Honeybee Detection and Pose Estimation using Convolutional Neural Networks

I. F. Rodríguez[1], K. Branson[2], E. Acuña[3], J. L. Agosto-Rivera[1], T. Giray[1], R. Mégret[1]

[1] University of Puerto Rico, Río Piedras campus
[2] Janelia Research Campus, Ashburn, Virginia
[3] University of Puerto Rico, Mayagüez campus

ivan.rodriguez5, jose.agosto1, tugrul.giray, edgar.acuna, remi.megret@upr.edu
kristin.branson@gmail.com

Keywords : Honeybee behavior analysis, pose estimation, Convolutional Neural Networks.

1 Introduction

The ability to automatize the analysis of video for monitoring animals and insects is of great interest for behavior science and ecology [1]. In particular, honeybees play a crucial role in agriculture as natural pollinators. However, recent studies has shown that phenomena such as colony collapse disorder are causing the loss of many colonies [2]. Due to the high number of interacting factors to explain these events, a multi-faceted analysis of the bees in their environment is required. We focus in our work in developing tools to help model and understand their behavior as individuals, in relation with the health and performance of the colony.

In this paper, we report the development of a new system for the detection, localization and tracking of honeybee body parts from video on the entrance ramp of the colony. The proposed system builds on the recent advances in Convolutional Neural Networks (CNN) for Human pose estimation and evaluates the suitability for the detection of honeybee pose as shown in Figure 1. This opens the door for novel animal behavior analysis systems that take advantage of the precise detection and tracking of the insect pose.

FIGURE 1 – Parts and parts connections used for pose estimation of honeybees

2 Related work

Traditional systems rely on background subtraction for bee detection. This was done in the Ctrax system [3] for fruit flies in lab conditions, and [4] for honeybees on the ramp at the entrance of the colony. The use of eigen models was proposed in [5] to track the position and orientation of bees inside the colony using a Rao-Blackwellized particle filter. In [6], a system was presented to tracking the 3D trajectory of bees in front of the colony using stereo vision. In [7] Convolutional Neural Network were used to track the antennas of constrained honeybees in a controlled setup. To the best of our knowledge, no system has yet shown the detection of multiple body parts of unconstrained honeybee insects with user trainable models.

In the context of human pose estimation, the recent Part Affinity Fields (PAF) approach [8] showed improved performance in multi-person pose detection by leveraging a deep Convolutional Neural Network for learning both parts detection and parts association. Taking as input an image the network simultaneously predicts a set of 2D confidence maps of body parts and 2D vectors fields of PAFs which encode the likely association between the points. The affinity fields provides information about the path and direction along which each pair of parts are more likely to be connected, and therefore is a key component to disambiguate the inference of associations once the parts have been detected. It simplifies the discrete matching step that has exponential complexity in the number of ambiguous matches when detecting multiple individuals.

3 Proposed approach

3.1 Method

The proposed approach aims at tracking the honeybees precisely when they move on the entrance ramp without intervening in their movements. The video capture system is designed to observe the ramp through which all foraging bees must pass to exit or enter the colony (see [9] for more details on the capture setup).

Following the Part Affinity Fields approach used in [8] for human pose, the neural network is composed on a first feature extraction part based on a pre trained VGG16 network, that serves as input for two convolutional network branches : the first branch predicts the part confidence map with one channel per part of interest, the second branch predict the part affinity fields with one vectorial channel per connection. These two branches are replicated to produce a sequence of 6 stages used to refine the estimation. The structural model used for the honeybees contains 5 parts which are user trainable : tip of the abdomen, thorax, tip of the head and tips of the left and right antennas. They are connected as illustrated in 1. The network is trained on synthetic smooth fields for the parts and PAFs fields built from the manual annotation. After detecting candidate parts as maxima of the parts fields, greedy inference is used to select the most likely predictions for the parts and use them as candidates for the PAFs to associate them. Under the constraint of 1-to-1 mapping, this process is very fast. For the inference of associations, a gaussian prior on the distance between parts was used, learned on the training dataset.

3.2 Evaluation

To evaluate the approach, a dataset was created using the videos acquired in June 2017 at the UPR Agricultural Experimental Station of Gurabo, Puerto Rico, producing videos of 2048x1600 pixels at 20 fps. We based our implementation of the PAF approach on the open source project [1] modified as discussed in 3.1 to allow a flexible structure definition adapted for honeybee morphology. Training was performed using 66 frames extracted randomly at different times. 5000 epochs were used for training using the Keras framework [10], taking 24h on a 12GB NVIDIA Titan X GPU. Testing was performed with 30 frames also extracted randomly. To avoid bias due to temporally close frames, it was ensured there was a difference of at least 2 seconds between the frames of the validation dataset and the training dataset, which is enough time for foraging bees to go though the ramp. The groundtruth for these frames was created by annotating all visible bees using an in-house Web application system.

Figure 2 shows a typical detection with the heat maps of the thorax detection and the head-abdomen affinity field. The final result after inference of the structure is shown in Figure 3 (left). It is interesting to note that despite very close and almost overlapping conditions in the lower left part of the image, all bees are correctly detected.

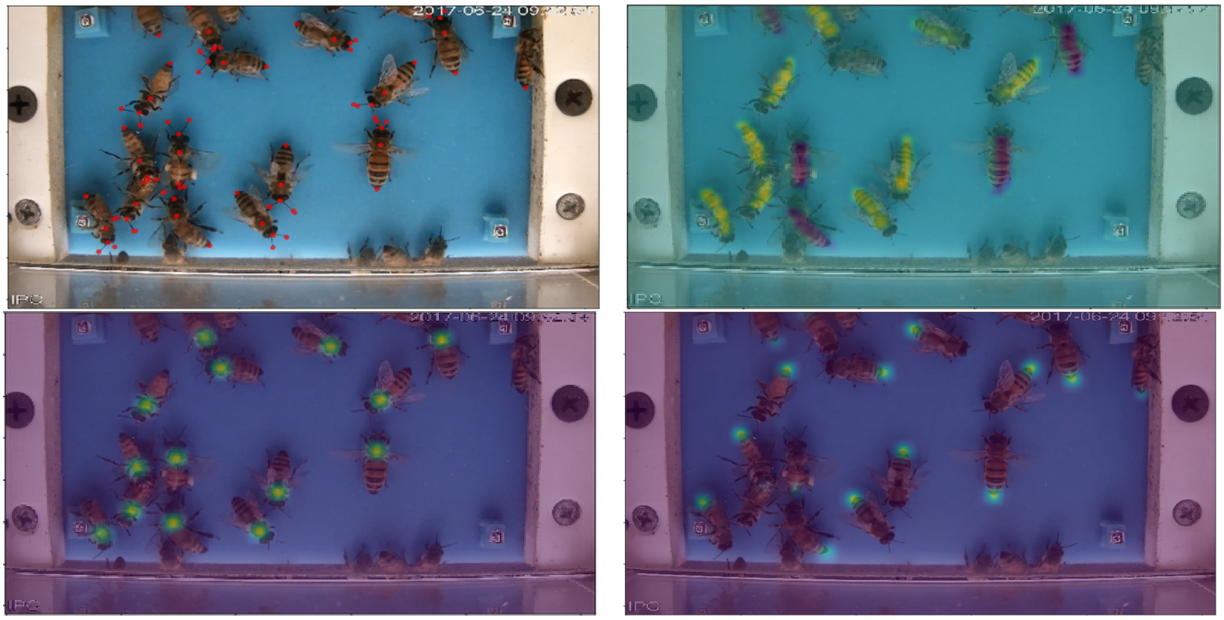

FIGURE 2 – Top Left : Input image with detected parts overlaid, Top Right : part affinity field for abdomen-thorax connection, Bottom : Heatmap for the thorax (left) and abdomen (right) part.

Quantitative evaluation was performed using a binary association of positive/false detection, as shown in Figure 3 (right). The Hungarian algorithm was applied to associate the detections to the groundtruth, taking a maximum distance of 30 pixels. The percentage of correct detection for each parts separately reached an average of 98.9%. The global matching score encapsulates the percentage of correct detections when all 5 parts are matched correctly, averaging 96.5%. For all measures, the median saturates at 100%, signaling that detection is perfect in more than half the frames.

1. https://github.com/michalfaber/keras_Realtime_Multi-Person_Pose_Estimation

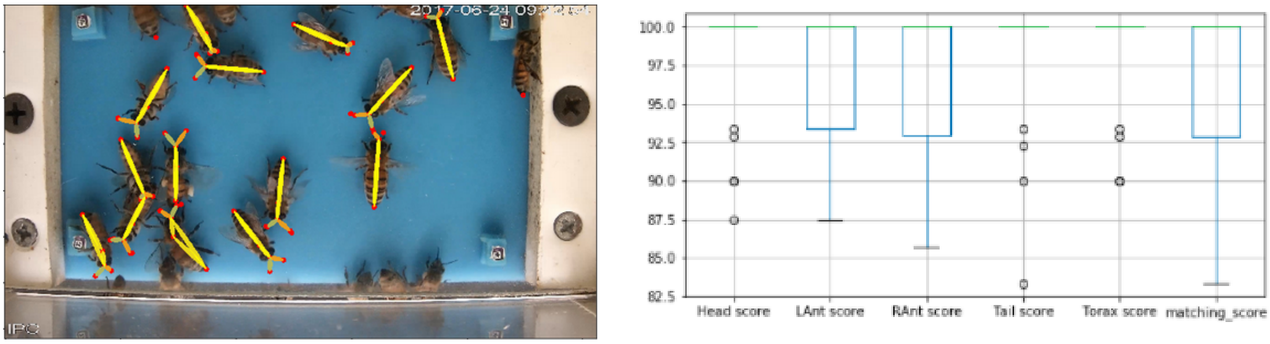

FIGURE 3 – Left : Example of inferred pose. Right : Percentage of correct detection of individual parts and global matching score over the test dataset (only 0, 25% and 50% quantile statistics are visible : median is 100% for all parts).

4 Conclusion

The proposed system shows very promising results with over 95% accuracy in pure detection of the complete bees on the entrance ramp without any tracking prior. Coupled with trait detection such as pollen bearing [9], this opens new possibilities for automated tracking of a rich set of honeybee behavior out of the laboratory in the field.
Future work will consider the evaluation of the method at large scale, which involves the reduction of the computational time and the evaluation in more challenging conditions such as low-light environment or the presence of markers used to identify a subset of individuals to be monitored more specifically.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grants No. 1633164 and 1633184. I. F. Rodriguez acknowledges support from the Janelia Visitor Program. T. Giray acknowledges support from Puerto Rico Science and Technology Trust (PRSTRT ; 2016-00161, 2017-00164). The authors would like to thank Stephanie Feliciano and Janpierre Aleman who helped with the acquisition at the UPR Agricultural Station of Gurabo of the videos used in the generation of the dataset and Jeffrey Chan who helped with the annotation of the videos.

References

[1] A. I. Dell, J. A. Bender, K. Branson, I. D. Couzin, G. G. de Polavieja, L. P. Noldus, A. Pérez-Escudero, P. Perona, A. D. Straw, M. Wikelski, and U. Brose, "Automated image-based tracking and its application in ecology," *Trends in Ecology & Evolution*, vol. 29, pp. 417–428, July 2014.

[2] T. Giray, M. Kence, D. Oskay, M. A. Döke, and A. Kence, "Scientific note : colony losses survey in Turkey and causes of bee deaths," *Apidologie*, vol. 41, pp. 451–453, July 2010.

[3] K. Branson, A. A. Robie, J. Bender, P. Perona, and M. H. Dickinson, "High-throughput ethomics in large groups of Drosophila," *Nature Methods*, vol. 6, pp. 451–457, June 2009.

[4] Z. Babic, R. Pilipovic, V. Risojevic, and G. Mirjanic, "Pollen Bearing Honey Bee Detection In Hive Entrance Video Recorded By Remote Embedded System For Pollination Monitoring," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III-7, pp. 51–57, June 2016.

[5] Z. Khan, T. Balch, and F. Dellaert, "A Rao-Blackwellized particle filter for eigentracking," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 980–986, 2004.

[6] G. Chiron, P. Gomez-Krämer, and M. Ménard, "Detecting and tracking honeybees in 3d at the beehive entrance using stereo vision," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, p. 59, 2013.

[7] L. Duan, M. Shen, W. Gao, S. Cui, and O. Deussen, "Bee Pose Estimation From From Single Images With Convolutional Neural Network," in *IEEE International Conference on Image Processing (ICIP)*, Sept. 2017.

[8] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[9] I. F. Rodriguez, R. Mégret, E. Acuña, J. L. Agosto-Rivera, and T. Giray, "Recognition of Pollen-bearing Bees from Video using Convolutional Neural Network," in *IEEE Winter Conf. on Applications of Computer Vision*, Mar. 2018.

[10] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015.