Visual navigation using ego-motion information

Marc Ebner and Andreas Zell
Eberhard-Karls-Universität Tübingen
Wilhelm-Schickard-Institut für Informatik
Arbeitsbereich Rechnerarchitektur
Köstlinstraße 6, 72074 Tübingen, Germany

Abstract

Ego-motion information which is available from the robot status information can be used for visual navigation. A system for visual navigation inside corridors is presented. Experimental results are reported both for a quantitative and a qualitative approach. For the quantitative system ego-motion information is used to determine depth information for interesting points in the image using a motion stereo approach. The qualitative approach is based on properties of the human visual system. In the qualitative system no explicit depth is computed. Instead, sparse optical flow from the left and right peripheral areas is used to achieve centering behavior using a complex logarithmic mapping. Ego-motion information is used to compensate for rotatory camera motions.

1 Motivation

Very often visual systems for mobile robots are constructed without taking into account additional information which might be available. The system only takes a series of images as input and produces the control commands as output. Valuable resources are used to compute information about the ego-motion of the robot directly from the images. We show that the status information (current angular and linear velocity of the robot and position and velocity of the camera) supplied by the robot can aid visual navigation to achieve vision during action. In addition the use of status information decouples the camera from the robot. Thus the camera can be used by an additional control algorithm for object recognition or visual attention.

2 Background

The importance of vision during action is discussed in detail by Sandini et al. [24]. Moravec [18] extracted interesting points from images and calculated distance information using a sliding camera. This information was then used to navigate the robot in its environment. Fossa et al. [7] used three cameras for visual navigation. Two were used to detect obstacles on the floor and one for self-localization. Brady and Wang [2] used stereo and structure from motion to reconstruct the environment from an image sequence for a mobile robot. They argued that it might not be necessary to compute depth explicitly. Tomasi and Kanade [32] have shown that the shape of an environment may be reconstructed from an image sequence using a factorization method without the necessity to compute depth as an intermediate step. In principle it is possible the completely reconstruct the environment from an image sequence, however these approaches are computationally very intensive. Horswill [11] has pointed out that it is not necessary to completely reconstruct the environment of the robot simply to compress this information down to a single number for a task such as corridor following. A vision system only has to compute what is required for the current task which leads to purposive vision [1]. Crespi et al. [6] used a memory-based approach to navigation. The robot is controlled depending on the similarity of current and previously seen images of known attitude and lateral displacement. Sobey [28] used a monocular robot moving in a zigzag motion to estimate the range to the objects in its environment. Neven and Schöner [20] extracted time-to-contact information from optical flow using piecewise linear paths. Jochem and Pomerleau [15] realized a very successful adaptive vehicle control system using a neural net approach. The network is trained by watching a human driver. Košecká [16] used visual servoing to position the robot in front of a door and for wall following.

Much may be learnt by looking at insect vision [10]. Some of this knowledge can be transferred to robot vision [8]. Indeed, several researchers have already built systems for corridor following based on properties of the visual system of bees [4, 25, 26]. Bees achieve centering behavior by balancing the speed of the retinal image in the left and right eye [30, 29]. However these approaches all assume small rotatory movements of the camera or rely on a camera setup such that the rotatory movements of the camera may be neglected.

Our approach differs from the above in that we use a foveated vision system together with status information from the robot to control the robot also during rotatory movements of the camera. We use only a monocular camera with the optical axis facing forward as opposed to sideways. Instead of being based on the bee's visual system we derive our inspiration from properties of the human visual system. An introduction to the human visual system is given by Tovée [33]. Humans could be using the complex logarithmic mapping to derive distance information without explicitly computing depth. Therefore we briefly summarize some of the important results about the complex logarithmic mapping. Jain et al. [13] showed that the radial component in ego-motion complex log space for a translating observer is a measure for the distance of the corresponding point. Vogelgesang et al. [34] also used radial optical flow fields for depth perception. Tistarelli and Sandini discuss the advantages of polar and log-polar mapping for the estimation of time-to-impact from optical flow [31]. Complex logarithmic mapping is pseudo-invariant to size, rotation, and projection scaling [27]. Due to these properties it has been used in a number of different areas such as extraction of moving objects [12, 9] and the centering of peripheral doors [21].

Information about ego-motion of the human head is supplied by the vestibular apparatus [3]. Also, information about the desired motion of the eyes could be used to compensate for the ego-motion of the eyes [22]. By compensating for the rotatory component one can assume a quasi-translating observer. In our case the information about the ego-motion of the robot is coming from the robot's sensors. By using the status information of the robot and the camera instead of their desired motion it is possible to use an independent algorithm for camera control. In this way camera and robot are effectively decoupled.

3 Visual navigation using ego-motion information

Our system is based on the computation of ego-motion from the available status information of the robot. Therefore we first describe how the ego-motion of the camera is computed.

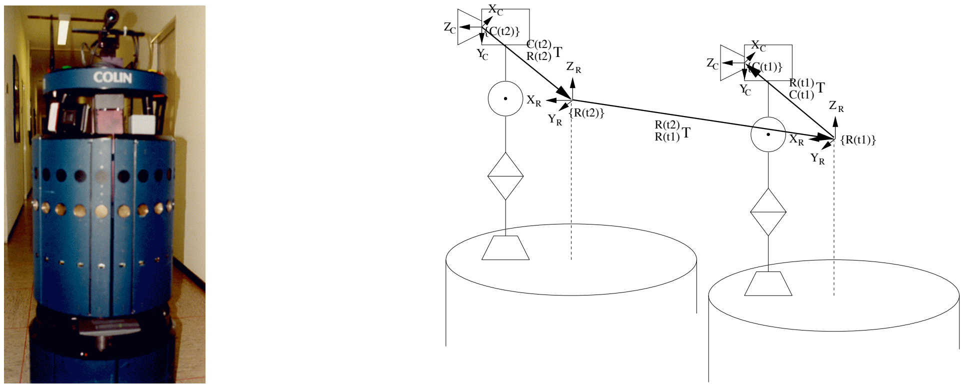

Figure 1: Colin, a Real World Interface B21 robot and movement of camera with transformations for motion due to pan/tilt unit and base of the robot.

3.1 Calculation of ego-motion

Let ${}^{C(t_2)}_{C(t_1)}\mathbf{T}$ be the homogeneous transformation [5] that describes the camera movement that occurred between time t_1 and time t_2.

$$
{}^{C(t_2)}_{C(t_1)}\mathbf{T} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1}
$$

For our robot this transformation is computed as ${}^{C(t_2)}_{C(t_1)}\mathbf{T} = {}^{C(t_2)}_{R(t_2)}\mathbf{T} \cdot {}^{R(t_2)}_{R(t_1)}\mathbf{T} \cdot {}^{R(t_1)}_{C(t_1)}\mathbf{T}$ where ${}^{C(t_2)}_{R(t_2)}\mathbf{T}$ describes the transformation from the camera frame to the robot base at time t_2, ${}^{R(t_2)}_{R(t_1)}\mathbf{T}$ describes the movement of the robot base from time t_1 to time t_2 and ${}^{R(t_1)}_{C(t_1)}\mathbf{T}$ describes the transformation from the robot base to the frame of the camera at time t_1 (figure 1). Each of the individual transformations may be calculated from status information available from the robot. The transformations from the robot base to the camera frame may be calculated using standard manipulator kinematics. Let $\mathbf{R}_Z$ be a rotational transformation about the Z-axis and $\mathbf{D}_X$ a linear transformation along the X-axis of the robot. The transformation describing the robot movement can be approximated as ${}^{R(t_2)}_{R(t_1)}\mathbf{T} = \mathbf{R}_Z(-w(t_2-t_1)) \cdot \mathbf{D}_X(-v(t_2-t_1))$ where w is the angular velocity and v is the linear velocity of the robot.

Let $(x(t_1), y(t_1))$ be the image coordinates of a point at time t_1 then the same point has the following coordinates at time t_2 assuming perspective projection [19]:

$$
x(t_2) = f \frac{r_{11}x(t_1) + r_{12}y(t_1) + f r_{13} + f\frac{t_x}{Z(t_1)}}{r_{31}x(t_1) + r_{32}y(t_1) + f r_{33} + f\frac{t_z}{Z(t_1)}} \tag{2}
$$

$$
y(t_2) = f \frac{r_{21}x(t_1) + r_{22}y(t_1) + f r_{23} + f\frac{t_y}{Z(t_1)}}{r_{31}x(t_1) + r_{32}y(t_1) + f r_{33} + f\frac{t_z}{Z(t_1)}} \tag{3}
$$

where f is the focal length of the camera and $Z(t_1)$ is the distance of the point in the coordinate system of the camera at time t_1.

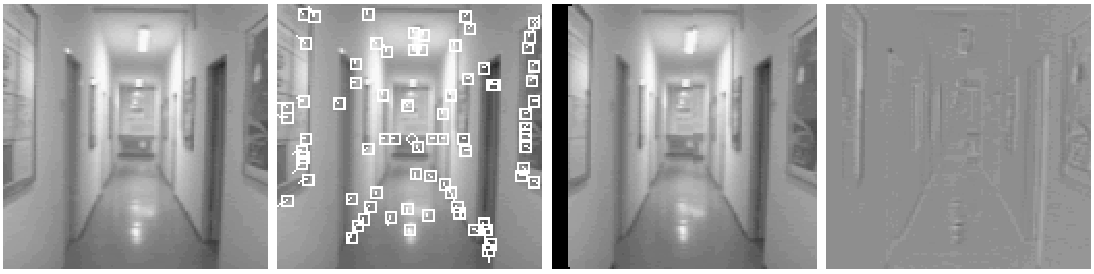

Figure 2: Image, image with sparse optical flow (after compensation for rotatory camera movement), previous image compensated for rotatory camera movement, difference picture between image and previous image compensated for rotatory camera movement.

3.2 A quantitative approach using motion stereo

Before we describe the approach to visual navigation based on properties of the human visual system we describe a quantitative approach which computes depth explicitly. First, interesting points are extracted. Currently we are using the Moravec interest operator [17, 18] to extract the points. Correspondence between two points in successive images is established by matching small areas around the interesting points. Equations 2 and 3 can be solved for $Z(t_1)$. The accuracy of the calculation depends on the amount of translatory motion which occurred during the time the two images were taken. Obviously, depth can only be determined provided that $|x(t_2) - \tilde{x}(t_2)| > 0$ or $|y(t_2) - \tilde{y}(t_2)| > 0$. Where $(\tilde{x}(t_2), \tilde{y}(t_2))$ are the coordinates of the point due to the rotatory motion of the camera. A larger displacement of a point in the image plane due to the translatory motion of the camera provides more accurate depth information. Therefore equation 2 is used to calculate depth if $|x(t_2) - \tilde{x}(t_2)| > |y(t_2) - \tilde{y}(t_2)|$ and equation 3 is used otherwise. This gives us knowledge of the point's 3D coordinates relative to the camera as $X(t_1) = \frac{1}{f}x(t_1)Z(t_1)$ and $Y(t_1) = \frac{1}{f}y(t_1)Z(t_1)$ using perspective projection. Let ${}^{C(t_1)}\mathbf{P} = [X(t_1), Y(t_1), Z(t_1), 1]^T$ be the point in the camera frame. Then the coordinates of the point in the robot frame can be calculated as ${}^{R(t_1)}\mathbf{P} = {}^{R(t_1)}_{C(t_1)}\mathbf{T}^{C(t_1)}\mathbf{P} = [X_r(t_1), Y_r(t_1), Z_r(t_1), 1]^T$. Distance to the points projected onto the floor plane is then calculated according to $d(t_1) = \sqrt{X_r^2(t_1) + Y_r^2(t_1)}$. This gives us quantitative data which can be used for visual navigation.

3.3 Qualitative approach based on properties of the human visual system

We realized a qualitative approach to visual navigation based on properties of the human visual system. Our qualitative system works as follows. We compensate for the rotatory motion of the robot using information about the ego-motion of the robot. For rotatory camera motions one can assume that the distance Z to the objects in view is large compared to the translatory movement of the camera $(t_x, t_y, t_z \ll Z)$. Under this assumption using perspective projection one can compensate for the rotatory motion of the camera by transforming the points $(x(t_1), y(t_1))$ in the image according to [19]:

$$\tilde{x}(t_2) = f\frac{r_{11}x(t_1) + r_{12}y(t_1) + fr_{13}}{r_{31}x(t_1) + r_{32}y(t_1) + fr_{33}} \quad \text{and} \quad \tilde{y}(t_2) = f\frac{r_{21}x(t_1) + r_{22}y(t_1) + fr_{23}}{r_{31}x(t_1) + r_{32}y(t_1) + fr_{33}} \tag{4}$$

Figure 3: Image, image in complex log space and its inverse.

where f is the focal length of the camera. Let $\tilde{I}(t_2)$ be the resulting image.

Sparse optical flow is computed between the predicted image $\tilde{I}(t_2)$ and image $I(t_2)$ for interesting points in the image (figure 2). After the interesting points are matched, the optical flow is transformed to ego-motion complex log space [14]. All points with coordinates (x, y) in the image are transformed to their radial and angular coordinates (r, θ)

$$r = s_1 \log \sqrt{(x - x_{\mathrm{FOE}})^2 + (y - y_{\mathrm{FOE}})^2} \qquad \text{and} \qquad \theta = s_2 \left(\pi + \tan^{-1} \left(-\frac{x - x_{\mathrm{FOE}}}{y - y_{\mathrm{FOE}}} \right) \right) \quad (5)$$

where s_1 and s_2 are scaling factors (figure 3). The transform is taken about the focus of expansion $(x_{\mathrm{FOE}}, y_{\mathrm{FOE}})$ which can be computed from the status information of the robot. This transform can be used to obtain distance information for a translatory moving observer. Jain et al. [13] showed that the radial component of the optical flow in ego-motion complex log space is a measure for the distance of the corresponding point. Using this method one would only be able to let the robot move along piecewise linear paths. Thus control of the robot would be done open loop whenever the robot turns. To close the loop during rotatory movements we use the robot's status information to subtract the component of the optical flow which was induced by rotatory motion of the camera.

3.4 Corridor following behavior

To achieve corridor following behavior we use the median of the data obtained from the sparse optical flow from the left and right peripheral visual areas. The quantitative approach calculates distance information for these points whereas the qualitative approach uses the optical flow directly. To stay in the center of the corridor the steering direction of the robot has to be adjusted such that the data from the left peripheral area and the right peripheral area is about equal. The difference between the data is our error signal used to control the robot.

4 Experiments

Two sets of experiments were performed with a Real World Interface B21 robot. The robot moves with constant forward velocity of $0.3\frac{m}{s}$. First a simple controller was realized to test the performance of the algorithm in the presence of rotations. Depending on the sign of the error signal the robot turns either right or left with constant velocity of $7\frac{\circ}{s}$. This produces oscillatory behavior which can be used to test our robot. In addition to this simple controller

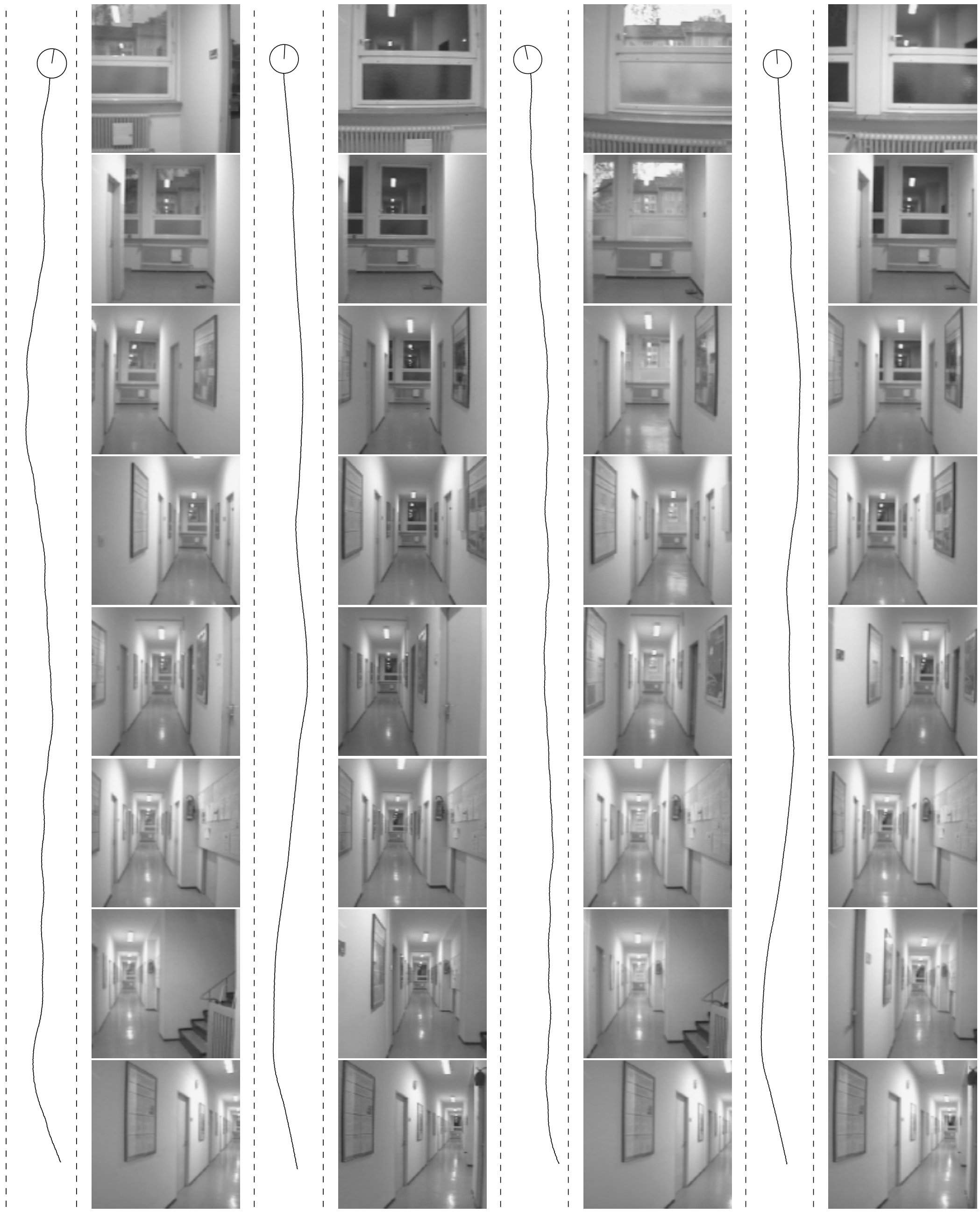

Figure 4: Path of the robot recorded from odometry data. The two on the left were recorded using the quantitative approach, the two on the right were recorded using the qualitative approach. The simple controller was used while the first and the third path was recorded. The PID controller was used while the second and the fourth path was recorded. The images shown along the paths are a subset of the images that were recorded during execution of the algorithm.

a PID controller was realized which eliminates the oscillatory behavior. If features are lacking on one side a turn is made towards the side with features. This is done because it is safer to move towards the direction for which distance information is known [28]. Experimental data from our system for both methods and both controllers is shown in figure 4.

5 Conclusion

A system for visual navigation inside corridors was presented. The quantitative approach uses the known ego-motion to explicitly calculate the depth of interesting points in the environment. The qualitative approach is based on properties of the human visual system and does not explicitly calculate depth. Here, information about ego-motion is used to compensate for rotatory camera motions. Sparse optical flow due to the translatory motion of the camera is transformed into complex log space which is then used to navigate inside a corridor. This shows that the use of ego-motion makes it possible to achieve centering behavior also during rotatory camera motions using the complex logarithmic mapping.

6 Acknowledgements

The first author is currently supported by a scholarship according to the Landesgraduiertenförderungsgesetz. For image processing we have used the Vista software environment [23].

References

[1] Yiannis Aloimonos, editor. *Active Perception.* Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey, 1993.

[2] Michael Brady and Han Wang. Vision for mobile robots. *Phil. Trans. R. Soc. Lond. B*, 337:341–350, 1992.

[3] Roger H. S. Carpenter. *Movements of the Eyes.* Pion Limited, London, second edition, 1988.

[4] David Coombs and Karen Roberts. "bee-bot": using peripheral optical flow to avoid obstacles. In David Casasent, editor, *Intelligent Robots and Computer Vision XI, Proceedings of the Society of Photo-Optical Instrumentation Engineers*, pages 714–721, 1992.

[5] John J. Craig. *Introduction to Robotics: Mechanics and Control.* Addison-Wesley Publishing Company, Reading, Massachusetts, second edition, 1989.

[6] B. Crespi, C. Furlanello, and L. Stringa. A memory-based approach to navigation. *Biological Cybernetics*, 69:385–393, 1993.

[7] M. Fossa, E. Grosso, F. Ferrari, M. Magrassi, G. Sandini, and M. Zapendouski. A visually guided mobile robot acting in indoor environments. In *Proceedings of the IEEE Workshop on Applications of Computer Vision, Palm Springs, USA, 1992*, pages 308–316. IEEE, 1992.

[8] N. Franceschini, J. M. Pichon, and C. Blanes. From insect vision to robot vision. *Phil. Trans. R. Soc. Lond. B*, 337:283–294, 1992.

[9] J. Frazier and R. Nevatia. Detecting moving objects from a moving platform. In *Proceedings of the DARPA Image Understanding Workshop, Pittsburgh, PA*, pages 348–355, 1990.

[10] G. A. Horridge. What can engineers learn from insect vision? *Phil. Trans. R. Soc. Lond. B*, 337:271–282, 1992.

[11] Ian Horswill. A simple, cheap, and robust visual navigation system. In Jean-Arcady Meyer, Herbert L. Roitblat, and Stewart W. Wilson, editors, *From animals to animats 2: Proceedings of the Second International Conference on Simulation of Adaptive Behavior, Honolulu, Hawaii, 1992*, pages 129–136. The MIT Press, 1993.

[12] Ramesh Jain. Segmenation of frame sequences obtained by a moving observer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(5):624–629, September 1984.
[13] Ramesh Jain, Sandra L. Bartlett, and Nancy O'Brian. Motion stereo using ego-motion complex logarithmic mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(3):356–369, 1987.
[14] Ramesh Jain, Rangachar Kasturi, and Brian G. Schunck. *Machine Vision*. McGraw-Hill, Inc., New York, 1995.
[15] Todd Jochem and Dean Pomerleau. Life in the fast lane: The evolution of an adaptive vehicle control system. *AI Magazine*, 17(2):11–50, 1996.
[16] Jana Košecká. Visually guided navigation. *Robotics and Autonomous Systems*, 21:37–50, 1997.
[17] Hans P. Moravec. Towards automatic visual obstacle avoidance. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, Vision–1: p. 584, 1977.
[18] Hans P. Moravec. *Ostacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*. PhD thesis, Computer Science Department, Stanford University, No. STAN-CS-80-813 and AIM-340, September 1980.
[19] Don Murray and Anup Basu. Motion tracking with an active camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5):449–459, May 1994.
[20] Hartmut Neven and Gregor Schöner. Dynamics parametrically controlled by image correlations organize robot navigation. *Biological Cybernetics*, 75:293–307, 1996.
[21] Richard Alan Peters II and Magued Bishay. Centering peripheral features in an indoor environment using a binocular log-polar 4 dof camera head. *Robotics and Autonomous Systems*, 18:271–281, 1996.
[22] Steven Pinker. *How the Mind Works*. W. W. Norton & Company, New York, 1997.
[23] Arthur R. Pope and David G. Lowe. Vista: A software environment for computer vision research. In *Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 768–772. IEEE, 1994.
[24] G. Sandini, F. Gandolfo, E. Grosso, and M. Tistarelli. *Vision During Action*, In Yiannis Aloimonos, editor, *Active Perception*, pages 151–190. Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey, 1993.
[25] José Santos-Victor, Giulio Sandini, Francesca Curotto, and Stefano Garibaldi. Divergent stereo for robot navigation: Learning from bees. In *Proceedings of Computer Vision and Pattern Recognition, New York*, pages 434–439, 1993.
[26] José Santos-Victor, Giulio Sandini, Francesca Curotto, and Stefano Garibaldi. Divergent stereo in autonomous navigation: From bees to robots. *International Journal of Computer Vision*, 14:159–177, 1995.
[27] Eric L. Schwartz. Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. *Vision Research*, 20:645–669, 1980.
[28] P. J. Sobey. Active navigation with a monocular robot. *Biological Cybernetics*, 71:433–440, 1994.
[29] Mandyam V. Srinivasan. Distance perception in insects. *Current Directions in Psychological Science*, 1(1):22–26, February 1992.
[30] Mandyam V. Srinivasan. How bees exploit optic flow: behavioural experiments and neural models. *Phil. Trans. R. Soc. Lond. B*, 337:253–259, 1992.
[31] Massimo Tistarelli and Giulio Sandini. On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):401–410, April 1993.
[32] Carlo Tomasi and Takeo Kanade. Factoring image sequences into shape and motion. In *Proceedings of the IEEE Workshop on Visual Motion, October 7-9, 1991, Nassau Inn, Priceton, New Jersey*, pages 21–28. IEEE Computer Society Press, 1991.
[33] Martin J. Tovée. *An introduction to the visual system*. Cambridge University Press, Cambridge, 1996.
[34] Jens Vogelgesang, Alex Cozzi, and Florentin Wörgötter. A parallel algorithm for depth perception from radial optical flow fields. In Christoph von der Malsburg, Jan C. Vorbrüggen, Werner von Seelen, and Bernhard Sendhoff, editors, *Artificial Neural Networks: 6th international conference; proceedings / ICANN 96*, pages 721–725. Springer-Verlag, 1996.