

Towards Automatic Honey Bee Flower-Patch Assays with Paint Marking Re-Identification

Luke Meyers* Rafael Meléndez-Ríos[†] Josué Rodríguez Cordero[†] Carlos Corrada Bravo[†]
Fanfan Noel[†] José Agosto-Rivera[†] Tugrul Giray[†] Rémi Mégret[†]

Abstract

In this paper, we show that paint markings are a feasible approach to automatize the analysis of behavioral assays involving honey bees in the field where marking has to be as lightweight as possible. We contribute a novel dataset for bees re-identification with paint-markings with 4392 images and 27 identities. Contrastive learning with a ResNet backbone and triplet loss led to identity representation features with almost perfect recognition in closed setting where identities are known in advance. Diverse experiments evaluate the capability to generalize to separate IDs, and show the impact of using different body parts for identification, such as using the unmarked abdomen only. In addition, we show the potential to fully automate the visit detection and provide preliminary results of compute time for future real-time deployment in the field on an edge device.

1. Introduction

Honey bees play an important role in our modern agriculture systems. As pollinators, the pollen they transfer between plants is critical to many crops’ ability to produce fruits [13]. Developing methods to monitor and further study bees and other pollinators is important for maintaining and further understanding their contributions to our agricultural systems. Field observations can maximize pollination’s agricultural benefit and lead to recommendations of pollinator and crop management choices. In particular, in individualized foraging decisions and strategies assays, visiting bees can choose between artificial flowers of different morphologies and rewards [8] [9]. Previous studies have shown variations of behavior at species level, which impact their choices when interacting with specific plants. However, to establish genetic linkage of behavioral traits demon-

strated in the assay would require a much larger number of behaviors at individual level.

Traditional experimental methods for such assays use visual inspection, which is limited to about 5 individual per day of experimentation due to humans limitations to monitor multiple individuals at once [9]. Computer Vision is a key element to improve the throughput of such studies, which in turn provide exciting challenges to push the state-of-the-art in insect monitoring: (i) Open setup in the field, with poorly controlled lighting conditions; (ii) Need for accurate re-identification with no or lightweight markings to be performed during the experimentation itself; (iii) Need for real-time analysis, as experimenters must physically catch individuals when specific conditions occur.

We propose several contributions towards these goals:

- Methods and results showing the suitability of paintcode markings as a re-ID method, validated on a new contributed dataset of 4392 high-quality images of honeybees comprising 27 manually labeled IDs.
- Method for detecting events during flower visits, validated on several videos taken in the field.
- Evaluation of the feasibility of integrating into an edge-computing system for the real-time acquisition and analysis in the field

The rest of the paper is organized as follows: after discussing related work (Sec. 2), we introduce the video collection system and visit detection approach (Sec. 3), present methods and results for re-ID (Sec. 4), and then provide preliminary results for a real-time implementation (Sec. 5).

2. Related Work

Computer vision has shown strong promise in automatic detection and monitoring of animals and more specifically pollinators in a variety of environments.

[20] have developed software that can effectively track insect movements in complex agricultural environments,

*Seattle University

[†]University of Puerto Rico, Río Piedras Campus

using a combination of background subtraction and deep learning models to address challenges such as occlusion. Their Polytrack algorithm includes a feature to identify flower position and extract relevant pollination events including number of visits and duration.

[16] developed a pipeline to characterize and identify bee behavior at the colony entrance based on pose estimation using deep convolutional networks [22], and [21] identified events such as entering and leaving. The SLEAP [19] and DeepLabCut [14] platforms leverage such deep learning models for multi-animal pose estimation and tracking in videos for a large variety of species.

In addition to short-term detection and tracking, recognizing the same individual across different timestamps or conditions, known as Re-Identification (reID), is an additional step to be able to monitor individual behavior over longer time periods.

In the case of bees, automatic Re-ID has traditionally used tags or bar-codes. [7] introduced the BEEtag system for individual bee tracking using fiducial markers, applied by [23] with SLEAP pose detection [19] to track and identify individual bumble bees to study antennal activity. Similarly, [3] proposed BeesBook, used by [24] to follow the lifetime behavior of thousands of bees.

Recent research has focused on markerless reID for various species, studied on datasets of seals [17], tigers [15], sea turtles [18], and cattle [2]. In the case of bees, [5] proposed a method for tracking bees inside a colony, using similarity between bee instances to resolve crossings. [4] trained a FaceNet model for bumblebee reID trained on the triplet loss using images from [25]. Finally, [6] presented a hybrid dataset of honey bees containing a larger set of unmarked bees used for unsupervised pre-training, and a smaller set of tagged bees for fine-tuning with 126 IDs for final evaluation. The effectiveness of the recognition from only the abdomen guided the design of our ablation study, comparing the effectiveness of full body images versus the abdomen and thorax crops (Sec. 4.2).

To our knowledge no study has evaluated the possibility of automatic reID using paint markings, which has the potential of offering a more reliable performance than unmarked reID, while being faster and more lightweight than tagging.

3. Video collection in the field

3.1. Experimental setup

The experimentation uses an artificial flower patch. Bees from a single hive are trained to forage at artificial flowers consisting of plexiglass squares with a small center well for nectar delivery. The assay considered in this work needs to evaluate the visits of each individual at each of the flowers for a specific number of visits, and then be captured to pre-

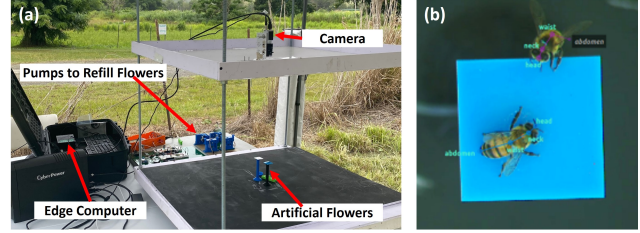


Figure 1. (a) Overview of prototype flower-patch system. (b) Example of bee pose detection on one artificial flower.

pare them for a new experimental phase. This requires both detection of the visits, recognized when the bees place their head on top of the center well, as well as individual identification. Each bee needs to be marked on their first visit, which is a delicate operation that justifies the use of paint, which is much easier to apply than other markers (numbered tags or bar-codes) and can be decoded by both human observers and the machine.

We used an experimental set up with two flowers, one blue and one white, with a Basler Gig-E camera mounted above the flower patch. White plexiglass was used to diffuse natural light and provided a surface to mount the camera using an L bracket next to a hole in the plexiglass. We tested the various frame rate and aperture settings of the camera to achieve optimum focus and lighting. The camera was connected to a NVIDIA Jetson Xavier edge computing system to record using a GStreamer pipeline utilizing onboard hardware converters to increase speed. The video analysis pipeline discussed next was validated offline using pre-recorded videos, and is intended to be run real time in the future (see section 5).

3.2. Visit detection

Images were generated from videos collected over 4 days of the experimental setup. During visits, researchers painted 1 to 2 small dots on each bee's thorax with various colors of enamel paint.

SLEAP [19] pose detection software was used to annotate a frames across several videos, train a neural network for bee pose estimation. The skeletons used contain a head, neck, waist, and abdomen keypoints. The final model was generated using a total of 170 annotated frames obtained incrementally by correcting the predictions from an initial model trained on 20 frames. The model and tracking was then run on all videos to obtain short-term tracks.

Flower coordinates were generated by detecting their square shape using OpenCV. This provided the the position of each flower and their center well for the visit detection. A visit to drink nectar was detected when the head keypoint was within a certain distance from the center well in each frame. Temporally consecutive detections were agglomerated to produce drinking events with a start and end frame,

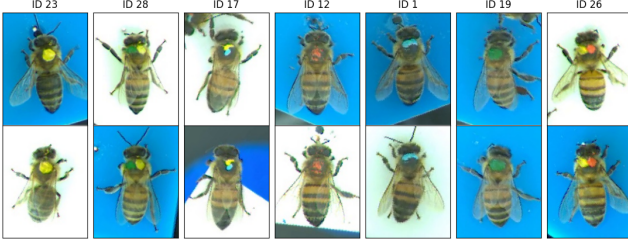


Figure 2. Sample images from the flowerpatch Re-ID dataset. Two images are shown for each individual bee.

associated to a flower ID and short-term track ID.

The evaluation of this module was performed on 3 videos, where a human annotated 98 drinking events. The automatic detection recalled 100% of the events, but duplicated 9% of them. (However, further review showed that in some cases bees had actually drank twice and moved slightly out of the well in between, suggesting the system can detect details that the human annotator had missed or ignored in the first visual inspection.)

4. Bee ReID

4.1. Re-ID dataset

From the tracks obtained, 150 x 200 pixel images of individual bees were extracted, centered on their waist key-point and aligned based on the neck and waist angle. Photos with multiple bees were removed, as well as wrong orientation estimates and occluded bees. Photos were then visually sorted based on individual ID and matching paint code. In creating the final dataset, tracks with less than 4 images of a single ID, and IDs with less than 2 tracks were removed. The resulting flowerpatch Re-ID dataset contains 4392 images with 27 identities. It was obtained from 787 short-term tracks, with an average of 5.6 images sampled per track. Sample images are shown in Figure 2. The dataset will be shared with the community.

4.2. General approach

When deployed in practice, a re-ID model will not have access to data for all possible identities that it will encounter. To this end, we framed the problem as a representation learning problem, trained using a supervised contrastive learning approach [12]. Each input image is projected into a 128 dimensional vector to be used as re-identification feature space using Euclidean distance.

We considered two data splits: *closed set setting* to evaluate the ideal supervised scenario; and *open set setting* to measure how well the model can generalize to unseen identities. (i) In the *closed set setting*, all identities are present in both the train and test sets. We do so by first sorting all images by time and grouping them into tracks. The first 70% of tracks for each individual bee is assigned to the train set

and the remaining to the test set. (ii) In the *open set setting*, the train and test sets have distinct identities. We do so by assigning 60% of the IDs (16) to the train set and the remaining IDs (11) to the test set. For evaluation, we further split the test set into a *reference* set and *query* set by first sorting the test images by time and grouping them into tracks, and then assigning the first 25% to the reference set and the rest to the query set.

We furthermore distinguish four different versions of the input images: 1) *aligned full body*; 2) *aligned abdomen* (bottom crop); 3) *aligned thorax* (top crop, which may include part of the head); and 4) *unaligned full body*. The first three will help evaluate how information from different parts of the image affect model performance. The last one allows us to evaluate the usefulness of vertically aligning the bee images as an additional pre-processing step.

4.3. Model Architecture and training

We use a convolutional neural network (CNN) that embeds the input images into a 128 dimensional feature space. This network consists of a ResNet50 backbone [10] with bottleneck residual blocks, truncated after the `conv3` stage (after 22 layers). The head is a fully connected layer with L2-normalization.

The model was trained using the semi-hard triplet loss with a margin of 0.2, optimized using Adam with a learning rate of 0.001 for a maximum of 1000 epochs and a patience of 100. Dropout was applied during training both before and after the fully connected layer with rates of 0.5 and 0.2 respectively.

Unaligned full body uses 200x200 pixel images trained with augmentation by a random rotation around the center.

4.4. Evaluation

Evaluation is performed using galleries and reporting the top-1 and top-3 Cumulative Matching Characteristic (CMC) metric. For each of the data splits, 10,000 galleries are constructed as follows:

(i) For the *closed set setting*, an anchor image is randomly sampled from the test set, while a positive image along with 9 negative images are randomly sampled from the train set.

(ii) For the *open set setting*, an anchor image is randomly sampled from the query set, while a positive image along with 9 negative images are randomly sampled from the reference set.

We additionally evaluate using k NN classifier trained on the complete reference data (train set for closed set setting, reference set for open set setting).

A simple *Pixel PCA* feature space is used as baseline: the raw image pixel RGB values flattened into a one-dimensional vector and projected into the principal components accounting for 95% of the variance from the train set.

4.5. Results

Closed Set Setting Tab. 1 shows that in the closed set setting, the baseline which uses pixel data without any model training can reach an accuracy of almost 0.83 1-NN accuracy when using the thorax, but only 0.38 Top-1 accuracy, which indicates a large intra-class variability that could be compensated with enough training data. The contrastively trained ResNet model performs very well regardless of input type, with the thorax crop or full image reaching almost perfect performance. The abdomen performance is lower, but still demonstrates the potential of using the unmarked abdomen for Re-ID, which aligns with previous work of [6].

Results for ResNet trained on un-aligned images were only slightly lower than for aligned images. Thus body alignment did not lead to a large improvement in the closed set setting: augmentation during training seemed to account for the variability in the input.

Open Set Setting Tab. 1 shows that in the open set setting, we see a larger gap between abdomen crop and thorax crop results, showing the importance of paint marking to obtain reliable re-identification in that context. It would be interesting for future work to evaluate the impact of using larger training sets on this gap. We note that all individual paint colors in the test set IDs are also present in the train set IDs, although in slightly different shapes and position, which may account for why thorax performed well.

Performance for the model trained on un-aligned images was lower than for aligned images. Thus, unlike in the closed set setting, vertically aligning the images was needed in the open set to help generalize to new IDs.

5. Towards real-time

Real-time analysis is critical for taking actionable decisions in the field during behavioral assays. In order to move from offline processing to real-time processing in the field, where the system may need to be battery powered, the hardware constraints of the system such as power consumption and computing capabilities impose additional limitations. Using the same NVIDIA Jetson Xavier AGX platform as was used for data collection, we trained a Yolov5 model [11] to detect the bees in our videos and benchmarked its performance on the videos cropped to 1184 x 1184 pixels.

As a baseline, frame-by-frame PyTorch inference took 60ms per frame (17 fps). Converting the model to a TensorRT engine and running inference within an NVIDIA DeepStream pipeline resulted in 13ms per frame (59 fps). Adding a tracker (NvDCF) and an Apache Kafka [1] message broker to export the trajectories added 1ms per frame. These preliminary results demonstrate that although a naive implementation of the models may not reach the 20 fps target, it is feasible to significantly exceed this framerate for

bee detection and tracking with the proper computational optimizations, leaving capacity to add the head pose estimation, re-id modules and visit detection.

6. Conclusion

In this paper, we presented a comprehensive set of methods to tackle the problem of individual behavior monitoring in flower patch assays in the field. We presented a novel dataset of honey bees with paint marks and trained a deep learning model for reID. Our results show that paint marks is a viable light-weight alternative to more traditional tagging, thus facilitating the automatic re-identification of individuals during field experiments. These results showed a strong benefit of vertically aligning images in the open set setting, but was not as strong in the closed set. They also showed promising performance of using only the unmarked abdomen, suggesting the prospect of usable unmarked Re-ID in the future with improved models. The CNN based detection was of good quality, leading to accurate drinking event detections. Integration of the various modules into a real-time system is in progress, with the potential to enable high-throughput individual honey bee behavior characterization in the field.

References

- [1] Apache Software Foundation. Apache Kafka. 4
- [2] Luca Bergamini, Angelo Porrello, Andrea Capobianco Don-dona, Ercole Del Negro, Mauro Mattioli, Nicola D’alterio, and Simone Calderara. Multi-views embedding for cattle re-identification. In *2018 14th international conference on signal-image technology & internet-based systems (SITIS)*, pages 184–191. IEEE, 2018. 2
- [3] Franziska Boenisch, Benjamin Rosemann, Benjamin Wild, David Dormagen, Fernando Wario, and Tim Landgraf. Tracking all members of a honey bee colony over their lifetime using learned models of correspondence. *Frontiers in Robotics and AI*, 5:35, 2018. 2
- [4] Parzival Borlinghaus, Frederic Tausch, and Luca Retten-berger. A purely visual re-id approach for bumblebees (*bombus terrestris*). *Smart Agricultural Technology*, 3:100135, 2023. 2
- [5] Katarzyna Bozek, Laetitia Hebert, Yoann Portugal, Alexander S Mikheyev, and Greg J Stephens. Markerless tracking of an entire honey bee colony. *Nature communications*, 12(1):1733, 2021. 2
- [6] Jeffrey Chan, Hector Carrión, Rémi Mégret, Jose L Agosto-Rivera, and Tugrul Giray. Honeybee re-identification in video: New datasets and impact of self-supervision. In *VISI-GRAPP (5: VISAPP)*, pages 517–525, 2022. 2, 4
- [7] James D Crall, Nick Gravish, Andrew M Mountcastle, and Stacey A Combes. Beetag: a low-cost, image-based tracking system for the study of animal behavior and locomotion. *PloS one*, 10(9):e0136487, 2015. 2

Features	Image Input	Closed Set Setting				Open Set Setting			
		Top-1	Top-3	1NN	3NN	Top-1	Top-3	1NN	3NN
PixPCA	Abdomen	0.18	0.44	0.42	0.40	0.25	0.51	0.56	0.53
	Thorax	0.38	0.57	0.83	0.79	0.45	0.64	0.79	0.77
	Full Img	0.23	0.47	0.59	0.58	0.26	0.55	0.70	0.68
ResNet	Abdomen	0.94	0.990	0.90	0.91	0.74	0.93	0.94	0.94
	Thorax	0.992	0.999	0.98	0.98	0.90	0.98	0.96	0.95
	Full Img	0.993	0.999	0.98	0.98	0.87	0.92	0.98	0.92
	Full Img Unaligned	0.98	0.997	0.95	0.95	0.78	0.95	0.93	0.91

Table 1. CMC Scores and kNN Accuracies for re-identification in closed and open set settings.

- [8] Jennifer H Fewell and Robert E Page Jr. Colony-level selection effects on individual and colony foraging task performance in honeybees, *apis mellifera* l. *Behavioral Ecology and Sociobiology*, 48:173–181, 2000. **1**
- [9] Tugrul Giray, Charles I Abramson, Ana Chicas-Mosier, Tiya Brewster, Christine Hayes, Karianne Rivera-Vega, Maya Williams, and Harrington Wells. Effect of octopamine manipulation on honeybee decision making: reward and cost differences associated with foraging. *Animal Behaviour*, 100:144–150, 2015. **1**
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. **3**
- [11] Glenn Jocher. YOLOv5 by Ultralytics, May 2020. **4**
- [12] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020. **3**
- [13] Alexandra-Maria Klein, Bernard E Vaissière, James H Cane, Ingolf Steffan-Dewenter, Saul A Cunningham, Claire Kremen, and Teja Tscharntke. Importance of pollinators in changing landscapes for world crops. *Proceedings of the royal society B: biological sciences*, 274(1608):303–313, 2007. **1**
- [14] Jessy Lauer, Mu Zhou, Shaokai Ye, William Menegas, Steffen Schneider, Tanmay Nath, Mohammed Mostafizur Rahman, Valentina Di Santo, Daniel Soberanes, Guoping Feng, Venkatesh N. Murthy, George Lauder, Catherine Dulac, Mackenzie Weygandt Mathis, and Alexander Mathis. Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nature Methods*, 19(4):496–504, Apr. 2022. Number: 4 Publisher: Nature Publishing Group. **2**
- [15] Shuyuan Li, Jianguo Li, Hanlin Tang, Rui Qian, and Weiyao Lin. Atrw: a benchmark for amur tiger re-identification in the wild. *arXiv preprint arXiv:1906.05586*, 2019. **2**
- [16] Rémi Mégret, Ivan F Rodriguez, Isada Claudio Ford, Edgar Acuña, Jose L Agosto-Rivera, and Tugrul Giray. Labelbee: a web platform for large-scale semi-automated analysis of honeybee behavior from video. In *Proceedings of the Conference on Artificial Intelligence for Data Discovery and Reuse*, pages 1–4, 2019. **2**
- [17] Ekaterina Nepovinnikh, Tuomas Eerola, Vincent Biard, Piia Mutka, Marja Niemi, Mervi Kunasranta, and Heikki Kälviäinen. Sealid: Saimaa ringed seal re-identification dataset. *Sensors*, 22(19):7602, 2022. **2**
- [18] Kostas Papafitsoros, Lukáš Adam, Vojtěch Čermák, and Lukáš Pícek. Seaturtleid: A novel long-span dataset highlighting the importance of timestamps in wildlife re-identification. *arXiv preprint arXiv:2211.10307*, 2022. **2**
- [19] Talmo D Pereira, Nathaniel Tabris, Arie Matsliah, David M Turner, Junyu Li, Shruthi Ravindranath, Eleni S Papadoyannis, Edna Normand, David S Deutsch, Z Yan Wang, et al. Sleep: A deep learning system for multi-animal pose tracking. *Nature methods*, 19(4):486–495, 2022. **2**
- [20] Malika Nisal Ratnayake, Adrian G Dyer, and Alan Dorin. Tracking individual honeybees among wildflower clusters with computer vision-facilitated pollinator monitoring. *Plos one*, 16(2):e0239504, 2021. **1**
- [21] Iván F Rodriguez, Jeffrey Chan, Manuel Alvarez Rios, Kristin Branson, José L Agosto-Rivera, Tugrul Giray, and Rémi Mégret. Automated video monitoring of unmarked and marked honey bees at the hive entrance. *Frontiers in Computer Science*, 3:142, 2022. **2**
- [22] Ivan F Rodriguez, Remi Megret, Edgar Acuna, Jose L Agosto-Rivera, and Tugrul Giray. Recognition of pollen-bearing bees from video using convolutional neural network. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 314–322. IEEE, 2018. **2**
- [23] Matthew A-Y Smith, August Easton-Calabria, Tony Zhang, Szymon Zmyslony, Jessie Thuma, Kayleigh Cronin, Cassandra L Pasadyn, Benjamin L de Bivort, and James D Crall. Long-term tracking and quantification of individual behavior in bumble bee colonies. *Artificial Life and Robotics*, 27(2):401–406, 2022. **2**
- [24] Michael L Smith, Jacob D Davidson, Benjamin Wild, David M Dormagen, Tim Landgraf, and Iain D Couzin. Behavioral variation across the days and lives of honey bees. *Isience*, 25(9):104842, 2022. **2**
- [25] Frederic Tausch, Simon Stock, Julian Fricke, and Olaf Klein. Bumblebee re-identification dataset. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, pages 35–37, 2020. **2**