Visual representations embodying spacetime structure

L. M. J. Florack

UU-CS-1999-07

Visual Representations Embodying Spacetime Structure

Luc Florack

Abstract

If one defines a visual front-end to be that part of a visual system in which the structure of the optical world is represented in some least committed form, then it is plausible to assert that it reflects the structural properties, notably symmetries, of the optical environment. We propose a methodology for constructing receptive field assemblies consistent with these symmetries, which contributes to our understanding of the visual front-end in a predictive and retrodictive way.

Keywords: Front-end vision, receptive fields, spacetime symmetries.

1 Introduction

If one defines a visual front-end to be the sensory stage of a visual system in which the structure of the optical world is represented in some least committed form, *i.e.* prior to semantical processing, then it is natural to subject it to the same constraints as those that govern the phenomenological environment, notably the structure of spacetime. In physics a common way to describe spacetime is in terms of the *symmetries* that underly the behaviour of spatiotemporal phenomena. In particular, classical spacetime can be described in terms of spatiotemporal shift invariance (homogeneity) and spatial rotation invariance (isotropy). One should add (isotropic) spatial and temporal scale invariance, which are instances of the universal law of scale invariance (the rationale behind dimensional analysis, *cf.* Olver [1]). An elegant way to combine the symmetry statements concerning "empty spacetime" with operationally definable concepts such as retinal irradiances, receptive fields and neural signals, is provided by the so-called "carry-along" principle. This is a precise recipe for expressing spacetime transformations in terms of actions on real objects, and captures the intuitive idea of a transformation as something carried out on things one can grasp rather than on physically void "points in spacetime".

Koenderink proposes a model of the front-end based on similar reasoning [2]. Central to this is a receptive field taxonomy [3], which is not merely a scheme for modelling receptive field profiles[1], but an ordering of various types in a hierarchical/heterarchical manner not unlike the periodic system of the elements used in physics and chemistry. Such a systematization is useful for both its predictive and retrodictive power (stipulating the existence of novel types

[1]Typical tolerances of single cell recordings in the visual system leave room for a variety of data models; decisions between these cannot be based on data evidence alone.

of receptive fields or interrelations as yet unknown, respectively adding structure to a hitherto unsorted body of empirical data).

The standard Gaussian, familiar for its role in scale-space theory [4, 5], constitutes the pivot of Koenderink's model. Ironically, there is hardly any empirical evidence for such a receptive field profile in mammalian vision, but its strength lies in the fact that certain basic operations carried out on it, such as differentiation, linear superposition, and particular types of affine transformations (*v.i.*), do generate templates that agree with physiological evidence [6, 7, 8, 9, 10, 11, 12, 13, 14]. Below a complete group of operations is deduced from the physics of classical spacetime consistent with available evidence.

2 Theory

Consider a tentative profile ϕ intended to model the shape of a putative receptive field disregarding attributes such as size, orientation, and base point. Let us call it a "template" for ease of reference. The idea is that if we "transform it in the right way" we obtain a function that may actually correspond to a receptive field profile found somewhere along the visual pathway. Similarly we introduce a scalar function ω representing a fiducial retinal irradiance distribution modulo such spacetime attributes. Of course this function could *a priori* have any form. Several fundamental questions arise, which will be addressed below:

1. How does a stimulus ω interact with a receptive field so as to produce a neural response?

2. What is a reasonable choice of template ϕ?

3. How should one transform it so as to become compatible with empirical findings?

In order to tackle these questions a simplifying assumption is made concerning the first, *viz.* that of *bilinearity*. That is to say, a neural response arises by linear filtering of ω using a suitably transformed ϕ as the correlating filter. This assumption is not as restrictive as might appear on first sight, since one is not compelled to identify ω with the physical photon flux impinging on the retina; any monotonic mapping will do. Typically one finds a logarithmic compression of photon flux at the retinal level. Accounting for this, linear signal production is appropriate within a limited range between threshold and saturation levels.

The assumption of bilinearity leads us into the realm of differential geometry, which deals with manifolds (like spacetime), vectors (receptive fields), covectors (local retinal irradiance patterns), *etc.* Although it is probably no harder to rebut any of the assertions made below than it is for bilinearity, it is interesting to put the visual front-end into the perspective of differential geometry and see what it leads to. (As the American statesman Samuel Rayburn once put it: "Any jackass can kick down a barn, but it takes a carpenter to build one.")

As for the second question, let us adopt the standard Gaussian as the basic template. Koenderink has motivated this choice by arguing that it is the only reasonable one in the context of a linear generalisation principle [4]: Only Gaussian filtering will prevent creation of spurious intensity levels that are above or below all those in the immediate neighbourhood of an extremal point ("spurious resolution"). Given the straticulate nature of the visual pathway, an entirely different but equally compelling demand is that of algebraic closure, stating that any concatenation of linear filterings by elements from a putative class should correspond to a single filtering by a member of the same class [15]. This is at least in agreement with findings that similar receptive field profiles show up at various locations along the visual pathway. The only *positive*, smooth function generating an autoconvolution algebra ("in-itself-closed") is, again, the normalised Gaussian. Anyway, for the procedure outlined below, pertaining to the third and final question, it does not so much matter what the template is, as long as it permits us to model actual receptive fields by simple transformations. The Gaussian indeed turns out to be a convenient template for this purpose, *cf.* the data modelling by Young [16, 17].

Let us turn to the details of the carry-along principle and investigate its implications for receptive field modelling. Fig. 1 illustrates the concept in terms of a commutative diagram. If we denote the neural signal produced by exposing a receptive field template ϕ to a reti-

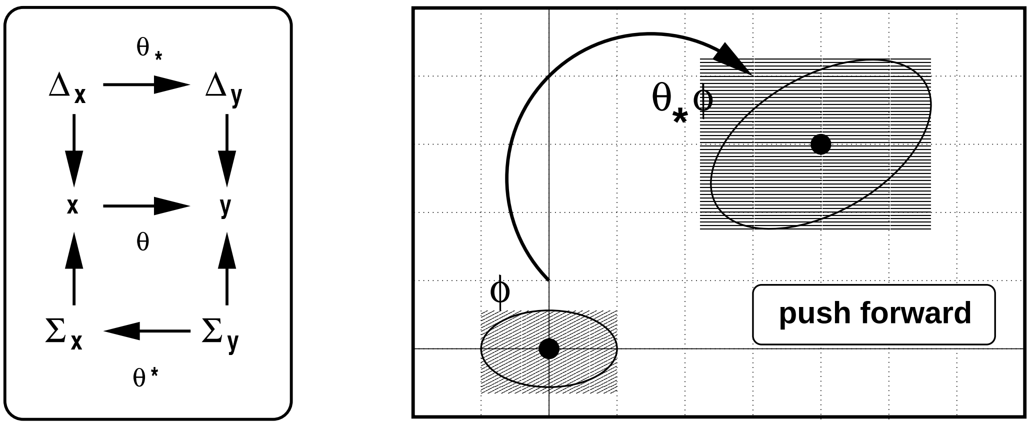

Figure 1: Left: The carry-along principle relates spacetime transformations θ to their corresponding actions on receptive field templates ϕ ("push forward") and retinal irradiance distributions ω ("pull back"). By construction one has $\omega[\theta_*\phi] = \theta^*\omega[\phi]$. Right: push forward of a receptive field template under an affine transformation.

nal irradiance distribution ω by $\omega[\phi]$, then the corresponding actions induced by a spacetime transformation θ on ϕ and ω are defined such that $\omega[\theta_*\phi] = \theta^*\omega[\phi]$. Here $\theta_*\phi$ denotes the θ-transformed template, and $\theta^*\omega$ the transformed retinal irradiance distribution equivalent to it. Thus any change of neural output can always be explained by a suitable change of either ω or ϕ ("duality"). Note that a label has been attached to the classes of receptive fields, Δ_x, and retinal irradiances, Σ_x, indicating the spacetime attribute x of the respective objects being transformed. From the directions of the arrows one appreciates why mathematicians are wont to call $\theta_*\phi$ and $\theta^*\omega$ the "push forward" and "pull back" of ϕ, respectively ω, by the mapping θ. By definition one takes $\theta^*\omega = \omega \circ \theta$, *i.e.* the original retinal irradiance function evaluated in θ-transformed points (a "warping"; physicists call this "scalar transformation"). By duality one then finds that $\theta_*\phi = |\det \nabla\theta^{\text{inv}}|\phi \circ \theta^{\text{inv}}$, *i.e.* the original template evaluated in inversely transformed points, with a Jacobian determinant that preserves normalisation ("filter tuning"; in the jargon of physics "density transformation"). It is understood that the attributes of the various objects and their respective transforms are θ-related as well, recall the labels in Fig. 1.

Since the nature of classical spacetime can be described in terms of a transformation group[2] [18], it is interesting to see what kind of profiles are obtained if we subject the Gaussian template to it. The collection of $\theta_* \phi$ for all θ in this group then reflects the very structure of the spacetime manifold. If the transformation is non-infinitesimal and can be described by a number of (Lie) parameters, then typically (but not necessarily) the result will also depend on these parameters. However, infinitesimal transformations are also of interest; *e.g.* an infinitesimal displacement induces an action which corresponds to operationally well-defined and well-posed differentiation in the sense of distribution theory [19]. In this way the standard Gaussian readily produces the Gaussian family of derivative profiles [3]: Fig 2.

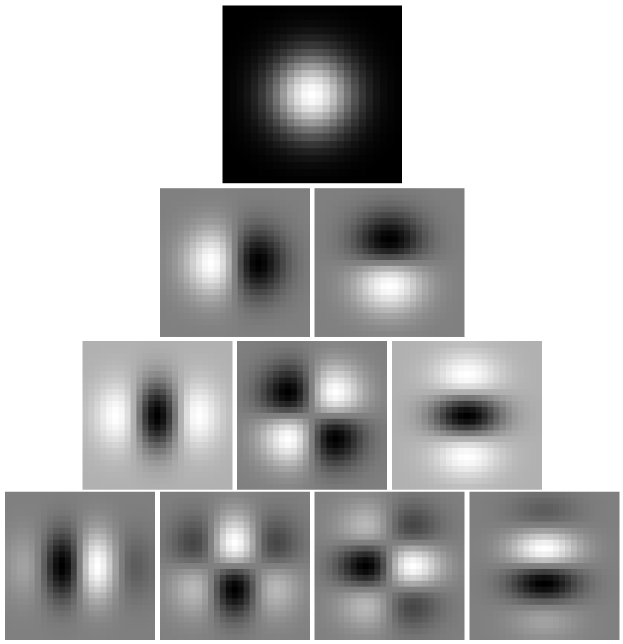

Figure 2: Cartesian representation of the Gaussian family up to third order. This scheme can be continued to any order, yielding increasingly complex shapes with one additional zero-crossing per order.

The Gaussian family constitutes a complete basis, which allows us, at least in principle, to model arbitrary profiles. However, its strength lies of course in the potential, to be verified or refuted by empirics, that one will need only *simple* combinations. (Otherwise any function basis would trivially be a perfect model!) Indeed, single cell recordings in mammalian visual front-ends reveal an abundance of spatial profiles like those in Fig. 2 up to order four or so, or linear combinations of these. Familiar are “centre-surround” profiles (much like Laplacians of a Gaussian) associated with ganglion cells, “edge” and “bar detectors” (directional derivatives), “end-stopped” and “grating detectors” (mixed derivatives), *etc.*

Although the carry-along recipe by infinitesimal transformations applied to the zeroth order Gaussian “explains” the entire Gaussian family, it remains to be shown how exactly empirical receptive field recordings relate to the templates of this family. For this we need to consider non-infinitesimal transformations. In practice one proceeds by picking an appropri-

[2]In fact, a so-called Lie group.

ate template, moving it to the retinal location of interest, orienting it in the proper way (if not isotropic), and scaling up its spatial extent and amplitude to the right size. But recall that this is just the push forward recipe applied to the appropriate template for the *spatial* part of the classical spacetime symmetry group! In addition this recipe explains why and how amplitude and spatial extent are confounded the way they are, why there must be a dense overlap of receptive fields, and why they occur replicated at multiple scales. It also explains their somatotopic organization, *e.g.* in the cortical columns discovered by the Nobel laureates Hubel and Wiesel.

It seems easy to extend the procedure to the time domain: Simply include time shifts (static model) and possibly Galilean boosts (kinematic model). The latter type of spacetime is one in which there is no notion of absolute velocity; systems behave alike if they move with constant relative speed. The static model introduces one parameter in the form of a temporal base point, *i.e.* the moment at—or rather, around—which a receptive field is most active, as well as one time scale parameter. It is a specific instance of the kinematic model, which brings in an additional parameter, *viz.* the tuning velocity of the receptive field[3] [20].

However, if we proceed along this line of reasoning too naively we are in for an unpleasant surprise. First of all a time shift applied to any member of the Gaussian family will never yield a profile that is completely within the causal part of the time axis, although larger shifts towards the past (given fixed temporal scale) will of course alleviate the degree of causality violation. Even worse, the profiles obtained simply fail to resemble the ones actually found. Apart from causality, one typically observes a peculiar time-skew; the profiles appear increasingly stretched towards the remote past. If one insists on the use of Gaussian templates, which are both of infinite support as well as of definite parity (symmetrical or anti-symmetrical), one wonders how one could ever "save the phenomena".

But it turns out that not only can one do so, the solution is a simple and rather elegant one. It was pointed out by Koenderink in his causal theory of "scale-time" for the zeroth order case [21], is in perfect agreement with an empirically tested universal scaling phenomenon known as Benford's law [22], and preserves all the symmetries of classical spacetime [15]. The "trick" is to map time logarithmically, such that the future part of the time axis is discarded from the outset and the present moment becomes a singularity. As the reader may verify, push forward of the 1D Gaussian family yields temporal profiles which depend on the present moment in addition to a characteristic scale and delay parameter. Varying scale and delay will affect the amount of skew, but the profiles are always manifestly causal. Spatiotemporal receptive fields arise by inclusion of the spatial domain in the usual way. See Fig. 3. Profiles like those of Fig. 3 are empirically confirmed, *cf.* the photoreceptor currents reported by Baylor [6] and the receptive field studies by DeAngelis *et al.* [9]. The Galilean-boosted profiles are similar to rotated copies of their static counterparts, and have indeed been modelled in this way, see *e.g.* Adelsen and Bergen's procedure [23]. However, unlike boosts, rotations in spacetime make no sense in physics; space and time "do not mix" so to speak. For this reason, and for its theoretical underpinning of time-skew and other fundamental spacetime phenomena, the carry-along principle outlined in this work is to be preferred.

[3] In the literature "velocity sensitivity" is sometimes attributed to profiles which are asymmetrical relative to a worldline aligned parallel to the time axis. This would qualify many nontrivial spatiotemoparal receptive fields as velocity sensitive, *e.g.* all odd time derivatives.

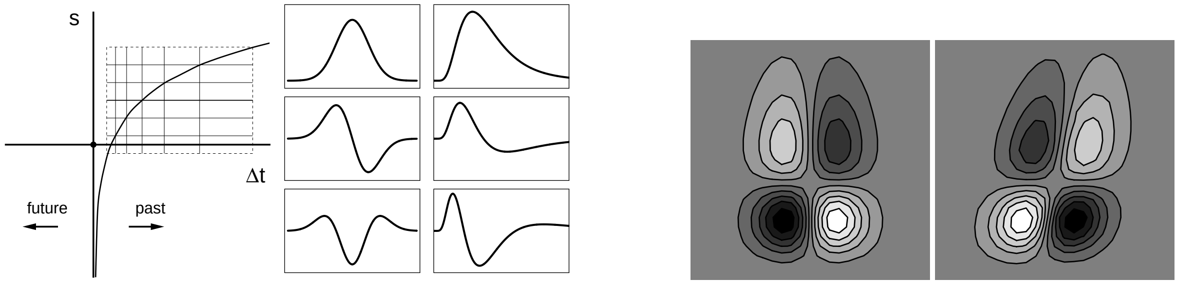

Figure 3: Left: Logarithmic time mapping and the induced push forward on 0th, 1st and 2nd order Gaussian derivative templates. Right: Push forward of mixed spacetime 2nd order Gaussian derivative template. Time varies along the vertical axis; the bottom line corresponds to the present moment. One spatial dimension has been integrated out, the other is presented horizontally. The left profile pertains to the static model, the right one to the kinematic model incorporating a Galilean boost of finite velocity.

3 Conclusion

The carry-along principle provides a rigorous mathematical recipe for adapting template receptive field profiles to apparent symmetries of the environment. Applied to the classical spacetime symmetry group and Gaussian derivative templates it yields profiles that are confirmed by a body of cell recordings in mammalian visual front-ends, puts these into a hierarchical/heterarchical order, and has the potential of predicting novel types of receptive fields and their relations to existing ones.

References

[1] P. J. Olver, *Applications of Lie Groups to Differential Equations*, vol. 107 of *Graduate Texts in Mathematics*. Springer-Verlag, 1986.

[2] J. J. Koenderink, "The brain a geometry engine," *Psychological Research*, vol. 52, pp. 122–127, 1990.

[3] J. J. Koenderink and A. J. van Doorn, "Receptive field families," *Biological Cybernetics*, vol. 63, pp. 291–298, 1990.

[4] J. J. Koenderink, "The structure of images," *Biological Cybernetics*, vol. 50, pp. 363–370, 1984.

[5] A. P. Witkin, "Scale-space filtering," in *Proceedings of the International Joint Conference on Artificial Intelligence*, (Karlsruhe, Germany), pp. 1019–1022, 1983.

[6] D. A. Baylor, "Photoreceptor signals and vision," *Investigative Ophtalmology and Visual Science*, vol. 28, pp. 34–49, January 1987. Proctor Lecture.

[7] D. Cai, G. C. DeAngelis, and R. D. Freeman, "Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens," *Journal of Neurophysiology*, vol. 78, pp. 1045–1061, August 1997.

[8] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman, “Depth is encoded in the visual cortex by a specialised receptive field structure,” *Nature*, vol. 352, pp. 156–159, July 1991.

[9] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman, “Receptive field dynamics in the central visual pathways,” *Trends in Neuroscience*, vol. 18, no. 10, pp. 451–458, 1995.

[10] D. H. Hubel and T. N. Wiesel, “Receptive fields of single neurons in the cat’s striate cortex,” *Journal of Physiology*, vol. 148, pp. 574–591, 1959.

[11] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” *Journal of Physiology*, vol. 160, pp. 106–154, 1962.

[12] D. H. Hubel and T. N. Wiesel, “Sequence regularity and geometry of orientation columns in the monkey striate cortex,” *Journal of Computational Neurology*, vol. 158, pp. 267–294, 1974.

[13] I. Ohzawa, G. C. DeAngelis, and R. D. Freeman, “Reply to TINS letter to the editor by Wang et al.,” *Trends in Neuroscience*, vol. 19, pp. 385–386, 1996.

[14] D. L. Ringach, G. Sapiro, and R. Shapley, “A subspace reverse correlation technique for the study of visual neurons,” *Vision Research*, vol. 37, no. 17, pp. 2455–2464, 1997.

[15] L. M. J. Florack, *Image Structure*, vol. 10 of *Computational Imaging and Vision Series*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1997.

[16] R. A. Young, “The Gaussian derivative model for machine vision: Visual cortex simulation,” *Journal of the Optical Society of America*, July 1986.

[17] R. A. Young, “The Gaussian derivative model for machine vision: I. retinal mechanisms,” *Spatial Vision*, vol. 2, no. 4, pp. 273–293, 1987.

[18] M. Friedman, *Foundations of Space-Time Theories: Relativistic Physics and Philosophy of Science*. Princeton: Princeton University Press, 1983.

[19] L. Schwartz, *Théorie des Distributions*. Publications de l’Institut Mathématique de l’Université de Strasbourg, Paris: Hermann, second ed., 1966.

[20] L. M. J. Florack, B. M. t. Haar Romeny, J. J. Koenderink, and M. A. Viergever, “Families of tuned scale-space kernels,” in *Proceedings of the Second European Conference on Computer Vision (Santa Margherita Ligure, Italy, May 1992)* (G. Sandini, ed.), vol. 588 of *Lecture Notes in Computer Science*, (Berlin), pp. 19–23, Springer-Verlag, 1992.

[21] J. J. Koenderink, “Scale-time,” *Biological Cybernetics*, vol. 58, pp. 159–162, 1988.

[22] B. Buck, A. C. Merchant, and S. M. Perez, “An illustration of Benford’s first digit law using alpha decay half lives,” *European Journal of Physics*, vol. 14, pp. 59–63, 1993.

[23] E. H. Adelson and J. R. Bergen, “Spatiotemporal energy models for the perception of motion,” *Journal of the Optical Society of America-A*, vol. 2, pp. 284–299, February 1985.