This is a repository copy of *Rotation invariant visual processing for spatial memory in insects*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/133447/

Version: Accepted Version

Article:

Rotation invariant visual processing for spatial memory in insects

Thomas Stone[1,4], **Michael Mangan**[2], **Antoine Wystrach**[3], **and Barbara Webb**[1]

[1]**School of Informatics, University of Edinburgh, 10 Crichton St, Edinburgh, EH8 9AB, UK**
[2]**Sheffield Robotics, Department of Computer Science, University of Sheffield, Regent Court, Sheffield, S1 4DP, UK**
[3]**CNRS, Universite Paul Sabatier, Toulouse, F-31062 cedex 09, France**
[4]**current affiliation: Google Inc**

ABSTRACT

Visual memory is crucial to navigation in many animals, including insects. Here we focus on the problem of visual homing, that is, using comparison of the view at a current location to a view stored at the home location to control movement towards home by a novel shortcut. Insects show several visual specialisations that appear advantageous for this task including almost panoramic field of view and ultraviolet light sensitivity that enhances the salience of the skyline. We discuss several proposals for subsequent processing of the image to obtain the required motion information, focussing on how each might deal with the problem of yaw rotation of the current view relative to the home view. Possible solutions include tagging of views with information from the celestial compass system, using multiple views pointing towards home, or rotation invariant encoding of the view. We illustrate briefly how a well known shape description method from computer vision, Zernike moments, could provide a compact and rotation invariant representation of sky shapes to enhance visual homing. We discuss the biological plausibility of this solution, and also a fourth strategy, based on observed behaviour of insects, that involves transfer of information from visual memory matching to the compass system.

1 INTRODUCTION

Spatial memory can be broadly defined as remembering information about a location that enables return to that location. For many animals, visual information is primary for this task. Potential advantages of visual cues are that they can be sensed passively over large distances, and that the structural information that can be extracted from an image tends to be a relatively stable and identifiable property of space (unlike, for example, olfactory cues). In order of increasing complexity, visual spatial navigation may involve simple recognition when a location is re-encountered, visual servoing towards a beacon (a clear visual cue at the desired location), visual homing that uses the configuration of surrounding cues, and at a higher level, linking together of multiple recognised locations, beacons or cue configurations into some form of map. Even the simplest function of recognising a location already raises significant problems for visual processing especially under natural conditions, such as dealing with varying illumination conditions and different viewpoints that will alter the image on the animal's retina. However, in this paper the focus will be on the problem of visual homing, which can be defined as the problem of determining, from comparison of the current view to a stored 'home' view, how to move so as to return to the home location (1; 2; 3).

This problem can be compared to the classic navigation process of triangulation, that is, using the bearing of landmarks to determine a position (4). In modern robotics, it forms a key part of visual simultaneous localisation and mapping (vSLAM): using the difference in bearing between identified landmarks in the current and previous views to estimate a movement vector. Theoretically, for movement on a plane, if three static landmarks are in view and correctly identified, the bearing difference between those landmarks is sufficient to uniquely identify the current location (4). In practice using more landmarks increases robustness, but there will always be some point beyond which disappearance of the familiar landmarks due to distance or occlusion will preclude successful homing. The region around the home view from which homing is possible is often referred to as the

catchment area, and its size is one way in which different homing algorithms can be compared. It is worth noting here that many algorithms for visual homing (including some of those discussed below) have relaxed the concept of 'landmark' to allow dense matching of image features, or even pixel-wise comparison of views, to form the basis of the calculation of the home direction.

In mammals, a classic demonstration of the ability to use surrounding visual cues to return via direct, novel paths to a desired location is the Morris water maze (5). In this widely used paradigm, a rodent is placed in a pool of opaque liquid and must swim until it discovers the location of a platform concealed under the water surface. On subsequent trials, starting from different locations, it finds the platform more rapidly. If conspicuous surrounding cues in the room are rotated, the search (for a platform that is now missing) will be centered on the position relative to these cues where the animal would experience the same view as it saw from the platform.

Many species of insects need to solve the equivalent task to pin-point the location of their nest (which can have an inconspicuous entrance) after foraging journeys. For example, ants displaced from their nest to novel locations within the catchment area are able to head directly back towards the nest, but are lost if moved beyond a visual boundary that occludes most of the features visible from the nest (6). Ants, bees or wasps trained with cues around their nest which are then translated while the animal forages will search for their nest in the appropriate location relative to the visual cues rather than at the actual nest position (for wasps, (7); for bees, (1); for ants, (8)). By manipulating landmark size and placement, it appears this search is principally based on obtaining a retinotopic match of the view, i.e. the bearing and size of landmarks (8; 1; 9; 10) , although in some situations, bees may also rely on the absolute distances of landmarks (11).

Perhaps more surprisingly, several species of insect that are not particularly known for their navigational prowess have been shown to be able to perform visual homing in a direct experimental analog of the Morris water maze. In this case, the aversive pool of water is replaced by an aversive heated floor, and the platform by a small area of cooler floor, which the animal can only discover by search. This paradigm was first used with cockroaches (12), then with crickets (13) and more recently with fruitflies (14; 15). After several trials, the animals move with increasing directness to the correct location, and if surrounding visual cues are rotated, they will search in the appropriate relative location.

A number of questions are still open regarding the mechanisms underlying this ability in insects (and for that matter, in vertebrates). What visual processing does the animal do to extract the information in a visual scene that will support homing, particularly under potentially variable conditions? What is the form of the stored memory? How is the memory and the current scene compared to recover a movement direction? What brain systems could support the required processing, memory storage, and direction recovery?

Insects have a number of specialisations in the periphery of their visual system that appear pertinent to the first question. Most insects have a near omnidirectional, but low-resolution, view of the world. Their apposition compound eyes are lattice-like structures comprised of many functionally independent yet physically interlocked light sensitive structures known as ommatidia, each complete with their own lens and photoreceptors. When stacked together these cone-shaped ommatidia form a bulbous outer contour producing an overall wide field of view but lacking focusing capacities (16). The combination of wide field of view with low-resolution has been demonstrated to be beneficial for visual navigation tasks providing sufficient information while preventing overfitting to details of the environment (17; 18). Sensitivity of insect photoreceptors to ultra-violet light is common (19) which has been shown to enhance skyline detection capacity (20; 21; 22; 23) underpinning skyline based navigation strategies (24; 25) and demonstrated to be advantageous for autonomous navigation (26; 27; 28; 21). Further, specially adapted photoreceptors in the dorsal rim (upwards facing) area of the eye show sensitivity to linearly polarised light allowing insects to derive their compass orientation from the pattern of polarised light present in the sky (29; 30; 31).

Clearly, the question of memory representation and recovery of a movement direction are linked, and a range of solutions have been developed (for a review see (3)). Most earlier theories focused on obtaining a compact representation, e.g., a one-dimensional binary description of surrounding landmarks and gaps between them (1), or a single vector average of the bearings of landmarks (32) or the intensity distribution (33). Reducing the memory to a vector allows very elegant calculation of the home vector through vector subtraction, but has found insufficient to account for insect homing in some of the experimental assays described above (34). More recent work has favoured approaches that assume storage of the raw, or minimally processed, full image (35; 36). In its most straightforward form, the agent stores an (omnidirectional) image at the home location, and use pixel-wise intensity comparison with the current image in a new location to obtain a difference measure. For natural

scenes, the difference tends to increase monotonically with distance from the home location, so a process of gradient descent on the difference surface will result in homing in both natural (35; 37; 38) and laboratory environments (34).

A problem that all of these approaches need to address is whether and how the animal can do a retinotopic image comparison when it is currently facing in a different direction. Most earlier algorithms incorporated the assumption that visual memories would be tagged by the compass system, and thus the first step to discover the heading direction would be to re-align the current image and the home image, either through mental rotation, or physical rotation to the desired compass angle (39). Yet, this cannot explain visual homing in the watermaze-inspired experiments described above where insects have no access to celestial compass cues and the arenas are specifically designed to provide no orientation information.

An alternative recent approach has been to use the fact that misaligned images are likely to mismatch as the basis for recovering the correct alignment. By rotating (it is usually assumed physically) while comparing images, a minima in the image difference will occur when the animal is aligned. In fact, this 'visual compass' approach has been used to argue that visual homing (or visual 'positional matching' (40)) can be subsumed under a general mechanism for route following by visual 'alignment matching' (41; 42; 43). If, rather than storing a single image at the home location, the animal stores a set of images from short displacements in different directions while facing the nest, then the best match it will obtain from a new location will be when aligned in the direction of the nearest stored image, i.e., facing the nest, the direction it needs to move (44; 6; 37). This explanation has several attractive features. The animal does not need to store any additional information about the images (the order in which they were experienced, their compass direction, etc.) and the matching process already provides the movement information, rather than requiring some additional mechanism to discover the local gradient of similarity. This lends itself to a neural network solution for memory storage in which all key images are input to a single associative net and the output is a general 'familiarity' measure (45), and such processing has been shown to be consistent with the circuit architecture of the insect mushroom body (46). This approach is also supported by some observations of insect behaviour. Central place foraging insects engage in learning flights (bees: (47); wasps: (48; 49; 50; 10; 51)) and walks (52; 53; 54) — on departure from the nest they rotate to view the nest from different positions. Ants also exhibit 'scanning' behaviours in which they rotate on the spot before moving off in a particularly direction, which suggests they may be attempting to physically align views (55; 56).

Overall, this algorithm explains well the behaviour of ants recapitulating their familiar route (45; 57; 58) and can also explain homing from novel locations assuming the use of nest-directed views stored during learning walks (44; 6). In the latter case, the recall of a useful nest-directed view requires that the stored views that lie between the agent and the nest match better than any other stored views. This can work over 10-15m in relatively open natural scenes (6; 37), however, it requires the agent to deal with small differences in mismatch and can be potentially brittle in other environments (57). In particular, during a scan, there will no longer be a single minima in the image difference function but rather multiple minima corresponding to each direction in which images were stored (c.f. figure 7 in (6)). This makes the task of finding the 'correct' minima more difficult, and more vulnerable to any factors that may distort the view or prevent regular, accurate alignment. Such factors include incomplete compensation for pitch (59) and roll (60) when moving over uneven terrain, as well as highly variable yaw relative to the direction of travel when dragging heavy food in a backwards motion (61). Ants in this situation may use forward peeks (62), but neither peeks nor scans are observed as often as has been assumed in most simulated evaluations of this algorithm. Finally, zero vector ants displaced from a familiar route do not orient as predicted by a visual alignment strategy (running parallel to the their familiar path) but instead move directly towards the familiar route, suggesting they can match and home to the nearest image while misaligned with its direction (63; 64; 57).

Here we present an alternative solution, exploiting an approach from computer vision for rotationally invariant recognition which is particularly suited to the problem insects need to solve in visual homing.

2 METHODS

2.1 Homing by matching sky shapes

As briefly described above (for more detail see (21; 22)), the visual system of the insect appears specialised to obtain an omnidirectional view that emphasises the skyline. If sky is segmented from ground and stored as a binary image, the shape of the sky forms a single blob or group of blobs with a

characteristic shape (Fig. 1a). Matching a segmented sky-image to another, despite yaw rotations, is essentially a form of shape recognition, a well established field (65). Various algorithms exist for matching shapes, despite variation in translation, scale and rotation. Common solutions include capturing properties of the contour (66; 67; 68) or the region itself, typically through use of geometric moments (69; 70).

The possibility of using a Fourier transform on a one dimensional skyline description to do more efficient visual homing was proposed in (71). This approach extracted a horizontal strip of greyscale values from the horizon pixels in an image from a panoramic camera mounted on a robot, and the array was converted to a frequency representation using a Fourier transform. However, rather than use this for rotation invariant image matching, their approach instead used phase components to recover an estimate of rotation combined with a frequency domain equivalent of the warping method (72) to determine the displacement from home. This was successfully used to home, with catchment areas being more pronounced and narrow for higher frequencies and wider for lower frequencies. Our first approach was to extend frequency coding to a two dimensional sky shape using Fourier descriptors (66). The outline of the shape is traced to form a complex signal representing the changing x and y coordinates (Fig. 1b), which then undergoes a Fourier transform to represent it in the frequency domain (Fig. 1c). Ignoring phase and storing the amplitude provides a shape signature that is invariant to yaw rotation. An additional benefit of this encoding is that the level of detail can be adjusted by only retaining the coefficients representing lower frequencies (Fig. 1d), allowing for easy scaling of storage and fast lookup of similar skylines. However this approach fails when the sky does not clearly consist of a single area, such as under foliage.

To mitigate these problems, a region-based shape recognition approach was applied. Zernike Moments (70) (ZM) are an orthogonal basis function set defined on the unit disk in polar co-ordinates, and have been widely used as global image descriptors. Like Fourier descriptors, the coefficients provide a frequency-based yaw-rotation-invariant representation of an image. As we are interested in performing homing based on the information in an image that consists of a 'sky disk' with a fixed centre and radius to the horizon (hence no translation and scaling) this representation seems especially suitable for our purposes.

2.2 Simulating visual homing

A virtual ant world previously made by (45) was used to create realistic panoramic greyscale views (Fig. 2a). Views were generated in a 10×10 m box at 2 cm intervals (Fig. 2b). Twenty evenly spaced locations were used as reference positions, to which all other locations within the box were compared. Two methods of comparison were used: a retinotopic pixel wise matching, and using amplitude coefficients of ZM of the panoramic images, remapped to a sky-centred disk (Fig. 2c).

The Euclidean distance between moments at each location was plotted on a 2D surface and gradient descent was used to determine which points would fall within the basin of attraction by iteratively moving a small step in the direction of steepest descent, where the local gradient was inferred using bicubic interpolation. If the starting location eventually led back to the nest it was deemed to be within the catchment area (Fig. 3). As we seek to assess the homing robustness of each encoding to changes in rotation it is the change in catchment area size that is most important rather than absolute performance per se, which could be strongly affected by multiple parameters not explored in this study.

3 RESULTS

Fig. 3 shows the catchment area obtained from comparison of images across the environment to a home image captured at one specific location. The images were either aligned with the reference image or taken at random headings. In Fig. 4 the results for 20 trials with different home image locations are shown.

As expected, these results clearly show that the use of random headings completely disrupt homing based on retinotopic matching (Fig. 3b) whereas the frequency representation is unaffected by the heading direction in which the picture are perceived (Fig. 3d). We also found that with a frequency representation, the catchment areas were considerably larger. The surprisingly poor performance obtained using retinotopic matching even with aligned images (Fig. 3a) is most likely due to a noisy image difference function, causing the agent to get stuck at local minima (similar results were reported by (73)). Storing multiple snapshots facing nestward from around the nest, which is one proposed solution to deal with random headings, will not fix this limitation. In contrast, the use of frequency representation seems to smooth out the image difference function (Fig. 3c), thus simplifying the problem of finding the global minima. We note that reducing the resolution of images

(using bilinear interpolation) did not improve results for retinotopic matching, but introducing blurring (low pass filtering) might reduce local noise in the gradient. Nevertheless, we expect parallax changes to cause landmarks across the horizon to be misaligned in moderately distant views, producing a poor retinotopic match (Fig. 5 c) and limiting the catchment area. It appears the frequency approach, especially at low frequencies is more robust (Fig. 5d).

4 DISCUSSION

Vision plays an important role in spatial navigation for many animals and the image processing that enables setting a course back towards a home location has been widely studied and modelled in insects. We have shown in previous work (21) that an image sensor based on the insect eye can be effective for recognition of previously experienced outdoor locations: combining ultraviolet light filtering and an omnidirectional view results in a distinctive, illumination invariant sky shapes. Here we show that using a frequency-based representation of the sky shape allows rotation invariant comparison of views in the context of visual homing. The increased catchment areas even for unrotated views suggests this encoding is also an effective way to capture global similarity in views in a way that alters smoothly with displacement. We note this contrasts with the results of frequency encoding of 1D horizon intensity used for visual homing in (71), who report a 20% reduction in the catchment area compared to pixel-based comparison. However, their approach was significantly different, as the aim was to show the efficiency of a frequency-based method for image-warping rather than image matching per se. Along with the different environment used (not representing a natural outdoor scene) there are multiple reasons why this different result could have been observed.

Using Zernike moments provides rotation invariance but is still vulnerable to factors that might cause the sky disk not to be centred. Specifically, in the insect (or robot) context, there might be additional variation in the image caused by pitch (59) and roll (60). We have explored the use of spherical harmonics (74) to address this issue. This method similarly uses the summation of basis functions to approximate a function corresponding to an image, but on the surface of a sphere rather than a disk. The results of a robot study applying this method for recognition of sky images are reported in (75). We note, however, that in contexts where variance is expected to be predominantly around the z-axis (yaw), e.g., for an animal moving on flat terrain or stabilising its head with respect to the horizon, Zernike moments will preserve more information about the shape with fewer components.

The use of ZM (or alternative basis function approaches) provides a potential advantage in terms of low dimensional storage and simple comparison of images using a small number of co-efficients. But this comes at the cost of more complex image processing. However, the long established analogy between basis function encoding and neural receptive fields (76; 77; 78) suggests that such processing may be biologically plausible. A mechanism to extract the amplitude at various frequencies would be through randomly aligned sinusoidal receptive fields, which is not inconsistent with known optics and early visual processing in insects, such as bar-sensitive, orientation-biased cells in insect lobula (79) or lateral triangle (80). Scene recognition in this manner would resemble the use of GIST descriptors used in computer vision, where the image is convolved with Gabor filters at different frequencies and orientations (81). However, we note that Zernike moments would be equivalent to very large receptive fields.

This problem has been approached in the other direction, i.e., from known properties of insect visual receptive fields to navigational behaviour (82; 83; 84). This work uses the receptive fields described in (80), or a generalised version of them, to encode omnidirectional views for an insect navigation task. However, the approach more closely resembles a phase filtering at a single spatial frequency and as such is shown to be effective for assessing rotational match, rather than encoding rotational invariance. The trend for many computer vision problems is towards using convolutional neural networks, where features are learnt rather than hand-designed, and thus potentially acquire invariance to many types of distortion including pose and lighting conditions (85). Previous results for this method in a navigation context can be seen in (86), where images depicting the same place but taken at different times are used to learn localisation-specific invariant features. A recent paper applies this approach in the ant navigation problem (87), using an autoencoder to compress 8100 pixel images taken in a forest ant habitat to 64 features that can still suffice for robust image comparison. As above, this work focuses on recovering information for rotational matching. It would be interesting to apply the approach to rotated and tilted views and to visualise how features learned by the network correspond to hand designed ones. At lower levels we certainly see the Gabor filter style tuning of units, which indicate some type of frequency approximation (88). It would be interesting to investigate whether any higher level features that emerge in such an approach approximate skyline contours.

As may be evident from the previous paragraph, the desirability of a rotation invariant encoding

depends on the assumed role of image comparison in navigation. In route following tasks the correct heading along a route can be recovered by rotational alignment with respect to a visual memory, in which case a rotation invariant encoding of view is not desired. This alignment strategy can also be combined with the use of multiple home facing images for direct control of movement in the homing task; this could be sufficient (6; 37) but may lack robustness compared to a rotation invariant approach. However, as discussed in the introduction, a purely alignment based approach does not capture the fact that displaced ants can return directly towards their familiar route by moving in a direction misaligned with stored views (57; 63; 64). Our frequency coding can provide an explanation for this behaviour as the gradient of familiarity, which slopes towards the closest familiar location, becomes accessible whatever the current body alignment.

Note that frequency coding could also be used to recover rotation information: by using the phase of the components, as was done in (71). Here we have ignored the phase information to obtain a rotation invariant image similarity measure that (within a catchment area) correlates with displacement distance. Choice of movement direction then requires estimation of the local gradient of familiarity, for which several possible solutions have been suggested. First, the familiarity of neighbor locations could be sampled, randomly or not, enabling different kinds of gradient descent (35). Local image warping has been suggested as a means to 'mentally' sample the consequences of moving to neighbouring locations (36). Another possibility is to adopt the klinokinesis strategy based on continuous oscillations that has been proposed for other animals negotiating a gradient with only a single point source measure, e.g., Drosophila larva in an odour gradient (89). This method, already demonstrated to be sufficient for route following in desert ant habitats (90), involves making regular lateral oscillations with an amplitude mediated by the change in stimulus intensity (here, in image similarity). Ants do appear to display lateral oscillation during visual guidance (91).

It also remains plausible that insects are able to implement either strategy, that is, rotation-based alignment or rotation-invariant image matching, depending on current conditions (57). An interesting possibility is that these mechanisms interact in visual homing, in a manner that would be consistent with the observation that ants dragging a large food item backward are observed to peek forward, followed by immediate course correction (62). That is, rather than homing by following the similarity gradient using any of the methods described in the previous paragraph, the animal may use a rotation invariant measure to monitor the visual familiarity of its current location and thus to control *when* it peeks; but still require physical alignment to obtain an estimate of the direction it needs to move. The direction may then be set relative to the celestial compass, allowing the animal to maintain the desired heading while facing another way when it returns to dragging the food (62). Whether or not the ant is still monitoring familiarity while facing backward could be readily tested in experiments that alter the view under these conditions.

As we have reviewed in this paper, visual homing in insects has been a rich field for investigation of the nature of biological image processing, and has been both an influence on, and influenced by, computer vision, especially in the context of robot navigation. A general conclusion that can be drawn is that biological 'image understanding' should be investigated within the context of behavioural control — what does the animal actually need to 'understand' in the image to solve the ecological task it is faced with? This approach has shifted the emphasis in insect navigational studies away from landmark extraction and towards holistic scene memory, leading to significant new insights into the mechanisms. A similar shift in perspective may be overdue for navigation in other animals.

AUTHOR CONTRIBUTIONS

Conception and analysis TS,MM,AW,BW. Simulation and figures TS. Writing BW,MM,AW.

ACKNOWLEDGMENTS

FUNDING

REFERENCES

[1] Cartwright B, Collett TS. Landmark learning in bees. Journal of comparative physiology. 1983;151(4):521–543.

[2] Vardy A, Möller R. Biologically plausible visual homing methods based on optical flow techniques. Connection Science. 2005;17(1-2):47–89.

[3] Zeil J. Visual homing: an insect perspective. Current opinion in neurobiology. 2012;22(2):285–293.

[4] Pierlot V, Van Droogenbroeck M. A new three object triangulation algorithm for mobile robot positioning. IEEE Transactions on Robotics. 2014;30(3):566–577.

[5] Morris R. Spatial localization does not require the presence of local cues. Learning and Motivation. 1981;12:239–260.

[6] Narendra A, Gourmaud S, Zeil J. Mapping the navigational knowledge of individually foraging ants, Myrmecia croslandi. Proceedings of the Royal Society of London B: Biological Sciences. 2013;280(1765):20130683.

[7] Tinbergen N. The study of instinct. Clarendon Press/Oxford University Press; 1951.

[8] Wehner R, Räber F. Visual spatial memory in desert ants, Cataglyphis bicolor (Hymenoptera: Formicidae). Cellular and Molecular Life Sciences. 1979;35(12):1569–1571.

[9] Brünnert U, Kelber A, Zeil J. Ground-nesting bees determine the location of their nest relative to a landmark by other than angular size cues. Journal of Comparative Physiology A. 1994;175(3):363–369.

[10] Zeil J. Orientation flights of solitary wasps (Cerceris; Sphecidae; Hymenoptera). Journal of Comparative Physiology A. 1993;172(2):207–222.

[11] Lehrer M, Collett T. Approaching and departing bees learn different cues to the distance of a landmark. Journal of Comparative Physiology A. 1994;175(2):171–177.

[12] Mizunami M, Weibrecht J, Strausfeld N. Mushroom bodies of the cockroach: their participation in place memory. Journal of Comparative Neurology. 1998;402(4):520–537.

[13] Wessnitzer J, Mangan M, Webb B. Place memory in crickets. Proceedings of the Royal Society of London B: Biological Sciences. 2008;275(1637):915–921.

[14] Foucaud J, Burns JG, Mery F. Use of spatial information and search strategies in a water maze analog in Drosophila melanogaster. PLoS One. 2010;5(12):e15231.

[15] Ofstad TA, Zuker CS, Reiser MB. Visual place learning in Drosophila melanogaster. Nature. 2011;474(7350):204–207.

[16] Land MF, Nilsson DE. Animal eyes. Oxford University Press; 2012.

[17] Milford M. Vision-based place recognition: how low can you go? The International Journal of Robotics Research. 2013;32(7):766–789.

[18] Wystrach A, Dewar A, Philippides A, Graham P. How do field of view and resolution affect the information content of panoramic scenes for visual navigation? A computational investigation. Journal of Comparative Physiology A. 2016;202(2):87–95.

[19] Menzel R, Ventura DF, Hertel H, De Souza J, Greggers U. Spectral sensitivity of photoreceptors in insect compound eyes: comparison of species and methods. Journal of Comparative Physiology A. 1986;158(2):165–177.

[20] Möller R. Insects could exploit UV–green contrast for landmark navigation. Journal of theoretical biology. 2002;214(4):619–631.

[21] Stone T, Mangan M, Ardin P, Webb B. Sky segmentation with ultraviolet images can be used for navigation. In: Proceedings of Robotics: Science and Systems. Berkeley, USA; 2014. .

[22] Differt D, Möller R. Insect models of illumination-invariant skyline extraction from UV and green channels. Journal of Theoretical Biology. 2015;380:444–462.

[23] Differt D, Möller R. Spectral skyline separation: Extended landmark databases and panoramic imaging. Sensors. 2016;16(10):1614.

[24] Graham P, Cheng K. Ants use the panoramic skyline as a visual cue during navigation. Current Biology. 2009;19(20):R935–R937.

[25] Schultheiss P, Wystrach A, Schwarz S, Tack A, Delor J, Nooten SS, et al. Crucial role of ultraviolet light for desert ants in determining direction from the terrestrial panorama. Animal Behaviour. 2016;115:19–28.

[26] Stein F, Medioni G. Map-based localization using the panoramic horizon. IEEE Transactions on Robotics and Automation. 1995;11(6):892–896.

[27] Meguro J, Murata T, Amano Y, Hasizume T, Takiguchi J. Development of a positioning technique for an urban area using omnidirectional infrared camera and aerial survey data. Advanced Robotics. 2008;22(6-7):731–747.

[28] Ramalingam S, Bouaziz S, Sturm P, Brand M. Geolocalization using skylines from omni-images. In: Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on. IEEE; 2009. p. 23–30.

[29] Wehner R, Labhart T. Polarization vision. Invertebrate Vision. 2006;p. 291–348.

[30] Homberg U, Heinze S, Pfeiffer K, Kinoshita M, El Jundi B. Central neural coding of sky polarization in insects. Philosophical Transactions of the Royal Society B: Biological Sciences. 2011;366(1565):680–687.

[31] Heinze S. Unraveling the neural basis of insect navigation. Current opinion in insect science. 2017;.

[32] Lambrinos D, Kobayashi H, Pfeifer R, Maris M, Labhart T, Wehner R. An autonomous agent navigating with a polarized light compass. Adaptive behavior. 1997;6(1):131–161.

[33] Hafner VV. Adaptive homing—robotic exploration tours. Adaptive Behavior. 2001;9(3-4):131–141.

[34] Mangan M, Webb B. Modelling place memory in crickets. Biological Cybernetics. 2009;101(4):307.

[35] Zeil J, Hofmann MI, Chahl JS. Catchment areas of panoramic snapshots in outdoor scenes. JOSA A. 2003;20(3):450–469.

[36] Möller R. Local visual homing by warping of two-dimensional images. Robotics and Autonomous Systems. 2009;57(1):87–101.

[37] Stürzl W, Grixa I, Mair E, Narendra A, Zeil J. Three-dimensional models of natural environments and the mapping of navigational information. Journal of Comparative Physiology A. 2015;201(6):563–584.

[38] Murray T, Zeil J. Quantifying navigational information: The catchment volumes of panoramic snapshots in outdoor scenes. PloS One. 2017;12(10):e0187226.

[39] Möller R. A model of ant navigation based on visual prediction. Journal of theoretical biology. 2012;305:118–130.

[40] Collett M, Chittka L, Collett TS. Spatial Memory in Insect Navigation. Current Biology. 2013;23(17):R789–R800.

[41] Baddeley B, Graham P, Philippides A, Husbands P. Holistic visual encoding of ant-like routes: Navigation without waypoints. Adaptive Behavior. 2011;19(1):3–15.

[42] Wystrach A, Mangan M, Philippides A, Graham P. Snapshots in ants? New interpretations of paradigmatic experiments. Journal of Experimental Biology. 2013;216(10):1766–1770.

[43] Dewar ADM, Philippides A, Graham P. What is the relationship between visual environment and the form of ant learning-walks? An in silico investigation of insect navigation. Adaptive Behavior. 2014 feb;22(3):163–179.

[44] Graham P, Philippides A, Baddeley B. Animal cognition: multi-modal interactions in ant learning. Current Biology. 2010;20(15):R639–R640.

[45] Baddeley B, Graham P, Husbands P, Philippides A. A model of ant route navigation driven by scene familiarity. PLoS Computational Biology. 2012;8(1):e1002336.

[46] Ardin P, Peng F, Mangan M, Lagogiannis K, Webb B. Using an insect mushroom body circuit to encode route memory in complex natural environments. PLoS Computational Biology. 2016;12(2):e1004683.

[47] Philippides A, de Ibarra NH, Riabinina O, Collett TS. Bumblebee calligraphy: the design and control of flight motifs in the learning and return flights of Bombus terrestris. Journal of Experimental Biology. 2013;216(6):1093–1104.

[48] Collett T, Lehrer M. Looking and learning: a spatial pattern in the orientation flight of the wasp Vespula vulgaris. Proceedings of the Royal Society of London B: Biological Sciences. 1993;252(1334):129–134.

[49] Lehrer M. Why do bees turn back and look? Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology. 1993;172(5):549–563.

[50] Collett T. Making learning easy: the acquisition of visual information during the orientation flights of social wasps. Journal of Comparative Physiology A. 1995;177(6):737–747.

[51] Stürzl W, Zeil J, Boeddeker N, Hemmi JM. How wasps acquire and use views for homing. Current Biology. 2016;26(4):470–482.

[52] Nicholson D, Judd S, Cartwright B, Collett T. Learning walks and landmark guidance in wood ants (Formica rufa). Journal of Experimental Biology. 1999;202(13):1831–1838.

[53] Müller M, Wehner R. Path integration provides a scaffold for landmark learning in desert ants. Current Biology. 2010;20(15):1368–1371.

[54] Fleischmann PN, Christian M, Müller VL, Rössler W, Wehner R. Ontogeny of learning walks and

the acquisition of landmark information in desert ants, Cataglyphis fortis. Journal of Experimental Biology. 2016;219(19):3137–3145.

[55] Wystrach A, Philippides A, Aurejac A, Cheng K, Graham P. Visual scanning behaviours and their role in the navigation of the Australian desert ant Melophorus bagoti. Journal of Comparative Physiology A. 2014;200(7):615–626.

[56] Zeil J, Narendra A, Stürzl W. Looking and homing: how displaced ants decide where to go. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2014;369(1636):20130034.

[57] Wystrach A, Beugnon G, Cheng K. Ants might use different view-matching strategies on and off the route. Journal of Experimental Biology. 2012;215(1):44–55.

[58] Wystrach A, Schwarz S, Schultheiss P, Beugnon G, Cheng K. Views, landmarks, and routes: how do desert ants negotiate an obstacle course? Journal of Comparative Physiology A. 2011;197(2):167–179.

[59] Ardin P, Mangan M, Wystrach A, Webb B. How variation in head pitch could affect image matching algorithms for ant navigation. Journal of Comparative Physiology A. 2015;201(6):585–597.

[60] Raderschall CA, Narendra A, Zeil J. Head roll stabilisation in the nocturnal bull ant Myrmecia pyriformis: implications for visual navigation. Journal of Experimental Biology. 2016;219(10):1449–1457.

[61] Ardin PB, Mangan M, Webb B. Ant Homing Ability Is Not Diminished When Traveling Backwards. Frontiers in Behavioral Neuroscience. 2016;10:69.

[62] Schwarz S, Mangan M, Zeil J, Webb B, Wystrach A. How Ants Use Vision When Homing Backward. Current Biology. 2017;27(3).

[63] Kohler M, Wehner R. Idiosyncratic route-based memories in desert ants, Melophorus bagoti: how do they interact with path-integration vectors? Neurobiology of Learning and Memory. 2005;83(1):1–12.

[64] Narendra A. Homing strategies of the Australian desert ant Melophorus bagoti II. Interaction of the path integrator with visual cue information. Journal of Experimental Biology. 2007;210(10):1804–1812.

[65] Latecki LJ, Lakamper R, Eckhardt T. Shape descriptors for non-rigid shapes with a single closed contour. In: IEEE Conference on Computer Vision and Pattern Recognition, 2000. Proceedings.. vol. 1. IEEE; 2000. p. 424–429.

[66] Persoon E, Fu KS. Shape discrimination using Fourier descriptors. IEEE Transactions on Systems, Man, and Cybernetics. 1977;7(3):170–179.

[67] Kaneko T, Okudaira M. Encoding of arbitrary curves based on the chain code representation. IEEE Transactions on Communications. 1985;33(7):697–707.

[68] Mokhtarian F, Abbasi S, Kittler J, et al. Efficient and robust retrieval by shape content through curvature scale space. Series on Software Engineering and Knowledge Engineering. 1997;8:51–58.

[69] Hu MK. Visual pattern recognition by moment invariants. IRE Transactions on Information Theory. 1962;8(2):179–187.

[70] Khotanzad A, Hong YH. Invariant image recognition by Zernike moments. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1990;12(5):489–497.

[71] Stürzl W, Mallot HA. Efficient visual homing based on Fourier transformed panoramic images. Robotics and Autonomous Systems. 2006;54(4):300–313.

[72] Franz MO, Schölkopf B, Mallot HA, Bülthoff HH. Where did I take that snapshot? Scene-based homing by image matching. Biological Cybernetics. 1998;79(3):191–202.

[73] Mangan M. Visual homing in field crickets and desert ants: A comparative behavioural and modelling study. The University of Edinburgh; 2011.

[74] Kazhdan M, Funkhouser T, Rusinkiewicz S. Rotation invariant spherical harmonic representation of 3D shape descriptors. In: Symposium on Geometry Processing. vol. 6; 2003. p. 156–164.

[75] Stone T, Differt D, Milford M, Webb B. Skyline-based localisation for aggressively manoeuvring robots using UV sensors and spherical harmonics. In: 2016 IEEE International Conference on Robotics and Automation (ICRA). IEEE; 2016. p. 5615–5622.

[76] Maffei L, Fiorentini A. The visual cortex as a spatial frequency analyser. Vision Research. 1973;13(7):1255–1267.

[77] Daugman JG. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. JOSA A. 1985;2(7):1160–1169.

[78] Bell AJ, Sejnowski TJ. The "independent components" of natural scenes are edge filters. Vision

research. 1997;37(23):3327–3338.

[79] O'Carroll D. Feature-detecting neurons in dragonflies. Nature. 1993;362:541–543.

[80] Seelig JD, Jayaraman V. Feature detection and orientation tuning in the Drosophila central complex. Nature. 2013;503(7475):262–266.

[81] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope. International journal of Computer Vision. 2001;42(3):145–175.

[82] Wystrach A, Dewar AD, Graham P. Insect vision: emergence of pattern recognition from coarse encoding. Current Biology. 2014;24(2):R78–R80.

[83] Dewar AD, Wystrach A, Graham P, Philippides A. Navigation-specific neural coding in the visual system of Drosophila. Biosystems. 2015;136:120–127.

[84] Dewar AD, Wystrach A, Philippides A, Graham P. Neural coding in the visual system of Drosophila melanogaster: How do small neural populations support visually guided behaviours? PLoS Computational Biology. 2017;13(10):e1005735.

[85] LeCun Y, Huang FJ, Bottou L. Learning methods for generic object recognition with invariance to pose and lighting. In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. vol. 2. IEEE; 2004. p. 97–104.

[86] Arandjelović R, Gronat P, Torii A, Pajdla T, Sivic J. NetVLAD: CNN architecture for weakly supervised place recognition. arXiv preprint arXiv:151107247. 2015;.

[87] Walker C, Graham P, Philippides A. Using deep autoencoders to investigate image matching in visual navigation. In: Conference on Biomimetic and Biohybrid Systems. Springer; 2017. p. 465–474.

[88] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems; 2012. p. 1097–1105.

[89] Wystrach A, Lagogiannis K, Webb B. Continuous lateral oscillations as a core mechanism for taxis in Drosophila larvae. Elife. 2016;5.

[90] Kodzhabashev A, Mangan M. Route following without scanning. In: Conference on Biomimetic and Biohybrid Systems. Springer; 2015. p. 199–210.

[91] Lent DD, Graham P, Collett TS. Phase-dependent visual control of the zigzag paths of navigating wood ants. Current Biology. 2013;23(23):2393–2399.

5 FIGURE CAPTIONS

Figure 1. Fourier descriptors to match skylines. (a) Segmented sky of an urban scene can be thought of as a closed shape. **(b)** x and y coordinates of the shape edge create a continuous signal. **(c)** The frequency components of this signal can describe the outline of the shape up to an arbitrary amount of detail. **(d)** Reconstruction of sky shape using only the first ten Fourier descriptors shows the rough shape signature, allowing high compression of skylines.

Figure 2. Simulated ant world. (a) A simulated world was populated with random vegetation at different densities creating an ant-like environment in which 360° panoramic views were generated every 2 cm in the test area (light orange) to test homing models. **(b)** Panoramic view from world. **(c)** Images were wrapped before converting to a Zernike moments representation.

Figure 3. Frequency based homing, catchment areas. All plots show similarity to a panoramic image from the same reference location measured at 2 cm intervals in the virtual environment. In *(a)* and *(c)* the images are taken facing north, in *(b)* and *(d)* with random headings. In *(a)* and *(b)* comparison is pixelwise, in *(c)* and *(d)* by taking the Euclidean distance between vectors made up by ZM amplitude coefficients. Blue contour lines indicate the border of the catchment area.

Figure 4. Homing performance. $N = 20$. Retinotopy provides guidance but only when aligned with memory. Frequency encoding performs well, regardless of orientation.

Figure 5. Explanation of poor performance of retinotopic matches between allocentrically aligned images. *(a,b)* Two panoramic views taken 3.6 metres apart in the virtual ant environment shown in figure 2 a. *(c)* Absolute image difference between the two images. Even a relatively small translational displacement in an open area with only distant landmarks can cause the majority of matching skyline features to misalign. *(d)* Conversely, normalised Zernike amplitude coefficients of each image remain relatively similar.

(a) Segmented sky shape
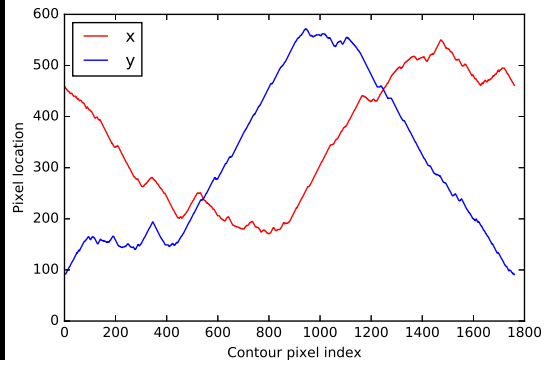
(b) Pixel coordinates
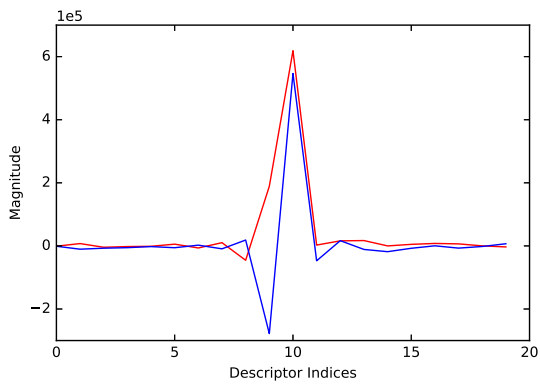
(c) Frequency components

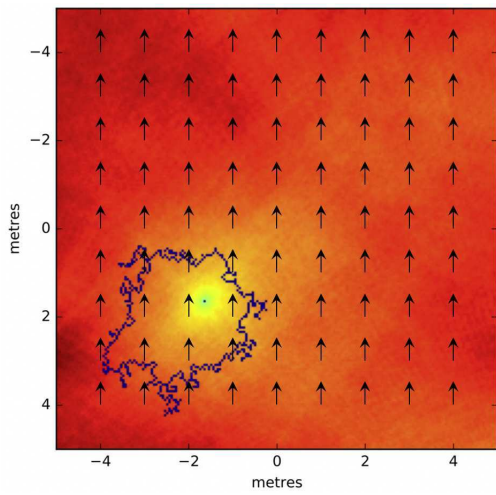**(d)** Reconstructed sky using 10 coefficients
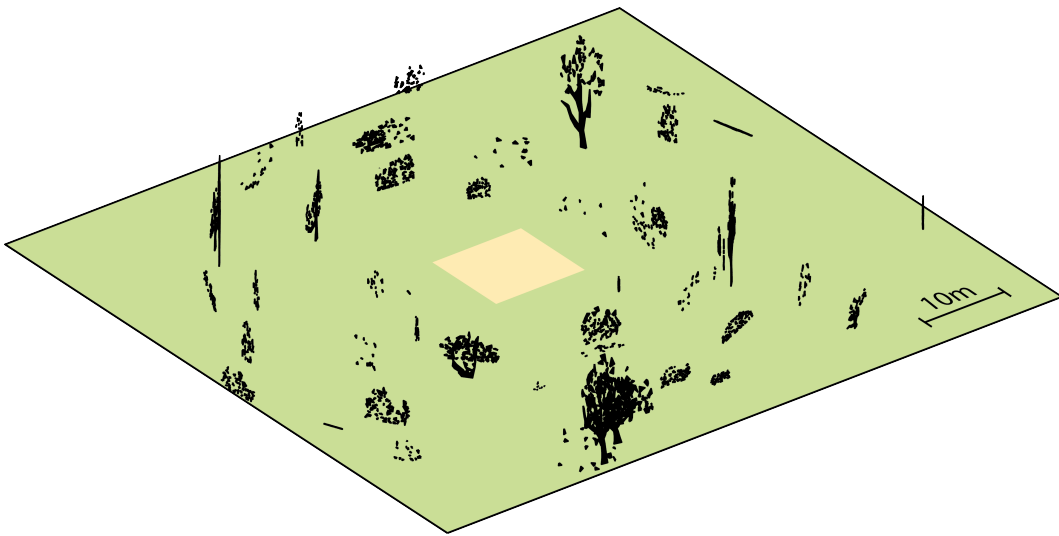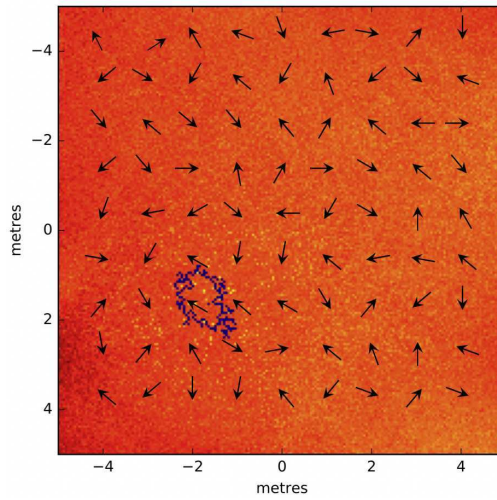
(a) 3D Computer model based on ant environment.
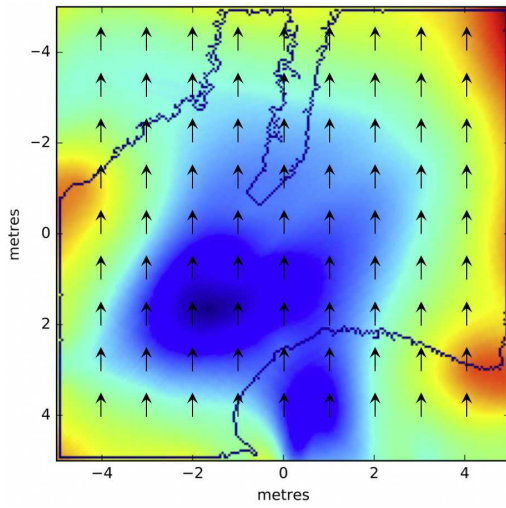

(b) Panoramic view in ant world

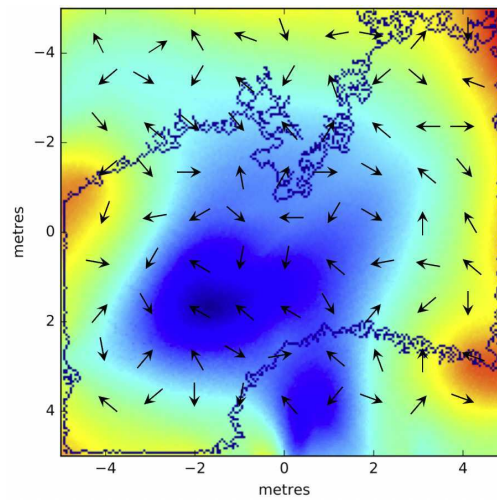

(c) Wrapped panorama.

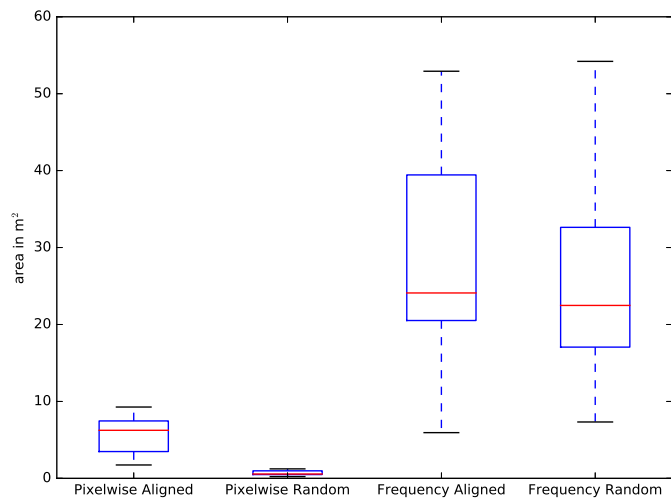**(a)** Retinotopic - aligned

(b) Retinotopic - random

(c) Frequency - aligned

(d) Frequency - random

Catchment area of gradient descent using different distance measures
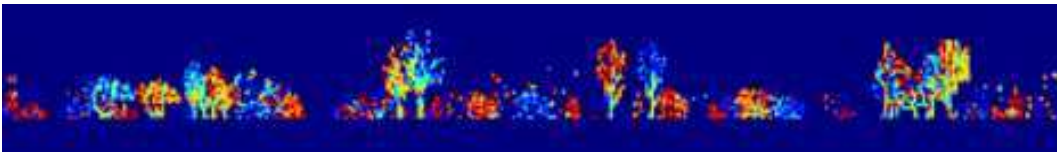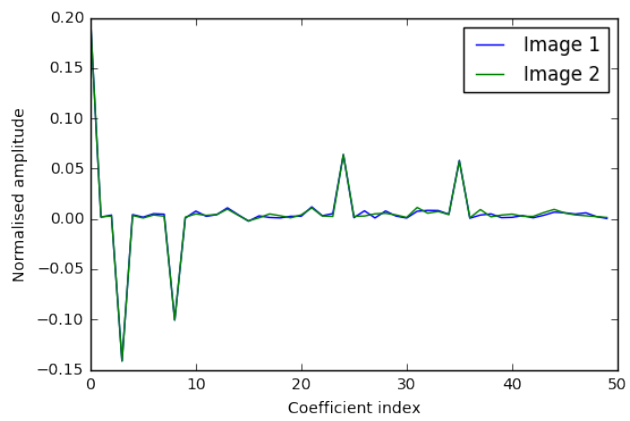
(a) North-aligned image 1

(b) North-aligned image 2 (displaced by 3.6 metres)

(c) Difference between images (red indicates the maximum possible difference)

(d) First 50 coefficients of Zernike transform amplitude of both images