# Ground-Nesting Insects Could Use Visual Tracking for Monitoring Nest Position during Learning Flights

Nermin Samet[1], Jochen Zeil[2], Elmar Mair[4],
Norbert Boeddeker[3], and Wolfgang Stürzl[4]

[1] Department of Computer Engineering, Bilkent University, Turkey
[2] Research School of Biology, The Australian National University
[3] Department of Cognitive Neuroscience, Bielefeld University, Germany
[4] Institute of Robotics and Mechatronics, German Aerospace Center (DLR)

**Abstract.** Ants, bees and wasps are central place foragers. They leave their nests to forage and routinely return to their home-base. Most are guided by memories of the visual panorama and the visual appearance of the local nest environment when pinpointing their nest. These memories are acquired during highly structured learning walks or flights that are performed when leaving the nest for the first time or whenever the insects had difficulties finding the nest during their previous return. Ground-nesting bees and wasps perform such learning flights daily when they depart for the first time. During these flights, the insects turn back to face the nest entrance and subsequently back away from the nest while flying along ever increasing arcs that are centred on the nest. Flying along these arcs, the insects counter-turn in such a way that the nest entrance is always seen in the frontal visual field at slightly lateral positions. Here we asked how the insects may achieve keeping track of the nest entrance location given that it is a small, inconspicuous hole in the ground, surrounded by complex natural structures that undergo unpredictable perspective transformations as the insect pivots around the area and gains distance from it. We reconstructed the natural visual scene experienced by wasps and bees during their learning flights and applied a number of template-based tracking methods to these image sequences. We find that tracking with a fixed template fails very quickly in the course of a learning flight, but that continuously updating the template allowed us to reliably estimate nest direction in reconstructed image sequences. This is true even for later sections of learning flights when the insects are so far away from the nest that they cannot resolve the nest entrance as a visual feature. We discuss why visual goal-anchoring is likely to be important during the acquisition of visual-spatial memories and describe experiments to test whether insects indeed update nest-related templates during their learning flights.

**Keywords:** Insect navigation, visual tracking, learning flights, homing.

# 1   Introduction

Many insects, in particular ants, bees and wasps, are competent navigators and are known to heavily rely on vision to memorize places and routes (reviewed in [1,2]). The landmark panorama [3,4], the sun, the pattern of polarized skylight [5] and even the Milky Way [6] provide them with an external compass reference. For visual homing, insects acquire scene memories at their nest or at newly discovered feeding sites during highly structured learning flights or learning walks [7,8]. There is evidence that insects acquire locale memory during this learning process, in particular visual information that allows them to subsequently return to the goal [9,10,1]. A common feature of learning flights and learning walks is the way, in which the insects carefully control where they see the nest entrance as they pivot around and back away from the nest (see references in [1]). This appears to be a crucial element of acquiring views for homing, most probably because it allows insects to memorize the visual panorama always in association with the goal direction. In ground-nesting wasps, the nest entrance position is clearly under visual control, because the insects track a small patterned disk when it is moved away from the nest entrance [10]. In contrast, ants use path integration information when they turn back and look across the nest during their learning walks [8].

For flying insects, there are several options to keep track of the nest entrance position: they may continuously update their position relative to the nest, based on estimates of their own movements, that is, use path integration information, like ants do [8]. However, it is not clear how accurately flying insects could employ path integration in this task, considering that they operate in three dimensions and at high speed. Learning flights have also been considered and modelled as a procedure akin to SLAM (simultaneous localization and mapping) [11], which, however, would appear to be quite computationally demanding.

Our aim here is to explore the possibility that insects visually track the nest entrance and its immediate visual environment (see [9,10,12]). We pay particular attention to the problem of tracking a location in the natural environment of ground-nesting wasps and bees that undergoes complex visual transformations as the insects pivot around it and gain distance from it. We will show that the nest direction can be estimated by means of a dynamic template update procedure, even in situations in which the nest entrance itself cannot be resolved due the limited resolution of the insect eye.

We proceeded in two steps: First, we used an existing data bank of reconstructed views during the learning flights of bees and wasps (e.g. [13,14,15]) and remapped these images according to an equidistant fisheye projection. In the second step, we developed and tested seven different template tracking methods and analysed how well each kept track of the nest location in these image sequences. We find that tracking is robust, provided templates are dynamically updated and suggest that the comparison of nest-registered snapshots with what a homing insect currently sees can in principle be used to predict the movement direction required to reach the goal position.
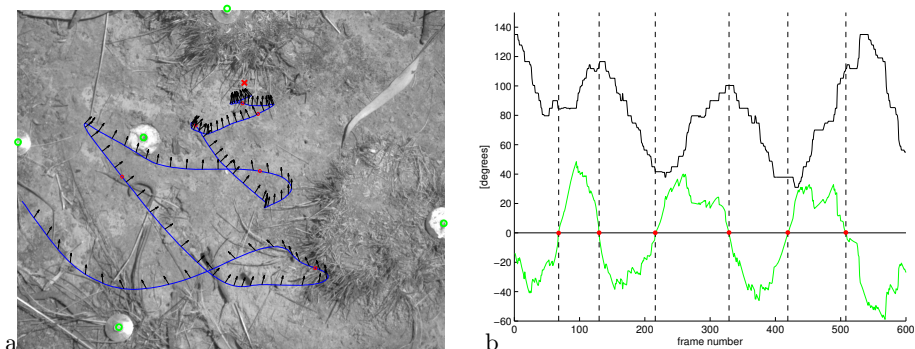
**Fig. 1.** a) The trajectory (blue curve) of a ground-nesting bee's learning flight, over-laid on a frame recorded with the downward looking high-speed camera. Black arrows indicate head orientation, plotted every 5th frame, i.e. every 20 ms; red dots highlight positions where the nest is "head-on", i.e. at 0° azimuth in the bee's visual field). The red cross indicates the nest position. The green dots highlight metal pins in the ground that were used to determine the transformation between the high-speed stereo system and the computer model of the environment. b) Head orientation (black curve) and azimuth angle of the nest in the visual field (green). The red dots (as in a) and the dashed vertical lines indicate 0° nest azimuth angle.

## 2    Reconstruction of the Visual Input Perceived by Insects

For this study we used existing learning flight data of three ground-nesting wasps (*Cerceris australis*) and of one ground-nesting bee (species not identified). In the following we describe briefly our approach for reconstructing the visual input the insects perceived during these flights, which forms the basis for the evaluation of our nest tracking hypothesis.

### 2.1    Recording and Path Reconstruction of Learning Flights

Wasps and bees were filmed with high-speed stereo cameras at 250 fps. The angle between the two cameras was about 90°: while one camera was viewing the recording area from above, the second camera was positioned close to the ground, viewing the scene from the side. The 3D flight path and the head yaw orientation were determined frame-by-frame using custom-made software (see [13,16] for details).

As example we show in figure 1 the flight path and the head orientation for the learning flight of a ground-nest bee. Interestingly, the nest is rarely seen directly in front (at 0° azimuth in the visual field) but is kept most of the time at 20°-50° in the lateral visual field. Furthermore, head orientation does not change smoothly but abruptly. Between these fast turns, which are called "saccades", the head orientation is kept virtually constant.
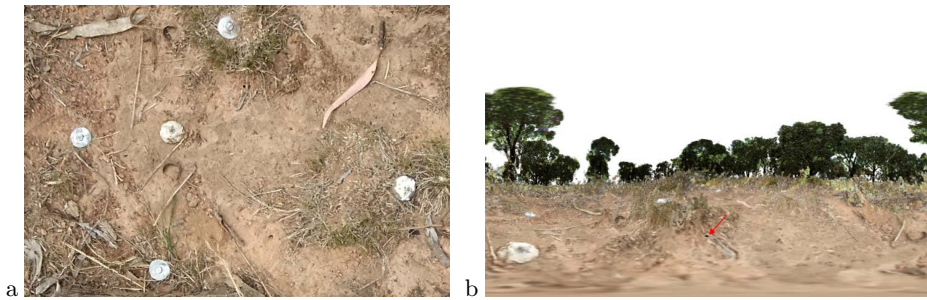
**Fig. 2.** a) Rendered image showing the part of scene that is seen by the downward looking high-speed camera (see figure 1 a). b) Panoramic image (covering $360° \times 180°$ in equi-rectangular projection) rendered with 2 pixels/degree at a position of the learning flight (frame 216, see figure 1 b), where the bee faces the nest entrance, which is highlighted by a red arrow.

## 2.2   Generating Computer Models of the Local Environment

We generated 3D models of the environments in which we had recorded the learning flights of insects. For the bee environment (Mount Majura Nature Park, Canberra, Australia; flight recorded January 2012) we used a combination of 3D reconstruction tools [14]. An area around the nest entrance covering about $1\,m^2$ was reconstructed using bundle adjustment with sub-sequential dense pairwise stereo processing on a set of 40 photos that were taken with a 10 mega-pixel camera with "locked focus" setting. This high resolution local model was complemented with point clouds acquired with a laser-range finder colour camera combination (Z+F IMAGER 5006). The different parts of the model were aligned using a set of large metal pins (some of them are visible in figure 1 a). Examples of rendered images are shown in figure 2.

The wasp learning flights were recorded in 2006, when we had no reconstruction equipment available. We thus modified our approach for generating a 3D model of that environment. Local models covering the nest entrances and their neighborhood (in an area of about $35\,cm \times 45\,cm$) were created by manually identifying about 270 corresponding points in stereo images and determining their 3D coordinates. The resulting 3D points were then triangulated and a video frame of the downward looking camera was mapped to the resulting wire-frame. In addition a 3D model of the surrounding scene was generated using the Z+F IMAGER 5006 in 2011. While the fine structure of the local scene had changed noticeably, the overall depth structure of scene remained, even after 5 years, basically the same. The model acquired with the laser-range finder and the local model were registered using nails hammered into the ground (which had remained there for more than 5 years).

## 2.3   Rendering Insect Views

After determining the transformation between the stereo camera system and the computer model by means of markers in the ground, images can be rendered

along insect flight paths. We use six virtual cameras oriented along the six normal vectors of a cube to cover the large field of view of insect eyes. The rendered views were converted to grey-level images and then remapped to a single planar image according to an equidistant fisheye projection ("f-theta lens") with radial resolution of either 1°/pixel or 2°/pixel (see examples in figure 3). The optical axis of this virtual fisheye lens was pitched by $-45°$ with respect to the horizontal, which helped to reduce distortions in the image regions relevant for tracking. For the learning flight of the ground-nesting bee we used in addition also a pixel mapping that resembles the spatial sampling of the eyes of a worker honeybee [17]. Although not necessarily an exact eye model for that particular species, this representation allowed us to study the "hand-over" of the tracked region between the two eyes (see figure 6).

## 3    Methods for Nest Tracking

In this section we introduce our methods for testing the nest tracking hypothesis. We use a comparatively simple template-based approach, which we consider biologically plausible because it shares similarities with current view-based models of insect navigation [2,18].

Suppose that we have the image sequence of a learning flight $\{I_n\}$ where $n = 1, 2, 3, ...$ is the frame number. We attempt to track the nest through the learning flight sequence by extracting an initial square template $T$ from the first frame $I_1$ with the nest in the centre. We then search for the region in the following frame(s) that matches the current template best. Assuming a reasonably high frame rate, the nest cannot change its location much from frame to frame and the search can be restricted to an area centred around the best matching region in the previous frame.

We used two different similarity functions for determining the best match between the template and image regions within the search area. We select the position $(x_n^{\text{opt}}, y_n^{\text{opt}})$ either of the minimum of the *Sum of Squared Differences* (SSD),

$$\text{SSD}_n(x, y) = \sum_{x', y'} \left( T_n(x', y') - I_n(x + x', y + y') \right)^2 \ ,$$

or of the maximum of the *Normalized Correlation Coefficient* (NCC),

$$\text{NCC}_n(x, y) = \frac{\sum_{x', y'} (T_n(x', y') - \bar{T}_n) \cdot (I_n(x + x', y + y') - \bar{I}_n(x, y))}{\sqrt{\sum_{x', y'} (T_n(x', y') - \bar{T}_n)^2 \cdot \sum_{x', y'} (I_n(x + x', y + y') - \bar{I}_n(x, y))^2}} \ .$$

With $n$ indicating the frame number and $\bar{T}_n$, $\bar{I}_n(x, y)$ the mean pixel value of template and image region.

We tested seven variants of template-based methods (M1-M7) for nest tracking that differ in the way the template is updated:

**M1: No Template Update.** We use the template extracted from the first frame of the learning flight to find matches in all the subsequent frames: $T_{n+1} = T_1$ for all $n \geq 1$.

**M2: Template Update (in each frame).** M1 is likely to fail when the insect's distance to the nest entrance increases. As a simple solution to this problem we update the template continuously in order to keep the similarity high between the template and the current nest region: $T_{n+1} = \mathrm{crop}(I_n, x_n^{\mathrm{opt}}, y_n^{\mathrm{opt}})$ for all $n \geq 1$, where 'crop$(I, x, y)$' describes the extraction of the template from image $I$ at position $(x, y)$.

**M3: Template Update on Rotated Image.** Due to the rapid changes in head orientation (see figure 1 b), even a template that is updated in each frame can give a poor match after a saccade and can cause significant deviation of the best matching image region from the image part containing the nest. However, saccadic head movements are initiated by the insects themselves and pure rotations, so that the image shifts they generate are predictable and can be accounted for by, for instance, an efference copy command. M3 is an extension of M2: It compensates for turns by counter-rotating both the previously updated template (as described for M2) and the centre of the search region.

**M4: Template Update on Rotated Image with Contour Detection.** A problem with M2 and M3 is that the area of best match tends to drift away from the nest region due to the accumulation of small errors with each update. In order to remove this template drift, at least as long as the nest is visible in the image, we added, as an extension of M3, a contour detection stage. Contours are determined around the position of the best match. Then, assuming that the contour closest to the position of the best match belongs to the nest, the template is updated with the image region centred at this contour.

**M5: Template Update on Rotated Image with Rotation Angle Threshold.** An alternative approach, in particular in case the nest entrance is not detectable at later stages of the learning flight, is to try to limit the number of template updates. M5 is a variation of M3: we compensate for rotations but keep the current template unless the rotation angle is larger than a defined threshold value.

**M6: Template Update on Rotated Image with Cumulative Angle Threshold.** Instead of considering just the turn angle between consecutive frames as in M5, we now update the template only if the cumulative turning angle, i.e. the change of head orientation since the last update, exceeds a certain threshold.

**M7: Template Update on Rotated Image with Matching Score Threshold.** This method is similar to M5 and M6. However, instead of the rotation angle we consider the matching score. The template is updated only if the similarity between the current template and the current best match falls below a certain threshold (for the dis-similarity measure SSD we update the template only if the matching score raises above a certain threshold).

Each of these tracking algorithms was implemented in C++ using the template matching methods provided by the OpenCV library (`http://opencv.org`). The implementation uses 40 by 40 pixel templates and the search area was restricted to a $70 \times 70$ pixel region. For M5 and M6, threshold angles for head rotation were fixed to 5° and 10° (cumulative angle), respectively. The matching
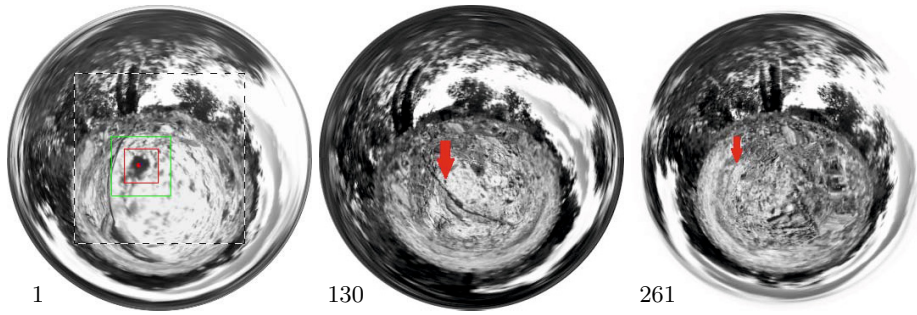
**Fig. 3.** The first, an intermediate and the last frame of *wasp learning flight 1*. The dashed rectangle overlaid on frame 1 depicts the central part of the image used to display the tracking sequences in figure 4. Red arrows point to the nest position.

score threshold of M7 was set empirically to $3 \times 10^6$ for SSD and to 0.70 for NCC.

## 4    Experiments and Results

In this section, we will first show detailed results for the different tracking methods focusing on *wasp learning flight 1*. We will then present results of the tracking methods for different learning flights and investigate the effects caused by reducing the resolution of the images and the precision with which the rotation angle is known.

Figure 3 shows three example frames from *wasp learning flight 1* which consists of 261 frames reconstructed at 50 fps (i.e. at every 5th position of the recorded flight path that was filmed at 250 fps). The red box in the first frame highlights the image region centred around the nest entrance that is used as initial template. The green square depicts the search area for the next frame. As can be seen from figure 3, apparent size of the nest entrance becomes smaller and eventually invisible. Images were rendered with 1°/pixel, which is still higher than the resolution of most insect eyes, including those of wasps and bees [19].

As illustrated in figure 4, the performance of individual tracking methods is quite different. We defined a tracking method to fail when the true nest position is located outside the best matching image region (depicted by the red box). The frames where this happened first are marked by a red cross in the lower right corner. For the results presented in figure 4 SSD was used as similarity measure. M1 has no template update and fails, as expected earlier than all other methods, at frame 56. The continuously updating template method M2 can track the nest region more than twice as long as M1, but fails at frame 138 due to drifts because of rotation induced template mismatches. M3 and M4, which both compensate for rotations, are successfully tracking the nest region for the entire learning flight. M4 has almost perfect tracking performance until about frame 150 after which the nest entrance is too small to be detected by the contour
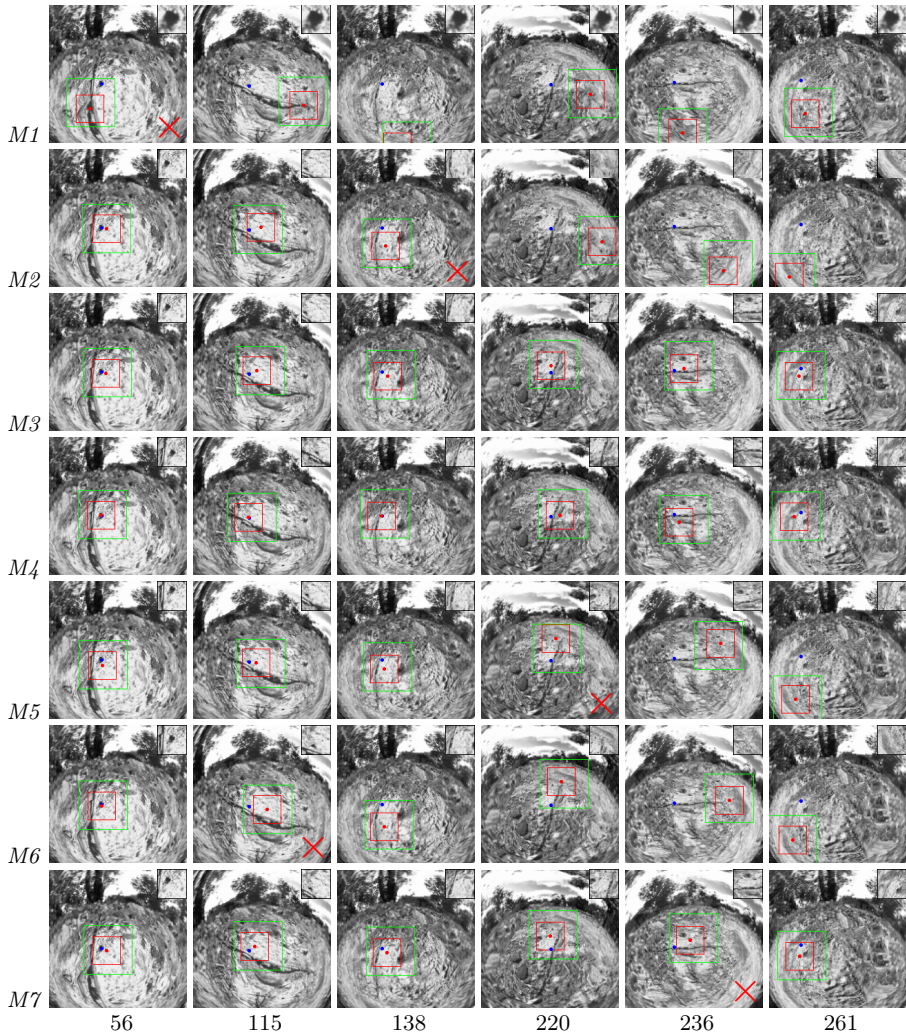
**Fig. 4.** Frames were some tracking methods fail while others succeed for *wasp learning flight 1*. Each row shows results for a different method, indicated by labels M1-M7 on the left. Frame numbers are given below. Insets in the upper right corner of each frame show the respective template. Red crosses in the lower right corners mark frames were individual tracking methods failed first. Blue dots mark the true nest position. Red boxes with a red dot in their centre show the best matching image regions. The green rectangle defines the search area.

finding algorithm. M5, M6 and M7 fail earlier because of the accumulating error between updates, which causes the best matching region to drift over time.

In figure 5 a we plot the pixel error, i.e. the distance (in pixels) of the centre of the best match from the true nest position in the image, for all methods over the full duration of the recorded flight. On average, tracking can be slightly enhanced by using the normalized cross correlation (NCC) instead of the sum of squared differences (SSD) for calculating the matching score (compare figure 5 a and b).

The proposed tracking methods were tested with 3 more learning flights including a bee learning flight. *Wasp learning flight 2* and *wasp learning flight 3* consist of 164 and 220 frames, respectively, reconstructed at 50 fps. The *bee learning flight* has 610 frames reconstructed at full frame rate of 250 fps. As shown in figure 5 c, tracking methods that regularly update the template and compensate for rotations performed also best for *wasp learning flight 2* due to the presence of structures with high contrast close to the nest. On average, tracking methods had the smallest error for *wasp learning flight 3* (figure 5 d), most likely because the entrance hole presented the only high contrast feature in the vicinity of the nest (see inset in upper left corner).

Tracking results for the *bee learning flight* are shown in figure 5 e,f. Most likely due to the higher frame rate, which reduces the amount of change between consecutive frames, the simple continuously updating tracking method 2 is performing much better for this flight (see yellow curve and compare with results for wasp learning flights in figure 5 a-d that were reconstructed with 50 fps).

In order to see the *effect of image resolution* we also tested the tracking methods with half resolution images, i.e. $2°$/pixel instead of $1°$/pixel. As shown in figure 5 f, error does not increase significantly despite the reduced image resolution. The same conclusion can be drawn from the results with half resolution images for wasp learning flights (data not shown).

Tracking methods M3-M7 compensate for head rotations by counter-rotating the template and the centre of the search region (see section 3). For the results presented so far we used the exact value of the turning angle. However, the insects may not be able to predict saccade-induced image shifts accurately. We confirmed that turning angles do not have to be known exactly, because adding 10% noise to the turning angles did not significantly affect performance (data not shown).

**Nest Tracking on Bee Eye Views.** So far we considered images with a fisheye projection that covered the full viewing sphere and thus the large field of view of both insect eyes combined without the discontinuity introduced by having two eyes. For modeling visual tracking of the nest in a more realistic way we created views according to a model that resembles the spatial sampling of the eyes of a worker honeybee [17]. Due to the binocular overlap, the nest, when located in the frontal visual field, will be visible in both eyes (see left side of figure 6), which may facilitate switching the tracking of the nest from one eye to the other.[1] For

---

[1] Interestingly, the binocular overlap is larger in the lower visual field (the region onto which the image of the nest will be projected) than in the frontal visual field.
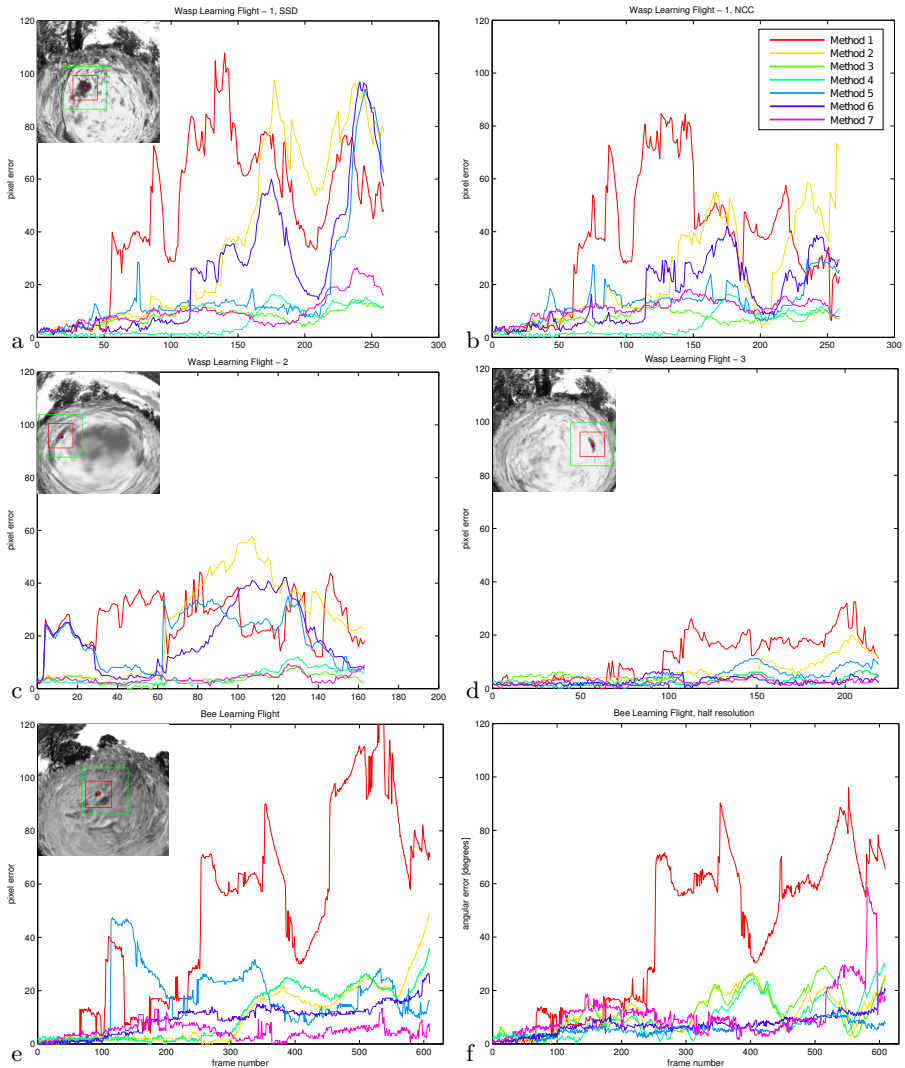
**Fig. 5.** Performance of different template tracking methods for three wasp learning flights (a-d) and one bee learning flight (e,f). SSD was used as similarity with the exception of b) which shows results for NCC. Insets in the upper left corners show the central part of the first frame of the respective learning flight; red boxes highlight the initial tracking template. f) Angular error for half image resolution.

implementing tracking on bee eye views we extended the search area to both eyes whenever the best match region found in the previous frames is close to the inner border of an eye. The right side of figure 6 shows six example frames from tracking method M2. The true nest location is kept within the region of the best match for the whole sequence of 610 frames.
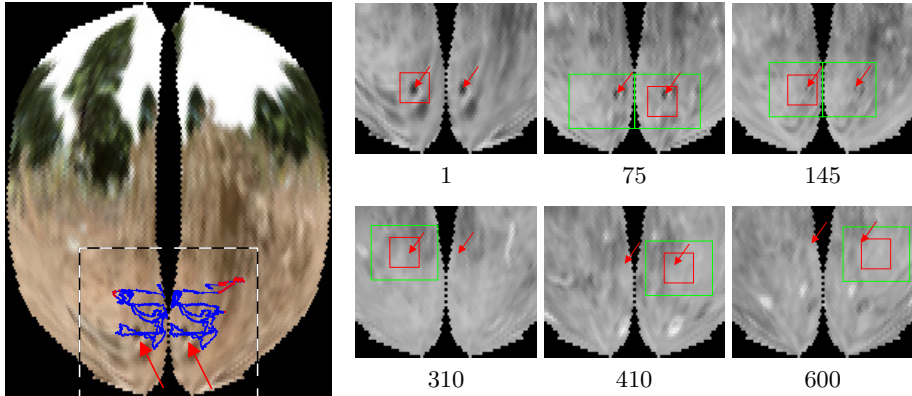
**Fig. 6.** Tracking on bee eye views. Left: First frame of the bee learning flight. The arrows highlight the nest entrance, which is seen in both eyes. The blue dotted curves illustrate how the nest position moves across the visual fields of both eyes during the learning flight, red dots highlight positions where the nest is seen by one eye only. The dashed rectangle depicts the part of the image used for displaying tracking results on the right side. Right: Example frames (with frame number below) illustrating tracking results using method M2.

## 5    Discussion

Ground-nesting insects acquire a visual representation of their nest environment during learning flights on departure. As the insects pivot around the nest entrance and back away from it, they carefully control where in the visual field they see the nest. This cannot be achieved by a simple position servo, because the visual appearance of the nest entrance and its immediate environment changes as the viewing direction and the distance of the insect changes during these flights. We have shown here, that we could track the image location of the nest in the reconstructed views that insects experience during learning flights, using updated template matching and a version of predictive tracking that accounts for the image shifts generated by the saccadic head movements of insects.

The possibility that wasps and bees use template matching when keeping track of the nest location can be tested by modifying high-contrast, artificial patterns around the nest entrance during learning flights. It is already known that the insects track such patterns when they are shifted [10] and a break-down of nest position control in the visual field in the presence of rapid pattern changes (not shifts) would indicate that the insects do employ template matching during their learning flights.

Why is visual goal-anchoring so important during the acquisition of visual-spatial memories? We suggest that it allows the insects to continuously form a strong association between changing views and the direction to the nest. After all, the purpose of this learning process is to ensure that sufficient information

has been acquired to allow the insect to pinpoint its nest on subsequent returns. The systematic, periodic structure of learning flights (in terms of the temporal sequence of bearing and orientation changes) indicates that the insects have several opportunities during these flights to check and re-check what they have learnt for consistency.

# References

1. Zeil, J., Boeddeker, N., Stürzl, W.: Visual homing in insects and robots. In: Floreano, D., Zufferey, J.C., Srinivasan, M.V., Ellington, C. (eds.) Flying Insects and Robots, pp. 87–100. Springer, Heidelberg (2009)
2. Zeil, J.: Visual homing – an insect perspective. Current Opinion in Neurobiology 22, 285–293 (2012)
3. Zeil, J., Hofmann, M., Chahl, J.: Catchment areas of panoramic snapshots in outdoor scenes. Journal of the Optical Society of America A 20(3), 450–469 (2003)
4. Graham, P., Cheng, K.: Ants use the panoramic skyline as a visual cue during navigation. Current Biology 19, R935–R937 (2009)
5. Evangelista, C., Kraft, P., Dacke, M., Labhart, T., Srinivasan, M.V.: Honeybee navigation: Critically examining the role of the polarization compass. Phil. Trans. R Soc. B 369, 20130037 (2014), doi:10.1098/rstb.2013.0037
6. Dacke, M., Baird, E., Byrne, M., Scholtz, C., Warrant, E.: Dung beetles use the milky way for orientation. Current Biology 23(4), 298–300 (2013)
7. Zeil, J., Kelber, A., Voss, R.: Structure and function of learning flights in bees and wasps. Journal of Experimental Biology 199, 245–252 (1996)
8. Müller, M., Wehner, R.: Path integration provides a scaffold for landmark learning in desert ants. Current Biology 20, 1368–1371 (2010)
9. Zeil, J.: Orientation flights of solitary wasps (Cerceris; Sphecidae; Hymenoptera). I. Description of flight. Journal of Comparative Physiology A 172, 189–205 (1993)
10. Zeil, J.: Orientation flights of solitary wasps (Cerceris; Sphecidae; Hymenoptera). II. Similarities between orientation and return flights and the use of motion parallax. Journal of Comparative Physiology A 172, 207–222 (1993)
11. Baddeley, B., Philippides, A., Graham, P., Hempel de Ibarra, N., Collett, T.S., Husbands, P.: What can be learnt from analysing insect orientation flights using probabilistic SLAM? Biol. Cybern. 101, 169–182 (2009)
12. Zeil, J.: The control of optic flow during learning flights. Journal of Comparative Physiology A 180, 25–37 (1997)
13. Zeil, J., Boeddeker, N., Hemmi, J., Stürzl, W.: Going wild: Toward an ecology of visual information processing. In: North, G., Greenspan, R. (eds.) Invertebrate Neurobiology. Cold Spring Harbor (2007)
14. Stürzl, W., Mair, E., Hirschmüller, H., Zeil, J.: Mapping the navigational information content of insect habitats. In: Front. Physiol. Conference Abstract: International Conference on Invertebrate Vision (2013), doi:10.3389/conf.fphys.2013.25.00085
15. Mair, E., Stürzl, W., Zeil, J.: Benchmark 3D models of natural navigation environments @ www.insectvision.org. In: Front. Physiol. Conference Abstract: International Conference on Invertebrate Vision (2013), doi:10.3389/conf.fphys.2013.25.00084

16. Zeil, J., Narendra, A., Stürzl, W.: Looking and homing: How displaced ants decide where to go. Philosophical Transactions B 369(1636), 20130034 (2014), doi:10.1098/rstb.2013.0034
17. Stürzl, W., Boeddeker, N., Dittmar, L., Egelhaaf, M.: Mimicking honeybee eyes with a 280° field of view catadioptric imaging system. Bioinspiration & Biomimetics 5, 36002 (2010)
18. Wystrach, A., Graham, P.: What can we learn from studies of insect navigation? Animal Behaviour 84, 13–20 (2012), doi:10.1016/j.anbehav.2012.04.017
19. Land, M.F.: Visual acuity in insects. Annual Review of Entomology 42, 147–177 (1997)