Biol. Cybern. 79, 191–202 (1998)

Biological Cybernetics


# Where did I take that snapshot? Scene-based homing by image matching

**Matthias O. Franz, Bernhard Schölkopf, Hanspeter A. Mallot, Heinrich H. Bülthoff**

Max-Planck-Institut für biologische Kybernetik Spemannstraße 38, D-72076 Tübingen, Germany
(e-mail: matthias.franz@tuebingen.mpg.de)

Received: 11 December 1997 / Accepted in revised form: 12 June 1998

**Abstract.** In homing tasks, the goal is often not marked by visible objects but must be inferred from the spatial relation to the visual cues in the surrounding scene. The exact computation of the goal direction would require knowledge about the distances to visible landmarks, information, which is not directly available to passive vision systems. However, if prior assumptions about typical distance distributions are used, a snapshot taken at the goal suffices to compute the goal direction from the current view. We show that most existing approaches to scene-based homing implicitly assume an isotropic landmark distribution. As an alternative, we propose a homing scheme that uses parameterized displacement fields. These are obtained from an approximation that incorporates prior knowledge about perspective distortions of the visual environment. A mathematical analysis proves that both approximations do not prevent the schemes from approaching the goal with arbitrary accuracy, but lead to different errors in the computed goal direction. Mobile robot experiments are used to test the theoretical predictions and to demonstrate the practical feasibility of the new approach.

## 1 Introduction

For many animal species it is vital to find their way back to a shelter or to a food source. In particular, flying animals cannot rely on idiothetic information for this task, as they are subject to wind 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 scene 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: (1) if the location itself is marked by salient visual cues, these may act as *beacons* that can be tracked until the goal is reached (e.g. Collett 1996); (2) if there are no beacons, the goal direction has to be inferred from the spatial relation to the visual cues in the surrounding scene. A protypical example for this navigation task is the Morris water-maze task (Morris 1981), where a rat has to find a platform hidden under an opaque water surface (cf. Fig. 1). If the animal moves so as to attain the same spatial relationship to the scene as the one recorded in the snapshot, it will eventually reach the goal. In this study, we refer to this behaviour as scene-based homing since it makes use of the whole scene rather than of 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 compute the goal direction from the azimuth and size change of landmarks near the goal. While vertebrates also seem to use landmark distances for scene-based homing tasks (Collett 1986), the proposed mechanism requires only the storing and processing of a single snapshot. Cartwright and Collett (1983) and Wittmann (1995) showed in computer simulations that a snapshot-based homing scheme is indeed sufficient to explain the search frequency patterns of honeybees. Experiments with mobile robots have demonstrated that similar mechanisms also work under real world conditions (Hong et al. 1991; Röfer 1995, 1997; Franz et al. 1997; Möller et al. 1998).

In the present work, we focus on the problems that any agent, animal or robot, has to face when using snapshots of the surrounding scene for homing tasks. We provide an analysis of the necessary computations which shows that several solutions to this problem are possible, depending on the basic assumptions. As an

*Correspondence to*: M.O. Franz, Max-Planck-Institut für biologische Kybernetik Spemannstraße 38, D-72076 Tübingen, Germany (e-mail: matthias.franz@tuebingen.mpg.de)

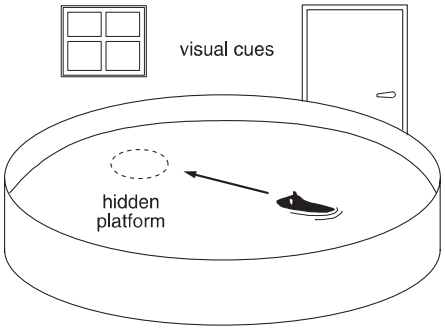



**Fig. 1.** Morris water-maze task: a rat has to swim to a platform hidden under an opaque water surface. The goal direction can only be inferred from the spatial relationship to the visual cues in the surrounding scene

alternative to existing computational models, we formulate a new scene-based homing algorithm that is able to cope with some of the shortfalls of previous approaches. We test the algorithm and our theoretical predictions on real robots to avoid the idealizations one necessarily has to accept when simulating an agent and its environment.

In the next section, we give a mathematical description of the basic task, followed by an investigation of the error and convergence properties of previous solutions. Using this computational framework, we propose a new algorithm in Sect.4. Section 5 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.

## 2 Inferring the goal direction from the surrounding scene

### *2.1 Mathematical description of the task*

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 chose a mobile omnidirectional 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 at the horizon of the surrounding panorama as a snapshot. Although this assumption is not an essential prerequisite, it simplifies the mathematical treatment and the practical implementation, since no computational resources are necessary to merge several restricted views into a common image of the panorama. Using a ring at the horizon has the additional advantage that, in a static environment, the optic flow will always be confined to this ring when moving in the plane. Imaged landmarks may move along that ring or become occluded, but will never leave the ring.

The position of an image point on the sensor ring is denoted by the angle $\theta$. 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 the sensor ring moves away from the home position $H$ in direction $\alpha$ by a distance $d$ to point $C$ and changes its orientation by the angle $\psi$ (see Fig. 2). As a consequence, the image of landmark $L$ at distance $r$ is shifted from $\theta$ to a new position $\theta + \delta$ (assuming a static environment). From the triangle $HLC$ in Fig. 2, we obtain

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

This relation can be used to compute the direction $\beta = \alpha - \psi + \pi$ back to the starting position $H$ from the change $\delta$ in the landmark position (the *displacement*) if $r/d$ and $\psi$ are known.

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

1. In order to compute the displacement $\delta$, a correspondence between image points in the snapshot and in the current view must be established.
2. If the snapshot and the current view are the only information available, the absolute distance $r$ of the landmark at $L$ is unknown. This lack of knowledge must be compensated by some additional assumption about the distance distribution of possible landmarks in the environment.

### *2.2 Computation of correspondences*

The problem of identifying corresponding regions in different images is a well-studied issue in computer

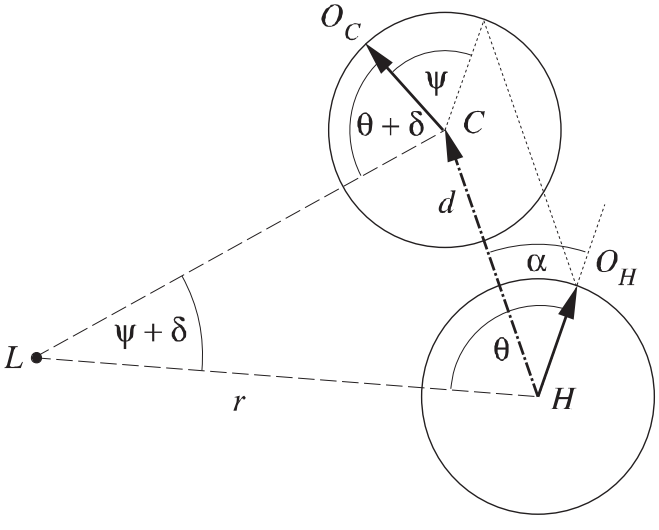



**Fig. 2.** Moving a ring sensor from the home position $H$ to $C$ in direction $\alpha$ (with respect to sensor's initial orientation $O_H$) and rotating it by $\psi$ leads to a change $\delta$ of the viewing angle of a landmark $L$

vision (Förstner 1993), e.g., in optic flow analysis or in stereo vision. This has led to the application of standard optic flow techniques to establish correspondences between the snapshot and the current view, such as feature matching (Hong et al. 1991) or correlation of image patches on a multi-scale pyramid (Wittmann 1995). In the general case of a translating and rotating agent, the entire image has to be searched for correspondences. This not only requires large computational resources, but also increases the danger of false matches. The search space for correspondences can be restricted if the agent knows its orientation with respect to an external reference (Cartwright and Collett 1983; Wittmann 1995; Möller et al. 1998)) or always keeps a constant orientation (Hong et al. 1991; Röfer 1995). In addition, the number of false matches can be reduced by using coloured images or assuming neighbourhood preservation of landmarks (Röfer 1995, 1997).

All previous approaches compute correspondences locally, i.e., they compare subregions of the image. This makes them susceptible to false matches since small image patches are not very distinctive. In Sect. 4, we will present a global image matching procedure that does not rely on local or feature correspondence. Similar schemes for global image comparison have been suggested for simple tasks in stereo vision by Mallot et al. (1996).

### 2.3 Assumptions about distance distributions

*Isotropic distance assumption*. Although never explicitly stated, most previous approaches to scene-based homing compensate the lack of distance knowledge by assuming an isotropic landmark distribution. This means that the frequency and distance of landmarks are assumed to be independent of the viewing direction. In this case, one obtains the correct goal direction by summing over all displacement vectors along the sensor ring, since all vector components orthogonal to the direction of the movement cancel each other (see Sect. 3). A homing algorithm based on this assumption does not need any object recognition mechanisms and can rely solely on local optical flow techniques. Although a violation of the isotropic distance assumption introduces considerable errors in the computed home direction, we show in Sect. 3.2 that it nevertheless leads to converging homing algorithms.

*Equal distance assumption*. In Sect. 4, we will introduce a new algorithm based on an approximation which we call equal distance assumption. The surrounding landmarks are assumed to have identical distances from the location of the snapshot. This approach provides constraints for the computation of the displacement field which will be used in a global image matching procedure. On first sight, the equal distance assumption appears not to be very realistic, but we will show in Sect. 4.2 that the effect of the resulting errors on homing performance remain small.

## 3 Homing with the average displacement vector

### 3.1 Displacement vector sum in isotropic environments

Most of the approaches described in the literature assume that the agent keeps a constant orientation ($\psi = 0$) or that any orientation change is corrected before computing the home direction. In this case, we obtain for the displacement from (1)

$$\tan\delta = \frac{d\sin(\theta-\alpha)}{r - d\cos(\theta-\alpha)} \tag{2}$$

Following Hong et al. (1991), we define the associated *displacement vector* $\vec{\delta}_i$ as a vector with absolute value $\delta$ pointing in direction $\theta_i + \delta/2 + \pi/2$ for $\delta > 0$, and in direction $\theta_i + \delta/2 - \pi/2$ for $\delta < 0$ (cf. Fig. 3a). It was assumed by Hong et al. (1991) that the distance to the goal can be most quickly reduced by moving in the direction of the displacement vector. By simply adding up all displacement vectors over the sensor ring and normalizing the result, one obtains an averaged unit home vector

$$\vec{\mathbf{h}} = \frac{\sum_i \vec{\delta}_i}{\left\|\sum_i \vec{\delta}_i\right\|} \tag{3}$$

which denotes the next movement direction. As can be seen from Fig. 3a, the displacement vector contains generally a component which is orthogonal to the direction of the movement and thus gives rise to the error $\eta$ in the computed driving direction. If one assumes an isotropic landmark distribution, these errors cancel each other in the vector sum (3). Note that vectors with a larger displacement stronger influence the computed home vector in (3). This has two advantages: First, large displacements are less affected by sensor noise. Second, the largest displacement vectors occur perpendicular to the goal direction and thus have the smallest error component (cf. Fig. 3).

The scheme of Hong et al. (1991) was used in the later implementations of Röfer (1995, 1997), Wittmann (1995), and Möller et al. (1998), but in each approach the displacements were computed using different methods (see Table 1). In the original scheme of Cartwright

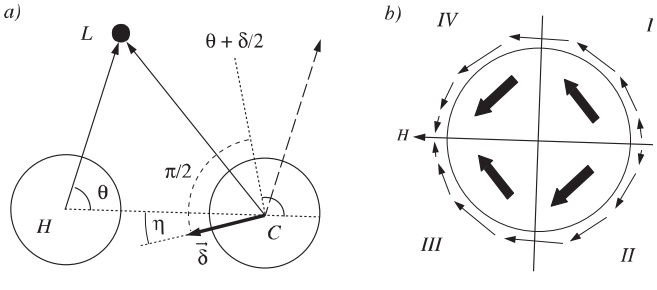



**Fig. 3. a** The direction of the displacement vector $\vec{\delta}$ is $\theta + \delta/2 + \pi/2$ for $\delta > 0$, and $\theta + \delta/2 - \pi/2$ for $\delta < 0$. The resulting error in the home direction is $\eta$. **b** The average displacement vector points exactly in the home direction, if the summed errors of quadrant I and III exactly match those of quadrant II and IV. The largest displacement vectors occur at viewing angles perpendicular to the home direction

and Collett (1983), the displacement vectors had unit length and were attached directly at $\theta$. They additionally included radial unit vectors which act to lessen the size discrepancy of visible objects.

In order to illustrate the properties of a homing algorithm based on the isotropic distance assumption, we have implemented a simplified version of the scheme by Hong et al. (1991). In their approach, local matches $m_{i,j}$ between an area of the current image $I$ centered at $i$ and an area of the stored snapshot $I^S$ centered at $j$ are calculated according to

$$m_{i,j} = \sum_{l=-L}^{L} \mathrm{w}_l \left| (I_{i+l} - \mu) - \frac{\sigma}{\sigma^{\mathrm{S}}} (I^{\mathrm{S}}_{j+l} - \mu^{\mathrm{S}}) \right| \tag{4}$$

where $\mathrm{w}_l = 1/(2 + |l|)$ is a weighting function[1], $[-L, L]$ the window, and $\mu$ and $\sigma$ the mean and the standard deviation of the pixel values inside the window. The angular difference between the image position of $i$ and that of the best match yielded the displacement of $i$ with respect to the snapshot. In contrast to the original implementation, we did not identify particularly suitable image points, but computed a dense displacement field at all image positions. This could possibly lead to a degradation in performance but suffices to illustrate the properties of the scheme as the applicability to robotic tasks was already shown by Hong et al. (1991).

Although the isotropic distance assumption has been implemented in several systems, none of these approaches provides an analysis of error or convergence properties. In the next section, we show that this approximation indeed leads to converging homing schemes, although the associated error does not necessarily decrease when approaching the home position.

### 3.2 Error and convergence

The angular deviation $\eta$ (cf. Fig. 3) between the displacement vector and true home direction can never exceed 90°. To visualise this, let us consider a landmark positioned left of the vector $\vec{\mathbf{HC}}$. Then, both $\theta$ and $\theta + \delta$ are confined to the interval $[0, \pi]$. Clearly, the same has to hold for their mean, $\theta + \delta/2$. Since the estimated home direction is given by

$$\beta = \theta + \frac{\delta}{2} + \frac{\pi}{2} \tag{5}$$

we have $\pi/2 \leq \beta \leq 3/2\pi$. Keeping in mind that the correct return angle is $\pi$, we find that the absolute error $|\eta|$ is less or equal to $\pi/2$ provided that there are no errors in the measurement of $\delta$. Equality is obtained if and only if $H$, $C$ and $L$ are all on one line. An analogous argument holds if the landmark is to the right of the movement direction.

Let us assume that we have solved the correspondence problem and that the computed displacements are error-free. Furthermore, let us assume that the sensor ring has a constant orientation and that the environment contains more than two non-aligned landmarks. If the homing mechanism described above is used, then the distance $d(t)$ to the home position tends to 0 in a general environment, provided that there are no locations with identical views. In practice, this means that the home position can be found from all points where correct displacements are computable.

*Proof.* Since the directional error is always smaller than or equal to 90°, $d(t)$ has to decrease monotonically during the homing procedure. An error of exactly 90° occurs only at two points in the visual field, namely in the home direction and in the opposite direction. As we assume there are at least three non-aligned landmarks, there must be at least one displacement vector within error smaller than 90°. Thus, $d(t)$ decreases strictly monotonically for $d(t) > 0$. In addition, $d(t)$ is always bounded from below by 0, such that $\lim_{t\to\infty} d(t)$ exists. This means that $d(t)$ converges to 0, which completes the proof.

In a non-isotropic environment, the configuration of sensor and environment will change after each movement and so will the error. Although the single displacement vectors become smaller when the sensor approaches the goal, it should be noted that the error in the estimated home direction does not necessarily decrease due to the normalization factor in (3). Since we have shown that the scheme converges, this should result in a spiral-like trajectory during the homing procedure.

## 4 Homing with parameterized displacement fields

### 4.1 A matched filter based on the equal distance assumption

Before applying the equal distance assumption, we will convert (1) into a suitable form for the subsequent analysis. We assume a typical landmark distance $R$ and denote the deviation from it by $r'$ so that $r = R + r'$. This leads to

$$\tan(\psi + \delta) = \frac{\frac{d}{R}\sin(\theta - \alpha)}{1 + \frac{r'}{R} - \frac{d}{R}\cos(\theta - \alpha)} \tag{6}$$

We replace $r'$ and $d$ by the ratios $\epsilon = r'/R$ and $\rho = d/R$. This leads to the simplified notation

$$\tan(\psi + \delta) = \frac{\rho \sin(\theta - \alpha)}{1 + \epsilon - \rho\cos(\theta - \alpha)} \tag{7}$$

The cases $\rho = 0$, $\epsilon = -1$, on the one hand, and $\rho = 1 + \epsilon$, $\theta = \alpha$ on the other 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 $\epsilon$ of the surrounding landmarks. When solved for $\delta$, the resulting expression

[1] The original paper uses $\mathrm{w}_l = 1/|2 + l|$ which we think is a printing error

$$\tan(\psi + \delta) = \frac{\rho \sin(\theta - \alpha)}{1 - \rho \cos(\theta - \alpha)} \tag{8}$$

describes the displacement field $\delta(\theta)|_{\epsilon=0}$ when all landmarks are located at a distance $R$ from the starting position. The displacement field is completely determined by only three parameters: $\alpha$, $\psi$ and $\rho$.

These simplified displacement fields can be used to estimate the real displacement field by a matching procedure. Since the direction of the sensor movement $\alpha$ is one of the matching parameters, the goal direction can be computed using the following algorithm:

1. For all parameter values of $\alpha$, $\psi$ and $\rho$, the current view is distorted by shifting the image positions $\theta$ of the single pixels according to (8). The result of this procedure is new images that would have been obtained if the sensor had moved according to the parameters in an environment where the constant distance assumption was perfectly valid.
2. The generated images are compared to the snapshot taken at the home position. To measure the degree of match, we use the dot product between the distorted image and the snapshot. The best match is produced by a displacement field which reconstructs the home view as accurately as possible.
3. The parameter value of $\alpha$ leading to the best match is selected to obtain an estimate $\beta = \alpha + \pi$ of the home direction.[2]
4. The agent moves in the estimated home direction, until the home position is reached. We use two independent criteria to detect the goal: either the dot product between the current image and the snapshot exceeds a pre-set threshold, or the computed home vector changes its direction about 180° after passing the goal.

In order to determine the unknowns in (8) 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. Note that the algorithm produces an estimate not only of the home direction, but also of the orientation $\psi$ and of the relative distance $\rho$. Although we do not use these estimates in our homing procedure, this information could be used, e.g., for visual odometry, without any additional computations.

The parameterized displacement field $\delta(\theta)|_{\epsilon=0}$ can be interpreted as a matched filter in the sense that the parameter set that reproduces the actual displacement field best is an estimate of the real one. Since the direction of movement $\alpha$ is one of the parameters, the best matching displacement field 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 and Hengstenberg 1996), and theoretically by Nelson and Aloimonos (1988). Mallot et al. (1991) have used motion templates for obstacle detection in a robot application.

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

### *4.2 Error due to the equal distance assumption*

Unlike the isotropic distance assumption, the error due to the equal distance assumption decreases when the sensor approaches the goal. This follows directly from (6) if we assume that the sensor is in the open space around the goal (i.e., $\rho < 1 + \epsilon$). In this case, the displacement $\delta$ is given by

$$\delta = \arctan\left(\frac{\rho \sin(\theta - \alpha)}{1 + \epsilon - \rho \cos(\theta - \alpha)}\right) - \psi \tag{9}$$

The error in the displacement $\delta$ due to neglecting the deviation of the landmark distance $r'$ from the averaged distance $R$ is

$$E(\epsilon, \rho) := \delta(\epsilon, \rho) - \delta(0, \rho) \tag{10}$$

Both $\delta$ and $E$ are continuous functions in $\rho$ and $\epsilon$ for $\epsilon > -1$. Moreover, $\delta$ and $E$ tend to zero for $\rho \to 0$. This means that for each $\epsilon > -1$, fixed $\theta$ and $\alpha$, and any desired accuracy bound $E_0 > 0$, there exists a $\rho_0$ such that $\rho < \rho_0$ implies $|E(\epsilon, \rho)| < E_0$. 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 every snapshot containing at least three landmarks, there exists an area in which the location of the snapshot can be approached arbitrarily closely.

The maximal area in which the goal can be found is called the *catchment area* of the snapshot (Cartwright and Collett 1987). In practice, the catchment areas tend to be larger than one might expect from the equal distance approximation, as there are several factors which effectively constrain the distances of the imaged landmarks. First, the error induced by an infinitely distant point is relatively small, compared to displacements generated by nearby landmarks. Second, very close points will not have an effect as adverse as might be expected from their large displacements, 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.

### *4.3 Limitation of spatial resolution by sensor noise*

As we have shown in the previous sections, neither the isotropic nor the equal distance assumption limit the spatial accuracy of homing. Therefore, the primary

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

limiting factor is the pixel and quantization noise of the sensor ring. In the following, we will determine the maximally achievable spatial resolution $\Delta\rho$, i.e. the minimal distance between places whose images can be reliably distinguished.

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 $\theta$, and spatial aliasing effects are eliminated.

If the variance of the noise in the intensity distribution is given by $\sigma^2$, the maximally resolvable intensity change is $2\sigma$ 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 movement of the sensor ring from the location of the snapshot induces a small change $\Delta\theta$ in the position of the landmarks. The resulting change of the detected intensity distribution at $\theta$ is, to a first-order approximation,

$$\Delta h(\theta) \approx h'(\theta) \cdot \Delta\theta \tag{11}$$

which leads to a maximally resolvable image displacement of

$$\delta_{\min} = \Delta\theta \approx \frac{2}{h'(\theta)}\sigma \tag{12}$$

From (6), we obtain the expression

$$\frac{\partial\rho}{\partial\delta} = \frac{(1+\epsilon)\sin(\theta-\alpha)}{\sin^2(\theta-\alpha+\psi+\delta)} \tag{13}$$

so that the maximal spatial accuracy is given by

$$\begin{aligned}\Delta\rho &= \rho(\delta+\Delta\theta) - \rho(\delta) \\ &\approx \frac{\partial\rho}{\partial\delta}\cdot\delta_{\min} \\ &= \frac{2(1+\epsilon)\sin(\theta-\alpha)}{h'(\theta)\sin^2(\theta-\alpha+\psi+\delta)}\sigma\end{aligned} \tag{14}$$

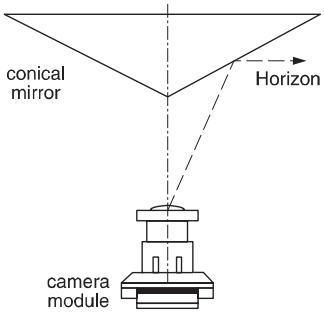
This shows that for extreme noise levels, low contrast and landmark positions near the movement direction, $\Delta\rho$ may become larger than the catchment area, so that in these cases a scene-based homing scheme is not applicable. Note, that the above limitation is derived for only one landmark. When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

## 5 Robot experiments

### *5.1 Experimental set-up*

The experiments were conducted in an arena with dimensions of $118 \times 102$ cm. Visual cues were provided by model houses in the arena (see Fig. 4). We used a modified Khepera miniature robot connected to an SGI Indy workstation via a serial and video transmission cable (Franz et al. 1997). Our scheme was also tested successfully in a real office environment on two other robot platforms. The imaging system on the robot comprises a conical mirror mounted above a small video camera which points up to the centre of the cone (Fig. 5). 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, and Yachida (1995). The video image was sampled at 25 Hz on four adjacent circles along the horizon with a resolution of 4.6° and averaged radially to provide robustness against inaccuracies in the imaging system and tilt of the robot platform. In a subsequent processing stage, a

**Fig. 4.** Test arena ($118 \times 102$ cm) with toy houses, used in the homing experiments (see Sect. 5.1). The modified Khepera robot is depicted in the right half of the arena



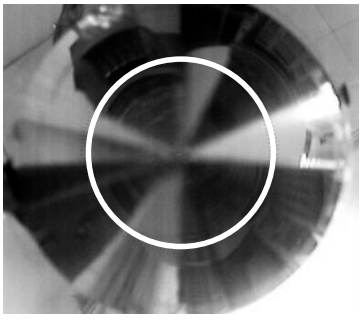



**Fig. 5.** The robot uses a camera module pointing at a conical mirror, which permits sampling of the environment over 360°, in a range of ±10° about the horizon. The photography shows the mirror as seen by the camera. The white circle marks the intensities at the horizon which are used as input for the homing procedure

spatiotemporal Wiener lowpass filter (e.g. Goldman 1953) was applied to the resulting one-dimensional array. To compensate for illumination changes, 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 these data were transmitted back to the robot using a serial data link with a maximal transmission rate of 12 commands per second.

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. Position information was not available to the robot during the experiments.

### *5.2 Performance of the homing scheme*

The feasibility of our approach was tested in an experiment with the Khepera robot in the 'toy house' arena (Fig. 4). During a test run, the robot covered the whole arena with 10 000 snapshots, while its position was recorded by the tracking device. The resulting view dataset samples the entire set of possible views (the view manifold) of this environment. The size of the catchment area can be visualized using the following procedure: For selected home views, we calculated the corresponding home vector at all possible positions, which leads to the maps 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. 6a, the catchment area can cover the entire open space around the goal position. The catchment area is somewhat smaller for goals that are closer to an object (Fig. 6c,d). However, an effective use of the scheme is possible in all areas of the arena where the robot does not collide with objects. Sample trajectories from typical homing runs are shown in Fig. 6b. During the homing runs, the robot computed the home direction relative to the current driving direction every 83 ms. The driving direction can only be corrected with a certain delay so that the trajectories do not follow the depicted home vectors exactly. That is, the shaded area shows the maximally achievable catchment area without taking into account the effects of the robot's hardware or control architecture.

Nevertheless, our algorithm can be succesfully applied for robot control as is demonstrated in the following experiment: for 20 different home positions, the robot was displaced relative to each home position by distances in the range of 5–25 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 s. The success rates in Fig. 7 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 and obstacles affected the performance. In the office environment, homing was successful up to 2 m away from the home position.

### *5.3 Improvements by independent parameter estimation*

The function over which the optimization in the three parameters $\alpha$, $\psi$ and $\rho$ has to be performed, has multiple local minima, and thus standard gradient descent methods cannot be used. Since a global search is very time consuming, it is convenient if $\psi$ or $\rho$ can be estimated independently.

*Spatial distance from image distance*. The image distance between snapshot and current view correlates with

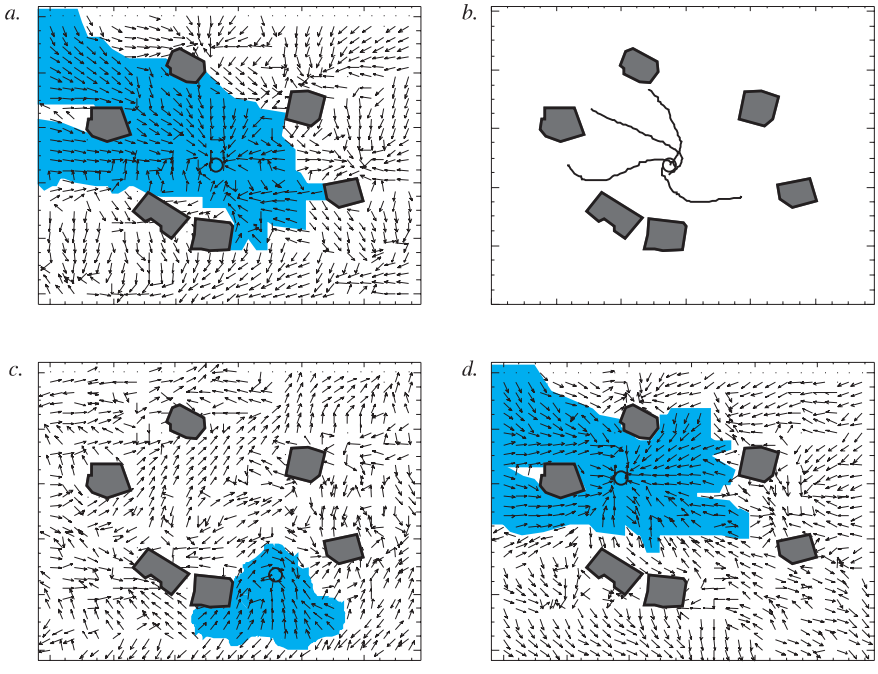



**Fig. 6.** Home vector fields in the arena shown in Fig. 4 for different home positions. **b** Actual trajectories of the homing robot for the home position shown in **a** The catchment area is depicted in *grey*, and the home position is marked by a *circle*

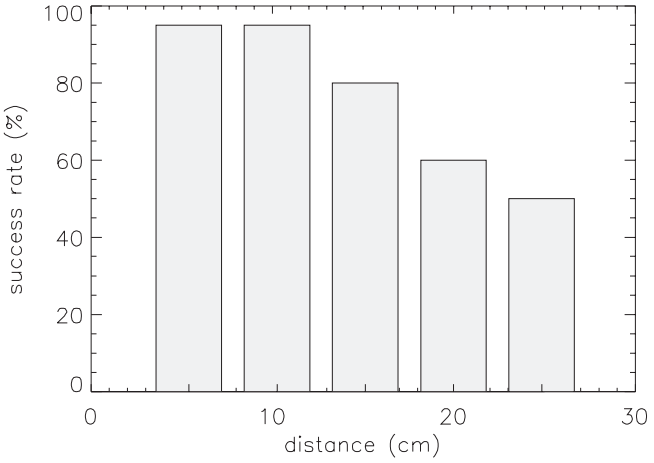



**Fig. 7.** Success rate for 100 homing runs, with starting distances between 5 and 25 cm

spatial distance, as can be seen from Fig. 8. The views are taken from the view dataset described in Sect. 5.2. We use the maximal pixel-wise cross-correlation $\Phi$ as a measure of image distance. This is equivalent to the dot product of two view vectors $\boldsymbol{a_i}$, $\boldsymbol{b_i}$, after first rotating one of them such as to maximize the overlap with the other one:

$$\Phi = \max_i \sum_j \boldsymbol{a_j} \boldsymbol{b_{j+i}} \tag{15}$$

Due to the correlation, a rough estimate of spatial distance may be obtained from the measured image distance. As the estimate of the other two parameters $\alpha$ and $\psi$ is very robust to variations in $\rho$, 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 an SGI Indy workstation (R4400 Processor at 100 MHz).

*Orientation estimation*. Similarly, the change of orientation $\psi$ may be estimated by shifting snapshot and current view until a minimal image distance is reached.

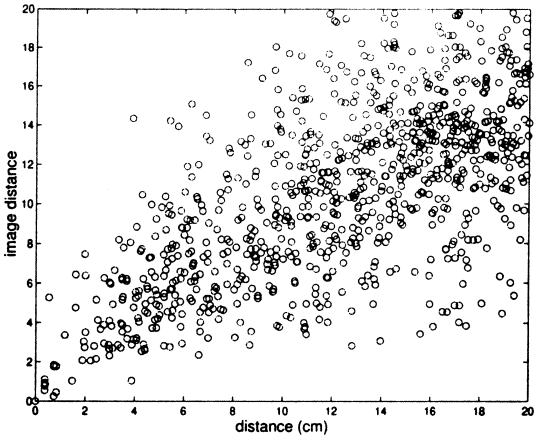



**Fig. 8.** Scatterplot of image distance vs spatial distance. The correlation can be used to estimate metric distance

Unfortunately, this works well only near the goal. Since the algorithm does not tolerate large errors in the estimate of $\psi$, 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 $\psi$. In particular, restricting the search space for $\psi$ to the neighbourhood of previous estimates of $\psi$ did not decrease accuracy.

### *5.4 Accuracy*

The accuracy of the computed home vector is directly linked to the error properties which we predicted in Sects. 3.2 and 4.2. As a measure of accuracy we use the *average homeward component* (Batschelet 1981). This measure characterizes both the accuracy and the angular dispersion of the computed home vectors and is often applied in homing experiments. As long as the homeward component stays significantly above zero, the robot moves nearer to the goal; if it is close to 1, the robot moves directly homeward.

In a first experiment (Fig. 9), we randomly selected 50 snapshots from the above mentioned 10 000 views. For each snapshot, all other views were divided into bins according to their distance from the snapshot, ranging from 2 to 30 cm in steps of 2 cm. From each bin, we selected a view at random and computed the home vector using both homing algorithms described above. The orthogonal projection of the estimated home vectors on the real home vector in each bin was averaged over all 50 snapshots.

We observe the predicted decrease in accuracy for the algorithm based on the constant distance approximation until a distance to the goal of 16 cm is exceeded. The decrease is due to two factors: first, the constant distance approximation becomes worse with increasing distance. Second, with increasing distance, occlusions become more frequent such that correspondences are harder to

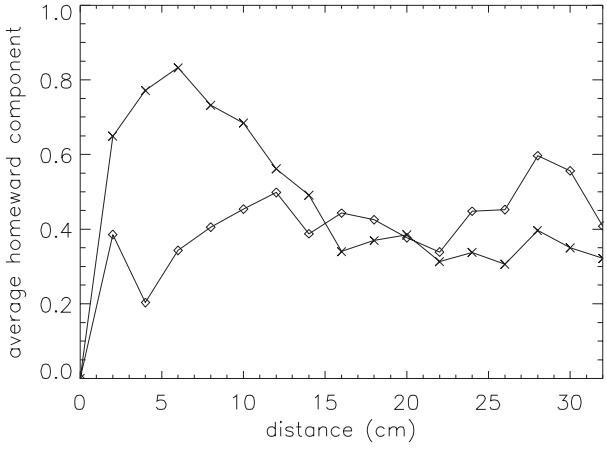



**Fig. 9.** Average homeward component of computed home vector for randomly chosen pairs of views. The values obtained from the constant distance assumption are marked by *crosses*, those from the isotropic distance assumption by *diamonds*

find. In contrast, the accuracy obtained from the algorithm based on the average displacement vector increases very weakly with distance, but remains at a low level. This is consistent with the prediction that no significant changes in accuracy are to be expected. The weak increase can be explained by the fact that displacement vectors are easier to compute for larger displacements which occur more often at larger distances from the goal. As a consequence, accuracy could probably be increased by using a higher image resolution. The decrease in accuracy of both algorithms for distances smaller than 2 cm is due to sensor noise as predicted by (14).

To assess the influence of occlusions on home vector accuracy, we recorded 450 pairs of views during a random walk and computed the respective home vectors for each pair (Fig. 10). 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. We again calculated the average homeward component using the constant distance approximation. The distances in the range of 1–15 cm were divided into 1-cm bins, each containing 30 samples. At distances larger than 15 cm, the database 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. The results in Fig. 10 indicate that the accuracy decrease due to the constant distance assumption is less pronounced when the path to the goal is free of obstacles.

## 6 Discussion

In this paper, we analysed the computational requirements for scene-based homing. We have shown that several solutions to this problem are possible, depending on the basic assumptions. As an alternative to existing computational models which rely on the isotropic distance assumption, we have proposed a novel approach to scene-based homing based on the equal distance assumption described in Sect. 2. We have shown that the accuracy with which these algorithms can approach a goal is limited only by sensor noise, not by the approximations, 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. The predicted error properties of the constant and isotropic distance assumption could be reproduced in the experiments. In the following sections, we discuss evidence from biological studies about scene-based homing in animals, and relate our approach to existing computational models.

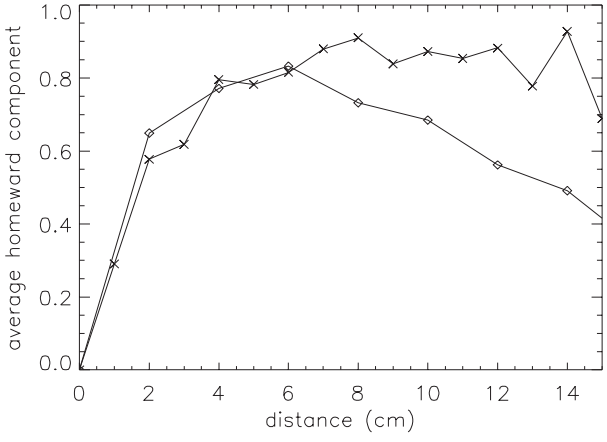



**Fig. 10.** Average homeward component based on the constant distance approximation for different choices of views. The curve marked by *crosses* is computed from view pairs connected by a direct line of sight, the curve marked by *diamonds* from randomly chosen view pairs

### 6.1 Relation to biological studies

In the biological literature, various terms have been used for scene-based homing. Collett (1996), e.g., describes this strategy as *image matching*, whereas Trullier et al. (1997) use the more general term *guidance*, which also includes non-visual strategies. Scene-based homing is a local navigation strategy, since the specific landmark configuration must be visible at any moment. Inside the visibility range, a home vector pointing towards the goal can be computed from the sensory input at all locations. This is different from *scene recognition-triggered responses* (Collett 1996; Trullier et al. 1997), where the recognition of the scene activates a previously stored home vector. Scene-based homing requires a working memory for the snapshot as the goal location itself is not specified by any visual cues. This is opposed to *aiming at beacons* (Collett 1996) or, as called by Trullier et al. (1997), *target approaching*, where the goal can always be perceived.

The ability for scene-based homing is widespread among animals. The most prominent example probably is found in the work on honeybees (Anderson 1977; Cartwright and Collett 1983). Cartwright and Collett showed that bees use only a snapshot to find the goal without recording the spatial layout of the surrounding scene. They proposed that by using a stack of distance-filtered snapshots, this mechanism could be extended to larger scale navigation (Cartwright and Collett 1987). More recent observations indicate, however, that scene-based homing in honeybees seems to be limited to the immediate vicinity of the goal, while other mechanisms are used for larger scale navigation (Collett 1996). Interestingly, the bee aligns its body in a specific compass direction during an approach (Collett and Baron 1994). From a computational point of view, this would greatly simplify the matching of snapshot and current view since only small image displacements have to be detected.

Scene-based homing abilities have been reported also for a number of other insect species such as hoverflies (Collett and Land 1975), waterstriders (Junger 1991), solitary wasps (Zeil 1993), ground nesting bees (Brünnert et al. 1994) and desert ants (Wehner et al. 1996). Wehner and Müller (1985) show that the snapshot recorded by desert ants remains fixed relative to retinal

coordinates and does not rotate to compensate for changes of the body orientation.

Vertebrates seem to have more highly developed scene-based homing abilities. Food-storing birds, for instance, retrieve hoarded food by remembering the location of thousands of caches (Sherry and Duff 1996). They rely mainly on visual information from nearby landmarks to locate the concealed caches. Gerbils (Collett et al. 1986) and rats (Morris 1981) store more than a mere snapshot, for these animals appear to remember also the spatial layout of the scene. If landmark distances are stored, additional assumptions about the distance distribution are no longer necessary. This allows for higher homing accuracies than in purely snapshot-based schemes.

Some experimental results have a simple explanation if one assumes that image displacement fields are used to compute the home direction. Cheng et al. (1987) report that the bee weights landmarks according to their distance from the goal. From (2), it is clear that distant landmarks lead only to small differences between the current image and the snapshot, while nearby landmarks cause larger displacement vectors. As a consequence, these vectors receive a higher weight in the displacement vector sum (3). Since in our algorithm (Sect. 4) a predefined displacement field is matched to the actual displacement field, a similar explanation holds: the algorithm mainly tries to reproduce the larger displacement vectors because these cause the largest mismatch between snapshot and current image. This can also account for the observation that bees assign higher weights to landmarks with high contrast (Pelzer 1985) since these lead to larger mismatches. It should be pointed out that other weighting mechanisms are conceivable. So far, our model uses only one-dimensional arrays of grey values as input, whereas other landmark properties such as their colour or height above ground might play an important role in assigning weights to them.

As we have seen, both the constant and the isotropic distance assumptions lead to working homing algorithms. While the single trajectories are often different, they result nonetheless in similar search density peaks for a given landmark array. This is due to the fact that both algorithms guide an agent towards a location where the match between the current image and the memorized snapshot is maximal. When looking only at the peaks of the search density pattern, all described algorithms agree well with the experimental data (cf. e.g. Cartwright and Collett 1983; or Wittmann 1995). As a consequence, differences between possible algorithms might be observed more easily on the level of single trajectories, or with respect to their catchment areas. As an example, one could build two landmark arrays in which only one of the assumptions described in our text is valid and the other is violated in order to decide between both assumptions (cf. Fig. 11). In both set-ups, the trajectories of the homing animals are recorded. If the animal uses, for example, the constant distance assumption, the average homeward component of its trajectories should be significantly higher in set-up *A* than in set-up *B*. The reverse would be obtained for the isotropic distance assumption.

### 6.2 Computational models

In this section, we will briefly discuss a number of scene–based homing schemes and point out some differences to the present approach. In doing so, we will mainly focus

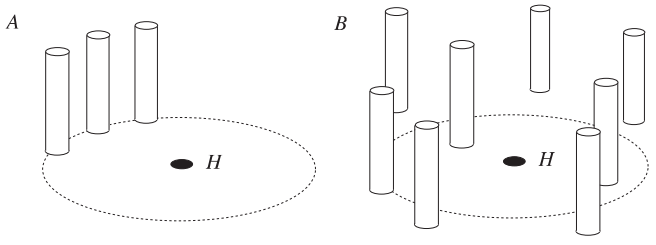



**Fig. 11.** In set-up A the isotropic distance is violated, while the constant distance assumption is valid. This condition is reversed in set-up B. An animal using, for instance, the constant distance assumption will show a significantly higher average homeward component in set-up A than in set-up B

**Table 1.** Overview of scene–based homing schemes (cf. Sect. 6)

| Reference | Approximation | Correspondence | Input | Constant orientation | Implementation |
|---|---|---|---|---|---|
| Cartwright and Collett 1983 | Isotropic landmark distribution | Region matching | Binary, 360° | Yes | Computer simulation |
| Hong et al. 1991 | Isotropic landmark distribution | Feature matching | Grey value, 360° | Yes | Mobile robot |
| Röfer 1995 | Isotropic landmark distribution | Kohonen network | Grey value, 360° | Yes | Mobile robot |
| Wittmann 1995 | Isotropic landmark distribution | Correlation on resolution pyramid | Grey value, 330° | Yes | Computer simulation |
| Röfer 1997 | Isotropic landmark distribution | Similarity to adjacent pixel pairs | RGB values and derivatives, 360° | No | Mobile robot |
| Franz et al. 1997 | Equal distance | Parameterized displacement fields | Grey value, 360° | No | Mobile robot |
| Möller et al. 1998 | Isotropic landmark distribution | Region matching | Binary, 360° | Yes | Mobile robot |

on the type of approximations and the correspondence mechanisms utilized (summarized in Table 1).

Most approaches use a 360° field of view. This simplifies the homing task considerably since, for an omnidirectional sensor, all non–occluded landmarks are permanently visible. As Nelson and Aloimonos (1988) pointed out, there is an additional advantage: in a 360° field of view, the rotatory and translatory part of the displacement field can easily be separated. In the case of limited fields of view, the decomposition becomes more difficult. Therefore, considerable effort has gone into technical implementations, including a camera pointing at a spherical (Hong et al. 1991; Röfer 1997) or conical mirror (Franz et al. 1997; Möller et al. 1998) and a rotating intensity sensor (Röfer 1995).

Cartwright and 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 the snapshots constant, either by 'mental' counterrotation or appropriate body orientation. The mobile robot of Möller et al. (1998) uses a polarized light compass (Lambrinos et al. 1997) to counterrotate a binary snapshot. Similarly, the camera platforms of the robot used by Hong et al. (1991) and Röfer (1995) do not rotate when the robot changes direction, so that all views have constant orientation. As pointed out in Sect. 5.3, this has the advantage of greatly reducing the computational cost. In addition, limited fields of view can be used without having to deal with the invisible parts, because the viewing direction always remains the same. The schemes of Cartwright and Collett (1983) and Wittmann (1995) are implemented in idealized computer models, so they do not have 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. (1991) and Röfer (1995). However, the orientation of the platforms is subject to cumulative errors, and thus these schemes may fail in large-scale environments.

Many schemes use additional assumptions for the computation of image displacements. Hong et al. (1991) assume that prominent features in the image can be attributed to objects around the robot. By observing only the displacement of these features, changes in the image due to different lighting or reflectance have less effect on homing performance. Röfer (1995, 1997) uses the assumption that landmarks maintain their order in the snapshot and the current view. While it restricts the search space for correspondences, this assumption is only true near the goal. It should be noted that although the above approaches differ in the way they establish correspondences between views, they all rely on the approximation of isotropic landmark distribution. The error in the computed home direction due to this approximation may be very large, even close to the goal (cf. Fig. 9). Thus, these schemes may converge very slowly in strongly non-isotropic environments, and even fail for higher noise levels.

Cartwright and Collett (1983) and the robot implementation of their scheme (Möller et al. 1998) included an additional feature to reproduce the experimental data: the vector sum for the computation of the home direction contains not only the tangential displacement 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 a segmentation of the image into objects and background.

### *6.3 Conclusion*

As the computation of displacement fields is an ill-posed problem, some additional assumption about the field has 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 displacement fields. The low–dimensionality leads to an optimization problem solvable in real time. All displacement 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 and Mallot 1995; Franz et al. 1998).

Since this work was largely inspired by biological principles, 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 and 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, three-dimensional information is only present implicitly, in the use of perspective distortion, and in the geometrical parameterization of displacement fields. Previous studies have shown that a variety of visual tasks (e.g., object recognition, see Bülthoff and Edelman 1992; navigation, Gillner and Mallot 1998) can be accomplished by humans without using explicit 3-D representations. Although these observations support our general approach, we emphasize that our homing algorithm is not an explicit model of animal behaviour. It aims rather at understanding possible solutions to a general problem which robots as well as animals have to solve.

*Acknowledgements*. The present work has greatly profited from discussions with and technical support by 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, the Human Frontier Science Program, and the Studienstiftung des deutschen Volkes.

## References


Anderson A (1977) A model for landmark learning in the honeybee. J Comp Physiol 114:335–355

Batschelet E (1981) Circular statistics in biology Academic Press, London

Brünnert U, Kelber A, Zeil J (1994) Ground-nesting bees determine the location of their nest relative to a landmark by other than angular size cues. J Comp Physiol A 175:363–369

Bülthoff HH, Edelman S (1992) Psychophysical support for a 2-D view interpolation theory of object recognition. Proc Nat Acad Sci 89:60–64

Cartwright BA, Collett TS (1983) Landmark learning in bees. J Comp Physiol A 151:521–543

Cartwright BA, Collett TS (1987) Landmark maps for honeybees. Biol Cybern 57:85–93

Chahl JS, Srinivasan MV (1996) Visual computation of egomotion using an image interpolation technique. Biol Cybern 74:405–411

Cheng K, Collett TS, Pickhard A, Wehner R (1987) The use of visual landmarks by honeybees: bees weight landmarks according to their distance from the goal. J Comp Physiol A 161:469–475

Collett TS (1992) Landmark learning and guidance in insects. Philos Trans R Soc Lond Biol 337:295–303

Collett TS (1996) Insect navigation en route to the goal: multiple strategies for the use of landmarks. J Exp Biol 199:227–235

Collett TS, Baron J (1994) Biological compasses and the coordinate frame of landmark memories in honey-bees. Nature 368:137–140

Collett TS, Cartwright BA, Smith BA (1986) Landmark learning and visuo-spatial memories in gerbils. J Comp Physiol A 158:835–851

Collett TS, Land MF (1975) Visual spatial memory in a hoverfly. J Comp Physiol 100:59–84

Förstner W (1993) Image matching. In: Haralick RM, Shapiro LG (eds) Computer and robot vision, Vol 2, Chapter 16. Addison Wesley, Reading, Ma.

Franz MO, Schölkopf B, Bülthoff HH (1997) Homing by parameterized scene matching. In: Husbands P, Harvey I (eds) Proc 4th Eur Conf Artif Life, pp 236–245 MIT Press, Cambridge, Ma.

Franz MO, Schölkopf B, Mallot HA, Bülthoff HH (1998) Learning view graphs for robot navigation. Autonomous Robots 5:111–125

Gillner S, Mallot HA (1998) Navigation and acquisition of spatial knowledge in a virtual maze. J Cogn Neurosci 10:445–463

Goldman S (1953) Information theory. Dover

Hong J, Tan X, Pinette B, Weiss R, Riseman EM (1991) Image-based homing. In: Proc IEEE Intl Conf on Robotics and Automation pp 620–625

Junger W (1991) Waterstriders ( *Gerris paludum* F.) compensate for drift with a discontinously working visual position servo. J Comp Physiol A 169:633–639

Krapp HG, Hengstenberg R (1996) Estimation of self-motion by optic flow processing in single visual interneurons. Nature 384:463–466

Lambrinos D, Maris M, Kobayashi H, Labhart T, Pfeifer R, Wehner R (1997) An autonomous agent navigating with a polarized light compass. Adaptive Behavior 6:131–161

Mallot HA, Arndt PA, Bülthoff HH (1996) A psychophysical and computational analysis of intensity-based stereo. Biol Cybern 75:187–198

Mallot HA, Bülthoff HH, Little JJ, Bohrer S (1991) Inverse perspective mapping simplifies optical flow computation. Biol Cybern 64:177–185

Möller R, Lambrinos D, Pfeifer R, Labhart T, Wehner R (1998) Modeling ant navigation with an autonomous agent. In: Pfeifer R, Blumberg B, Meyer J-A, Wilson SW (eds) From animals to animals 5. MIT Press Cambridge, Ma pp 185–194

Morris RGM (1981) Spatial localization does not require the presence of local cues. Learning Motiv 12:239–260

Nelson RC, Aloimonos J (1988) Finding motion parameters from spherical motion fields (or the advantages of having eyes in the back of your head). Biol Cybern 1988:261–273

Pelzer V (1985) Die Bedeutung von Helligkeits- und Farbkontrast bei der Erkennung und Lokalisation der Futterquelle anhand künstlicher Nahmarken: Verhaltensexperimente mit *Apis mellifera*. PhD thesis, University of Zürich

Röfer T (1995) Controlling a robot with image-based homing. (Report 3/95) Center for Cognitive Sciences, Bremen

Röfer T (1997) Controlling a wheelchair with image-based homing. Proc. AISB workshop on spatial reasoning in mobile robots and animals. Dept Computer Sci, Manchester University (Technical Report UMCS-97-4-1)

Schölkopf B, Mallot HA (1995) View-based cognitive mapping and path planning. Adapt Behav 3:311–348

Sherry DF, Duff SJ (1996) Behavioural and neural bases of orientation in food-storing birds. J Exp Biol 199:165–172

Trullier O, Wiener SI, Berthoz A, Meyer J-A (1997) Biologically based artificial navigation systems: review and prospects. Prog Neurobiol 51:483–544

Wehner R, Michel B, Antonsen P (1996) Visual navigation in insects: coupling of egocentric and geocentric information. J Exp Biol 199:129–140

Wehner R, Müller M (1985) Does interocular transfer occur in visual navigation by ants? Nature 315:228–229

Wittmann T (1995) Modeling landmark navigation. (Report 3/95) Center for Cognitive Sciences, Bremen

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

Zeil J (1993) Orientation flights in solitary wasps. 2. Similarities between orientation and return flights and the use of motion parallax. J Comp Physiol A 172:207–222