Erschienen in: ICIP 2013 : IEEE International Conference on Image Processing, Sept. 15 - 18, 2013, Melbourne, Australia / International Conference on Image Processing <2013, Melbourne>. - Piscataway, NJ : IEEE, 2013. - S. 4112 - 4116. - ISBN 978-1-4673-2532-5

AUTOMATIC FRAMEWORK FOR TRACKING HONEYBEE'S ANTENNAE AND MOUTHPARTS FROM LOW FRAMERATE VIDEO

Minmin Shen[⋆] *Paul Szyszka*[†] *C. Giovanni Galizia*[†] *Dorit Merhof*[⋆]

[⋆] INCIDE Center, University of Konstanz
[†] Institute of Neurobiology, University of Konstanz

ABSTRACT

Automatic tracking of the movement of bee's antennae and mouthparts is necessary for studying associative learning of individuals. However, the problem of tracking them is challenging: First, the different classes of objects possess similar appearance and are close to each other. Second, tracking gaps are often present, due to the low frame-rate of the acquired video and the fast motion of the objects. Most existing insect tracking approaches have been developed for slow moving objects, and are not suitable for this application. In this paper, a novel Bayesian framework is proposed to automatically track bees' antennae and their mouthparts. This framework incorporates information about their kinematics, shape, order and temporal correlation between neighboring frames. Experimental evaluation demonstrates the effectiveness and efficiency of the proposed framework.

Index Terms— multi-target tracking, bee antennae and mandibles and proboscis, splitted detections, merged detections

1. INTRODUCTION

Honeybees are a powerful model to study the neuronal mechanisms of learning and memory. 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 learnt the association between odorant and sugar is usually assessed by its proboscis extension reflex [1]. The proboscis is the tongue of a bee, and is extended reflexively when the bee is stimulated with sugar water or with a previously conditioned odorant. We imaged honeybees' heads during a learning and memory task and analyzed the movement of their proboscis, antenna and mandibles (each object is shown in Fig. 1a). Automatic tracking of the movement of bee's antennae, mandibles and proboscis provides quantitative information about antenna and proboscis movement, thus enables a more fine-grained analysis of learning and memory performance and opens the way to new questions in behavioural insect studies. Tracking of insects antennae and mouthparts is a difficult problem, due to the low frame-rate of the acquired video (30 frames/s), motion blur resulting from the high speed of movement, and the complicated motion model of the bee's antenna. In this paper, a Bayesian-based framework is proposed to automatically track the movements of individual bee's antennae, mandibles and proboscis.

Fig. 1. Example of bee video and the segmented bee head.

Only few methods addressing this problem are reported in the literature. A method for antennae tracking is proposed in [2], but it requires initial manual labelling for each video. In recent work [3], the movements of antennae only are tracked by selecting the two largest clusters, but mandibles and proboscis are not considered. Some research on tracking bees uses particle filtering to maintain identity through video sequence [4], which is not applicable here. As pointed out in [5], particle filtering is often only effective over short tracking gaps and the search space becomes significantly larger over long gaps. The main problem of our data is that the tracking gap of each moving object is relatively long given the low frame-rate. The antennae move rather fast, and they cannot be detected when they move above the bee's head due to the low contrast. The mandibles and proboscis move infrequently, thus their tracklets are short. The resulting gaps incur 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. Moreover, the detection errors produced by typical moving object detectors such as false, missing, splitted or merged measurements increase the difficulties of assigning correct identity and maintaining identity.

In this paper, we propose a novel tracking framework that incorporates prior information about the kinematics and shapes of antennae, mandibles and proboscis. The overall tracking framework consists of three levels: object level,

frame level and temporal level. At object level, a Naive Bayesian Classifier is used to compute the prior probability that an object is identified as antenna, mandibles or proboscis. At frame level, the identification of each object is assigned according to the sequence in which the objects 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. The experimental results show that the proposed framework is capable of reliably detecting and tracking individual bee's antennae, mandibles and proboscis.

2. PRELIMINARIES

Each individual bee is imaged using a "FMVU-03MTM/C" firefly camera which is capable of acquiring a video at 30 frames/s, in order to record the head with proboscis, mandibles and antennae. Stimulus delivery (odour) is monitored by lighting an LED within the field of view of the camera, so that data analysis can be done relative to stimulus delivery. Sixteen bees are harnessed on a platform, with their head in fixed positions, but able to move antennae and mouthparts freely. For each trial, the position of the platform is adjusted manually. The camera is set on top of an individual bee. The camera is fixed, and the platform where the bees are fixed is moved when changing a new bee for recording. Unlike the high speed camera used in [6], which is capable of capturing video at 500 frames/s, the frame-rate of the acquired bee movies in this paper is only 30 frames/s. Although it would be possible to record with a high speed camera, we aim at developing a framework that uses affordable cameras such as web-cam or consumer level cameras, which keeps the data volume low. Each video is about 30 minutes long and consists of four trials, for which each has 16 individual honeybees. Thus, a single video to be processed is approximately 30 s and the frame size is 480×640 pixels.

To extract the information of the relative position of each object, it is required to set up the coordinate system. As the base is not static during the changing of bees, the scene change is detected to ensure a static background before the actual tracking procedure starts. After the background has stabilized, the mean of the first 10 frames is used to estimate the bee head's position. After thresholding, a dark region with the greatest circularity value and an area within the range of [1000, 8000] pixels is selected as the segmented bee head. Then, the position of the mandibles is estimated as the most left point of the segmented bee head (as shown in Fig. 1b). With the mandible position (marked as point "o") and the centroid of the bee's 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.

3. TRACKING FRAMEWORK

3.1. Object detection

For detecting moving objects, Gaussian Mixture Model (GMM) background modelling [7] is used. The object detector generates an unordered set of false measurements (e.g. shadows, reflection and bee's legs), missing measurements (motion blurred antennae or antennae above the head), splitted measurements (splitted bounding boxes of the same antenna), merged measurements (one bounding box including two or three objects) as shown in Fig. 3, which make the following tracking task difficult. Therefore, pre-processing operations include shadow removal [7], exclusion of undesired objects by incorporating position information, merging splitted measurements, and splitting merged measurements. These pre-processing operations greatly reduce the undesired detection measurement, but some false, missing, splitted and merged measurements may still remain. Thus the tracking algorithm is required to tackle this problem.

3.2. 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, texture and speed, where n_j is the number of the detected objects in $\mathbf{Z}_j$ and N is the number of frames. To represent the position of each bounding box, the vertices nearest or furthest to point "c" are extracted. Whether point "o" is within the bounding box is also included in the feature vector to identify the object, as the bounding box of an antenna seldom includes point "o". The shape of each object is indicated by its area. Also the top-hat filter is used as a ridge detector to identify an antenna: after thresholding and grayscale reversion, the top-hat filter is performed on the image block within its bounding box. However, an antenna with severe motion blur cannot be detected by the top-hat filter (e.g. the left antenna in Fig. 1a). 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 [8]. The seven features used to represent the appearance model are: distance between the nearest vertex and mandible $\mathbf{f}_{i,j}(1)$, distance between the furthest vertex and the tongue line $\mathbf{f}_{i,j}(2)$, 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 indicating whether the mandible is within the bounding box $\mathbf{f}_{i,j}(7)$.

3.3. Object Level

The objective of the proposed tracking algorithm is to assign each object $z_{i,j}$ with labels $l_{i,j}$, where $l_{i,j} \in \{$1:right antenna; 2:right mandible; 3:proboscis; 4:left mandible; 5:left antenna; 6:false positive$\}$.

At object level, the probability $P(c_{i,j}|\mathbf{f}_{i,j})$ of each object $z_{i,j}$ for each class $c_{i,j}$ (where $c_{i,j} \in \{1$:antenna; 2:mandible; 3:proboscis$\}$) is computed given its feature vector $\mathbf{f}_{i,j}$. Among the seven features, $\mathbf{f}_{i,j}(1), \ldots, \mathbf{f}_{i,j}(3)$ are assumed to follow a Gaussian distribution whose means and standard deviations are learned from the training set, i.e. a set of annotated objects. Let us pack the three features into a vector and denote it as $\bar{\mathbf{f}}_{i,j} = [\mathbf{f}_{i,j}(1), \ldots, \mathbf{f}_{i,j}(3)]^T$. Thus the conditional probability is

$$P(c_{i,j}|\bar{\mathbf{f}}_{i,j}) = \frac{1}{(2\pi)^{3/2}|\mathbf{C}_{i,j}|^{1/2}} exp[-\frac{1}{2}(\bar{\mathbf{f}}_{i,j} - \mathbf{u}_{i,j})^T \cdot \mathbf{C}_{i,j}^{-1}(\bar{\mathbf{f}}_{i,j} - \mathbf{u}_{i,j})] \quad (1)$$

where $\mathbf{u}_{i,j}$ and $\mathbf{C}_{i,j}$ are the mean vector and the covariance matrix of $\bar{\mathbf{f}}_{i,j}$, respectively.

The other features $\mathbf{f}_{i,j}(4), \ldots, \mathbf{f}_{i,j}(7)$ are modeled as discrete variables with constant prior probabilities assumed to be known. The class-conditional probability density function 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, e.g. $P(\mathbf{f}_{i,j}(6) > 0|c_{i,j} = 1) = 1$, since an antenna must have top-hat filtered pixels, while $P(\mathbf{f}_{i,j}(6) > 0|c_{i,j} = 3) = 0.05$, since the probability that a bounding box of proboscis may include antennae or its reflection, which may also pass through the top-hat filter, is rather low. The other unknowns of Eq. 2 can be obtained by solving the equations combining the constraint that each object must be antenna, mandible or proboscis:

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

Given estimates for $P(c_{i,j}|\mathbf{f}_{i,j})$, a Naive Bayesian Classifier is performed for each object 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 position and speed. The proposed framework improves the tracking results by incorporating information of the sequence in which the objects are ordered in the same frame (frame level) and the temporal correlation between neighbouring frames (temporal level).

3.4. 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 a bee, i.e. the position and the order of $z_{i,j}$. As a result, an ordered collection $L_j = \{l_{1,j}, \ldots, l_{i,j}, \ldots, l_{n_j}\}$ is constructed, where n_j is the number of the detected objects in the jth frame. $P(L_j|C_j)$, where $C_j = \{c_{1,j}, \ldots, c_{i,j}, \ldots, c_{n_j,j}\}$ is computed as

$$P(L_j|C_j) = \begin{cases} 0 & \text{if } \|c_1\| > 2 \text{ or } \|c_2\| > 1 \text{ or } \|c_3\| > 2 \\ & \text{or } \exists l_{i,j} > l_{k,j}, \forall k < i \\ 1 & \text{if } L_j = \{1, 5\} \text{ or } L_j = \{2, 4\} \\ \binom{5}{n_j} & \text{otherwise} \end{cases} \quad (4)$$

where $\|c_i\|$ is the number of c_i.

3.5. Temporal Level

At temporal level, the temporal 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 most confident frames, and their less confident neighbours $L_{c\pm1}$ are updated by minimizing the pairwise linking costs between L_c and $L_{c\pm1}$. The optimal assignment for all frames $\widehat{\mathbf{L}} = \{\widehat{L_j}, j = 1, \ldots, N\}$ is found by solving the optimization problem

$$\widehat{\mathbf{L}} = \arg\max_{\mathbf{L}} \prod_{c \in \Psi} P(L_{c\pm1}|L_c) \quad (5)$$

where Ψ is a set of frames such that $c \in \Psi : P(L_j|C_j) = 1 \& P(L_{j\pm1}|C_{j\pm1}) \neq 1$.

To solve Eq. 5, the assignment of each frame and the corresponding probability is updated as follows:

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

- Find the L_c with the highest probability $P(L_c|C_c = 1)$.
- The frame-to-frame linking between L_c and $L_{c\pm1}$ is found by forming a $n \times n$ cost matrix $M = \{M_{i,j}\}$ with

$$\begin{aligned} M_{i,j} &= -logP(l_{i,c} \mapsto l_{j,c\pm1}) \\ &= \left\|\bar{\mathbf{f}}_{i,c} - \bar{\mathbf{f}}_{j,c\pm1}\right\| \end{aligned}$$

  where $n = \max(n_c, n_{c\pm1})$ and the sign "$\mapsto$" denotes a correspondence. The Hungarian algorithm [9] is applied to find the optimal linking.

- Update $P(L_{c\pm1}|C_{c\pm1})$ as

$$P(L_j|C_j)^t = \begin{cases} 0 & \text{if } \|c_1\| > 2 \text{ or } \|c_2\| > 1 \text{ or } \|c_3\| > 2 \\ & \text{or } \exists l_{i,c\pm1} > l_{k,c\pm1}, \forall k < i \\ & \text{or } \exists l_{i,c\pm1} \text{ is null match} \\ 1 & \text{otherwise} \end{cases}$$

- If $P(L_{c\pm1}|C_{c\pm1}) = 1$, update $L^t_{c\pm1}$ by the linking of the Hungarian algorithm. Otherwise, $L^t_{c\pm1} = L^{t-1}_{c\pm1}$.
- Mark $L_c, L_{c\pm1}$ as *updated*.

Output $\widehat{\mathbf{L}}$.

4. EXPERIMENTAL RESULTS

In our experiments, 16 bee videos are used to test the performance of the proposed framework. "bee C", "bee E", "bee G", "bee L" and "bee M" are selected as representative testing videos because they are the most challenging cases. The appearance model is trained on 40 manually annotated antennae, 10 mandibles, and 18 proboscis from the video "bee E".

To evaluate the proposed framework, we manually generated ground truth tracks for each object. The tracking performance is measured in two ways: the no. of tracking errors **TE** (the no. of incorrectly labeled objects) defined in [10] and the total no. of times that a tracked trajectory changes its matched ground truth identity **IDS** defined in [11]. The description of each video is characterized by five values: the no. of frames N, the no. of groundtruth trajectories **GT**, the detection errors after pre-processing operation described in Sec. 3.1 including the no. of splitted detections **SD**, the no. of merged detections **MD**, the no. of false detections **FD**.

The complexity 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, and the overall processing time is only about 7.5 s per frame, while the computation in Sec. 3.4 and Sec. 3.5 takes 0.5 sec for all the videos (10781 frames in total). Table 1 shows the results of data association method used in [12] (implemented using Hugarian algorithm), Naive Bayesian Classifier (**NB**) and the final outputs on the tested videos of bee C, E, G, L, M. It is seen that the method in [12] cannot handle the case of variable number of moving objects, and is not robust to detection errors. Moreover, the final results significantly improve the tracking performance over **NB** even in "bee M", which is the most challenging case (**GT**=5).

To show the pattern of movement, the position of the tip (which is estimated as the furthest point of the object to the point "c") of each identified object by the proposed framework is demonstrated in Fig. 2 for "bee L". The proposed framework accurately tracks each object despite of several long tracking gaps for the left antenna (magenta), the longest amounting to about 50 frames.

Illustrative examples of the tracking procedure from bee C, G and M are shown in Fig. 3. The first row shows the moving regions output (white regions) by the detector. The second row shows the tracking results with a bounding box around the moving object and the corresponding label $l_{i,j}$. The tip of each object is marked by "*". It is shown that moving objects are correctly tracked despite of the false detection in "bee C",

Table 1. Tracking performance on tested videos

Bee		TE	IDS	GT	N	SD	MD	FD
C	**[12]**	484	20	3	652	1	9	1
	NB	21	11					
	Final	1	1					
E	**[12]**	1045	78	5	688	0	36	1
	NB	41	17					
	Final	3	5					
G	**[12]**	844	11	4	702	5	2	7
	NB	12	4					
	Final	0	0					
L	**[12]**	362	2	2	658	0	0	0
	NB	1	1					
	Final	0	0					
M	**[12]**	837	63	5	693	3	54	1
	NB	44	20					
	Final	8	8					

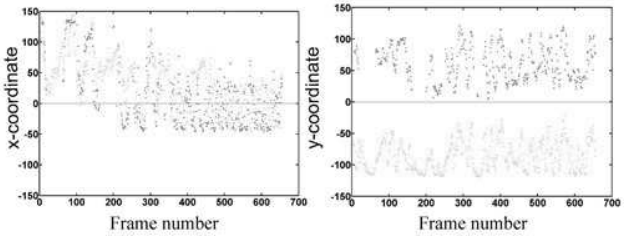

Fig. 2. Position of the tip of each object (green: right antenna, magenta: left antenna) in "bee L".

splitted detection in "bee G" and merged detection in "bee M" (highlighted by red rectangles in the first row).

5. CONCLUSION

It is a challenging task to track the movements of individual bee antennae, mandibles and proboscis in a video at low frame-rate. The main problems result from long tracking gaps and false, missing, splitted or merged detections. In this paper, a Bayesian framework is proposed to address these issues by incorporating the prior information about the appearance model (object level), the order of the objects (frame level) and temporal correlation (temporal level). Experimental results prove the efficacy and efficiency of the proposed framework.

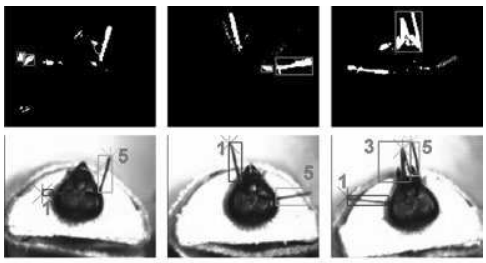

Fig. 3. Samples from bee C (left), G (middle) and M (right) with **FD**, **SD** or **MD**.

6. REFERENCES

[1] Y. Matsumoto, R. Menzel, J.C. Sandoze, and M. Giurfa, "Revisiting olfactory classical conditioning of the proboscis extension response in honey bees: a step toward standardized procedures," in *J. Neuroscience Methods*, 2012, pp. 159–167.

[2] S.A. Hussaini, L. Bogusch, T. Landgraf, and R. Menzel, "Sleep deprivation affects extinction but no acquisition memory in honeybees," in *Lean. Mem.*, 2009, vol. 16, pp. 698–705.

[3] S. Mujagic, S.M. Wuerth, S. Hellbach, and V. Duerr, "Tactile conditioning and movement analysis of antennal sampling strategies in honey bees (apis mellifera l.)," in *J. of Visualized Experiments*, 2013, in press.

[4] A. Veeraraghavan, R. Chellappa, and M. Srinivasan, "Shapeand-behavior encoded tracking of bee dances," in *IEEE PAMI*, 2008, vol. 3, pp. 463–476.

[5] A.G.A. Perera, C. Srinivas, A. Hoogs, G. Brooksby, and W. Hu, "Multi-object tracking through simultaneous long occlusions and split-merge conditions," in *IEEE CVPR*, 2006, vol. 1, pp. 666–673.

[6] J. Voigts, B. Sakmann, and T. Celikel, "Unsupervised whisker tracking in unrestrained behaving animals," in *J. Neurophysiol*, 2008, vol. 100, pp. 803–806.

[7] P. Kaewtrakulpong and R. Bowden, "An improved adaptive background mixture model for realtime tracking with shadow detection," in *Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems*, 2001, pp. 135–144.

[8] J. Yu, J. Amores, N. Sebe, and Q. Tian, "A new study on distance metrics as similarity measurement," in *IEEE ICME*, 2006, pp. 533 – 536.

[9] J. Munkres, "Algorithms for assignment and transportation problems," in *J. SIAM*, 1957, vol. 5, pp. 32–38.

[10] C.R. del Blanco, F. Jaureguizar, and N. Garcia, "Bayesian visual surveillance: A model for detecting and tracking a variable number of moving objects," in *IEEE ICIP*, 2011, pp. 1437–1440.

[11] Y. Li, C. Huang, and R. Nevatia, "Learning to associate: Hybridboosted multi-target tracker for crowded scene," in *IEEE CVPR*, 2009, pp. 2953–2960.

[12] T. Balch, Z. Khan, and M. Veloso, "Automatically tracking and analyzing the behavior of live insect colonies," in *Proceedings of the fifth international conference on Autonomous agents*. ACM, 2001, pp. 521–528.