A geometric view on early and middle level visual coding

ERHARDT BARTH *

Institute for Signal Processing, Ratzeburger Allee 160, D-23538 Lübeck, Germany

Received 2 June 1999; revised 15 March 2000; accepted 16 March 2000

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.

Keywords: Motion; MT neurons; nonlinear features; curvature tensor.

1. 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 (1961) a few authors have related visual computations to the statistics of natural images (Field, 1987, 1994; Zetzsche
and Schönecker, 1987; Zetzsche et al., 1993; Rao and Ballard, 1999). 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 (Barth et al., 1993; Zetzsche et al., 1993). 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 (Barth et al., 1993). 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.

2. 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 \(R\). The strategy we attribute to the visual system is that it reduces redundancy by computing deviations from flatness.

Before considering the \(R\) 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 (\(R\) 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.