

A geometric view on early and middle level visual coding

Erhardt Barth

Institute for Signal Processing, Ratzeburger Allee 160, D-23538 Lübeck, Germany.
barth@isip.mu-luebeck.de,
<http://www.isip.mu-luebeck.de>

Keywords: motion, MT neurons, nonlinear features, curvature tensor

Abstract

As opposed to dealing with the geometry of objects in the 3D world, this paper considers the geometry of the visual input itself, i.e. the geometry of the spatio-temporal hypersurface defined by image intensity as a function of two spatial coordinates and time. The results show how the Riemann curvature tensor of this hypersurface represents speed and direction of motion, and thereby allows to predict global motion percepts and properties of MT neurons. It is argued that important aspects of early and middle level visual coding may be understood as resulting from basic geometric processing of the spatio-temporal visual input. Finally, applications show that the approach can improve the computation of motion.

Introduction

Traditionally it is assumed that visual computation deals, to a large extent, with the problem of reconstructing the 3D world from the image-intensity input $f(x,y,t)$. A good example is visual motion: objects move and thereby induce a retinal optical flow which, with certain restrictions, can be used to estimate the true motion of the objects in the world. The assumption is then validated by noting that human observers and cortical neurons are sensitive to parameters of the motion in a way which is more or less consistent with algorithms used to compute the motion. Moreover, the optical-flow paradigm has been useful in technical applications like robot navigation, video compression, tracking, and shape from motion.

We consider a different framework for dealing with visual motion selectivity. Instead of assuming that visual computations are meant to recover the 3D world, we assume that the purpose of early and mid-level vision is to perform an efficient coding of the input. In the tradition of Barlow [1] a few authors have related visual computations to the statistics of natural images [2-6]. In parallel it has been shown how redundancy and predictability can be related to the intrinsic geometry of the visual input if we think of images as surfaces [4, 7]. In particular, it has been shown that curved image regions are highly significant: although in natural images curved regions (where the Gaussian curvature of the associated surface differs from zero) are rare events, images can be reconstructed from generalized curvature measures [7]. Therefore, as an alternative view to predictive coding based on the statistics of natural images, we can hypothesize that low- and mid-level vision deals with the intrinsic geometry of the visual input. In this paper we extend this view from images to image sequences by showing that it is consistent with the notion of motion

selectivity. Moreover, it provides new insights into the problem of motion estimation and explains a few results from the psychophysics and neurophysiology of motion in cases where the notion of motion selectivity is less useful.

Curvature of movie hypersurfaces

In this paper we deal with the geometry of the visual input itself, i.e., the geometry of the hypersurface defined by

$$(x, y, t, f(x, y, t)) \quad (1)$$

where f denotes image intensity at position (x, y) and time t . The geometry of this movie hypersurface is hard to visualize and differs considerably from the geometry of the image surface but, as with images, we will assume that the curvature of the hypersurface is the feature of interest. Curvature as a measure of deviation from flatness should not be confounded with mean curvature, which is a very different concept (not further explained here). Deviations from flatness 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 curvature tensor \mathbf{R} . The strategy we attribute to the visual system is that it reduces redundancy by computing deviations from flatness.

Before considering the \mathbf{R} components, we should note that the tensor itself is the geometric object of interest because, unlike its components, it is invariant. 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 (\mathbf{R} has rank four). We should mention here that the Gaussian curvature K does not measure curvature in 3D (the hypersurface can be curved even if $K=0$). For images a typical curved feature is a corner and it can be characterized by a scalar measure. A spatio-temporal "corner", however, must be characterized by a tensor.

In case of three-dimensional manifolds, \mathbf{R} has 81 components, but only 6 are independent and given below for a hypersurface of type (1):

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

$$R_{3121} = \frac{f_{xx}f_{yt} - f_{xt}f_{xy}}{1 + f_x^2 + f_y^2 + f_t^2}; R_{3221} = \frac{f_{xy}f_{yt} - f_{yy}f_{xt}}{1 + f_x^2 + f_y^2 + f_t^2}; R_{3231} = \frac{f_{xy}f_{tt} - f_{xt}f_{yt}}{1 + f_x^2 + f_y^2 + f_t^2} \quad (3)$$

The derivatives f_{xt} , ... 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. The component R_{2121} is similar to the curvature in the 2D case but for the f_t term in the denominator. We can therefore think of R_{2121} as a (sectional) curvature in (x, y) . By analogy the components R_{3131} and R_{3232} are curvatures in (x, t) and (y, t) respectively. If the mixed derivatives are equal to zero (the coordinate frame coincides with the main axis) only these three sectional curvatures will differ from zero. A misalignment of the coordinate frame (this is the case of motion) will activate the remaining components also.

Curvature and motion

We now assume rigid motion, i.e., an image-intensity function restricted by:

$$f: f'(x - tv \cos \theta, y - tv \sin \theta) \quad (4)$$

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

Now we insert the constraint (4) in the expressions (2) and (3) for the \mathbf{R} components and find the following relations (by symbolic simplification [8]):

$$\frac{R_{3221}}{R_{2121}} = \frac{R_{3231}}{R_{3121}} = \frac{R_{3232}}{R_{3221}} = v \cos(\theta); \quad \frac{R_{3232}}{R_{2121}} = v^2 \cos(\theta)^2; \quad (5)$$

$$\frac{R_{3121}}{R_{2121}} = \frac{R_{3131}}{R_{3121}} = \frac{R_{3231}}{R_{3221}} = -v \sin(\theta); \quad \frac{R_{3131}}{R_{2121}} = v^2 \sin(\theta)^2; \quad (6)$$

We note that motion introduces dependencies among the otherwise independent \mathbf{R} components, which thereby provide distributed representations of motion. 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. One of the possibilities (the left-hand terms in(5) and (6), i.e., the velocity vector $(R_{3221}/R_{2121}, R_{3121}/R_{2121})$) has been derived previously, and by other means, as a by now standard method for optical flow estimation [9]. Note that the relations in (5) and (6) also hold for the numerators in (2) and (3).

What are the benefits of having different expressions for the estimation of motion parameters? In case of rigid motion the different estimates will yield the same result and the mean of the estimates a more robust measure. If the assumption of rigid motion is violated, however, the results will differ. Therefor the differences between different motion estimates can be used as an indicator of non-rigid motion. An example of such an application is given in Figure 1.

Insert Figure 1 about here.

Global motion percepts and motion-selective neurons

We now briefly present a few examples of percepts and results from neurophysiology that favor our view on motion selectivity. The results were first presented in [10] and details of the simulations are given in [11].

In the barber-pole illusion we see lines moving in a direction defined by the motion of the line ends and small changes in the shape of the aperture can change the perceived direction [12]. The only assumption we need to make in order to explain this percept in our geometric framework is, that it results from spatially integrated \mathbf{R} components [10] - like the motion sketch in Figure 1. 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 [13].

As shown in [14], neurons in monkey cortical area MT can have a spatial-orientation tuning that is orthogonal to the direction tuning of those same neurons - see Figure 2. This property and the selectivity of MT neurons to multiple motions [15] have been modeled by assuming that these neurons evaluate the mean of different motion frames - see Figures 2 and 3.

Insert Figures 2 and 3 about here.

Discussion

We have related (intrinsic) geometric properties of the visual input, in particular the Riemann tensor, to the problems of motion detection and the estimation of optic flow. Previously Zetsche and Barth [16] have argued that an operator based on the Gaussian curvature of the hypersurface can be used to detect flow-field discontinuities. It has also been shown that computations possibly involved in flow-field estimation are related to endstopping and detectors for discontinuities due to spatially 1D occlusions have been proposed [17]. A specific relationship between Riemann-tensor components and flow field estimation has been pointed out [18]. Differential techniques have often been criticized for being sensitive to noise. However, we have extended the geometric framework to what we called *geometric signal processing* [4, 7].

To summarize, the strategy we attribute to the visual system is that it evaluates intrinsic geometrical properties 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 but this is not a necessary condition. 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. As we have shown, this stage will involve multiple, distributed representations of motion. Of course, these representations can be used for navigation and higher-order interpretations of motion.

Acknowledgement

This work has been supported by the Deutsche Forschungsgemeinschaft, grant Ba 1176/4-1. Preliminary results have been presented before [10].

References

1. Barlow, H.B., *Possible principles underlying the transformation of sensory messages*, in *Sensory Communication*, W. Rosenblith, Editor. 1961, MIT Press: Cambridge, MA. p. 217--234.
2. Field, D.J., *Relations between the statistics of natural images and the response properties of cortical cells*. *J Opt Soc Am A*, 1987. **A4**: p. 2379--2394.
3. Zetsche, C. and W. Schönecker, *Orientation selective filters lead to entropy reduction in the processing of natural images*. *Perception*, 1987. **16**: p. 229.
4. Zetsche, C., 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, Editor. 1993, MIT Press: Cambridge, MA. p. 109--138.
5. Field, D.J., *What is the goal of sensory coding?* *Neural computation*, 1994. **6**: p. 559-601.
6. Rao, R.P. and D.H. Ballard, *Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects*. *Nat Neurosci*, 1999. **2**(1): p. 79-87.
7. Barth, E., T. Caelli, and C. Zetsche, *Image encoding, labelling and reconstruction from differential geometry*. *CVGIP:GRAPHICAL MODELS AND IMAGE PROCESSING*, 1993. **55**(6): p. 428--446.
8. Wolfram, S., *Mathematica: A system for doing mathematics by computer*. Second Edition ed. 1991, New York: Addison-Wesley.
9. Tretiak, O. and L. Pastor. *Velocity estimation from image sequences with second order differential operators*. in *Proc. 7th Int. Conf. Pattern Recognition*. 1984. Montreal, Canada: IEEE Computer Society Press.
10. Barth, E. and A.B. Watson, *Nonlinear spatio-temporal model based on the geometry of the visual input*. *Investigative Ophthalmology and Visual Science*, 1998. **39**(4): p. S2110.
11. Barth, E. and A.B. Watson, *A geometric framework for nonlinear visual coding*. in *submission*.
12. Kooi, F.L., *Local direction of edge motion causes and abolishes the barberpole illusion*. *Vision Res*, 1993. **33**(16): p. 2347-51.
13. Barth, E., C. Zetsche, and I. Rentschler, *Intrinsic two-dimensional features as textons*. *J Opt Soc Am A Opt Image Sci Vis*, 1998. **15**(7): p. 1723-32.
14. Albright, T.D., *Direction and orientation selectivity of neurons in visual area MT of the macaque*. *J Neurophysiol*, 1984. **52**(6): p. 1106-30.
15. Recanzone, G.H., R.H. Wurtz, and U. Schwarz, *Responses of MT and MST neurons to one and two moving objects in the receptive field*. *J Neurophysiol*, 1997. **78**(6): p. 2904-15.
16. Zetsche, C. and E. Barth, *Direct detection of flow discontinuities by 3D-curvature operators*. *Pattern Recognition Letters*, 1991. **12**: p. 771--779.
17. Zetsche, C., E. Barth, and J. Berkman, *Spatio-temporal curvature measures for flow field analysis*. *Geometric Methods in Computer Vision*, B. Vemuri Ed., 1991. **SPIE 1590**: p. 337--350.
18. Barth, E., C. Zetsche, and G. Krieger, *Curvature measures in visual information processing*. *Open Systems and Information Dynamics*, 1998. **5**: p. 25-39.

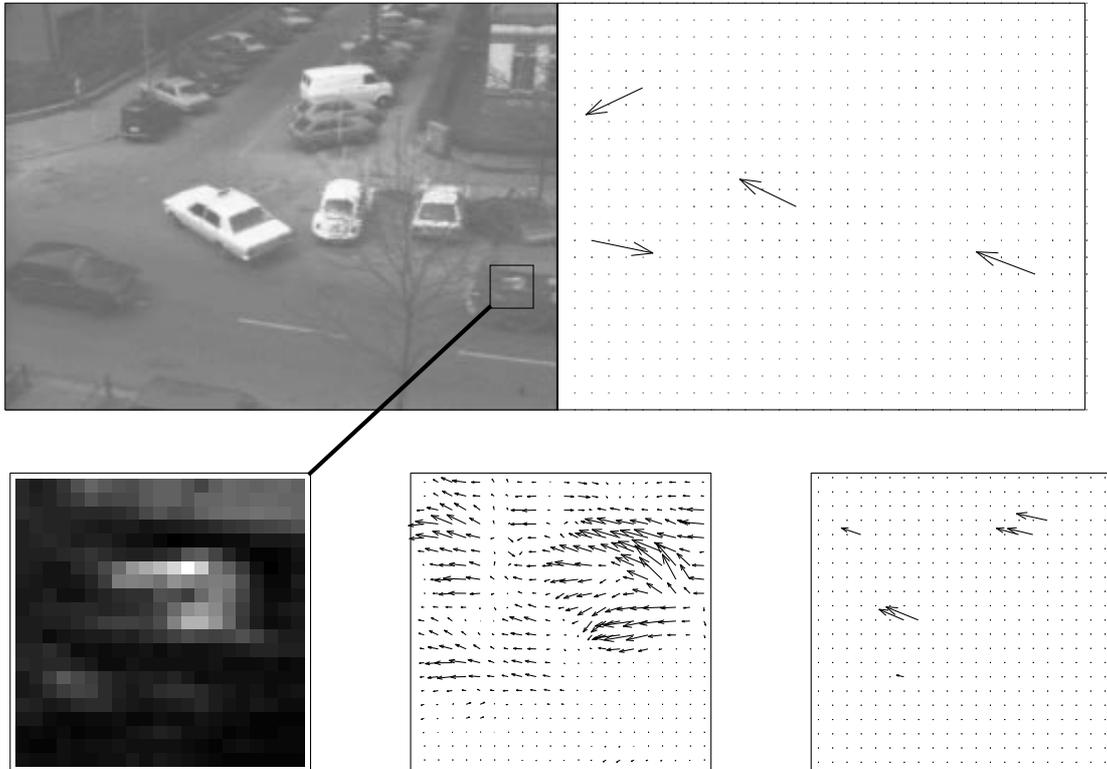


Figure 1. Frame seven of the Hamburg-taxi movie is shown top left. In this scene the bright taxi in the center turns around the corner, a small dark car moves from left to right, a dark van from right to left, and a pedestrian walks along the top left sidewalk (the movie can be viewed at www.isip.mu-luebeck.de/~barth/papers/badhomburg.html). The sequence has been pre-processed by a ROG (Ratio of Gaussians) filter, and the result of this operation is shown enlarged at the bottom left for the section indicated by a dark frame in the original image (depicting an area around the window of the van). The results of computing the optical flow from the first terms in (5) and (6), i.e., based on the traditional constant gradient approach, are shown in the middle bottom (for the same section within the dark frame). On the bottom, right, the mean of the 4 flow fields is shown at those locations, where the variance in direction was less than 5 deg. Note that incorrect motion vectors (that would be generated by a number of motion algorithms) are being eliminated. The motion sketch shown top right has been obtained by low-pass filtering the corrected local flow field (over the whole image). Subsequently the result has been sub-sampled and connected regions have been reduced to single values by a morphological image operation (erosion). The resulting vectors are shown with unit length.

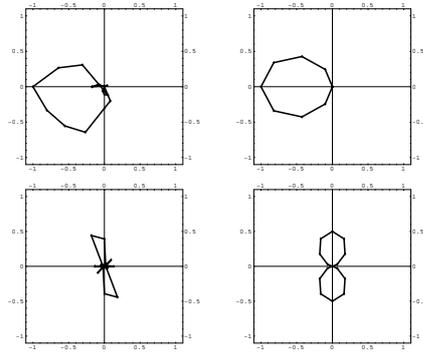


Figure 2 Data by Albright [14] are shown in the left columns for direction (top) and orientation (bottom) selectivity of macaque MT neurons (that are selective to motion along the preferred spatial orientation). The corresponding simulation results are shown in the right columns, and have been obtained 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 and a flickering grating parameterized by spatial orientation - see [11] for details. 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.

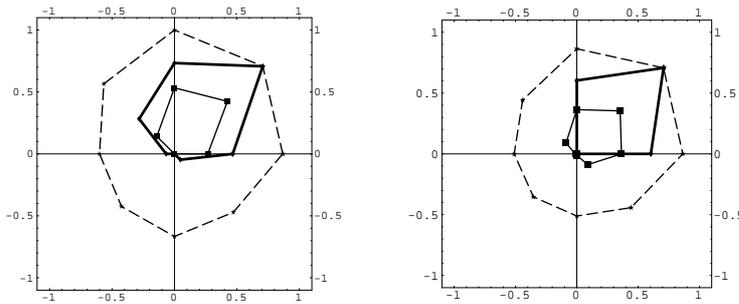


Figure 3 Data by Recanzone et al. illustrating the selectivity of MT neurons to multiple motions [15] are shown on the left and simulation results obtained as in Figure 2 on the right (with appropriate input functions). Thick lines are chosen for the case of a single moving dot, thin lines for the case with an additional dot moving opposite to the preferred direction (the inner curve), and dashed lines for the case with an additional dot moving along the preferred direction.