T20-6B

Universität Bielefeld

Faculty of Biology
Biological Cybernetics

CITEC

Marker-less Motion Capture of Antennal Movement Kinematics in Honeybees and other Hymenopterans

Volker Dürr, Florian P. Schmidt, Tristan Walter, Simon M. Würth, and Mario Botsch

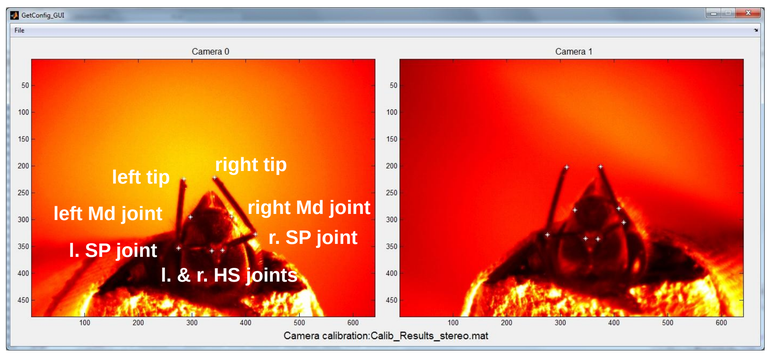
Correspondence :volker.duerr@uni-bielefeld.de

Introduction

Bees extensively use their antennae for active tactile exploration, pattern recognition and learning. To date, the honeybee (*Apis mellifera*) is the only insect species that has been investigated systematically in both non-associative motor learning and associative conditioning paradigms. In each of these paradigms, bees actively sample the ambient space with both antennae. For investigating the role of antennal movement sequences in tactile learning, a posture tracking system is needed for resolving antennal kinematics.
In our earlier work, we proposed a system for automatic tracking of the antennal tip position [1]. Here, we present a method for automated and reliable marker-less posture-tracking of both antennae in Hymenopterans with geniculate antennae, honeybees in particular.

Experiments

For testing and evaluation of the tracking algorithm, we captured stereo videos of spontaneous antennal movements in honeybees and bumblebees under various conditions. Bees were fixed in metal tubes, strongly restraining head movements. Spontaneous antennal movements were recorded for 60 s at 100 fps, using a stereo video camera system with macro lenses. Cameras were calibrated for triangulation of



3D positions. For each trial, a head-fixed reference coordinate system was determined by manual labelling of the anterior mandible joints (Md) and the antennal head-scape joints (HS) in a single stereo frame pair (see above). Furthermore, the scape-pedicel joints (SP) and antennal tips were labelled to set the segment lengths of the model.
The antenna model had 4 degrees of freedom (DoF) in each antenna: three yaw-pitch-roll angles in the HS joint and an extension/flexion angle in the SP joint.

Model-based Tracking Algorithm

The tracking algorithm uses a virtual model with 4 DoF per antenna and calculates the correspondence between the projections of the model into both camera views and the pixel clusters of the processed video. The correspondence is optimised by a particle swarm algorithm [2] that adjusts the 2x4 DoF. To begin with, all particles of the swarm start at random positions in an 8D "search space", moving with a velocity that is updated interatively by

$$v_i \leftarrow \underbrace{\omega}_{inertia}\left(v_i + \underbrace{\phi_r \Theta}_{random\ velocity} + \underbrace{\phi_p r_p (p_i - x_i)}_{cognitive} + \underbrace{\phi_g r_g (g - x_i)}_{social}\right)$$

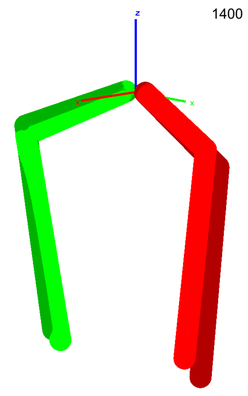
where x_i is the position of the particle, p_i is the best position the particle has reached and g_i is the best position any particle of the swarm has reached. The constants r_x control the relative influences of the individual terms, and ϕ_x are random values between 0 and 1. The inertia constant ω ranges between 0 and 1. It slows down the changes from iteration *n* to *n+1*. The random velocity Θ term helps the particles to escape from local minima.
The particle positions and velocities are updated iteratively, making them "move" through the search space. A position in the search space is a vector containing all 2x4 DoF of the antenna model.
In each iteration the antenna model is projected into the camera images. The *error function*

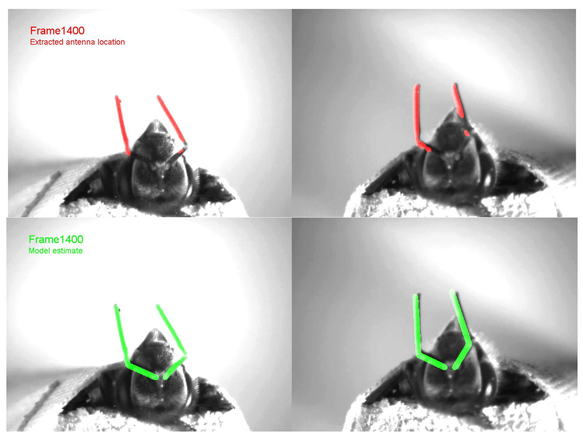
$$E(w,f_i) = \frac{\sum_{x=0}^{w}\sum_{y=0}^{h} and(w(x,y),f_i(x,y))}{\sum_{x=0}^{w}\sum_{y=0}^{h} w(x,y)}$$

quantifies the mismatch between the binarized video image *w* and the projected model images *f*. For every pixel in each rendered image, it performs a logical *AND* operation and adds all the ones together, normalized by the overall number of white pixels in the original. Once this value is sufficiently small, a good configuration has (most likely) been reached. Error values are averaged over two (or more, if available) camera perspectives. This yields the overall error E_{ges}.
Greyscale information of the original videos may be considered in the error function above, if the binary image is not obtained by simple thresholding but rather by brightness-dependent dithering .

The Tracking Algorithm at Work

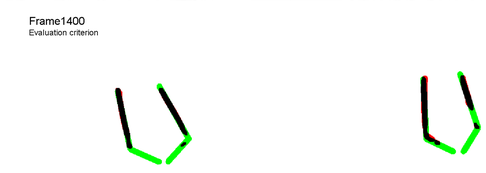
The image to the right shows both the antenna model (lighter colours) and the true antennal posture (darker colours) as inferred from triangulation of manually tracked samples (green: right antenna; red: left antenna).
For automated posture tracking, videos were processed by subtracting a mean image, thresholding and filtering, yielding a binarized image with pixel clusters that correspond to parts of the antennae (and other moving image parts, like shadows or mouthparts).





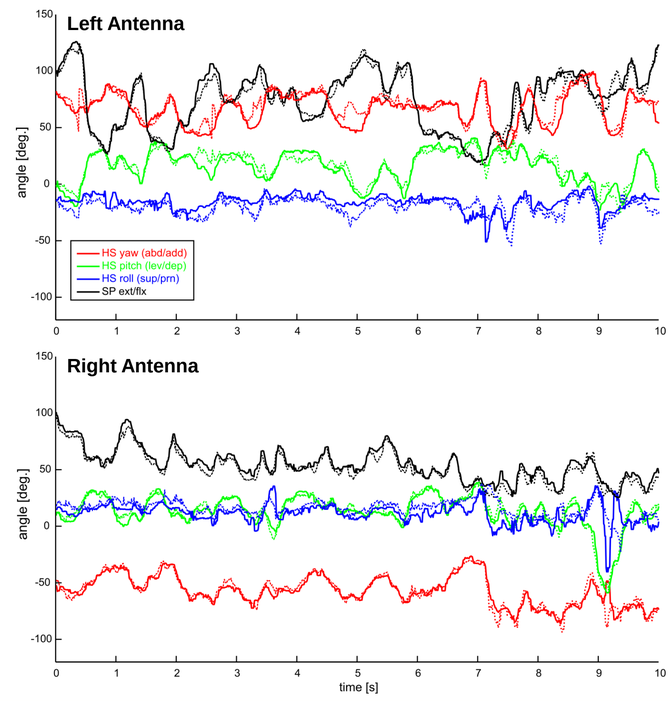
Red areas show the overlap of the binarized and the original images.

Green shows the overlay of the back-projected model and the original image.

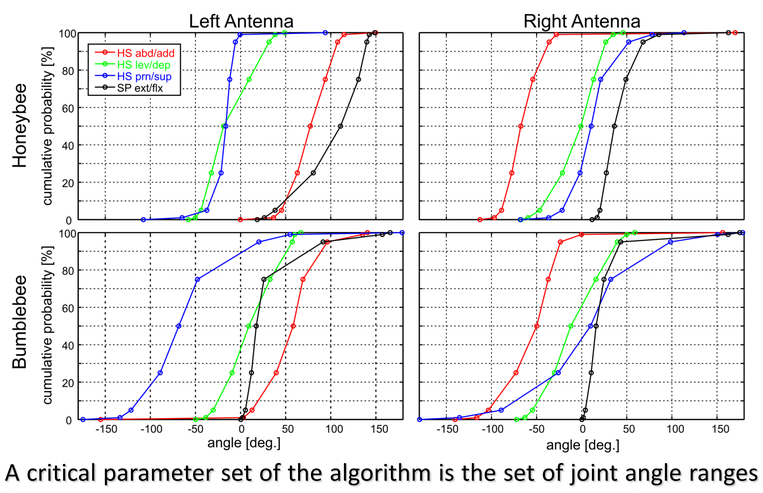


Black indicates superposition of the binarized images and the back-projected model.

Accuracy and robustness of the algorithm was evaluated by comparison with three manually tracked videos from one honeybee and two bumblebees. The graphs below demonstrate the good correspondence between the automated tracking result (64 particles) and the true joint angles (dotted lines) that were inferred from manual tracking .



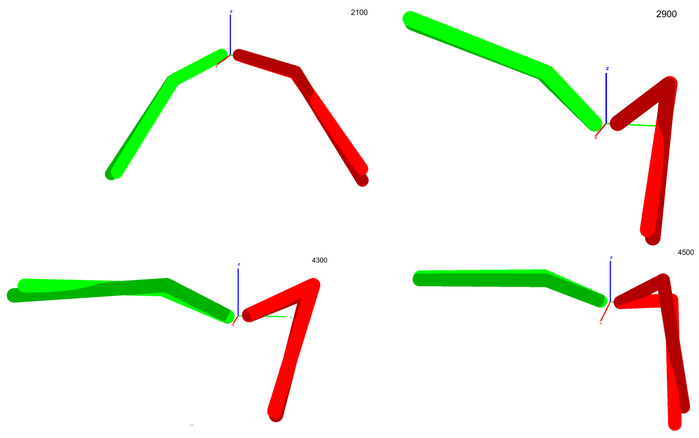
Joint Angle Ranges Differ among Species

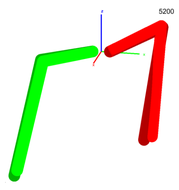


A critical parameter set of the algorithm is the set of joint angle ranges for each DoF. The figure above shows how the joint angle distributions of manually tracked antennae differ between a honeybee (top) and two bumblebees (bottom). Here, the bumblebee antennae were less abducted (red), more extended (black) and supination/pronation of the scape (blue) varied much more than in the honeybee.

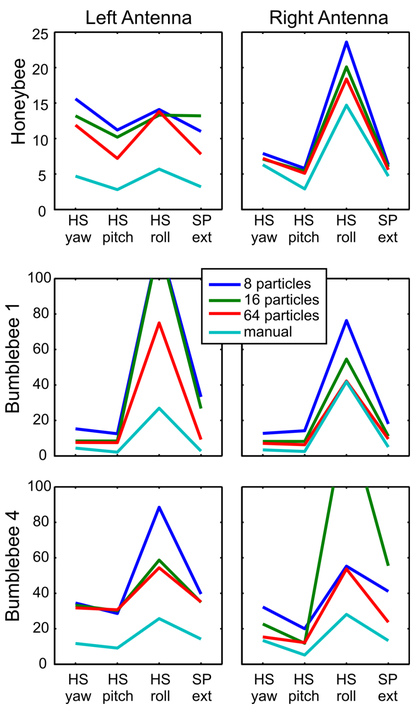
Evaluation of the Algorithm

Generally, the accuracy of the algorithm was good, being better for honeybees than for bumblebees. Problems occurred whenever (i) one or both antennae did not move much for some time (average image contains the antenna), (ii) an antenna was strongly extended (supination angle = HS roll becomes arbitrary) , (iii) an antenna moves too fast (image blurring), or (iv) if an antenna casts a shadow (the shadow might get tracked)





Frame	MSQ left	MSQ right	Correct-ness
1400	**5.0**	**6.8**	**0.925**
2100	**103.4**	11.9	0.874
2900	3.0	**737.6**	0.855
4300	55.8	44.3	**0.735**
4500	**139.1**	**134.8**	0.911
5200	**3.5**	**9.4**	**0.896**



Overall evaluation reveals that the honeybee trial was tracked much better than the bumblebee trials. In all cases, supination/pronation of the scape (HS roll) tends to cause the largest error, whereas levation/depression (HS pitch) and extension/flexion of the SP joint are most accurate.
The figure to the left shows the rms error between all manually and automatically tracked postures, per DoF. This was done for three variants of the algorithm, differing in particle number. The rmse may be compared to the jitter caused by manual inaccuracy.

Further tests suggest that the tracking performance may be improved if the greyscale of the background-subtracted image is considered, e.g., by brightness-dependent dithering of the binary image. In case of the bumblebee videos, this led to an rmse reduction by up to 25%.

Conclusions

- **Based on stereo-videos, the antennal movement of honeybee antennae can be tracked accurately by a marker-less model-based particle swarm algorithm.**
- **Generally, frequent strong extension of the antenna hampers tracking accuracy, because HS roll becomes arbitrary. This problem occurs more often in bumble-bees than in honeybees**
- **Tracking is also sensitive to video quality, extreme movement speed, and shadows caused by sample illumination. Tracking may be improved by including greyscale information into the model fit estimate.**

References

[1] S. Mujagic, S.M. Würth, S. Hellbach, V. Dürr (2012) JoVE: e50179
[2] J. Kennedy, R. Eberhart (1995) Particle swarm optimization. In: *Neural Networks* 4: 1942–1948.