A geometric framework for nonlinear visual coding

Erhardt Barth* and Andrew B. Watson

NASA Ames Research Center, Moffett Field, CA 94035-1000, http://vision.arc.nasa.gov/

*Present, corresponding address: Institute for Signal Processing, University of Lübeck, Ratzeburger Allee 160, D-23538 Lübeck, Germany. http://www.isip.mu-luebeck.de/
barth@isip.mu-luebeck.de

Abstract: It is argued that important aspects of early and middle level visual coding may be understood as resulting from basic geometric processing of the visual input. The input is treated as a hypersurface defined by image intensity as a function of two spatial coordinates and time. Analytical results show how the Riemann curvature tensor **R** of this hypersurface represents speed and direction of motion. Moreover, the results can predict the selectivity of MT neurons for multiple motions and for motion in a direction along the optimal spatial orientation. Finally, a model based on integrated **R** components predicts global-motion percepts related to the barber-pole illusion.

References and links

1. H. R. Wilson and J. R. Bergen, "A four mechanisms model for threshold spatial vision," Vision Research **19** , 19-33 (1979).
2. A. B. Watson, "Detection and recognition of simple spatial forms," in *Physical and biological processing of images*, O. J. Braddick and A. C. Sleigh, eds. (Springer-Verlag, Berlin, 1983).
3. A. B. Watson and A. J. Ahumada, Jr., "Model of human visual-motion sensing," J. Opt. Soc. Am A **2**, 322-342 (1985).
4. E. H. Adelson and J. R. Bergen, "The Plenoptic Function and the Elements of Early Vision," in *Computational Models of Visual Processing*, M. Landy and J. A. Movshon, eds. (MIT Press, Cambridge, MA, 1991).
5. C. Zetzsche and E. Barth, "Fundamental limits of linear filters in the visual processing of two-dimensional signals," Vision Research **30**, 1111-1117 (1990).
6. C. Zetzsche, E. Barth, and B. Wegmann, "The importance of intrinsically two-dimensional image features in biological vision and picture coding," in *Digital images and human vision*, A. B. Watson, ed. (MIT Press, Cambridge, MA, 1993).
7. R. M. Haralick, L. T. Watson, and T. J. Laffey, "The topographic primal sketch," International J. of Robotic Research **2**, 50-72 (1983).
8. P. J. Besl and R. C. Jain, "Segmentation through variable-order surface fitting," IEEE Trans. Pattern Anal. Mach. Intell. **10**, 167-192 (1988).
9. J. Shi and C. Tomasi, "Good features to track," Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 593-600 (1994).
10. J. J. Koenderink and A. J. v. Doorn, "Representation of local geometry in the visual system," Biol. Cybern. **55**, 367-375 (1987).
11. D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," J. Physiol. **195**, 215-243 (1968).
12. J. Y. Lettvin, H. R. Maturana, W. S. McCulloch, and W. H. Pitts, "What the frog's eye tells the frog's brain," Proceedings IRE **47**, 1940-1951 (1959).
13. G. A. Orban, *Neuronal operations in the visual cortex*, (Springer, Heidelberg, 1984).
14. E. Peterhans and R. von der Heydt, "Functional organization of area V2 in the alert macaque," European Journal of Neuroscience **5**, 509-24 (1993).
15. J. B. Levitt, D. C. Kiper, and J. A. Movshon, "Receptive fields and functional architecture of macaque V2," J Neurophysiol **71**, 2517-42 (1994).
16. C. Yu and D. M. Levi, "End stopping and length tuning in psychophysical spatial filters," J. Opt. Soc. Am. A **14**, 2346-54 (1997).
17. A. Dobbins, S. W. Zucker, and M. S. Cynader, "Endstopping and curvature," Vision Res **29**, 1371-87 (1989).
18. F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kubler, "Simulation of neural contour mechanisms: from simple to end-stopped cells," Vision Res **32**, 963-81 (1992).

19. H. R. Wilson and W. A. Richards, "Mechanisms of contour curvature discrimination," J. Opt. Soc. Am. A **6**, 106-115 (1989).
20. S. P. Liou and R. C. Jain, "Motion detection in spatio-temporal space," Computer Vision, Graphics, and Image Processing **45**, 227-250 (1989).
21. C. Zetzsche and E. Barth, "Direct detection of flow discontinuities by 3D-curvature operators," Pattern Recognition Letters **12**, 771--779 (1991).
22. C. Zetzsche, E. Barth, and J. Berkmann, "Spatio-temporal curvature measures for flow field analysis," Geometric Methods in Computer Vision, B. Vemuri Ed. **SPIE 1590**, 337--350 (1991).
23. M. P. Do Carmo, *Riemannian Geometry*, (Birkhäuser, Boston, 1992).
24. S. Weinberg, *Gravitation and Cosmology*, (Wiley and Sons, New York, 1972).
25. B. Schutz, *A first course in general relativity*, (Cambridge University Press, Cambridge, 1985).
26. E. Barth, T. Caelli, and C. Zetzsche, "Image encoding, labelling and reconstruction from differential geometry," CVGIP:GRAPHICAL MODELS AND IMAGE PROCESSING **55**, 428--446 (1993).
27. C. Mota and J. Gomes, "Curvature Operators in Geometric Image Processing," presented at Brasilian Symposium On Computer Graphics and Image Processing, (Campinas, Brazil, 1999).
28. E. Barth, C. Zetzsche, and G. Krieger, "Curvature measures in visual information processing," Open Systems and Information Dynamics **5**, 25-39 (1998).
29. E. Barth, *Riemann-tensor motion analysis*, (2000), http://www.visionscience.com/vsDemos.html.
30. O. Tretiak and L. Pastor, "Velocity estimation from image sequences with second order differential operators," presented at Proc. 7th Int. Conf. Pattern Recognition, (Montreal, Canada, 1984).
31. T. S. Huang and A. N. Netravali, "Motion and structure from feature correspondence: a review," Proceedings of the IEEE **82**, 252-268 (1994).
32. E. Barth, "Spatio-temporal curvature and the visual coding of motion," in *Neural Computation (NC'2000)*, vol. 1404-093, H. Bothe and R. Rojas, eds. (ICSC Academic Press, Berlin, 2000).
33. H. Haußecker and H. Spies, "Motion," in *Handbook of Computer Vision and Applications*, B. Jahne, H. Haußecker, and P. Geissler, eds., 1999).
34. S. J. Nowlan and T. J. Sejnowski, "A selection model for motion processing in area MT of primates," J Neurosci **15**, 1195-214 (1995).
35. S. Wuerger, R. Shapley, and N. Rubin, ""On the visually perceived direction of motion" by Hans Wallach: 60 years later," Perception **25**, 1317-1367 (1996).
36. F. L. Kooi, "Local direction of edge motion causes and abolishes the barberpole illusion," Vision Res **33**, 2347-51 (1993).
37. E. Barth, C. Zetzsche, and I. Rentschler, "Intrinsic two-dimensional features as textons," J Opt Soc Am A Opt Image Sci Vis **15**, 1723-32 (1998).
38. T. D. Albright, "Direction and orientation selectivity of neurons in visual area MT of the macaque," J Neurophysiol **52**, 1106-30 (1984).
39. G. H. Recanzone, R. H. Wurtz, and U. Schwarz, "Responses of MT and MST neurons to one and two moving objects in the receptive field," J Neurophysiol **78**, 2904-15 (1997).
40. E. P. Simoncelli and D. J. Heeger, "A model of neuronal responses in visual area MT," Vision Res **38**, 743-61 (1998).
41. E. Barth and A. B. Watson, "Nonlinear spatio-temporal model based on the geometry of the visual input," Investigative Ophthalmology and Visual Science **39**, S2110 (1998).

Introduction

Rigorous mathematical models of visual processing have been used successfully to predict a variety of psychophysical results, e.g. [1-4]. Furthermore, model components often mimic properties of neurons in the visual pathway, which are likely to be band-pass and may be oriented. These models are essentially linear and thus linear models can account for important aspects of visual processing. However, the amount of evidence against linear behavior of subjects and neurons is significant [5, 6]. In this paper we define specific nonlinearities that are useful from a geometrical, but also information-theoretical, perspective.

For still images, geometric approaches have been used in the computer vision literature to define a first stage of object recognition systems, e.g. [7, 8]. More specific problems involve corner detection and feature extraction for motion estimation, e.g. [9]. In human vision research, few authors have looked at "the brain as a geometric engine" [10]. A central geometric concept is that of curvature, which becomes important in the context of vision because it is related to the property of endstopping. The term endstopping has been introduced in [11] to describe a decrease in neural activity with an increase in the length of a line pattern. But the basic property of endstopped neurons is that the response to straight, elongated patterns is significantly reduced relative to the response to short or curved patterns, line ends, corners,

and other non-straight stimuli. Type II ganglion cells in the frog exhibit such selectivity and have been described earlier [12]. Meanwhile it seems likely that the majority of neurons in areas V2 (and V1) are endstopped to different degrees [13-15]. A number of perceptual phenomena that can be related to endstopping have been reported (see [6] for a brief review) including some more recent results on psychophysical endstopping [16]. Models for endstopping and related phenomena, however, need not be based on geometry. Different, more or less heuristic, nonlinearities have been proposed [17-19]. However, any model which is insensitive to straight patterns must be nonlinear, an assertion that is easily proven by considering elementary properties of linear systems. Zetzsche and Barth have attempted to characterize the general class of nonlinear systems that suppress straight patterns [5, 6]. A central concept of their framework is that of intrinsic dimensionality. Accordingly, a pixel (x,y) can be classified as (i) intrinsically 0D (i0D) (ii) intrinsically 1D (i1D), or (iii) intrinsically 2D (i2D) depending on the two-dimensional image-intensity function f(x,y) in the neighborhood of that pixel, which can (i) be constant in all directions (ii) be constant in one direction, or (iii) vary in all directions (it is straightforward to extend the *inD* nomenclature to higher dimensions). The task then becomes to define the properties of nonlinear systems selective to i2D features (i2D operators). The approach first laid out in [5] is based on differential geometry and filter theory and identifies the basic nonlinearity as an AND-type combination of linear oriented filters that need to satisfy a compensation equation.

For the analysis of image sequences the available geometric tools are generally restricted to the geometry of objects and their motions, the geometry of the eye or camera and of the projection. We should stress here that in this paper we deal with the geometry of the visual input itself, i.e., the geometry of the hypersurface associated with image intensity f(x,y,t). The extrinsic geometry of this hypersurface has been used for the purpose of motion detection in computer vision [20]. We will relate (intrinsic) geometric properties of the hypersurface, in particular the Riemann tensor, to the problems of motion detection and the estimation of optic flow. Previously it has been argued that an operator based on the Gaussian curvature of the hypersurface can be used to detect flow-field discontinuities [21] and that computations possibly involved in flow-field estimation are related to endstopping [22].

Curvature as deviation from flatness

In geometry there are two fundamentally different types of surface properties: intrinsic properties depend only on the metric, i.e., the distances between points on the surface, whereas extrinsic properties depend on the way the points are placed in the embedding space. The common geometric thinking is extrinsic because important properties of the world around us are extrinsic (especially as sensed by touch). Nevertheless, intrinsic properties are relevant in everyday life. Consider, for example, cloth. The shape of a shirt is partially extrinsic as determined by the person wearing it but it also has shape as defined by the way the designer has cut it. The latter type of shape is intrinsic and is preserved as the shirt changes from being worn to being stored on a hook or crumpled in a suitcase. Before being tailored to be a shirt the fabric is (in general) flat, i.e., it can be fitted to a plane. Such surfaces are called developable. Deviations from flatness are measured by curvature and are important because they define the intrinsic shape of a surface independent of the actual embedding. In differential geometry, deviations from flatness are measured by the Riemann tensor. Beyond the intuition given above, textbooks on differential geometry, e.g. [23], and mathematical physics [24, 25] provide a clear account for the fact that the Riemann tensor is the most important property of a surface since it fixes the structure of the manifold.

Curvature and redundancy

Because i0D signals are constant and i1D signals are fully specified by a one-dimensional function they are both redundant. This simple fact, however, would not be sufficient for a coding strategy to be efficient if based on a distinction between i0D-, i1D-, and i2D features. For example, no benefits would be expected if the images to be coded would be generated from a white noise source. But analysis of natural-image statistics reveals that the probability

of i1D features to occur is lower than that for i0D features, and that the probability for i2D features is lowest. Therefore a visual representation will be more efficient if it distinguishes different intrinsic dimensions [6].

The geometric view of the redundancy of i0D- and i1D features is based on the observation that such features correspond to developable surfaces. It is known that the curvature considerably constrains possible embeddings, but it is also known that, in general, it cannot completely specify a surface. It is less clear, however, under which additional constraints it could. The view that developable surfaces are redundant and curved features are the most significant has been developed in [26] and it has been shown that images can be approximately reconstructed from only i2D features. More recently, a mathematical proof of the uniqueness of curved image regions has been given [27].

Geometry of movie hypersurfaces

The hypersurface associated with an image sequence is defined by

$$(x, y, t, f(x, y, t)) \tag{1}$$

i.e., image intensity f at position (x,y) and time t defines the set of points which belong to the hypersurface. The geometry of this movie hypersurface is harder to visualize and differs considerably from the geometry of an image surface. In geometry there is less interest in the specialties of the 3D case since surfaces are 2D and space-time is 4D. The 3D case, however, has special properties that are beyond the scope of this paper [25].

Riemann tensor

The curvature of hypersurfaces is measured by the Riemann tensor **R**. To appreciate the role of this tensor in differential geometry, the above mentioned textbooks may be consulted, but for the scope of this paper it should suffice to know that the Riemann tensor measures deviations from flatness and that we associate geometric flatness with redundancy. The strategy we attribute to the visual system is that it *reduces redundancy* by computing *deviations from flatness*.

Before considering the components of **R**, we should note that the tensor itself is the geometric object of interest because, unlike its components, it is invariant in the sense that it does not depend on the choice of coordinates. In fact, it is part of the difficulties encountered in differential geometry, that in 3D no scalar measure (an invariant of rank one) for curvature exists (**R** has rank four). We should mention here that the Gaussian curvature K is a scalar, but it does not measure curvature in 3D (the hypersurface can be curved even if $K=0$). Neither does the scalar curvature, which is a contraction of the Riemann tensor [28].

Riemann-tensor components

With the above considerations in mind we shall now consider the **R** components. In 3D, **R** has 81 components, but only 6 are independent. Although Cartesian coordinates might not be the optimal choice for visual processing, we will express the 6 components in Cartesian (x,y,t) coordinates (**R** itself does not depend on a particular choice of coordinates):

$$R_{2121} = \frac{f_{xx}f_{yy} - f_{xy}^2}{1 + f_x^2 + f_y^2 + f_t^2}; R_{3131} = \frac{f_{xx}f_{tt} - f_{xt}^2}{1 + f_x^2 + f_y^2 + f_t^2}; R_{3232} = \frac{f_{yy}f_{tt} - f_{yt}^2}{1 + f_x^2 + f_y^2 + f_t^2};$$

$$R_{3231} = \frac{f_{xy}f_{tt} - f_{xt}f_{yt}}{1 + f_x^2 + f_y^2 + f_t^2}; R_{3121} = \frac{f_{xx}f_{yt} - f_{xt}f_{xy}}{1 + f_x^2 + f_y^2 + f_t^2}; R_{3221} = \frac{f_{xy}f_{yt} - f_{yy}f_{xt}}{1 + f_x^2 + f_y^2 + f_t^2}. \tag{2}$$

Subscripts denote differentiation, e.g., f_{xx} is the second order derivative of f with respect to x, and we note that the computation of the **R** components involve first- and second-order derivatives of the luminance function and, more importantly, nonlinear combinations of the derivatives. In 3D, the derivatives can be thought of as being linear filters oriented in space

and time. We notice that if the luminance does not change with time (all derivatives with respect to t are zero) only the component R_{2121} differs from zero. We can think of R_{2121} as a two-dimensional (sectional) curvature in (x,y) since the same expression is obtained for the curvature of a single image (but for the term f_t). By analogy the components R_{3131} and R_{3232} are curvatures in (x,t) and (y,t) respectively. Fig. 1 illustrates the selectivity of the **R** components for different input patterns - see also [29]. If we regard only stationary images, corners are among the most prominent (curved) i2D features. But what are the analogous features in 3D, i.e. for image sequences? As illustrated in Fig. 1, stationary and translated corners, but also discontinuous (e.g. occluded) straight edges activate the Riemann Tensor and are thus i2D features in 3D. In 3D, however, curvature and "cornerness" cannot be measured by a single number. Instead, we have to consider the **R** components as an entity (the tensor). The figure also illustrates the motion selectivity of the **R** components, an issue that will be treated in the next sections.
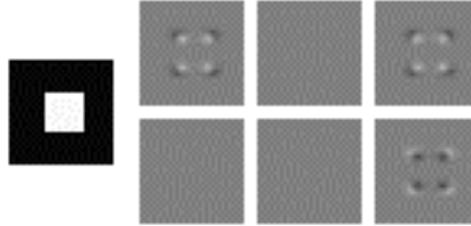

Fig. 1. (180 K) The movie shows the six Riemann-tensor components for a square (shown on the left) that appears and moves in different directions. The arrangement of the components is as in Eq. (2).

Curvature and motion

We are now ready to establish a relationship between the geometric properties defined in the previous section and the computation of visual motion. We assume rigid motion, i.e., an image intensity function restricted by:

$$f: \quad f(x - tv\cos\theta, y - tv\sin\theta) \tag{3}$$

i.e., at any time t image intensity as a function $f(x,y)$ is given by a translation of image intensity at a previous time. The parameters of the motion are speed v and direction θ.

Motion in terms of R components

Now we insert the constraint (3) in the expressions (2) for the **R** components and find the following relations:

$$\frac{R_{3221}}{R_{2121}} = \frac{R_{3232}}{R_{3221}} = \frac{R_{3231}}{R_{3121}} = v\cos(\theta); \quad \frac{R_{3232}}{R_{2121}} = v^2\cos(\theta)^2; \tag{4}$$

$$\frac{R_{3121}}{R_{2121}} = \frac{R_{3131}}{R_{3121}} = \frac{R_{3231}}{R_{3221}} = -v\sin(\theta); \quad \frac{R_{3131}}{R_{2121}} = v^2\sin(\theta)^2. \tag{5}$$

We note that motion introduces strong dependencies among the otherwise independent **R** components. The above relations suggest a number of ways for computing the image-flow field, i.e., for estimating the motion parameters from a given intensity function. Only one of the possibilities, namely $\mathbf{v} = (R_{3221}, -R_{3121})/R_{2121}$, had been derived previously as a by-now-standard method for computing the motion vector **v** [30]. In principle, of course, the above relations can be derived in a purely algebraic framework. Algebraic methods yielded powerful

algorithms for computer vision, we should note, however, that the traditional computer-vision problem of recovering the motion of objects in 3D, see [31] for a review, is not considered here.

The following relations show more explicitly that pairs of **R** components yield representations of direction:

$$\frac{R_{3131}}{R_{3231}} = -\frac{R_{3121}}{R_{3221}} = -\frac{R_{3231}}{R_{3232}} = \tan(\theta); \frac{R_{3131}}{R_{3232}} = \tan(\theta)^2 \ . \tag{6}$$

From Eqs. (4) and (5) we can also obtain an expression that relates the three sectional curvatures by velocity:

$$R_{3131} + R_{3232} = v^2 R_{2121} \ . \tag{7}$$

We conclude that the hypothetical strategy of the visual system, which is to eliminate redundancies by computing curvature, involves representations of motion. We will show in later sections that these relations are relevant to the perception of motion direction and certain properties of motion-selective neurons.

Motion detection

We have seen how motion parameters can be expressed in terms of **R** components when rigid motion is assumed. We now address the problem of inferring the likelihood of rigid motion from the geometry of the movie hypersurface. Computationally, this motion-detection problem is the more difficult one.

If **R** vanishes (all **R** components are equal to zero) the motion can only be zero or ambiguous, a typical case being that of a translating straight pattern (the aperture problem). This relationship can be understood by considering the Fourier transform (FT) of the intensity function f(x,y,t). In case of rigid motion the Fourier transform is in a plane and the parameters of this plane are determined by the motion parameters [3]. It is sufficient that the first 3 components of **R** (the (x,y), (x,t), (y,t) sectional curvatures) be zero to conclude that the FT of *f* is restricted to a line and thus the parameters of the plane and the motion are undefined. The above condition is sufficient because if a 2D curvature is zero, as it is for i1D (and i0D) signals, the FT of that 2D section will be restricted to a line, and if the FTs of 3 orthogonal 2D sections are on a line, then the 3D FT cannot be in a plane or volume. The FT of an i1D signal is restricted to a line because an i1D signal can be written as the product of a one-dimensional function and a constant function, and the latter transforms to a Dirac distribution that restricts the FT to a line. Exceptions from the above considerations are cones and tilted cylinders that are i2D but not curved (this problem may not be that relevant for practical applications but it will be solved in a forthcoming paper by considering another metric than the one induced by Eq. 1; for now we put the left arrow in the first row of Table 1 in parenthesis).

If **R** differs from zero the motion can be of different types. Examples are accelerated straight patterns, translating corners, occlusions, and motion discontinuities. The FT of the input can be on a plane or in a volume and we would like to distinguish between these two cases. With arguments similar to the ones above, it is possible to show that the detection of volumes in the FT is equivalent to the detection of i3D features [28]. A straightforward geometric way of detecting i3D features is to compute the Gaussian curvature of the movie hypersurface. An operator based on Gaussian curvature has been used successfully to detect motion discontinuities [21]. However, at this point the benefits of having different expressions for the estimation of motion parameters become evident. Consider the relations in (6). In case of rigid motion the different estimates of direction will yield the same result. If the assumption of rigid motion is violated, however, the results differ. Thus the differences between different motion estimates can be used as an indicator of non-rigid motion [32]. In the Appendix we proof that if Eqs. (4) and (5) hold, the input must be a translation with constant velocity (3), but for an additive term that depends linearly on time only (because of this restriction the left arrow in the last row of Table 1 is in parenthesis).

Such indicators are of particular interest since for an unambiguous detection of motion it is not sufficient to exclude the cases where the FT of f(x,y,t) is on a line or in a volume. Accelerated straight patterns (i1D occlusions and discontinuities) and translating corners both have a FT on a plane. One way of dealing with this problem is to extract spatial i2D features (corners) and estimate the motion only there. Some optical flow algorithms do so more or less explicitly by the use of confidence measures - see [33] for a recent review on motion algorithms. A closer look at the relations in (4) and (5) reveals that in case of rigid motion all **R** components can be expressed as scaled (by speed and/or direction or their inverse) versions of one of the components. Therefore, in case of rigid motion, if one of the components is zero, all components will be zero. As a consequence, the spatial-filtering pattern is identical for all components, and since R_{2121} is clearly selective to only spatially-i2D features, all other components will be so. Thus, in cases where local motion signals are integrated, no additional selection mechanism, as in [34], is needed. A complementary approach could be to detect 1D discontinuities instead of the continuously moving i2D features. It has been shown before that two-dimensional i2D operators applied to space-time sections of the input sequence can detect such discontinuities [22]. The **R** components R_{3131} and R_{3232} are such detectors of 1D discontinuities. Therefor we can be confident that if motion estimates based on R_{3131} and R_{3232} differ from those based on other components, the difference will indicate a violation of Equation (3) even in the cases where the FT of the input is in a plane. To be more precise, we need the following logical expression to be true in order to infer rigid motion: ($R \neq 0$) *AND (K=0) AND NOT ((R$_{2121}$=0) AND* (($R_{3131} \neq 0$) OR ($R_{3232} \neq 0$))). After expanding and reorganizing this expression we obtain: ((*K=0*) AND ($R \neq 0$) AND ($R_{2121} \neq 0$) OR (((*K=0*) AND ($R \neq 0$) AND (*R$_{3131}$=0) AND (R$_{3232}$=0)*). But since the second part of this expression cannot be true in case of rigid motion (as we have shown, if one components is equal to zero, the tensor will vanish) the first part ((*K=0*) AND ($R \neq 0$) AND ($R_{2121} \neq 0$) is a necessary and sufficient condition.

We conclude that in case of rigid motion we can express the motion parameters in terms of **R** components and, in turn, we can use the geometric properties of the movie hypersurface to detect rigid motion (and thereby avoid fitting lines and volumes to motion planes). We have thus established a few correspondences between rigid motion and the (intrinsic) geometry of the input. The correspondences are summarized in Table 1 and the different intrinsic dimensions of a simple image sequence are illustrated in Fig. 2.

Ambiguous motion	(<-) ->	*R=0*
Occlusions, discontinuities, multiple motions	<->	*K≠0*
Defined motion (Eq. 3)	->	*R≠0, K=0*
Defined motion (Eq. 3)	<->	*R≠0, K=0,* all **R** components *≠0*
Defined motion (Eq. 3)	<->	*R≠0, K=0, R$_{2121}$≠0*
Defined motion (Eq. 3)	(<-) ->	Different direction in **(6)** defined and equal

Table 1. Summary of correspondences between motions and curvatures.

Fig. 2. (196 K) The movie shows a moving square with colored circles (red, green, blue for intrinsic dimensions 3, 2, 1) indicating the intrinsic dimension of different movie regions.

Simulations of global motion percepts

In the barber-pole illusion we see lines moving in a direction defined by the motion of the line ends [35] and small changes in the shape of the aperture can change the perceived direction [36] – see Fig. 3.
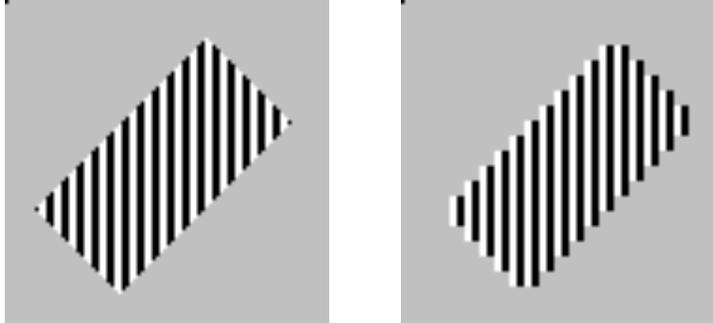

Fig. 3. (2*148 K) Two movies showing the Barber-pole illusion (left) and the Kooi effect (right). While in both cases the grating moves horizontally to the left behind the gray aperture, we see it moving down left in the direction of the oblique aperture in left movie. If the shape of the aperture is changed as in the right movie, the grating is seen to move horizontally.

The only assumption we need to make in order to explain these percepts in our geometric framework is that they result from spatially integrated **R** components. Since all components are endstopped (equal to zero for translating straight patterns) the motion estimated at line ends will determine the direction of global motion (interestingly, the same principle, i.e., the integration of curved features over space, has been used to explain texture segregation [37]). To obtain the simulation results shown in Fig. 4 we have low-pass filtered the numerators of the **R** components R_{3221} and R_{3231} and treated the results as components of vectors that are plotted on a sub-sampled grid only for those pixels where the absolute value of the temporal derivative is larger than a small threshold. (These simulations are only meant to illustrate the resulting direction of perceived motion at different locations, and not to suggest the perception of a dense flow field. Similar results would have been obtained by using the full components including the normalization by the gradient in the denominator. The issue of normalization has been discussed in previous work, e.g. in [26]. Also, we could have used other pairs of components according to Equations (4) and (5) since for translated line ends these pairs yield the same direction). It is assumed that motion is detected (segmented) at a higher spatial resolution, and that the direction of motion is estimated at lower resolution and "attached" to the moving pattern.
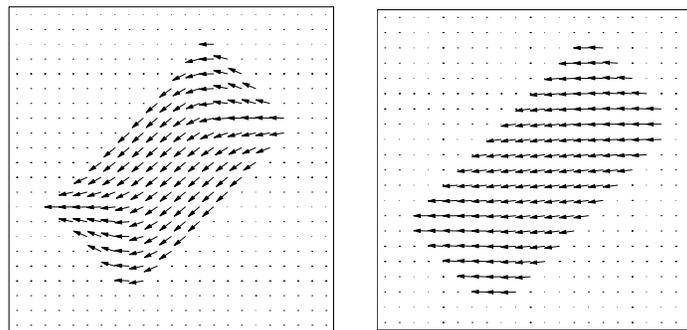

Fig. 4. Simulation results for the two movies in Fig. 3. Note that small changes in the shape of the aperture change the resulting global direction of motion.

It seems that the visual system deals with the aperture problem by simply ignoring regions of undefined motion (and not by assigning a direction perpendicular to the spatial pattern, i.e., the direction of the spatial gradient). Another way of putting it is to say that seeing the straight lines moving is an illusion induced by the line ends.

Curvature and motion-selective neurons

We have shown that curvature-selective operators (or neurons) would also be motion selective. MT neurons, however, have properties that are not sufficiently described by the term "motion selectivity". To test hypothetical mechanisms of motion selectivity, dynamic input patterns other than rigid motions must be considered. We will now give two examples of data from the literature that have been obtained for such patterns, and which favor our view on motion selectivity.

Orthogonal direction and orientation tunings

As shown in [38], neurons in monkey cortical area MT can have a spatial-orientation tuning that is orthogonal to the direction tuning of those same neurons. The authors measured the tuning curves of neurons for a dot moving in different directions (top right of Fig. 5) and for bars flashed at different orientations (bottom right).
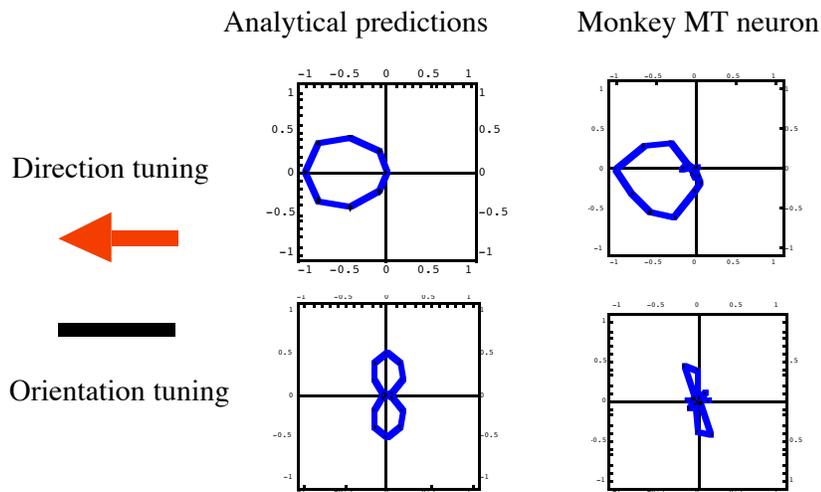

Fig. 5. Data by Albright [38] are shown in the right columns for direction (top) and orientation (bottom) selectivity of macaque MT neurons (that are selective to motion along the preferred spatial orientation) and simulation results in the left columns - see text.

The corresponding simulation results are shown in the left columns, and have been obtained by analytically evaluating the normalized mean of the two vectors $(-R_{3221}, R_{3121})$ and (R_{3232}, R_{3131}) for a Gaussian blob parameterized by direction of motion, i.e., $e^{-\pi(x-tCos(\theta)+y-tSin(\theta))^2}$ and a flickering grating parameterized by spatial orientation, i.e., $Cos(2\pi t) \cdot Cos(2\pi(xCos(\varphi)+ySin(\varphi)))$ (these inputs and those in the next section have been used for computational convenience only).

The key to the simulation results is that the components R_{3131} and R_{3232} (the sectional curvatures in x,t and y,t) are selective to a certain spatial orientation of a transient edge or line and to motion along that edge or line - see Fig. 1. Although this is not relevant for the presented results, we should mention that only the numerators of the R components have been evaluated. In order to avoid the insensitivity of R_{3131} and R_{3232} to the sign of motion, these two components have been clipped, i.e., computed as $(Max[0, R_{3232}], Max[0, R_{3131}])$. Finally the

resulting expressions have been evaluated at the location x=y=t=0 and plotted as a function of direction θ and orientation φ respectively (polar plots).

Multiple motions

Recanzone et al. have measured the selectivity of MT neurons to multiple motions [39]. While moving one dot in the (previously determined) preferred direction and opposite to this direction respectively, tuning curves were measured for a second dot. Representative data are shown on the right in Fig. 6. The simulation results are shown on the left in Fig. 6, and have been obtained by evaluating the same analytical function as in the previous section. In this case, the input functions were defined as $e^{-\pi(x-tCos(\theta)+y-tSin(\theta))^2}$ (for the case one dot), $e^{-\pi(x-tCos(\theta)+y-tSin(\theta))^2} + e^{-\pi(x+y-t)^2}$ (for an additional dot moving in the preferred direction), and $e^{-\pi(x-tCos(\theta)+y-tSin(\theta))^2} + e^{-\pi(x+y+t)^2}$ (for an additional dot moving opposite to the preferred direction) respectively. Again, the resulting expressions have been evaluated at the location x=y=t=0 and plotted as function of direction θ.

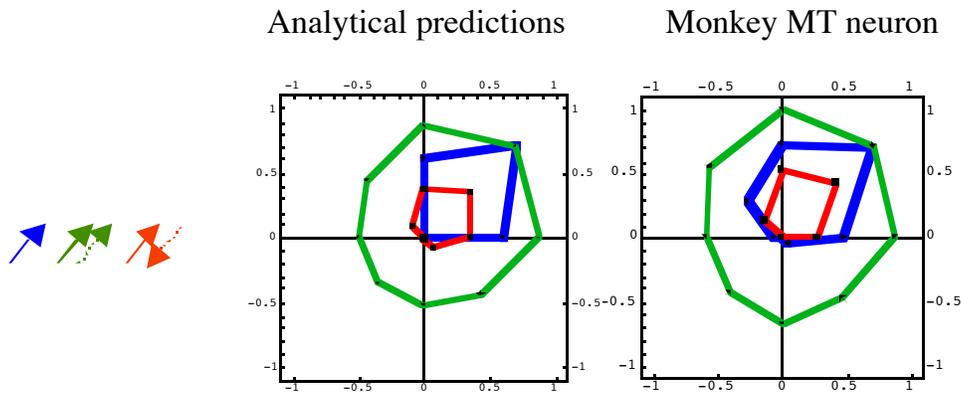Analytical predictions   Monkey MT neuron

Fig. 6. Data by Recanzone et al. illustrating the selectivity of MT neurons to multiple motions are shown on the right and simulation results obtained as in Fig. 5 (but for a rotation by 135 degree) on the left. As indicated by the arrows on the left, blue is chosen for the case of a single moving dot, red for the case with an additional dot moving opposite to the preferred direction, and green for the case with an additional dot moving along the preferred direction.

Certainly, the above simulations are not based on elaborated models designed to account for many data sets (as for example [40] is). However, it seems an important result that simple analytical evaluations in a new framework result in such excellent predictions.

Computational aspects

A frequent concern with differential methods is the sensitivity to noise, and algorithms for motion estimation incorporate more or less extensive regularization procedures to cope with noise and the aperture problem [33]. Our approach offers the possibility of regularization not only in space and/or time, but also *within* the four motion frames, with the benefit that such inter-frame regularization does not reduce spatial resolution. In addition, the frames can be used to detect occlusions. Regarding the above simulations, a benefit of evaluating the mean of different motion frames is that it yields a more robust estimate of rigid-motion parameters. Conversely, the differences between motion frames can be used as indicators of non-rigid motion [32].

Discussion

The strategy we attribute to the visual system is that it evaluates the intrinsic dimension of the input and, by doing so, reduces redundancy in that input. Such a strategy can be implemented with a first stage consisting of linear filters oriented in space and time. Such a stage is common to most vision models. Should the description of the next stages be based on differential geometry, the linear filters must be derivatives. Problems with noise sensitivity of differentials can be avoided by first low-pass filtering the input. In previous work, however, we have already shown ways of computing *inD* features based on other filter functions and different nonlinearities. Whatever the shape of the linear filters, the model we propose has a second stage where the nonlinearities suppress flat regions of the movie hypersurface (i0D and i1D signals). The hypothetical purpose of stage 2 is to eliminate redundancy but, as we have shown, this stage will thereby involve representations of motion.

Conclusion

The main benefits of our approach are: (i) it is in accordance with how we perceive the global direction of motion (ii) it predicts the motion selectivity of MT neurons including more general features (iii) it relates the intrinsic geometry of the visual input to motion parameters, thereby providing a new geometric framework for motion selectivity.

Acknowledgements

Appendix

If a and b are given by Eqs. (4) and (5) and are constant in a neighborhood of a point, then $\frac{\partial}{\partial x}F(x,y,t) = \frac{\partial}{\partial y}F(x,y,t) = \frac{\partial}{\partial t}F(x,y,t) = 0$ with $F(x,y,t) = f_x a + f_y b + f_t$. This can easily be verified by substitution. Therefore, we have $f_x a + f_y b + f_t = c$ with c a constant. The solution of the above equation is $f(x,y,t) = f(x-at, y-bt) + ct$.