Homing by Parameterized Scene Matching

Matthias O. Franz Bernhard Schölkopf Heinrich H. Bülthoff
Max–Planck–Institut für biologische Kybernetik
Spemannstraße 38, D-72076 Tübingen, Germany
franz/bs/hhb@mpik-tueb.mpg.de

Abstract

In visual homing tasks, animals as well as robots can compute their movements from the current view and a snapshot taken at a home position. Solving this problem exactly would require knowledge about the distances to visible landmarks, information, which is not directly available to passive vision systems. We propose a homing scheme that dispenses with accurate distance information by using parameterized disparity fields. These are obtained from an approximation that incorporates prior knowledge about perspective distortions of the visual environment. A mathematical analysis proves that the approximation does not prevent the scheme from approaching the goal with arbitrary accuracy. Mobile robot experiments are used to demonstrate the practical feasibility of the approach.

1 Introduction

For many animal species it is vital to be able to find their way back to a shelter or to a rewarding food source. In particular, flying animals cannot rely on idiothetic information for this task, as they are subject to drift. Thus they have to use external information, often provided by *vision*. A location may be identified visually using one of two methods: first, by association with an image *of* the location (recorded while approaching or leaving it), or second, by association with an image of the panorama as seen *from* the location. These two methods depend on the visual characteristics of the location, and determine how such a snapshot can be used to recover its associated spatial position: if the location itself is marked by salient optical features, these can be *tracked* until the goal is reached (e.g. Collett 1996). If there are no such features, the goal direction after a displacement has to be inferred by comparing the current visual input to the snapshot: image regions in the direction of the displacement are expanded while the image in the goal direction is contracted. Driving into the direction of maximal image contraction finally leads to the goal position (cf. Figure 1).

Note that even though neither of the two possible approaches necessarily requires recognition mechanisms, our distinction already foreshadows another distinction which is known in the domain of recognition, namely the one between object recognition and scene recognition (see e.g. Tarr & Bülthoff, 1996). As in our case, these two domains differ in that the former deals with something localized in space that the observer is not part of, whereas the latter deals with something non–local, surrounding the observer. Inspired by this distinction, we shall use the term *scene–based homing* to refer to visual homing strategies which make use of the whole scene, rather than tracking single objects.

A number of experiments have shown that invertebrates such as bees or ants are able to pinpoint a location defined by an array of nearby landmarks (see Collett 1992 for a review). Apparently, these insects search for their goal at places where the retinal image forms the best match to a memorized snapshot. Cartwright and Collett (1983) have put forward the hypothesis that bees might be able to actively extract the goal direction by a mechanism using the azimuth and size change of visible objects after a displacement.

In the present work, we want to approach homing from a different viewpoint: Rather than proposing a model or mechanisms underlying actual insect behaviour we will focus in this study on the problems that any agent has to face when using perspective distortion to infer the goal direction from a snapshot. To that end we use real robots to avoid the idealizations one necessarily has to accept when simulating an agent and its environment. We limit ourselves to mechanisms that are computationally inexpensive, both to afford robotic implementation and to avoid overly complex explanations.

In the next section, we give a mathematical description of the basic task, followed by the introduction of a new algorithm (Section 3) that is able to cope with some of the shortfalls of previous approaches. Section 4 describes our implementation on a mobile robot and presents experimental results. We conclude our study with a discussion of the results and relate them to previous approaches taken by researchers in biology and robotics.

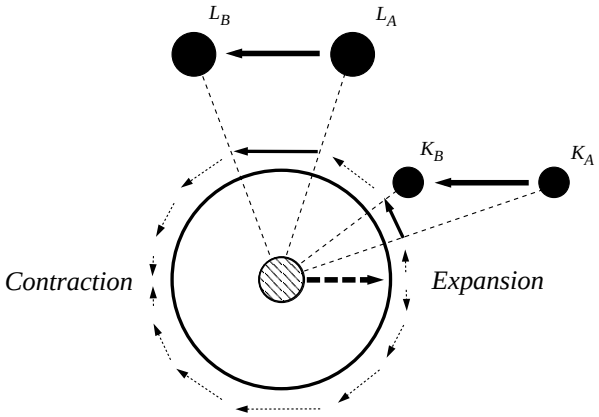

Figure 1: By displacing the sensor ring along the dashed arrow, the landmarks L_A and K_A are shifted to new positions L_B and K_B relative to the sensor. The image region in the direction of displacement expands, while the opposite region contracts. If the landmarks are isotropically distributed, then the vector sum of the disparities always points towards the starting point.

2 Inferring the home direction from perspective distortion

To characterize the basic task mathematically, we start by giving some definitions which will be used throughout the paper. As an idealized model of an agent we choose a mobile sensor ring measuring the surrounding light intensity. If the allowed movements of the sensor ring are restricted to two dimensions then a ring parallel to the movement plane suffices in principle, to determine the relevant motion parameters. The agent is able to record a 360° view of the surrounding panorama as a snapshot. The position of an image point on the sensor ring is denoted by the angle θ. All points in the environment giving rise to identifiable points in the image are called landmarks. This should not be confused with the usual notion of a landmark as a physical object. In our terminology, a visible object may contain several landmarks.

Suppose we displace the sensor ring in direction α at position A by a distance d to point B and change its orientation by the angle ψ (see Fig.2). As a consequence, the image of landmark L at distance r is shifted from θ to a new position $\theta + \delta$ (assuming a static environment). From the triangle ALB in Fig.2 we obtain

$$\frac{r}{d} = \frac{\sin(\theta - \alpha + \psi + \delta)}{\sin(\psi + \delta)}. \qquad (1)$$

This relation can be used to compute the direction $\beta = \alpha + \pi$ back to the starting position A from the change δ in the landmark position (the *disparity*), if r/d and ψ are known.

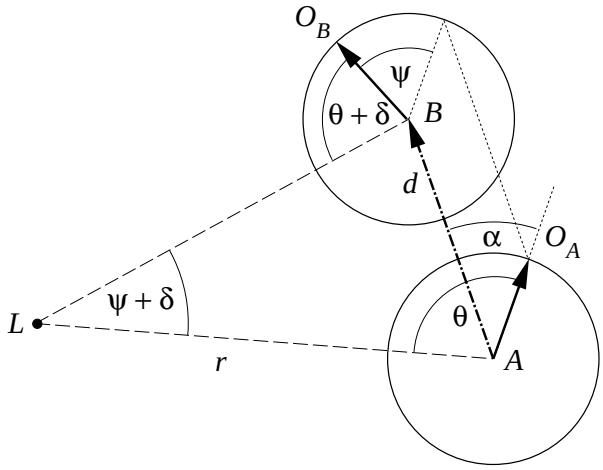

Figure 2: Displacing a ring sensor from A to B in direction α (with respect to sensor's initial orientation O_A) and rotating it by ψ leads to a change δ of the viewing angle of a landmark L.

Before relation Eq. (1) can be applied for homing tasks, two basic problems have to be solved:

1. In order to compute the disparity δ, a correspondence between image points in the snapshot and in the current view must be established. Since the displacement is the result of an arbitrary movement, we are not allowed to assume that image points belonging to the same landmark occur at similar locations in both the snapshot and current view.

2. A visually navigating agent has no access to the actual distance r of the landmark at L. Therefore, this lack of knowledge must be compensated by some additional assumption about the distance distribution of possible landmarks in the environment.

The computation of disparities requires techniques used for optical flow analysis, or mechanisms for object recognition that allow to identify regions in different images as belonging to the same object. The problem can be alleviated if the agent either knows its orientation with respect to an external reference (Cartwright & Collett 1983, Wittmann 1995) or always keeps a constant orientation (Hong et al. 1991, Röfer 1995).

In previous approaches, two basic assumptions have been used to compensate for the lack of distance know/-ledge:

Isotropic landmark distribution. If the surrounding landmarks are distributed isotropically (i.e. frequency and distance of landmarks does not depend on the viewing direction), one obtains the correct goal direction by summing over all disparity vectors along the sensor ring, since all disparity components orthogonal to direction of the displacement cancel each other (see Figure 1).

Weak perspective. The weak perspective projection approach is based on the assumption that the projected objects are sufficiently far away from the sensor, so that the distance differences of the individual landmarks belonging to a single object become negligible.[1] The individual disparities of the landmarks can be measured while the unknowns α, ψ and r/d in Eq. (1) remain the same all over the object. Thus, an object containing at least three landmarks suffices to determine the home direction. To use this approximation, the agent must be able to segment image regions belonging to an object from the background, and to identify the object in different images.

In the following section, we will introduce an algorithm based on an approximation which we call *equal distance assumption*. The surrounding landmarks are assumed to have approximately identical distances from the location of the snapshot. Similar to the approximation of an isotropic landmark distribution, a homing algorithm based on this assumption does not need any object recognition mechanisms and can rely on optical flow techniques. Additionally, this approach provides constraints for the computation of the disparity field which will be used in our algorithm. As the equal distance assumption is, in the strictest sense, unrealistic, we will show in Section 3.2 that the effect of the resulting errors on homing performance remain small.

3 Homing with parameterized disparity fields

3.1 A matched filter based on the equal distance assumption

Before applying the equal distance assumption we will convert Eq. (1) into a suitable form for the subsequent analysis. Solving Eq. (1) for ψ, we have

$$\tan(\psi + \delta) = \frac{d \sin(\theta - \alpha)}{r - d \cos(\theta - \alpha)}. \qquad (2)$$

We assume a typical landmark distance R and denote the deviation from it by r' so that $r = R + r'$. This leads to

$$\tan(\psi + \delta) = \frac{\frac{d}{R} \sin(\theta - \alpha)}{1 + \frac{r'}{R} - \frac{d}{R} \cos(\theta - \alpha)}. \qquad (3)$$

By setting $\epsilon = r'/R$, $\nu = d/R$ and $\gamma = \theta - \alpha$ to simplify the notation, we obtain from Eq. (3)

$$\delta = \arctan\left(\frac{\nu \sin\gamma}{1 + \epsilon - \nu \cos\gamma}\right) - \psi. \qquad (4)$$

[1] Note that the second assumption of weak perspective is automatically satisfied for a sensor ring: the object is always near the optical axis of some sensors.

The cases $\nu = 0$, $\epsilon = -1$ and $\gamma = 0, \pi$, $\nu = 1 + \epsilon$ have to be excluded which means that the agent is not allowed to occupy the same position as a landmark while homing or taking a snapshot.

We now apply the equal distance assumption by neglecting the individual distance differences ϵ of the surrounding landmarks. The resulting expression

$$\delta = \arctan\left(\frac{\nu \sin\gamma}{1 - \nu \cos\gamma}\right) - \psi. \qquad (5)$$

describes the disparity field when all landmarks are located at a distance R from the starting position. The disparity field is completely determined by only three parameters α, ψ and ν. This allows us to estimate the real disparity field and the home direction using the following algorithm:

1. For all parameter values of α, ψ and ν the current view is distorted by shifting the image positions θ of the single pixels according to Eq. (5). The result of this procedure are new images that would have been obtained if the sensor was displaced in an environment where the constant distance assumption was perfectly valid.

2. The generated images are compared to a snapshot taken at the home position. To measure the degree of match between both images we use the Euclidian distance between the grey values at each pixel. The best match is produced by a disparity field which reconstructs the home view as accurately as possible.

3. In a final step, the parameter value of α leading to the best match is selected as an estimate $\beta = \alpha + \pi$ of the home direction.[2]

Note that in order to determine Eq. (5) completely, at least three landmarks must be visible. Otherwise, the home direction can only be estimated if additional information sources such as compasses or odometers are available. The parameterized disparity field $\delta(\theta)|_{\epsilon=0}$ can be interpreted as a matched filter in the sense that the parameter set that reproduces best the actual disparity field can be assumed to approximate the real one. Since the direction of the displacement α is one of the parameters, the best matching disparity immediately gives the goal direction. Similar motion templates for determining egomotion parameters from given optical flow fields have been described for the visual system of the blowfly *Calliphora* (Krapp & Hengstenberg 1996), and theoretically by Nelson & Aloimonos (1988).

Although the equal distance assumption is hardly ever valid in its strictest sense, the estimate of the disparity field is quite robust as will be demonstrated in the next section.

[2] The relative distance ν obtained by this matching process is generally not the mean relative distance of the surrounding landmarks, but a weighted average according to the disparity caused by each individual landmark.

3.2 Error due to the equal distance assumption

Let

$$E(\epsilon,\nu) := \delta(\epsilon,\nu) - \delta(0,\nu) \tag{6}$$

denote the error in the disparity δ due to neglecting the deviation of the landmark distance r' from the averaged distance R. We want to show that for each $\epsilon > -1$, fixed γ, and any desired accuracy bound $E_0 > 0$, there exists a ν_0 such that $\nu < \nu_0$ implies $|E(\epsilon,\nu)| < E_0$. In other words, even if the equal distance approximation does not hold, we can reach any desired accuracy level, provided that we are close enough to the goal.

For the proof, we use the Taylor expansion of E (cf. Eqs. (6), (5)) in ϵ,

$$\begin{aligned} E(\epsilon,\nu) &= \frac{-\nu\sin\gamma}{1+\nu^2-2\nu\cos\gamma}\cdot\epsilon \\ &+ \frac{(1+\theta\epsilon-\nu\cos\gamma)\nu\sin\gamma}{\left(\nu^2\sin^2\gamma+(1+\theta\epsilon-\nu\cos\gamma)^2\right)^2}\cdot\epsilon^2, \end{aligned} \tag{7}$$

where $0 < \vartheta < 1$. Clearly, the first term tends to 0 as ν tends to 0. If $\epsilon \geq 0$, then the second term can also easily be seen to tend to 0: its denominator will not tend to zero, while the denumerator always does. If $\epsilon < 0$, we need to take a closer look at the denominator: First note that even though ϑ might depend on ν, we know that for all γ,

$$1+\theta\epsilon > 1+\epsilon > 0. \tag{8}$$

Thus we can choose $\nu_0 > 0$ such that for $\nu < \nu_0$,

$$\begin{aligned} 1+\theta\epsilon-\nu\cos\gamma &> 1+\epsilon-\nu\cos\gamma \\ &\geq \kappa > 0 \end{aligned} \tag{9}$$

for some suitable κ. Hence for $\nu < \nu_0$ the denominator is bounded from below by κ^4, whereas the numerator approaches 0, which completes the proof.

Therefore, for every snapshot containing at least three landmarks, there exists a catchment area in which the location of the snapshot can be approached arbitrarily closely. In practise, the catchment areas tend to be larger than one might expect from the equal distance approximation, as there are several factors which effectively select a ring–shaped area for the matching procedure. First, the error induced by an infinitely distant point is relatively small, compared to disparities generated by nearby landmarks. Second, very close points will not have an effect as adverse as might be expected from their large disparities, since commonly used obstacle avoidance systems make them less likely to occur. In addition, the vision system's limited depth of field will cause both very close and very distant landmarks to be blurred and reduced in contrast, which decreases their effect on the matching procedure. This serendipitous bonus makes our approximation all the more suitable for real world applications.

3.3 Limitation of accuracy by sensor noise

As we have shown in the previous section, the equal distance assumption does not influence the spatial accuracy of the homing scheme. Therefore, the primary limiting factor is the pixel and quantization noise of the sensor ring. In the following, we will determine the influence of noise on the spatial accuracy, namely the smallest achievable relative distance from the goal $\Delta\nu$.

We assume that the intensity distribution $h(\theta)$ sampled by the sensor ring is low-pass filtered in a subsequent processing stage so that the derivative of the intensity distribution $h'(\theta)$ is well defined for all sensor coordinates θ and spatial aliasing effects are eliminated.

The variance of the noise in the intensity distribution is given by $\Delta h^2_{noise} = \sigma_{sensor} + \Delta^2/12$, where σ_{sensor} is the sensor noise and the second term results from the quantization error (here, Δ is the step size of the quantizer, cf. Oppenheim & Schafer 1989). As a consequence, the maximally resolvable intensity change is $2\Delta h_{noise}$ according to the usual reliability criterion for communication systems which is analogous to assuming that the threshold signal to noise ratio is unity (Goldman 1953).

A small displacement of the sensor ring from the location of the snapshot induces a small change $\Delta\theta$ in the position of the image features. The resulting change of the detected intensity distribution at θ is, to a first order approximation,

$$\Delta h(\theta) = h'(\theta)\cdot\Delta\theta. \tag{10}$$

From (10), we find that the maximally resolvable image displacement is

$$\Delta\theta = \frac{2}{h'(\theta)}\Delta h_{noise}. \tag{11}$$

From Eq. (5), we obtain the expression

$$\frac{\partial\nu}{\partial\delta} = \frac{(1+\epsilon)\sin\gamma}{\sin^2(\gamma+\psi+\delta)}, \tag{12}$$

so that the maximal spatial accuracy is given by

$$\begin{aligned} \Delta\nu &= \nu(\delta+\Delta\theta)-\nu(\delta) \\ &\approx \frac{\partial\nu}{\partial\delta}\cdot\Delta\theta \\ &= \frac{2(1+\epsilon)\sin\gamma}{h'(\theta)\sin^2(\gamma+\psi+\delta)}\Delta h_{noise}. \end{aligned} \tag{13}$$

This shows that for extreme noise levels, low contrast and sensor positions very close to landmarks, $\Delta\nu$ may become larger than the catchment area, so that in these cases the described homing scheme is not applicable. Note, that the above limitation is derived for only one landmark. When more landmarks are visible, the spatial accuracy becomes higher due to the effect of statistical averaging.

4 Results

4.1 Experimental setup

Most experiments were conducted in an arena with dimensions 118 × 102 cm. Visual cues were provided by model houses and landmarks surrounding the arena (see Fig. 3). In these experiments, we used a modified Khepera miniature robot (Fig. 4) connected to an SGI Indy workstation via a serial and video transmission cable. Our scheme was also tested in a real office environment on a six-wheeled platform with wireless modem and video transmission.

The imaging system on the robot comprises a conical mirror mounted above a small video camera which points up to the center of the cone (Fig. 4). This configuration allows for a 360° horizontal field of view extending from 10° below to 10° above the horizon. A similar imaging technique was used by Chahl and Srinivasan (1996) and Yagi, Nishizawa, & Yachida (1995). The video image was sampled on four circles along the horizon with a resolution of 2.5° and averaged vertically to provide robustness against inaccuracies in the imaging system and tilt of the robot platform. In a subsequent processing stage, a spatiotemporal Wiener lowpass filter (e.g. Goldman, 1953) was applied to the resulting one-dimensional array. To remove changes in the illumination, the average background component was subtracted and, in a final step, the contrast of the array was maximized via histogram equalization. The movement commands calculated from this data were transmitted back to the robot using a serial data link with a maximal transmission rate of 12 commands per second for the Khepera (5 for the wireless modem).

The Khepera's position was tracked with a colour camera mounted above the arena, tuned to a red marker attached to the robot. Position and image data were recorded with a time stamp and synchronized offline.

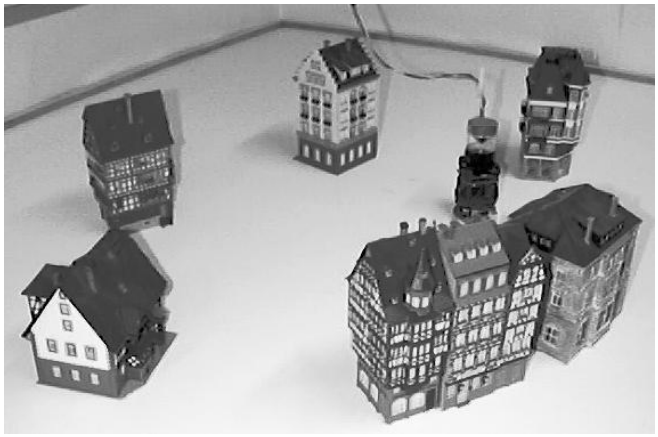

Figure 3: Test arena with toy houses, used in the homing experiments (see Sec. 4.1).

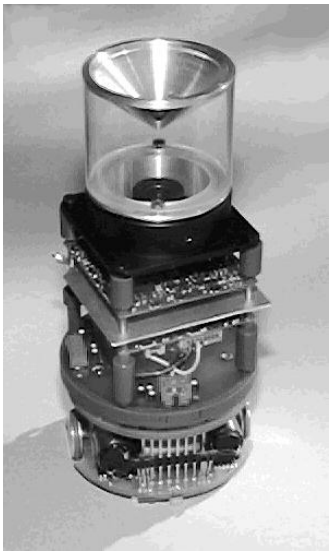

Figure 4: Khepera^{TM} robot with camera module and custom made conical mirror, which permits sampling of the environment over 360°, in a range of ±10° about the horizon.

4.2 Performance of the homing scheme

The viability of our approach was tested in an experiment with the Khepera robot in the "toy house" arena. During the homing runs, the robot computed the home direction every 83 ms relative to the current driving direction. The home vector was used to set the new driving direction. For 20 different home positions, the robot was displaced relative to each home position by distances in the range of 5 to 25 cm in random directions. A trial was counted as a success if the robot reached the home position within a radius of 1 cm without colliding with an obstacle or exceeding a search time limit of 30 sec-

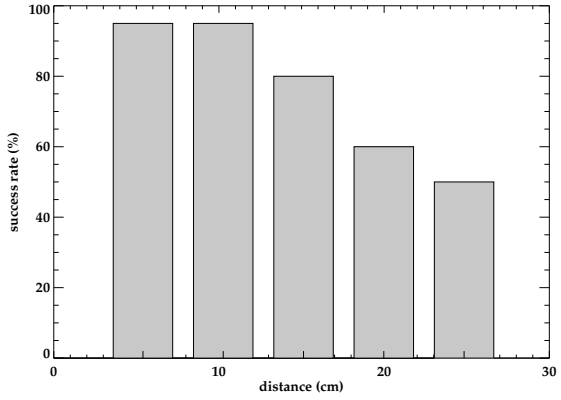

Figure 5: Success rate for 100 homing runs, with starting distances between 5 and 25 cm.

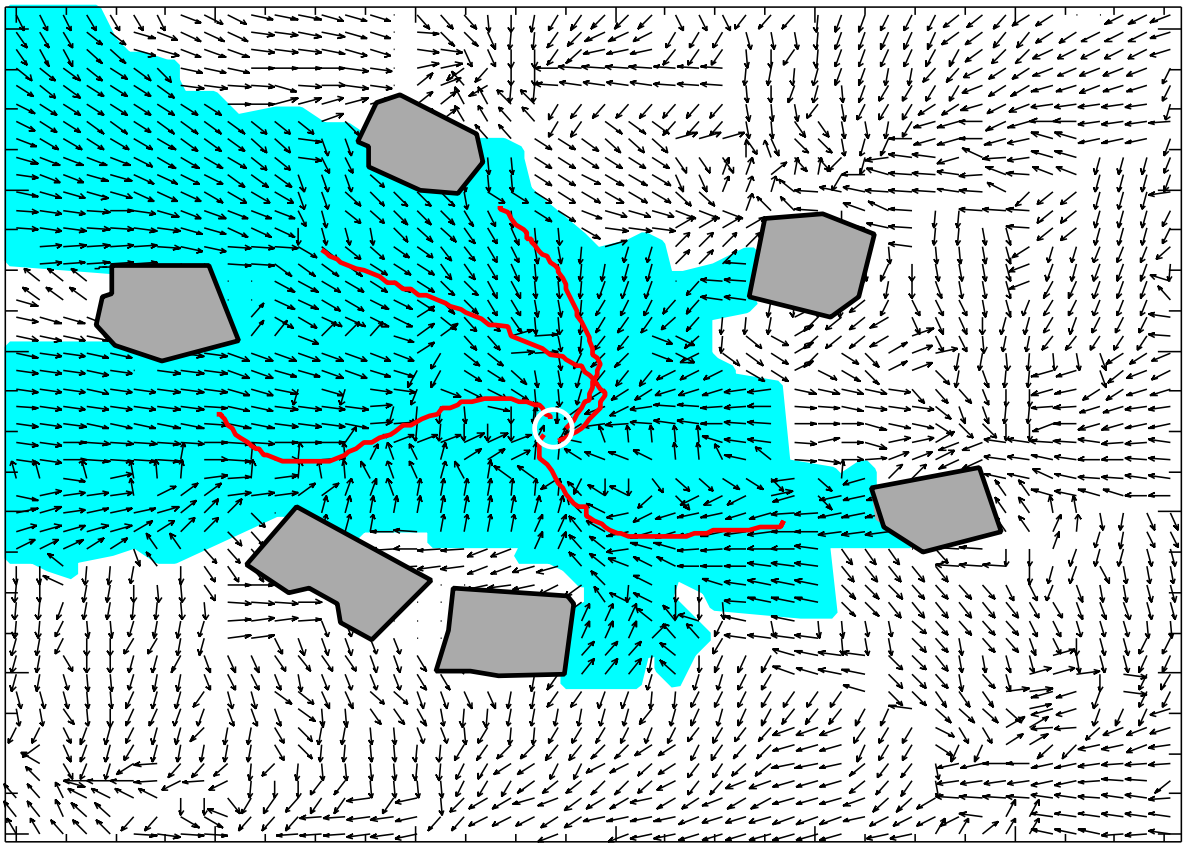

Figure 6: Home vector field for a central view in the arena (118 × 102 cm) shown in Fig. 3, with sample trajectories of a homing Khepera robot. The catchment area is depicted in grey, and the home position is marked by a circle.

onds. The success rates in Fig. 5 show that the algorithm performs robustly up to an average distance of 15 cm from the home position. For larger distances, the start position was often outside the open space around the home position, so that occlusions affected the performance. Sample trajectories from the homing runs are shown in Fig. 6. In the office environment, homing was successful up to 2 m away from the home position.

With the help of the tracking device the size of the *catchment area* can be visualized (Fig. 6) using the following procedure: During a test run, the robot covered the whole arena with 10000 snapshots thus approximating the entire set of possible views (the view manifold) of this environment. For a selected home view, we calculated the corresponding home vector at all possible positions which lead to the map in Fig. 6. A point is considered part of the catchment area, if there is a path along the goal vectors leading to the goal. As can be seen from Fig. 6, the catchment area can cover the entire open space around the goal position. If the goal is placed nearer to an object, the catchment area decreases in size only moderately, which allows the effective use of the homing scheme in all areas of the arena where the robot does not collide with objects.

To assess the directional *accuracy* of the computed home vectors, we recorded 450 pairs of views during a random walk and computed the respective home vectors for each pair (Fig. 7). The pairs were required to be connected by a direct line of sight, and no snapshots were taken within $\approx 2\ cm$ reach of the obstacles. To evaluate the accuracy of our method, we calculated the average homeward component for distances in the range of 1 to 15 cm in 1 cm bins, each containing 30 samples. This measure characterizes both the accuracy and the angular dispersion of the computed home vectors and is often applied in homing experiments (Batschelet, 1981). As long as the homeward component stays significantly above zero, the robot moves nearer to the goal; if it is close to one, the robot moves directly homeward. The decrease in accuracy for distances smaller than 2 cm is due to sensor noise as predicted by Eq. (13). At distances larger than 15 cm the data base was too small, because pairs with larger distances fulfilling the imposed conditions occurred very rarely during the random walk,

due to the cluttered structure of the arena. Other experiments indicate that the accuracy decreases rapidly due to occlusions beyond this value. This should not be considered a limitation of the homing scheme, but a characteristic of the environment.

4.3 Improvements by independent parameter estimation

The function over which the optimization in the three parameters α, ψ and ν has to be performed, has multiple local minima, thus standard gradient descent methods cannot be used. Since a global search is very time consuming it is convenient if ψ or ν can be estimated independently. A possible solution would be the use of an external compass to estimate ψ, but it can also be extracted from image data alone, as described below.

Spatial distance from image distance Due to the occurence of multiple local minima, the Euclidean image distance cannot be used to home by direct gradient descent. Nevertheless, the image distance between snapshot and current view correlates well with the relative distance ν, as can be seen from Fig. 8. The map was calculated from the view manifold data described in Sec. 4.2. In fact, up to a certain distance, spatial distance may be inferred from measured image distance. As the estimate of the other two parameters α and ψ is very robust to variations in ν (in particular, at large distances, cf. Sec. 3.2), we use a linear approximation for the relationship between spatial and image distance. This speeds up the algorithm considerably, so that home vectors can be computed in our $C{+}{+}$ implementation at a frame rate of 25 Hz on the SGI Indy workstation (R4400 Processor at 100 MHz).

Orientation estimation Similarly, the change of orientation ψ may be estimated by shifting snapshot and current view until a minimal image distance is reached. Unfortunately, this only works well near the goal. Since the algorithm does not tolerate large errors in the estimate of ψ, this method is not directly applicable in our scheme. A different approach, however, which we have successfully tested in other experiments, involves using previously acquired information to speed up the estimation of ψ. In particular, restricting the search space for ψ to the neighbourhood of previous estimates of ψ did not decrease accuracy.

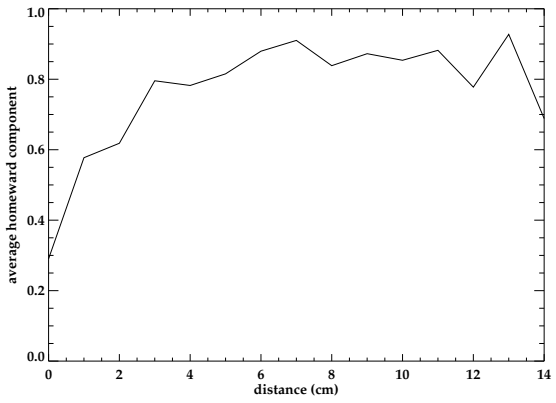

Figure 7: Average homeward component of computed home vector.

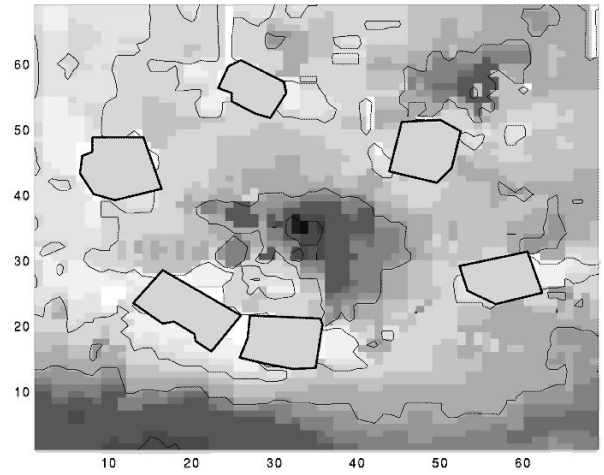

Figure 8: Euclidean image distance map for the home in Fig. 6. The displayed grey values correspond to the minimum distance which can be obtained by rotating the respective views. Darker areas represent smaller image distances.

5 Discussion

Previous approaches to scene–based homing. In the following, we will relate our approach to previously published approaches by briefly discussing a number of scene–based homing schemes and pointing out some differences to the present approach. In doing so, we will mainly focus on the type of approximations and the correspondence mechanisms utilized (summarized in Table 1).

Most approaches use a 360° field of view, greatly reducing computational cost: for limited fields of view, invisible parts must be kept in some internal representation, whereas for an omnidirectional sensor, all non–occluded landmarks are permanently visible. As Nelson & Aloimonos (1988) pointed out, there is an additional advantage: In a 360° field of view, the rotatory and translatory part of the disparity field can easily be separated, while in the case of limited fields of view, there is no unique decomposition. Therefore, considerable effort has gone into technical implementations, including a camera pointing at a spherical mirror (Hong et al. 1991), and a rotating intensity sensor (Röfer 1995).

Author	Approx-imation	Correspon-dence	Input	Constant orientation	Implemen-tation
Cartwright & Collett 1983	Isotropic landmark distribution	Region matching	Binary, one-dim., 360°.	yes	Computer simulation
Hong et al. 1991	Isotropic landmark distribution	Feature matching	Grey value, one-dim., 360°.	yes	Mobile robot
Röfer 1995	Isotropic landmark distribution	Kohonen network	Grey value, one-dim., 360°.	yes	Mobile robot
Wittmann 1995	Isotropic landmark distribution	Correlation on resolution pyramid	Grey value, one-dim., 360°.	yes	Computer simulation
Basri & Rivli 1995	Weak perspective	Linear combination of model images	Video images	no	Computer simulation
Franz et al. 1997	Equal distance	Parameterized disparity fields	Grey value, one-dim., 360°.	no	Mobile robot

Table 1: Overview of scene–based homing schemes (cf. Sec. 5)

Cartwright & Collett (1983) and Wittmann (1995) proposed models for honey bee landmark navigation. Both assume that the bee stores its orientation with respect to an external compass reference provided by the sun or the earth's magnetic field. This allows the bee to keep the orientation of snapshots constant, either by "mental" counterrotation or appropriate body orientation. Similarly, the camera platforms of the robot used by Hong et al. and Röfer do not rotate when the robot changes direction, so that all views have constant orientation. As pointed out in Sec. 4.3, this has the advantage of greatly reducing computational cost. The schemes of Cartwright & Collett and Wittmann are implemented in idealized computer models, so they do not have to provide solutions on how to deal with noisy orientation estimates. Since these errors may result in large deviations in the estimation of the home direction, small rotatory deviations are compensated for in the robotic implementations of Hong et al. and Röfer. However, the orientation of the platforms is subject to cumulative errors, thus their schemes may fail in large scale environments.

It should be noted that although the above approaches differ in the way they establish correspondences between views, they all rely on constant orientation and the approximation of isotropic landmark distribution. Apart from the fact that the latter is rarely realized, the error in the computed home direction due to this approximation may be very large, even close to the goal (e.g., if all landmarks were concentrated near K_A in Figure 1). Thus, these schemes may converge very slowly in strongly non-isotropic environments, and even fail for higher noise levels. Cartwright & Collett included therefore an additional feature in their scheme to reproduce the experimental data: The vector sum for the computation of the home direction contains not only the tangential disparity vectors but also radial vectors which act to lessen the size discrepancy of the visible objects. This makes their scheme less sensitive to non-isotropic landmark distributions (and even works if only one single object is visible), but requires mechanisms for object recognition.

The scheme of Basri & Rivli (1995), unlike the above approaches, operates with a limited field of view. It uses images of objects with identifiable features, taken from different view points, as model images. Under weak perspective conditions any other view of the object can be generated by a linear combination of the model views. The goal direction is computed from the transformation coefficients of the current view and the snapshot. In addition to the mathematical discussion, Basri & Rivli validate their method on real–world test images; however, they do not give results obtained for simulated vehicles or robots.

The present approach: parameterized disparity fields. In this paper, we have proposed a novel approach to scene–based homing, based on the equal distance assumption described in Sec. 2. The accuracy with which the algorithm can approach a goal was shown to be limited only by sensor noise, not by the approximation, and that every snapshot is surrounded by a catchment

area. Robot experiments demonstrated the validity of our method for real world applications and provided a quantitative assessment of its performance.

As the computation of disparity fields is an ill-posed problem, some additional assumption about the field had to be included. Our scheme makes explicit use of the underlying geometry of the task. Together with the equal distance assumption, this yields a low–dimensional parameterization of the possible disparity fields. The low–dimensionality leads to an optimization problem solvable in real time. All disparity fields defined by the parameterization, in particular the result of the optimization, are such that they can occur in real–world situations. This, however, is not guaranteed for general optical flow methods such as feature matching or correlation.

Clearly, our homing scheme is limited to the immediately accessible surroundings of a snapshot. Elsewhere, we have described how to deal with navigation in large–scale environments by combining several snapshots into a graph-like structure (Schölkopf & Mallot 1995, Franz et al. 1997).

Since this work was largely inspired by biology we want to conclude with a few remarks concerning the biological relevance of our scheme. The proposed algorithm could be implemented with matched filters in very simple neural circuitry. As Krapp & Hengstenberg (1996) have recently shown, flies use matched filters for complex stimuli such as generic optical flow fields. Moreover, we note that in our approach, 3–D information is only present implicitly, in the use of perspective distortion, and in the geometrical parameterization of disparity fields. Previous studies have shown that a variety of visual tasks (e.g. object recognition, see Bülthoff & Edelman, 1992) can be accomplished by biological systems without using explicit 3–D representations. Although these observations support our approach, we emphasize that it is not an explicit model of animal behaviour. It rather aims at understanding possible solutions to a general problem which robots as well as animals have to solve.

Acknowledgements. The present work has profited from discussions and technical support by Hanspeter Mallot, Philipp Georg, Susanne Huber, and Titus Neumann. We thank Ralf Möller and Guy Wallis for helpful comments on the manuscript. Financial support was provided by the Max–Planck–Gesellschaft and the Studienstiftung des deutschen Volkes.

References

[1] R. Basri and E. Rivlin. Localization and homing using combinations of model views. *Artificial Intelligence*, 78:327 – 354, 1995.

[2] E. Batschelet. *Circular statistics in biology.* Academic Press, London, 1981.

[3] H. H. Bülthoff and S. Edelman. Psychophysical support for a 2–D view interpolation theory of object recognition. *Proceedings of the National Academy of Science*, 89:60 – 64, 1992.

[4] B. A. Cartwright and T. S. Collett. Landmark learning in bees. *J. comp. Physiol. A*, 151:521 – 543, 1983.

[5] J. S. Chahl and M. V. Srinivasan. Visual computation of egomotion using an image interpolation technique. *Biol. Cybern.*, 74:405 – 411, 1996.

[6] T. S. Collett. Landmark learning and guidance in insects. *Phil. Trans. R. Soc. Lond. B*, 337:295 – 303, 1992.

[7] T. S. Collett. Insect navigation en route to the goal: Multiple strategies for the use of landmarks. *J. exp. Biol.*, 199:227 – 235, 1996.

[8] M. O. Franz, B. Schölkopf, P. Georg, H. A. Mallot, and H. H. Bülthoff. Learning view graphs for robot navigation. In W. L. Johnson, editor, *Proc. 1. Intl. Conf. on Autonomous Agents*, pages 138 – 147, New York, 1997. ACM Press.

[9] S. Goldman. *Information theory.* Dover, New York, 1953.

[10] J. Hong, X. Tan, B. Pinette, R. Weiss, and E. M. Riseman. Image-based homing. In *Proc. IEEE Intl. Conf. on Robotics and Automation 1991*, pages 620 – 625, 1991.

[11] H. G. Krapp and R. Hengstenberg. Estimation of self–motion by optic flow processing in single visual interneurons. *Nature*, 384:463 – 466, 1996.

[12] R. C. Nelson and J. Aloimonos. Finding motion parameters from spherical motion fields (or the advantages of having eyes in the back of you head). *Biol. Cybern.*, 1988:261 – 273, 1988.

[13] A. V. Oppenheim and R. W. Schafer. *Discrete–time signal processing.* Prentice Hall, Englewood Cliffs, NJ, 1989.

[14] T. Röfer. Controlling a robot with image-based homing. In B. Krieg-Brückner and C. Herwig, editors, *Kognitive Robotik (ZKW Bericht Nr. 3)*, Zentrum für Kognitionswissenschaften, Universität Bremen, 1995.

[15] B. Schölkopf and H. A. Mallot. View–based cognitive mapping and path planning. *Adaptive Behavior*, 3:311 – 348, 1995.

[16] M. J. Tarr and H. H. Bülthoff (eds.). Abstracts of the scene recognition workshop Tübingen. Technical Report 30, Max–Planck–Institut für biologische Kybernetik, 1996. Available as /pub/mpi-memos/TR-30.ps.Z via anonymous ftp from ftp.mpik-tueb.mpg.de.

[17] T. Wittmann. Modeling landmark navigation. In B. Krieg-Brückner and C. Herwig, editors, *Kognitive Robotik (ZKW Bericht Nr. 3)*, Zentrum für Kognitionswissenschaften, Universität Bremen, 1995.

[18] Y. Yagi, Y. Nishizawa, and M. Yachida. Map-based navigation for a mobile robot with omnidirectional image sensor COPIS. *IEEE Trans. Robotics Automat.*, 11:634 – 648, 1995.