Motion Segmentation in Artificial and Biological Systems*

Walter J. Gillner

Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany

Z. Naturforsch. **53c**, 622–627 (1998); received April 22, 1998

Motion Detection, Segmentation, Motion Analysis, Independent Motion, Active Vision

In the early steps of visual information processing motion is one of the most important queues for the development of spatial representations. Obstacle detection and egomotion estimation are only two examples of the powerfulness of visual motion detection systems. The underlying process of information extraction has to be active due to the observer's capabilities of egomotion. This means that the observer's motion has an impact on the projected retinal motion field. Therefore one of the challenging tasks for biological as well as for technical vision systems is to couple retinal motion and egomotion and to uncouple egomotion and object motion. The following sections describe a model that couples visual motion processing with the egomotion parameters of a moving observer. Beneath a theoretical introduction of the model an application to traffic scene analysis is presented. At last the paper relates the model to biological motion processing systems.

Introduction

The classical theoretic approaches for the estimation of egomotion rely on the computation of the optical flow. This apparent retinal motion field is mostly evaluated in differential approaches by using the local spatiotemporal image gradient. This is formulated together with the searched motion vector $(u_x,u_y)^T$ in the „Brightness Constancy Equation„ (BCE).

$$I_x u_x + I_y u_y + I_t = 0 \ ,$$

with $\boldsymbol{u}=(dx/dt \ , \ dy/dt)^T$, where I_x, I_y and I_t denote the partial differentials of the image intensity I (Horn and Schunk, 1981). The equation is valid under the assumption that the total change in the image intensity is zero. The solution of the BCE yields the optical flow.

The BCE suffers from the aperture problem (Hildreth, 1984), which means that its solution is ambiguous. This is because it yields the motion field only in the direction of the image gradient. The perpendicular direction rests ambiguous. Trying to eliminate these ambiguities by regularisation reduces the spatial resolution of the resulting motion field.

The classical theoretical and technical approaches use optical flow to estimate the position of the focus of expansion (FOE) (Gibson, 1950), which is a measure for the direction of egomotion. But secure detection of this point is not guaranteed if an observer is moving not only in translational degrees of freedom. This is because the FOE can be masked by a rotational field (see Fig. 1). On the other hand, if the FOE can be found correctly, then the absolute values of the motion vectors along the direction of motion give a measure of the relative scene depth. Vectors perpendicular to the direction of motion indicate independent object motion.

Coupling Environmental Knowledge and Motion Measurement

In the following we present a way to couple pre-acquired knowledge about the environment with the estimation of the motion field. A model for the synchronous description and measurement of the expected egomotion trajectory is presented. To simplify the development of the model it is assumed that the observer's motion describes a cir-

---

* This communication is a contribution to the workshop on "Natural Organisms, Artificial Organisms, and Their Brains" at the Zentrum für interdisziplinäre Forschung (ZiF) in Bielefeld (Germany) on March 8–12, 1998.

Reprint requests to W. Gilner,
Daimler Benz Forschungszentrum,
Wilhelm-Runge-Str. 11, 89081 Ulm.
Fax: +497315054224.
E-mail: Walter.Gillner@dbag.ulm.daimlerbenz.com

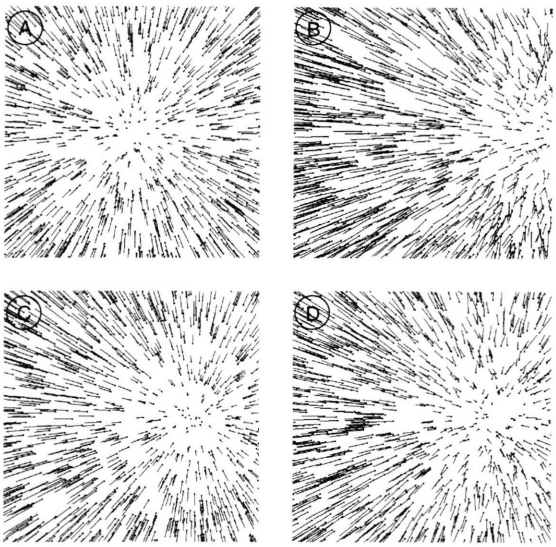

Fig. 1. Projected synthetic motion fields. A: Pure translation in Z-direction. The FOE is located on the optical axis. B: Combination of translation and rotation. In the front the translation is dominating. C: Combination of two translations (ahead and parallel to image plane). D: Combination of translation and rotation, which yields a field which is similar to C.

cular arc on a plane surface, which can be described by a simple circle function

$$R_c^2=(X-X_0)^2+(Z-Z_0)^2$$

The observer's motion then consists of a progressive translational component W and a rotational component Ω_Y about the vertical axis which together yield the radius

$$R_c = \frac{W}{\Omega_Y}$$

of the circle function. To receive more than one circle function it is necessary to add an additional parameter ΔX_0 which denotes a lateral distance and yields

$$(R_c-\Delta X_0)^2=(X-R_c)^2+(Z-Z_0)^2$$

where $X_0 = R_c$.

To describe the observer's retinal flow it is necessary to map this circle function by perspective mapping into the observer's eye coordinate system, i.e. into the image plane. A scheme of this mapping and the declaration of the variables can be found in figure 2.

Solving for the image coordinate x_{tr} gives

$$x_{tr} = \frac{y}{H}\left(R_c - \sqrt{(R_c - \Delta X_0)^2 - \left(\frac{fH}{y} - Z_0\right)^2}\right),$$

which is an analytic description of the retinal flow in the image plane through perspectively projected circle functions. These functions describe the trajectory of the image points and the underlying grey values, which depend on the parameters of the observer's egomotion. Given this descriptive set of trajectories the velocity measurement necessary for the estimation of scene depth can be done on the basis of a three-dimensional spatiotemporal sampling function

$$b(t',y') = \sum_x \sum_y \sum_t I(x,y,t)\,\delta\,(t'-t)\delta\,(x_{tr}-x)\delta\,(y'-y)\,,$$

which maps the grey values underlying the trajectories into coordinates of a spatiotemporal image. A scheme of this mapping procedure is shown in figure 3.

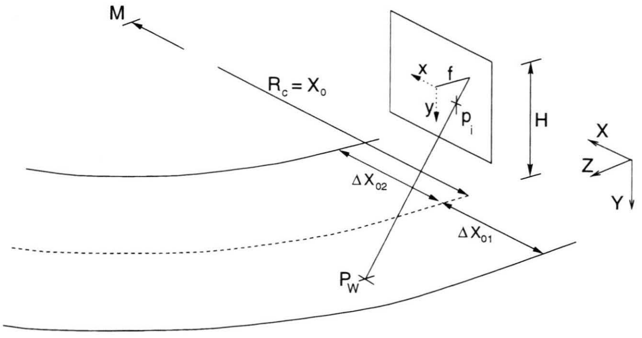

Fig. 2. Trajectories on the ground mapped into the image plane. The observer's movement describes a circle function in world coordinates (f: focal distance, H: observers height above the ground plane, x,y: image plane coordinates). A world point P_w is mapped into an image point p_i.

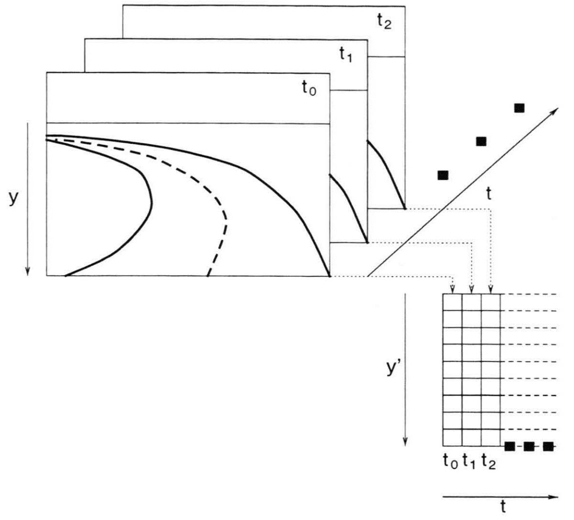

Fig. 3. Sampling of the grey values of the image sequence underlying the trajectory (dotted line). The grey values are mapped into a spatiotemporal image where the horizontal axis is the time axis and the vertical axis is related to the route along the trajectory.

The rows of this image are equivalent to the route along the trajectory whereas the columns define the time axis. The velocity of the objects in the image sequence is represented by the slope of structures and edges in the spatiotemporal image. Fig. 4 shows a schematic description of the mapping result. The object velocities in world coordinates result from a backprojection of the vertical axis of the spatiotemporal image into world coordinates.

Egomotion Detection

In classical approaches as mentioned before the detection of egomotion is done by estimating the position of the focus of expansion. The detection of the observer's egomotion by using the sampling function is a nonlinear problem since we have to minimize a nonlinear error function. The definition of this error function becomes clear if we look at a motion sequence with wrongly parameterized sampling functions (see Fig. 5). The resulting spatiotemporal images are depicted in Fig. 6.

This wrong parametrized sampling function results in spatiotemporal images, which are disturbed by horizontal structures. The amount of disturbance decreases with the accuracy of the sampling function, but these structures give no information about the direction in which the parameters of the sampling function have to be changed in order to reduce the number of disturbing horizontal structures and therefore to reach the global minimum of the error function.

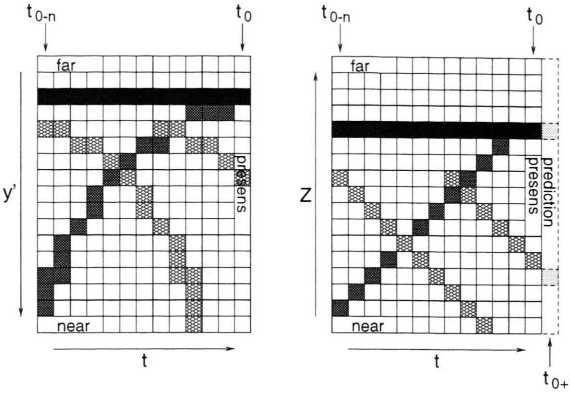

Fig. 4. Schematic description of the mapping result. A rigid environment is assumed. The sampling function yields the spatiotemporal image on the left. Nonlinear structures in y'-t-coordinates on the left are projected through perspective backprojection into Z-t-coordinates on the right. Here the slope of the image structures represent the object velocities.

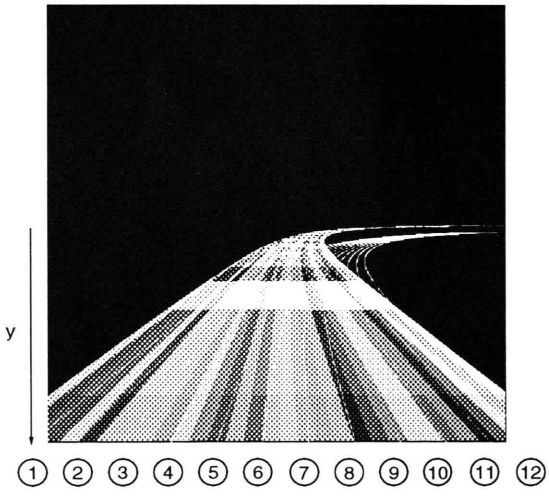

Fig. 5. Synthetic scene with wrongly parameterized sampling function.The sampling function crosses the expected trajectories.

Independent Motion

It can be shown that motion of independent objects also results in disturbances in the spatiotemporal images, but normally they are not of global

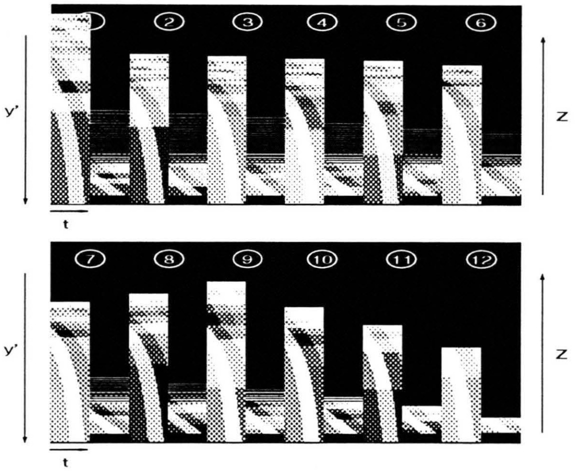

Fig. 6. Spatiotemporal images from Fig. 5. The 12 spatiotemporal images are disturbed by horizontal structures which result from a wrongly parameterized set of sampling functions. Each image is depicted in y'-t-coordinates (left) and Z-t-coordinates (right).

nature. A special case can be shown for the movement of the observer over the ground plane. It results from the spatial derivative dx/dy of the trajectory functions in image coordinates. (see Fig. 7).

The minima of each of the trajectories on the horizontal image axis define a curve where the horizontal components of the velocity field have a zero crossing, which means that they change the sign. A consistency test can find independently moving objects, if they move in the opposite direction of the expected horizontal field.

Application to Traffic Scene Analysis

One of the main goals in traffic scene analysis is to get a representation of the environment in which the observer is moving. To solve complex tasks like object recognition (Noll, 1996) especially segmentation of obstacles, overtaking cars and other moving objects from the moving background is one of the most important preprocessing steps that an artificial vision system has to solve.

The egomotion parameters can be extracted by using the trajectory description as a model for a lane estimator (Gillner, 1997), which is mainly based on fitting the model to the position and the direction of detected lane markings. The observer's translational velocity can easily be estimated from the spatiotemporal structure of the periodic stripes of the lane markings. The trajectory model allows the integration of external sensors like velocity indicators and gyroscopes to give an egomotion estimation.

Fig. 8 shows an image of a motorway scenario and the resulting spatiotemporal images. The spatiotemporal structures of the lane markings indicate the observers velocity. By simple elimination of the translational ground velocity a segmentation of rigid obstacles and independently moving objects can be reached. The result for the motorway scenario is depicted in figure 9.

Biological Implications and Discussion

The functionality and the signal processing steps discussed in the first sections of this paper can be related to signal processing functions found in higher vertebrates. Visual motion processing is located in the dorsal pathway. The most known areas are V1, MT and MST. V1 (cortex striatum) is responsible for orientation coding (Hubel and Wiesel, 1977)

Orientation and direction sensitive cells can be found in in area MT (Maunsell and van Essen, 1983) of the macaque monkey. MT cells show reactions to certain directions of motion. With higher excentricity of the visual field they are well-

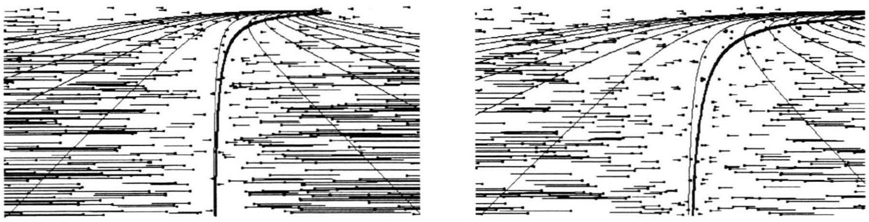

Fig. 7. Horizontal motion field on the ground plane. The minimum function of the spatial differential dx/dy of the trajectories in image coordinates (thick line) define the zero crossing of the horizontal component of the motion field. Only the horizontal component of the motion field is shown (A: R_c=800 m , B: R_c=300 m).

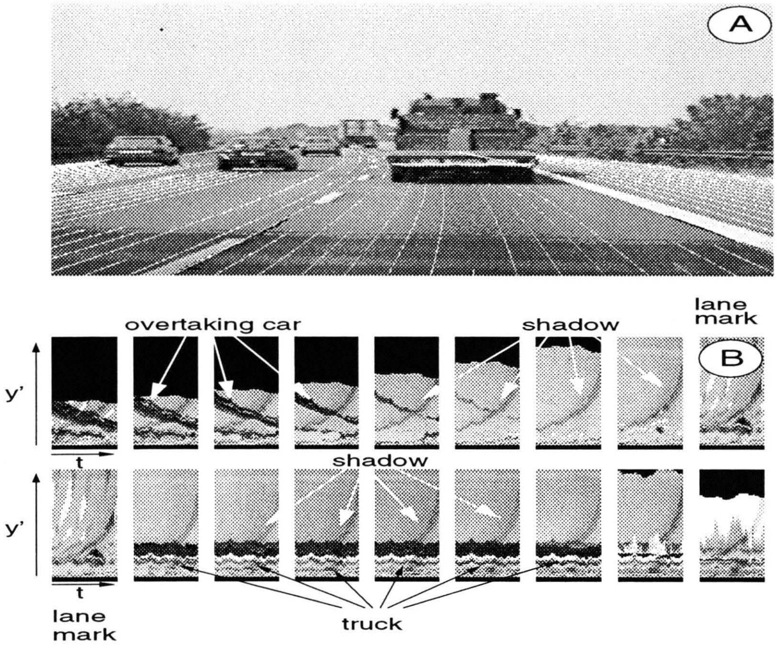

Fig. 8. An image of a motorway scenario with sampling functions (white lines) and the resulting spatiotemporal image. Structures like shadows, lane markings and scene objects can easily by identified. Spatiotemporal images are depicted in *y*-*t*-coordinates.

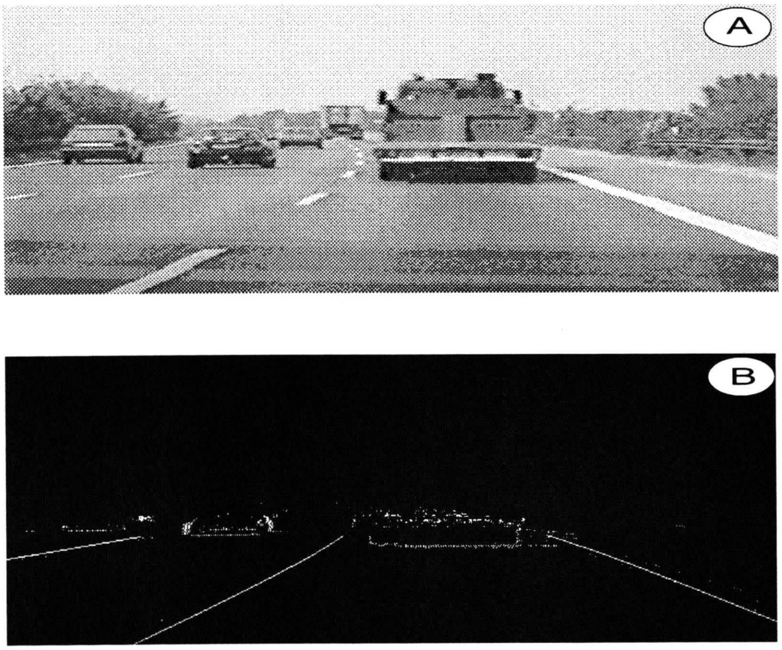

Fig. 9. Segmentation result of the motorway scenario. Fig. A: Original grey value image. Fig. B: The corresponding binary image indicating obstacles or independently moving objects.

tuned for higher absolute velocity values. But the receptive fields of the MT cells are small and so they are able to integrate motion fields like it is necessary for egomotion detection.

First investigations by Tanaka and Saito (1989a; 1989b) showed that Area MST (mediosuperior-temporal) reacts specific to certain motion stimuli like expansion, rotation and contraction. Later Duffy and Wurtz (1991a; 1991b) found cells in MST that where sensitive to combinations of the first mentioned stimuli. In a more recent work Lappe (Lappe, 1996) presented a model that postulates a population coding mechanism for integration done by area MST. He proposes a superposition of MT-stimuli in MST. This signal integration could yield an estimation of the direction of motion.

This model is comparable with the trajectory model introduced in the previous sections since it postulates an integration of local, ambiguous motion measurements which is inherent to the system model. The whole space of possible answers is represented in a neural network. With a general trajectory model such a space can be represented. With the circle model it is possible to integrate external sensor information. For MST this is also possible, since there exist connections between the vestibular system and MST. This integration of additional sensor modalities is a prerequisite to depth estimation and detection of independent motion. A possible model for the solution of the independent motion detection problem has not yet been found in biological systems. If MST solves this problem then there must exist coincident answers in MST for two different motion stimuli. Such an experiment could be designed with the trajectory model.

A more technical question is how an artificial system (for example an artificial neural network) can solve the tasks of area MST avoiding the vast amount of computational effort to calculate and represent spatiotemporal structures.

Duffy C. J. and Wurtz R. H. (1991a), Sensitivity of {MST} neurons to optic flow stimuli. I. A continuum of response selectivity to large field stimuli. J. Neurophysiol. **65**, 1329 –1345.

Duffy C. J. and Wurtz R. H. (1991b), Sensitivity of {MST} neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli, J. Neurophysiol., **65**, 1346–1359.

Gibson J. J. (1950), The Perception of the Visual World. Houghton Mifflin, Boston, MA.

Gillner W. (1997), Bewegungsgekoppelte Segmentierung in technischen und biologischen Systemen. Diss. RWTH Aachen.

Hildreth E. C. (1984), The computation of the velocity field. Proc. R. Soc. London, B **221**,189–220.

Horn B. K. P. and Schunk B. G. (1981), Determining optical flow. Artificial Intelligence **17**, 185–204.

Hubel D. H. and Wiesel T. N. (1977), Functional architecture of macaque monkey visual cortex. Proc. R. Soc. London, B **198**, 1–59.

Maunsell J. H. R. and van Essen D. C. (1983), Functional properties of neurons in the middle temporal area of the macaque monkey. J. Neurophysiol., **49**, 1127–1147.

Lappe M. (1996), Analysis of self-motion by parietal neurons : Parietal Lobe Contributions to Orientation in 3D-Space. (Thier, P., Karnath, H.-O. ed.). Springer, Berlin.

Noll D. (1996), Ein Optimierungsansatz zur Objekterkennung, VDI-Verlag, Düsseldorf.

Tanaka K. and Saito H.-A. (1989a), Analysis of motion of the visual field by direction, expansion/contraction and rotation cells clustered in the dorsal part of the medial temporal area of the macaque monkey. J. Neurophysiol. **62**, 626–641.

Tanaka K. and Saito H.-A. (1989b), Underlying mechanisms of the response specifity of expansion/contraction and rotation cells clustered in the dorsal part of the medial temporal area of the macaque monkey. J. Neurophysiol. **62**, 642–656.