Erschienen in: Journal of Neuroscience Methods ; 239 (2015). - S. 194-205
https://dx.doi.org/10.1016/j.jneumeth.2014.10.021

Automated tracking and analysis of behavior in restrained insects

Minmin Shen[a,d,*], Paul Szyszka[b], Oliver Deussen[a], C. Giovanni Galizia[b], Dorit Merhof[c]

[a] *INCIDE Center (Interdisciplinary Center for Interactive Data Analysis, Modelling and Visual Exploration), University of Konstanz, Germany*
[b] *Institute of Neurobiology, University of Konstanz, Germany*
[c] *Institute of Imaging & Computer Vision, RWTH Aachen University, Aachen, Germany*
[d] *School of Software Engineering, South China University of Technology, PR China*

HIGHLIGHTS

- We present an algorithm for tracking the movement body parts of restrained animals.
- The tracking algorithm works with low frame-rate videos.
- The tracking algorithm automatically segments and tracks multiple body parts.
- We demonstrate the power of the algorithm in analysing insect behaviour.

Insect
Behavior
Honey bee
Classical conditioning
Multi-target tracking
Antenna

ABSTRACT

Background: Insect behavior is often monitored by human observers and measured in the form of binary responses. This procedure is time costly and does not allow a fine graded measurement of behavioral performance in individual animals. To overcome this limitation, we have developed a computer vision system which allows the automated tracking of body parts of restrained insects.
New method: Our system crops a continuous video into separate shots with a static background. It then segments out the insect's head and preprocesses the detected moving objects to exclude detection errors. A Bayesian-based algorithm is proposed to identify the trajectory of each body part.
Results: We demonstrate the application of this novel tracking algorithm by monitoring movements of the mouthparts and antennae of honey bees and ants, and demonstrate its suitability for analyzing the behavioral performance of individual bees using a common associative learning paradigm.
Comparison with existing methods: Our tracking system differs from existing systems in that it does not require each video to be labeled manually and is capable of tracking insects' body parts even when working with low frame-rate videos. Our system can be generalized for other insect tracking applications.
Conclusions: Our system paves the ground for fully automated monitoring of the behavior of restrained insects and accounts for individual variations in graded behavior.

1. Introduction

Insects are often used to study the neuronal mechanisms that underly behaviors ranging from sleep to higher-order associative learning (Sauer et al., 2003; Matsumoto et al., 2012; Menzel, 2012). When controlled stimulus conditions are needed, insects are often restrained and their behavior is monitored as movements of body parts such as their antenna or mouthparts. Insect behavior is often measured by human observers and recorded in the form of binary responses to prevent the introduction of subjective biases by the observer. This procedure is time consuming and it does not allow a fine graded measure of behavioral performance in individual animals.

In neuroscience the honey bee is a particularly powerful model animal for learning and memory research (Menzel, 2012). Associative learning of individual, fixed bees can easily be studied by classical conditioning, where an odorant is paired with a sugar reward. Whether a bee has learned the association is usually assessed by its proboscis (i.e. the mouthpart of the bee) extension response (binary all-or-nothing measure) (Bitterman et al., 1983). A bee extends the proboscis reflexively when stimulated with sugar

* Corresponding author. Tel.: +49 7531 88 5108; fax: +49 7531 88 4715.
E-mail addresses: minmin.shen@uni-konstanz.de (M. Shen), paul.szyszka@uni-konstanz.de (P. Szyszka), oliver.deussen@uni-konstanz.de (O. Deussen), giovanni.galizia@uni-konstanz.de (C.G. Galizia), Dorit.Merhof@lfb.rwth-aachen.de (D. Merhof).

water or with a previously conditioned odorant. Up to now learning and memory have been mainly assessed by a crude all-or-nothing measure (whether a bee reacts to a learned stimulus, or not). This binary measurement is not suited to reveal individual differences in learning and memory performance, for this purpose a graded performance measurement is required (Pamir et al., 2014).

A graded measure for learning and memory can be extracted from the temporal characteristic of the proboscis extension response, which contains information about whether a bee has learned an association or not (Rehder, 1987; Smith et al., 1991; Gil et al., 2009). Moreover, temporal patterns of antennae movement change upon sensary stimulation (Erber et al., 1993) and reveal internal states such as sleep and wakefulness (Hussaini et al., 2009; Sauer et al., 2003). To precisely analyze such dynamic behavioral monitors, tracking systems are required. However, available insect tracking systems often have the weakness that they require prior marking of the animal (Hussaini et al., 2009), and are often capable only of tracking single insects (Veeraraghavan et al., 2008; Landgraf and Rojas, 2007), working with slowly-moving insects only (Balch et al., 2001; Ying, 2004), or can track only one type of body part, i.e. bee's antennae (Hussaini et al., 2009; Mujagić et al., 2011).

We addressed this issue and developed a computer vision system which allows the automated tracking of the body parts of restrained insects while providing quantitative information about the movements of their mouthparts and antennae. This system can easily be adopted to other insects, and it allows one to implement novel approaches to analyze insect behavior using graded measures of behavioral performance.

2. Materials and methods

We will elaborate our system as follows. We firstly perform moving object detection by subtracting the static background (Section 2.3). The moving object detector generates a set of bounding boxes (BBs), which are rectangles that bound detected objects. We then preprocess the input frame to reduce undesired BBs including false, missing, splitted and merged ones (Section 2.4). The appearance model is constructed in Section 2.5. Finally we propose a tracking algorithm in Section 2.6, which is able to identify the label of each of the five moving objects: "1" for right antenna, "2" for right mandible, "3" for proboscis, "4" for left antenna and "5" for left mandible as shown in Fig. 1c. For the sake of clarity, in Table 1 we list all abbreviations and notations used in the paper.

2.1. Video acquisition

Honey bee foragers (*Apis mellifera*) were caught from outdoor hives and prepared as described in Szyszka et al. (2011). Small ant workers (*Camponotus floridanus*) were provided by C.J. Kleineidam. Colonies were reared in a climate chamber at 50–60% relative humidity and 26 °C. The founding queens were collected by A. Endler and S. Diedering in Florida Keys (USA). The ant's neck was pushed through a slit in plastic foil, and its head was fixed dorsally to the plastic foil with a low temperature melting, equal-weight mixture of dental wax (Deiberit 502; Dr. Böhme und Schöps Dental), and n-eicosan and myristic acid (both Sigma–Aldrich). Each individual insect was imaged at 30 frames per second using a CCD camera ("FMVU-03MTM/C" Point grey, Richmon, Canda) in order to record the head with proboscis, mandibles and antennae. The setup of the bee experiment is shown in Fig. 1a. Insects were recorded with or without odor stimulation and sugar feeding. Odor stimulus delivery was monitored by lighting an LED within the field of view of the camera, so that data analysis can be done relative to stimulus delivery. Insects were harnessed on a platform, with their head in fixed positions, but able to move antennae and mouthparts freely. The camera was set on top of an individual insect. The camera was fixed, and the platform to which the insects were fixed was moved when changing to a new insect for recording. Unlike the high speed camera used in (Voigts et al., 2008), which is capable of capturing videos at 500 frames/s, the frame-rate of the acquired movies in this paper was only 30 frames/s. Although it would be possible to record with a high speed camera, we aim at developing a system that uses affordable cameras such as web-cam or consumer level cameras and keeps the data volume low. Each video was about 30 min long and consists of 12 trials, with 16 individual honey bees each. For each trial, a single video to be processed was approximately 10–30 s long and had a frame size of 480 × 640 pixels.

2.2. Coordinate system setup

To extract the information of the relative position of each object to the insect head, it is required to set up the coordinate system. As the platform is not static during the changing of insects, the scene change is detected to ensure a static background before the actual tracking procedure starts. For scene change detection, the edges in each frame were detected using a Sobel Filter. The mean of all the blocks within the edge image is computed and compared to the mean of all the blocks of the previous frame. If the absolute difference of means between two blocks in consecutive frames is greater than a predefined value, the block is assumed to be changed. The scene is detected to be changed if the number of changed blocks is greater than a predefined number. The video is cropped into several shots automatically according to the scene change detection.

For each shot, the mean of the first ten frames is used to estimate the insect head's position. After thresholding, a dark region with the greatest circularity value and an area within the range of 0.33–2.6% of the whole image is selected as the segmented head, and the position of the origin is estimated as the left-most point of the segmented head (as shown in Fig. 1b). With the origin (marked as point "o") and the centroid of the head (marked as point "c") estimated, a new coordinate system is established by using the mandible as the origin, line "oc" as x-axis and the line orthogonal to "oc" as y-axis.

2.3. Object detection

For detecting moving objects, Gaussian Mixture Model (GMM) background modelling (KaewTraKulPong and Bowden, 2002) is used. The first five frames of each shot are used for training the initial model parameters of the GMM background model. As in KaewTraKulPong and Bowden (2002), background subtraction is performed by marking a pixel as a foreground pixel if it is more than 2.5 standard deviations away from any of the distributions of the background model. The background model is updated for each frame; and a static object staying long enough will be determined as part of the background. The model is suitable for our case, where a static background exists in each shot.

2.3.1. LED and sugar stick detection

As the LED is used to indicate when the odor is released, detection of the LED is part of our task. Due to the nature of the GMM background model, the detection of the LED fails when it is on for a few seconds. To address this problem, we store the BB of the LED when it is detected for the first time, and measure the intensity within the BB. If the intensity is greater than the average of the image, the LED is determined to be on.

The time when the sugar stick touches the insect is required for assessing the latency of its proboscis extension response. A BB that is attached to the dilated head having a width or height greater than 100 pixels is assumed to be the sugar stick.

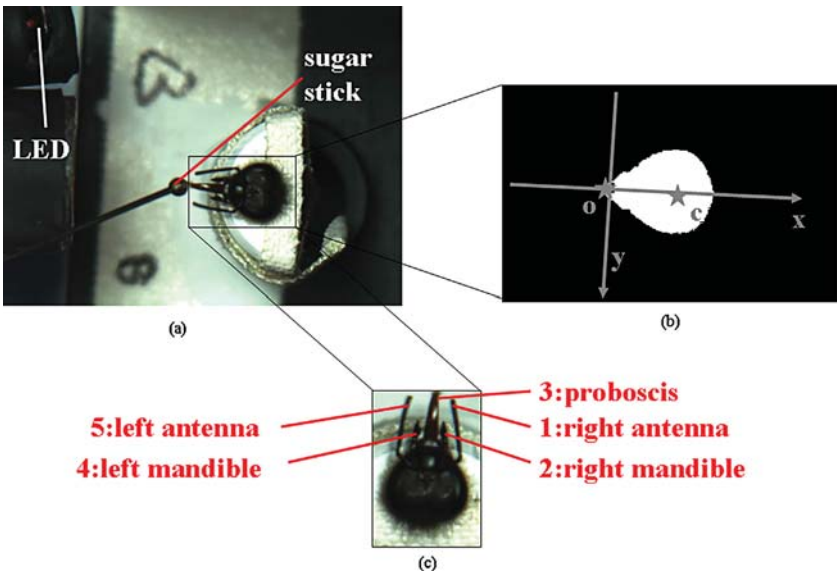

Fig. 1. Illustration of the coordinate system setup: (a) Example of the experimental setup, (b) constructed coordinate system: the left-most point of the bee's head is marked as point "o" and the centroid of the bee's head is marked as point "c". The new coordinate system uses the point "o" as the origin, line "oc" as x-axis and the line orthogonal to "oc" as y-axis. (c) label of each object which needs to be identified.

2.4. Preprocessing

The object detector generates a set of false BBs (e.g. shadows, reflection and legs), missing BBs (motion blurred antenna or antenna above the head), splitted BBs (splitted BBs of the same antenna), merged BBs (one BB including two or three objects), which make the following tracking task difficult. Therefore, preprocessing operations include exclusion of undesired BBs by incorporating position information, shadow removal (KaewTraKulPong and Bowden, 2002), merging splitted BBs, and splitting merged BBs.

We will show in Section 3.1 that these preprocessing operations greatly reduce the detection of undesired BBs, but some false, missing, splitted and merged BBs may still remain. Thus the tracking algorithm is required to tackle this problem.

2.5. Appearance model

A feature vector $\mathbf{f}_{i,j} = [\mathbf{f}_{i,j}(1), \ldots, \mathbf{f}_{i,j}(7)]^T$ is extracted for the ith object $z_{i,j}$, $i = 1, \ldots, n_j$ in the jth frame $\mathbf{Z}_j$, $j = 1, \ldots, N$ to indicate its position, shape, geometry and speed, where n_j is the number of the detected objects in $\mathbf{Z}_j$ and N is the number of frames. Seven features are used to represent the appearance model and are listed in Table 1: distance between the nearest vertex and mandible $\mathbf{f}_{i,j}(1)$, distance between the furthest vertex and the y-axis $\mathbf{f}_{i,j}(2)$, area of the object $\mathbf{f}_{i,j}(3)$, motion vector ($\mathbf{f}_{i,j}(4)$,$\mathbf{f}_{i,j}(5)$), area of top-hat filtered output $\mathbf{f}_{i,j}(6)$, and a logical variable $\mathbf{f}_{i,j}(7)$ indicating whether the furthest vertex is within the BB.

To represent the position of each BB, the vertices nearest or furthest to point "c" are extracted and denoted as point "n" and "f" in Fig. 2, respectively. The distance between point "n" and "o" and the x-coordinate of point "f" are used as features. The shape of each object is indicated by the area of the black region, since each object is black. A top-hat filter is used as a ridge detector for identifying antennae: after thresholding and greyscale reversion, the top-hat filter is applied to the image block within its BB. However, the output of the top-hat filter is not a unique feature for antennae. As illustrated in Fig. 2, there are three BBs including an antenna without motion blur, a proboscis with reflection of light, and an antenna with severe motion blur. Their outputs of the top-hat filter are also shown. It can be seen that the area detected by the top-hat filter may be significantly different if the antenna has severe motion blur. On the other hand, the area of the top-hat filter output of a proboscis with reflection may be comparable to an antenna, so we have to distinguish between these two cases by other features. Therefore, whether the area of the top-hat filter output of the image patch within the BB is greater than 0 serves as a condition when calculating the conditional probability. Similarly, whether point "o" is

Table 1
Abbreviation and Notations

BB	Bounding box	GMM	Gaussian mixture model
$\mathbf{Z}_j$	jth frame	$z_{i,j}$	ith BB in $\mathbf{Z}_j$
N	number of frames	n_j	number of the detections in $\mathbf{Z}_j$
C_j	set of $c_{i,j}$	L_j	ordered set of $l_{i,j}$
$l_{i,j}$	label of $z_{i,j}$,1: right antenna; 2: right mandible; 3: proboscis; 4: left mandible; 5: left antenna; 6: false positive.		
$c_{i,j}$	class of $z_{i,j}$,1: antenna; 2: mandible; 3: proboscis.		
$\mathbf{f}_{i,j}$	feature vector		
$\mathbf{f}_{i,j}(1)$	distance between the nearest vertex "n" and "o"		
$\mathbf{f}_{i,j}(2)$	x-coordinate of the furthest vertex "f"		
$\mathbf{f}_{i,j}(3)$	area of $z_{i,j}$		
$\mathbf{f}_{i,j}(4)$	x-component of motion vector		
$\mathbf{f}_{i,j}(5)$	y-component of motion vector		
$\mathbf{f}_{i,j}(6)$	area of top-hat filter output		
$\mathbf{f}_{i,j}(7)$	"1" represents point "o" is within the BB; "0" otherwise		

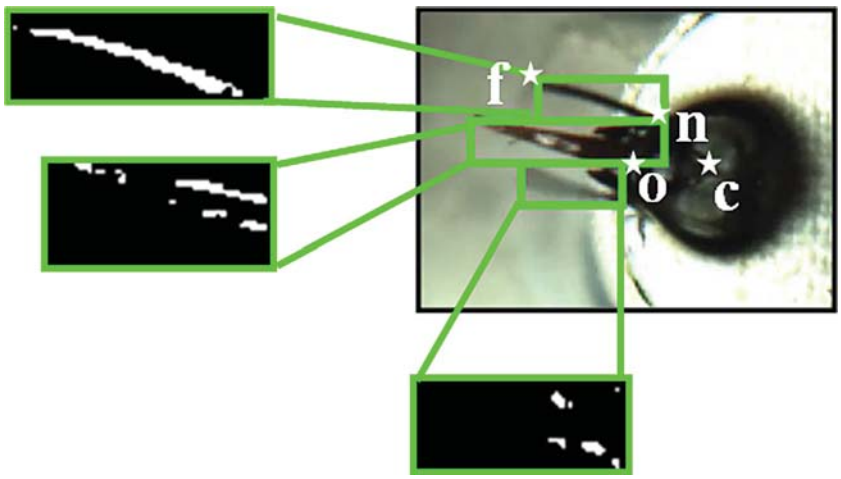

Fig. 2. Appearance model for classifying body part of a bee: the closest vertex of a BB (the top-most BB in this figure) to point "c" is denoted as point "n" and the furthest one as "f". Three BBs including an antenna without motion blur (top), a proboscis with reflection (middle), an antenna with severe motion blur (bottom) and their outputs of the top-hat filter are highlighted to show the property of the top-hat filter as a feature.

within the BB is also used as a feature, as the BB of an antenna seldom includes point "o". The motion vector, which is the relative displacement between each bounding box in its previous frame and current frame is estimated by the template matching method (Yu et al., 2006).

2.6. Tracking algorithm

We propose a novel tracking algorithm that incorporates prior information about the kinematics and shapes of antenna, mandibles and proboscis. The objective of the algorithm is to assign each BB $z_{i,j}$ with labels $l_{i,j}$, where $l_{i,j} \in \{1$:right antenna; 2:right mandible; 3:proboscis; 4:left mandible; 5:left antenna; 6:false positive$\}$. We guide the tracking by using the preceeding frames. The overall tracking algorithm consists of three levels: object level, frame level and temporal level. At object level, the prior probability of the class of each BB (i.e. antenna, mandibles or proboscis) is computed. At frame level, the identification of each BB is assigned according to the sequence in which the BBs are arranged, and the probability that the assignment corresponds to its ground truth is computed based on the prior probability and the prior information of all the objects' order. The frames with the highest probability are treated as benchmarks. The final assignment is fulfilled by frame-to-frame linking between benchmarks and their temporal neighbours. As a result, the transitive update of the assignment generates the most probable identifications.

2.6.1. Object level

At object level, the probability $P(c_{i,j}|\mathbf{f}_{i,j})$ of each BB $z_{i,j}$ for each class $c_{i,j}$ (where $c_{i,j} \in \{1$:antenna; 2:mandible; 3:proboscis$\}$) is computed given its feature vector $\mathbf{f}_{i,j}$. They are further classified as $l_{i,j}$ at frame level described in the following section. Among the seven features, $\mathbf{f}_{i,j}(1), \ldots, \mathbf{f}_{i,j}(3)$ are assumed to follow a Gaussian distribution whose mean μ and covariance matrix Σ are learned from the training set, i.e. a set of annotated BBs. Let us pack the three features into a vector and denote it as $\bar{\mathbf{f}}_{i,j} = [\mathbf{f}_{i,j}(1), \ldots, \mathbf{f}_{i,j}(3)]^T$. The conditional probability $P(c_{i,j}|\bar{\mathbf{f}}_{i,j})$ is computed by

$$P(c_{i,j}|\bar{\mathbf{f}}_{i,j}) = \frac{1}{(2\pi)^{\frac{1}{2}N}|\Sigma|^{\frac{1}{2}}} \exp\{-\frac{1}{2}(\bar{\mathbf{f}}_{i,j}-\mu)^T|\Sigma|^{-1}(\bar{\mathbf{f}}_{i,j}-\mu)\}. \quad (1)$$

The other features $\mathbf{f}_{i,j}(4), \ldots, \mathbf{f}_{i,j}(7)$ are modelled as discrete variables with constant prior probabilities assumed to be known. The class-conditional probability density function $P(c_{i,j}|\mathbf{f}_{i,j})$ of a feature $\mathbf{f}_{i,j}$ is computed based on Bayes' rule.

$$\begin{aligned} & P(c_{i,j}|\mathbf{f}_{i,j}) \\ = \; & P(c_{i,j}|\bar{\mathbf{f}}_{i,j}, \mathbf{f}_{i,j}(4) \in \Phi_4, \ldots, \mathbf{f}_{i,j}(7) \in \Phi_7) \\ = \; & \frac{P(c_{i,j}|\bar{\mathbf{f}}_{i,j})P(\mathbf{f}_{i,j}(4) \in \Phi_4|c_{i,j})}{P(\bar{\mathbf{f}}_{i,j})P(\mathbf{f}_{i,j}(4) \in \Phi_4)} \\ & \cdot \prod_{p=5}^{7} \frac{P(\mathbf{f}_{i,j}(p) \in \Phi_p|c_{i,j})}{P(\bar{\mathbf{f}}_{i,j}, \mathbf{f}_{i,j}(4), \ldots, \mathbf{f}_{i,j}(p-1))P(\mathbf{f}_{i,j}(p) \in \Phi_p)} \end{aligned} \quad (2)$$

where Φ_p is the set that represents the constraint of $\mathbf{f}_{i,j}(p)$, and the conditional probability $P(\mathbf{f}_{i,j}(p) \in \Phi_p|c_{i,j} = k)$ is assumed to be known and set as a constant. For example, $P(\mathbf{f}_{i,j}(6) > 0|c_{i,j} = 1) = 1$, since an antenna must have top-hat filtered pixels. The other unknowns of Eq. (2) can be computed by solving the equations combining the constraint that each object must be an antenna, mandible or proboscis, thus we have:

$$\sum_{k=1}^{3} P(c_{i,j} = k|\mathbf{f}_{i,j}) = 1. \quad (3)$$

Given estimates for $P(c_{i,j}|\mathbf{f}_{i,j})$, a Naïve Bayesian Classifier is performed for each BB to decide which class it belongs to according to the highest conditional probability.

However, a high accuracy is not guaranteed using this approach due to the similarity of the shape of different classes, and in some cases different objects possess similar positions and speed. The proposed algorithm improves the tracking results by incorporating information of the sequence in which the BBs are ordered in the same frame (frame level) and the temporal correlation between neighbouring frames (temporal level).

2.6.2. Frame level

At frame level, $l_{i,j}$ is assigned to $z_{i,j}$ based on its estimated class $c_{i,j}$ in the jth frame $\mathbf{Z}_j$ incorporating the appearance information of an insect head, i.e. the position and the order of $z_{i,j}$. As a result, an ordered collection $L_j = \{l_{1,j}, \ldots, l_{i,j}, \ldots, l_{n_j,j}\}$ is constructed, where n_j is the number of the detected objects in the jth frame.

The conditional probability $P(L_j|C_j)$ of the assignments C_j in frame j given their estimated classes L_j is computed as the fidelity of the assignment at frame level. Applying Bayes' theorem, we have

$$P(L_j|C_j) = P(L_j)P(C_j|L_j) \quad (4)$$

where $P(L_j)$ is the frequency of the sequence in which the objects are arranged, and $P(C_j|L_j)$ is the likelihood of C_j generated from the assignments L_j. They are estimated following two assumptions based on the observation that the objects maintain a consistent sequence shown in Fig. 1c, except for occasionally missing objects:

1. If the number of antenna BBs is greater than 2, or the number of mandibles BBs is greater than 2, or the number of proboscis BBs is greater than 1, $P(C_j|L_j) = 0$; otherwise, $P(C_j|L_j) = 1$.
2. If L_j is not in ascending order, $P(L_j) = 0$; otherwise, $P(L_j)$ is the likelihood of a permutation of L_j and is computed as $1/\binom{n}{n_j}$.

where n is the number of objects, i.e. $n = 5$ in the case of honey bee. $P(L_j|C_j)$ is computed following Eq. (4) and normalized over N frames. As a result, the highest $P(L_j|C_j) = 1$.

2.6.3. Temporal Level

At temporal level, the correlation between neighbouring frames is taken into account to generate the final assignment. The frames L_c with the highest conditional probability $P(L_j|C_j) = 1$ are regarded as the benchmark frames, and their less confident neighbours $L_{c\pm k}$ are updated by minimizing the pairwise linking costs between L_c and $L_{c\pm k}$. The optimal assignments are found as follows:

While $\exists L_j, j = 1, \ldots, N$ that is *not updated* do:

1. Find the L_c with the highest probability $P(L_c|C_c = 1)$.
2. The frame-to-frame linking between L_c and $L_{c\pm k}$ is found by applying the Hungarian algorithm (Munkres, 1957).
3. Update $P(L_{c\pm k}|C_{c\pm k})$ as the following scheme:
   - If the number of antenna BBs is greater than 2, or the number of mandibles BBs is greater than 2, or the number of proboscis BBs is greater than 1, $P(L_{c\pm k}|C_{c\pm k}) = 0$;
   - If $L_{c\pm k}$ is not in ascending order, $P(L_{c\pm k}|C_{c\pm k}) = 0$;
   - If there is no match of $l_{i,c\pm k}$ found in L_c, $P(L_{c\pm k}|C_{c\pm k}) = 0$;
   - Otherwise, $P(L_{c\pm k}|C_{c\pm k}) = 1$.
4. Mark L_c, $L_{c\pm k}$ as *updated*.

Output **L**.

2.7. Software implementation

A set of graphical user interfaces (GUIs) and processing algorithms were developed using Matlab with the Computer Vision System Toolbox. For those who would like to acquire a copy of the software implementation described here, further information can be obtained from the authors via email.

The user interface for the developed software is shown in Fig. 3. Users can input videos, set/adjust parameters and operate functions that implement the proposed algorithm. For example, the parameter of the top-hat (TH) filter is important for feature extraction (see Fig. 2). Users could view its influence on the filtered image in the window of Fig. 3 by selecting the region through a video player in Fig. 4, and adjusting the TH parameters through Fig. 3. The other interface in Fig. 5 is used for selecting a few training samples and its corresponding class labels within the classification (Section 2.6.1). Given the user inputs, a table of feature vectors $\mathbf{f}_{i,j}$ of selected objects and their corresponding $c_{i,j}$ is stored for training the Naïve Bayesian Classifier. Finally, for evaluating and viewing the tracking results, the label $l_{i,j}$, the BB and the tip of each object is added to the output video (see Fig. 6). The final output of the tracking procedure is the set of positions and angles for each object for each frame in Excel file format, based on which subsequent analysis is performed. The complexity of the tracking system is measured by processing time. The proposed algorithm is run using Matlab on an Intel Core i7-2600K CPU at 3.4 GHz with 16 GB RAM, the overall processing time is only about 7.5 s per frame. The main computational load comes from the feature extraction in Section 2.5, while the computations in Sections 2.6.2 and 2.6.3 are negligible (it takes 0.5 s for 10000 frames).

3. Results

We tested the proposed tracking algorithm on a set of movies of honey bee heads (*Apis mellifera*) and an ant (*Camponotus floridanus*) during odor stimulation and sugar feeding (Fig. 7). Across the different movies, the patterns of moving objects were different and such was the tracking success (Table 2).

3.1. Preprocessing

First, we show the efficacy of preprocessing operations stated in Section 2.4 in detail.

3.1.1. Exclusion of false BBs

To exclude the legs of the insect (Fig. 8A) or false BBs which are caused by reflection (Fig. 8C), a mask is obtained by segmenting the insect head (shown as the green region). Given the insect head mask, the BBs which are not attached to the mask or totally contained within it are excluded. Results with false BBs excluded are shown in Fig. 8B and D.

3.1.2. Shadow removal

To further exclude detection errors due to shadows, we applied a shadow removal algorithm provided by KaewTraKulPong and Bowden (2002). In this algorithm, the pixel is considered as a shadow if the difference in both chromatic and brightness components is within some predefined threshold. As an example, the shadow of the antenna in Fig. 8E is effectively removed by the algorithm (Fig. 8F).

3.1.3. Merging splitted BBs

This scheme is applied to merge the BBs that belong to the same objects but are detected as two distinct BBs due to the reflection (Fig. 8G). Two BBs are merged to one if they have the approximately the same angle, which is shown in Fig. 8H.

3.1.4. Splitting merged BBs

A BB including both antenna and probascis (or mandible) is split in this algorithm (shown in Fig. 8I), so that only a BB including the point "o" is considered to be split. A top-hat filter is applied to the BB to identify the antenna, and a new BB is obtained based on the result. The old BB is split into two or three BBs according to the position of the new BB, as shown in Fig. 8J.

3.2. Tracking performance

In the following we show the capability of the proposed algorithm for rectifying the incorrect classification output produced at object level (Section 2.6.1). An example is shown in Fig. 9, the classsification result at the 649th frame is $c_{1,649} = 1$, $c_{2,649} = 3$, which indicates $c_{2,649}$ is incorrectly classified as proboscis. Given $C_{649} = \{1, 3\}$ (where $C_j = \{c_{1,j}, \ldots, c_{i,j}, \ldots, c_{n_j,j}\}$), the upper BB is assigned as $l_{1,649} = 1$ indicating it is the right antenna, and the lower one is

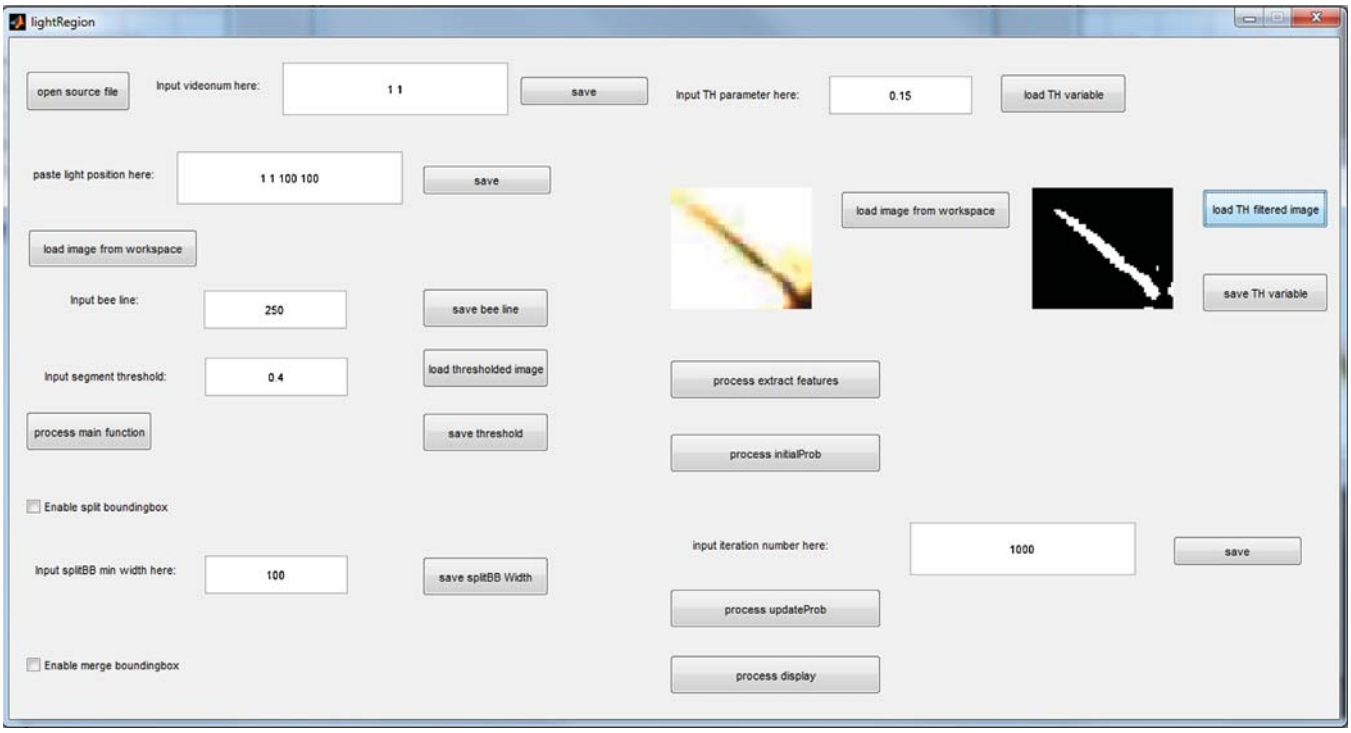

Fig. 3. A screenshot of the user interface of the software. Users can input videos, set/adjust parameters and operate functions that implement the proposed algorithm. For example, users could view its influence on the filtered image in the window by selecting the region though a video player in Fig. 4, and adjust the TH parameters.

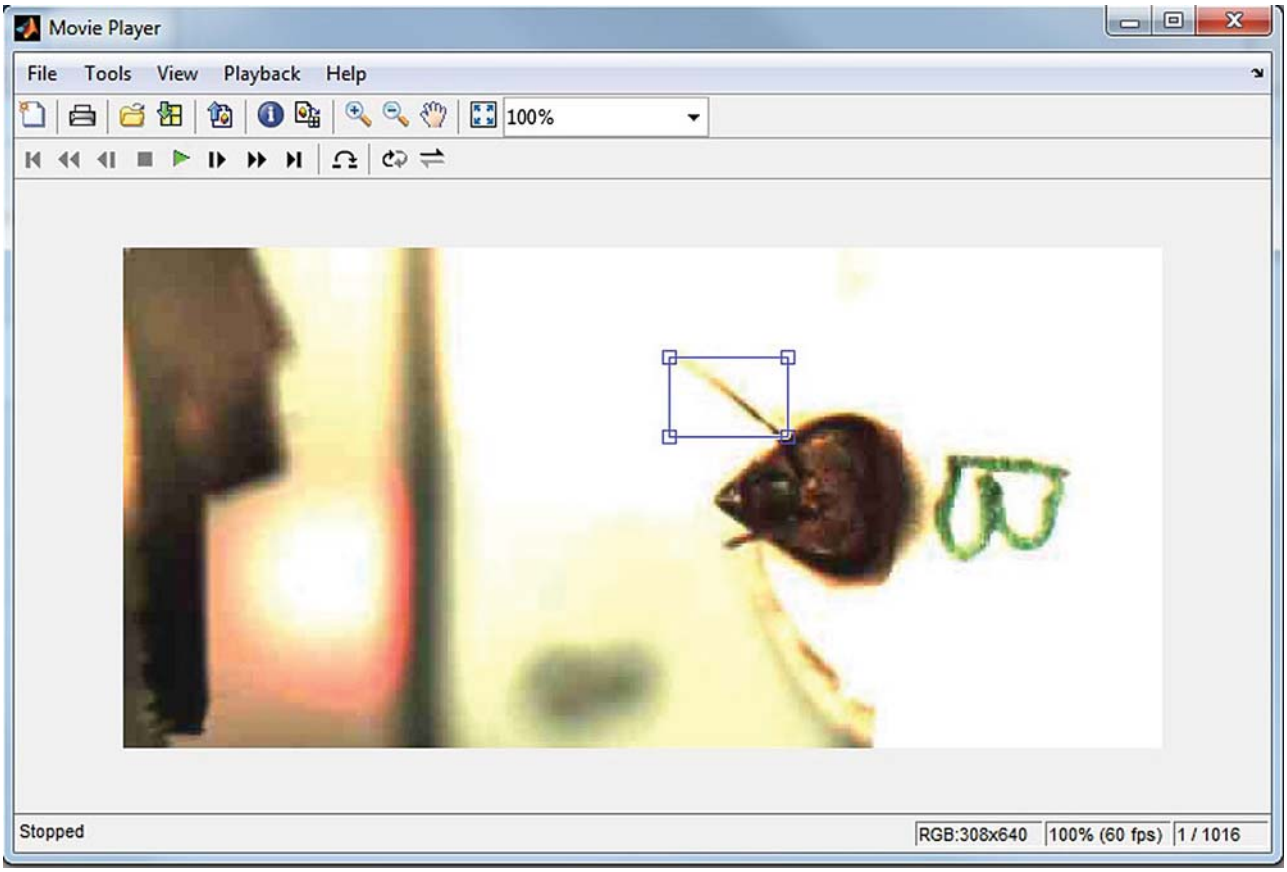

Fig. 4. A screenshot of the movie player in Matlab. The selected region could be stored and exported to other functions and GUIs.

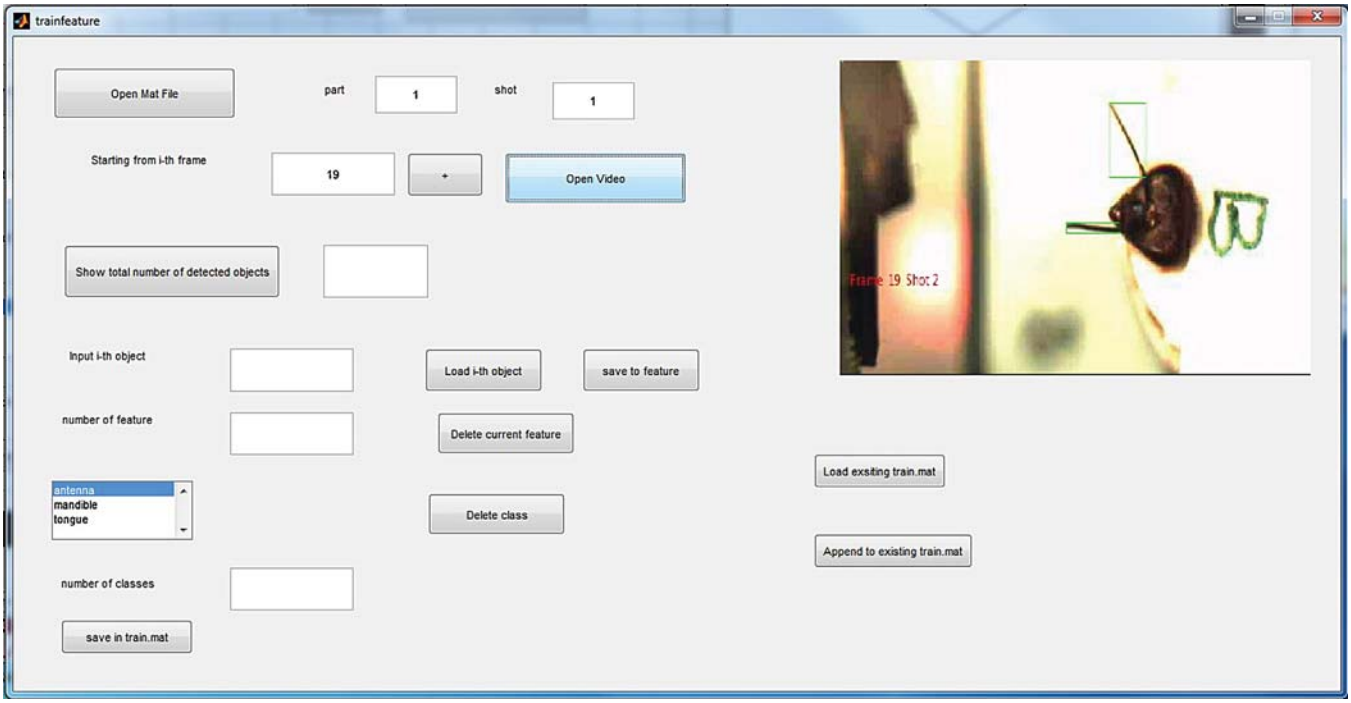

Fig. 5. A screenshot of the classification module of the software. Users can select training samples and their corresponding class labels for classification (Section 2.6.1). Given the user inputs, a table of feature vectors $\mathbf{f}_{i,j}$ of selected objects and their corresponding $c_{i,j}$ is saved for training the Naïve Bayesian Classifier.

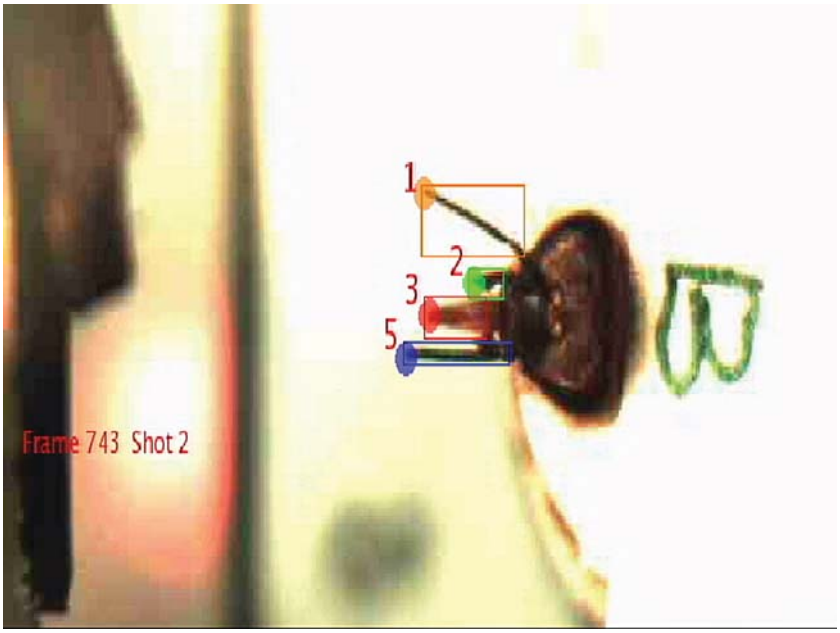

Fig. 6. An example frame with tracking labels $l_{i,j}$, the BB and the tip of each target.

Table 2
Tracking performance on six tested videos during different stimulation protocols and behaviors (Length: the number of frames N, TE: tracking errors, MD: missing detection, GT: groundtruth trajectories).

Video	Length	Animal	Stimulus or behavior		TE(%)	MD(%)	GT
1	5150	Bee1	Odor stimulation		3.9	14.1	5
2	3600	Bee2		Non-stimulated	0	20.2	2
3	3600	Bee2		Sleeping	0	0	2
4	7200	Bee3	Conditioning	Odor stimulation and feeding during classical odor-sugar conditioning	7.4	22.2	5
5	4245	Bee3		Odor stimulation during memory retention	0	0.2	5
6	4357	Ant		Odor stimulation feeding during classical odor-sugar conditioning	5.5	17.4	3

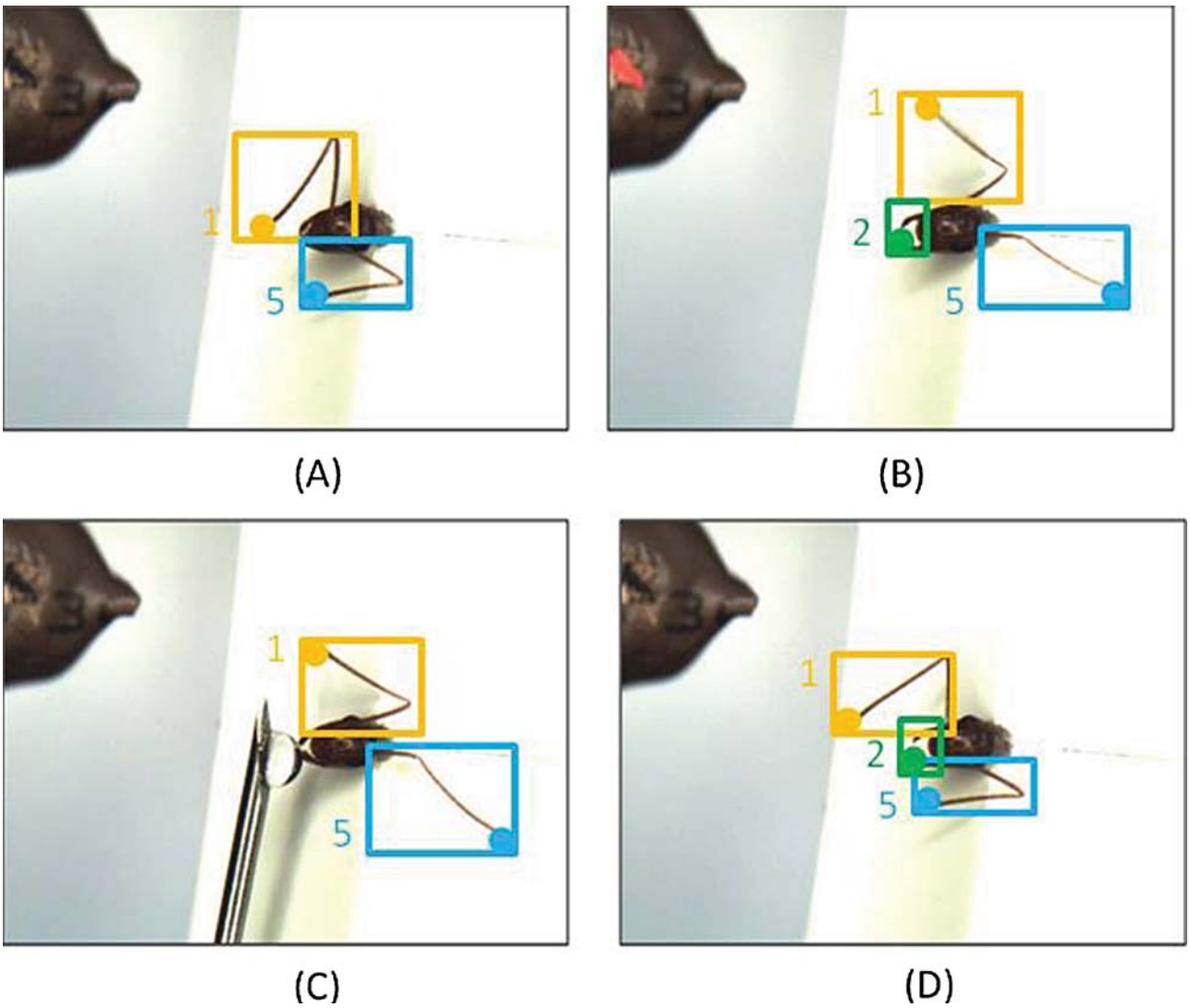

Fig. 7. Example of an ant head (A) before and (B) during odor stimulation, (C) sugar rewarding and (D) after odor stimulation. Three body parts (i.e. two antennae and mouthpart, mouthpart is detected in a single BB) of an ant are tracking targets.

assigned as $l_{1,649} = 3$ indicating it is a proboscis (see Fig. 10). According to Eq. (4), we have the following values: $P(L_{649} = \{1, 3\}|C_{649} = \{1, 3\}) = 0.5$, and $P(L_{648} = \{1, 5\}|C_{649} = \{1, 1\}) = 1$. Then L_{649} is corrected with the help of the benchmark frame L_{648} in the temporal level. As the refined result generated in the temporal level described in Section 2.6.2, L_{649} is updated as $L_{649} = \{1, 5\}$.

For measuring the overall performance of the proposed algorithm on different experiments, we manually evaluate the labels on the tested videos, as shown in Table 2. The ratio of tracking errors (TE, the number of frames containing incorrectly labeled objects) to total frames N is listed in Table 2. The description of each video is characterized by three values: Length (the number of frames N), the number of groundtruth trajectories GT (i.e. trajectories of actual moving objects) and the ratio of missing detections MD. The difficulty of tracking increases with a larger number of GT, as identity switching tends to occur more frequently. MD occurs due to severe motion blur when the antennae are moving quickly or cannot be detected when they move above the bee's head due to low contrast. A larger ratio of MD leads to a more challenging tracking problem.

It is shown in Table 2 that the tracking performance is satisfactory, since the ratio of TE to N is below 10% in all experiments. The bees did not move when they are asleep, thus both TE and MD are zero. For Videos 4 and 6, the sugar stick used for feeding disturbs the background model, thus producing significantly higher TE and MD.

3.3. Behavioral analysis

We tested the tracking algorithm on four videos of three bees' heads and one ant head, and tracked the movement of their proboscis, antennae and mandibles (Table 2). One bee was recorded during odor stimulation, a second bee was recorded during sleep, and a third bee was recorded during classical conditioning and memory retention. The ant was recorded during classical conditioning. The tracking performance differed between the videos; the tracking error rate ranged from 0 to 7.4%, the missed detections rate ranged from 0 to 22.2 %, as shown in Table 2.

We then used the tracking data to evaluate different behavioral monitors during associative odor-sugar learning of an individual bee (Fig. 11A). We trained a honey bee to extend its proboscis to an odor by pairing this odor with a sugar reward (Szyszka et al., 2011). The training consisted of 10 trials which were spaced by 11 min. During each training trial the bee received a 6 s long odor stimulus (1-hexanol) and a 3 s long sugar reward which started 5 seconds after odor onset. In a memory test 30 min after the training the bee was stimulated with the trained odor and a novel odor (1-nonanol).

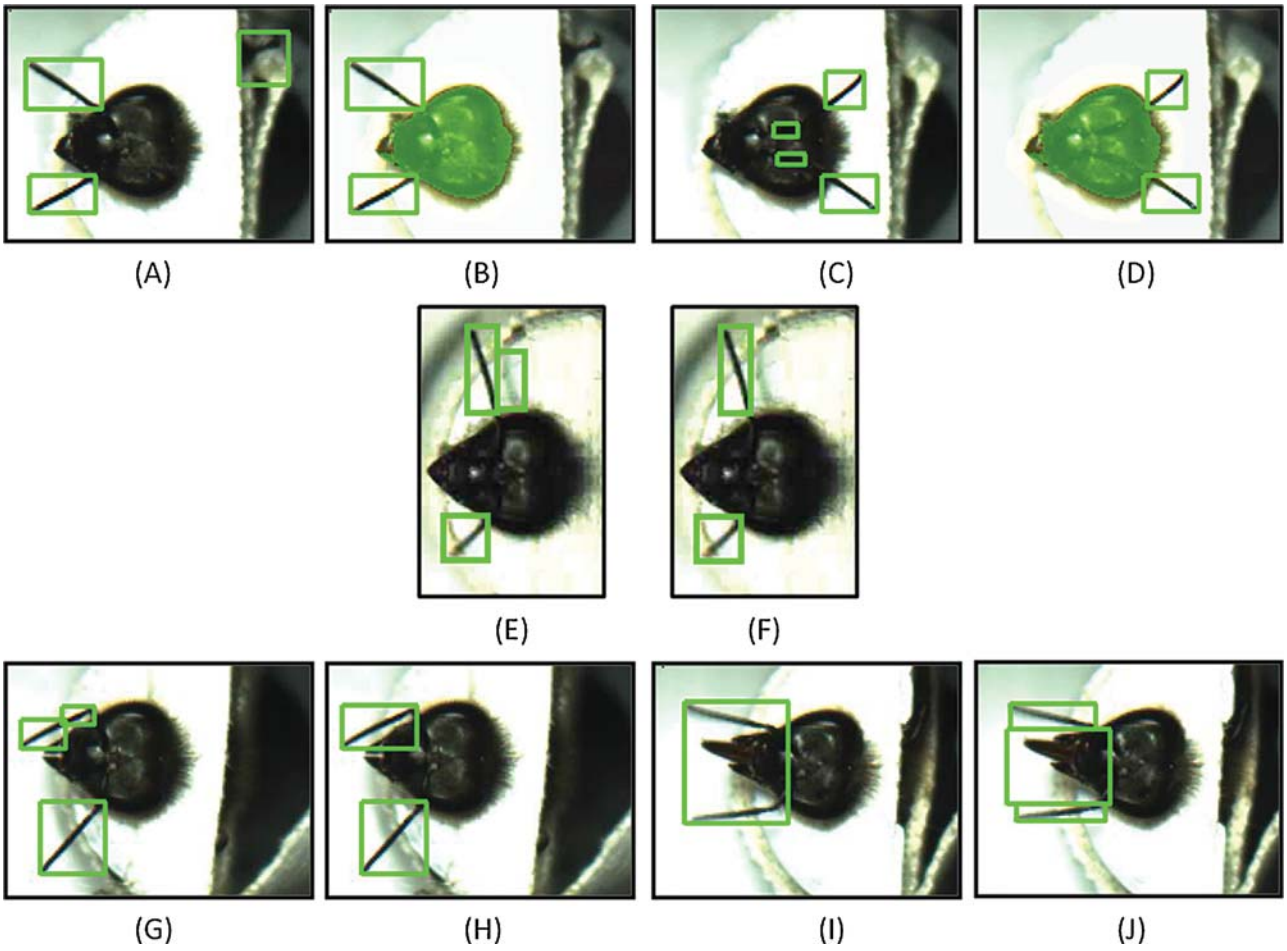

Fig. 8. Examples of excluding false measurements (A) and (C) and the results after exclusion (B) and (D); shadow removal (E) and (F); merging split measurements (G) and (H); splitting merged mearsurements (I) and (J).

The common way to analyze the bees' performance during this paradigm is to note whether it extends the proboscis during the odor stimulus (but before the sugar stimulus) in anticipation of the sugar reward (Matsumoto et al., 2012). This monitor yields binary data: "1" for proboscis extension response, "0" for no response. The bee started responding to the trained odor during the third training trial and continued responding during subsequent trials. During the test it responded to the trained but not to the novel odor, indicating that it formed an odor specific associative odor-sugar memory. This stable behavioral performance in individual bees is typical: once bees start responding during training they continue to respond (Pamir et al., 2011, 2014). However, it is currently unclear whether this abrupt behavioral performance change reflects abrupt learning or whether learning is a more gradual process (Gallistel et al., 2004; Pamir et al., 2011, 2014). In fact, this abrupt behavioral performance change might be due to the binary monitor of the proboscis extension response which does not allow monitoring gradual changes in behavior. Therefore, we analyzed other graded parameters which we extracted from the videos. The onset of movements of the proboscis (Fig. 11C), for example started already during the second trial, while the full proboscis extension response started during the third trial (Fig. 11B). The onset of the proboscis movement occurred three seconds after odor onset and became shorter during the third and fourth trial. Thus, during training there is a gradual behavioral change in the odor response which is not detectable in the binary proboscis extension response (Fig. 11B). During the test the proboscis movement onset occurred earlier for the trained than for the novel odor. The proboscis movement response to the novel odor indicates that the bee partly generalized the learned response to the novel odor. This information is lost in the binary proboscis extension response (Fig. 11B). Similarly the elongation of the proboscis (Fig. 11C) shows a gradual change during learning and memory test, as it progressively increases for the trained odor. Bees constantly move their antennae, both in the absence and presence of odor stimuli. We asked whether the mean pointing direction

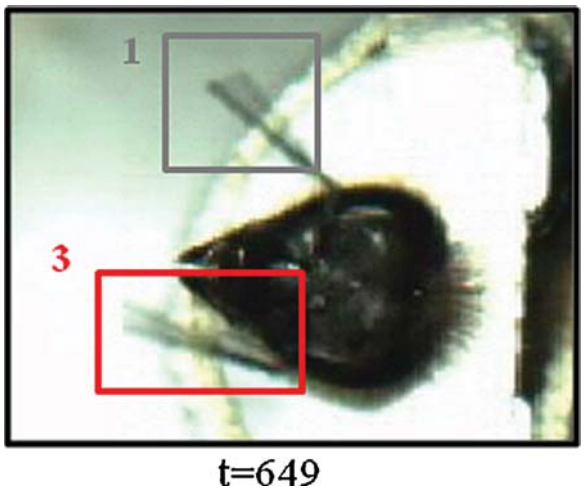

Fig. 9. Sample frame $t = 649$: Example of false classification. The left antenna is incorrectly classified as proboscis, the classes of two detections are $c_{1,649} = 1$, $c_{2,649} = 3$.

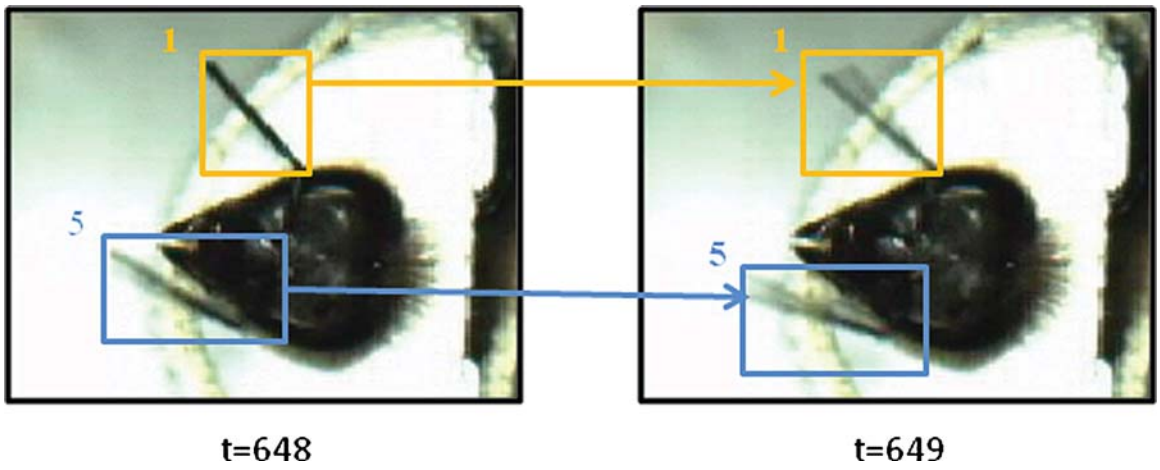

Fig. 10. Illustration of temporal level: The conditional probability of the frame 648 and frame 649 are $P(L_{648} = \{1, 5\}|C_{649} = \{1, 1\}) = 1$ and $P(L_{649} = \{1, 3\}|C_{649} = \{1, 3\}) = 0.5$, respectively. The labels are updated according to the frame-to-frame linking between L_{649} and the benchmark L_{648}. The label $l_{2,649}$ is corrected as 5.

of the antennae differ during the absence and presence of an odor and whether there is a change in pointing direction during training (Fig. 11E). Before odor stimulation the mean angle of the bee's antenna was around 32°. During the first odor presentation before receiving the sugar stimulus in trial 1 the bee moved its antennae backwards. During the following training trials the bee pointed the antennae further and further forwards; however, during the memory test there was no difference in the pointing direction between the trained and the novel odor. Next we asked whether odor stimulation and odor-sugar training changes the correlation between the left and right antenna movements (Fig. 11F). We quantified antenna movement correlations by calculating the Pearson correlation between the angles of both antennae during four seconds before or during odor onset. Correlated forward-backward movements of both antennae would yield positive correlation values; correlated left-right movements would yield negative correlation values. Antenna movements were generally negatively correlated (correlated left-right movements). During odor stimulation the bee exhibited fewer correlated left-right movements than before odor stimulation. However, there was no apparent change in correlation in the course of the training. Taken together, the tracking data of antennae and proboscis provide a gradual measure of behavioral performance in individual bees. These behavioral monitors could allow detecting and quantifying gradual changes in behavioral performance in individuals which would not be accessible using the binary proboscis extension response.

4. Discussion

We presented a novel computer vision system for the automated analysis of behavior of restrained insects, by tracking their antennae and mouthparts.

4.1. Comparison to existing tracking systems

Automatically tracking the movement of insects provides researchers quantitative information about the movement of the insect body or body parts such as antennae and mouthparts, thus it allows for a more fine-grained analysis and paves the way to address open questions in behavioral insect studies. However, new challenges arise from the specific requirements of the biological experiments, and addressing them by simply applying existing generic image/video processing algorithms leads to suboptimal results.

There has been intensive work on tracking objects in video sequences. However, most of these algorithms do not directly adapt well to tracking insects, which exhibit very specific forms of motion. Some existing research on tracking insect bodies (e.g. bee dance Veeraraghavan et al. (2008), Landgraf and Rojas (2007), ants Balch et al. (2001), Ying (2004)) and body parts (e.g. bees' antennae Hussaini et al. (2009), Mujagić et al. (2011), mice's whiskers (Voigts et al., 2008)) has been reported recently. A method for antennae tracking is proposed by Hussaini et al. (2009), but it requires initial manual labelling for each video. In another recent work by Mujagić et al. (2011), the movements of antennae are tracked by selecting the two largest clusters only. In both Hussaini et al. (2009) and Mujagić et al. (2011), mandibles and proboscis are not considered, which make them not applicable for our study.

Many state-of-art tracking approaches estimate the posterior distribution of the position of the object in the current frame by using a Particle Filter (Zhou et al., 2004; Khan and Dellaert, 2004), and some studies also exploit its usage in insect tracking (Veeraraghavan et al., 2008; Landgraf and Rojas, 2007; Ying, 2004). For example, the algorithm proposed by Veeraraghavan et al. (2008) tracks a single bee using Particle Filtering to maintain its identity throughout the video sequence. However, as pointed out by Perera et al. (2006), Particle Filtering is often only effective for short tracking gaps and the search space becomes significantly larger for long gaps. This is applicable only for the videos that were captured by a high-speed camera (Voigts et al., 2008; Petrou and Webb, 2012) or inlude slow moving objects (Balch et al., 2001). The main problem of our videos is that the tracking gap of each moving object is relatively long due to the lower frame-rate, while the antennae move rather fast. Another problem is that antennae cannot be detected when they move above the head due to the low contrast. The mandibles and proboscis move infrequently, thus their tracklets are short. The resulting gaps give rise to an issue similar to long gaps in that the frame-rate of the recorded videos is usually low and thus the potential matches on the far side of the gap are difficult to predict. Similarly, the algorithm proposed by Balch et al. (2001) is not able to tackle the tracking gap, which tracks multiple ants by merely applying data association techniques. Moreover, the detection errors produced by typical moving object detectors such as false, missing, split or merged measurements increase the difficulties of assigning correct identity and maintaining identity. In Voigts et al. (2008), a statistical model is used to assign each whisker of a mice the most probable identity under the constraint that whiskers are ordered along the face. Inspired by Voigts et al. (2008), we construct a Bayesian algorithm in this paper, which computes the

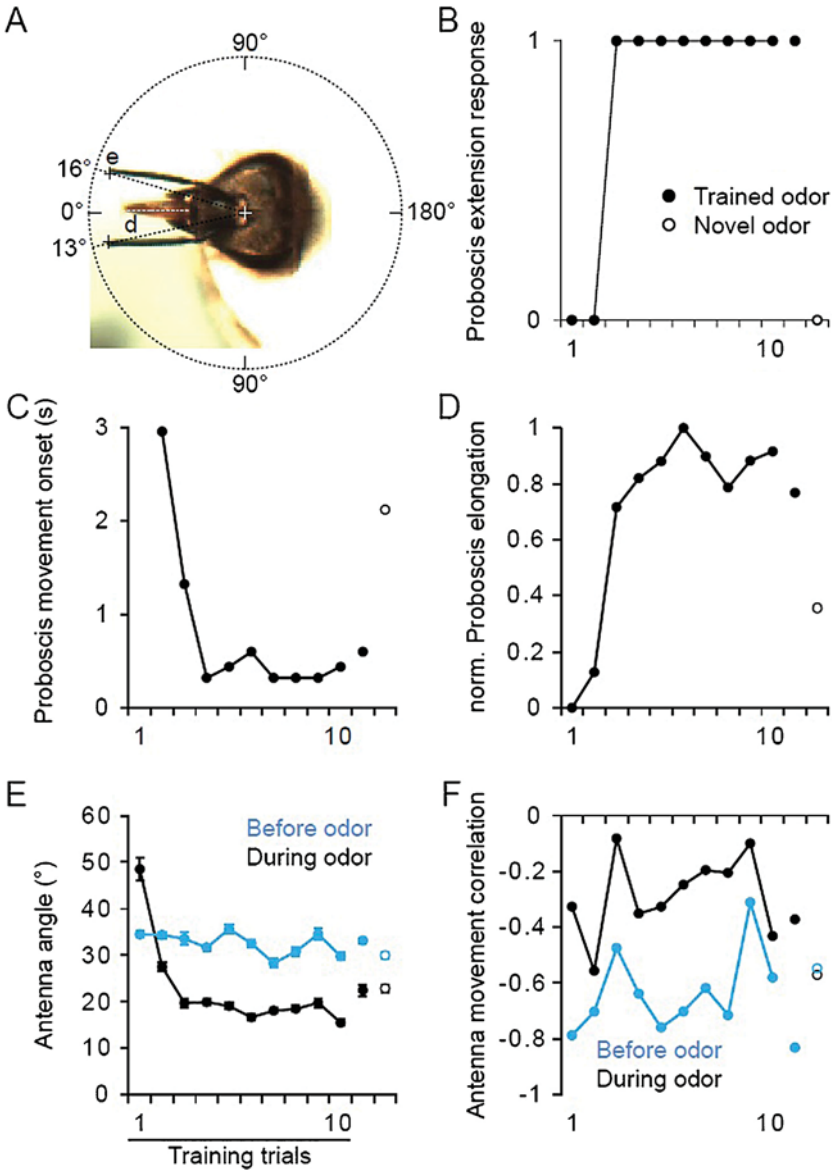

Fig. 11. Antenna and probscis movements reveal gradual performance changes during odor learning. Behavioral performance of a single bee during associative odor-sugar learning (trials 1–10) and memory test with the trained odor (trial 11) and a novel odor (trial 12). (A) Imaged bee head. The parameter "antenna angle" (E) reads as follows: 0°: Antenna is pointing straight forward; 180°: antenna points straight backward; negative values: Antenna crosses the midline. The parameter "proboscis elongation" (D) shows the length of the proboscis normalized to the maximum length. (B) Behavioral performance monitored as binary proboscis extension response (full extension). The bee started responding to the trained odor during the 3rd training trial and continued responding to it throughout the training and test. (C) Onset latency of the proboscis movement. The onset latency decreased from the 2nd to 4th trial. During the test the onset latency was shorter for the trained odor than for the novel odor. (D) Elongation of the proboscis during odor stimulation (mean during the initial 4 s of the odor stimulation). (E) Antennae angle during 4 s before (blue) and during odor stimulation (mean ± SEM). Angles of the left and right antennae were averaged. (F) Correlation between the angular antenna movements of the left and right antennae during 4 s before (blue) and during odor stimulation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

probability of the assignments of objects in each video frame, given the estimation of their classes (antennae, mandible or proboscis) using a Naïve Bayesian Classifier. The proposed algorithm exploits temporal correlation between benchmark frames with high probability and their less confident neighbors, and generates the most probable labels. We verify the efficacy of the proposed system on different type of bees' behavioral experiments.

4.2. Possible applications

Our tracking system provides quantitative measures of the movements of multiple body parts of restrained insects. This behavioral read-out allows us to obtain a graded performance measure for individual insects and opens up a method to overcome problems of traditional experimental procedures and to address novel questions in insect behavioral research.

For example, the common experimental procedure of pooling binary performance measures of groups of identically treated animals often confounds the interpretation of behavioral data, as the group average is not representative for all individuals (Gallistel et al., 2004; Pamir et al., 2011, 2014). Our approach helps to overcome this problem, as it allows the analysis of graded behavior in individuals.

To give another example: The observation that once having responded for the first time, honey bees continue to respond with a high probability during training and memory test could suggest that learning results in abrupt performance changes (Pamir et al., 2011, 2014). However, learning related gradual changes might

exist, and might have been masked by the binary behavioral read out. Thus, our tracking approach can help to reveal the dynamic of behavioral performance changes within individuals.

Finally, our tracking system might help to investigate how the individual learning and memory performance depends on training parameters, genetic factors and internal states, such as arousal and attention.

Acknowledgements

Thanks to Oliver Kühn, Christopher Dieter Reinkemeier and Manuel Wildner for help with the behavioral experiments and software evaluation. This work was funded by Bundesministerium für Bildung und Forschung (01GQ0931 to PS and CGG), with partial support also from the National Natural Science Foundation of China under Grant No. 61302121.

References

Balch T, Khan Z, Veloso M. Automatically tracking and analyzing the behavior of live insect colonies. In: Proceedings of the fifth international conference on autonomous agents. ACM; 2001]. p. 521–8.

Bitterman M, Menzel R, Fietz A, Schäfer S. Classical conditioning of proboscis extension in honeybees (*Apis mellifera*). J Comp Physiol 1983];97(2):107.

Erber J, Pribbenow B, Bauer A, Kloppenburg P, et al. Antennal reflexes in the honeybee: tools for studying the nervous system. Apidologie 1993];24(3):283–96.

Gallistel CR, Fairhurst S, Balsam P. The learning curve: implications of a quantitative analysis. Proc Natl Acad Sci U S A 2004];101(36):13124–31.

Gil M, Menzel R, De Marco RJ. Side-specific reward memories in honeybees. Learn Memory 2009];16(7):426–32.

Hussaini SA, Bogusch L, Landgraf T, Menzel R. Sleep deprivation affects extinction but not acquisition memory in honeybees. Learn Memory 2009];16(11):698–705.

KaewTraKulPong P, Bowden R. An improved adaptive background mixture model for real-time tracking with shadow detection. In: Video-based surveillance systems. Springer; 2002]. p. 135–44.

Landgraf T, Rojas R. Tracking honey bee dances from sparse optical flow fields. FB Mathematik und Informatik FU 2007]:1–37.

Matsumoto Y, Menzel R, Sandoz J-C, Giurfa M. Revisiting olfactory classical conditioning of the proboscis extension response in honey bees: a step towards standardized procedures. J Neurosci Methods 2012];211(1):159–67.

Menzel R. The honeybee as a model for understanding the basis of cognition. Nat Rev Neurosci 2012];13(11):758–68.

Mujagić S, Würth SM, Hellbach S, Dürr V. Tactile conditioning and movement analysis of antennal sampling strategies in honey bees (*Apis mellifera* L.). J Vis Exp 2011];70:e50179.

Munkres J. Algorithms for the assignment and transportation problems. J Soc Ind Appl Math 1957];5(1):32–8.

Pamir E, Chakroborty NK, Stollhoff N, Gehring KB, Antemann V, Morgenstern L, Felsenberg J, Eisenhardt D, Menzel R, Nawrot MP. Average group behavior does not represent individual behavior in classical conditioning of the honeybee. Learn Memory 2011];18(11):733–41.

Pamir E, Szyszka P, Scheiner R, Nawrot MP. Rapid learning dynamics in individual honeybees during classical conditioning. Front Behav Neurosci 2014];8:313.

Perera AA, Srinivas C, Hoogs A, Brooksby G, Hu W. Multi-object tracking through simultaneous long occlusions and split-merge conditions. In: 2006 IEEE computer society conference on computer vision and pattern recognition, vol. 1. IEEE; 2006]. p. 666–73.

Petrou G, Webb B. Detailed tracking of body and leg movements of a freely walking female cricket during phonotaxis. J Neurosci Methods 2012];203(1):56–68.

Rehder V. Quantification of the honeybee's proboscis reflex by electromyographic recordings. J Insect Physiol 1987];33(7):501–7.

Sauer S, Kinkelin M, Herrmann E, Kaiser W. The dynamics of sleep-like behaviour in honey bees. J Comp Physiol A 2003];189(8):599–607.

Smith BH, Abramson CI, Tobin TR. Conditional withholding of proboscis extension in honeybees (apis mellifera) during discriminative punishment. J Comp Physiol 1991];105(4):345.

Szyszka P, Demmler C, Oemisch M, Sommer L, Biergans S, Birnbach B. Mind the gap: olfactory trace conditioning in honeybees. J Neurosci 2011];31:7229–39.

Veeraraghavan A, Chellappa R, Srinivasan M. Shape-and-behavior encoded tracking of bee dances. IEEE Trans Pattern Anal Mach Intell 2008];30(3):463–76.

Voigts J, Sakmann B, Celikel T. Unsupervised whisker tracking in unrestrained behaving animals. J Neurophysiol 2008];100(1):504–15.

Ying F. Visual ants tracking. University of Bristol; 2004], Ph.D. thesis.

Yu J, Amores J, Sebe N, Tian Q. A new study on distance metrics as similarity measurement. In: 2006 IEEE international conference on multimedia and expo. IEEE; 2006]. p. 533–6.

Khan Z, Dellaert F. TB. A rao-blackwellized particle filter for eigen tracking. In: Proceedings of IEEE international conference on computer vision and pattern recognition; 2004].

Zhou SK, Chellappa R, Moghaddam B. Visual tracking and recognition using appearance-adaptive models in particle filters. IEEE Trans Image Process 2004];13(11):1491–506.