INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460392 - I

SMR/302-6

COLLEGE ON NEUROPHYSICS:
"DEVELOPMENT AND ORGANIZATION OF THE BRAIN"
7 November - 2 December 1988

"Neural Network Strategies for Robot Navigation"

Class 3

Hans Peter MALLOT
Institut für Zoologie III
Gutenberg Universität
Mainz, W. Germany

Proc. nEuro '88 (Paris, June 6 - 9, 1988)
ed. by G. Dreyfus & L. Personnaz

NEURAL NETWORK STRATEGIES FOR ROBOT NAVIGATION *

Hanspeter A. Mallot, Ekkehart Schulze, and Kai Storjohann
Institut für Zoologie III
Johannes Gutenberg-Universität
D-6500 Mainz, West Germany

ABSTRACT

A new paradigm is proposed for deriving from visual information certain features of the 3D structure of the environment that are important for obstacle avoidance. For a moving observer with two translatory degrees of freedom, obstacles are defined as anything raising above the horizontal plane. Detection of such obstacles either via the optical flow induced by egomotion or via stereopsis can be facilitated by the application of a coordinate transform which will be called *inverse perspective mapping*. The formalism of inverse perspective mapping is presented in terms of projective geometry. The method is related to principles of biological information processing such as retinotopic mapping.

1 INTRODUCTION

Recent interest in neural networks is largely due to the information processing capabilities of both real and artificial nets. If we understand why it is that neural networks come in different anatomical types (such as nuclei, cortices, or formatio reticularis) or are endowed with different physiological properties, we might be able to learn something about the computational theory used by the brain. In David Marr's [1] classical distinction of the levels in which information processing can be analysed, this direction of inference, i.e. from the hardware to the computational level is not included. The reason for us to pursue this line of thought is that the brain, unlike a typical computer, is highly adapted to the tasks it performs. By virtue of these structural adaptations, conclusions on computational issues can be derived from anatomical and physiological considerations.

How can one find in neural networks structural adaptations that correspond to computational strategies? There is a well known answer to that question in classical

*Supported by the Deutsche Forschungsgemeinschaft (Grant Se251/30-1), by the German Federal Department of Research and Technology (Grant ITM 8503-C1), and by the NATO Scientific Affairs Division (Collaborative Research Grant 0403/87)

biology: if a certain anatomical, physiological, or, for that matter, behavioural characteristic is correlated to an information processing task, it should show up as an analogue or convergent development in various otherwise unrelated structures subserving this same task. Therefore, comparative investigation of visual systems can be used to identify structural principles that in turn hint to computational mechanisms for vision. In this paper, we present an example of how this approach leads to novel strategies for autonomous navigation. Neural networks are dealt with on the level of architectonics rather than single neurons, since it is this level of organization in which the differences and similarities of brain structures are apparent.

Neural networks processing sensory information in the vertebrate brain are typically organized as cortices, i.e. as stacks of twodimensional layers. We have compiled a preliminary list of *structural principles* of the visual cortex and their possible computational implications [2, 3]. At least two of these are found in the optic tectum and the torus semicircularis [4] as well; the criterion of analogue development is therefore satisfied.

Average anatomy Most cortical neurons (e.g., the pyramidal cells) have strong vertical connectivity with dendrites in the upper layers and both descending and recurrent axons that constitute positive feedback. In the horizontal direction, the connectivity is rather uniform. For a model based on this average anatomy and related physiological results, see [2, 5].

Retinotopic mapping In the horizontally extended cortex, the retinal image is represented in a roughly continuous (neighbourhood preserving) way [6, 7].

Patchy connectivity If connectivity from several areas converges in one area, it segregates to patches, stripes, blobs, etc.

Columnar organization Intrinsic processing varies almost periodically with horizontal position (e.g, orientation columns in the striate cortex).

Each of these features corresponds to a basic operation that appears to be important for the way the brain works. For example, the principles listed above correspond to local intrinsic operations such as spatio-temporal filtering, thresholding, two-dimensional coordinate transforms, the interlacing of multiple images in stripes or patches, and a spatial code for a variety of stimulus characteristics. Taken together, these basic operations can be interpreted as an *instruction set for neural computing*. The knowledge of this instruction set would considerably increase the understanding of biological information processing.

2 APPLYING THE NEURAL INSTRUCTION SET

2.1 Obstacle Avoidance

In previous work [3], we have presented a number of results on the information processing mechanisms associated with one or another of the structural principles listed above. Here, we focus on the implementation of a vision task in terms of the assumed neural instruction set. The selection of a suitable task is again motivated by biological considerations: vision and visual behaviour have followed an evolution

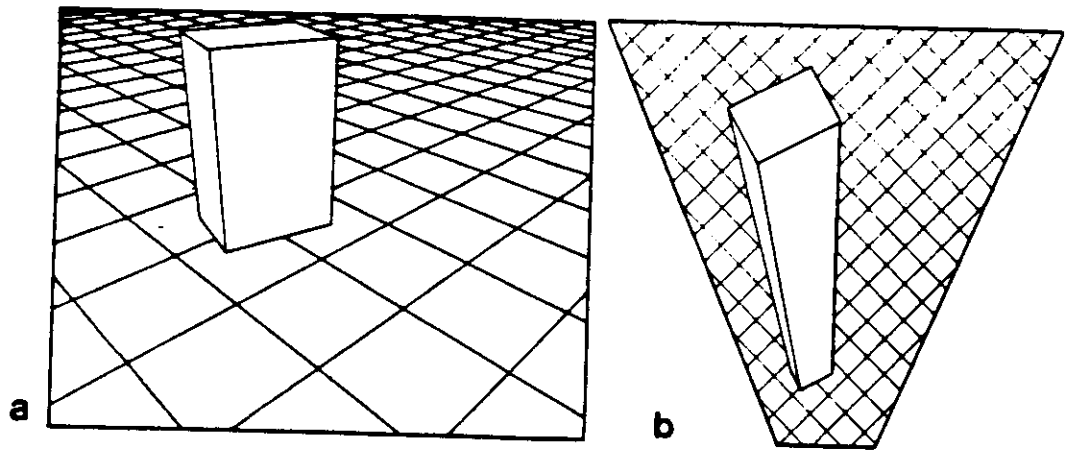

Figure 1: Two perspective views of the same 3D scene. a. Projection onto the image plane, $\mathcal{P}_C$. b. Projection onto the horizontal plane of the world coordinates, $\mathcal{P}_H$. Inverse perspective mapping directly transforms image a into image b and thus undoes the perspective distortion of the plane

from simple but useful to highly sophisticated capabilities. The tasks for which the structural principles have evolved are the simple ones that had to be solved early in the evolution of visual behaviour. Therefore, it seems promising to study the problem of obstacle avoidance and autonomous navigation. A discussion of the sequence in which visual behaviour evolves or should be implemented in a mobile robot, is given by Horridge [8] and Brooks [9], respectively.

Consider an observer with two degrees of freedom to move. As for a typical mammal this means that movement is confined to a horizontal plane. In this situation, which of course is familiar in autonomous navigation as well, an obstacle can be defined as anything raising above this plane. This is a minimal definition of an obstacle which does not require additional information about the nature of the object. We will now show how the proposed neural instruction set can be used to detect obstacles, i.e. elevated points, in motion sequences or stereoscopic image pairs. The proposed algorithms are minimal in the sense that no information processing capacities are wasted to derive superfluous informations.

2.2 Inverse Perspective Mapping

With no obstacles around, both the optical flow generated by translatory egomotion and stereoscopic disparities take a simple form which is determined by the camera (eye) geometry. Any deviation from this expected pattern must be due to an obstacle which is the more important, the larger the deviation is. Unfortunately, variations of optical flow or disparity can be due to both, perspective foreshortening and the 3D structure of the scene. Since we are only interested in the latter part, i.e. the presence of elevated points, we could try to use a coordinate transform to eliminate the effects of perspective (cf. Fig. 1).

Fig. 2 shows how this can be done: consider a point E somewhere in 3D space. Perspective mapping means that we draw a line through this point and the center of projection N and intersect it with the image plane to find the corresponding image point. In order to remove the distortions of the horizontal plane, we now want to undo this perspective map for points in the plane. We define an 'inverse perspective mapping' by the following procedure: for a point E' in the image plane, we trace

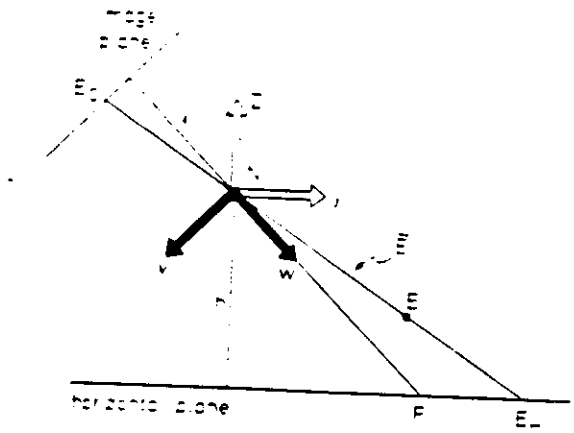

Figure 2: Coordinate systems for inverse perspective mapping. N: center of projection (nodal point). v, w camera coordinate axes. y, z world coordinate axes. The additional axes u and x are perpendicular to the paper plane and are omitted in the Figure. F: fixation point. E: a point in 3D space; E'_C, E'_H: its representations in the image and the horizontal plane; $\tilde{E}$: homogeneous representation of E'.

the associated ray through N towards the horizontal plane. The intersection is the result of inverse perspective mapping applied to the image point E'. If we compose both perspective and inverse perspective, the horizontal plane is mapped unto itself, while elevated parts of the scene are distorted.

Another way to look at inverse perspective mapping is that we assume a fixed relationship between the position of a point in the image and the distance of the corresponding 3D point from the observer. For a void plane, such a relation exists. Next we magnify the parts of the image depicting distant points and compress the parts depicting close points. For the plane, perspective forshortening would thus be undone. However, for an elevated point, the assumed relation between image position and distance does not hold. Therefore, inverse perspective results in a magnification of elevated points, or obstacles.

Mathematically, inverse perspective mapping is a projective collineation [10] which can be written as a linear mapping in homogeneous (or projective) coordinates [12]. For a coordinate system $B := \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ and a plane at distance d from the origin parallel to a, b, perspective mapping is described by :

$$\begin{aligned} \mathcal{P}_{B,d} : \mathbf{R}^3 &\mapsto \mathbf{R}^2 \\ E &\mapsto E'_B = \begin{pmatrix} a' \\ b' \end{pmatrix} := \frac{-d}{(E \cdot \mathbf{c})} \cdot \begin{pmatrix} (E \cdot \mathbf{a}) \\ (E \cdot \mathbf{b}) \end{pmatrix}, \end{aligned} \tag{1}$$

where $(\cdot)$ denotes the inner product. The 3D-line of all points projected to $E'_B = (a', b')^\top$ is given by

$$E'_B \mapsto \tilde{E}_B := \begin{pmatrix} \lambda a' \\ \lambda b' \\ -\lambda d \end{pmatrix} \quad for\ \lambda \in \mathbf{R}. \tag{2}$$

$\tilde{E}_B$ is the homogeneous representation of E in the frame B. By construction, we have $\mathcal{P}_B(\tilde{E}_B) = E'_B$ for all $\lambda \neq 0$.

Next, we introduce a world coordinate system $H := \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ where x and y span the horizontal plane while z points in the upward direction. The camera

model is described by a second coordinate system $C := \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ where $\mathbf{u}$ and $\mathbf{v}$ span the image plane and $\mathbf{w}$ is the optical axis. Both frames share a common origin, the center of projection or nodal point, N at distance h (height) and f (focal length) from the horizontal and image planes, respectively. The coordinate transform from the camera centered system to the world system is described by an orthogonal matrix T which is composed out of the column vectors $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ (cf. Fig. 2).

In inverse perspective mapping, we start with a point E'_C in the image plane and want to find the corresponding point E'_H in the horizontal plane, i.e., the $\mathbf{x}, \mathbf{y}$-plane of the world coordinate system H. In homogeneous coordinates, this is a linear mapping of a point $\tilde{E}_C$ to a point $\tilde{E}_H$, characterized by the orthogonal 3×3 matrix T:

$$\begin{array}{rcl} \tilde{\mathcal{Q}} : \mathbf{R}^3 & \mapsto & \mathbf{R}^3; \quad \tilde{E}_H := T \cdot \tilde{E}_C \\ \begin{pmatrix} \mu x' \\ \mu y' \\ -\mu h \end{pmatrix} & = & \begin{pmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{pmatrix} \cdot \begin{pmatrix} \lambda u' \\ \lambda v' \\ -\lambda f \end{pmatrix} \end{array} \tag{3}$$

Here, $u_1, \ldots, w_3$ denote the components of the axes of the camera frame C expressed in world coordinates. We project $\tilde{E}_H$ onto the horizontal plane (Eq. 1):

$$\begin{array}{rcl} \mathcal{Q} : \mathbf{R}^2 & \mapsto & \mathbf{R}^2 \\ \begin{pmatrix} x' \\ y' \end{pmatrix} & := & \dfrac{-h}{u_3 u' + v_3 v' - w_3 f} \cdot \begin{pmatrix} u_1 u' + v_1 v' - w_1 f \\ u_2 u' + v_2 v' - w_2 f \end{pmatrix}. \end{array} \tag{4}$$

By construction, the composition of perspective and inverse perspective mapping is identical to the projection through the camera nodal point onto the horizontal plane. It is only this projection $\mathcal{Q} \circ \mathcal{P}_C$ that we need to discuss in the applications.

$$\begin{array}{rcl} \mathcal{Q} \circ \mathcal{P}_C(E) & = & \mathcal{P}_H(E) \\ \mathcal{Q} \circ \mathcal{P}_C : \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} & \mapsto & -\dfrac{h}{e_3} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \end{array} \tag{5}$$

Here, $e_1, \ldots, e_3$ denote the components of E, expressed in the world coordinate system H. In the final result $\mathcal{P}_H$, the dependence on the camera coordinate system was removed by the inverse perspective map. In practice $\mathcal{P}_H$ can not be obtained directly, since the original image acquisition used the projection $\mathcal{P}_C$ in the first place.

In applications it is convenient to shift the center of the world coordinate system into the intersection point of the horizontal plane and the optical axis, i.e., the fixation point $F := \mathcal{Q}(0,0)$ if such an intersection exists (i.e, if $w_3 \neq 0$). In homogeneous coordinates, shifts can again be accomplished by linear matrices. We state the result for the example presented in Figs. 1 and 2, where the camera is tilted by an angle $90° + \varphi$ towards the horizontal plane:

$$\begin{array}{rcl} \mathcal{Q}^* : \mathbf{R}^2 & \mapsto & \mathbf{R}^2 \\ \begin{pmatrix} x' \\ y' \end{pmatrix} & = & \dfrac{h}{-v' \cos\varphi + f \sin\varphi} \cdot \begin{pmatrix} u' \\ v'/\sin\varphi \end{pmatrix}. \end{array} \tag{6}$$

Note that $\mathcal{Q}^*(0,0) = (0,0)$, i.e., the map is centered around the fixation point.

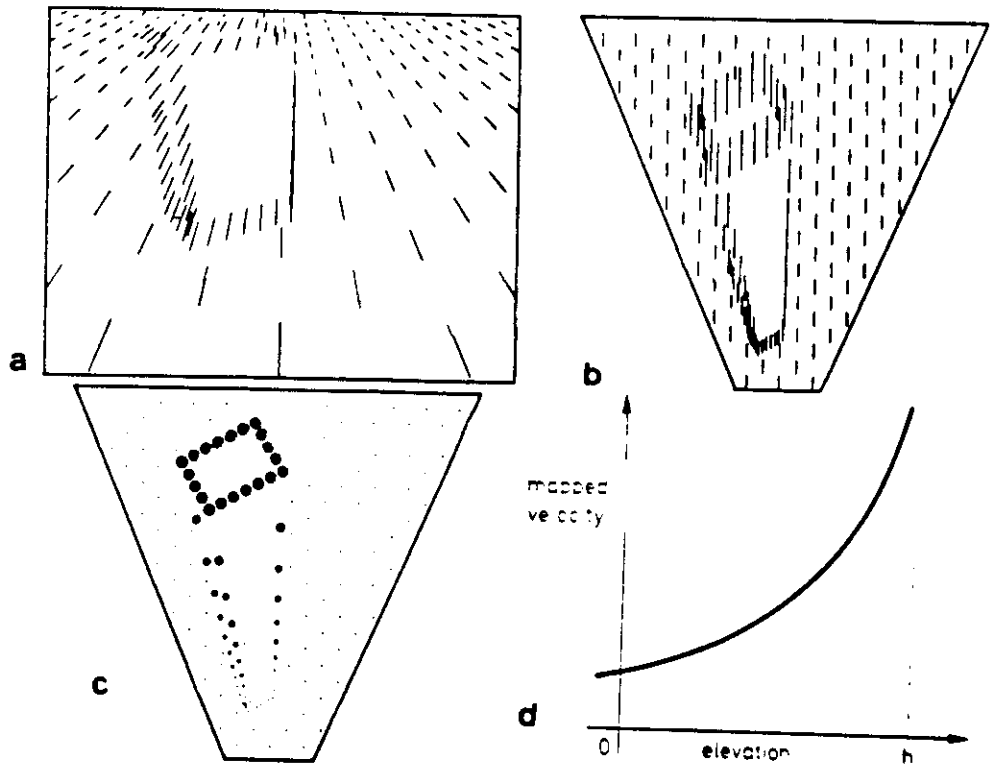

Figure 3: Flow field of the projected image. **a.** without inverse perspective (cf. Fig. 1a) **b.** with inverse perspective (cf. Fig. 1b) **c.** detected obstacle after thresholding **d.** relation of obstacle elevation and motion distortion in the inversely mapped image (cf. Eq. 10)

2.3 Optical Flow

If the camera frame, i.e., the observer is moving in the horizontal plane at a constant speed m, a stationary 3D point E will move relative to the camera frame with a motion vector $dE/dt = -\mathbf{m}$. In the image plane, its motion $\mathbf{m}'_C$ is determined by

$$\mathbf{m}'_C := \frac{d\mathcal{P}_C(E)}{dt} = -J_{\mathcal{P}_C}(E) \cdot \mathbf{m} \tag{7}$$

where $J_{\mathcal{P}_C}$ denotes the Jacobian matrix of the projection (cf. Eq. 1). Fig. 3a shows the resulting motion field for the scene depicted in Fig. 1a and translatory egomotion. If we apply the inverse perspective mapping $\mathcal{Q}$ prior to the computation of the image flow, i.e., if we compute the image flow from the transformed image shown in Fig. 1b, the result is:

$$\begin{aligned} \mathbf{m}'_H &= -J_{\mathcal{Q}\circ\mathcal{P}_C}(E) \cdot \mathbf{m} = -J_{\mathcal{P}_H}(E) \cdot \mathbf{m} \\ &= -h \begin{pmatrix} -1/e_3 & 0 & e_1/e_3^2 \\ 0 & -1/e_3 & e_2/e_3^2 \end{pmatrix} \cdot \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} \end{aligned} \tag{8}$$

The Jacobian was obtained by differentiation of Eq. 5. Since egomotion is bound to the plane, we have $m_3 = 0$. From this, we obtain

$$\mathbf{m}'_H = \frac{h}{e_3} \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \tag{9}$$

$$\|\mathbf{m}'_H\| = \left| \frac{h}{h - elev.} \right| \cdot \|\mathbf{m}\|, \tag{10}$$

Figure 4: Application of the algorithm outlined in Eq. 12 for obstacle avoidance in a custom autonomous vehicle. **a.** Left view ($\mathcal{P}_L$). **b.** Right view ($\mathcal{P}_R$). **c.** Left view (a) mapped to the right camera coordinates by $\mathcal{Q}_{L\to R} \circ \mathcal{P}_L$. **d.** Thresholded difference image of b and c. The obstacle can be clearly distinguished from background textures

where $elev. := h + e_3$ is the elevation of the point E above the ground plane, i.e., its importance as an obstacle. (Note that $e_3 < 0$ in typical cases.) From here, it is easy to detect the obstacle by a local uniform operation, such as an unidirectional motion detector. Thresholding the result to cut off the egomotion vector itself, the obstacle can be made stand out clearly (Fig. 3c). No further information concerning the obstacle is required to trigger some 'avoidance behaviour'.

The relation of the velocity of the mapped flow and the elevation of the obstacle (Eq. 10) is shown in Fig. 3d. The effect scales with the elevation, i.e. the importance of the obstacle, and approaches infinity for objects at the height of the observer. As compared to complex logarithmic mapping [11], inverse perspective has two advantages:

- Not only are the motion vectors in the mapped image unidirectional, for points in in the ground plane they are also of constant length.
- The position of the focus of expansion is not required to apply inverse perspective mapping.

2.4 Stereopsis

The second paradigm applies similar ideas to stereopsis. Again, if no obstacles

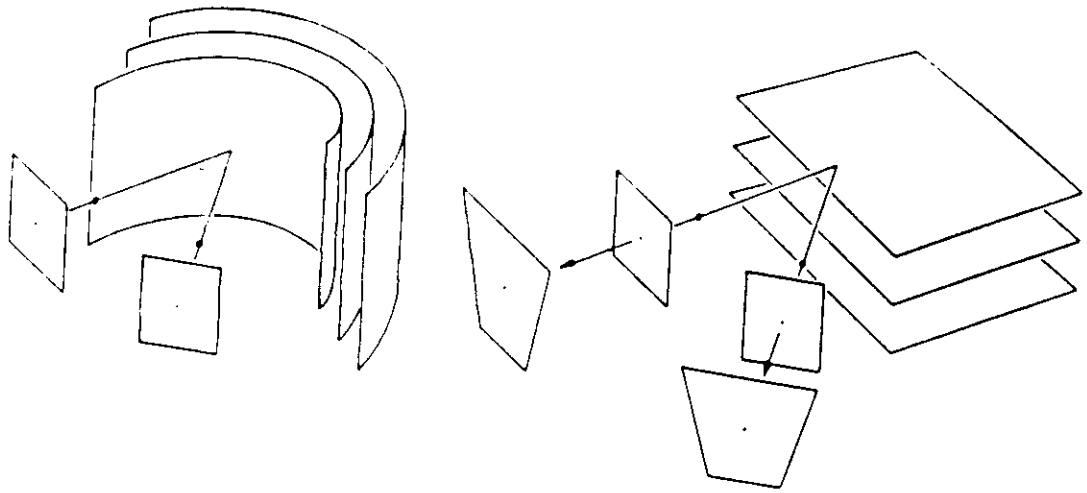

Figure 5: Iso-disparity surfaces for normal and inverse perspective stereopsis. **Left.** In normal stereopsis, two images with slightly differing perspective are acquired. Disparities between corresponding image points translate into distance from the observer. Points with constant disparities correspond to horopter surfaces in space, namely vertical cylinders on the corresponding *Vieth-Müller*-circles. **Right.** If we map both the left and the right image by their inverse perspective mappings, $\mathcal{Q}_L$ and $\mathcal{Q}_R$, disparities in the mapped images correspond directly to elevation above the horizontal plane. Constant disparity in the mapped images translates into positions on isoelevation planes.

were present, the image generated in the left camera would be identical to that in the right one, up to a coordinate transform that includes the inverse perspective mappings for both cameras. Thus, an obstacle can be defined as a deviation from that rule without any higher level information involved. The comparison of the two images is performed via a topographic mapping and a subsequent subtraction which leaves only those regions above threshold that display an obstacle.

Let $L := \{\mathbf{u}_L, \mathbf{v}_L, \mathbf{w}_L\}$ and $R := \{\mathbf{u}_R, \mathbf{v}_R, \mathbf{w}_R\}$ denote the left and right camera coordinate systems, respectively. We assume that both camera nodal points N_L, N_R are at the same height h above the horizontal plane. The separation of the nodal points is given by a vector in the horizontal plane, $(s_1, s_2)^\top$. In homogeneous coordinates, this shift corresponds to a matrix

$$S := \begin{pmatrix} 1 & 0 & s_1/h \\ 0 & 1 & s_2/h \\ 0 & 0 & 1 \end{pmatrix}. \tag{11}$$

The mapping of the image obtained by the left camera onto that obtained by the right can thus be written

$$\begin{aligned} \tilde{\mathcal{Q}}_{L\to R} : \mathbf{R}^3 &\mapsto \mathbf{R}^3 \\ \tilde{E}'_L &\mapsto \tilde{E}'_R = T_R^\top \cdot S \cdot T_L \cdot \tilde{E}'_L. \end{aligned} \tag{12}$$

This algorithm is currently being implemented in a custom autonomous vehicle [13]. The two upper frames in Fig. 4 show the normal, i.e., perspective views of the two cameras. At the lower left, the view of the left camera is distorted in such a way that its image appears under the perspective of the right camera. If

no elevated points are present, the difference of the two frames is small (Fig. 4d). However, if an obstacle comes into sight, the true right view and the one predicted from the left view (under the assumption that no obstacle is present) differ.

While Eq. 12 is convenient in practical applications, since only one image transformation has to be computed, we will use an equivalent description in the sequel, where both image planes are mapped to a common 'cyclopean' plane. We will now derive the disparity of the two images of a 3D point E in this common plane. We choose the origin of the world coordinate system to be at N_L and obtain $N_R = (s_1, s_2, 0)^\top$. Therefore, the point E appears at a position $E - N_R$ for the projection onto the right image plane. Next, we define a disparity vector D in the common horizontal plane by

$$\begin{aligned} D : \mathbf{R}^3 &\mapsto \mathbf{R}^2 \\ D(E) &:= \mathcal{Q}_L \circ \mathcal{P}_L(E) - \mathcal{Q}_R \circ \mathcal{P}_R(E - N_R) - \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}. \end{aligned} \tag{13}$$

From Eq. 5, it follows

$$D = -\frac{h}{e_3}\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} + \frac{h}{e_3}\begin{pmatrix} e_1 - s_1 \\ e_2 - s_2 \end{pmatrix} - \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \tag{14}$$

$$\|D\| = \left|\frac{elev.}{h - elev}\right| \|N_R - N_L\|. \tag{15}$$

As in Eq. 10, $elev. := h + e_3$ is the elevation of the obstacle above the ground plane, i.e. a measure for its importance.

Eq. 15 shows that disparity in the mapped image depends only on the z-component of the imaged point. Surfaces of constant disparity are therefore the horizontal planes. The situation can be made clear by Fig. 5. By choosing other mapping functions, different three-dimensional surfaces can be made 'horopter'-surfaces as well.

2.5 Inverse perspective mapping in biology

We have shown that coordinate transforms such as the inverse perspective mapping are powerful tools for visual information processing. In neurophysiology, on the other hand, pure inverse perspective mapping has not been described as a retinotopic map. However, as Epstein [14] has noted, the deviation of the area 17 map in the cat cortex from a conformal mapping can be accounted for by the inversion of a perspective as it occurs when the cat fixates a point about one meter in front of it on the floor. Interestingly, no such correction is found in animals living in a three-dimensional surround, where it in fact would be useless.

3 CONCLUSION

We propose a low-level detection scheme for obstacles where obstacles are defined as deviations from some expected image. Topographic mappings, as are common in the visual cortex, are used to facilitate the detection of such deviations. In future research, we plan to explore the value of other principles both of brain anatomy and explorative behaviour.

In more general terms, we think that our results give an example of how hardware considerations can have fruitful implications on computational theory if the hardware is adapted to the task it performs. A strong contribution that biology can make to the science of information processing is the formulation of what exactly these structural adaptation are.

References

[1] D. Marr *Vision.* San Francisco (W. H. Freeman), 1982

[2] W. von Seelen, H. A. Mallot, F. Giannakopoulos: Characteristics of neuronal systems in the visual cortex. *Biol. Cybern.* **56**: 37-49, 1987

[3] H. A. Mallot, W. von Seelen: Why cortices? Neural networks for visual information processing. In: J.-P. Ewert & M. A. Arbib (eds.) *Visuomotor integration: Amphibians, comparisons, models, and robots.* New York (Plenum Press), 1988

[4] W. Heiligenberg: this volume

[5] J. Best, H. A. Mallot, K. Krüger, H. R. O. Dinse: this volume

[6] H. A. Mallot: An overall description of retinotopic mapping in the cat's visual cortex areas 17, 18, and 19. *Biol. Cybern.* **52**: 45-51, 1985

[7] H. A. Mallot: Point images, receptive fields, and retinotopic mapping. *Trends in Neurosci.* **10**: 310-311, 1987

[8] G. A. Horridge: The evolution of visual processing and the construction of seeing systems. *Proc. Roy. Soc. Lond.* B **230**: 279-292, 1987

[9] R. Brooks: Autonomous mobile robots. In: W. E. L. Grimson & R. S. Patil (eds.) *AI in the 1980s and beyond.* Cambridge, Mass. (MIT Press), 1987

[10] H. S. M. Coxeter: *Projective geometry. (Second ed.)* New York (Springer Verlag), 1987

[11] R. Jain, S. L. Barlett, N. O'Brien: Motion stereo using ego-motion complex logarithmic mapping. *IEEE Trans. Pattern Anal. Machine Intel.* **PAMI-9**: 356-369, 1987

[12] R. O. Duda, P. E. Hart: *Pattern classification and scene analysis.* New York (John Wiley & Sons), 1973

[13] K. Storjohann, E. Schulze, W. von Seelen: Segmentierung dreidimensionaler Szenen mittels perspektiver Kartierungen. In: H. Bunke, O. Kübler, P. Stucki (eds.): *Mustererkennung 1988, Proc. 10. DAGM-Symposion, Zürich, Switzerland, 1988.* To be published in *Informatik-Fachberichte*, Berlin (Springer-Verlag), 1988

[14] L. I. Epstein: An attempt to explain the differences between the upper and lower halves of the striate cortical map in the cat's field of view. *Biol. Cybern.* **49**: 175-177, 1984