Available online at www.sciencedirect.com
SCIENCE DIRECT
Robotics and Autonomous Systems 54 (2006) 300–313
Robotics and Autonomous Systems
www.elsevier.com/locate/robot

Efficient visual homing based on Fourier transformed panoramic images☆

W. Stürzl[a,b,*], H.A. Mallot[a]

[a] Cognitive Neuroscience, Department of Zoology, Tübingen University, Auf der Morgenstelle 28, 72076 Tübingen, Germany
[b] Centre for Visual Sciences, Research School of Biological Sciences, Australian National University, Biology Place, Canberra, ACT 0200, Australia

Received 8 December 2004; received in revised form 28 November 2005; accepted 8 December 2005
Available online 9 February 2006

Abstract

We present a fast and efficient homing algorithm based on Fourier transformed panoramic images. By continuously comparing Fourier coefficients calculated from the current view with coefficients representing the goal location, a mobile robot is able to find its way back to known locations. No prior knowledge about the orientation with respect to the goal location is required, since the Fourier phase is used for a fast sub-pixel orientation estimation. We present homing runs performed by an autonomous mobile robot in an office environment. In a more comprehensive investigation the algorithm is tested on an image data base recorded by a small mobile robot in a toy house arena. Catchment areas for the proposed algorithm are calculated and compared to results of a homing scheme described in [M. Franz, B. Schölkopf, H. Mallot, H. Bülthoff, Where did I take that snapshot? Scene based homing by image matching, Biological Cybernetics 79 (1998) 191–202] and a simple homing strategy using neighbouring views. The results show that a small number of coefficients is sufficient to achieve a good homing performance. Also, a coarse-to-fine homing strategy is proposed in order to achieve both a large catchment area and a high homing accuracy: the number of Fourier coefficients used is increased during the homing run.

Keywords: Visual homing; Panoramic images; Orientation estimation; Fourier transformation

1. Introduction

The ability to return to known places is important for most animals and also for mobile robots. Snapshot or template-based models have been developed to explain the homing behaviour of insects, for instance honey bees [1,2] and desert ants [3]. In order to estimate the "home direction", the current image is continuously compared to the image ("snapshot") obtained at the goal location ("home") and stored in memory. It is assumed that only low-level image processing operations are required.

Although visual homing is confined to a certain region around the snapshot location called "catchment area" [2], large areas can be covered by using several snapshots taken at different locations. Between neighbouring positions, one can then move using visual homing.

In this paper we address the following fundamental aspects of view-based homing:

- size of the catchment area, i.e. region of the environment in which visual homing using a single snapshot succeeds — the larger the catchment area of snapshots the smaller the number of images that have to be stored;
- accuracy, i.e. final distance to the goal location;
- memory usage of a single snapshot — since many snapshots usually have to be stored for navigation in large cluttered environments, this can be a limiting factor;
- computational effort for the estimation of the "home vector", i.e. the direction to the goal.

1.1. Connection and comparison to other work

Several homing algorithms can be found in the literature; for reviews, see [4–8]. Probably the most parsimonious model is the average landmark vector (ALV) model [9,10], where only two vector components have to be stored. Like some other algorithms, e.g. [11–13], the ALV approach assumes that the current orientation relative to the goal position is known

☆ Supported by the Deutsche Forschungsgemeinschaft, SFB 550 and STU 413/1-1.

* Corresponding author at: Australian National University, Research School of Biological Sciences, Centre for Visual Sciences, Building 46, Biology Place, ACT 0200, Canberra, Australia. Fax: +61 02 612 53808.
E-mail addresses: wolfgang.stuerzl@uni-tuebingen.de (W. Stürzl), hanspeter.mallot@uni-tuebingen.de (H.A. Mallot).

0921-8890/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2005.12.001

(e.g. by using a compass). This is a reasonable assumption since a compass sense has been found in many animals, e.g. in desert ants and honey bees which use the polarisation pattern of the sky[1] or pigeons which have a magnetic sense; see [14–17]. However, relying on a compass limits the usage of these algorithms on mobile robots to applications where the orientation can be estimated by other means.

Several researchers have used image-based localisation combined with principal component analysis to reduce memory usage and to speed up image comparison, e.g. [18,19]. In the paper by Jogan and Leonardis [20] an orientation invariant localisation method is described that can also deal with a high degree of occlusion in images. Recently, Menegatti et al. [21] use Fourier transformed images for image-based localisation. The number of Fourier coefficients is chosen according to the required accuracy. Pajdla and Hlavac [22] proposed an orientation invariant representation for panoramic images based on the phase of the first Fourier coefficient ("zero phase representation") to allow for image comparison without knowledge of relative orientation. In this paper we take a different approach. In the first step the orientation is estimated using more than one coefficient. Then the similarity is calculated for the estimated orientation.

The homing algorithm proposed in this paper can be understood as an extension of the algorithm of Franz et al. [23] to the Fourier or frequency domain. The main advantages are:

- significantly faster estimation of home direction and relative orientation (with sub-pixel accuracy);
- efficient memory usage, since only a few Fourier coefficients are necessary to achieve good homing performance.

To achieve fast estimation of home vectors, approximations are used that may result in smaller average catchment area compared to the algorithm of Franz et al. [23].

2. Theoretical framework: Derivation of the Fourier-based homing algorithm

2.1. Image difference function

As a similarity measure of two images consisting of N pixels, $\mathbf{I}^{\mathrm{A/B}} = (I^{\mathrm{A/B}}[0], I^{\mathrm{A/B}}[1], \ldots, I^{\mathrm{A/B}}[N-1])^{\top}$, we use the sum of squared differences (SSD),

$$\mathrm{E_I}(\mathbf{I}^{\mathrm{A}}, \mathbf{I}^{\mathrm{B}}) = \sum_i \left((I^{\mathrm{A}}[i] - \bar{I}^{\mathrm{A}}) - (I^{\mathrm{B}}[i] - \bar{I}^{\mathrm{B}}) \right)^2, \tag{1}$$

where the mean pixel value ($\bar{I}^{\mathrm{A/B}}$) is subtracted in order to account for possible changes in brightness. In the following, only panoramic images consisting of a one dimensional (1D) array of gray values are considered. The image acquisition is shown in Fig. 1. Like many other robotic systems, we apply a catadioptric sensor consisting of a video camera directed towards a convex mirror to achieve a large field of view [24–26].

[1] The "Sahabot 2" of Lambrinos et al. (2000) uses a custom-build polarisation compass [9].

Visual homing using images is based on the assumption that the place, where the image difference reaches its minimum, is the goal position.[2] This can be described mathematically as finding (with respect to the current pose) position x^*, y^* and orientation difference ς^* that minimise $\mathrm{E_I}\left(\mathbf{I}^{\mathrm{h}}, \mathbf{I}(x, y, \varsigma)\right)$. The superscript 'h' denotes the image at the goal position ("home").

2.2. The homing algorithm of "Franz et al., 1998"

The homing algorithm described in [23] can be understood as replacing $\mathbf{I}(x, y, \varsigma)$ by hypothetical or expected images $\mathbf{I}^{\mathrm{e}}(x, y, \varsigma)$. These images are calculated from the current image using the assumption that all surrounding objects have the same distance to the agent.[3] This "equal distance assumption" allows estimation of the home direction without having to compute distances to objects explicitly or having to search for correspondences of image patches or image features; see [8] for a detailed discussion. Although it may be a very coarse approximation to the actual distance distribution of the surrounding scene, it has been shown that the error due to this approximation decreases with decreasing displacement from the goal position [23].

If the current (continuous) panoramic image is given by $I(\phi)$, then by means of simple trigonometry (Fig. 2) the expected image can be computed according to

$$I^{\mathrm{e}}(\phi|x, y, \varsigma) = I\left(\phi + \varsigma - \arcsin\left(\frac{x}{R}\sin(\phi+\varsigma) - \frac{y}{R}\cos(\phi+\varsigma)\right)\right), \tag{2}$$

where R denotes the assumed equal distance.[4] Even without any knowledge about the magnitude of R, the direction described by the parameters $\xi = x/R$ and $\eta = y/R$ can be estimated by the algorithm using

$$I^{\mathrm{e}}(\phi|\xi, \eta, \varsigma) = I(\phi + \varsigma - \arcsin(\xi\sin(\phi+\varsigma) - \eta\cos(\phi+\varsigma))). \tag{3}$$

This can be simplified using an approximation for $\xi^2 + \eta^2 \ll 1$, i.e.

$$I^{\mathrm{e}}(\phi|\xi, \eta) \approx I(\phi - \xi\sin\phi + \eta\cos\phi). \tag{4}$$

Put together, the home vector is estimated according to

$$(\xi^*, \eta^*) \approx \arg\min_{\xi,\eta,\varsigma} \tilde{\mathrm{E}}_{\mathrm{I}}\left(I^{\mathrm{h}}(\phi), I^{\mathrm{e}}(\phi|\xi,\eta,\varsigma)\right) \tag{5}$$

$$= \arg\min_{\xi,\eta,\varsigma} \int_{-\pi}^{\pi} \left((I^{\mathrm{h}}(\phi-\varsigma) - \bar{I}^{\mathrm{h}}) - (I^{\mathrm{e}}(\phi|\xi,\eta) - \bar{I}^{\mathrm{e}}) \right)^2 \mathrm{d}\phi. \tag{6}$$

[2] Other constraints can also be considered, e.g. that image difference should be below a certain threshold.

[3] If distances are known, images at all positions of a convex environment, i.e. an environment without occlusions, can be calculated from a single image — with the precision limited only by the resolution of the vision sensor. Of course, this is also true for the calculation of distances; see [28].

[4] To simplify notation, we assume that images are 2π-periodic functions.

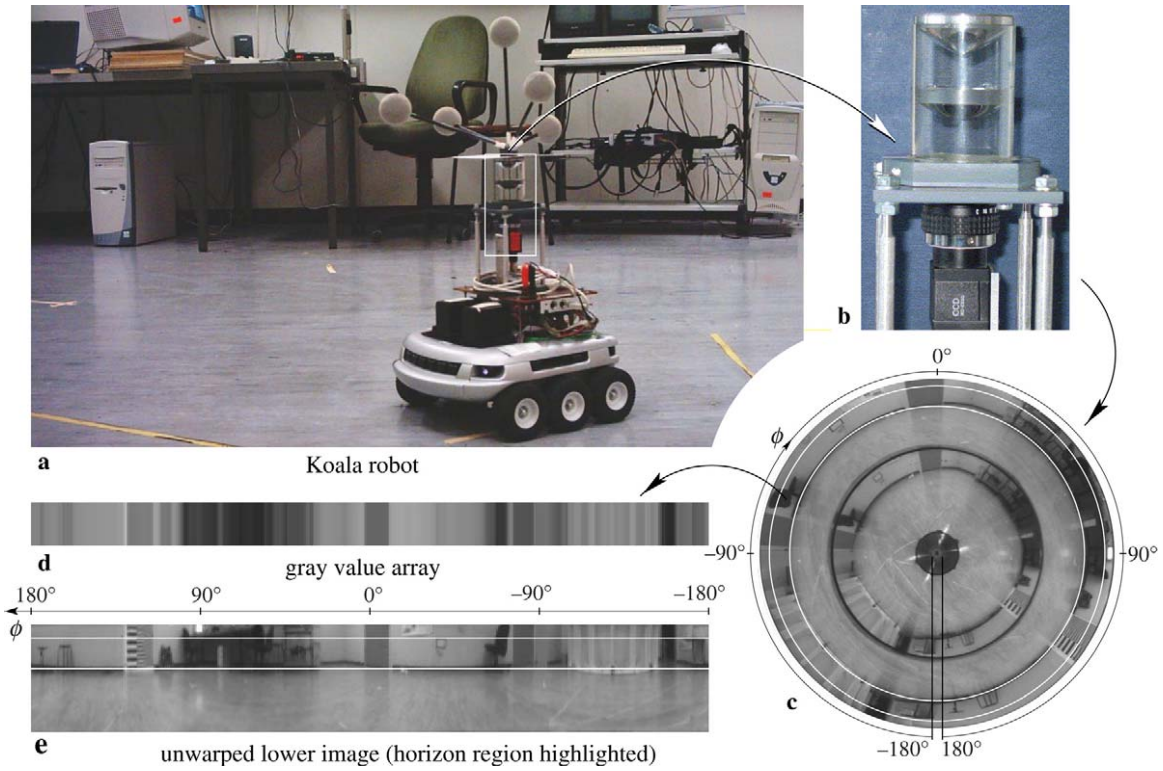

Fig. 1. a: Koala robot with panoramic vision. Image processing and control of the robot is performed by a PC/104 on-board computer (Pentium III, 700 MHz). On top of the vision sensor a reflective target is mounted which allows for high precision position and orientation estimation by a tracking system (A.R.T., http://www.ar-tracking.de). b: Panoramic stereo sensor consisting of two separate mirrors. A hole in the lower mirror permits imaging via the upper mirror. For further details, see [27]. The stereo information is not used for image-based homing described in this paper. c, d: The image part, where the $\pm 5°$-region around the horizon is mapped to, is converted into an array of $N = 360$ gray values. An unwarped version of the image created by the lower mirror is shown in e.

For pixel images, Franz et al. estimated the minimum over ς simply by testing all possible integer pixel shifts, i.e. $\varsigma = \frac{2\pi}{N}s$, $s = 0, 1, \ldots, N-1$:

$$(\xi^*, \eta^*) \approx \arg\min_{\xi,\eta,\varsigma} \mathrm{E_I}\left(\mathbf{I}^{\mathrm{h}}, \mathbf{I}^{\mathrm{e}}(\xi, \eta, \varsigma)\right) \tag{7}$$

$$= \arg\min_{\xi,\eta,s} \sum_i \Big((I^{\mathrm{h}}[(i-s) \bmod N] - \bar{I}^{\mathrm{h}}) - (I^{\mathrm{e}}[i|\xi, \eta] - \bar{I}^{\mathrm{e}})\Big)^2. \tag{8}$$

Also, for (ξ, η), a discrete set of P positions $\{(\xi_p, \eta_p)\}_{p=0,1,\ldots,P-1}$ is evaluated. For fast calculation of the expected images, we use a discrete version of approximation (4),

$$I^{\mathrm{e}}[i|\xi_p, \eta_p] \approx I\left[\mathrm{rint}\left(i - \xi_p \sin\left(\frac{2\pi}{N}i\right)\frac{N}{2\pi} + \eta_p \cos\left(\frac{2\pi}{N}i\right)\frac{N}{2\pi}\right)\right], \tag{9}$$

where 'rint( )' means "round to nearest integer". The pixel index on the right side of Eq. (9) can be easily computed using a look-up table of size $P \times N$. In our experiments we found no negative effect of this approximation on the homing performance of our robots. Since the SSD is evaluated for all N possible orientations, the number of operations for the algorithm of Franz et al. is of order $O(P \times N^2)$.

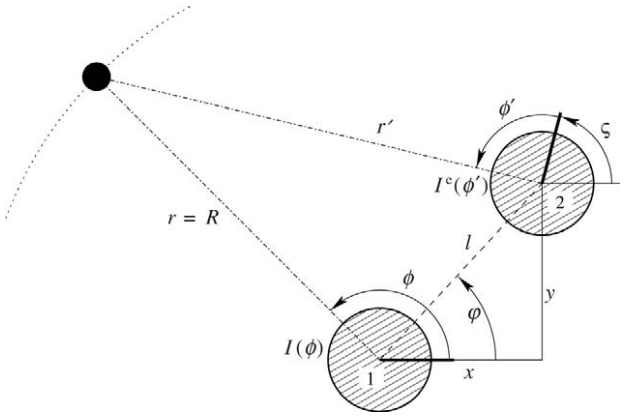

Fig. 2. Derivation of Eq. (2), i.e. the calculation of an expected image $I^{\mathrm{e}}(\phi')$ at position 2 based on the current image $I(\phi)$ at position 1 for a heading difference ς (the heading of the robot is marked by thick lines; the black point depicts an object in distance R to the robot). From the drawing we find $(r'\cos(\phi'+\varsigma)+x, r'\sin(\phi'+\varsigma)+y)^\top = (R\cos\phi, R\sin\phi)^\top$. Solving for ϕ yields $\phi = \Phi(\phi') = \phi' + \varsigma - \arcsin(x/R\sin(\phi'+\varsigma) - y/R\cos(\phi'+\varsigma))$. Hence, the expected image can be calculated using $I^{\mathrm{e}}(\phi') = I(\Phi(\phi'))$, which is identical to (2) after renaming $\phi' \to \phi$.

2.3. Transformation to Fourier space

In this section we will first express Eq. (6) in the Fourier domain, which will then lead – by means of approximations – to an efficient homing algorithm.

Since gray values are real numbers, we use the following two real-valued Fourier representations of images ("sine–cosine-representation" and "amplitude-phase-representation"):

$$I(\phi) \approx \frac{1}{2}\mathfrak{a}_0 + \sum_{k=1}^{K} \mathfrak{a}_k \cos(k\phi) + \mathfrak{b}_k \sin(k\phi) \tag{10}$$

$$= \frac{1}{2}\mathfrak{a}_0 + \sum_{k=1}^{K} \mathcal{A}_k \cos(k\phi + \psi_k), \tag{11}$$

where $K \leq N/2$ is the number of frequencies used in the approximation of $I(\phi)$.[5] The coefficients, amplitudes and phases can be calculated according to

$$\mathcal{A}_k \cos\psi_k = \mathfrak{a}_k = \frac{2}{N}\sum_j I[j] \cos\left(k\frac{2\pi}{N}j\right), \tag{12}$$

$$-\mathcal{A}_k \sin\psi_k = \mathfrak{b}_k = \frac{2}{N}\sum_j I[j] \sin\left(k\frac{2\pi}{N}j\right). \tag{13}$$

Using Parseval's theorem [29], the image difference function (1) can be transformed into (see footnote 5)

$$\mathrm{E_I}(\mathbf{I}^\mathrm{A}, \mathbf{I}^\mathrm{B}) \approx \frac{N}{2}\mathcal{E}_\mathrm{I}(\mathbf{I}^\mathrm{A}, \mathbf{I}^\mathrm{B}, K), \tag{14}$$

$$\mathcal{E}_\mathrm{I}(\mathbf{I}^\mathrm{A}, \mathbf{I}^\mathrm{B}, K) = \sum_{k=1}^{K} (\mathfrak{a}_k^\mathrm{A} - \mathfrak{a}_k^\mathrm{B})^2 + (\mathfrak{b}_k^\mathrm{A} - \mathfrak{b}_k^\mathrm{B})^2. \tag{15}$$

$\mathcal{E}_\mathrm{I}(\mathbf{I}^\mathrm{A}, \mathbf{I}^\mathrm{B}, K)$ is a similarity measure for the first K Fourier coefficients of two images. Thus, Eq. (6) can be replaced by

$$(\xi^*, \eta^*) \approx \arg\min_{\xi,\eta,\varsigma} \mathcal{E}_\mathrm{I}\left(\mathbf{I}^\mathrm{h}, \mathbf{I}^\mathrm{e}(\xi, \eta, \varsigma), K\right) \tag{16}$$

$$= \arg\min_{\xi,\eta,\varsigma} \left(\sum_k (\mathfrak{a}_k^\mathrm{h}(\varsigma) - \mathfrak{a}_k^\mathrm{e}(\xi, \eta))^2 + (\mathfrak{b}_k^\mathrm{h}(\varsigma) - \mathfrak{b}_k^\mathrm{e}(\xi, \eta))^2 \right), \tag{17}$$

where $\mathfrak{a}_k(\varsigma) = \mathfrak{a}_k \cos(k\varsigma) + \mathfrak{b}_k \sin(k\varsigma)$, $\mathfrak{b}_k(\varsigma) = \mathfrak{b}_k \cos(k\varsigma) - \mathfrak{a}_k \sin(k\varsigma)$. If an estimation ς_0 of the relative orientation exists, i.e. $\varsigma = \varsigma_0 + \Delta\varsigma$, then for $k|\Delta\varsigma| \ll 1$ we can use the approximations

$$\mathfrak{a}_k^\mathrm{h}(\varsigma) \approx \mathfrak{a}_k^\mathrm{h}(\varsigma_0) + \mathfrak{b}_k^\mathrm{h}(\varsigma_0) k\Delta\varsigma, \tag{18}$$

$$\mathfrak{b}_k^\mathrm{h}(\varsigma) \approx \mathfrak{b}_k^\mathrm{h}(\varsigma_0) - \mathfrak{a}_k^\mathrm{h}(\varsigma_0) k\Delta\varsigma. \tag{19}$$

Note that the constraint $k|\Delta\varsigma| \ll 1$ restricts the use of high frequencies.

To simplify further, we consider a linearisation of Eq. (4):

$$I^\mathrm{e}(\phi|\xi, \eta) \approx I(\phi) - \partial_\phi I(\phi)(\xi \sin\phi - \eta\cos\phi). \tag{20}$$

Since the derivative of $I(\phi)$ can be calculated easily, substituting (10) on both sides of Eq. (20) leads to

$$\mathfrak{a}_k^\mathrm{e}(\xi, \eta) = \mathfrak{a}_k + \mathfrak{a}_k^x \xi + \mathfrak{a}_k^y \eta, \tag{21}$$

$$\mathfrak{b}_k^\mathrm{e}(\xi, \eta) = \mathfrak{b}_k + \mathfrak{b}_k^x \xi + \mathfrak{b}_k^y \eta, \tag{22}$$

$$\mathfrak{a}_k^x = \frac{1}{2}[-(k-1)\mathfrak{a}_{k-1} + (k+1)\mathfrak{a}_{k+1}], \tag{23}$$

$$\mathfrak{a}_k^y = \frac{1}{2}[(k-1)\mathfrak{b}_{k-1} + (k+1)\mathfrak{b}_{k+1}], \tag{24}$$

$$\mathfrak{b}_k^x = \frac{1}{2}[-(k-1)\mathfrak{b}_{k-1} + (k+1)\mathfrak{b}_{k+1}], \tag{25}$$

$$\mathfrak{b}_k^y = \frac{1}{2}[-(k-1)\mathfrak{a}_{k-1} - (k+1)\mathfrak{a}_{k+1}]. \tag{26}$$

Using Eqs. (18), (19), (21) and (22), we obtain

$$\mathcal{E}_\mathrm{I}\left(\mathbf{I}^\mathrm{h}, \mathbf{I}^\mathrm{e}(\xi, \eta, \varsigma), K\right) \approx \mathcal{E}_1(\varsigma_0) + \mathcal{E}_2(\xi, \eta, \Delta\varsigma|\varsigma_0), \tag{27}$$

where

$$\mathcal{E}_1(\varsigma_0) = \sum_k [\mathfrak{a}_k^\mathrm{h}(\varsigma_0) - \mathfrak{a}_k]^2 + [\mathfrak{b}_k^\mathrm{h}(\varsigma_0) - \mathfrak{b}_k]^2 \tag{28}$$

$$= \sum_k (\mathcal{A}_k^\mathrm{h})^2 + \mathcal{A}_k^2 - 2\mathcal{A}_k^\mathrm{h}\mathcal{A}_k \cos(\psi_k^\mathrm{h} - \psi_k + k\varsigma_0), \tag{29}$$

$$\begin{aligned}\mathcal{E}_2(\xi, \eta, \Delta\varsigma|\varsigma_0) &= \sum_k 2[\mathfrak{a}_k^\mathrm{h}(\varsigma_0) - \mathfrak{a}_k][\mathfrak{b}_k^\mathrm{h}(\varsigma_0)k\Delta\varsigma - \mathfrak{a}_k^x\xi - \mathfrak{a}_k^y\eta] \\ &- 2[\mathfrak{b}_k^\mathrm{h}(\varsigma_0) - \mathfrak{b}_k][\mathfrak{a}_k^\mathrm{h}(\varsigma_0)k\Delta\varsigma + \mathfrak{b}_k^x\xi + \mathfrak{b}_k^y\eta] \\ &+ [\mathfrak{b}_k^\mathrm{h}(\varsigma_0)k\Delta\varsigma - \mathfrak{a}_k^x\xi - \mathfrak{a}_k^y\eta]^2 \\ &- [\mathfrak{a}_k^\mathrm{h}(\varsigma_0)k\Delta\varsigma + \mathfrak{b}_k^x\xi + \mathfrak{b}_k^y\eta]^2.\end{aligned} \tag{30}$$

In (27) the similarity function $\mathcal{E}_\mathrm{I}$ is split up into two parts: $\mathcal{E}_1$ depends only – however non-linearly – on the relative orientation, and $\mathcal{E}_2$ consists only of expressions up to second order in the positional parameters ξ, η and the orientation correction $\Delta\varsigma$, assuming that a coarse orientation estimate ς_0 is known. This representation can be used to minimise $\mathcal{E}_\mathrm{I}$ in a computationally efficient way.

Algorithm for Fourier-based home vector estimation

For an efficient estimation of the home vector ξ^*, η^* and the relative orientation ς_0^*, we propose the following two-step procedure:

1. Minimise $\mathcal{E}_1(\varsigma_0)$ with respect to ς_0 using a fast rotation estimation (with sub-pixel accuracy) based on Fourier phases described in Section 2.4. This yields an estimation of the orientation difference ς_0^*.
2. Minimise $\mathcal{E}_2(\xi, \eta, \Delta\varsigma|\varsigma_0^*)$ with respect to $\xi, \eta, \Delta\varsigma$. As can be seen from Eq. (30), this leads to a system of linear equations. The solution of this system is the estimated home vector ξ^*, η^* (and an orientation correction $\Delta\varsigma^*$, i.e. the estimated total orientation difference is $\varsigma^* = \varsigma_0^* + \Delta\varsigma^*$).

Steps 1 and 2 can be iterated easily using the update $\mathfrak{a}_k + \mathfrak{a}_k^x\xi^* + \mathfrak{a}_k^y\eta^* \to \mathfrak{a}_k$, $\mathfrak{b}_k + \mathfrak{b}_k^x\xi^* + \mathfrak{b}_k^y\eta^* \to \mathfrak{b}_k$. The estimated home vector is then the sum of all shifts: $\sum_{i=1}^{N_\mathrm{it}+1} \xi_i^*$, $\sum_{i=1}^{N_\mathrm{it}+1} \eta_i^*$, where N_it is the number of iterations.[6]

[5] The approximations in Eqs. (10) and (14) differ from true equality due to the number of reconstructing frequencies (K) and sampling effects.

[6] Alternatively, one could start with small K for orientation and home vector estimation, and iterate while increasing K. This approach has not yet been investigated.

Since the number of calculations needed in steps 1 and 2 are of the order $O(K)$, the proposed algorithm needs $O((N_{\text{it}}+1)\times K)$ operations. In addition, K Fourier coefficients have to be calculated resulting in an overall complexity of $O((N+N_{\text{it}}+1)\times K)$. Due to the approximations, smaller catchment areas are expected for the Fourier-based homing compared to the algorithm of Franz et al. [23]; see Section 4.

2.4. Fourier phase-based correlation and orientation estimation[7]

To estimate the relative orientation of two images, we consider the dependence of their correlation on image shift ς[8]:

$$\mathrm{C_I}(\mathbf{I}^{\mathrm{A}},\mathbf{I}^{\mathrm{B}}(\varsigma)) = \sum_i (I^{\mathrm{A}}[i]-\bar{I}^{\mathrm{A}})\left(I^{\mathrm{B}}\left[i+\frac{N}{2\pi}\varsigma\right]-\bar{I}^{\mathrm{B}}\right) \quad (31)$$

$$\overset{(*)}{\approx} \frac{N}{2}\mathcal{C}_{\mathrm{I}}(\mathbf{I}^{\mathrm{A}},\mathbf{I}^{\mathrm{B}}(\varsigma),K), \quad (32)$$

$$\mathcal{C}_{\mathrm{I}}(\mathbf{I}^{\mathrm{A}},\mathbf{I}^{\mathrm{B}}(\varsigma),K) = \sum_{k=1}^{K}\mathcal{A}_k^{\mathrm{A}}\mathcal{A}_k^{\mathrm{B}}\cos(\psi_k^{\mathrm{A}}-\psi_k^{\mathrm{B}}-k\varsigma). \quad (33)$$

In $(*)$, Eq. (11) and $I[i]\approx I(\frac{2\pi}{N}i)$ have been substituted. In Fig. 3, Eq. (33) is illustrated using two images recorded by the Khepera robot (see Section 4) at the same position but with a different heading. Note that, in Fig. 3a, there are local maxima in addition to the global maximum, preventing an orientation estimation using simple gradient descent. In Fig. 3b, the maximum of correlation differs for $K=1$ by more than 10° from the real image shift. The reason for this is probably sensor noise and the occluded 12° image region (see Section 4). Thus, for a better orientation estimation, more coefficients have to be used.

To estimate the optimal shift ς^*, we have to solve

$$\frac{\partial}{\partial\varsigma}\mathcal{C}_{\mathrm{I}}(\mathbf{I}^{\mathrm{A}},\mathbf{I}^{\mathrm{B}}(\varsigma),K) = \sum_k \mathcal{A}_k^{\mathrm{A}}\mathcal{A}_k^{\mathrm{B}}\sin(\psi_k^{\mathrm{A}}-\psi_k^{\mathrm{B}}-k\varsigma)k = 0. \quad (34)$$

If the difference in $\mathbf{I}^{\mathrm{A}}$ and $\mathbf{I}^{\mathrm{B}}$ is mainly due to the orientation difference ς, then (for all k): $\mathcal{A}_k^{\mathrm{B}}\approx\mathcal{A}_k^{\mathrm{A}}$ and

$$\psi_k^{\mathrm{A}}-\psi_k^{\mathrm{B}}-k\varsigma+2\pi n_k\approx 0, \quad (35)$$

where n_k is an integer number. Therefore, an estimation of ς can be found for each frequency:

$$\varsigma\approx\varsigma_k = \frac{\psi_k^{\mathrm{A}}-\psi_k^{\mathrm{B}}+2\pi n_k}{k}. \quad (36)$$

[7] An earlier version of this section using a different notation has been published in [30].

[8] Since $\mathrm{E_I}(\mathbf{I}^{\mathrm{A}},\mathbf{I}^{\mathrm{B}}(\varsigma)) = \sum_i (I^{\mathrm{A}}[i]-\bar{I}^{\mathrm{A}})^2+(I^{\mathrm{B}}[i|\varsigma]-\bar{I}^{\mathrm{B}})^2-2\,\mathrm{C_I}(\mathbf{I}^{\mathrm{A}},\mathbf{I}^{\mathrm{B}}(\varsigma))$, the rotation estimation by minimising the SSD leads to the same result as maximising the correlation.

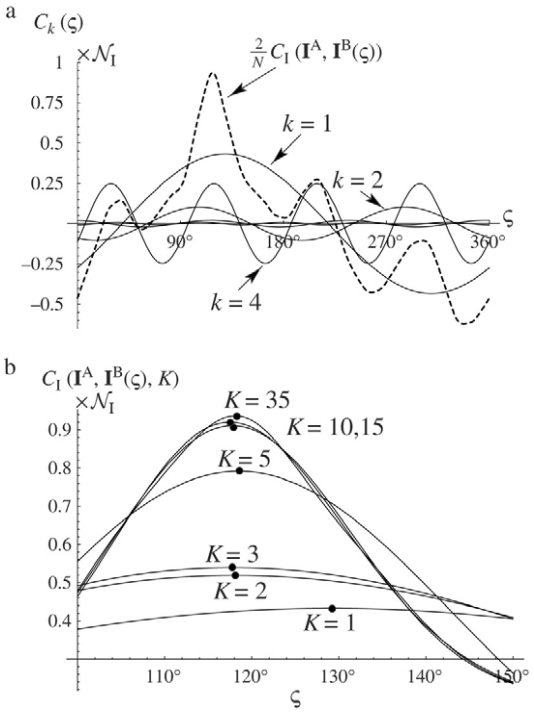

Fig. 3. a: First five terms $\mathcal{C}_k(\varsigma)=\mathcal{A}_k^{\mathrm{A}}\mathcal{A}_k^{\mathrm{B}}\cos(\psi_k^{\mathrm{A}}-\psi_k^{\mathrm{B}}-k\varsigma)$ of the sum in Eq. (33) for two panoramic images with an orientation difference of approximately 120° ($\mathcal{C}_3(\varsigma)$ and $\mathcal{C}_5(\varsigma)$ have very small amplitudes). The scaling constant of the y-axis is $\mathcal{N}_{\mathrm{I}}=\frac{2}{N}\sqrt{\mathrm{C_I}(\mathbf{I}^{\mathrm{A}},\mathbf{I}^{\mathrm{A}})\mathrm{C_I}(\mathbf{I}^{\mathrm{B}},\mathbf{I}^{\mathrm{B}})}$. The dashed curve shows the dependence of the correlation, Eq. (31), on the image shift ς. b: Maximum of the correlation for different numbers of Fourier coefficients.

However, a unique solution with $\varsigma_k\in[0,2\pi)$ exists only for $k=1$; see Fig. 3a. Substitution of (35) into (34) yields[9]

$$0\approx\sum_k\mathcal{A}_k^{\mathrm{A}}\mathcal{A}_k^{\mathrm{B}}k(\psi_k^{\mathrm{A}}-\psi_k^{\mathrm{B}}-k\varsigma+2\pi n_k) \quad (37)$$

$$\Longleftrightarrow \varsigma\approx\bar{\varsigma}=\frac{\sum_k\mathcal{A}_k^{\mathrm{A}}\mathcal{A}_k^{\mathrm{B}}k(\psi_k^{\mathrm{A}}-\psi_k^{\mathrm{B}}+2\pi n_k)}{\sum_k\mathcal{A}_k^{\mathrm{A}}\mathcal{A}_k^{\mathrm{B}}k^2} \quad (38)$$

$$\overset{(36)}{=}\frac{\sum_k\mathcal{A}_k^{\mathrm{A}}\mathcal{A}_k^{\mathrm{B}}k^2\varsigma_k}{\sum_k\mathcal{A}_k^{\mathrm{A}}\mathcal{A}_k^{\mathrm{B}}k^2}=\frac{\sum_k w_k\varsigma_k}{\sum_k w_k},\quad w_k=\mathcal{A}_k^{\mathrm{A}}\mathcal{A}_k^{\mathrm{B}}k^2. \quad (39)$$

Algorithm for fast orientation estimation

In order to arrive at a correct estimation of ς from Eq. (38) or (39), the integers $\{n_k\}$ have to be known for frequencies $k>1$. To achieve this, we propose the following "coarse-to-fine" algorithm for orientation estimation with sub-pixel accuracy:

$$\bar{\varsigma}_1=\varsigma_1=\psi_1^{\mathrm{A}}-\psi_1^{\mathrm{B}},\quad w_1=\mathcal{A}_1^{\mathrm{A}}\mathcal{A}_1^{\mathrm{B}}, \quad (40)$$

`for` $k=2,3,\ldots,K$ `do:`

$$n_k=\mathrm{rint}\left(\frac{\psi_k^{\mathrm{A}}-\psi_k^{\mathrm{B}}+k\bar{\varsigma}_{k-1}}{2\pi}\right)\in\mathbb{Z}, \quad (41)$$

[9] For $\psi_k^{\mathrm{A}}-\psi_k^{\mathrm{B}}-k\varsigma+2\pi n_k+\pi$ we get the minima of the image correlation, since $\cos(\pi+2\pi n)=-1, n\in\mathbb{Z}$.

$$\varsigma_k = \frac{\psi_k^{\mathrm{A}} - \psi_k^{\mathrm{B}} + 2\pi n_k}{k}, \quad w_k = \mathcal{A}_k^{\mathrm{A}} \mathcal{A}_k^{\mathrm{B}} k^2, \tag{42}$$

$$\bar{\varsigma}_k = \frac{\sum_{l=1}^{k} w_l \varsigma_l}{\sum_{l=1}^{k} w_l} = \frac{\left(\sum_{l=1}^{k-1} w_l\right) \bar{\varsigma}_{k-1} + w_k \varsigma_k}{\sum_{l=1}^{k-1} w_l + w_k} \tag{43}$$

`endfor`

'rint( )' in (41) means – as in Eq. (9) – "round to nearest integer". The estimated shift $\bar{\varsigma}_K$ can now be used to calculate the image correlation, Eq. (33).

The algorithm depends on a roughly correct estimation of ς_1. Therefore, although occurring rarely (since images usually have high spectral power in low frequencies; see Section 4.1.1), very small and noise-prone amplitudes $\mathcal{A}_1^{\mathrm{A}}$, $\mathcal{A}_1^{\mathrm{B}}$ are critical. If $w_1 = \mathcal{A}_1^{\mathrm{A}} \mathcal{A}_1^{\mathrm{B}}$ is smaller than a certain threshold ($0.1\ \frac{1}{K} \sum_k w_k$ in the current implementation), we compute a second estimate of ς starting with $\bar{\varsigma}_1' = \bar{\varsigma}_1 + \pi$. The shift with the larger correlation is then assumed to be the correct estimation.[10] However, individual small values of w_k, $k > 1$ are not critical, since $\bar{\varsigma}_K$ is calculated as a weighted sum. Of course, the weights w_k can easily be substituted by a different weighting function. $w_k = \mathcal{A}_k^{\mathrm{A}} \mathcal{A}_k^{\mathrm{B}} k^2$ has been derived directly from maximising image correlation.

3. Experiments with "Koala" robot

We performed several homing runs on a mobile robot ("Koala", K-team, http://www.k-team.org) in an office environment; see Fig. 1. In the following, we denote the algorithm of Franz et al. [23] by `hs_ei` (homing strategy using expected images) and the algorithm derived in Section 2 by `hs_Fi` (homing strategy based on Fourier transformed images).

First, an image $\mathbf{I}^{\mathrm{h}}$ was recorded at the goal position (and Fourier coefficients $\{\mathfrak{a}_k^{\mathrm{h}}, \mathfrak{b}_k^{\mathrm{h}}\}$ were calculated). Then the robot was driven manually to several places in the environment. From each starting position, a homing run was performed by the robot according to the following procedure:

(1) take image at current position (calculate $\{\mathfrak{a}_k, \mathfrak{b}_k\}$ for Fourier-based homing) and estimate home vector (ξ^*, η^*)
(2) if $(\xi^*)^2 + (\eta^*)^2 < \varepsilon$ stop ($\varepsilon = 10^{-4}$)
(3) rotate by $\varphi = \mathrm{atan2}(\eta^*, \xi^*)$ and move by $l = \min[l^{\max}, R^{\mathrm{typ}} \sqrt{(\xi^*)^2 + (\eta^*)^2}]$ (R^{typ} was set to a typical distance to surrounding objects, i.e. $R^{\mathrm{typ}} = 3$ m; translation was limited to $l^{\max} = 0.3$ m in order to reduce the influence of a single wrong estimate.[11])
(4) goto 1.

Fig. 4 shows homing runs for different goal positions using the proposed algorithm `hs_Fi` with $K = 5$ spatial frequencies and no iteration. Fig. 4b also shows homing runs performed by means of algorithm `hs_ei` (size of pixel array $N = 360$) which seems to have a larger catchment area (see next section).

[10] If, in addition to w_1, successive weights w_k, $l = 2, 3, \ldots$ are small, a larger number of starting estimates for ς would have to be considered.

[11] Of course, other values of R^{typ} are possible — a smaller value would lead to more homing steps

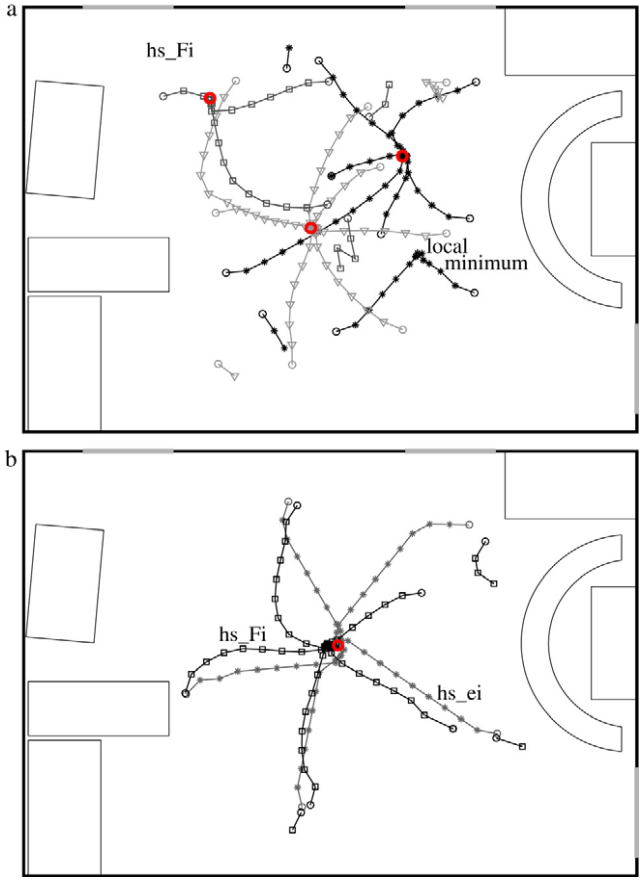

Fig. 4. Homing runs of the Koala robot using the Fourier-based algorithm `hs_Fi` with $K = 5$ coefficients, no iteration (room size $\approx 7 \times 5$ m^2). Also, some unsuccessful trials (corresponding start positions are not in catchment area) are plotted leading away from the goal or to a different position (label "local minimum"). a: Homing trials for three different goal positions. The symbols $\triangledown$, $\square$, $*$ indicate homing runs to different goals, which are marked by large circles. Start positions are marked by small circles. b: Comparison of homing runs using `hs_Fi` (indicated by $\square$) with homing runs using the algorithm of Franz et al. `hs_ei` (indicated by $*$) for approximately the same goal position.

As a coarse measure of the performance of both algorithms, we measured the time needed to compute a single home vector estimate on the Koala's PC/104 (Pentium 3,700 MHz). We found the following computation times:

- approximately 100 ms using the algorithm `hs_ei` (current implementation: $N = 360$ pixel, $P \approx 200$ test positions $\{(\xi_p, \eta_p)\}$ with $\xi_p^2 + \eta_p^2 < (\frac{2}{3})^2$);
- approximately 1 ms using the proposed Fourier-based algorithm `hs_Fi` ($K = 5$, $N_{\mathrm{it}} = 0$).

4. Evaluation using image data sets

To evaluate the homing performance in a more systematic way, two data sets (1250 images each; size of the extracted gray value arrays is $N = 72$ pixel) were autonomously recorded by a small Khepera robot in a toy house arena on two consecutive days; see Fig. 5. Illumination conditions and recording positions were approximately the same but, due to sensor noise and the limited accuracy of the tracking system, corresponding images of the two sets differ slightly. In

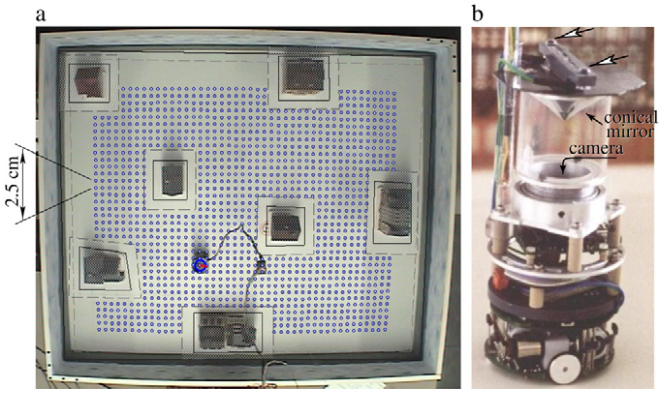

Fig. 5. Recording of image data sets. a: Top view of the toy house arena (140 cm × 120 cm) seen by the tracking camera at the ceiling. Superimposed are the recording positions (dots) which lie on a rectangular grid (cell size 2.5×2.5 cm^2). b: Position and heading of the Khepera robot ($\varnothing \approx 5.5$ cm; H $\approx$ 13 cm) was measured by a tracking system using two light-emitting diodes (accuracy: position $\approx$3 mm; heading $\approx 2°$). On the left, the cable can be seen that connects the robot to an external computer where image processing and the control of the robot is done.

addition, the cable to an external computer occluded part of the image ($\approx 12°$ horizontally). The corresponding pixel values are estimated using linear interpolation between the neighbouring gray values.

Using these image data sets, different homing algorithms can be compared easily based on identical input data. Since they allow off-line computation of home vectors, catchment areas – which are difficult to investigate using homing runs of a mobile robot – can be estimated; see Section 4.4.

4.1. Properties of the image sets

We first discuss two properties of the image sets that are relevant for the Fourier-based algorithm.

4.1.1. Power spectrum

The power spectrum of the image data sets is shown in Fig. 6. As in most natural images, low frequencies have high amplitudes [31,32]. This fact is usually described by a power law, i.e. $\langle \mathcal{A}_k^2 \rangle \propto k^{-\alpha}$. For the 1D images in the toy house arena, we find $\alpha \approx 1.8$. This finding supports the use of only a few coefficients (low frequencies) for rotation and home vector estimation.

4.1.2. Change of Fourier coefficients and phases with distance

To estimate the influence of the approximations used for the derivation of `hs_Fi`, we computed the change of Fourier coefficients with distance, i.e.

$$\mathcal{E}_k(l) = (\mathfrak{a}_k^{\mathrm{h}} - \mathfrak{a}_k(l))^2 + (\mathfrak{b}_k^{\mathrm{h}} - \mathfrak{b}_k(l))^2. \tag{44}$$

In Fig. 7a the mean value of $\mathcal{E}_k(l)$ is shown for different k. If the linearisation in Eqs. (21) and (22) were exact, one would get $\langle \mathcal{E}_k(l) \rangle \propto l^2$. As can be seen, this is true in a wider range of l for lower frequencies. Therefore, one expects better home vector estimations using only low frequencies if the distance to the goal is large.

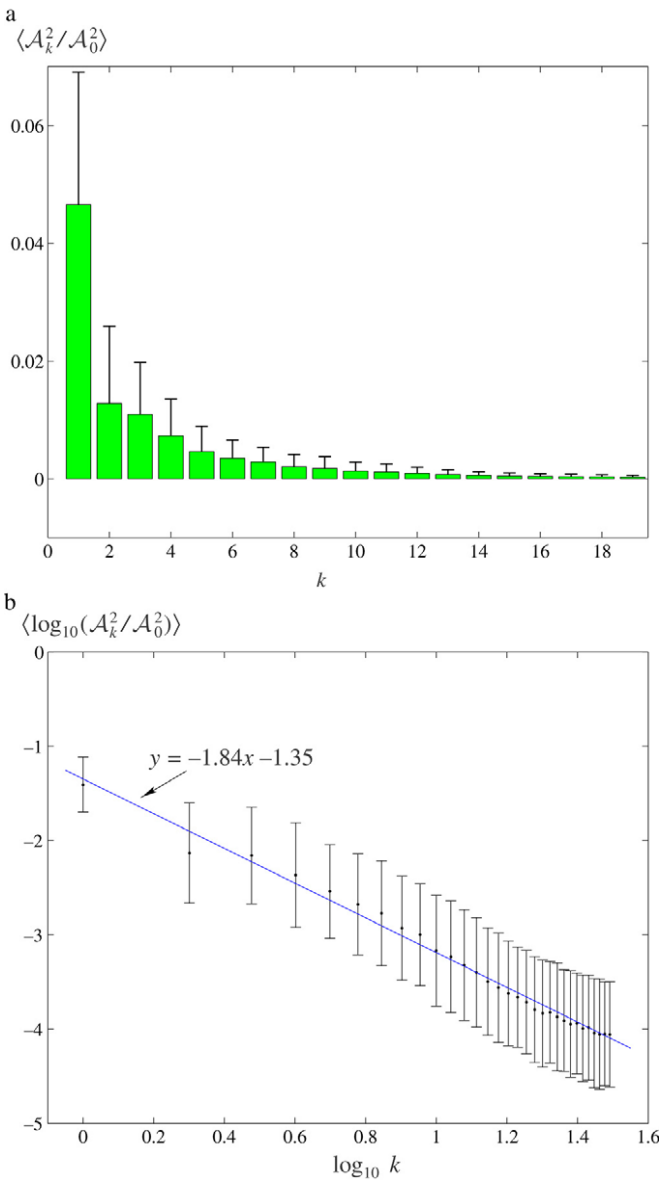

Fig. 6. Frequency power spectrum for image data sets of toy house arena (error bars show standard deviation).

Since, in the first step of `hs_Fi`, the orientation relative to the goal is estimated based on Fourier phases, we also calculated

$$\Psi_k(l) = \left(\frac{(\mathfrak{a}_k^{\mathrm{h}}, \mathfrak{b}_k^{\mathrm{h}})^\top}{\sqrt{(\mathfrak{a}_k^{\mathrm{h}})^2 + (\mathfrak{b}_k^{\mathrm{h}})^2}} - \frac{(\mathfrak{a}_k(l), \mathfrak{b}_k(l))^\top}{\sqrt{(\mathfrak{a}_k(l))^2 + (\mathfrak{b}_k(l))^2}} \right)^2 \tag{45}$$

$$= ((\cos\psi_k^{\mathrm{h}}, \sin\psi_k^{\mathrm{h}})^\top - (\cos\psi_k^{\mathrm{h}}(l), \sin\psi_k^{\mathrm{h}}(l))^\top)^2 \tag{46}$$

$$= 2\left(1 - \cos\left(\psi_k^{\mathrm{h}} - \psi_k(l)\right)\right), \tag{47}$$

which is a measure of phase change with distance. The mean value of $\Psi_k(l)$ for different k is plotted in Fig. 7b. $\langle \Psi_k(l) \rangle$ increases much more strongly for high frequencies and soon reaches the point where phases are unrelated, i.e. $\langle \Psi_k(l) \rangle \approx 2$, since $2(1 - \langle \cos\delta\psi \rangle_{\delta\psi}) = 2$.

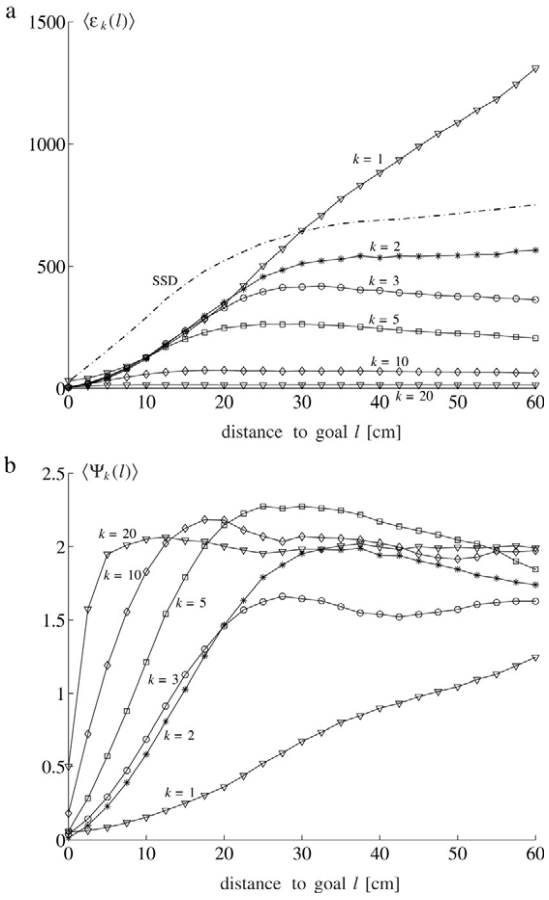

Fig. 7. Change of Fourier coefficients with spatial distance (all 1250 positions of the image data were evaluated using the same image orientation). a: Shown is $\langle \mathcal{E}_k(l) \rangle = \langle (\mathfrak{a}_k^{\mathrm{h}} - \mathfrak{a}_k(l))^2 + (\mathfrak{b}_k^{\mathrm{h}} - \mathfrak{b}_k(l))^2 \rangle$ for $k = 1, 2, 3, 5, 10, 20$. The dash–dotted curve shows the sum $\sum_{k=1}^{N/2} \langle \mathcal{E}_k(l) \rangle$ (scaled by the factor 1/5) which is equal to $\frac{2}{N} \langle E_{\mathrm{I}}(\mathbf{I}^{\mathrm{h}}, \mathbf{I}(l)) \rangle$. The small offset at $l = 0$, visible for $k = 1$, is due to the use of two image data sets. b: Change in Fourier phase $\langle \Psi_k(l) \rangle = 2(1 - \langle \cos(\psi_k^{\mathrm{h}} - \psi_k(l)) \rangle)$ for different k. The large offset for $k = 20$ is due to the use of two image sets and the limited accuracy of the tracking system – an orientation difference of $\Delta\varsigma \approx 2^\circ$ results in $\delta\psi_{20} \approx \frac{2\pi}{9}$; see Eq. (36), and $2(1 - \cos\delta\psi_{20}) \approx 0.5$.

4.2. Orientation estimation

The dependence of the error in the first step of `hs_Fi`, i.e. the orientation estimation based on Fourier phases (described in Section 2.4), on the distance between two images is shown in Fig. 8. All possible combinations of image position were evaluated and the circular standard deviation [33] calculated. For small distances, the error decreases with the number of coefficients[12] (better signal to noise ratio); for larger distances a smaller K results in a slightly smaller error. The reason for this is that the rotation estimation was derived assuming only orientation differences (and no position differences), and Fourier phases for high frequencies usually change more rapidly with position (see Fig. 7b).

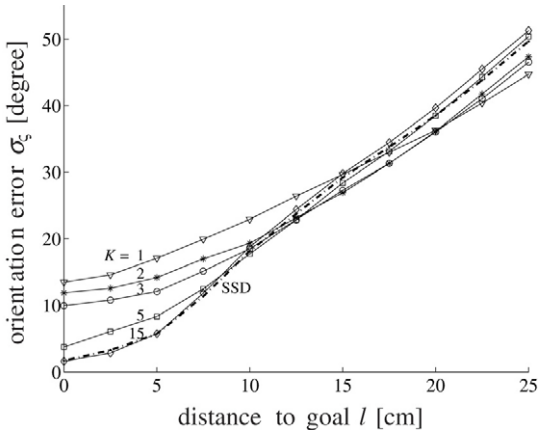

Fig. 8. Error of orientation estimation, depending on distance for different numbers of Fourier coefficients: $K = 1$ (▽), $K = 2$ (∗), $K = 3$ (○), $K = 5$ (□) and $K = 15$ (◇) (plots for $K \geq 10$ are indistinguishable from the $K = 15$ curve at this resolution). The dash–dotted curve shows the result if all integer shifts for the SSD are tested, i.e. $\varsigma^* = 2\pi/N\, s^*$, $s^* = \arg\min_s \sum_i (I^{\mathrm{A}}[i] - \bar{I}^{\mathrm{A}})(I^{\mathrm{B}}[(i+s) \bmod N] - \bar{I}^{\mathrm{B}})$.

4.3. Home vector estimation

An example of home vectors calculated using the proposed Fourier-based algorithm is shown in Fig. 9. Most vectors point into a direction that locally increases the image similarity. The appearance of local minima shows clearly the limitations of view-based homing (and also of appearance-based localisation).[13] A homing trial started from an arbitrary position may end at the wrong place. The similarity in the neighbourhood of these locations is sometimes much higher than at locations inside the catchment area of the true goal location, making it difficult to find a threshold with respect to image similarity that prevents such failures without reducing the catchment area significantly. However, it may be possible to detect a mistake by using a threshold for the image similarity at the final position. Of course, some coarse metric information, e.g. from path integration, can help to avoid this problem. The high image similarity at "wrong places" is caused by the low number of pixels used, and also by the regular layout of the toy house arena.

4.3.1. Quality measure "average homeward component" (AHC)

Since the homing procedure described above is an iterative process, a single home vector estimate does not have to be very accurate. Deviations can be corrected easily in the following steps. However, it is important that (at least in the majority of steps) the estimated home vector has a component in the direction of the true home direction. If the home vector is estimated continuously, the agent will get closer to the goal position as long as the angle between the estimated and the true home directions ($\mathbf{v}^* = (\cos\varphi^*, \sin\varphi^*)^\top = \frac{1}{\sqrt{(\xi^*)^2+(\eta^*)^2}}(\xi^*, \eta^*)^\top$ and $\mathbf{v}^{\mathrm{h}}$, respectively) is less than

[12] The error value of $\sigma_\varphi \approx 1.5^\circ$ for $l = 0$ is already below the accuracy of the tracking system that was used to determine the orientation differences of the images.

[13] Increasing the amount of sensor input (image size) will of course reduce the probability of high image similarity at different places.

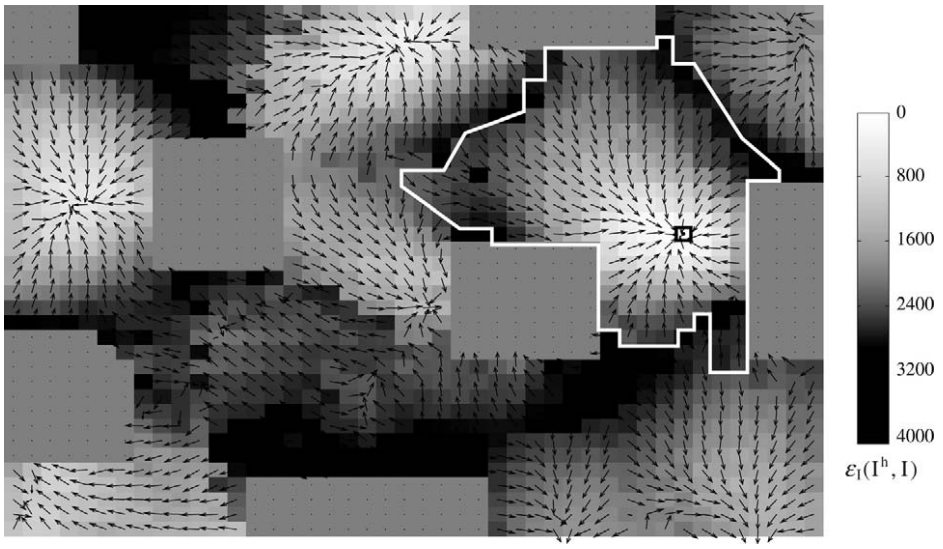

Fig. 9. Fourier-based home vector estimation (`hs_Fi` with $K = 5$, $N_{it} + 1 = 3$) using image data sets from the toy house arena (uniform gray areas depict positions of toy houses where no images were recorded; see Fig. 5). The gray values code the difference, Eq. (15), of the image at each position compared to the image at the home position (marked by a square). The white curve shows the borders of the catchment area. Note the local minima in other parts of the arena.

90°. Therefore, the dependence of the "average homeward component" (AHC) on the distance l to the goal can be used to assess the quality of the calculated home vectors:

$$\mathrm{AHC}(l) = \langle \mathbf{v}^*(l)\mathbf{v}^h \rangle = \langle \cos(\varphi^*(l) - \varphi^h) \rangle, \tag{48}$$

where $\mathbf{v}^*$ has a component along $\mathbf{v}^h \iff \mathbf{v}^*\mathbf{v}^h > 0$.

Fig. 10 shows AHCs for different numbers of Fourier coefficients K and iterations N_{it}. From 10b it can be seen that a few iterations can improve the estimation of the home direction. It is obvious that a small K, i.e. using only low spatial frequencies, leads to higher AHC for larger distances, but to lower AHC for close positions. Reasons for this are:

- Fourier phases for low spatial frequencies vary less with position than do high frequencies, which leads to a smaller error in rotation estimation (which assumes identical positions); see Section 4.2.
- Linearisations in Eqs. (20)–(22) and (18), (19) have a larger valid range for low frequencies; see Section 4.1.2.
- A larger number of coefficients yields better signal to noise ratio close to goal, where linearisation errors are small.

The linearisations are also responsible for smaller AHCs of `hs_Fi` compared to `hs_ei` for distances in the range of 15 cm $< l <$ 25 cm.

The strong decrease in the AHC for small distances for `hs_ei` is mainly due to the use of a fixed grid of positions where expected images are calculated (increasing the number of positions on the grid close to the current position will therefore improve the AHC for small l). In addition, the direction of the estimated home vector is very sensitive to noise close to the goal, since the length of the vector tends towards zero.

4.4. Catchment areas of simulated homing runs

In order to estimate the size of catchment areas, homing runs were simulated using the image data sets. To adapt the homing

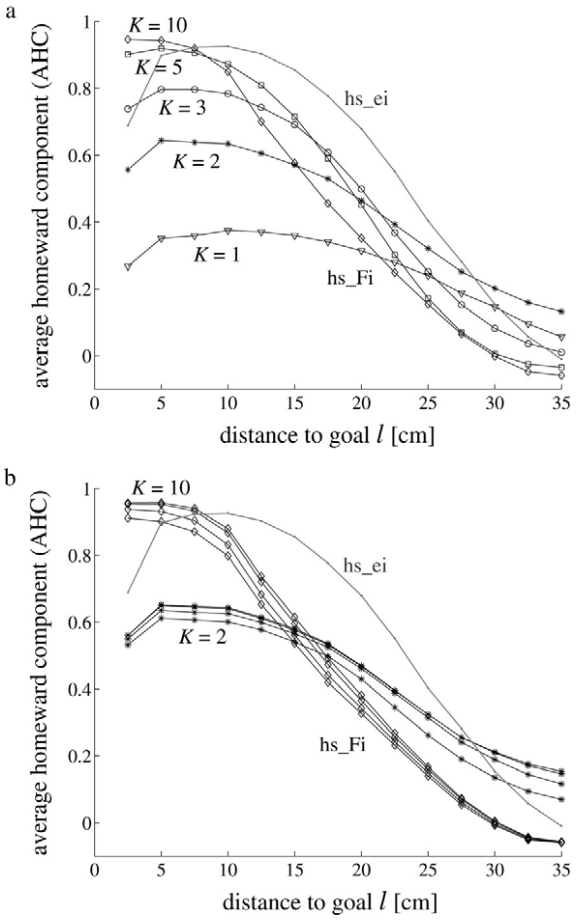

Fig. 10. Average homeward component depending on distance to the goal for `hs_Fi` and `hs_ei` (gray curve), averaged over all 1250 home positions of one set of images. The "current" images are taken from the other set (bin width is 2.5 cm). a: Dependence on number of coefficients for `hs_Fi` (no iteration): $K = 1$ (▽), $K = 2$ (∗), $K = 3$ (○), $K = 5$ (□) and $K = 10$ (◇). b: Dependence on the number of iterations for `hs_Fi` ($K = 2$ and $K = 10$): $N_{it}+1 = 1$ (black), $N_{it} + 1 = 2$ (dashed), $N_{it} + 1 = 5$ (dotted), $N_{it} + 1 = 10$ (dash–dotted).

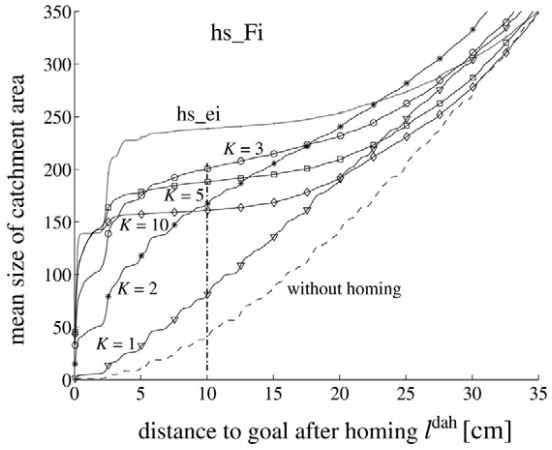

Fig. 11. Mean size of catchment area (average of 1250 homing runs) for `hs_Fi` ($N_{it}+1=3$) for different numbers of Fourier coefficients: $K=1$ (▽), $K=2$ (∗), $K=3$ (○), $K=5$ (□) and $K=10$ (◇). The gray curve shows the mean catchment area for `hs_ei`. The vertical line marks $l^{dah}=10$ cm; see Fig. 12. The steps appearing at $l^{dah}\approx 2.5$ cm stem from homing runs that stopped at a position in the direct neighbourhood of the goal.

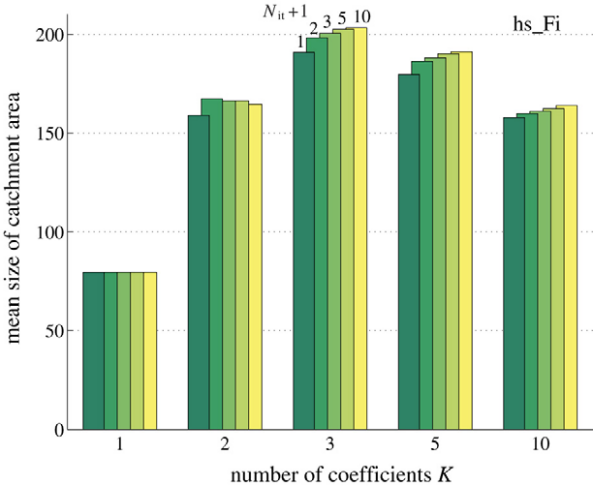

Fig. 12. Dependence of the mean size of the catchment area ($l^{dah}=10$ cm) on the number of iterations N_{it} for different K. The values for $N_{it}+1=3$ correspond to the values along the dash–dotted line in Fig. 11.

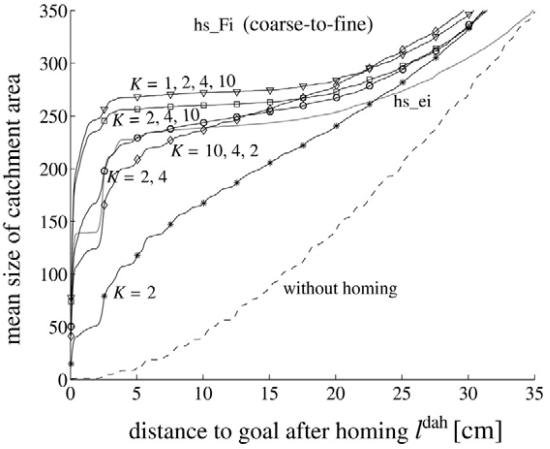

Fig. 13. The dependence of the mean size of the catchment area on the final deviation from the goal for `hs_Fi` ($N_{it}+1=3$) for different combinations of frequencies: $K=2$ (∗), $K=2,4$ (○), $K=2,4,10$ (□), $K=1,2,4,10$ (□), $K=10,4,2$ (◇). For better comparison, the gray curve shows the result again for `hs_ei` (as in Fig. 11).

procedure described in Section 3 to the discrete grid of image positions, some modifications had to be made:

(1) calculate home vector at current position (parameters set to $R^{typ}=50$ cm, $l^{max}=5$ cm)
(2) move to grid position closest to estimated home vector
(3) stop homing run if
  - no change in position,
  - number of homing steps >30,
  - distance to a position on grid >2 cm (“obstacle”)
(4) goto 1.

Although the catchment area can be defined simply as the area where homing succeeds, one has to specify what accuracy is needed for a homing run in order to be counted as successful. Due to sensor noise, small misalignment of the mirror, minor changes in the environment or in illumination, etc., it may happen that a homing run does not end exactly at the goal position but at some distance to the goal. For the simulated homing runs we define the dependence of the mean size of the catchment area (MCA) on the accepted maximum deviation l^{dah} as the mean number of grid positions (out of 1250) for which the distance to the goal position after the homing run is below l^{dah}. The area belonging to one position on the grid is $(2.5\text{ cm})^2$. For each of the 1250 goal positions of one image set, homing is started for all 1250 positions of the other image set.[14]

Fig. 11 compares the mean catchment area of `hs_ei` to `hs_Fi` for different numbers of Fourier coefficients. If no homing run is performed, that is, when l^{dah} is simply the distance from the starting position to the goal position, one expects that the mean catchment area increases according to $\text{MCA}\propto (l^{dah})^2$; see dashed curves in Fig. 11. Since the size of the environment is limited, this is true only for small l^{dah} (of course, the limit is $\text{MCA}=1250$ for $l^{dah}\gtrsim 145$ cm). As could be expected from the AHC results, the homing accuracy increases with the number K of Fourier coefficients. However, for medium level accuracy, smaller K can lead to larger MCA.

Fig. 12 shows the improvement in the mean catchment area if the Fourier-based algorithm is iterated. The allowed maximum homing error is set to $l^{dah}\approx 10$ cm. In addition to the fast estimation of home vectors, another advantage of the Fourier-based algorithm (`hs_Fi`) is now obvious: for $K=3$, only six real numbers have to be stored.

The approximations used for `hs_Fi` cause (at least using the image data sets) a smaller average catchment area ($\approx -20\%$) compared to `hs_ei`. This effect can easily be compensated for by storing more Fourier images, e.g. using $K=3$ frequencies (i.e. 6 coefficients $\{a_k, b_k\}$) and `float` variables (4 bytes), three images can be stored, which have the same size (72 bytes) as a 72 pixel gray image using `unsigned char` (1 byte) but –

[14] If a position, visited in a previous homing run, is encountered, the current homing run is stopped and the final deviation to the home is copied.

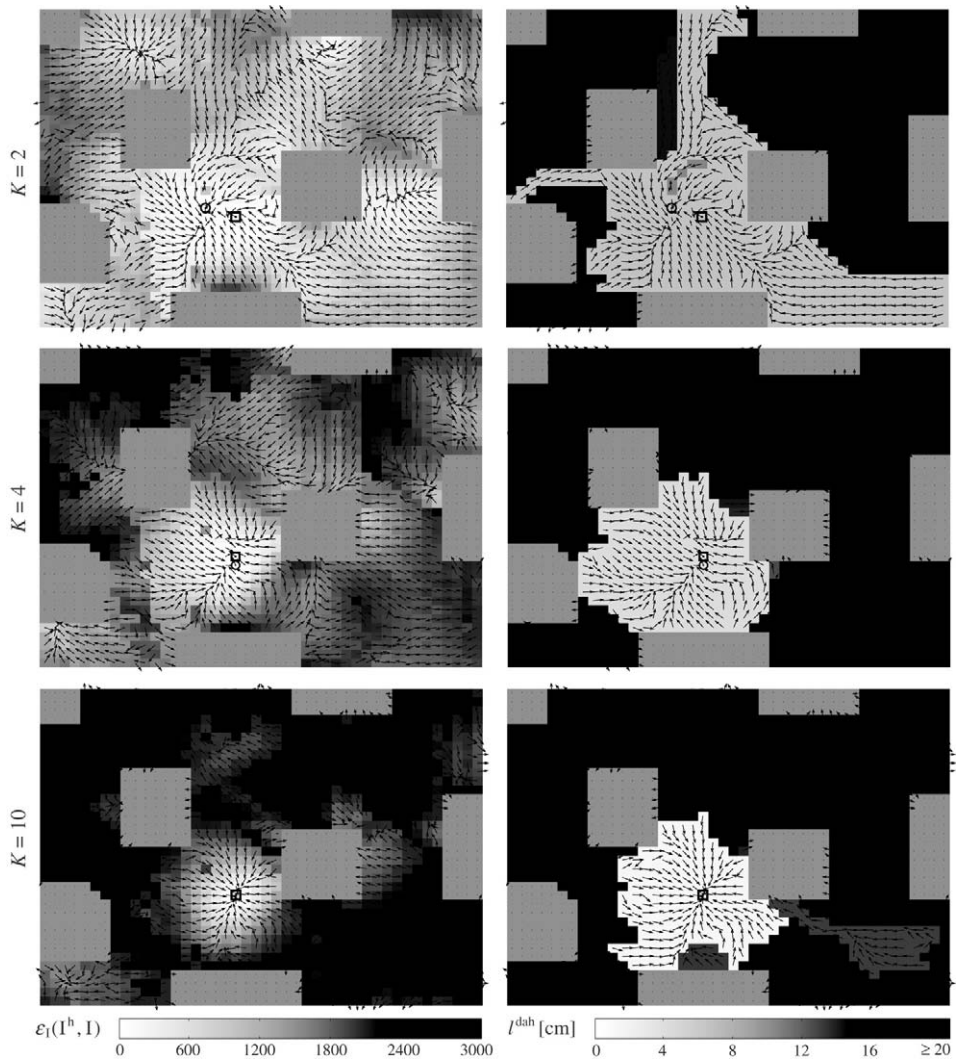

Fig. 14. Example of coarse-to-fine homing using `hs_Fi` for a single goal position (marked by a square; the position of highest image similarity is marked by a circle if different). Left: Image similarity (gray coded) and estimated home vectors (black arrows) for different numbers of Fourier coefficients, i.e. $K = 2, 4, 10$ (from top to bottom). At the position 3 grid units above and 4 grid units to the left of the goal position, there seems to be a "larger error" in the image data set, probably caused by an accidental switching on of an additional light source during the recording of one of the image data sets. Right: Catchment area and distance after homing run l^{dah} (gray coded). Usually, catchment areas decrease with K whereas the homing accuracy increases.

depending on the selection of the recording positions – can have a larger "effective catchment area".

5. Coarse-to-fine Fourier-based homing

With the results presented in the previous section in mind – especially the dependence of the AHC and the MCA on the number of Fourier coefficients and the distance to the goal – we propose the following approach in order to achieve both a large catchment area and a small deviation from the goal position: "Start homing with a small number of Fourier coefficients (low frequencies) and increase the number as the distance to the goal decreases". Since the absolute distance to the goal may not be easily estimated, one has to rely on other measures, e.g. the change in image similarity. For our experiments, we tested an even simpler approach: "Start with small K (low frequencies); when homing stops, increase K and continue".

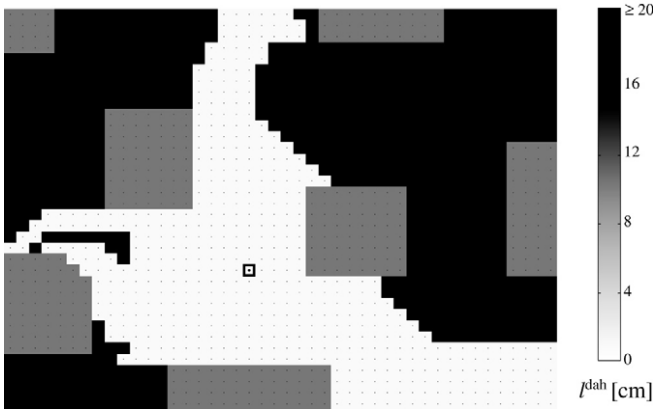

Fig. 15. Total catchment area for "coarse-to-fine homing" using the sequence "$K = 2, 4, 10$". For all starting points inside the catchment area (white) the goal grid cell is reached (no home vectors are shown since their directions depend on K, i.e. there is no unique vector at each position).

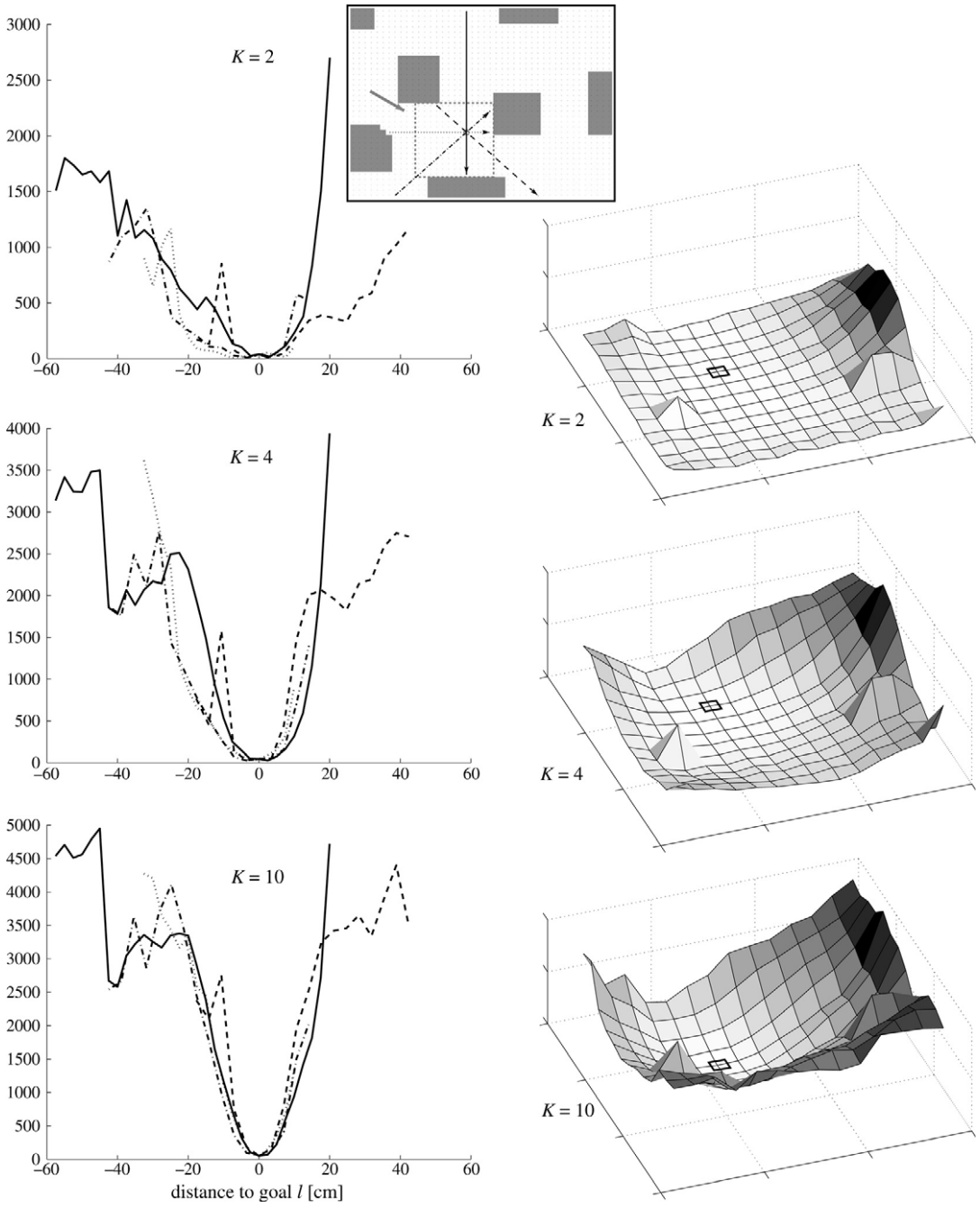

Fig. 16. Plots of image similarity function in the vicinity of the goal (position as in Fig. 14). Top centre: The arrows mark different transects for which the image difference is shown on the left. The dashed rectangle highlights the part of the arena for which the image difference function is plotted on the right (the gray arrow indicates the viewing direction). Left: Transects of the image similarity function for (from top to bottom) $K = 2, 4, 10$. Home position is at $l = 0$. Right: Three dimensional plot of image similarity function for $K = 2, 4, 10$. Home position is marked by a rectangle.

The catchment areas achieved by this coarse-to-fine strategy using `hs_Fi` are shown in Fig. 13. By comparison to Fig. 11, it is clear that this approach enhances the average size of catchment areas. Part of the improvement is due simply to the increased number of homing steps. However, comparison of the sequence "$K = 2, 4, 10$" and the reverse sequence "$K = 10, 4, 2$" shows that the coarse-to-fine approach achieves a much higher homing accuracy.

Fig. 14 shows an example of home vector fields and catchment areas for the different stages of "$K = 2, 4, 10$". For each grid cell of the catchment area, the distance after the homing runs is shown gray coded on the right side. While homing accuracy increases with K, the catchment area becomes smaller. The "combined coarse-to-fine catchment area" is shown in Fig. 15. As can be seen, both a large catchment area and a small deviation from the goal is achieved. Transects and three dimensional plots of the image difference function are shown in Fig. 16. With increasing K, the minimum becomes more pronounced and steeper.

To our knowledge, it is not yet clear if insects use a similar mechanism for visual homing. However, it has been suggested by Cartwright and Collett [2] that honey bees use images that are filtered according to object distances: image parts containing distant objects can be used when the bee is far away from the goal. While approaching the goal, closer image parts are added to achieve a more accurate estimate of the home direction.

6. Extensions of the Fourier-based homing scheme and future work

We have presented a Fourier-based homing algorithm that uses 1D images representing a 10° region around the horizon. As long as the vertical field of view is small, this algorithm can be directly applied row-wise to two dimensional images. However, for a larger field of view, vertical image motion has to be taken into account when calculating expected images. We also work on an extension to non-planar ground and movements in three dimensions. This will allow the study of homing performance in natural environments. Interestingly – although biologically inspired – most homing algorithms have been tested only in simulations or on mobile robots in man-made environments. The recent work of Zeil et al. [7], who investigated catchment areas of outdoor scenes, showed that view-based homing can be quite reliable in natural environments which are usually very cluttered and – e.g. because of wind or movement of the sun – non-static. With respect to the proposed coarse-to-fine strategy, the depth structure of the scene is also of interest. As mentioned in the last section, segmenting image parts according to the distance of the imaged objects could improve homing performance further. For navigation in complex environments, the visual homing algorithm will be integrated in a navigation system that uses a graph-like internal representation of the environment also containing metrical information estimated from odometry or optical flow [34].

References

[1] B. Cartwright, T. Collett, Landmark learning in bees, Journal of Comparative Physiology A 151 (1983) 521–543.
[2] B. Cartwright, T. Collet, Landmark maps for honeybees, Biological Cybernetics 57 (1987) 85–93.
[3] R. Wehner, B. Michel, P. Antonsen, Visual navigation in insects: Coupling of egocentric and geocentric information, Journal of Experimental Biology 199 (1996) 129–140.
[4] O. Trullier, S. Wiener, A. Berthoz, J.-A. Meyer, Biologically based artificial navigation systems: Review and prospects, Progress in Neurobiology 51 (1997) 483–544.
[5] M. Franz, H. Mallot, Biomimetic robot navigation, Robotics and Autonomous Systems 30 (2000) 133–153.
[6] G. DeSouza, A. Kak, Vision for mobile robot navigation: A survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2) (2002) 237–267.
[7] J. Zeil, M. Hofmann, J. Chahl, Catchment areas of panoramic snapshots in outdoor scenes, Journal of the Optical Society of America A 20 (3) (2003) 450–469.
[8] A. Vardy, R. Möller, Biologically plausible visual homing methods based on optical flow techniques, Connection Science 17 (1–2) (2005) 47–89.
[9] D. Lambrinos, R. Möller, T. Labhart, R. Pfeifer, R. Wehner, A mobile robot employing insect strategies for navigation, Robotics and Autonomous Systems 30 (2000) 39–64.
[10] R. Möller, Insect visual homing strategies in a robot with analog processing, Biological Cybernetics 83 (2000) 231–243.
[11] K. Weber, S. Venkatesh, M. Srinivasan, Insect-inspired robotic homing, Adaptive Behavior 7 (1) (1999) 65–96.
[12] R. Cassinis, D. Duina, S. Inelli, A. Rizzi, Unsupervised matching of visual landmarks for robotic homing using fourier–mellin transform, Robotics and Autonomous Systems 40 (2002) 131–138.
[13] A. Vardy, F. Oppacher, Low-level visual homing, in: Seventh European Conference on Artificial Life, ECAL03, 2003.
[14] R. Wehner, The ant's celestial compass system: spectral and polarization channels, in: Orientation and Communication in Arthropods, Birkhäuser Verlag, Basel, 1997, pp. 145–185.
[15] S. Rossel, Navigation by bees using polarized skylight, Comparative Biochemistry and Physiology 104A (1993) 695–708.
[16] G. Horvath, D. Varju, Polarized Light in Animal Vision, Springer, Berlin, 2004.
[17] R. Wiltschko, W. Wiltschko, Magnetic Orientation in Animals, Springer Verlag, Berlin, New York, 1995.
[18] B. Kröse, R. Bunschoten, Probabilistic localization by appearance models and active vision, in: IEEE International Conference on Robotics and Automation, ICRA, 1999, pp. 2255–2260.
[19] L. Paletta, S. Frintrop, J. Hertzberg, Robust localization using context in omnidirectional imaging, in: IEEE International Conference on Robotics and Automation, ICRA, 2001, pp. 2072–2077.
[20] M. Jogan, A. Leonardis, Robust localization using an omnidirectional appearance-based subspace model of environment, Robotics and Autonomous Systems 45 (2003) 51–72.
[21] E. Menegatti, T. Maeda, H. Ishiguro, Image-based memory for robot navigation using properties of omnidirectional images, Robotics and Autonomous Systems 47 (2004) 251–267.
[22] T. Pajdla, V. Hlavac, Zero phase representation of panoramic images for image based localization, in: F. Solina, A. Leonardis (Eds.), CAIP 1999, in: LNCS, vol. 1689, Springer, 1999, pp. 550–557.
[23] M. Franz, B. Schölkopf, H. Mallot, H. Bülthoff, Where did I take that snapshot? Scene based homing by image matching, Biological Cybernetics 79 (1998) 191–202.
[24] J. Chahl, M. Srinivasan, Reflective surfaces for panoramic imaging, Applied Optics 36 (1997) 8275–8285.
[25] S. Nayar, Catadioptric omnidirectional cameras, in: Proc. of IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 1997, pp. 482–488.
[26] H. Ishiguro, Development of low-cost compact omnidirectional vision sensors, in: R. Benosman, S. Kang (Eds.), Panoramic Vision, Springer, 2001, pp. 23–38.
[27] W. Stürzl, H.-J. Dahmen, H. Mallot, The quality of catadioptric imaging – application to omnidirectional stereo, in: T. Pajdla, J. Matas (Eds.), ECCV 2004, in: LNCS, vol. 3021, Springer, 2004, pp. 614–627.
[28] B. Kröse, An effcient representation of the robot's environment, in: E. Pagello et al. (Eds.), Intelligent Autonomous Systems, vol. 6, 2000, pp. 589–595.
[29] E. Weisstein, "Parseval's theorem." from MathWorld — a Wolfram web resource, http://mathworld.wolfram.com/ParsevalsTheorem.html.
[30] W. Stürzl, H. Mallot, Fast phase-based orientation estimation for panoramic images, in: R. Würtz, M. Lappe (Eds.), Dynamic Perception, Aka GmbH, Berlin, 2002, pp. 289–294.
[31] A. van der Schaaf, H. van Hateren, Modelling the power spectra of natural images: statistics and information, Vision Research 36 (1996) 2759–2770.
[32] D. Ruderman, Origins of scaling in natural images, Vision Research 37 (1997) 3385–3398.
[33] K. Mardia, P. Jupp, Directional Statistics, Wiley, 2000.

[34] W. Hübner, H. Mallot, Integration of metric place relations in a landmark graph, in: J. Dorronsoro (Ed.), ICANN 2002, in: LNCS, vol. 2415, Springer, 2002, pp. 825–830.

Wolfgang Stürzl received his Diploma in Physics from the Technical University in Munich, Germany in 1999. He worked as a research assistant in the Cognitive and Computational Psychophysics department at the Max-Planck-Institute for Biological Cybernetics and in the Cognitive Neuroscience department at the Eberhard-Karls-University in Tübingen, Germany, where he finished his Ph.D. in 2004. He is currently a postdoctoral fellow at the Research School of Biological Sciences, Australian National University in Canberra, Australia. His research interests include autonomous navigation of mobile robots, modelling of animal behaviour, computer vision, and catadioptric sensor design.

Hanspeter Mallot received his Ph.D. at the Department of Zoology, University of Mainz, Germany in 1986. In the following years, he held postdoctoral and research positions at the Massachusetts Institute of Technology, the Ruhr-University Bochum, the Max-Planck-Institute for Biological Cybernetics in Tübingen and the Institute for Advanced Study, Berlin. In 2000, he was appointed Professor of Cognitive Neuroscience at the Eberhard-Karls-University, Tübingen. Research focusses on spatial cognition in rats, humans and robots, using behavioral experiments in virtual reality, eye-movement recordings, and simulated agents in hardware and software.

Hanspeter Mallot is editorial board member of the journals “Biological Cybernetics” and “Neural Networks”. He is also a past president of the European Neural Network Society (ENNS) and the German Society of Cognitive Science (GK).