according to Equation (25), where $v = 0$ corresponds to space-time separable receptive fields and $v \neq 0$ to inseparable receptive fields. The temporal smoothing function $h(t; \tau)$ is modelled as a set of first-order integrators/truncated exponential kernels of the form (27) coupled in cascade and using a logarithmic distribution of the intermediate temporal scale levels. (upper left) Separable receptive fields corresponding to first-order derivatives with respect to space and time. (upper right) Separable receptive fields corresponding to second-order derivatives with respect to space and first-order derivatives with respect to time. (lower left) Inseparable velocity-adapted receptive fields corresponding to second-order derivatives over space. (lower right) Inseparable velocity-adapted receptive fields corresponding to third-order derivatives over space. Parameter values with $\sigma_x = \sqrt{s}$ and $\sigma_t = \sqrt{\tau}$: (a) h_{xt}: $\sigma_x = 0.6$ degrees, $\sigma_t = 80$ ms, $c = 2$. (b) h_{xxt}: $\sigma_x = 0.6$ degrees, $\sigma_t = 120$ ms, $c = 2$. (c) h_{xx}: $\sigma_x = 0.7$ degrees, $\sigma_t = 70$ ms, $v = 0.007$ degrees/ms, $c = 2$. (d) h_{xxx}: $\sigma_x = 0.5$ degrees, $\sigma_t = 100$ ms, $v = 0.004$ degrees/ms, $c = 2$. (Horizontal axis: Space x in degrees of visual angle. Vertical axis: Time t in ms.) (The figures in the top and third rows reprinted with permission.)

Gaussian spatio-temporal concept presented in this article, as well as in our closely related earlier work [61, 62]. Young does, however, use a different type of parameterization.

The normative theory for visual receptive fields presented in [13, 19, 20, 21] and here does first of all provide additional theoretical foundation for Young's spatial modelling work based on Koenderink and van Doorn's theory [8, 9]. It does additionally extend that model from regular (isotropic) Gaussian derivatives to affine Gaussian derivatives, and does also provide a conceptual extension to a time-causal spatio-temporal domain that takes into explicit account the fact that the future cannot be accessed. Furthermore, our model provides a better parameterization of the spatio-temporal receptive field model over a non-causal spatio-temporal domain based on the Gaussian spatio-temporal scale-space concept.

This model, or earlier versions of it, has in turn been exploited for modelling of biological receptive fields by Lowe [63], May and Georgeson [64], Hesse and Georgeson [65], Georgeson et al. [66], Wallis and Georgeson [67], Hansen and Neumann [68], Wang and Spratling [69], Mahmoodi [70, 71] and Pei et al. [72].
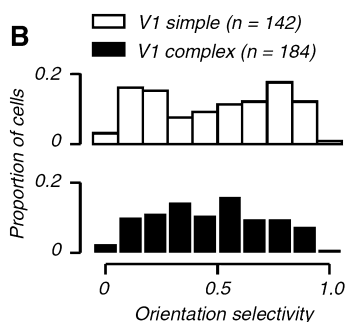
5.1. Relations to modelling by Gabor functions

Motivated by the property of Gabor functions [73]

$$G(x; s, \omega) = e^{-i\omega x} g(x; s), \tag{43}$$

that they minimize the uncertainty relation, Marčelja [24], Jones and Palmer [25, 26], Ringach [27, 28] and others have proposed to use Gabor functions to model spatial dependencies of visual receptive fields.
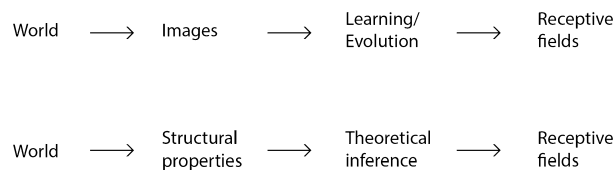
Fig. 19. Measurements of the orientation selectivity of simple cells and complex cells in the primary visual cortex of the Macaque monkey as reported by Goris et al. [50]. Interpreted with regard to the affine Gaussian derivative model for the receptive fields of simple cells (23), this large variability in orientation selectivity implies that we should consider covariance matrices Σ for a large range of eccentricities, as can be quantified by ratio between their eigenvalues λ_1 and λ_2. (The orientation selectivity of an affine Gaussian derivative kernel increases with the eccentricity.)

Fig. 20. Two structurally different ways of deriving receptive field shapes for a vision system intended to infer properties of the world by either biological or artificial visual perception. (top row) A traditional model for learning receptive fields shapes consists of collecting real-world image data from the environment, and then applying learning algorithms possibly in combination with evolution over multiple generations of the organism that the vision system is a part of. (bottom row) With the normative theory for receptive fields presented in this paper, a short-cut is made in the sense that the derivation of receptive field shapes starts from structural properties of the world (corresponding to symmetry properties in theoretical physics) from which receptive field shapes are constrained by theoretical mathematical inference.

There are, however, reasons to question this motivation, both on theoretical and empirical grounds: Following the arguments by Stork and Wilson [74]: (i) only the full complex-valued Gabor functions (treated as pairs) minimize the uncertainty relation, the single real or imaginary components do not, (ii) the real-valued functions that minimize the uncertainty relation are Gaussian kernels and Gaussian derivatives, not Gabor functions, (iii) quantitative comparisons between Gabor functions and other functions to physiological and psychophysical data have demonstrated that other functions, such as Gaussian derivatives, may enable better fits between the model and the data than for Gabor functions.

There are conceptual similarities between Gabor functions and Gaussian derivatives in terms of the ripples they have. For Gabor functions, the ripples are given by the zero-crossings of complex sine waves. For Gaussian derivatives, the ripples are given by the zero-crossings of Hermite functions of different orders. To specify a Gabor function, does, however, require two parameters; a scale parameter representing the spatial extent and a frequency. To specify a Gaussian derivative requires a scale parameter and the order of (spatial or temporal) differentiation, The Gaussian derivative model has the theoretically attractive properties that the receptive fields satisfy the diffusion equation and that derivatives of different orders can be mutually related by derivative operators, and can be computed from local nearest-neighbour operations over image space, implying that they can be implemented in biological wetware by connections between neighbouring computational units (neurons).

Regarding invariance properties to natural image transformations, it holds that the family of affine Gaussian kernels is closed under the full group affine image deformations. The family of Gabor functions based on multiplications of rotationally symmetric Gaussians with sine and cosine waves is not closed under general affine image deformations. This implies that we cannot compute truly affine invariant image representations from such families of traditional Gabor functions. If we have a pair of images that are related by a non-uniform scaling transformation, then the lack of affine covariance means that there will be systematic errors if we attempt to match image representations that are computed from such Gabor functions. If we compute receptive field responses based on directional derivatives of affine Gaussian kernels, it will on the other hand be possible to compute fully affine invariant features [20], in turn providing better internal consistency between receptive field responses computed from different views of objects in the world.

Concerning invariance to multiplicative illumination variations, it holds that the integral of the even cosine component of a Gabor function will, in general, not be equal to zero. This implies that the illumination invariant properties under local multiplicative illumination transforma-

tions or multiplicative exposure control mechanisms outlined in Section 3.4 will not hold for receptive field responses that are computed from such Gabor functions.

In these respects, the proposed Gaussian derivative model is conceptually simpler, the image measurements can be theoretically modelled using tools in differential geometry, it can be derived by necessity from symmetry principles in an axiomatic manner, its receptive field responses can be computed from local connections, and it enables provable invariance properties under local linearized image deformations (affine transformations) as well as to local multiplicative illumination variations and multiplicative exposure control mechanisms.

5.2. Relations to approaches for learning receptive fields from natural image statistics

A more data-driven approach to defining receptive field models that has been explored in the field consists of learning them from the statistics of natural image data (Field [75]; van der Schaaf and van Hateren [76]; Olshausen and Field [77]; Rao and Ballard [78]; Simoncelli and Olshausen [79]; Geisler [80]; Hyvärinen et al. [81]; Lörincz [82]). This approach also leads to visual receptive fields with similar shapes as those found in biological vision. The presented theory of visual receptive fields can in this context be seen as constituting a meta theory that describes the fundamental physical constraints under which different learning based method will operate. The physical structure of the world will determine what types of natural images can be generated. Field [75] and Doi and Lewicki [83] have described how "natural images are not random, instead they exhibit statistical regularities". If a learning system is assumed to be optimally adapted, we could expect that the receptive field shapes it learns should agree with the theoretical predictions, provided that the data sets for learning are sufficiently large and sufficiently representative with regard to the properties of natural image data (see Fig. 20).

The theory proposed in this paper can thus be interpreted as a theory at a higher level of abstraction, formulated based on basic principles that reflect properties of the environment, which in turn determine properties of the natural image data, and with no need for explicit statistical modelling of the image data. Specifically, the presented theory explains why statistical approaches for learning receptive fields can be expected to lead to qualitatively similar models for receptive fields as the idealized functional models of receptive fields.

From the observation that the receptive field profiles in the retina, the LGN and the primary visual cortex of higher mammals are very close to *ideal*, in the sense that the biologically measured receptive fields are very similar to the predictions from the idealized theory, we can regard biological vision as having adapted very well to the transformation properties of the outside world, specifically the transformations corresponding to the mapping of the three-dimensional world to two-

dimensional images. This property may be regarded as highly desirable for a biological organism, if there is or has been sufficient evolutionary pressure on its vision system.

5.3. Logarithmic brightness scale

Concerning the concept of a *logarithmic brightness scale*, already the Greek astronomer Hipparchus implicitly made use of this notion, when he defined a subjective scale for the brightness of stars. In his brightness scale, divided into six levels, the brightest stars were referred to as of the first magnitude, whereas the faintest stars, near to what can be perceived by a human observer without additional lenses, were said to be of the sixth magnitude. Later, when it became possible to accurately measure the physical intensities of stars, it was noted that Hipparchus subjective scale corresponded to logarithmic intensity values. Today, in astronomy, the *apparent brightness* of stars is still quantified in terms of logarithmic intensities, although over a much wider range of brightness. The retinex theory of early vision (Land [84, 85]) does also make use of a logarithmic brightness scale.

A logarithmic relationship between the perceived intensity and the physical magnitude of stimuli does more generally occur in the *Weber-Fechner law* in psychophysics. Consider a background intensity I that is subject to an increment threshold ΔI corresponding to a just noticeable difference. Then, the Weber-Fechner law states that the Weber ratio intensity I

$$\frac{\Delta I}{I} = k, \tag{44}$$

is constant over large ranges of magnitude variations [86, Pages 671–672]. The theoretical analysis in Section 3.4, regarding invariance properties of a logarithmic brightness scale under multiplicative intensity transformations and multiplicative exposure control mechanisms, is in excellent agreement with these psychophysical findings. If one considers an adaptive image exposure mechanism in the retina that adapts the diameter of the pupil and the sensitivity of the photopigments, such that relative range variability in the signal divided by the mean illumination is held constant (44) (see *e.g.* Peli [87]), then such an adaptation mechanism can be seen as implementing an approximation of the derivative of a logarithmic transformation

$$d(\log z) = \frac{dz}{z}. \tag{45}$$

This result is also closely related to information theoretic arguments by (Jaynes [88]) to use $\log z$ as a default parameterization of a strictly positive entity, in the absence of further information. Then, the ratio dz/z becomes a dimensionless integration measure.

The physical model in Section 3.4 provides a formal justification for transforming brightness values in a logarithmic way in connection to receptive field measurements, and how such a transformation relates to inherent physical properties of object surfaces in the environment.

6. Summary

From neurophysiological cell recordings we know that mammalian vision has developed receptive fields with characteristic properties: The first layers of visual receptive fields are tuned to different sizes and orientations over the spatial domain, and to different image velocities over joint space-time. In this article, we have presented an overview of a normative theory that shows how it is possible to derive such receptive field profiles *by necessity*, starting from a set of structural requirements of an *idealized vision system*, and whose functionality is determined by set of mathematical and physical assumptions (see Fig. 20).

These structural requirements reflect *structural properties of the world* for the receptive fields to be compatible with natural image transformations including: (i) variations in the sizes of objects in the world, (ii) variations in the viewing distance, (iii) variations in the viewing direction, (iv) variations in the relative motion between objects in the

world and the observer, (v) variations in the speed by which temporal events occur and (vi) local multiplicative illumination variations or multiplicative exposure control mechanisms.

We argue that it is natural for vision system, that is to *interact with the world* in a successful manner, to adapt to these structural requirements. If there is sufficient evolutionary pressure on an organism, in competition between different individuals of the same species or between individuals of different species, adaptation to the principles that determine structural properties of the environment may constitute an *evolutionary advantage*.

The proposed *normative theory* provides a way to derive *functional models of linear receptive fields* from first principles, leading to receptive field shapes in terms of affine Gaussian derivatives and closely related operators. Specifically, the presented theory can *explain* the different shapes of receptive field profiles that are found in biological vision from a requirement that the visual system should be able to compute covariant receptive field responses under the natural types of image transformations that occur in the environment, to enable the computation of invariant representations for perception at higher levels in the visual hierarchy [20] (see Appendix I in the supplement for a description about how covariant receptive fields at lower layers in the visual hierarchy enable invariances to geometric image transformations at higher levels in the visual hierarchy).

Such a view, that V1 performs an expansion of image data over the parameters of natural image transformations, is consistent with the substantial expansion of measurement data that is performed from the LGN[8] with about 1 M neurons and 1 M output channels to V1 with 190 M neurons and 37 M output channels [89, Fig. 3].

We have shown that the predictions from the presented theory are in good qualitative agreement with receptive fields found by neurophysiological cell recordings in mammalian vision. Specifically, we have presented idealized functional models (i) for space-time separable receptive fields in the retina and the LGN and (ii) for both space-time separable and non-separable simple cells in the primary visual cortex (V1).

The qualitatively very good agreement between the predicted receptive field profiles from the normative axiomatic theory with the receptive field profiles found by neurophysiological measurements indicates that the earliest receptive fields in higher mammal vision can be interpreted as having reached a state that can be seen as very close to *ideal* in view of the stated structural requirements/symmetry properties. From this viewpoint, mammalian vision can be interpreted as having adapted very well to the transformation properties of the outside world and to the transformations that occur when a three-dimensional world is projected onto a two-dimensional image domain.

In relation to other approaches of learning receptive field profiles from natural image statistics, the presented theory determines receptive field shapes without any need for training data. The presented theoretical approach also adds explanatory value in terms of underlying covariance and invariance properties, in the sense that requiring the first layers of receptive fields to be provably covariant under scaling transformations, rotations, perspective transformations and Galilean transformations makes it possible to define invariant properties with respect to these essential transformation groups at higher levels in the visual hierarchy. If the underlying first layers of visual receptive fields would not obey such covariance properties, then there would be a systematic bias in the visual operations, corresponding to the amount of mismatch between the backprojected receptive fields.

Corresponding types of arguments applied to the area of hearing, lead to the formulation of a normative theory of auditory receptive fields (Lindeberg and Friberg [90, 91]).

---

[8] In the retina, there are about 100 M photoreceptors with about 1 M output channels to the LGN.

Declarations

Author contribution statement

Tony Lindeberg: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This work was supported by the Swedish Research Council (Grant Numbers 2014-4083, 2018-03586).

Data availability statement

Data included in article/supplementary material/referenced in article.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2021.e05897.

References

[1] D.H. Hubel, T.N. Wiesel, Receptive fields of single neurones in the cat's striate cortex, J. Physiol. 147 (1959) 226–238.

[2] D.H. Hubel, T.N. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, J. Physiol. 160 (1962) 106–154.

[3] D.H. Hubel, T.N. Wiesel, Brain and Visual Perception: The Story of a 25-Year Collaboration, Oxford University Press, 2005.

[4] G.C. DeAngelis, I. Ohzawa, R.D. Freeman, Receptive field dynamics in the central visual pathways, Trends Neurosci. 18 (10) (1995) 451–457.

[5] G.C. DeAngelis, A. Anzai, A modern view of the classical receptive field: linear and non-linear spatio-temporal processing by V1 neurons, in: L.M. Chalupa, J.S. Werner (Eds.), The Visual Neurosciences, vol. 1, MIT Press, 2004, pp. 704–719.

[6] T. Iijima, Basic theory on normalization of pattern (in case of typical one-dimensional pattern), Bull. Electrotechn. Lab. 26 (1962) 368–388 (in Japanese).

[7] A.P. Witkin, Scale-space filtering, in: Proc. 8th Int. Joint Conf. Art. Intell, Karlsruhe, Germany, Aug. 1983, pp. 1019–1022.

[8] J.J. Koenderink, The structure of images, Biol. Cybern. 50 (1984) 363–370.

[9] J.J. Koenderink, A.J. van Doorn, Representation of local geometry in the visual system, Biol. Cybern. 55 (1987) 367–375.

[10] J.J. Koenderink, A.J. van Doorn, Generic neighborhood operators, IEEE Trans. Pattern Anal. Mach. Intell. 14 (6) (Jun. 1992) 597–605.

[11] T. Lindeberg, Scale-Space Theory in Computer Vision, Springer, 1993.

[12] T. Lindeberg, Scale-space theory: a basic tool for analysing structures at different scales, J. Appl. Stat. 21 (2) (1994) 225–270, also available from http://www.csc.kth.se/~tony/abstracts/Lin94-SI-abstract.html.

[13] T. Lindeberg, Generalized Gaussian scale-space axiomatics comprising linear scale-space, affine scale-space and spatio-temporal scale-space, J. Math. Imaging Vis. 40 (1) (2011) 36–81.

[14] T. Lindeberg, Generalized axiomatic scale-space theory, in: P. Hawkes (Ed.), Advances in Imaging and Electron Physics, vol. 178, Elsevier, 2013, pp. 1–96.

[15] L.M.J. Florack, Image Structure, Series in Mathematical Imaging and Vision, Springer, 1997.

[16] J. Sporring, M. Nielsen, L. Florack, P. Johansen (Eds.), Gaussian Scale-Space Theory: Proc. PhD School on Scale-Space Theory, Series in Mathematical Imaging and Vision, Springer, Copenhagen, Denmark, 1997.

[17] J. Weickert, S. Ishikawa, A. Imiya, Linear scale-space has first been proposed in Japan, J. Math. Imaging Vis. 10 (3) (1999) 237–252.

[18] B. ter Haar Romeny, Front-End Vision and Multi-Scale Image Analysis, Springer, 2003.

[19] T. Lindeberg, A computational theory of visual receptive fields, Biol. Cybern. 107 (6) (2013) 589–635.

[20] T. Lindeberg, Invariance of visual operations at the level of receptive fields, PLoS ONE 8 (7) (2013) e66990.

[21] T. Lindeberg, Time-causal and time-recursive spatio-temporal receptive fields, J. Math. Imaging Vis. 55 (1) (2016) 50–88.

[22] B.R. Conway, M.S. Livingstone, Spatial and temporal properties of cone signals in alert macaque primary visual cortex, J. Neurosci. 26 (42) (2006) 826–10 846.

[23] E.N. Johnson, M.J. Hawken, R. Shapley, The orientation selectivity of color-responsive neurons in Macaque V1, J. Neurosci. 28 (32) (2008) 8096–8106.

[24] S. Marcelja, Mathematical description of the responses of simple cortical cells, J. Opt. Soc. Am. 70 (11) (1980) 1297–1300.

[25] J. Jones, L. Palmer, The two-dimensional spatial structure of simple receptive fields in cat striate cortex, J. Neurophysiol. 58 (1987) 1187–1211.

[26] J. Jones, L. Palmer, An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex, J. Neurophysiol. 58 (1987) 1233–1258.

[27] D.L. Ringach, Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex, J. Neurophysiol. 88 (2002) 455–463.

[28] D.L. Ringach, Mapping receptive fields in primary visual cortex, J. Physiol. 558 (3) (2004) 717–728.

[29] R.W. Rodieck, Quantitative analysis of cat retinal ganglion cell response to visual stimuli, Vis. Res. 5 (11) (1965) 583–601.

[30] R.A. Young, The Gaussian derivative model for spatial vision: I. Retinal mechanisms, Spat. Vis. 2 (1987) 273–293.

[31] R.A. Young, R.M. Lesperance, W.W. Meyer, The Gaussian derivative model for spatio-temporal vision: I. Cortical model, Spat. Vis. 14 (3, 4) (2001) 261–319.

[32] R.A. Young, R.M. Lesperance, The Gaussian derivative model for spatio-temporal vision: II. Cortical data, Spat. Vis. 14 (3, 4) (2001) 321–389.

[33] A. Omurtag, B.W. Knight, L. Sirovich, On the simulation of large populations of neurons, J. Comput. Neurosci. 8 (2000) 51–63.

[34] M. Mattia, P.D. Giudice, Population dynamics of interacting spiking neurons, Phys. Rev. E 66 (5) (2002) 051917.

[35] O. Faugeras, J. Toubol, B. Cessac, A constructive mean-field analysis of multi-population neural networks with random synaptic weights and stochastic inputs, Front. Comput. Neurosci. 3 (1) (2009).

[36] T. Lindeberg, Feature detection with automatic scale selection, Int. J. Comput. Vis. 30 (2) (1998) 77–116.

[37] T. Lindeberg, Image matching using generalized scale-space interest points, J. Math. Imaging Vis. 52 (1) (2015) 3–36.

[38] T. Lindeberg, Spatio-temporal scale selection in video data, J. Math. Imaging Vis. 60 (4) (2018) 525–562.

[39] J.A. Movshon, E.D. Thompson, D.J. Tolhurst, Receptive field organization of complex cells in the cat's striate cortex, J. Physiol. 283 (1) (1978) 79–99.

[40] E. Adelson, J. Bergen, Spatiotemporal energy models for the perception of motion, J. Opt. Soc. Am. A 2 (1985) 284–299.

[41] R.C. Emerson, M.C. Citron, W.J. Vaughn, S.A. Klein, Nonlinear directionally selective subunits in complex cells of cat striate cortex, J. Neurophysiol. 58 (1) (1987) 33–65.

[42] D.J. Heeger, Normalization of cell responses in cat striate cortex, Vis. Neurosci. 9 (1992) 181–197.

[43] J. Touryan, B. Lau, Y. Dan, Isolation of relevant visual features from random stimuli for cortical complex cells, J. Neurosci. 22 (24) (2002) 811–10 818.

[44] T. Serre, M. Riesenhuber, Realistic modeling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex, Tech. Rep. AI Memo 2004-017, MIT Computer Science and Artifical Intelligence Laboratory, 2004.

[45] K.P. Kording, C. Kayser, W. Einhäuser, P. Konig, How are complex cell properties adapted to the statistics of natural stimuli?, J. Neurophysiol. 91 (1) (2004) 206–212.

[46] J. Touryan, G. Felsen, Y. Dan, Spatial structure of complex cell receptive fields measured with natural images, Neuron 45 (5) (2005) 781–791.

[47] N.C. Rust, O. Schwartz, J.A. Movshon, E.P. Simoncelli, Spatiotemporal elements of macaque V1 receptive fields, Neuron 46 (6) (2005) 945–956.

[48] M. Carandini, What simple and complex cells compute, J. Physiol. 577 (2) (2006) 463–466.

[49] M. Carandini, D.J. Heeger, Normalization as a canonical neural computation, Nat. Rev. Neurosci. 13 (2012) 51–62.

[50] R.L.T. Goris, E.P. Simoncelli, J.A. Movshon, Origin and function of tuning diversity in Macaque visual cortex, Neuron 88 (4) (2015) 819–831.

[51] T. Lindeberg, Provably scale-covariant continuous hierarchical networks based on scale-normalized differential expressions coupled in cascade, J. Math. Imaging Vis. 62 (1) (2020) 120–148.

[52] T. Gollisch, M. Meister, Eye smarter than scientists believed: neural computations in circuits of the retina, Neuron 65 (2) (2010) 150–164.

[53] M. Ghodrati, S.-M. Khaligh-Razavi, S.R. Lehky, Towards building a more complex view of the lateral geniculate nucleus: recent advances in understanding its role, Prog. Neurobiol. 156 (2017) 214–255.

[54] I.I. Hirschmann, D.V. Widder, The Convolution Transform, Princeton University Press, Princeton, New Jersey, 1955.

[55] T. Lindeberg, Temporal scale selection in time-causal scale space, J. Math. Imaging Vis. 58 (1) (2017) 57–101.

[56] T. Lindeberg, Scale-space for discrete signals, IEEE Trans. Pattern Anal. Mach. Intell. 12 (3) (Mar. 1990) 234–254.

[57] T. Lindeberg, D. Fagerström, Scale-space with causal time direction, in: Proc. European Conf. on Computer Vision (ECCV'96), in: LNCS, vol. 1064, Springer, Cambridge, UK, Apr. 1996, pp. 229–240.

[58] T. Lindeberg, Edge detection and ridge detection with automatic scale selection, Int. J. Comput. Vis. 30 (2) (1998) 117–154.

[59] T. Lindeberg, Scale selection, in: K. Ikeuchi (Ed.), Computer Vision, Springer, 2021.

[60] V. Bonin, V. Mante, M. Carandini, The suppressive field of neurons in the lateral geniculate nucleus, J. Neurosci. 25 (47) (2005) 844–10 856.

[61] T. Lindeberg, Linear spatio-temporal scale-space, in: Proc. International Conference on Scale-Space Theory in Computer Vision (Scale-Space'97), in: Springer LNCS, vol. 1252, Springer, 1997, pp. 113–127.

[62] T. Lindeberg, Linear spatio-temporal scale-space, Dept. Of Numerical Analysis and Computer Science, KTH, Tech. Rep. ISRN KTH/NA/P–01/22–SE, Nov. 2001, available from http://www.csc.kth.se/cvap/abstracts/cvap257.html.

[63] D.G. Lowe, Towards a computational model for object recognition in IT cortex, in: Biologically Motivated Computer Vision, in: Springer LNCS, vol. 1811, Springer, 2000, pp. 20–31.

[64] K.A. May, M.A. Georgeson, Blurred edges look faint, and faint edges look sharp: the effect of a gradient threshold in a multi-scale edge coding model, Vis. Res. 47 (13) (2007) 1705–1720.

[65] G.S. Hesse, M.A. Georgeson, Edges and bars: where do people see features in 1-D images?, Vis. Res. 45 (4) (2005) 507–525.

[66] M.A. Georgeson, K.A. May, T.C.A. Freeman, G.S. Hesse, From filters to features: scale-space analysis of edge and blur coding in human vision, J. Vis. 7 (13) (2007) 7.

[67] S.A. Wallis, M.A. Georgeson, Mach edges: local features predicted by 3rd derivative spatial filtering, Vis. Res. 49 (14) (2009) 1886–1893.

[68] T. Hansen, H. Neumann, A recurrent model of contour integration in primary visual cortex, J. Vis. 8 (8) (2008) 8.

[69] Q. Wang, M.W. Spratling, Contour detection in colour images using a neurophysiologically inspired model, Cogn. Comput. 8 (6) (2016) 1027–1035.

[70] S. Mahmoodi, Linear neural circuitry model for visual receptive fields, J. Math. Imaging Vis. 54 (2) (2016) 1–24.

[71] S. Mahmoodi, Nonlinearity in simple and complex cells in early biological visual systems, J. Math. Imaging Vis. (2017) 1–10.

[72] Z.-J. Pei, G.-X. Gao, B. Hao, Q.-L. Qiao, H.-J. Ai, A cascade model of information processing and encoding for retinal prosthesis, Neural Regen. Res. 11 (4) (2016) 646.

[73] D. Gabor, Theory of communication, J. IEE 93 (1946) 429–457.

[74] D.G. Stork, H.R. Wilson, Do Gabor functions provide appropriate descriptions of visual cortical receptive fields, J. Opt. Soc. Am. 7 (8) (1990) 1362–1373.

[75] D.J. Field, Relations between the statistics of natural images and the response properties of cortical cells, J. Opt. Soc. Am. 4 (1987) 2379–2394.

[76] A. van der Schaaf, J.H. van Hateren, Modelling the power spectra of natural images: statistics and information, Vis. Res. 36 (17) (1996) 2759–2770.

[77] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, J. Opt. Soc. Am. 381 (1996) 607–609.

[78] R.P.N. Rao, D.H. Ballard, Development of localized oriented receptive fields by learning a translation-invariant code for natural images, Comput. Neural Syst. 9 (2) (1998) 219–234.

[79] E.P. Simoncelli, B.A. Olshausen, Natural image statistics and neural representations, Annu. Rev. Neurosci. 24 (2001) 1193–1216.

[80] W.S. Geisler, Visual perception and the statistical properties of natural scenes, Annu. Rev. Psychol. 59 (2008) 10.1–10.26.

[81] A. Hyvärinen, J. Hurri, P.O. Hoyer, Natural Image Statistics: A Probabilistic Approach to Early Computational Vision, Computational Imaging and Vision, Springer, 2009.

[82] A. Lörincz, Z. Palotal, G. Szirtes, Efficient sparse coding in early sensory processing: Lessons from signal recovery, PLoS Comput. Biol. 8 (3) (2012) e1002372.

[83] E. Doi, M.S. Lewicki, Relations between the statistical regularities of natural images and the response properties of the early visual system, in: In Japanese Cognitive Science Society: Sig P & P, Kyoto University, 2005, pp. 1–8.

[84] E.H. Land, The retinex theory of vision, Proc. R. Inst. G. B. 57 (1974) 23–58.

[85] E.H. Land, Recent advances in retinex theory, Vis. Res. 26 (1) (1986) 7–21.

[86] S.E. Palmer, Vision Science: Photons to Phenomenology, MIT Press, 1999, first edition.

[87] E. Peli, Contrast in complex images, J. Opt.l Soc. Am. (JOSA A) 7 (10) (1990) 2032–2040.

[88] E.T. Jaynes, Prior probabilities, Trans. Syst. Sci. Cybern. 4 (3) (1968) 227–241.

[89] J.J. DiCarlo, D. Zoccolan, N.C. Rust, How does the brain solve visual object recognition? Neuron 73 (3) (2012) 415–434.

[90] T. Lindeberg, A. Friberg, Idealized computational models of auditory receptive fields, PLoS ONE 10 (3) (2015) e0 119 032:1–58.

[91] T. Lindeberg, A. Friberg, Scale-space theory for auditory signals, in: Proc. Scale Space and Variational Methods in Computer Vision (SSVM 2015), in: LNCS, vol. 9087, Springer, 2015, pp. 3–15.