The 'catchment areas' of panoramic snapshots in outdoor scenes

Jochen Zeil, Martin I. Hofmann, Javaan S. Chahl

Centre for Visual Sciences
Research School of Biological Sciences
Australian National University
P.O. Box 475
Canberra ACT 2601
Australia

23-10-2002

Running Title: Catchment areas of panoramic snapshots

Keywords: View-based homing, Panoramic snapshots, Landmark guidance, Insects

Corresponding author:
Jochen Zeil
phone: 0061 2 6125 5066
fax: 0061 2 6125 3808
email: zeil@rsbs.anu.edu.au

Summary

We ask the fundamental question of how well a position in natural space is defined by the scene viewed from that position. We took panoramic snapshots in outdoor scenes at regular intervals in two- or three-dimensional grids covering 1 m^2 or 1 m^3. We subsequently determined how the root mean squared (r.m.s.) pixel differences between each of the images and a reference image acquired at one of the locations in the grid develop over distance from the reference position. We then asked, whether the reference position can be pinpointed from a random starting position, by moving the panoramic imaging device in such a way that the image differences relative to the reference image are minimized.

We find that on time-scales of minutes to hours, outdoor locations in space are accurately defined by a clear, sharp minimum in a smooth three-dimensional volume of image differences (the 3D-difference function). 3D-difference functions depend on the spatial frequency content of natural scenes and on the spatial layout of objects therein. They become steeper in the vicinity of dominant objects. Their shape and smoothness, however, are affected by changes in illumination and shadows. The difference functions generated by rotation are similar in shape to those generated by translation, but their plateau values are higher. Rotational difference functions change little with distance from the reference location. Simple gradient descent methods are surprisingly successful in recovering a goal location, even if faced with large, transient changes in illumination.

Our results show that view-based homing with panoramic images is in principle feasible in natural environments, and does not require the identification of individual landmarks. We discuss the relevance of our findings to the study of robot and insect homing.

Introduction

The wealth of evidence showing that insects use view-based homing mechanisms has been extensively reviewed over the last few years[1-3]. Based on this evidence, a number of models have been proposed to explain how the comparison of a scene as viewed from a goal location with scenes viewed from vantage points some distance away can be used to derive instructions on how to return to the goal area (for a recent review see Ref. 4). Initially these models were tested in computer simulations, but more and more attempts are being made to implement them on mobile robotic platforms, with the aim of testing their performance in more realistic environments. However, these environments are still largely restricted to indoor settings and the question arises, how do view-based, insect-inspired navigation schemes perform under biologically relevant conditions (for a rare example of outdoor robotic navigation using view-based strategies see Ref. 5). To answer this question we need to know more about the natural operating conditions navigating animals are confronted with, and eventually to have insect-inspired robots operating in the field. We will only then be able to assess, whether the information currently fed to model navigation systems is available, sufficient and reliable for navigation in a visually noisy, cluttered, and time-varying natural environment.

Let us first briefly review the computational structure of the main models for view-based homing with special emphasis on the visual cues used in the models[4]. As far as input requirements are concerned, current models fall into two broad categories: models that require extraction of some features and/or the identification of objects (landmarks) in a scene, and models that neither require feature extraction nor the identification of landmarks.

In most cases in which panoramic images have been used for view-based robot navigation[6-11], image information has been greatly reduced, mainly because this generates a noise-free

pattern of signals in the indoor worlds of experimental robots where man-made structures and artificial lighting produce clearly identifiable, and mostly vertical edges.

Hong and colleagues[6], for instance, used a one-dimensional strip of a panoramic image, sampled along the horizon circle with 1^{o} resolution and averaged over 5^{o} degrees in elevation. Landmark features, or 'characteristic points' were then extracted, by segmenting the image strip into regions of monotonically increasing or decreasing intensity, by locating the intensity zero crossings in these segments, and by subsequently selecting the 15 most conspicuous points by ranking them according to the magnitude of intensity change within the segment. For the purpose of navigation, a matching procedure based on correlation was employed.

Franz and colleagues[9] also sampled a strip along the horizon with horizontal resolution of 4.6^{o} and some averaging in elevation. The resulting one-dimensional array was low-pass filtered, the background component was subtracted, and the contrast was finally maximised by histogram equalisation. The authors then compare two homing schemes which differ in the assumptions that are being made about the distance distribution of landmarks in a scene. One scheme is a variant of the navigation procedure suggested by Hong *et al*.[6] in which an average displacement vector is calculated from the differences in angular positions of local features in the current view compared to the reference view. In the second procedure, Franz *et al*.[9] construct a matched filter which predicts the displacement field of landmarks based on the assumption that objects are all at the same distance. Views are then distorted according to this simplified displacement field and are subsequently compared to the reference image using the dot product between the two images as a measure of match. From this an estimate of the home direction is derived and used to guide a robot platform. The latter procedure appears to improve performance slightly very close to the

goal, compared to other view-based homing schemes[10-13] which implicitly (and realistically) assume that the distances of objects are distributed independently of viewing direction.

Lambrinos and colleagues[10] and Möller[11] also reduced image information significantly when testing landmark guidance by their robots. They were able to threshold the image, because it basically contained only high-contrast artificial landmarks. They then extracted a horizontal panoramic strip from it, and finally arrived at a 1 pixel wide segmented horizon which was used as a template for image-matching, or for generating an average landmark vector, which is more parsimonious computationally but performs just as well as the image-matching procedures proposed previously.

In contrast to the models described above, Lehrer and Bianco[14] and Gaussier *et al*.[15] extracted selected image regions and used correlation techniques to find best matches for local features or 'landmarks'. The correlation coefficients then serve to determine which image regions offer the most reliable cues for the landmark guidance of their robots.

It is important to note that all the models we briefly summarized above where either tested in computer simulations (e.g. Refs. 12-13), with robots in laboratory environments without deep depth structure (e.g. Refs. 6,9,11,14,15; for a selective review see Ref. 4), or with robots operating in featureless environments, like a salt-pan desert in which artificial, high-contrast, and sparse landmarks were conveniently placed[10]. There are at least two reasons why it is difficult at present to assess how these homing mechanisms, which have 'evolved' in artificial worlds would perform under the normal natural conditions in which animals typically operate. First of all, most natural scenes are much more cluttered in texture, colour, luminance contrast, and depth compared to indoor scenes. These properties may either be advantageous for view-based homing mechanisms because visual information content is higher, or conversely may add an amount of

visual noise, which the current homing schemes may not be able to tolerate. It may be difficult to extract useful high-contrast features from natural scenes by applying a threshold and their deep depth structure introduces large discontinuities in parallax and generates occlusions, which are likely to cause problems for image matching schemes that rely on information derived from low-level image processing. Natural scenes are in addition subject to large temporal variations of luminance, of the direction of illumination, of shadow contours, and of background motion, which together will have the effect of dramatically changing the appearance of a scene, especially over time-scales of hours.

We attempt here to approach this problem, not by testing specific models under more natural conditions, but by asking a fundamental question which is relevant to all the models proposed so far: how well is a location in space defined by the surrounding scene as viewed from that location? Or to state the question more accurately, how different is the visual world when viewed from neighbouring vantage points, is this difference correlated with, and does it vary smoothly with physical distance? Our approach is inspired by the image interpolation technique developed by Srinivasan[7] and Chahl and Srinivasan[8], who have shown that over small distances and for small angles of rotation, the position and orientation of an image can be determined by interpolation from reference images taken at two different locations. The technique has been successfully used to measure egomotion, range and surface orientation[7,8,16-19] in simulated and in indoor environments. We apply a simple variant of the technique here to analyse the principal constraints of view-based homing in outdoor scenes by studying how global image differences develop with distance from a reference position. The resulting spatial difference functions can be understood as defining the 'catchment area' (*sensu* Cartwright and Collett[13]) of view-based homing schemes that operate on global image differences.

Materials and methods

We used a custom-designed 3D-positioning platform which is mounted on a trolley (robotic gantry, Fig. 1a) to move a panoramic imaging device[20] (Fig. 1b) outdoors into accurately defined positions in space. The robotic gantry consists of three perpendicular axes (horizontal x- and y-axis and a vertical z-axis) the movements of which are controlled by servo motors (components from Isel, Germany). The gantry can service an area of one cubic metre with a positioning accuracy of 0.01 cm. We did not implement rotational degrees of freedom for the panoramic imaging device, since yaw-axis rotations, which are most relevant for the questions we ask here, can more conveniently be simulated by software rotation of the panoramic images. During experiments, the gantry was levelled by adjusting heavy-duty set screws on the frame of the trolley. We took care to choose sites for our analysis in such a way that the panoramic images contained as few artificial structures as possible. A small cardboard screen prevented direct sunlight falling on the reflective cone. Panoramic images were recorded with a black and white or a colour CCD camera (Samsung BW-410CA, JVC TK860E) the gain control of which was switched off. They were digitised to 768x576 pixels at 8 bit resolution by a frame grabber and stored directly on the hard disk of a computer for analysis. We used panoramic imaging surfaces with a vertical field of view of 120° to 150°, which were seen by about 260000 pixels in a circular area at the centre of the rectangular video image. The panoramic images therefore were sampled with an average resolution of 0.2˚, decreasing with elevation from below the horizon to the zenith. The root mean squared (r.m.s.) differences in pixel values were determined off-line between images after a mask had been applied to remove the image regions outside the reflecting cone and in most cases also those containing dominant gantry and camera structures (Fig. 1c and

d). Fig. 1e shows an unfolded version of the image in Fig. 1c, with the image of the camera lens removed.

The r.m.s. pixel differences between images in any one scene and at any one time are determined by the specific scene composition, by the overall contrast and by the intensity of incident illumination. To derive an estimate of the maximum difference that can be expected for a given scene, apart from differences caused by changing the viewpoint of the imaging device or the orientation of a snapshot, we first selected an image, which in most cases was the reference image at the centre of a grid. By using the same distribution of pixel values in this image, we then created a second image, in which the location of these pixel values was randomised. We finally calculated the r.m.s. pixel difference between the original image and its randomised self to arrive at an estimate of the maximal image difference (max_{rms}) for that particular scene after all spatial correlation has been removed. The values of max_{rms} in our images ranged from 22 (Fig. 6c) to 102 (Fig. 9a) with a mean of 59.38 ± 15.57std.

Further details of the analysis and of the gradient descent experiments are explained in the text.

Results

The catchment areas of panoramic snapshots

We initially positioned the gantry in an area, which was enclosed by low bushes standing in front of a group of larger trees (Fig. 1e). We recorded snapshots in a cube of 0.7 m side length with grid points spaced at 10 cm intervals (white dots in Fig. 2a). Recording one horizontal grid plane took about two minutes. The process started from a position close to the left base of the gantry (Fig. 2b) and proceeded along transects parallel to its x-axis to end at the furthest reach of

the gantry's x- and y-axes on the right. The lowest grid plane was approximately 30 cm above ground, which was covered with irregular patches of grass. After applying a mask to the images (see Fig. 1d), we calculated the r.m.s. pixel differences between each of the images and a reference image recorded at the centre of the cube. This procedure provides us with a difference value for each location in the grid. The difference values for the lowest plane of the three-dimensional grid are shown in Fig. 2c as a function of their spatial position in the grid. We call this distribution of difference values the two-dimensional *difference function* for this particular snapshot location. Fig. 2d shows horizontal transects through the difference function in Fig. 2c, along the x-direction (full grey dots), the y-direction (full black dots), and along the two diagonals (open grey and black dots) as indicated by arrows and symbols in Fig. 2b. Note that the image differences increase smoothly with distance from the reference position and that the difference function has not levelled out at the edges of the recording grid. Since the images were recorded sequentially, there is the possibility that some of these differences are caused by changes in illumination. We took care to find stable illumination conditions for our recordings, but as we will show later, r.m.s. pixel differences did increase slowly with changes in the direction of illumination. Such temporal effects would lead to consistently larger r.m.s. difference values along the y-transect (black dots, Fig. 2d), compared to the x-transect through difference functions (grey dots, Fig. 2d), because the images along the y-transect were taken at time intervals that were seven times larger than those along the x-transect (cf. Fig. 2b). Although we do see these systematic differences in this case and in many others we will discuss later, they are nearly always restricted to the half of the grid which is closest to the gantry (see Fig. 2d). This suggests that systematic temporal effects on r.m.s. pixel differences are negligible compared to spatial effects, at least for the time it takes to record a single two-dimensional grid.

The 2D difference functions for snapshots at different heights above ground have very similar shapes and dimensions (Fig. 3a), most probably because this particular location had a densely cluttered depth structure in all directions. To investigate the vertical extent of the 3D difference function, we calculated the differences between the reference image RI at x = 0 m, y = 0 m, z = 0.3 m and the images recorded in horizontal grid planes at 0.3 m, 0.4 m, 0.5 m, and 0.6 m height above ground (Fig. 3b). Clearly, image differences also vary smoothly with vertical distance from the reference location and the 2D difference functions, although loosing their distinct cusp shape, still possess a detectable minimum up to 20 cm above the reference location.

As a first result we thus note that there is a smooth three-dimensional volume of image differences with a clear minimum at the location where the reference image was recorded, at least over a time-scale of minutes. The shape of horizontal or vertical sections through the centre of this volume is cusped and the image differences have not reached their maximal values at the edges of the recording area. This suggests that the distance at which the slope of the difference function for this particular location could first be detected - that is the size of the catchment area as it was defined by Cartwright and Collett[12,13] – is larger than the distance between the reference position and the borders of the recording area (30-50 cm in this particular case).

To discover whether and where these difference functions level out, we recorded at a different, more open location images at 10 cm intervals in a horizontal plane of 1 m x 3 m side length by moving the gantry one meter sideways after each 1 m x 1 m grid was recorded. The resulting difference functions are shown in Fig. 4. For this more open place, the difference function levels out at distances beyond one meter from the location at which the reference image was taken. Note that the difference functions for two different reference locations in the same

scene are quite similar in shape (compare Fig. 4a and b). They are both smooth, they lack pronounced local minima and reach approximately the same values for a given distance from the reference location. We ask next, what determines the size and the shape of difference functions in outdoor scenes and how robust the gradients of image differences are against changes in illumination and the movements of wind driven vegetation.

The dependence of difference functions on location and spatial layout

We first wanted to investigate, how invariant these properties of difference functions are and how much they are influenced by the specific spatial layout of a scene. We repeated the measurements described in the previous section in several locations, which differed in their depth structure. We reasoned that difference functions should be steeper in a scene containing large, close and high-contrast objects, compared to an open space, since for a given displacement, motion parallax should generate larger differences in images containing close objects. The resulting two-dimensional difference functions are shown in Fig. 5 for reference images taken close to the ground in four locations: The scene in Fig. 5a contains gantry contours and the brick wall of a barbecue area, Fig. 5b was recorded within a small stand of trees, Fig. 5c at the edge of this stand of trees, and Fig. 5d in an open area approximately 10 m away from the trees. Note that gantry contours loom large in the images, especially close to the trolley (at y = -0.5) and that masking them as they appear in these positions mimics the effect of removing close objects from a scene.

As expected, the gradient of the difference function is steeper in confined spaces compared to an open area with objects such as trees more than 10 m away, indicating that for a given displacement, image differences are proportional to the distance of objects in a scene. The

strange shape of the difference function we recorded in open terrain is a case in point (Fig. 5d): the function is shallower and rises to a lower plateau value, compared to the one shown in Fig. 5a, but it actually appears to be a combination of a very steep narrow and a shallow broad profile. As we show below, this is a consequence of the fact that the depth structure of the scene is not uniform with very close and very distant contours dominating the image.

This fundamental constraint of view-based navigation can be documented in three ways. We first introduced a large conspicuous landmark in an area where a difference function had been recorded previously. The results of this experiment are shown in Fig. 6a and b. The difference function without the landmark (Fig. 6a) is indeed shallower than in the presence of a landmark (Fig. 6b), although the introduced object covers only a small part of the image. The landmark has a surprising 'range of influence' on the difference function. In its presence, image differences are still elevated at 1 m distance from the reference position (Fig. 6b) compared to the scene without landmark.

We secondly analysed the contributions of close and distant objects to image differences in the same scene, by masking different parts of the image before calculating the difference function. The results are shown in Fig. 6c and d for the scene we have already encountered in Fig. 5d. We masked those parts of the images, which either contain features above the horizon (Fig. 6c), or features on the ground below the horizon (Fig. 6d). Close features on the ground clearly generate a steeper, narrower difference function, which rises to a flat plateau, compared to more distant features above the horizon, which generate a wide, shallow function.

In a third experiment we moved the imaging device inside a 20 cm wide and 20 cm high tunnel, the walls of which were lined with a random dot pattern. The elements of the pattern were 1 cm squares. Tunnels of this kind have been extensively used to study honeybee

navigation, in particular visual odometry, whereby the random dot patterns serve the function to provide the bees with optic flow information, but no information on location along the tunnel (see Ref. 19 for a review). The difference functions in this spatially restricted environment are very narrow and appear to be determined by the spatial frequency of the pattern lining the walls of the tunnel (Fig. 7a), as can be demonstrated by blending out the distant contours outside the tunnel (Fig. 7b) or the close ones offered by the tunnel walls (Fig. 7c). The distant objects above the tunnel contribute small image differences for a given displacement and thus on their own generate a shallow difference function. Note that since we calculate r.m.s. pixel differences, distant and close objects determine the shape of the difference functions, not in a purely additive manner, but depending on the size and the contrast of the image region they occupy.

The shape of translational and rotational difference functions

As we have seen, the detailed shape of difference functions is influenced by the spatial layout of a scene. Another way of showing how the depth structure of the world around a specific location influences the shape of the difference function is to compare the difference functions generated by a translational displacement of the recording device (translational difference functions) with those generated by a pure rotation around the yaw axis (rotational difference functions). The rationale for this procedure is as follows. The differences between images taken at two neighbouring locations in space are caused partly by parallax and occlusion, i.e. by the differential displacement of contours depending on their distance from the camera. Translational difference functions should therefore be determined by the spatial frequency content of the scene, the distance of objects, their contrast and their location in the 'visual field'. When images are rotated relative to the reference image, all objects will contribute to the r.m.s.

pixel differences depending on their angular size, but regardless of their distance from the imaging system and of their location in the 'visual field'. A distant mountain, for instance, will not change its position in images taken at locations one meter apart and thus will contribute nothing or only very small values to the overall image difference. An angular displacement of 10 degrees, however, will shift its image position by the same amount as those of any other object in the scene. Rotational difference functions should therefore be determined only by the spatial frequency content and the contrast of a scene. For a given scene, rotational difference functions should be deeper, with higher plateau values than translational difference functions. Although in a strict sense, rotational and translational difference functions cannot be directly compared, one of their properties, the ratio between the plateau values they can reach to their minimum, or their depth, is functionally significant in the context of view-based homing.

We explored the properties of rotational difference functions in several ways. We first compared the rotational functions at different locations of a three-dimensional grid with the image taken at each of these locations serving as reference (Fig. 8a and b). Note that to generate these rotational difference functions we applied rotationally symmetrical circular masks to only remove the image regions outside the reflecting cone. Rotational difference functions are indeed much deeper than the translational ones we have seen before for the same scene (see Fig. 5c). Their shape is very similar for each of the 10 locations shown, their depth, however, increases when comparing locations close to the ground (Fig. 8b) with those at the top of the grid, approximately 1 m off the ground (Fig. 8a).

We next used the image at the centre of the bottom plane as a reference image and calculated the difference functions it produced when subtracted from the rotated images at nine different locations around the cube (Fig. 8c and d). The rotational difference functions at these

locations are now shallower and the minimum is different from zero, corresponding to the value of the translational difference function at the respective location. The finding that the difference functions in so widely separated locations contain robust information on the orientation of a snapshot, was unexpected. The distances of corner locations on the bottom plane from the centre reference location were 0.7 m, and the corner locations on the top plane were 1.22 m away from the centre reference location. An animal sensitive to image differences could minimize these differences first by yaw rotations at any distance from the goal and thus align itself with the compass bearing it had during the acquisition of a snapshot. It could then pinpoint the goal by finding the minimum of the translational difference function using translational movements only.

Rotational difference functions are invariant against the depth structure of the environment they are determined in. Fig. 9 shows the shape of the rotational difference function in the sparse scene of a tropical mudflat. Although the function is slightly steeper than the one we recorded in a cluttered terrestrial scene (cf. Fig. 8), its shape and plateau values are quite similar and apparently not significantly influenced by the part of the scene contributing to it. However, the rotational difference function of the image of the celestial hemisphere (Fig. 9c) levels out at smaller angles of rotation to lower plateau values, compared to the difference functions for the whole image (Fig. 9a), or that for the part imaging the ground (Fig. 9b), suggesting that the spatial frequency distribution affects the shape of these functions.

The dependence of difference functions on illumination conditions and environmental motion

It is intuitively clear that the simple measure of r.m.s. pixel differences between images is subject to large variations due to changes in illumination, caused at different time-scales by the

movement of the sun and of clouds, and by the movement of wind-driven vegetation and its shadows. We call these movements environmental motion, because they are likely to generate responses in biological motion detectors which are unrelated to an animal's own movements and those of other creatures[21]. We attempted to analyse the effects of changes in the direction of illumination and the concomitant changes in the image positions of shadow contours and of specular reflections, first by recording a two-dimensional grid of images at the same location at about hourly intervals throughout the day on a cloudless and on an overcast day. The recording site was located at the edge of a small forest and was subject to large variations in shadow contours, but also in environmental motion generated by wind driven vegetation from branches overhanging the area.

These long-term recordings illustrate that difference functions can be both extremely volatile, but also quite stable over time (Fig. 10). The functions shown in Fig. 10 were calculated relative to the reference image recorded at the beginning of the experiment at the centre of the grid (at 13:00 on the first, clear day and at 13:44 on the second, overcast day). Although the shape of the difference function at this location can break down completely at certain times of day, in the sense that the function ceases to be cusped and to have a minimum at the reference position, it can also recover (left column Fig. 10; compare the functions recorded at 14:04, 15:02 and 16:02). Depending on illumination conditions, difference functions thus can retain their basic properties over quite long periods of time. On a partially overcast day, for instance, the difference function at the same location experiences a 'DC-shift', over a period of some hours, it also becomes more corrugated and shallower between 14:00 and 16:00, but generally maintains its overall shape (Fig. 10 right column). Most of these 'instabilities' arise from rapid changes in illumination caused by cloud movements covering and uncovering the sun, and by wind-driven

vegetation having a direct effect on the scene and an indirect one through changes in shadow contours. How stable a difference function can be over time, is shown in Fig. 11 for an open location on a calm, clear, cloudless day. Fig. 11 also allows a comparison to be made between the shape of difference functions as they are determined at different times of day with the reference image recorded at the same time (right panels in Fig. 11) or at the beginning of the experiment (left panels in Fig. 11).

To investigate the long-term and short-term properties of difference functions, we recorded images at the same location in space intermittently over several hours. We then determined the image differences for this location at 10-second intervals in 10-minute blocks with the image recorded at the beginning serving as reference. As can be seen from a ten-minute record in Fig. 12a, temporal variation of illumination due to moving clouds causes large and rapid variations in image differences on the time-scale of minutes. These rapid variations are accompanied by a slow increase in image differences over time, suggesting that changes in the direction of illumination caused by the movement of the sun, also contribute to the temporal variation of image differences. At this sampling rate, temporal variations are very similar above and below the horizon (see insets Fig. 12a). These different time-courses of variations in image differences can be more clearly seen in recordings of image series, which we interspersed with the recordings of difference functions (shown in Fig. 10) on a clear and on a partially overcast day (Fig. 12b and c). The 'diurnal' component was particularly pronounced on the cloudless day (Fig. 12b), with image differences increasing from 13:00 to about 15:00 for a reference image recorded at 13:14, and then decreasing again later in the afternoon. We recorded a similar pattern of slow 'diurnal' changes and rapid fluctuations of image differences, that is of scene similarity, on the second, overcast day (Fig. 12c), but the slow component was much weaker, presumably

because the scene was less affected by the changes in direction of illumination due to the movement of the sun. Shadow contours are either absent or reduced in contrast when the day is overcast. It is more difficult to identify the causes of rapid fluctuations in scene similarity. Cloud movements have clearly the largest effect (see Fig. 12a), but how much wind-driven vegetation and the shadow movements it produces contribute to image differences can only be quantified at higher temporal resolution of scene variation in different parts of the image.

Homing by gradient descent

Are the difference functions we measured in outdoor scenes useful for navigation? Their cusped shape and their smoothness suggest that an animal, which is sensitive to the differences of views relative to a remembered one could in principle pinpoint a reference location by moving in such a way that image differences are minimized.

To investigate whether image differences are a useful means of guidance, we implemented two simple gradient descent algorithms on our robotic gantry. Both algorithms move the imaging device to the centre of the active space of the robotic gantry and acquire a reference image. The imaging device is then moved in an arbitrary direction and distance away from that position, and its movements are subsequently controlled by the image differences relative to the reference image. For technical reasons, we used images with constant orientation throughout in these experiments and calculated the mean squared (m.s.) pixel differences over whole, unmasked images, but excluded image regions outside the reflective surface with a physical mask made out of black cardboard. In one of the gradient descent methods, called 'RunDown', which is a form of the Gauss-Seidel strategy, the gantry is instructed to start moving in one direction until image differences increase. As soon as this happens, movement direction is changed by 90^{o} and the new

course is followed until image differences increase again. In the second algorithm, 'Triangular', a version of Evolutionary Operation"[22] the imaging device is moved to three positions at the corners of a triangle to acquire images (for two-dimensional gradient descents) or to four positions at the vertices of a small tetrahedron (for three-dimensional gradient descents). Images taken at the endpoints of these search positions are subtracted from the reference image, thus generating three or four m.s. difference values that are then compared. The result is used to generate a vector pointing in the direction of the maximal gradient. The imaging device is then moved along this vector to a new location where the sampling is repeated.

In many situations, both procedures were quite successful in returning the imaging device to the location at which the reference image was recorded, by moving down the gradient of the difference function to reach its minimum. Fig. 13 shows the results of an experiment for which we admittedly chose ideal conditions: the experiment was run on a cloudless, clear day at an open site, well away from tall vegetation. The 2D difference function for the site is shown in Fig. 13a. Over a period of 2 hours we conducted 40 homing runs, half of them with 'RunDown' and half with 'Triangular', in a randomised sequence, always starting with a reference image at the centre of a plane which had an area of 1 m^2 and was approximately 20 cm above ground. We stopped a run either after 3 minutes or after the imaging device had reached to within 5 cm of the reference location in the centre. Both algorithms successfully returned the imaging device to the reference locations in 18 out of 20 runs, as is documented by the paths of the imaging device in Fig. 13b and c, and by the plots of m.s. pixel differences over the distance from the reference location in Fig. 13d and e. It is interesting to note, that it would in principle also be possible to terminate homing runs by using a m.s. difference cut-off directly. On this occasion, for instance,

terminating the runs when image differences had reached a value of 100, would in most cases have brought the imaging device to within 10 cm of the goal.

In the course of this and other experiments, which took place at other sites, we made the remarkable observation that these primitive gradient descent schemes can be quite robust and successful even when faced with changes in illumination. The homing runs in Fig. 14 show several examples of this achievement for both algorithms. In hindsight it becomes clear why this is so: when illumination changes, scene similarity as measured by m.s. pixel difference relative to a reference image taken at a different state of illumination decreases dramatically (see for instance black curve in Fig. 14b and d). However, in this situation, a gradient descent algorithm, which compares image differences between successive time steps, as the ones we employed do, will not be able to escape the location it has already reached, because wherever it moves, image differences will not become systematically smaller (e.g. Fig. 14a and c). Transient changes in illumination, therefore, will only slow down the progress of gradient descent, but not destabilise it. This can be seen in two of the examples in Fig. 14c and d, where the imaging device had not reached the goal location when the change in illumination occurred. The algorithm 'hunts' around the location where the change in illumination occurred, but does not break out. It continues its descent, when illumination has returned to the situation prevalent at the time the reference image was recorded. However, gradient descent on raw pixel differences will most certainly fail, whenever the reference image is acquired at a rare state of general illumination.

In spite of the obvious limitations of simple gradient descent strategies when faced with variable illumination, this result tells us, that in principle, panoramic image differences can be used by an agent, which is sensitive to them, to relocate a goal position in complex and visually cluttered outdoor scenes.

Discussion

We have documented that the simple measure of r.m.s. pixel differences between panoramic images can be a reliable cue to location in outdoor scenes. This is so, because image differences change in a regular and smooth fashion with distance from a reference position, meaning that they are correlated with position in space. We have also shown that image differences can be used to recover a reference *orientation*, even if the observer is some distance away from the location at which the reference image was acquired. We finally demonstrated that the global and simple measure of panoramic image difference does supply enough information to re-locate a position in space by means of simple gradient descent algorithms. Using panoramic snapshots for homing thus does seem to be an option for animals under the complex natural conditions they are operating in.

However, at any one location, natural scenes also change with time because the direction of illumination changes slowly with the movement of the sun and because environmental motion generated by wind driven vegetation, the movement of shadows and the movement of clouds, alter their appearance on a short time-scale. These sources of temporal variation can override the spatial correlation of image differences, degrading the cusped shape of difference functions to such an extent, that they become useless for view-based homing.

We discuss our results in three steps: We ask what determines the size and shape of difference functions, what strategies animals or robots could employ to cope with the temporal variations of natural scenes, and what relevance our results have for the study of view-based homing in animals and robots.

What determines the shape, size and depth of difference functions?

Our comparison between difference functions in scenes with different depth structure suggests that the shape and depth of these functions depend on the spatial frequency content of images, on the degree of occlusion and on the depth structure of the scenes they are recorded in (see also Refs. 7,8, 16-19). Our evidence for this conjecture is at this stage still qualitative, but the following considerations do give it some weight. Let us assume that objects at different distances contribute about equal amount of *spatial frequencies* and *angular sizes* to the image taken at a reference position. As the imaging device moves away from the reference position, those image regions, which view nearby objects, will generate the largest image differences. Their contribution to the r.m.s. pixel difference calculated over the whole panoramic view will saturate, however, at some value proportional to the relative size of the image region they occupy in the reference image, when each pixel has been replaced by a background pixel due to motion parallax or through occlusion (see Fig. 6d and Fig. 7). The r.m.s. pixel differences are thus a complex function of the distance distribution of objects in the world and of their angular size in the image. For a given displacement, objects generate fewer and fewer image differences the further away they are. The contributions to the difference image of objects in each depth plane (normalised to the displacement distance) will level out at some maximal difference value until a plateau is reached when most of the pixels differ from the equivalent ones in the reference image. During pure translation, the plateau value of the difference function is determined by the relative contribution of image regions occupied by distant objects, which stay the same in all images, and those that are occupied by nearby objects. The comparison between translational and rotational difference functions is interesting in this respect: since image differences generated by rotation are independent of the depth structure of a scene, they plateau at much higher values

compared to image differences generated by translation. The reason being that in rotational difference functions, all image regions contribute equally to the r.m.s. pixel difference, independent of the distance of objects in the scene. The comparison between rotational and translational difference functions also tells us, that their smoothness is most likely caused by the 'scene density' outdoors, in terms of the spatial frequency distribution, the broad distribution of contrast, the wide distribution of object distances and of angular sizes, and by the absence of sharp vertical contours. Compared with indoor scenes, occluding surfaces which would cause abrupt changes in translational, but not in rotational difference functions, do not appear to play a major role in outdoor scenes.

The dependence of image differences on the depth structure of a scene, allows us to predict that the pixel differences contributing to a given difference function are not distributed equally across the 'visual field'. Much like the image velocity vectors in the optic flow field experienced by a moving optical system, the image differences generated by a displacement, depend on the direction of translation, with largest differences occurring in directions of view perpendicular to the heading direction[8,9]. We would also predict that, if the depth, shape and smoothness of difference functions depend on the spatial frequency content of a scene, low-pass filtering the images should make the functions shallower and smoother. The retinal topography of panoramic image differences and the effect of resolution on difference functions, especially non-uniform resolution across the visual field, are thus interesting topics for further investigation.

The temporal stability of difference functions

We have shown that difference functions can be quite stable over time, but that changes in the direction of illumination and environmental motion often seriously degrade their shape. How

could animals relying on view-based homing cope with these scene variations? Visual systems are known to employ a variety of strategies to alleviate the effects of changes in illumination ranging from receptor adaptation, contrast normalisation, to motion and colour processing. Compared to our description of difference functions on the level of raw pixel values, any additional processing will make difference functions more immune to changes in illumination and as a consequence, will make view-based navigation more robust. The contribution further processing can make to scene stability and view-based homing can now be systematically explored. It would be quite simple, for instance, to compare difference functions and gradient descent performance with the automatic gain control of the camera switched on or off. Experiments could also be carried out off-line, either by normalising images to their average brightness or by normalising contrast over the image, before calculating r.m.s. pixel differences.

Since gradient descent based on global image differences is fairly unaffected by changes in illumination, some relatively simple pre-processing of images may be all that is required to make view-based homing immune to the temporal variations in natural scenes. It is more difficult to see how insects or robots might cope with the more rapid temporal variations, which are generated by wind-driven vegetation and the movement of shadows. One way to reduce the contribution of shadows to image differences may be to remove shadows by colour processing the images before comparing them. The visual effects of wind-induced motion could be reduced, by spatial and temporal low-pass filtering of images. We are currently developing ways to separate the effects of environmental motion from those generated by changes in illumination, which will enable us to investigate these strategies in detail.

View-based homing

In surveying the machine vision and robotics literature, it was surprising for us to notice that image differences, as a simple cue for position in space, have to date apparently not been found useful for view-based homing, although they have been shown to provide robust information on egomotion[7,8,16,19] and range[17,18]. We can see three reasons for this neglect: first, in conventional robotics applications, image differences have probably been discarded as potential cues for long range navigation, because the indoor environments in which robots are normally developed and tested are characterised by repetitive spatial arrangements of very self-similar structures and sharp edges. The hallways of robotics laboratories come to mind, in which the only dominant landmarks are doors, doorways, desks, and chairs. Their similarities in 2D images taken at different locations must be a prime source of local minima, which are difficult to overcome. In addition, sharp, occluding edges may limit the range over which image differences increase monotonically. A last, and probably more crucial reason for rejecting image differences and gradient descent as potent navigational aids for pinpointing goals is the fact that they do not enable a navigating agent to compute its position relative to a goal in absolute coordinates and even in relative coordinates it cannot determine its position without moving and comparing[8]. This however is exactly what insects appear to do.

Flying insects do not just take a snapshot when they leave a location they wish to return to. Instead, they go through an elaborate sequence of behaviour, called an orientation or learning flight, in the course of which they turn towards the goal and move away from it backwards, in a series of increasing arcs, while pivoting about the goal location[23-25] (for reviews see Refs. 1,2,14,26,27). There are two likely reasons why insects perform learning flights: one is the need to segregate foreground from background contours and to identify close three-dimensional

objects by means of motion parallax. This would allow insects to filter out shadow contours, to parse images and to restrict the reference image and the subsequent matching process to motion defined contours[23]. Secondly, departing insects may need to measure how reliable the currently acquired visual representation is for the subsequent homing task in real-time during the learning process. Already during the learning phase, therefore, insects move and have ample opportunity to compare what they have already learnt to what they are currently experiencing. When they subsequently return to the goal, insects also do not fly in a straight path. Although their mean orientation correlates well with their orientation during learning[1,25,28], they approach the goal not directly, but in a series of sideways movements. In the presence of a distinct landmark close to the goal, they may approach this landmark first, thus using it as a beacon[29] and subsequently turn and move to approach the goal position in the orientation they had during the learning phase on departure.

Insects are also known to be able to extract features like the average orientation of contours in a pattern[30-32], apart from apparently memorising snapshot-like images[12,33,34]. In the context of pattern recognition, at least, we note a lively recent discussion about whether image matching and/or feature extraction best explain the feats of pattern recognition in flies and bees[30-32,35-39]. The question of interest in the present context of view-based navigation, however, is whether general features, such as the distribution of contour orientations[40], or the spatial frequency content of natural scenes[41,42] carry reliable and robust information, not on the *identity of a pattern* to be recognised, but on the *identity of a location* in space. We believe that such general features are unlikely to be useful in homing tasks. For instance, the second-order statistics of natural scenes are so similar across different viewpoints and different habitats[41], that they cannot serve as a signature for individual locations in the world.

In spite of the fact that the simple measure of panoramic image differences provides a robust cue to location in space, the behaviour of homing insects tells us that animals do also attend to individual objects on their own. Landmarks are used in at least three different ways, as beacons, as route landmarks, and as part of a scene specifying a location, whereby a small displacement of a landmark can have marked effects on the path an insect follows, or on the location where it searches for a goal (for review see Refs. 1, 2). Why do insects behave like this, although we have shown here that overall image differences would in principle be sufficient to guide them back to a goal position? We can think of a number of reasons why landmarks as distinct objects may offer additional and crucial information for navigation: landmarks may help in navigating open terrain where image differences are small and probably fairly constant over large distances (beacons, route landmarks). Landmarks are also important for accurate pinpointing, especially with relation to the goal direction from an object: for instance, except at very close range, it may be impossible to decide by determining panoramic image differences alone, on which side of a small landmark a nest entrance lies. Recognition of landmarks and their use as a reference may in addition help to alleviate the problems of shadows and varying illumination in outdoor scenes.

So far, the most parsimonious model for view-based homing is probably the average landmark vector scheme, proposed by Lambrinos *et al.*[10] and Möller[11]. The model is more parsimonious compared to conventional image matching schemes, because it requires only one vector to be stored, rather than an image or its derivatives. In this respect, the average landmark vector model resembles the simple measure of global image differences, which we investigated here. It remains to be shown, however, that the good performance of the average landmark vector scheme in the sparse and high-contrast landmark environments, in which it was tested, is predictive of its performance under real-life conditions. It is not clear to us, for instance, how to

choose appropriate parts of a complex natural scene for generating the component vectors needed in the scheme, both during acquisition and during homing. This task seems especially difficult in the three-dimensionally cluttered worlds through which animals navigate.

We used a gradient descent scheme based on the m.s. pixel differences between panoramic snapshots to test whether the difference gradients towards a goal location provide enough information for homing. We demonstrate that this is indeed the case, at least for snapshots with constant orientation, which raises the question how biologically plausible gradient descent methods are. The problem can be broken down into two levels. The first is behavioural, and the question is how animals may sample the gradient of image differences before deciding on their next move. One crucial issue that we did not explore in our gradient descent experiments, is the need for correcting for orientation errors before using image differences as a guide for homing[9,12]. As we mentioned before, the shape and depth of rotational and translational difference functions suggest, that flying insects could, in addition to using celestial or magnetic compass cues to control orientation, minimise image differences first by rotating and then by using sideway movements to navigate towards the minimum of the translational difference function. The pivoting and sideways flight characteristics of homing insects, both during learning and during the approach to a goal, may be reflecting such a strategy[26,27,43,44]. The second level involves neural processing and the question here is how a nervous system may determine and store image differences on a pixel-by-pixel basis[32]. It is not clear, for instance, whether the overall pixel difference between images is a computation more easily performed by the brain of an insect than feature difference or object recognition. We know too little about the constraints operating on insect neural networks involved in processing and storing visual information, to be able to assess whether a certain operation is 'simple' or not. Parsing the retinal image through

edge detection, motion segmentation or colour processing, is likely to aid processing, retention and recall of visual information. However, it remains to be seen, whether reduction of the information potentially available in the retinal image improves homing performance and results in a representation that is easier to store.

Acknowledgements

The work was initially supported by the Australian Defence Science and Technology Organization (DSTO) and the United States Air Force (Eglin AFB). It was subsequently continued with the help of funds from the Human Frontiers Science Program (HFSP 84/97) and from an IAS Planning and Performance grant to the Research School of Biological Sciences, Australian National University. We thank Walter Junger for his help and advice in the design phase of the robotic gantry, and Katharina Siebke for her expert help as gantry operator and trouble-shooter. Cole Gilbert, Jan Hemmi, Daniel Osorio, and Mandyam Srinivasan read an earlier version of the manuscript and we are grateful for their constructive criticism, their discoveries of flaws and their suggestions for improvement. We finally acknowledge the suggestions for improvement made by two anonymous referees.

References

1. T.S. Collett and J. Zeil, "Selection and use of landmarks by insects," in *Orientation and Communication in Arthropods*, M. Lehrer, ed. (Birkhäuser Verlag, Basel, 1997), pp. 41-65.

2. T.S. Collett and J. Zeil, "Places and landmarks: an Arthropod perspective," in *Spatial representation in animals*, S. Healy, ed. (Oxford University Press, Oxford, 1998), pp. 18-53.

3. M. Giurfa and E.A. Capaldi, "Vectors, routes and maps: new discoveries about navigation in insects," Trends Neurosci. **22**, 237-242 (1999).

4. M.O. Franz and H.A. Mallot, "Biomimetic robot navigation," Robotics Autonom. Syst. **30**, 133-153 (2000).

5. E.M. Riseman, A.R. Hanson, J.R. Beveridge, R. Kumar, and H. Sawhney, "Landmark-based navigation and the acquisition of environmental models," in *Visual navigation*, Y. Aloimonos, ed. (Lawrence Erlbaum Ass. Publ., New Jersey, 1997), pp. 317-374.

6. J. Hong, X. Tan, B. Pinette, R. Weiss, and E.M. Riseman, "Image-based homing," IEEE Control Systems, Special Issue on Robotics and Automation, Vol **12**, 38-45 (1992). (or Proc. 1991 IEEE Int. Conf. on Robotics and Automation, pp. 620-625)

7. M.V. Srinivasan, "An image interpolation technique for the computation of optic flow and egomotion," Biol. Cybern. **71**, 401-415 (1994).

8. J.S. Chahl and M.V. Srinivasan, "Visual computation of egomotion using an image interpolation technique," Biol. Cybern. **74**, 405-411 (1996).

9. M.O. Franz, B. Schölkopf, H.A. Mallot, and H.H. Bülthoff, "Where did I take that snapshot? Scene-based homing by image matching," Biol. Cybern. **79**, 191-202 (1998).

10. D. Lambrinos, R. Möller, T. Labhart, R. Pfeifer, and R. Wehner, "A mobile robot employing insect strategies for navigation," Robotics Autonom. Syst. **30**, 39-64 (2000).

11. R. Möller, "Insect visual homing strategies in a robot with analog processing," Biol. Cybern. **83**, 231-243 (2000).

12. B.A. Cartwright and T.S. Collett, "Landmark learning in bees: experiments and models," J. Comp. Physiol. **151**, 521-543 (1983).

13. B.A. Cartwright and T.S. Collett, "Landmark maps for honeybees," Biol. Cybern. **57**, 85-93 (1987).

14. M. Lehrer and G. Bianco, "The turn-back-and-look behaviour: bee versus robot," Biol. Cybern. **83**, 211-229 (2000).

15. P. Gaussier, C. Joulain, J.P. Banquet, S. Leprêtre, and A. Revel, "The visual homing problem: An example of robotics/biology cross fertilization," Robotics Autonom. Syst. **30**, 155-180 (2000).

16. M.G. Nagle, M.V. Srinivasan, and D.L. Wilson, "Image interpolation technique for measurement of egomotion in 6 degrees of freedom," J. Opt. Soc. Am. A **14**, 3233-3241 (1997).

17. J.S. Chahl and M.V. Srinivasan, "Range estimation with a panoramic visual sensor", J. Opt. Soc. Am. A **14**, 2144-2151 (1997).

18. M.G. Nagle and M.V. Srinivasan, "Structure from motion: determining the range and orientation of surfaces by image interpolation," J. Opt. Soc. Am. A **13**, 25-34 (1996).

19. M.V. Srinivasan, J.S. Chahl, and S.W. Zhang, "Robot navigation by visual dead-reckoning: Inspiration from insects," Int. J. Pattern Recognition and Artificial Intelligence **11**, 35-47 (1997).

20. J.S. Chahl and M.V. Srinivasan, "Reflective surfaces for panoramic imaging," Appl. Opt. **36**, 8275-8285 (1997).

21. M.P. Eckert and J. Zeil, "Towards an ecology of motion vision," in *Motion Vision: Computational, neural and ecological constraints*, J.M. Zanker and J. Zeil, eds. (Springer Verlag, Berlin, 2001), pp. 333-369.

22. G.E.P Box and N.R. Draper, *Evolutionary Operation.* John Wiley & Sons, New York.

23. J. Zeil, "Orientation flights of solitary wasps (*Cerceris*; Sphecidae; Hymenoptera): I. Description of flight," J. Comp. Physiol. A **172**, 189-205 (1993).

24. M. Lehrer, "Why do bees turn back and look?," J. Comp. Physiol. A **172**, 549-563 (1993).

25. T.S. Collett and M. Lehrer, "Looking and learning: a spatial pattern in the orientation flight of the wasp *Vespula vulgaris*," Proc. R. Soc. Lond. B **252**, 129-134 (1993).

26. J. Zeil, A. Kelber, and R. Voss, "Structure and function of learning flights in bees and wasps," J. Exp. Biol. **199**, 245-252 (1996).

27. T.S. Collett and J. Zeil, "Flights of learning," Current Directions in Psychol. Sci. **5**, 149-155 (1996).

28. J. Zeil, "Orientation flights of solitary wasps (*Cerceris*; Sphecidae; Hymenoptera): II. Similarities between orientation and return flights and the use of motion parallax," J. Comp. Physiol. A **172**, 207-222 (1993).

29. T.S. Collett and J.A. Rees JA, "View-based navigation in Hymenoptera: multiple strategies of landmark guidance in the approach to a feeder," J. Comp. Physiol. A **181**, 47-58 (1997).

30. J.H. van Hateren, M.V. Srinivasan, and P.B. Wait, "Pattern recognition in bees: orientation discrimination," J. Comp. Physiol. A **167**, 649-654 (1990).

31. D. Efler and B. Ronacher, "Evidence against a retinotopic-template matching in honeybees' pattern recognition," Vision Res. **40**, 3391-3403 (2000).

32. M. Dill and M. Heisenberg, "Visual pattern memory without shape recognition," Phil. Trans. R. Soc. Lond. B **349**, 143-152 (1995).

33. T.S. Collett and M.F. Land, "Visual spatial memory in a hoverfly," J. Comp. Physiol. **100**, 59-84 (1975).

34. R. Wehner and F. Räber, "Visual spatial memory in desert ants, *Cataglyphis bicolor* (Hymenoptera: Formicidae)," Experientia **35**, 1569-1571 (1979).

35. M. Dill, R. Wolf, and M. Heisenberg, "Visual pattern recognition in *Drosophila* involves retinotopic matching," Nature **365**, 751-753 (1993).

36. M. Heisenberg, "Pattern recognition in insects," Current Opinion Neurobiol. **5**, 475-481 (1995).

37. B. Ronacher, and U. Duft, "An image-matching mechanism describes a generalization task in honeybees," J. Comp. Physiol. A **178**, 803-812 (1996).

38. B. Ronacher, "How do bees learn and recognize visual patterns?," Biol. Cybern. **79**, 477-485 (1998).

39. R. Ernst and M. Heisenberg, "The memory template in *Drosophila* pattern vision at the flight simulator," Vision Res. **39**, 3920-3933 (1999).

40. D.M. Coppola, H.R. Purves, A.N. McCoy, and D. Purves, "The distribution of oriented contours in the real world," Proc. Natl. Acad. Sci. **95**, 4002-4006 (1998).

41. A. van der Schaaf and H. van Hateren, "Modelling the power spectra of natural images: statistics and information," Vision Res. **36**, 2759-2770 (1996).

42. D. Ruderman, "Origins of scaling in natural images," Vision Res. **23**, 3385-3398 (1997).

43. R. Voss and J. Zeil, "Active vision in insects: An analysis of object-directed zig-zag flights in a ground-nesting wasp (*Odynerus spinipes*, Eumenidae)," J. Comp. Physiol. A **182**, 377-387 (1998).

44. K. Dale and T.S. Collett, "Using artificial evolution and selection to model insect navigation," Current Biology **11**, 1305-1316 (2001).

Figure legends

Fig. 1 (a) The robotic gantry in its natural habitat. The panoramic imaging device, consisting of a video camera and a reflective surface can be seen at the end of the horizontal y-axis arm at the far right of the picture. (b) Close-up of the panoramic imaging surface and the camera lens. (c) Panoramic image after a circular mask was applied to the original video image. (d) Panoramic image after applying an additional mask blocking the main gantry and the image of the camera and the camera lens. (e) An unwarped version of the panoramic image shown in c, after removing the image regions containing the camera lens.

Fig. 2 The difference function in a densely vegetated area. (a) The position of a three-dimensional grid of image positions and the orientation of gantry axes in the scene. (b) The two-dimensional grid of 7x7 spatial positions at which panoramic images were taken for each horizontal plane of the three-dimensional grid shown in a. The recording sequence starts at x = -0.3m, y = -0.3m and ends at x = 0.3m, y = 0.3m. Coordinates are given relative to the reference location at x = 0, y = 0. Transects are labelled with different symbols (see d). (c) The 2D difference function for the lowest plane of the three-dimensional grid. The r.m.s. pixel differences are shown along the z-axis for each image position in the 7x7 grid, as compared with the image taken at the reference position in the centre. (d) Transects along the x- and y-direction (solid dots) and along the two diagonals (open dots) through the 2D difference function shown in c. Directions of transects and their symbols are indicated in b.

Fig. 3 The vertical spatial extent of difference functions and their dependence on reference image location. (a) Transects through the 2D difference functions at z = 0.4m, z = 0.5m and z =

0.6m above ground for reference images RI at the centre of the same planes of the 7x7x7 three-dimensional grid shown in Fig. 2a. See inset for definition of symbols. Otherwise conventions as in Fig. 2d. (b) The difference functions for the same planes (at z = 0.4m, z = 0.5m and z = 0.6m), calculated relative to the reference image at the centre of the bottom plane (at x = 0m, y = 0m, z = 0.3m; see diagram at bottom centre).

Fig. 4 The horizontal extent of difference functions. The difference functions for two locations in an open area at the edge of a stand of tall eucalyptus trees. The difference functions were determined over an area of 1m x 3m for two different reference locations (top and centre) by moving the gantry one metre at a time along the x-direction. Images were taken approximately 20 cm above ground and the grid spacing was 10 cm. Transects along the x-axis at y = 0 are shown in the bottom graph. Otherwise conventions as before.

Fig. 5 Transects through the 2D difference functions in four outdoor scenes with different depth structure. Conventions as before. Images were recorded at 10 cm intervals in a 11x11 grid approximately 20 cm above ground (a) in a location close to the brick wall of a barbecue area, (b) within a small stand of trees, (c) at the edge of the small stand of trees and (d) approximately 10 m away from the trees in an open area. The oval shape at the right masks one of the trolley wheels which looms large in the image at position x = 0.5, y = -0.5, z = 0.

Fig. 6 The influence of depth structure on the shape and extent of difference functions. The difference functions for panoramic images before (a) and after (b) a cylindrical landmark had been placed a few centimetres beyond position x = -0.5m, y = 0m (marked by dot in lower

diagrams). The masked panoramic images of the scene are shown on top and the respective 2D difference functions with the reference image at x = 0m, y = 0m below. Otherwise conventions as before. (c) and (d) Difference functions for the location shown in Fig. 5d, but with different image regions masked. (c) Image region viewing exclusively objects above the horizon; (d) image region viewing the ground. Transects through the 2D difference functions are shown below. Conventions as before (see inset).

Fig. 7 The influence of depth structure on the extent and depth of difference functions. The graph shows the image differences along a 1 m stretch of narrow tunnel, the walls of which were lined with a random dot pattern with 1 cm element size. The tunnel was 20 cm wide and 20 cm high. To determine the contributions of different image regions, differences are shown for the full scene (a, black line), for the part of the images viewing the tunnel only (b, dark grey line) and for the part of the images viewing the scene beyond the tunnel (c, light grey line). The reference image was recorded at a position 50 cm along the tunnel. Otherwise conventions as before.

Fig. 8 The properties of rotational difference functions. Images were recorded every 10 cm in a 11x11x11 grid at the edge of a small stand of trees. Image differences were calculated for 5 locations each on the bottom and the top plane of the three-dimensional grid (see inset in the centre) for different degrees of rotation of the same images ((a) and (b)) or between a reference image at the centre of the bottom plane and rotated images at all other locations ((c) and (d)). Note the difference in scale compared to the translational difference functions shown in previous

figures. Images were rotated in 9 degree steps after a circular mask had been applied to remove image regions outside the reflective cone.

Fig. 9 Rotational difference functions in a flat world. A panoramic image taken low to the ground in a tropical mudflat was used to analyse the dependence of rotational difference functions on the spatial structure of a scene. The three curves were calculated after masking different parts of the image with circular masks (see insets). Images were rotated in steps of 9 degrees.

Fig. 10 The temporal stability of difference functions. Images were recorded every 10 cm in a 11x11 grid approximately 20 cm above ground repeatedly over 3 hours on two consecutive days in the same location at the edge of a small stand of trees. Recording the 121 images in one grid plane took about 5 minutes. Transects in the left column are through 2D difference functions recorded on a clear and windy day in the same location, the first one at 13:00 hrs, the last one at 16:02 hrs. The reference image used throughout was the one recorded at the centre of the grid at 13:00 hrs. The aperture setting of the camera lens had to be adjusted to prevent camera saturation and is shown in each panel together with the time of recording. The right column shows transects through 2D difference functions recorded at the same location on the following day, which was predominantly overcast and windy. The reference image used to calculate the difference functions was recorded at 13:44 hrs at the centre of the grid. The recording area lies at the north-west edge of the small forest where the shadows from overhanging branches can change the scene significantly depending on the wind and the movements of clouds and the sun (see Fig. 12 below). Conventions as before (see inset).

Fig. 11 The temporal stability of difference functions: calm, clear, cloudless day in an open area. Images were recorded with 10 cm spacing in a 11x11 grid approximately 20 cm above ground on a calm, mostly cloudless day at an open site over 10 m away from the edge of a small forest. The transects through the 2D difference functions on the left were calculated with images recorded at different times of the day, using the reference image recorded at the centre of the grid at 13:10 hrs. Functions on the right were calculated with the reference image recorded at the same time of the day. All other conventions as before (see inset).

Fig. 12 Short and long-term changes of image differences in outdoor scenes. Images were recorded at the same location with a sampling rate of 6 per minute. Image differences were calculated with the image at t = 0 as a reference. (a) The trace shows the r.m.s. image differences over time, separately for the whole image (dark grey, see insets) and for image regions viewing the world below (black) and above the horizon (light grey). The large variations are due to the movements of clouds, as can be seen by the sample images on top, which were recorded at 2 minute intervals. (b) Image differences at the same location over a period of three hours (same scene as Fig. 10, left panels). Images were recorded at 10-second intervals intermittently over 10 minute periods. Different grey-levels indicate the aperture settings of the camera lens during the recording. The reference image was recorded at 13:14 hrs (t = 0). Dots on the x-axis mark the times at which 2D difference functions were recorded (see Fig. 10). The rapid variations in image differences are probably due to the movement of clouds, wind-driven vegetation and shadows, the slow change is due to the change in the direction of illumination. (c) Long and short term variation of image differences at the same location on a windy, overcast day (same

scene as in Fig. 10, right panels). The reference image was recorded at 13:50 hrs (t = 0). Note the large variations due to environmental motion and clouds and the comparatively minor change in image differences due to changes in the direction of illumination. Other conventions as before.

Fig. 13 Homing by gradient descent: Comparison of two algorithms which were tested in a plane approximately 20 cm above ground in an open area on a calm, clear day. For details see text. The difference function relative to the reference image at the centre of this location is shown in (a). The performance of the two algorithms is shown in (b) and (c): in "***RunDown***" the gantry moved in one direction as long as mean squared (m.s.) pixel differences became smaller. If image differences increased, the direction of movement was changed by 90 degrees. In the second algorithm, "***Triangular***", the gantry determines m.s. pixel differences relative to the reference image at three positions at the corners of a small triangle, and subsequently moves in the direction of the minimum. We tested the performance in a randomised sequence of 40 runs. (b) and (c) show the two-dimensional paths with starting positions marked by dots. (d) and (e) show the m.s. pixel differences plotted over the distance from the reference location. (b) The results of 18 gradient descents using the '***RunDown***' algorithm (two out of the 20 runs are shown on their own in Fig. 14. Note that only two of the runs do not reach the goal (thick lines). (c) The results of 18 gradient descents using the '***Triangular***' algorithm (two of the 20 runs are shown on their own in Fig. 14). Otherwise conventions as in (b).

Fig 14 Homing by gradient descent: Effects of changes in illumination. The figure shows examples in which illumination changed during the execution of a gradient descent run. Both the "RunDown" and the "Triangular" algorithms appear to be immune to and able to recover from

changes in illumination since the large image differences they cause also arrest the gradient descent schemes at the position they currently occupy. Return to normal conditions allows the schemes to progress to the goal position. All traces except the light-grey one are from the experimental session shown in detail in Fig. 13. Other conventions as before.

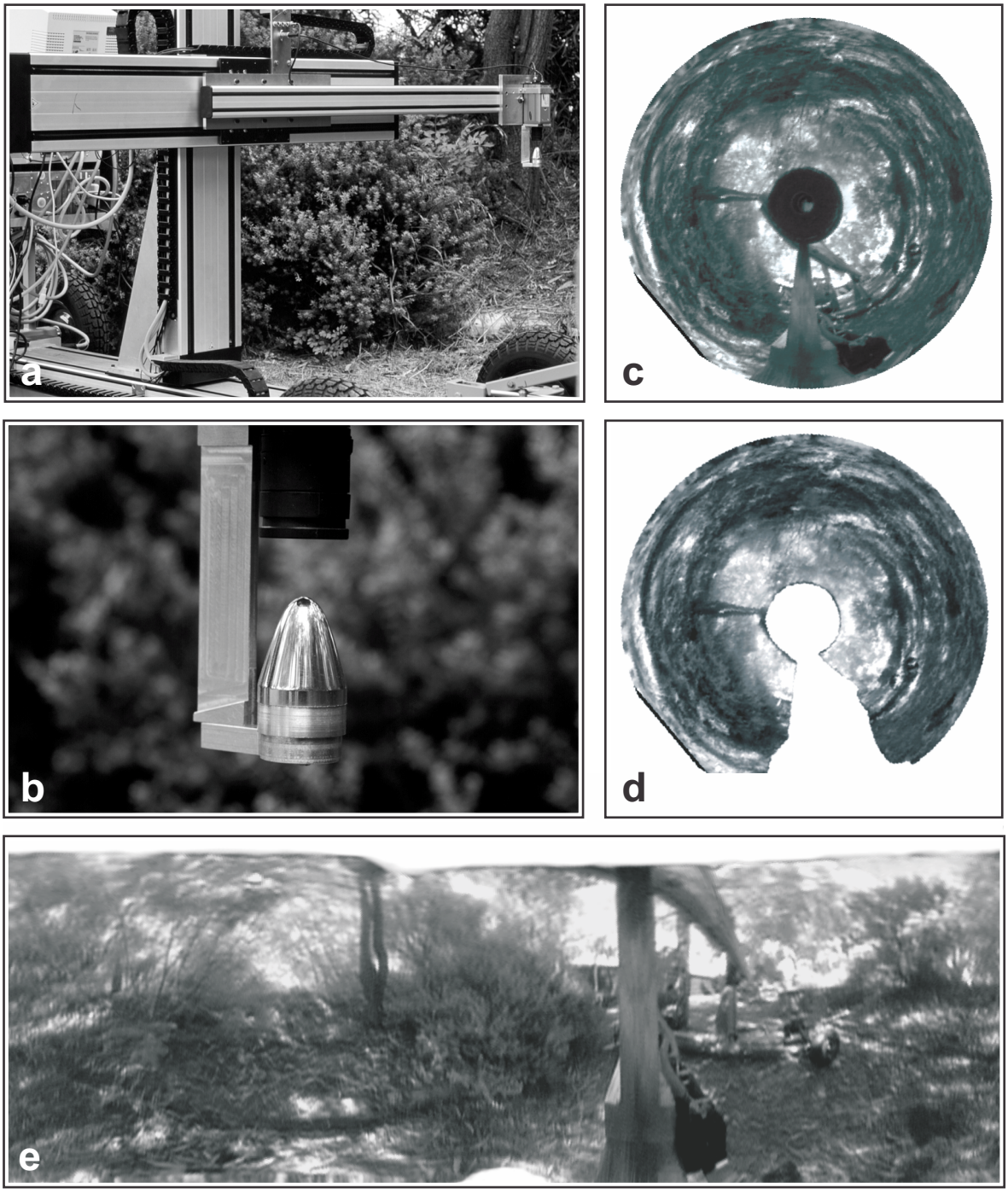

Fig. 1 (a) The robotic gantry in its natural habitat. The panoramic imaging device, consisting of a video camera and a reflective surface can be seen at the end of the horizontal y-axis arm at the far right of the picture. (b) Close-up of the panoramic imaging surface and the camera lens. (c) Panoramic image after a circular mask was applied to the original video image. (d) Panoramic image after applying an additional mask blocking the main gantry and the image of the camera and the camera lens. (e) An unwarped version of the panoramic image shown in c, after removing the image regions containing the camera lens.

Fig. 1
Zeil Hofmann Chahl