

Vision-Based Homing with a Panoramic Stereo Sensor

Wolfgang Stürzl and Hanspeter A. Mallot

Universität Tübingen, Zoologisches Institut, Kognitive Neurowissenschaften,
72076 Tübingen, Germany
wolfgang.stuerzl@uni-tuebingen.de
<http://www.uni-tuebingen.de/cog/>

Abstract. A panoramic stereo sensor is presented which enables a Khepera robot to extract geometric landmark information (“disparity signatures”) of its surroundings. By comparing the current panoramic disparity signature to memorized signatures the robot is able to return to already visited places (“homing”). Evaluating a database of panoramic stereo images recorded by the robot we compare homing performance of the proposed disparity-based approach with a solely image-based method considering the size of catchment areas. Although the image-based technique yields larger catchment areas, disparity-based homing achieves a much higher degree of invariance with respect to illumination changes. To increase homing performance, we suggest an extended homing scheme integrating both landmark types.

1 Introduction

There is strong evidence that rodents, see e.g. [1], [2], and also humans, e.g. [3], use memorized geometric cues to return to already visited places. Various mechanisms for visually estimating distances to surrounding objects are known, e.g. motion parallax, stereopsis, shape from shading (see [4] for an overview). It has been claimed that geometric landmark information has certain advantages compared to pure image or texture information because of its higher invariance regarding illumination changes or seasonal variations. In robotics, localization is usually based on distance measurements acquired with active sensors like ultrasonic or laser range finder. In this paper we investigate the use of geometric landmark information for recognizing known places and finding back to them (“homing navigation”) using a small Khepera robot with a passive panoramic stereo sensor (see Fig. 1 a).

2 Panoramic Stereo Sensor

Panoramic visual information has been shown to be advantageous for various navigation tasks in insects, e.g. [5], [6], and in robotics, e.g. [7],[8]. In order to acquire geometric information of the robot’s current place we have built a

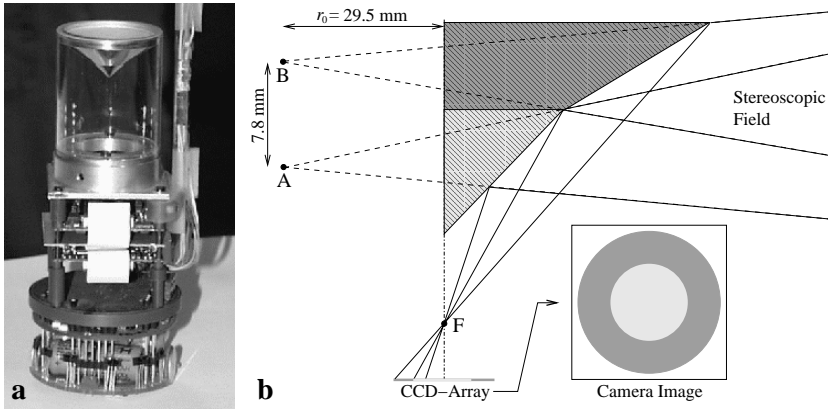


Fig. 1. **a:** Khepera with panoramic stereo camera on top (diameter ≈ 5 cm, height ≈ 13 cm). **b:** Schematic diagram of the bipartite mirror for an axial plane (not to scale). The imaging can be considered as “looking” through two vertically separated points (A, B) which are mirror images of the nodal point of the camera (F). The inset shows the resulting panoramic stereo image: The inner filled circle (light grey) depicts the part imaged through the lower cone; the outer part (dark grey) is imaged through the upper cone.

panoramic stereo sensor. A similar, but much larger omnidirectional stereo system with two cameras and parabolic mirrors has been presented in [9]. A single camera system is described e.g. in [10] where stereo panoramas are created from a video stream captured by a rotating video camera. However, to the authors’ knowledge, none of these panoramic stereo imaging systems were used for navigation purposes on a robot.

Mounted on top of the Khepera, a CCD-camera is directed vertically towards a bipartite conic mirror (see Fig. 1 a). It consists of two conical parts with slightly different slopes (48.5° and 54.5° respectively) yielding an effective vertical stereo base line of ≈ 8 mm (Fig. 1 b).

3 Disparity Signatures of Places

As depicted in Fig. 2 a, raw stereo images, taken by the panoramic stereo sensor, are divided into $N = 72$ sectors (representing a 5° range horizontally). Each sector is subdivided into radial elements resulting in an array of 100 grey-scale pixels $I(x)$, $x = 0, 1, \dots, 99$ (Fig. 2 b).

We have implemented a simple correlation based stereo algorithm to estimate the mean shift d (disparity) of the two image parts by minimization of the matching error (see Fig. 2 b,c),

$$d_{\min} := \arg \min_d E_m(d) \quad (1)$$

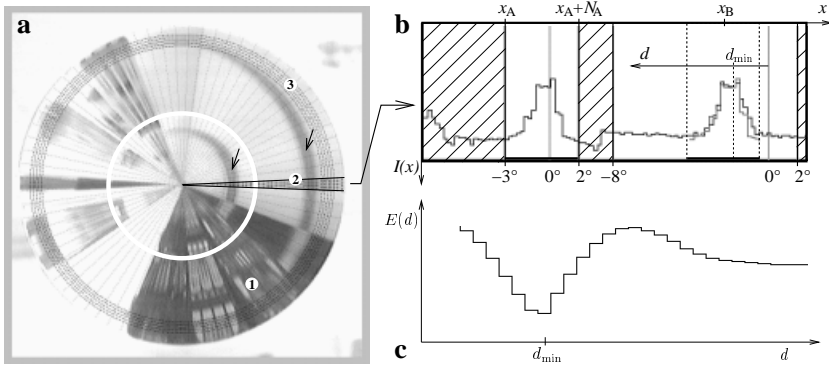


Fig. 2. **a:** Raw stereo image. Images of the toy houses can be seen in the lower right part (1). In the marked sector element (2), a horizontal line on the arena wall is imaged twice (arrows). The marked pixels on the 5 circles (3) are used to compute the image signature for image-based homing as described in Sect. 5. **b:** Grey values corresponding to the sector element in **a**. Linear search for maximum correlation (error function plotted in **c**) between the inner and outer part yields the disparity. The hatched parts are excluded because of low horizontal resolution in the image center (left) and because of imaging distortions at the transition area of the two different slopes of the mirror (middle).

$$E_m(d) := \sum_{x=0}^{N_A-1} (I(x_A + x) - I(x_B - d + x))^2, \quad (2)$$

where $N_A = 20$ is the width of a window taken from the inner image, x_B is the outer image which has zero disparity with respect to x_A (start of inner image). Due to the set-up of the imaging mirrors only a one dimensional correspondence search is needed yielding a disparity range of $N_d = 30$ pixels, i.e. $d \in [0, 29]$. For each estimated disparity $d_{\min,i}$, $i = 0, 1, \dots, N-1$, we compute a quality value $q \in [0, 1]$ depending on uniqueness and reliability of the found match.

After the stereo computation, the current place can be represented in memory by $N = 72$ disparities and their corresponding quality values¹, $(\mathbf{d}, \mathbf{q}) = \{(d_i, q_i), i = 0, 1, \dots, N-1\}$, which we call a “disparity signature” of the considered location. Due to occlusion caused by the cable for video transmission to the host computer (as can be seen in Fig. 1 a), no disparity calculation is possible in a range of 15° and the corresponding three quality values are set to zero.

Using elementary trigonometry, distances r to surrounding objects can be computed according to

$$r(d) \approx \alpha/d - r_0, \quad \alpha \approx 2100 \text{ mm} \times \text{pixel}, \quad (3)$$

¹ To simplify notation we omit the index ‘min’ in the following.

where $r_0 = 29.5$ mm is the distance between the virtual nodal points (A,B) and the robot axis (see Fig. 1 b). The corresponding error,

$$\Delta r(d) = \left| \frac{\partial r(d)}{\partial d} \right| \Delta d \approx \alpha/d^2 \Delta d \approx r(d)^2/\alpha \Delta d , \quad (4)$$

increases approximately with the square of distance.

4 Homing Algorithm

By comparing the current signature with a stored one, it should be possible to return to the place where the signature has been memorized within a certain neighborhood. To investigate this we have extended the homing algorithm described in [7] for the use of disparities:

Using the current disparity signature (\mathbf{d}, \mathbf{q}) , we compute for several possible movements of the robot (turns by angle φ followed by a straight move of length l) predicted signatures $\{(\mathbf{d}^c(\varphi_i, l_i), \mathbf{q}^c(\varphi_i, l_i)), i = 0, 1, \dots, N_c - 1\}$ using (3) and trigonometric calculus. Occlusions are dealt with by setting the corresponding quality values to zero. To avoid wrong disparity predictions due to uncertain disparities (low quality value) we have excluded disparities with $q < 0.7$. The similarities of the predicted signatures to the stored signature at the home position, $(\mathbf{d}^h, \mathbf{q}^h)$, are estimated according to²

$$E_h^d(\varphi_i, l_i) := \min_s \sum_{k=0}^{N-1} q_k^h q_{k_s}^c(\varphi_i, l_i) (d_k^h - d_{k_s}^c(\varphi_i, l_i))^2 / \sum_{k=0}^{N-1} q_k^h q_{k_s}^c(\varphi_i, l_i) , \quad (5)$$

where $k_s := (k + s) \bmod N$, $s = 0, 1, \dots, N - 1$ ³. In the current implementation the considered positions ($N_c = 132$), lie on a hexagonal grid within a radius of approximately 10 cm. Subsequently the robot moves to the position $(\varphi_{\text{opt}}, l_{\text{opt}})$, which minimizes (5). We will call $(\varphi_{\text{opt}}, l_{\text{opt}})$ the ‘‘homing vector’’. To reduce influence of single wrong decisions, the covered distance is limited to $l < 5$ cm. These steps are repeated until the position of highest similarity deviates only marginally from the current position, i.e. $l_{\text{opt}} < l_{\text{thresh}} = 5$ mm.

5 Results

In order to test the panoramic stereo sensor and the proposed homing algorithm systematically, a database consisting of 1250 panoramic stereo images was automatically recorded by the Khepera robot inside a toy house arena with an approximate size of 140 cm \times 120 cm, see Fig. 3 a. Recording positions were approximately on a rectangular 44 \times 36 grid with cell size 2.5 \times 2.5 cm². Minimum

² Since, as can be seen from (4) and Fig. 3, large distances are prone to errors, we compare disparities directly.

³ If additional information, e.g. from odometry, about the robot’s position or orientation relative to the landmark is available the range of φ , l and s can be restricted.

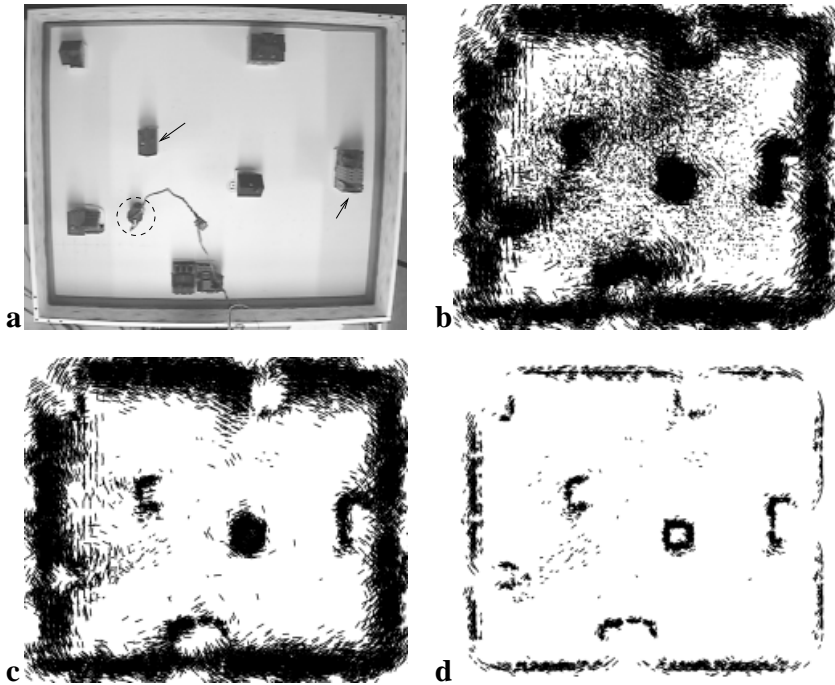


Fig. 3. **a:** The toy-house arena (approximate size of $140 \text{ cm} \times 120 \text{ cm}$) seen from the tracking camera at the ceiling. Arrows indicate non-textured walls of toy-houses; the Khepera robot is marked by the circle. **b-d:** Superposition of object positions measured from all robot positions of the database using (3). **b:** Distances $r(d_k)$ up to 40 cm are shown. **c:** Restricting the plotting to distances, where the quality values of corresponding disparities $q_k > 0.9$, reduces the number of false distance estimations. **d:** Additional restriction to distances $r(d_k) \leq 20$ cm decreases the scattering caused by low distance resolution for larger distances, Eq. (4).

distance of recording positions to walls were ≈ 15 cm, minimum distances to houses were ≈ 5 cm. The accuracy of the tracking system used for the estimation of the robot's pose is ≈ 2 mm (position) and 1.5° (orientation).

5.1 Geometry from Panoramic Stereo

A superposition of object positions calculated from disparity signatures using (3) is shown in Fig. 3 b-d. Increasing the quality threshold significantly reduces the number of wrong distance estimates (Fig. 3 c), although the low resolution for large distances is still obvious⁴. Since the basic geometry of the environment is revealed, the panoramic stereo sensor could be used for map building e.g. based on an occupancy grid.

⁴ We have made no attempt to achieve sub-pixel accuracy from the stereo matching since it is unnecessary for the proposed homing algorithm.

5.2 Catchment Area Size and Homing Accuracy

We have simulated homing runs for all stereo images in the database, starting from every recording position.

Each time the simulated agent has traveled the homing vector, as described in Sect. 4, the distance to the nearest neighboring node in the data base is considered. If the distance exceeds 2 cm, the homing run is stopped (“run into obstacle”), otherwise the agent is placed at the position of the nearest node. If a single homing movement has not brought the agent to a different node the homing run is also stopped (“homing finished”). Hence small homing vectors in the correct direction, which – if occurring repeatedly – could have led the robot closer to the correct position in the real arena, go unnoticed.

We define the *catchment area*, which is usually a coherent region, as the set of all starting nodes, for which the home was reached up to a residual error of 10 cm or less, and their surrounding grid cells.

Comparison with imaged-based homing: In order to investigate homing performance for different signatures, we have also implemented a pure image-based homing scheme, similar to that introduced in [7]: 1D-images consisting of $N = 72$ pixels, each of them representing a 5° region (vertically) around the horizon, were extracted from panoramic stereo images (see Fig.2 a) and normalized using histogram equalization. The three grey values in the range occluded by the video cable were estimated using linear interpolation between the neighboring grey values. To calculate homing movements, (5) was replaced by the sum of squared (grey value) distances (I^h denotes image at home position),

$$E_h^i(\varphi_i, l_i) := \min_s \sum_{k=0}^{N-1} (I_k^h - I_{k_s}^c(\varphi_i, l_i))^2 . \quad (6)$$

Predicted images I^c were computed using the assumption that all surrounding objects are located at the same distance D (D was set to 15 cm).

In Tab. 1 homing performance with respect to catchment area size are compared⁵. Two different illumination conditions are considered: Same illumination during recording and homing (“constant illum.”) and strongly different illumination (“changed illum.”) caused by a position change of the light source from above the arena to the “north”.

As can be seen from Tab. 1, without illumination changes mean catchment area size of solely image-based homing is larger but breaks down under strong illumination changes whereas the disparity-based approach stays approximately at the same performance level. Examples of catchment areas and homing vectors for both illumination conditions are shown in Fig. 4.

⁵ Values for homing runs of the Khepera are in a comparable range.

Table 1. Comparison of disparity and image-based homing at different illumination conditions (simulated homing runs with stereo images recorded by the Khepera): Large catchment areas are desirable. Mean values and standard deviation are given. For the illumination changes also the number of nodes without catchment area is listed (which did not occur for constant illumination).

		disparity-based	image-based	combination
constant illum.	catchment area size (number of grid cells)	198 ± 71	242 ± 59	223 ± 68
changed illum.	catchment area size (number of grid cells)	198 ± 64	23 ± 45	195 ± 68
	number of nodes (out of 1250) without catchment area	0	580	0

Combining image and disparity-based homing: In order to exploit the advantages of both signature types we combine both homing strategies:

$$(\varphi_{\text{opt}}, l_{\text{opt}}) := \arg \max_{(\varphi_i, l_i)} \Pi\left(\frac{E_h^d(\varphi_i, l_i)}{\sigma_h^d}\right) \Pi\left(\frac{E_h^i(\varphi_i, l_i)}{\sigma_h^i}\right) \quad (7)$$

$$\Pi(x) := (1 - \epsilon) \frac{1}{1 + x^2} + \epsilon . \quad (8)$$

For the results listed in Tab. 1 we have used the parameter values $\sigma_h^i = 10^4$, $\sigma_h^d = 1$ and $\epsilon = 0.01$.

6 Concluding Remarks and Future Work

Seeking for invariance properties of landmark information is essential for an efficient spatial representation of an environment. We have shown that robustness regarding illumination changes can be achieved using disparity signatures acquired by a panoramic stereo sensor. Since its effective stereo base line is only ≈ 8 mm, the panoramic stereo sensor is limited to environments where distances are comparatively small (< 1 m). For navigation tasks in office environments we have built another single camera stereo sensor consisting of two separate mirrors with an effective stereo base line of ≈ 36 mm.

Although we have shown that stereoscopic panoramic vision can be used as a rather minimalistic approach for acquiring useful spatial representations, stereo vision in animals is usually restricted to a small field of view. Dealing with a limited field of view for image and disparity based homing will be an integral part of future work. In addition, since many animals have no or only limited stereoscopic vision we will consider the use of optical flow for acquiring geometric landmark information. In both cases, strategies based on limited fields of view rather than on panoramic signatures will become necessary. Future work

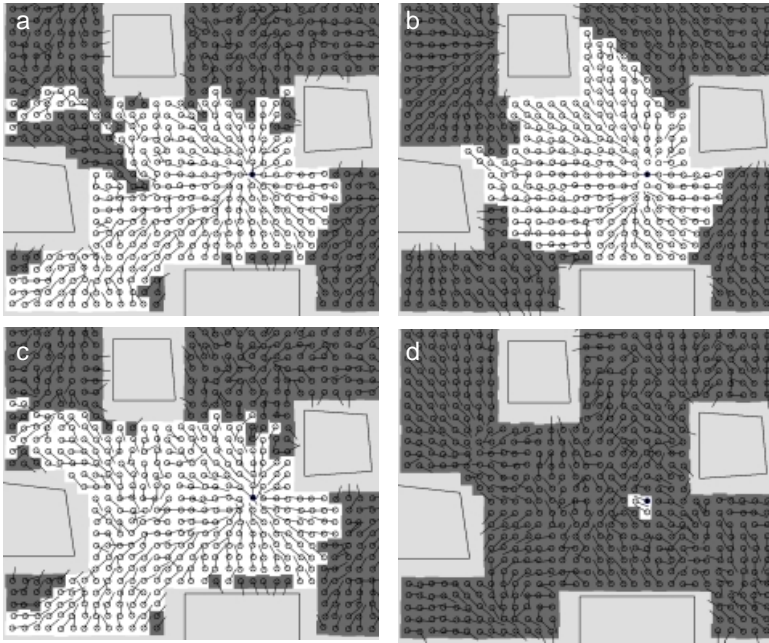


Fig. 4. Comparison of catchment areas (white regions) for a home node (black dot) in the lower left part of the arena: Same illumination during recoding and homing: **a**: disparity-based, **b**: image-based. Different illumination during recoding and homing: **c**: disparity-based, **d**: image-based (catchment area consists of only 3 grid cells). Houses are marked by polygons, small circles mark recording positions of images in the data base. Orientation and length (limited to 3 cm) of the homing vector is depicted as a line starting from the center of each node.

will therefore use a view based approach to spatial memory, as was introduced by [11].

References

1. Cheng, K.: A purely geometric module in the rat's spatial representation. *Cognition* **23** (1986) 149–178
2. Collett, T.S., Cartwright, B.A., Smith, B.A.: Landmark learning and visuo-spatial memories in gerbils. *J. Comp. Physiol. A* **158** (1986) 835–851
3. Hermer, L., Spelke, E.S.: A geometric process for spatial reorientation in young children. *Nature* **370** (1994) 57–59
4. Mallot, H.: *Computational Vision. Information Processing in Perception and Visual Behavior*. MIT Press, Cambridge, MA (2000)
5. Cartwright, B.A., Collett, T.S.: Landmark learning in bees. *J. Comp. Physiol. A* **151** (1983) 521–543
6. Dahmen, H., Wüst, R.W., Zeil, J.: Extracting egomotion parameters from optic flow: principal limits for animals and machines. In Srinivasan, M.V., Venkatesh, S., eds.: *From living eyes to seeing machines*. Oxford University Press (1997) 174–198

7. Franz, M.O., Schölkopf, B., Mallot, H.A., Bühlhoff, H.H.: Where did I take that snapshot? Scene-based homing by image matching. *Biol. Cybern.* **79** (1998) 191–202
8. Adorni, G., Cagnoni, S., Enderle, S., Kraetzschmar, G.K., Mordonini, M., Plagge, M., Ritter, M., Sablatnög, S., Zell, A.: Vision-based localization for mobile robots. *Robotics and Autonomous Systems* **36** (2001) 103–119
9. Gluckman, J., Nayar, S., Thorek, K.: Real-time omnidirectional and panoramic stereo. In: *DARPA Image Understanding Workshop*. (1998) 299–303
10. Peleg, S., Ben-Ezra, M.: Stereo panorama with a single camera. In: *IEEE Conference on Computer Vision and Pattern Recognition*. (1999) 395–401
11. Schölkopf, B., Mallot, H.A.: View-based cognitive mapping and path planning. *Adaptive Behavior* **3** (1995) 311–348