Biol Cybern (2007) 96:519–531
DOI 10.1007/s00422-007-0147-3

ORIGINAL PAPER

Biological
Cybernetics

Depth, contrast and view-based homing in outdoor scenes

Wolfgang Stürzl · Jochen Zeil

Received: 1 August 2006 / Accepted: 13 February 2007 / Published online: 19 April 2007

Abstract Panoramic image differences can be used for view-based homing under natural outdoor conditions, because they increase smoothly with distance from a reference location (Zeil et al., J Opt Soc Am A 20(3):450–469, 2003). The particular shape, slope and depth of such image difference functions (IDFs) recorded at any one place, however, depend on a number of factors that so far have only been qualitatively identified. Here we show how the shape of difference functions depends on the depth structure and the contrast of natural scenes, by quantifying the depth-distribution of different outdoor scenes and by comparing it to the difference functions calculated with differently processed panoramic images, which were recorded at the same locations. We find (1) that IDFs and catchment areas become systematically wider as the average distance of objects increases, (2) that simple image processing operations—like subtracting the local mean, difference-of-Gaussian filtering and local contrast normalization—make difference functions robust against changes in illumination and the spurious effects of shadows, and (3) by comparing depth-dependent translational and depth-independent rotational difference functions, we show that IDFs of contrast-normalized snapshots are predominantly determined by the depth-structure and possibly also by occluding contours in a scene. We propose a model for the shape of IDFs as a tool for quantitative comparisons between the shapes of these functions in different scenes.

1 Introduction

Many pieces of evidence and theoretical considerations suggest that places are uniquely defined by the visual appearance of the world as viewed from these places. Experiments on homing insects show that they return to nest sites and food sources with the aid of remembered views (Cartwright and Collett 1983; for reviews see Collett and Zeil 1998; Giurfa and Capaldi 1999). The properties and limits of view-based homing have recently received renewed attention in theoretical studies, backed up by simulations and experimental robotics (for reviews see Franz et al. 1998; Zeil et al. 2003; Vardy and Möller 2005). The principal limits of view-based homing have been recognized early on (e.g., Cartwright and Collett 1983; Zeil 1993a,b): for instance, in a featureless landscape, in which objects are far away, the appearance of a scene does not change, except over large distances of travel. Also, scenes in dense vegetation habitats are subject to large changes in appearance due to changes in the direction of illumination and the ever-changing position of shadows (e.g., Zeil et al. 2003). The accuracy and robustness of view-based homing thus depends on the depth structure of scenes and on the stability of the light environment.

Homing animals like ground-nesting bees and wasps when acquiring a representation of their nest environment during their learning flights on departure (see Zeil et al. 1996; Collett and Zeil 1997a) may, therefore, need ways of accounting for these conditions. For instance, if nesting in open country,

W. Stürzl (✉) · J. Zeil
ARC Centre of Excellence in Vision Science
and Centre for Visual Sciences,
Research School of Biological Sciences,
The Australian National University,
PO Box 475, Canberra, ACT 2601, Australia
e-mail: wolfgang.stuerzl@anu.edu.au

J. Zeil
e-mail: jochen.zeil@anu.edu.au

Present Address:
W. Stürzl
Robotics and Embedded Systems,
Computer Science Department,
Technical University Munich,
Boltzmannstr. 3, 85748 Garching, Germany

they would need to store more reference images or snapshots, because the distant scene above the horizon contributes only very small image differences per distance travelled, while close textures on the ground change very rapidly with distance. In addition, regardless of depth structure, homing insects are likely to need ways of making their representation immune against changes of illumination and the movements of shadows (Zeil et al. 2003; Möller 2002).

Here we explore these issues by showing how panoramic image differences depend on the depth structure of outdoor scenes, how lateral inhibition and simple local contrast normalization make image differences immune against changes of illumination and how the shape of difference functions can be modelled for quantitative comparison.

2 Methods

Image acquisition: We recorded panoramic images with a colour FireWire CCD-camera (Marlin MF-046C, Allied Vision Technologies, image size 640 × 480 pixels) viewing a convex mirror with constant gain in elevation (see Chahl and Srinivasan 1997). To achieve minimum obstruction and a rigid assembly, the mirror was held by four thin metal blades below the camera (see Fig. 1b). This panoramic imaging arrangement was attached to a 3D positioning platform (robotic gantry), which was mounted on a trolley and could be moved into different outdoor locations (Fig. 1a). The gantry allowed us to move the camera with high precision (re-positioning accuracy <0.01 mm) within the space of 1 m^3 (see Fig. 1a). The gantry was levelled with a spirit level before each experiment. The camera settings were kept constant during recordings, with the automatic gain control switched off and the built-in gamma correction enabled to increase the dynamic range. In order to avoid very low image contrast, high sensor noise, or saturation in very bright image parts, the lens aperture was changed between recordings if necessary.

Quantifying depth: We measured the depth structure of outdoor scenes with a laser range finder (SICK LMS 200) mounted sideways on a large turntable driven by a stepper motor (see Fig. 1c). The range finder was centred above the rotation axis of the turntable at a height above ground of about 75 cm and the whole arrangement was levelled using a spirit level. The range finder was oriented and programmed in such a way that its measuring beam scanned through a vertical slice of 180° in 1° steps in synchrony with the turntable rotation of 1° step at a time. A complete rotation of the turntable took about 5 min, producing a range map containing 360 × 181 distance measurements $\{r_i\}$ in the range from 10 cm to 80 m with a resolution of 1 cm. The turntable itself covered about 30° in the lower part of the range map, so that the useful depth array ended 60° below the horizontal. To relate this range map to vision, where image shifts due to observer movement are inversely proportional to distance, we plot depth as disparity values (pixel shifts or 'nearness'; see Eqs. (9), (10) in the Appendix):

$$d_i = \alpha / r_i \ , \tag{1}$$

with $\alpha = 180/\pi \times 0.1\,\mathrm{m}$. d_i is the expected image shift (or disparity) in pixels when moving the camera 0.1 m orthogonal to the direction of an object with distance r_i assuming an image resolution of 1°/pixel. For completely flat, textured ground, d_i is approximately linearly related to elevation ε_i below the horizon since $d_i = \alpha|\sin\varepsilon_i|/z \approx \alpha|\varepsilon_i|/z$ for $|\varepsilon_i| \ll \pi/2$ (with z being the height of the camera above ground).

Recording procedure: In preparation of a recording session we moved the camera to the centre of the gantry's horizontal range (gantry coordinates $x = y = 50$ cm) about 1 m above ground (gantry coordinate $z = 100$ cm). The horizontal distance from the body of the gantry of this reference location was about 2.25 m throughout this study. We then centred and levelled the laser range finder directly below the camera at a height above ground of between 70 and 85 cm, depending on the slope of the terrain (see Fig. 1a). After recording the distance map range finder and turntable were removed and the effective viewpoint of the camera was moved to the height above ground of the viewpoint of the range finder to record a reference image. Since we were using a mirror with constant angular gain there is strictly speaking no single viewpoint of the panoramic imaging system (e.g., Chahl and Srinivasan 1997). However, all principal rays cross the symmetry axis of the camera system within 0.7 cm of a point close to the tip of the mirror. We then recorded images at regular 10 cm intervals in an 11 × 11 element grid, defining a horizontal plane of 1 m^2. To avoid the effect of vibrations generated by the movement of the gantry, the images were recorded with a delay of about 1 s, after moving to the next grid location.

Calculating image difference functions: Differences between the images $\{\mathbf{I}(\mathbf{x})\}$ recorded at regularly spaced positions in a 11×11 grid and the reference image $\mathbf{I}^{\mathrm{r}}$ were calculated using the sum of squared pixel differences (SSD). Grid spacing was 10 cm. In the following we will use the term (translational) image difference function (IDF) to describe the dependence of the image difference on location, i.e.,

$$\mathrm{IDF}(\mathbf{x}) = \sum_i \left(I_i(\mathbf{x}) - I_i^{\mathrm{r}}\right)^2 . \tag{2}$$

Image processing: Before further processing, original images (Fig. 1d) were un-warped to a rectangular size of 360 × 104 pixels (Fig. 1e) and converted to grey level images. To avoid

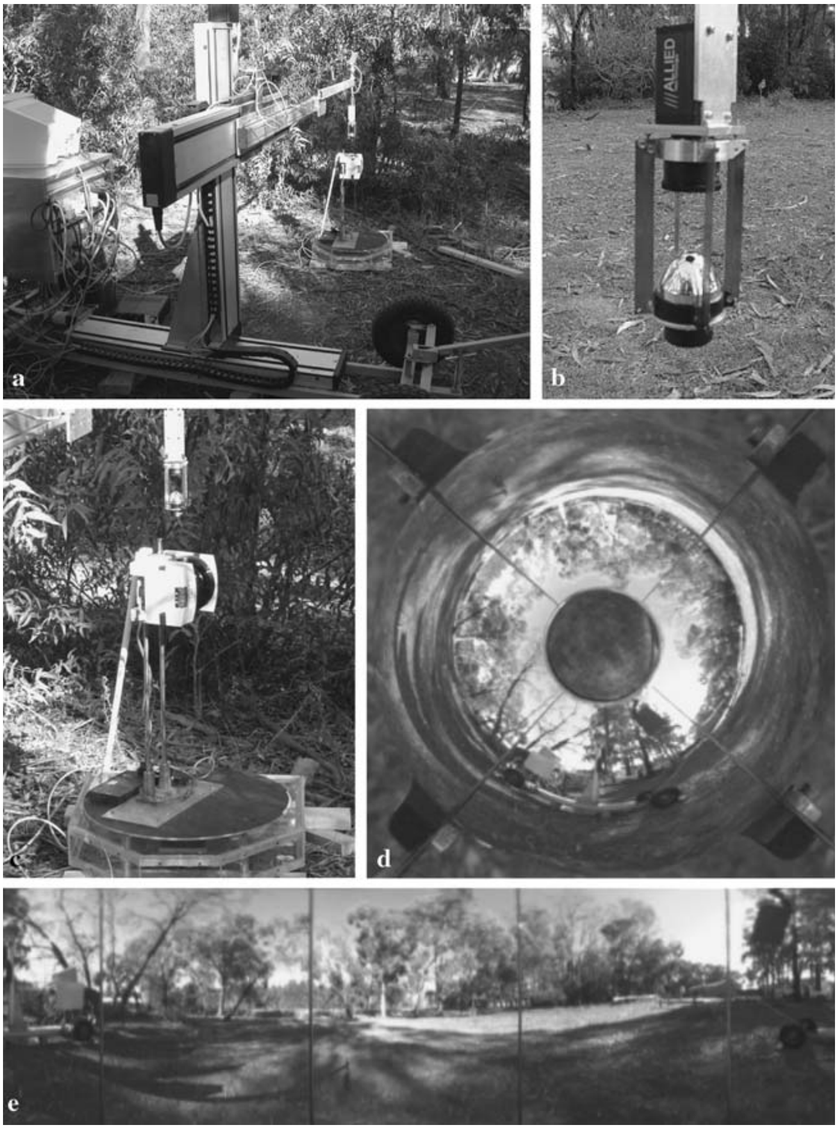

Fig. 1 The experimental setup. **a** View of the 3D positioning platform (gantry) together with the laser scanner on its rotating table. **b** Panoramic imaging device consisting of a Marlin FireWire camera looking down onto a reflective surface, which is held in place by four thin radial metal plates. **c** Sick Laser Scanner mounted sideways on a stepper-motor driven rotating table. **d** Original panoramic image of an open field site imaged via the reflective surface. **e** Un-warped panoramic image. The thin vertical strips are the images of the holding plates

aliasing, images were first un-warped to $1,800 \times 208$ pixels using bilinear interpolation, before being scaled down using super-sampling. The images cover approximately a range of elevation between $-50°$ and $+50°$ (plus a 2° border that is used when filtering the images, see below), except for some cases where we had to restrict image size to $-37°$ and $+50°$, due to scratches on the mirror surface. We decided to un-warp the images before calculating image differences or before applying filters because they provide a better approximation of a spherical image projection than the original camera image. The resolution of un-warped panoramic images was 1°/pixel in azimuth and elevation.

We investigated the effects of manipulating the image frequency spectrum and image contrast on difference functions by applying the following filter to the images (with pixel values $\mathbf{I} = (I_1,\ I_2,\ \ldots,\ I_N)^\top$) before calculating the SSD:

$$I'_i = \frac{I_i - \langle I\rangle_i}{1 + \sigma_{1/2}^{-1}\sigma_i}, \tag{3}$$

with $$\sigma_i = \sqrt{\langle (I_i - \langle I\rangle_i)^2\rangle_i}. \tag{4}$$

$\langle X\rangle_i = \sum_j g_{ij} X_j$ is the Gaussian weighted mean of X (the standard deviation of the Gaussian was 1°) in the neighbourhood of the ith pixel; σ_i is the local standard deviation of the intensity values over the same Gaussian window and $\sigma_{1/2}^{-1}$ is a normalization constant.

For $\sigma_{1/2}^{-1} = 0$, the local mean is subtracted from each intensity value. This “local background differencing” operation enhancing local contrast is equivalent to lateral inhibition and mimics the responses of laminar monopolar cells in insects (e.g., Srinivasan et al. 1982; van Hateren 1992, 1993). For $\sigma_{1/2}^{-1} > 0$, Eq. (3) becomes a variant of “divisive contrast normalization" models, which have been proposed to explain the properties of simple and complex cells in the mammalian visual cortex (e.g., Fleet et al. 1996; Carandini et al. 1997). With local contrast defined as $c_i = \sigma_i/\langle I \rangle$, where $\langle I \rangle$ is the mean intensity value of the image, Eq. (3) equalizes local contrast. For two image patches, for instance, with low and high local contrast c_1 and c_2, the ratio c_1/c_2 will be closer to 1 after normalization. The higher the value of $\sigma_{1/2}^{-1}$ the stronger is the effect. In the present study, we will be using values of $\sigma_{1/2}^{-1} = 0.5$ and $\sigma_{1/2}^{-1} = 100$ to achieve different degrees of local contrast normalization.

3 Results

3.1 Image differences and the depth structure of natural scenes

We first describe the relationship between the depth structure and the IDFs recorded in natural scenes, using an open and a forest habitat as examples. The colour-coded disparity (depth) maps of the two scenes are shown in Fig. 2a, left and right panels, together with depth histograms at different elevations in the scenes (Fig. 2b, as indicated by the angular width in elevation of horizontal slices through the laser scans in Fig. 2a), which extend from −50° below to +37° above the horizon. Figure 2c shows the IDFs for the same elevations and Fig. 2d x and y transects through the equivalent IDFs. Comparing the distance histograms for the two scenes (left and right panels, Fig. 2b) shows that there is a clear division between above and below horizon distances in the open habitat. Above the horizon large distances are represented about equally, and below the horizon, the distribution is practically flat across a wide range of close distances because of perspective foreshortening. On flat ground, each distance slice from the horizon downwards, contributes about equally. The histogram of the densely vegetated habitat is different in that above horizon and below horizon distances contribute much more equally to the distribution compared to the open habitat. Furthermore, the depth distributions of the whole image and of different elevations are multi-modal, except for very low elevations, firstly because close objects are not restricted to the parts of the scene below the horizon, and secondly because below the horizon, the effects of perspective foreshortening are "contaminated" by close objects of different sizes that are seen against the background of the ground plane.

The two-dimensional IDFs (Fig. 2c) and the transects through their centre in x and y direction (Fig. 2d) for the two scenes show characteristic differences, both with respect to scene differences and with respect to elevation. The first result to note is that the IDF of the whole image (top left panels in Fig. 2c and d) is shallower in the open scene (left), compared to the forest scene (right). These differences seem to be mainly produced by objects above the horizon (top row panels in Fig. 2c, d), while the IDFs calculated for elevations below the horizon do not differ significantly between the two scenes (bottom row in Fig. 2c, d). Those of the forest scene, however, are corrupted by shadow contours that have changed between the recording of the reference image and the images in the horizontal grid. Note that the image differences at the reference image location are not zero in Fig. 2 and the following figures, because the images on the grid of 11 × 11 locations were recorded up to 10 min after the reference image, so that the image at the centre of the grid is not identical to the reference image.

The differences between the shape and depth of IDFs in these two scenes are unlikely to be due to differences in local image contrast. We quantified image contrast by applying 3 × 3-degree-wide horizontal and vertical Sobel filters[1] which approximate the local image gradient at each pixel by determining the differences between neighbouring pixels. The histograms of the local slope of intensity changes as estimated by Sobel filters applied to the reference images recorded in the two scenes (see Fig. 3a) are practically indistinguishable between the open and the forest habitat (Fig. 3b). The same holds true for the different elevation ranges. In both scenes, the distributions are broadest, meaning that contrast is highest, above and at the horizon.

3.2 Image differences and the local contrast in natural scenes

For a number of reasons, we wanted to explore how IDFs depend on pre-processing of images. First, we looked for ways to alleviate the effects of changes in illumination and the movement of shadows (e.g., Zeil et al. 2003; Möller 2002); secondly, we wanted to mimic the early stages of insect visual processing, which are known to involve lateral inhibition (e.g., Srinivasan et al. 1982; van Hateren 1992, 1993); and thirdly we looked for ways of modifying the spatial frequency spectrum of images, to investigate how IDFs depend on second-order natural scene statistics. We chose to compare three pre-processing strategies (see Fig. 4 and Sect. 2): Lateral inhibition or difference-of-Gaussian (DoG) filtering $\left(\sigma_{1/2}^{-1} = 0\right)$, and two different degrees of local contrast normalization ($\sigma_{1/2}^{-1} = 0.5$ and 100, see Sect. 2). The degree to which these filtering operations reduce the dependence of the Fourier amplitudes on the spatial frequency of the images (i.e., lead to a whitening of the spatial frequency spectrum)

[1] Changes in pixel value due to translation $\overrightarrow{\Delta x}$ can be estimated by $\Delta I_i(\overrightarrow{\Delta x}) \approx \partial_\phi I_i \Delta\phi_i(\overrightarrow{\Delta x}) + \partial_\varepsilon I_i \Delta\varepsilon_i(\overrightarrow{\Delta x})$ for $|\overrightarrow{\Delta x}/r_i| \ll 1$. $\partial_\phi I_i$ and $\partial_\varepsilon I_i$ are local image gradients (in azimuth ϕ and elevation ε) that can be approximated by the output of vertical and horizontal Sobel filters. $\Delta\phi_i(\overrightarrow{\Delta x})$ and $\Delta\varepsilon_i(\overrightarrow{\Delta x})$ are the corresponding image shifts, see Eqs. (9),(10) in the Appendix.

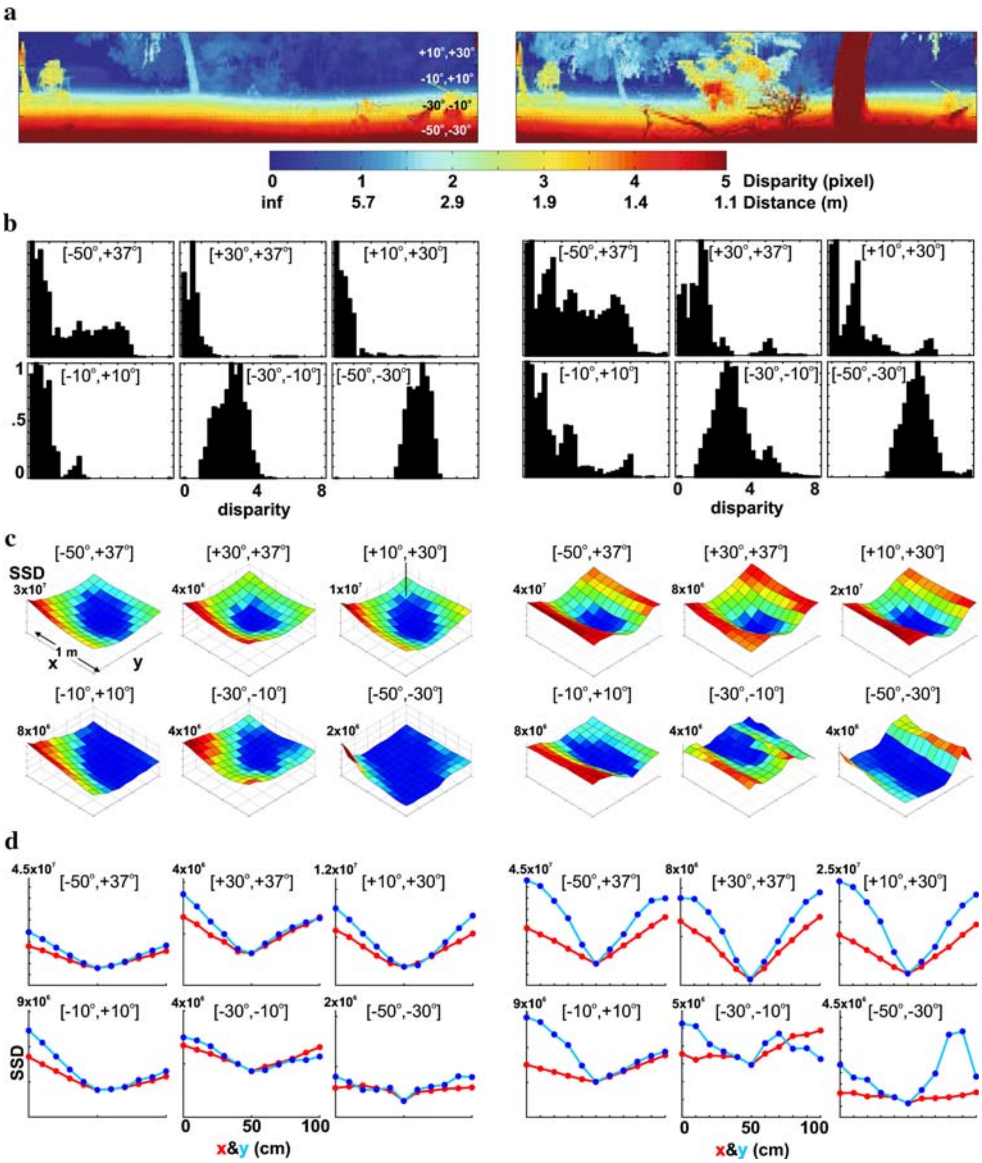

Fig. 2 The depth structure of outdoor scenes. **a** Colour-coded range maps for an open (*left*) and a forest site (*right*). Colour bar shows coding of distances in units of disparity (pixels) and of absolute distance (m). Angular elevation ranges relative to the horizon are shown in 20° steps in the left range map. **b** Normalized histograms for the two scenes are shown for the whole scenes (*top left histogram*) and for each of the 20° elevation ranges indicated in the left range map in **a**. **c** Image difference functions, calculated as sum of square pixel differences (*SSD*), Eq. (2), for a horizontal plane using the image of the whole scene (*top left surface*) and image segments corresponding to the different elevation slices shown in the *left* and *right* range maps in **a**, as indicated above each IDF. **d** Transects along the x- (*red*) and y-direction (*blue*) through the IDFs shown in **c**. Otherwise conventions as in **c**. Note that subplots in **c** and **d** are plotted at different scales to emphasize the shape of IDFs. Colour online only

can be seen in Fig. 4a, which shows the mean amplitudes of the row-wise Fourier transform for unfiltered and filtered images in one example scene. Clearly, local contrast normalization also equalizes the amplitudes of the Fourier spectra for different elevations, which means that the contribution each elevation makes to the overall IDF becomes almost the same (see below).

These pre-processing strategies have quite dramatic effects on IDFs (see transects through IDFs of the open scene (light grey) and the forest scene (black) in Fig. 4c–f): compared with IDFs determined with the raw images (Fig. 4c), pre-processing leads to steeper, more symmetrical IDFs with more pronounced slope and the degrading effects of shadows on the ground increasingly disappear with increasing degrees of local contrast normalization (Fig. 4e, f). The IDFs at the two sites become very similar, mainly because the slope and minimum of the IDF for the open site become more pronounced (e.g., left column Fig. 4c–f). It remains to

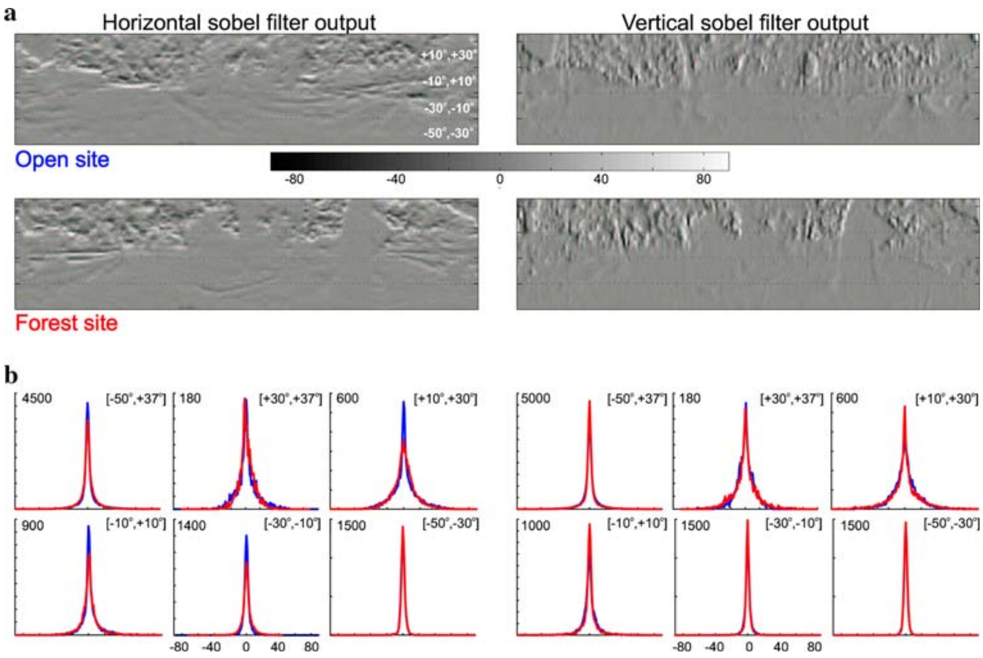

Fig. 3 Image contrast in two different outdoor scenes. **a** Horizontal (*left*) and vertical Sobel filtered images (*right*) of the open (*top*) and the forest scene (*bottom*). **b** Probability density functions of horizontal (*left panels*) and vertical Sobel contrast (*right panels*) for the whole scene (*top left panels*) and horizontal slices through the scenes as indicated by the elevation ranges in *square brackets*, referring to the numbers given in the *top left image* in **a**. Colour online only

be the case, however, that the IDFs at the forest site are narrower and steeper (black transects), compared with the ones recorded at the open site (light grey transects, Fig. 4c–f). These results are robust against varying illumination conditions and against slight modifications of view-points, as can be seen in Fig. 5 which shows panoramic images, laser scans, distance distributions (Fig. 5a, b) and transects (Fig. 5c, d) through the IDFs on two different days (Fig. 5a, b) at about, but not exactly, the same locations as before. New reference images were recorded on each day. The shallow and "noisy" IDFs we recorded for both locations on both days (left panels in Fig. 5c, d) develop a clear minimum and steep slopes by pre-processing (Fig. 5c, d, columns two to four).

We note that pre-processing thus makes IDFs robust against changes in illumination and the influence of shadows. The IDFs are narrower because low frequencies are reduced as a result of DoG-filtering, which is a high pass in our implementation and more generally a bandpass. However, the IDFs of the open scene (e.g., grey transects in Fig. 5), are still shallower compared to the IDFs of the forest scene (black transects in Fig. 5). Although the IDFs for the different elevation slices look quite similar for $\sigma_{1/2}^{-1} = 0$, 0.5 and 100, except for some distortions caused by changes in illumination, the overall IDFs for contrast normalized images are more cusp-shaped because lower elevation ranges ($[-50^\circ, -30^\circ]$ and $[-10^\circ, -30^\circ]$) contribute more to the overall IDF as a result of their higher contrast after normalization.

3.3 The shape of IDFs and the depth structure of natural scenes

Are the shape and depth of the pre-processed IDFs thus solely determined by the depth structure of different scenes and possibly by the effects of occlusion? We approached this question in two ways: firstly, we calculated the rotational IDFs for the same scenes by shifting the un-warped images, with and without pre-processing and we secondly developed a model of the translational IDFs to determine shape values, which we then could test for their dependence on the different distance distributions in outdoor scenes.

The rationale for investigating the properties of rotational IDFs in this context was the following: image differences for panoramic image rotations are independent of the distance of objects in the scene and, because there is no motion parallax, there are no effects of occlusion (e.g., Zeil et al. 2003). The image differences caused by image rotation depend exclusively on the second-order statistics (the spatial frequency spectrum) of the image. They are equivalent to the cross-correlation function calculated by shifting image patches across their image neighbourhood (e.g., Baddeley 1997). We hoped that investigating the effects of pre-processing on rotational difference functions would allow us to disentangle the contributions of second-order statistics and the contributions of depth and occlusion to the properties of translational difference functions.

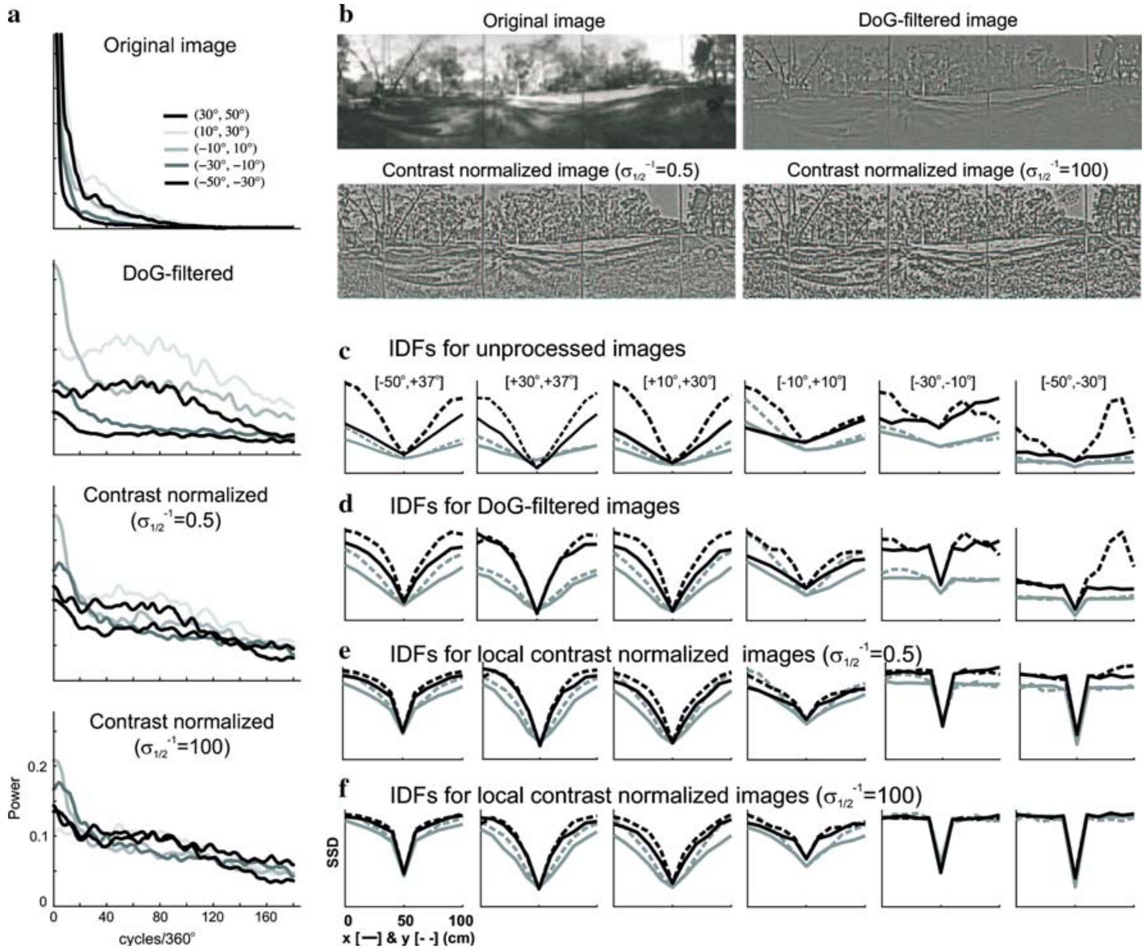

Fig. 4 The effect of image pre-processing on IDFs. **a** Horizontal Fourier amplitude spectrum for different horizontal image slices (as indicated in the *top* diagram) for differently processed images (original, difference-of-Gaussian filtered, weakly and strongly contrast normalized). To obtain the curves we used the following processing steps: images were scaled to length 1, i.e., $I_i \to I_i / \left(\sum_i I_i^2\right)^{1/2}$; amplitudes of one-dimensional (*row-wise*) discrete Fourier-coefficients were calculated; amplitudes were averaged for each of the five horizontal slices; the curves were smoothed with a Gaussian filter with standard deviation 2.5 cycles/360°. We used the *row-wise* Fourier transform in order to make comparison between different elevation ranges easier. Note the increasing equalization of the Fourier amplitudes with increasing contrast normalization. **b** The original and three differently filtered images for the open scene. **c**–**f** Transects along the *x*- (*continuous lines*) and the *y*-direction (*dashed lines*) through the IDFs determined with unprocessed and processed images for the forest (*black*) and the open scene (*grey*) for the whole image (*left panels*) and five horizontal 20°-wide slices as indicated by *numbers* in *square brackets* in **c**

The results are surprisingly clear: the typical cusp-shaped rotational IDFs, calculated with raw images, collapse into very narrow, steep and deep shapes that have a very small catchment area, regardless of elevation (Fig. 6). This property of rotational IDFs is independent of scene composition and variation in illumination (compare different rows in Fig. 6). The reason for this stark effect is quite obvious: Lateral inhibition removes most low spatial frequencies from the image and contrast normalization reduces the differences in the contributions the different elevation slices make to the overall function (parameter A, see Fig. 7b and Eq. (5) below). The rotational IDF of an image with low spatial frequencies would be very broad and shallow. The IDF of an image with high spatial frequencies only—as we produced it by lateral inhibition and local contrast normalization—would be very steep and narrow. However, the most interesting observation in the present context is that—as we have seen—translational IDFs are not reduced to steep and narrow functions by local contrast normalization. Their shape, thus, must reflect other image properties than second-order statistics, because it is independent of the spatial frequency content of the images. The image properties that determine the shape of translational IDFs after local contrast normalization thus must depend on motion parallax, which in turn depends on only two remaining factors: the depth distribution of natural scenes and the effects of occlusion. We have not found a good way of quantifying the latter, but can do so with the former.

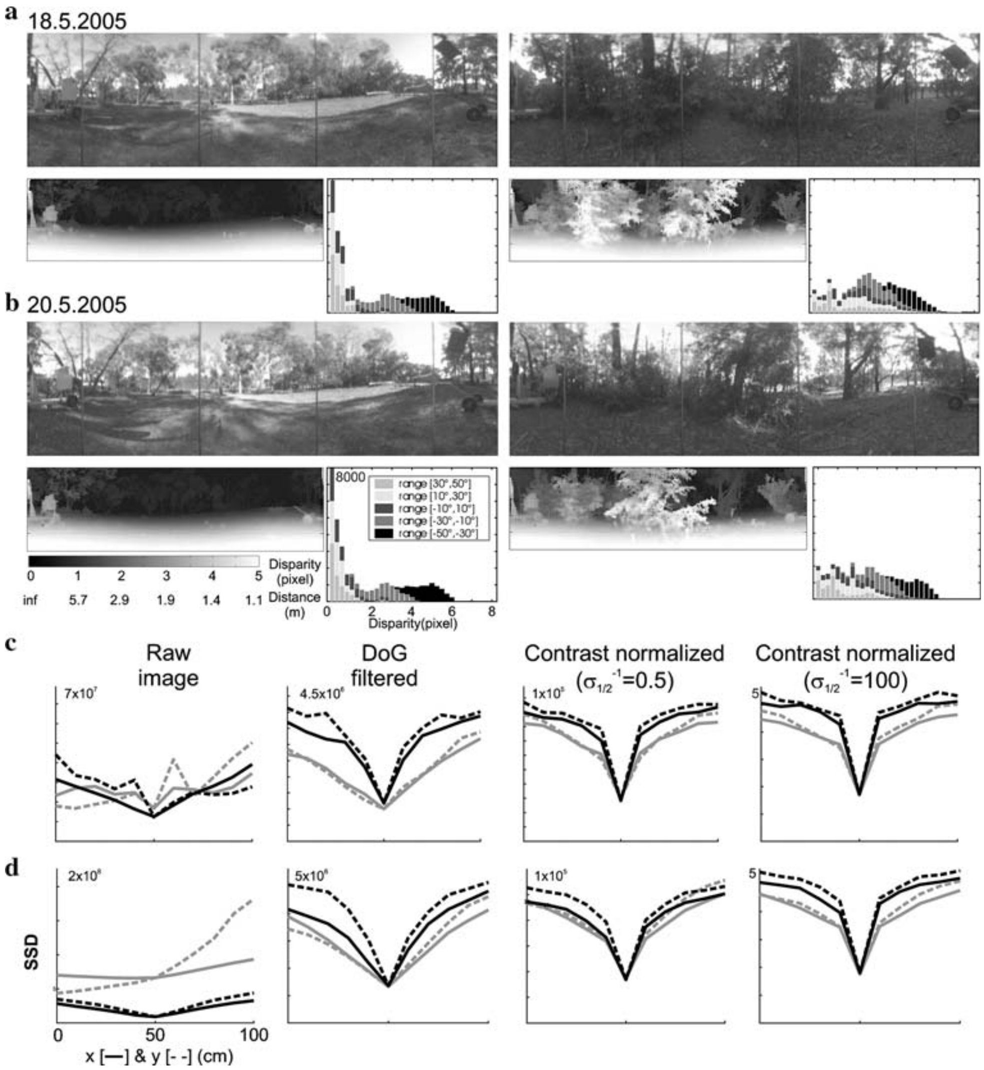

Fig. 5 Comparison of IDFs for two sites across time. **a** Panoramic images (*top*), range maps (*bottom left*), and range probability density functions (*bottom right*) for an open (*left*) and a forest site (*right*) recorded on 18 May 2005. **b** Similar, but not identical sites recorded on 20 May 2005. **c** Transects through IDFs for open (*grey*) and forest scenes (*black*) recorded on 18 May 2005 (see **a**), determined with unprocessed images (*left panel*) and differently processed images (*left to right*) as indicated. **d** Same as **c** for images recorded on 20 May 2005. New reference images were recorded on each day

To model IDFs, we used the function

$$\mathrm{IDF}(\mathbf{x}|A, C, \mathbf{S})$$

$$= A\left(1 - \frac{1}{1 + (\mathbf{x}-\mathbf{x}^{\mathrm{r}})^{\top}\mathbf{S}\,(\mathbf{x}-\mathbf{x}^{\mathrm{r}})}\right) + C, \qquad (5)$$

where A is the maximum depth of the IDF and C is the SSD value at the reference position $\mathbf{x}^{\mathrm{r}}$. The symmetric matrix $\mathbf{S}$, describes the spread of the IDF, see Fig. 7b. Since we compare images recorded at positions in the x-y plane, $\mathbf{S}$ is a 2×2 matrix. Eq. (5) can be approximated by the quadratic function $(\mathbf{x}-\mathbf{x}^{\mathrm{r}})^{\top} A\mathbf{S}\,(\mathbf{x}-\mathbf{x}^{\mathrm{r}}) + C$ for $\left|(\mathbf{x}-\mathbf{x}^{\mathrm{r}})^{\top}\mathbf{S}\,(\mathbf{x}-\mathbf{x}^{\mathrm{r}})\right| \ll 1$, and approaches the constant value $A + C$ for $\left|(\mathbf{x}-\mathbf{x}^{\mathrm{r}})^{\top}\mathbf{S}\,(\mathbf{x}-\mathbf{x}^{\mathrm{r}})\right| \gg 1$. A justification for and derivation of the model function is given in the Appendix where the translational IDF is calculated from the rotational IDF for contrast normalized images.

To describe the average width of the IDF by a single value, we introduce the parameter w defined by

$$w^2 = \det(\mathbf{S})^{-1/2} = w_1 w_2 \ . \qquad (6)$$

w_1 and w_2 describe the width of the IDF in the directions of the principal axes of $\mathbf{S}$.

The function in Eq. (5) provides a good fit to most of the IDFs we recorded, as documented for two examples in Fig. 7a, which shows for whole images (top row) and individual horizontal slices (row 2–6) the transects through the IDFs of the open and forest site recorded on 31 March 2005 (dashed lines and dots) and the least square fit of our model (solid lines). Fits were usually somewhat worse when-

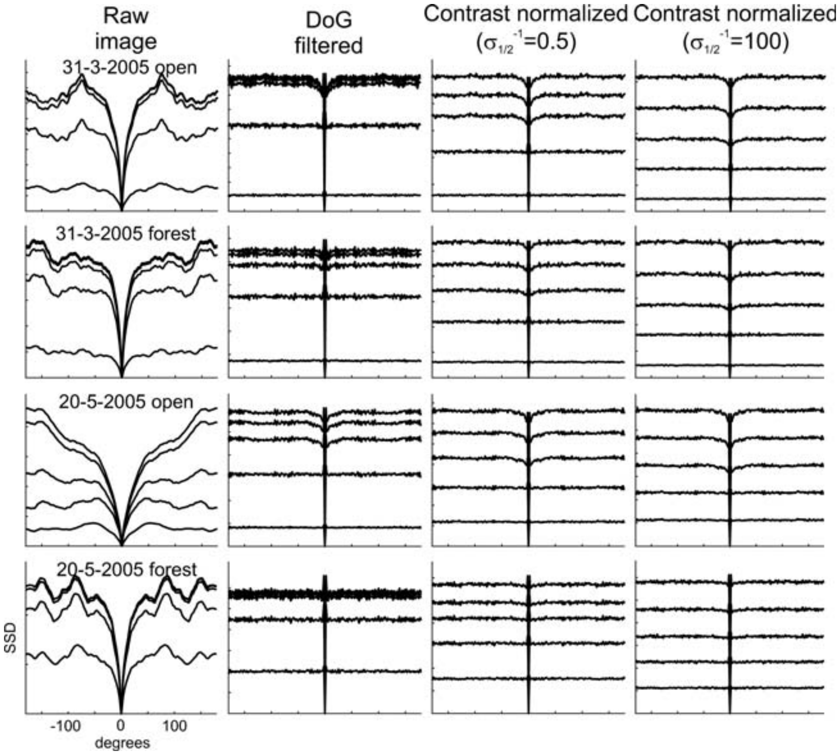

Fig. 6 The effects of image pre-processing (*left to right*) on rotational difference functions (IDFs) for different scenes recorded on different days (*top to bottom*, as indicated in *left panels*). Each *panel* shows five IDFs from bottom to top, which were determined using increasingly larger vertical parts of the image, starting below the horizon: the first IDF (*bottom*) shows the rotational IDF for the lowest 20° slice, the next one the lowest 40° slice, the third one 60°, aso, in such a way that the top IDF was calculated using the whole panoramic image. Note the dramatic change of shape of IDFs with different amounts of contrast normalization (*left to right*), which is very similar for the different scenes recorded on different days

ever the corresponding disparity distributions were broad, because the model function depends on the assumption that disparity distributions are narrow. Although in principle, IDFs for broad depth distributions can be modelled by a sum of two or more model functions, single model fits adequately serve our main present purpose of correlating IDF shape parameters with the depth structure of the scenes in which they were recorded. After fitting the model to the IDFs calculated for five horizontal slices through un-warped panoramic images, as shown in Figs. 2 and 5, we determined their average width w using Eq. (6) and plotted these values over $\langle 1/r\rangle^{-1} = \alpha/\langle d\rangle$, where $\langle d\rangle$ is the mean of the respective disparity distribution. The results are shown in Fig. 7c for the contrast normalized IDFs in the forest and the open scenes recorded on two different days (see Fig. 5). Clearly, the width of DoG-filtered and local contrast-normalized IDFs depends systematically on the depth structure of natural scenes. The values for w in Fig. 7c are also in good agreement with the calculations for contrast normalized images presented in the Appendix that predict $w \propto \langle 1/r\rangle^{-1}$.

4 Discussion

We confirmed that image differences in outdoor scenes develop smoothly with distance from a reference location (Zeil et al. 2003). We show here for the first time that image differences outdoors become immune against changes in illumination and shadow contours by local contrast normalization and that their shape after normalization depends almost entirely on the depth structure of scenes. Image difference functions generated by translation become systematically broader as the mean distance of objects increases. In contrast, image differences generated by changes in orientation [rotational image difference functions (rIDF)] become very narrow after local contrast normalization, regardless of the depth structure of scenes, and are, therefore, solely determined by the spatial frequency composition of natural scenes. We developed a model of IDFs, which can now be used to compare them quantitatively in different environments.

Our results reveal an interesting property of the natural world, namely that locations in it are uniquely defined by the views taken at these locations and that this property depends on the distribution of objects in natural scenes. Our results further confirm the earlier observation (Zeil et al. 2003) that the orientation of a view is also uniquely determined by that view itself, without the need for additional compass information. We conclude that any agent sensitive to image differences, be it an insect or a robot, would be able to pinpoint locations in the natural world with the aid of remembered views (Cartwright and Collett 1983, 1987; Franz et al. 1998; Franz and Mallot 2000; Zeil et al. 2003; Vardy and Möller 2005).

However, insect behaviour tells a different story and whether insects actually do store panoramic images is

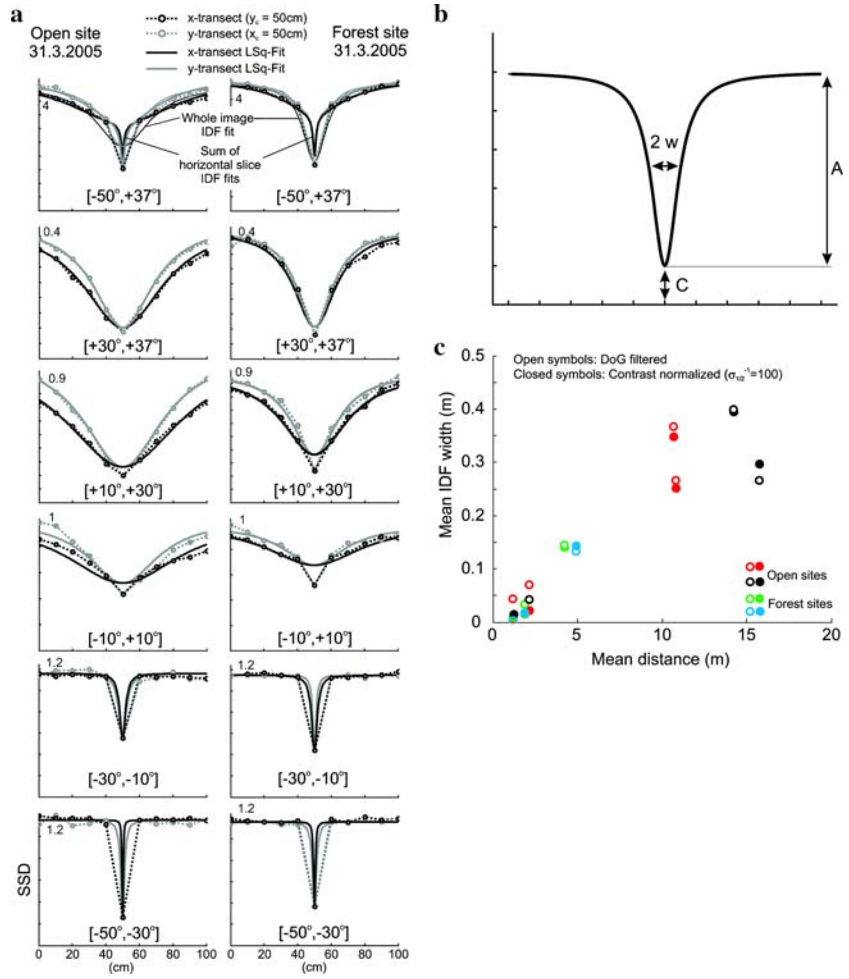

Fig. 7 Fitting the IDF model function to measured image differences. **a** Shown are x and y-transects through IDFs recorded on 31 March 2005 at an open (*left*) and a forest site (*right*) for contrast normalized images $\left(\sigma_{1/2}^{-1} = 100\right)$. IDF transects in x- and y-direction are shown as *dotted lines* as indicated in the *top inset*. *Continuous curves* show model fits to these transects. In the *top row* the thick continuous curves show the sum of the model fits for the five individual slices plotted below, the elevation range of which is given in *square brackets*. *Thin lines* are model fits for the whole image IDF. **b** Shown is the 1D illustration of the 2D model function IDF($\mathbf{x}|A, C, \mathbf{S}$) defined in Eq. (5) with parameters A: depth; C: offset; $2w$: width at half maximum. See text for details. **c** The width of IDFs depends on the depth structure of natural scenes. The half-width w of the fitted model IDF is plotted over the distance $\langle 1/r \rangle^{-1} = \alpha/\langle d \rangle$ in open and forest sites. Fits for the elevation range $[-10°, +10°]$ (*fourth row* in **a**) were excluded, as was the fit for the $[+10°, +30°]$ range of the forest site (*third row*, right), because a single model function fails to adequately fit broad depth distributions (see text). Colour online only

unknown. First of all, homing insects on departure from a place of significance, like a nest or a food source, do not simply take a snapshot, but instead employ an elaborate sequence of learning behaviour, before departing the location (e.g., Zeil 1993a,b; Lehrer 1993; Zeil et al. 1996; Collett and Zeil 1997a; Nicholson et al. 1999). The need for these learning flights, or turn-back-and-look behaviours, in both flying and walking insects has been interpreted in different ways. One possibility is that the insects cannot predict how large the catchment area of any given snapshot is, and therefore need to move and compare following acquisition, in order to decide when to take the next one (e.g., Gaussier et al. 2000). Sequences of snapshots recorded at different locations may also need to be linked in a systematic fashion, requiring clearly structured and systematic movements between the recording locations (e.g., Collett and Lehrer 1993). The surprisingly invariant dynamics of pivoting and rotational movements of flying insects, like wasps and bees, during learning flights may also indicate that this behaviour could serve an image processing function, namely depth filtering by creating a particular pivoting motion parallax field, emphasizing objects close to the goal (Cartwright and Collett 1987; Zeil 1993b; Collett 1995; Collett and Zeil 1997b; Voss and Zeil 1998). Support for this last conjecture comes from experiments showing that both wasps and bees acquire information on the absolute distance of landmarks relative to the goal during these flights (Zeil 1993b;

Brünnert et al. 1994; Lehrer and Collett 1994). Lastly, these elaborate behaviours during acquisition of a visual representation may reflect a need for "quality assurance": the insects may have to continuously check—again by moving and comparing—whether the representation they have acquired is robust and informative enough for a successful return. Because these learning flights are an example of active acquisition of visual information, they are hard to experimentally interfere with under the natural conditions in which they occur (but see Nicholson et al. 1999). We have recently begun to reconstruct the views experienced by wasps departing from and returning to their nests in the ground, in order to identify whether and how much both view acquisition and homing are driven by systematic view-based components.

The second reason why insects do not seem to only rely on panoramic image matching is that when searching for a goal, they are clearly guided by individual, visually salient objects (e.g., Collett and Zeil 1997b, 1998). Yet if these objects are removed (e.g., Zeil 1993b; Graham et al. 2003), the insects are still able to home into the general area in which the goal lies, or follow a route that was previously guided and influenced by the dominant landmark. These observations demonstrate that the insects do not only memorize individual salient features, but also the wider visual context in which they are seen. A number of other observations support this conjecture: bees that have been trained to fly through a narrow textured tunnel to find a food source some distance into the tunnel, do not fly beyond a novel overhead object (landmark) that is introduced during tests (Vladusich et al. 2005); the visual environment of waterstriders has to be reduced to a single overhead point of light, before their ability to maintain their position on a flowing water surface is compromised (Junger 1991); the homing ability of stingless bees is affected when the apparent position of distant landmarks beyond the nest and to the side of their normal flight path is changed (Zeil and Wittmann 1993). Although experiments such as these do suggest that at least insects do memorize the panoramic view of their environment, direct evidence is still lacking.

Potentially the most serious constraint on using memorized images for homing in the natural world is the temporal variability of views due to the movement of the sun, the movements of clouds, the movements of wind-driven vegetation and the resulting movements of shadows (Zeil et al. 2003; Zanker and Zeil 2005). Möller (2002) recently suggested a colour opponent representation as one possible way in which visual representations can be made immune against changes in illumination, by emphasising the contrast between terrestrial objects and the sky. We add here another possibility, namely lateral inhibition or local contrast normalization, which have the additional advantage of not discarding information below the horizon. All these operations involve early visual processing routines that biological vision systems are known to or are likely to employ (e.g., Srinivasan et al. 1982; van Hateren 1992, 1993; Osorio and Vorobyev 2005), thus indicating that the information on location we have documented here, is also available to biological agents moving through the natural world.

Acknowledgments A German Science Foundation (DFG) fellowship and a Centre for Visual Sciences (CVS) fellowship to WS supported this work. We acknowledge additional financial support from the ARC Centre for Excellence in Vision Science, the Human Frontier Science Project (JZ), the Australian Defence Science and Technology Organization (DSTO) and Eglin Airforce Base (JZ). We thank David Austin, Research School for Information Science and Engineering (RSISE) at the Australian National University for the loan of the Sick Laser Scanner, and Hartmut Surmann at the Fraunhofer Institute for Autonomous Intelligent Systems (AIS), Sankt Augustin, Germany, for revealing some crucial secrets of the laser range finder, Martin Hofmann and Katharina Siebke for their help with operating the gantry, and Dean Soccol for his expertise and care in constructing the panoramic imaging system. We are grateful to Mandyam Srinivasan for advice and to Norbert Boeddeker and Jan Hemmi for their continuous and constructive input to this project.

Appendix

Image shifts from translations: Objects seen at an image position defined by (ϕ_i, ε_i) (azimuth, elevation) will shift their position by $\Delta\phi_i = \phi'_i - \phi_i$ and $\Delta\varepsilon_i = \varepsilon'_i - \varepsilon_i$ depending on object distance r and translation $\boldsymbol{\Delta}\mathbf{x} = (\Delta x, \Delta y, \Delta y)^\top$. Ignoring occlusion, $\Delta\phi_i, \Delta\varepsilon_i$ can therefore be calculated from the coordinate transformation

$$\mathbf{x}'_i := \mathbf{X}\left(r'_i, \phi'_i, \varepsilon'_i\right) = \mathbf{X}\left(r_i, \phi_i, \varepsilon_i\right) - \boldsymbol{\Delta}\mathbf{x}, \tag{7}$$

$$\mathbf{X}(r, \phi, \varepsilon) := r\,(\cos\phi\ \cos\varepsilon, \sin\phi\ \cos\varepsilon, \sin\varepsilon)^\top, \tag{8}$$

using $\phi'_i = \operatorname{arctan2}\left(y'_i, x'_i\right)$ and $\varepsilon'_i = \arctan\left(z'_i\left(x'^2_i + y'^2_i\right)^{-\frac{1}{2}}\right)$.

The non-linear mapping $(\phi_i, \varepsilon_i) \rightarrow \left(\phi'_i, \varepsilon'_i\right)$ can be simplified for $|\boldsymbol{\Delta}\mathbf{x}|/r_i \ll 1$ and $|\varepsilon_i| \ll \frac{\pi}{2}$. The image shifts are then given by

$$\Delta\phi_i \approx \frac{\Delta x \sin\phi_i - \Delta y \cos\phi_i}{r_i \cos\varepsilon_i}, \tag{9}$$

$$\Delta\varepsilon_i \approx \frac{(\Delta x \cos\phi_i + \Delta y \cos\phi_i)\sin\varepsilon_i}{r_i} - \frac{\Delta z \cos\varepsilon}{r_i}. \tag{10}$$

Modelling translational IDFs for contrast normalized images: Motivated by the almost constant shape of the rIDF of different images and horizontal slices for contrast normalized images $\left(\text{with } \sigma_{1/2}^{-1} = 100\right)$ we will estimate the translational IDF. A model for IDFs for a perspective camera moving parallel to objects at a constant distance, for instance a textured wall, has recently been proposed by Szenher (2005). Although the rIDF (between an image and its rotated version) for a local patch usually differs slightly from the overall rIDF, we make the following assumption: The local rIDF depends only on the rotation angle $\Delta\alpha$. More specifically we model the local rIDF by the function (a flipped Gaussian)

$$\mathrm{rIDF}(\Delta\alpha) = A\left(1 - e^{-\frac{\Delta\alpha^2}{2\eta^2}}\right) + C. \tag{11}$$

We found reasonable fits for $\eta = 0.65°$ when using the overall rIDF between the un-warped reference image and an un-warped image recorded later at the same position. Fits to different horizontal and vertical slices usually gave values for η in the range $[0.55°, 0.75°]$.

Since translations $\overrightarrow{\Delta x}$ cause image shifts $\Delta\phi_i$, $\Delta\varepsilon_i$ defined in Eqs. (7)–(10), the translational IDF can be calculated according to

$$\mathrm{IDF}(\overrightarrow{\Delta x}) = \frac{1}{N}\sum_i \mathrm{rIDF}\left(\Delta\alpha_i(\overrightarrow{\Delta x})\right), \tag{12}$$

where the rotation angle for un-warped images is given by $\Delta\alpha_i^2 \approx \Delta\phi_i^2 + \Delta\varepsilon_i^2$. The sum is over all pixels that are used for calculating the image difference, e.g., all pixels within a slice.

If we consider a translation in x-direction we have

$$\mathrm{IDF}(\Delta x) = A\left(1 - \frac{1}{N}\sum_i e^{-\frac{\Delta\phi_i^2(\Delta x)+\Delta\varepsilon_i^2(\Delta x)}{2\eta^2}}\right) + C. \tag{13}$$

This model of a translational IDF for contrast normalized images depends solely on the depth structure of the scene $\{r_i\}$ (and the translation Δx), i.e., the IDF can be estimated by calculating for each Δx the mapping $(\phi_i, \varepsilon_i) \to \left(\phi_i', \varepsilon_i'\right)$ for the given $\{r_i\}$. Using the linearizations (9), (10), Eq. (13) can be approximated for $|\Delta x|/r_i \ll 1$ by

$$\mathrm{IDF}(\Delta x) = A\left(1 - \frac{1}{N}\sum_i e^{-\frac{\Delta x^2}{2\eta^2 r_i^2}\left(\frac{\sin^2\phi_i}{\cos^2\varepsilon_i}+\cos^2\phi_i\,\sin^2\varepsilon_i\right)}\right) + C. \tag{14}$$

If we assume almost constant distances in a horizontal slice, i.e., $1/r_i \approx 1/R := \langle 1/r_j\rangle_j$, $\varepsilon_i \approx \varepsilon$ for all i, we can integrate over azimuth ϕ and find

$$\mathrm{IDF}(\Delta x) \approx A\left(1 - e^{-b(\Delta x)}\frac{1}{\pi}\int_{-1}^{1} e^{-a(\Delta x)\,\xi^2}\frac{d\xi}{\sqrt{1-\xi^2}}\right) + C \tag{15}$$

$$= A\left(1 - e^{-b(\Delta x)}e^{-a(\Delta x)/2}\,\mathcal{I}_0(a(\Delta x)/2)\right) + C\ , \tag{16}$$

where we have used the substitutions $a(\Delta x) := \frac{\Delta x^2}{2\eta^2 R^2} \times \left(\frac{1}{\cos^2\varepsilon_i} - \sin^2\varepsilon_i\right)$, $b(\Delta x) := \frac{\Delta x^2}{2\eta^2 R^2}\sin^2\varepsilon_i$, and $\xi := \sin\phi$. $\mathcal{I}_0$ is a modified Bessel function of the first kind. For $|\varepsilon| \ll \frac{\pi}{2}$, Eq. (16) can be approximated by

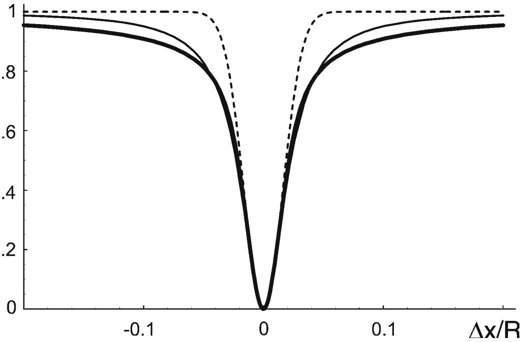

Fig. 8 Comparison of three functions for modelling translational IDFs. *Thick line*: $y = 1 - \exp\left(-\frac{\Delta x^2}{4\eta^2 R^2}\right)\mathcal{I}_0\left(\frac{\Delta x^2}{4\eta^2 R^2}\right)$; *dashed line*: $y = 1 - \exp\left(-\frac{\Delta x^2}{4\eta^2 R^2}\right)$; *thin line*: $y = -\left(1 + \frac{\Delta x^2}{4\eta^2 R^2}\right)^{-1}$; all plotted for $\eta = 0.65° = 0.011$ rad

$$\mathrm{IDF}(\Delta x) \approx A\left(1 - e^{-\frac{\Delta x^2}{4\eta^2 R^2}\left(1+2\varepsilon^2\right)}\mathcal{I}_0\left(\frac{\Delta x^2}{4\eta^2 R^2}\right)\right) + C\ , \tag{17}$$

which is also useful if one wishes to integrate over elevation ε.

Figure 8 shows $y = 1 - \exp\left(-\frac{\Delta x^2}{4\eta^2 R^2}\right)\mathcal{I}_0\left(\frac{\Delta x^2}{4\eta^2 R^2}\right)$ (thick line) for $\eta = 0.65° = 0.011$ rad. It approaches the asymptote $y = 1$ much more slowly than $y = 1 - \exp\left(-\frac{\Delta x^2}{4\eta^2 R^2}\right)$ (dashed line)—an effect of $\Delta x \propto \sin\phi$ and the integration over ϕ. Also shown is $1 - \left(1 + \frac{\Delta x^2}{4\eta^2 R^2}\right)^{-1}$ (thin line), which is our IDF model function from Eq. (5). This function is preferable for parameter fitting because $\exp\left(-\xi^2\right)\mathcal{I}_0\left(\xi^2\right)$ is numerically difficult to handle for $|\xi| \gg 1$. For $\left|\frac{\xi}{2\eta R}\right| \ll 1$, all three functions can be approximated by $\frac{\Delta x^2}{4\eta^2 R^2}$. Thus, the width parameter w in Eq. (6) can be estimated by

$$w^2 \approx \frac{4\eta^2 R^2}{1+2\varepsilon^2} = \frac{4\eta^2\langle 1/r\rangle^{-2}}{1+2\varepsilon^2}\ . \tag{18}$$

References

Baddeley R (1997) The correlational structure of natural images and the calibration of spatial representations. Cogn Sci 21:351–372

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

Carandini M, Heeger D, Movshon J (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. J Neurosci 17:8621–8644

Cartwright B, Collett T (1983) Landmark learning in bees: experiments and models. J Comp Physiol 15:521–543

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

Chahl J, Srinivasan M (1997) Reflective surfaces for panoramic imaging. Appl Optics 36:8275–8285

Collett T (1995) Making learning easy: the acquisition of visual information during the orientation flights of social wasps. J Comp Physiol A 177:737–747

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

Collett T, Zeil J (1997a) Flights of learning. Curr Dir Psych Sci 5: 149–155

Collett T, Zeil J (1997b) Selection and use of landmarks by insects. In: Lehrer M (ed) Orientation and communication in arthropods. Birkhäuser Verlag, Basel, pp 41–65

Collett T, Zeil J (1998) Places and landmarks: an arthropod perspective. In: Healy S (ed) Spatial representation in animals. Oxford University Press, Oxford, pp 18–53

Fleet D, Heeger D, Wagner H (1996) Modeling binocular neurons in primary visual cortex. In: Jenkin M, Harris L (eds) Computational and biological mechanisms of visual coding. Cambridge University Press, London, pp 103–130

Franz M, Mallot H (2000) Biomimetic robot navigation. Rob Auton Syst 30:133–153

Franz M, Schölkopf B, Mallot H, Bülthoff H (1998) Where did I take that snapshot? Scene based homing by image matching. Biol Cybern 79:191–202

Gaussier P, Joulain C, Banquet J, Leprêtre S, Revel A (2000) The visual homing problem: an example of robotics/biology cross fertilization. Rob Auton Syst 30:155–180

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

van Hateren H (1992) Theoretical predictions of spatiotemporal receptive fields of fly LMCs. J Comp Physiol A 171:157–170

van Hateren H (1993) Three modes of spatiotemporal processing by eyes. J Comp Physiol A 172:583–591

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

Lehrer M (1993) Why do bees turn back and look? J Comp Physiol A 172:549–563

Lehrer M, Collett T (1994) Approaching and departing bees learn different cues to the distance of a landmark. J Comp Physiol A 175: 171–177

Möller R (2002) Insects could exploit UV-green contrast for landmark navigation. J Theor Biol 214:619–663

Nicholson D, Judd S, Cartwritght B, Collett T (1999) Learning walks and landmark guidance in wood ants (*Formica rufa*). J Exp Biol 202:1831–1838

Osorio D, Vorobyev M (2005) Photoreceptor spectral sensitivities in terrestrial animals: adaptations for luminance and colour vision. Proc R Soc B 272:1745–1752

Srinivasan M, Laughlin S, Dubs A (1982) Predictive coding: a fresh view of inhibition in the retina. Proc R Soc Lond B 216:427–459

Szenher M (2005) Visual homing in natural environments. In: Nehmzow U, Melhuish C, Witkowski M (eds) Towards autonomous robotic systems (TAROS-05), p 221

Vardy A, Möller R (2005) Biologically plausible visual homing methods based on optical flow techniques. Connect Sci 17:47–89

Vladusich T, Hemmi J, Srinivasan M, Zeil J (2005) Interactions of visual odometry and landmark guidance during food search in honeybees. J Exp Biol 208:4123–4135

Voss R, Zeil J (1998) 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

Zanker J, Zeil J (2005) Movement-induced motion signal distributions in outdoor scenes. Netw Comp Neural Syst 16:357–376

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

Zeil J (1993b) 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

Zeil J, Wittmann D (1993) Landmark orientation during the approach to the nest in the stingless bee *Trigona (Tetragonisca) angustula* (Apidae, Meliponinae). Insectes Sociaux 40:381–389

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

Zeil J, Hofmann M, Chahl J (2003) Catchment areas of panoramic snapshots in outdoor scenes. J Opt Soc Am A 20(3):450–469