A Vision-based System for Social Insect Tracking

Kristi Žampachů
Faculty of Electrical Engineering
Czech Technical University in Prague
Prague, Czech Republic
zampakry@fel.cvut.cz

Jiří Ulrich
Faculty of Electrical Engineering
Czech Technical University in Prague
Prague, Czech Republic
ulricji1@fel.cvut.cz

Tomáš Rouček
Faculty of Electrical Engineering
Czech Technical University in Prague
Prague, Czech Republic
rouceto1@fel.cvut.cz

Martin Stefanec
Artificial Life lab
Univerzity of Graz
Graz, Austria
martin.stefanec@uni-graz.at

Dominik Dvořáček
Faculty of Electrical Engineering
Czech Technical University in Prague
Prague, Czech Republic
dvorado6@fel.cvut.cz

Laurenz Fedotoff
Artificial Life lab
Univerzity of Graz
Graz, Austria
laurenz.fedotoff@uni-graz.at

Daniel Nicolas Hofstadler
Artificial Life lab
Univerzity of Graz
Graz, Austria
daniel.hofstadler@uni-graz.at

Fatemeh Rekabi-Bana
Swarm & Comp. Int. Lab., Dept. of CS
Univerzity of Durham
Durham, United Kingdom
fatemeh.rekabi-bana@durham.ac.uk

George Broughton
Faculty of Electrical Engineering
Czech Technical University in Prague
Prague, Czech Republic
brouggeo@fel.cvut.cz

Farshad Arvin
Swarm & Comp. Int. Lab., Dept. of CS
Univerzity of Durham
Durham, United Kingdom
farshad.arvin@durham.ac.uk

Thomas Schmickl
Artificial Life lab
Univerzity of Graz
Graz, Austria
thomas.schmickl@uni-graz.at

Tomáš Krajník
Faculty of Electrical Engineering
Czech Technical University in Prague
Prague, Czech Republic
krajnt1@fel.cvut.cz

***Abstract*—Social insects, especially honeybees, play an essential role in nature, and their recent decline threatens the stability of many ecosystems. The behaviour of social insect colonies is typically governed by a central individual, e.g., by the honeybee queen. The RoboRoyale project aims to use robots to interact with the queen to affect her behaviour and the entire colony's activity. This paper presents a necessary component of such a robotic system, a method capable of real-time detection, localisation, and tracking of the honeybee queen inside a large colony. To overcome problems with occlusions and computational complexity, we propose to combine two vision-based methods for fiducial marker localisation and tracking. The experiments performed on the data captured from inside the beehives demonstrate that the resulting algorithm outperforms its predecessors in terms of detection precision, recall, and localisation accuracy. The achieved performance allowed us to integrate the method into a larger system capable of physically tracking a honeybee queen inside its colony. The ability to observe the queen in fine detail for prolonged periods of time already resulted in unique observations of queen-worker interactions. The knowledge will be crucial in designing a system capable of interacting with the honeybee queen and affecting her activity.**

***Keywords*—fiducial markers, experimental biology, robot vision**

This work was supported by EU H2020 project RoboRoyale no. 964492. Zampachu and Ulrich share first authorship.

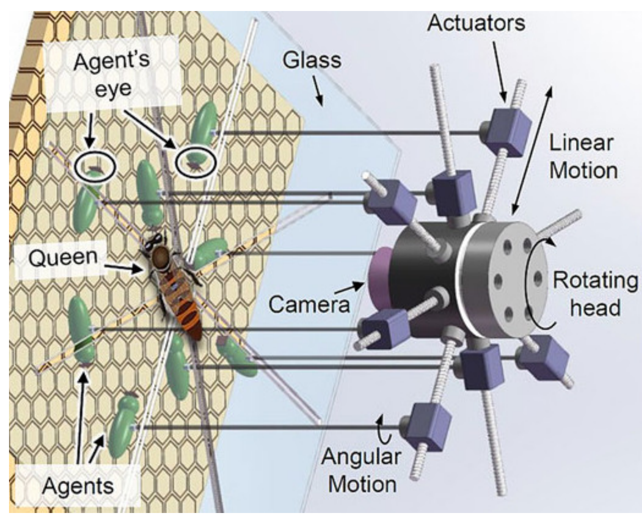

Fig. 1. Concept of the RoboRoyale system: Artificial agents are used to infiltrate the honeybee queen's court to affect her behaviour and, thus, the activity of the honeybee colony. Courtesy of [1]

I. Introduction

Honeybees are the primary pollinators of flowering plants and, thus, are a crucial species to the earth's ecosystem. Around 400 billion US dollars worth of annual global food production relies on their contribution [2]. Individual bees are capable of pollinating plants in an area of 300 km^2 around their hive, and it is estimated that a colony visits more than 500,000 flowers per year [3]. In natural ecosystems, bees support the reproduction and spread of plants that provide food, building and nesting materials, and shelter for animals [1].

All this is possible because honeybees are eusocial insects, and the bee colony acts like a singular superorganism. Much of the coordination is decentralised [4], for example, through the waggle dance, where a successful worker bee forager transmits information about a food source to other worker bees [5], [6]. However, an important central component of the superorganism is the honeybee queen, which is responsible for the growth of the bee colony and the reproduction of the superorganism. Since these essential functions depend on a single individual, it is crucial that the bee colony and the queen are in a constant exchange of information. This information exchange, as well as behavioural control from the queen to the workers, is regulated to large parts through pheromones [7]. The essential pheromone produced by the queen is the Queen Mandibular Pheromone (QMP). This pheromone transmits information about the presence and status of the queen to other bees in the colony [8]. It attracts drones during their mating flight, suppresses the development of ovaries in workers [9], [10], and influences the behaviour of the other bees in the colony in many additional ways. [7], [11]. The worker bees that interact with the queen bee directly are called retinue bees or court bees. They feed the queen, groom her, and ensure that the queen's pheromones are distributed sufficiently throughout the hive. When the queen shows little locomotive activity, they converge around her to form a characteristic pattern called the queen's court.

These worker bees are the main recipients of communication from the queen; they act as the interface between the queen and the rest of the colony.

The EU Future and Emerging Technologies project 'RoboRoyale' aims to integrate a set of biomimetic robots into the queen's court to affect the queen's behaviour and the transmission rate of her pheromones [1]. Affecting the queen allows influencing the brood production, as well as the activity of the entire colony, which, in turn, impacts the surrounding ecosystem through foraging and pollination. In short, the integration of robotic agents into the court would allow for optimising the hive's macroscopic variables by regulating the queen's pheromone transfer to the colony.

To perform the integration, we will deploy a robotic manipulator that will continuously track the queen. The manipulator's head will consist of a mechatronic system capable of moving the artificial agents in the queen's vicinity, see Figure 1. The concept of the robotic system is described in [1], and a more detailed description of its mechanics is described in [12]. To prevent any harm to the queen, the manipulator must be able to track her precisely and reliably. Moreover, the artificial agents must be accepted by the queen as well as the other worker bees. This requires not only that their shape and material are compliant but also that their motion patterns and behaviour match those of the court bees.

Therefore, prior to the deployment of the manipulator, we constructed a system capable of continuous observation of the queen's vicinity. These observations would provide the data necessary to analyse the court bees' interactions with the queen so that the robotic system could imitate these. This, again, requires precise and reliable tracking of the honeybee queen inside an observation hive, which is the focus of this paper.

II. Systems used for bee tracking

Much research on honeybee behaviour is done using observation hives, a scaled-down version of the hives utilised for commercial purposes in apiaries [13]. They usually consist of one to three vertically arranged honeycombs, which are enclosed by two glass panes so that the bees can only access the honeycomb area via a designated entrance which leads to the outside. Computerised observation hives have additional cameras pointed at the combs and connected to equipment for data storage and analysis, see Figure 2. To provide bees with naturally dark conditions, the hive is illuminated with infrared or near-infrared light that is invisible to bees [14].

Fig. 2. Observation hive setup - two infrared cameras on the left are recording two combs on the right.

Honeybee colonies present a particular challenge for detection and tracking systems. The core problem is the high density of nearly identical specimens that are constantly moving and occluding each other. Another issue is that the observations are performed through the glass that covers the combs. As the temperature and humidity inside the hive change, fog forms on the glass, significantly reducing the image quality. This is further amplified by the wax or other residuals that bees sometimes deposit on the glass. Bees also tend to slightly alter the shape of the combs, causing a loss of focus in the affected areas. Fast movement of the bees, e.g. during their waggle dance, combined with low-intensity (near-)infrared lights, results in motion blur. Finally, the level of detail needed to analyse the behaviour of small specimens combined with the requirement to monitor the entire comb results in the need to use high-resolution cameras. This, in turn, requires large storage and significant computational power if the analysis should be performed online. Conventional, general-purpose detection methods, based on YOLO [15] or SURF [16] , struggle to achieve the required performance in these situations.

However, there are a few systems capable of tracking individual bees in an observation hive setting. The system proposed in [17] exploits the subtle differences (or 'pixel-personalities') in the individual bees. The resulting system [18] uses a segmentation convolutional neural network (CNN) built upon the U-Net architecture. Based on the segmentation, individual detections are generated. It adds a temporal component, therefore reducing the size of the network to 6% of the original [17]. The previous frame is used to leverage the spatio-temporal information encoded in a video sequence. This allows one to reduce the size of a CNN while preserving its accuracy. From small sequences of detections that are spatially linked through subsequent frames, pixel personalities are learnt about individual bees. The personalities are used to link the bee detections with the temporal detections of individual bees throughout the recording [17]. This markerless system is capable of detecting the positions of individual bees and their orientations, but it cannot differentiate between a worker bee and a queen.

One of the ways to alleviate the aforementioned problems and increase tracking accuracy is the usage of easy-to-detect markers. Many of these methods are inspired by the fiducial markers commonly used in robotics research, such as the square fiducials AprilTag [19] or ArUco [20] or the circular ones like TRIP [21] or WhyCode [22], [23].

The BEEtag marker [24], which was designed to track small animals, is a derivative of the AprilTag [19]. The tag features a black square with a white border with an inscribed identification matrix on it. The 5×5 matrix offers a 15-bit encoding with an additional 10-bit correction. Theoretically, the 15-bit encoding allows one to distinguish over 30 thousand individuals. However, the BEEtag authors limit the possible code combinations to resolve the orientation estimation ambiguity of the square marker. Moreover, they introduce additional constraints on the code's Hamming distance, producing two BEEtag versions with 7000 and 110 unique tags, respectively. Reliable recognition of the BEEtag marker requires that the width of its diagonal is at least 50 pixels. Given that the size of the marker on the queen's thorax is limited to about 2 mm and the size of the monitored combs is around 400 mm, the minimal required image resolution would exceed 10000×10000 pixels.

Another fiducial marker Circulatrix was specifically designed for honeybees [25]. The marker, which is printed on polyester film, is slightly curved to adjust for the shape of the worker bee's thorax, as shown in Figure 4. The marker's central element is designed to provide not only the position but also the orientation of the bee to which it is attached. The elements used for identification are arranged around the central pattern, achieving 12-bit encoding that allows distinguishing over 4000 individuals. An impressive BeesBook system was based on the Circulatrix marker [26]. The BeesBook system first constructs tracklets, which are trajectories of individual bees in consecutive frames. The tracklets are subsequently merged into longer tracks, capturing the movement of each marked bee in the honeybee colony. Although the circularity of the marker means that it can be slightly smaller than BEEtag, achieving its reliable detection across a standard-sized comb still requires significant camera resolution.

Fig. 3. BEEtag markers attached to a group of cockroaches. Courtesy of [24]

Fig. 4. Circulatrix markers glued onto the bee's thoraxes. Courtesy of [26]

None of the aforementioned systems achieves real-time performance when processing high-resolution images. Unfortunately, high resolution is required because the monitored area of a honeycomb is large, and the level of detail needed is high.

To deal with the problem, we opted to build our detection on the WhyCon marker [22] and its derivative WhyCode [23].

To reliably detect the honey bee queen, we need a system that is capable of detection even at lower resolution in the observation hives, and that is accurate in differentiating the queen and worker bees. ArUco markers must be quite large compared to the size of a bee and are highly susceptible to occlusions. The resolution of BeesBook (Circulatrix) and BEEtag markers used in [27] would, in our case, require the processing of images with resolutions exceeding 7000×7000 pixels. The markerless system cannot track the queen in the long term because the pixel personality is not reliable enough to re-identify the queen after the loss of tracking, for example, due to occlusions or the queen's absence. Moreover, the U-net network, on which the markerless tracking is based, is not meant for real-time applications. In particular, real-time tracking of the queen in four 12Mpx images (our configuration) would require 24 NVIDIA GeForce RTX 3070Ti graphics cards per hive. Therefore, we chose WhyCode because of its ability to detect markers even at a lower resolution [22] and in real time, and the markers can be easily printed on paper.

III. Method description

The proposed system, WhyComb, is based on the WhyCode [23], [28] system, and it is enhanced to improve the detection and localisation of a marker attached to a bee in an observation hive. While the WhyCode offers high pose estimation precision even in challenging illumination conditions, it can track a signification number of markers in real time without specialised hardware. One of the dominant features of the marker is the high detection distance, which directly translates into good segment detection even when the marker has low resolution. This feature is advantageous, especially when only a single camera is used to observe the entire comb, resulting in a marker diameter smaller than 25 pixels. Although the marker supports a versatile encoding system for unique identification, it is not necessary to use any marker with a higher number of bits than two because there is only one queen in a colony. Also, such a low-bit encoding does not restrict the detection conditions compared to a higher-bit encoding while still providing orientation estimation.

A. Dealing with occlusions

The prominent characteristic of any eusocial insect is their mutual interaction which causes partial marker occlusions. In our case, marker occlusions are caused by the worker bees' interactions with the queen. As the original WhyCode detection method was not designed to be resistant to occlusion, even minor obstructions can hamper successful detection depending on the immediate segmentation threshold. Examples of such occlusions, where the marker is partially visible, but the WhyCode method cannot detect it, are shown in Figure 5.

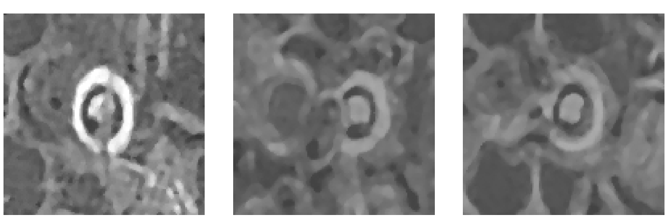

Fig. 5. Examples of partial marker occlusions

To ensure the best continuous tracking possible, the WhyCode system is extended as indicated in Figure 7. The image captured by the camera is processed using the WhyCode method, and the detected marker parameters are passed for further analysis to filter out false positive detections. If the detection is evaluated as successful, it is passed to an 'estimated position' buffer. The position from the buffer is then used as an initial position for a subsequent convolution step, which further refines the marker position and places the result back in the position buffer. In other words, the convolution continuously updates the position buffer with its last estimates, tracking the marker during its occlusions. However, when the WhyCode method detects the marker, the position buffer is updated with the WhyCode detection result, reinitialising the convolution-based tracker.

The convolution kernel has to capture the shape of the marker as precisely as possible. However, for computational efficiency, the unique marker encoding used for orientation estimation is suppressed in the convolution kernel, as shown in Figure 6. This kernel allows rotationally invariant detection, meaning that the convolution can only be performed in the 2D domain. This avoids the need to calculate the convolution for all possible marker orientations, which would be computationally inefficient.

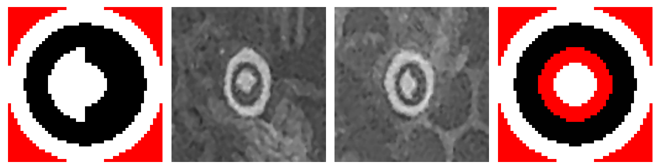

Fig. 6. Ideal shape of the WhyCode marker, its real examples in the captured images and the convolution kernel used to refine its position. The red colour indicates areas which are not part of the marker or the convolution kernel.

Suppose the queen's marker is completely obscured. Then, the maximum of the convolution could drift away from the queen due to the marker's structural resemblance to the surrounding comb and because the convolution will always produce some maximum when applied to an image. This is why we have to incorporate the continuous reinitialisation of the convolution using WhyCode estimates, as shown in Figure 7. Moreover, we need to suppress situations where the WhyCode method would provide a false positive detection, as the convolution would start to produce false detections until it is reinitialised correctly.

B. Suppressing false positive detections

The combs of the hive consist of slightly elliptical cells, and the larvae in their centre also form elliptical shapes. This makes the cells susceptible to being mistaken for the queen's marker. Moreover, comb cells are not the only elements that can be falsely detected, as circular bits of debris often appear scattered across the comb.

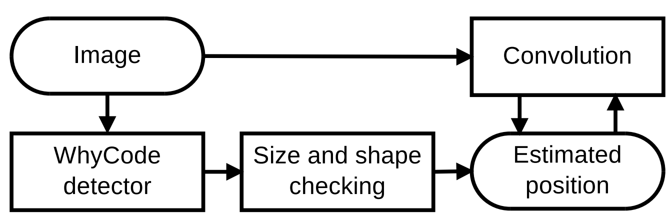

Fig. 7. Method overview: The positions of the WhyCode detector are filtered and used to reinitialise the convolution-based tracker.

To deal with the problem, we take advantage of WhyCode's eigenvalues and eigenvectors, which describe the detected segment. This allowed us to construct an additional filtering step based on the marker's expected shape of the inner ellipse. In particular, we verify if the outermost pixels of the inner white segment of the marker are located within an expected range from the detected marker centre. First, we transform the coordinates of each pixel of the inner segment into a canonical coordinate system as follows:

$$x_p = \frac{v_2\,(x_o - x_c) - v_1\,(y_o - y_c)}{2\sqrt{\lambda_2}}, y_p = \frac{v_2\,(y_o - y_c) + v_1\,(x_o - x_c)}{2\sqrt{\lambda_1}}, \tag{1}$$

where λ_0, λ_1 are the eigenvalues of the segment, v_1 and v_2 are elements of the dominant eigenvector, x_c, y_c are the coordinates of the centre of the segment, and x_o, y_o are the original coordinates of the transformed pixel. Equation 1 corresponds to centring the pixel coordinates, rotating them to align with the coordinate system and scaling them by the segment size. Thus, if applied to a correctly detected segment, Equation 1 would transform all its pixel coordinates into a unit circle. This means that the transformed outermost pixels of the inner segment circle, as shown in the left part of Figure 6, would appear only at a specific distance. By setting a range of allowed distances, we can filter out detections with inner segments that do not reflect the expected marker shape. Thus, our filter calculates the squared distance d of the outermost pixel as

$$d = \max_{p \in \mathscr{P}} (x_p^2 + y_p^2), \tag{2}$$

where $\mathscr{P}$ are all pixels in the inner segment. After calculation of the parameter d, we simply discard all segments with $d <$ 0.4 (inner segment too small) or $d > 0.6$ (inner segment too large). Furthermore, we also filter out all detections with a radius smaller than 10 pixels.

For the sake of interoperability, the aforementioned improvements were integrated into the localisation system as a Robot Operating System (ROS) node.

IV. Experiments

To evaluate the impact of the modifications, we compare the performance of the WhyComb method with the original WhyCode algorithm, as well as with standard convolution. First, we measured the localisation accuracies of both methods and compared them using a Wilcoxon pairwise test. We also calculated the cumulative distribution function of the localisation errors to provide further insight. Finally, we compared the precision and recall of the methods.

A. Evaluation datasets

Our evaluation datasets were recorded at the University of Graz, Austria. The setup consisted of a single observation hive 2 containing two combs. Four 4000×3000 px infrared cameras were installed to observe both sides of the two combs. Each camera was connected to an NVIDIA Jetson Nano computer that was running the WhyComb method. These four computers were streaming their results to a central PC that stores the images. We took approximately 10 hours of recording from the system, containing around 27000 frames and used them to create two evaluation datasets.

The 'Occluded' dataset contains 875 images located at the end of the 27000 frame sequence. By that time, the worker bees had accepted the marker on the queen, so they treated it as a part of her body and frequently crawled over it, causing frequent occlusions. This dataset was used primarily to evaluate the impact of the improvements proposed in Section III-A onto the robustness of the system to occlusions.

The second 'FalsePos' dataset contained 635 frames with false positive detections spread across the original 27000 frame sequence. This dataset was used to verify the efficiency of the filtering techniques proposed in section III-B.

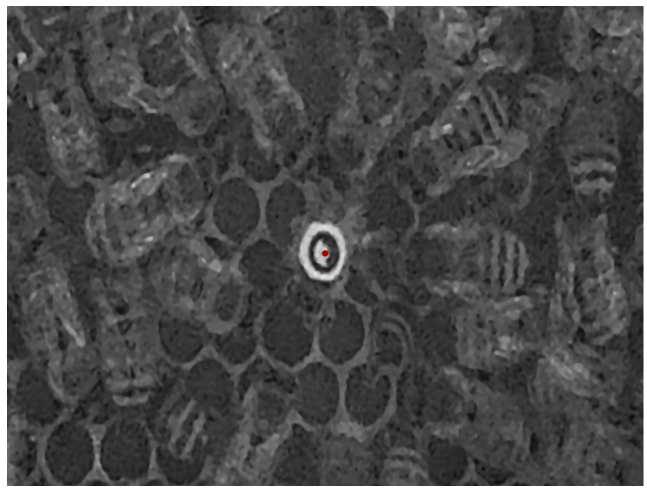

Fig. 8. Detail of an annotated frame of the 'Occluded' dataset.

The 'Occlusions' dataset was manually annotated in order to get the positions of the marker in pixels in each frame. We denote the manually obtained position of the marker in the i-th frame of the dataset as $\mathbf{x}_i^{gt} = (x_i^{gt}, y_i^{gt})$ - the position is in pixels.

B. Localisation error

To calculate the localisation error, we analysed the 'Occluded' dataset by the WhyCode and WhyComb methods, as well as by a convolution-based detection described in Section III-A. The positions of the marker provided by a particular method m are denoted as $\mathbf{x}_i^m = (x_i^m, y_i^m)$ In the case that the method failed to detect the marker in the ith frame, we carry the last detection over from the previous one, that is, $\mathbf{x}_i = \mathbf{x}_{i-1}$ in the case of no detection. The localisation error in the ith frame of a method m is calculated as $e_i = ||\mathbf{x}_i^m - \mathbf{x}_i^{gt}||$. Thus, each method m produced a vector $\mathbf{e}^m$ consisting of the values of e_i^m. The error vectors $\mathbf{e}^m$ are the primary data entry for the subsequent analysis.

C. Statistical testing

To conclude which method performs statistically significantly better, we used the Wilcoxon pairwise test on pairs of error vectors $\mathbf{e}^m$ created from the same dataset. The Wilcoxon pairwise test is a non-parametric test used for two independent samples that do not need to have a Gaussian distribution. Its null hypothesis is that, for randomly selected values X and Y from two populations, the probability that X is greater than Y is equal to the probability that Y is greater than X. Rejecting the null hypothesis means that the two samples were drawn from different distributions.

We performed the Wilcoxon test on the WhyCode-WhyComb and WhyCode-convolution error vector pairs. In both cases, the p-value of the Wilcoxon pairwise test was less than 0.001, indicating that the precision of these methods differs in a statistically significant way. Comparison of the mean error values of the methods, as shown in I, indicates that the WhyComb method outperforms the original WhyCode method in terms of accuracy.

TABLE I
AVERAGE ERROR OF THE PROPOSED WHYCOMB METHOD COMPARED TO THE ORIGINAL WHYCODE AND CONVOLUTION-BASED DETECTION.

Method	WhyComb	WhyCode	Convolution
Average error [px]	4.3	10.5	134.2

D. Qualitative analysis

For a qualitative comparison of the methods, we computed the cumulative distribution functions of the error vectors. The cumulative distribution functions show the probability that the localisation error of a particular method is lower than a certain value. This allows a better qualitative comparison of the evaluated methods and provides a better insight into their performance.

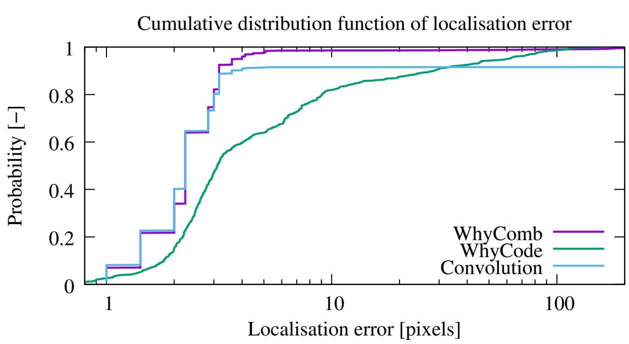

Fig. 9. The graph depicts cumulative distribution function errors of WhyCode, WhyComb and Convolution on the Occluded dataset. The function shows the probability that the method will have an error equal to or less than the threshold on the x axis.

Figure 9 shows that the WhyCode generally produces a higher localisation error, which is mainly caused by lower detection rates due to occlusions. It also indicates that, while the convolution is able to track the queen's position, it tends to produce errors higher than 100 pixels due to misdetections. The figure shows that the WhyComb method can track partially occluded markers, similarly to the convolution, and the WhyComb's misdetection rate is much lower than the other methods. Unlike WhyCode, the WhyComb and convolution were not designed to achieve subpixel resolution, which is reflected in the quantization of the results in the Figure 9.

E. Precision and recall

Finally, we calculate the precision and recall of the compared methods. Both the WhyComb and convolution will always retrieve an estimation of the position of the honey bee queen due to the convolution pattern matching will always find a local maximum. Therefore, the recall of these methods will be 100%. To determine the precision, we have set a threshold of localisation error to 7 pixels, which is the minimal marker radius.

TABLE II
COMPARISON OF THE PRECISION AND RECALL OF THE INDIVIDUAL METHODS

Method	Precision	Recall
WhyComb	98.7%	100.0%
WhyCode	97.6%	40.5%
Convolution	92.8%	100.0%

Table II summarises the precision and recall of the evaluated methods. The results indicate that WhyComb achieved higher recall and precision compared to WhyCode. The WhyComb's precision also exceeded convolution-based detection, making WhyComb superior to both baseline methods.

F. Testing the filters

Finally, to evaluate the filters implemented in Section III-B, we used a subset of the 'FalsePos' dataset. The inner ellipse filter removed 81.2% of the false detections, and a subsequent size check improved the rejection rate to 98.7%. Thus, only eight images remained out of the original 635 false positive detections produced by the WhyCode method on the 27000 image sequence. This indicates that the precision of the WhyComb method, as shown in Table II, is caused by imperfect results of the convolution rather than by the WhyCode-based detection step.

V. CONCLUSION

This paper presented a vision-based system for detecting and tracking a honeybee queen. To achieve real-time performance despite the large image resolution and the small marker size, we decided to base our system on the WhyCode method [23], [28]. We combined the WhyCode method with convolution-based detection and additional filtering steps to resolve issues with frequent occlusions and misdetections. This resulted in novel detection and localisation methods, dubbed WhyComb. Our experiments, based on infrared images gathered in real beehive colonies, showed that compared to the other methods, the WhyComb achieves higher accuracy, precision and recall while maintaining real-time performance.

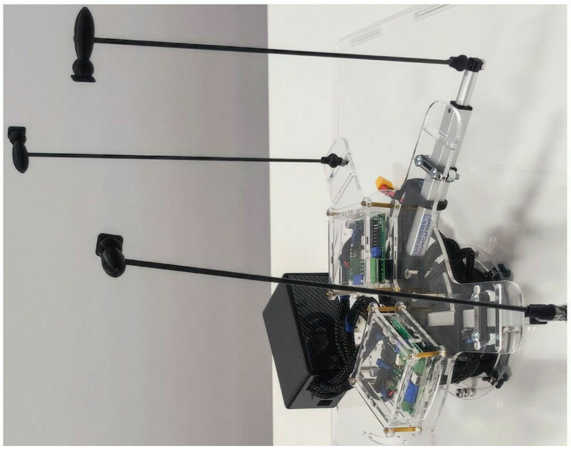

Fig. 10. Prototype RoboRoyale manipulator with passive agents.

To ensure interoperability with the other system components and easy data handling, we implemented it as a Robot Operating System (ROS) node. This allowed its smooth integration into a complex system for long-term observations of honeybee queen activity inside a honeybee colony. The system started to operate in March 2021, and over its lifetime, it produced more than 100 million images of honeybee queens, showing how they interact with the rest of the colony.

Furthermore, individual components of the RoboRoyale manipulator prototype, see Figure 10, were deployed to the observation beehive [12], and the tracking method was used to guide the manipulator head in a closed-loop manner, i.e., the manipulator's head was physically tracking the queen. This allowed us to boost further the resolution and quality of the images obtained.

REFERENCES

[1] Stefanec *et al.*, "A minimally invasive approach towards "ecosystem hacking" with honeybees," *Frontiers in Robotics and AI*, vol. 9, 2022.

[2] B. Danforth, "Bees," *Current biology*, vol. 17, no. 5, pp. R156–R161, 2007.

[3] J. Van Der Steen, "The foraging honey bee," *BBKA News-The British Bee Journal*, vol. 2015, no. February, pp. 43–46, 2015.

[4] T. D. Seeley, *The wisdom of the hive: the social physiology of honey bee colonies*. Harvard University Press, 2009.

[5] C. Rowe, "Receiver psychology and the evolution of multicomponent signals," *Animal behaviour*, vol. 58, no. 5, pp. 921–931, 1999.

[6] K. Von Frisch, "The dance language and orientation of bees," in *The dance language and orientation of bees*. Harvard University Press, 2013.

[7] L. Bortolotti and C. Costa, "Chemical communication in the honey bee society," *Neurobiology of chemical communication*, 2014.

[8] A. Maisonnasse, C. Alaux, D. Beslay, D. Crauser, C. Gines, E. Plettner, and Y. Le Conte, "New insights into honey bee (apis mellifera) pheromone communication. is the queen mandibular pheromone alone in colony regulation?" *Frontiers in Zoology*, vol. 7, no. 1, pp. 1–8, 2010.

[10] S. E. Hoover, C. I. Keeling, M. L. Winston, and K. N. Slessor, "The effect of queen pheromones on worker honey bee ovary development," *Naturwissenschaften*, vol. 90, no. 10, pp. 477–480, 2003.

[9] N. E. Gary, "Chemical mating attractants in the queen honey bee," *Science*, vol. 136, no. 3518, pp. 773–774, 1962.

[11] C. I. Keeling, K. N. Slessor, H. A. Higo, and M. L. Winston, "New components of the honey bee (apis mellifera l.) queen retinue pheromone," *Proceedings of the National Academy of Sciences*, vol. 100, no. 8, pp. 4486–4491, 2003.

[12] F. Rekabi-Bana, M. Stefanec, J. Ulrich, E. E. Keyvan, T. Rouček, G. Broughton, B. Y. Gunderer, O. Şahin, A. E. Turgut, S. Erol, T. Krajník, T. Schmickl, and F. Arvin, "Mechatronics designs for multi robots-insect swarms interactions," in *IEEE International Conference on Mechatronics*, 2023, in review.

[13] R. Scheiner, C. I. Abramson, R. Brodschneider, K. Crailsheim, W. M. Farina, S. Fuchs, B. Gruenewald, S. Hahshold, M. Karrer, G. Koeniger *et al.*, "Standard methods for behavioural studies of apis mellifera," *Journal of Apicultural Research*, vol. 52, no. 4, pp. 1–58, 2013.

[14] R. Menzel and W. Backhaus, "Color vision honey bees: Phenomena and physiological mechanisms," in *Facets of vision*. Springer, 1989, pp. 281–297.

[15] A. Corovic, V. Ilic, S. Duric, M. Marijan, and B. Pavkovic, "The real-time detection of traffic participants using yolo algorithm," in *2018 26th Telecommunications Forum (TELFOR)*, 2018, pp. 1–4.

[16] S. A. Yaseen and S. Sasi, "Robust algorithm for object detection and tracking in a dynamic scene," *Journal of Image and Graphics*, vol. 2, no. 1, pp. 41–45, 2014.

[17] K. Bozek, L. Hebert, A. S. Mikheyev, and G. J. Stephens, "Pixel personality for dense object tracking in a 2D honeybee hive," *CoRR*, vol. abs/1812.11797, 2018. [Online]. Available: http://arxiv.org/abs/1812.11797

[18] K. Bozek, L. Hebert, Y. Portugal, A. S. Mikheyev, and G. J. Stephens, "Markerless tracking of an entire honey bee colony - nature communications," *Nature*, May 2021. [Online]. Available: https://doi.org/10.1038/s41467-021-21769-1

[19] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *International Conference on Robotics and Automation*, 2011.

[20] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[21] P. R. Mendonça, A. Hopper *et al.*, "Trip: A low-cost vision-based location system for ubiquitous computing," *Personal and Ubiquitous Computing*, vol. 6, no. 3, pp. 206–219, 2002.

[22] T. Krajník, M. Nitsche, J. Faigl, P. Vaněk, M. Saska, L. Přeučil, T. Duckett, and M. Mejail, "A practical multirobot localization system," *Journal of Intelligent & Robotic Systems*, 2014. [Online]. Available: http://dx.doi.org/10.1007/s10846-014-0041-x

[23] J. Ulrich, A. Alsayed, F. Arvin, and T. Krajník, "Towards fast fiducial marker with full 6 dof pose estimation," in *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, 2022, pp. 723–730.

[24] J. D. Crall, N. Gravish, A. M. Mountcastle, and S. A. Combes, "Beetag: a low-cost, image-based tracking system for the study of animal behavior and locomotion," *PloS one*, vol. 10, no. 9, p. e0136487, 2015.

[25] B. Wild, L. Sixt, and T. Landgraf, "Automatic localization and decoding of honeybee markers using deep convolutional neural networks," *arXiv preprint arXiv:1802.04557*, 2018.

[26] F. Wario, B. Wild, M. Couvillon, R. Rojas, and T. Landgraf, "Automatic methods for long-term tracking and the detection and decoding of communication dances in honeybees," *Frontiers in Ecology and Evolution*, vol. 3, 2015.

[27] F. Boenisch, B. Rosemann, B. Wild, D. Dormagen, F. Wario, and T. Landgraf, "Tracking all members of a honey bee colony over their lifetime using learned models of correspondence," *Frontiers in Robotics and AI*, vol. 5, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2018.00035

[28] P. Lightbody, T. Krajník, and M. Hanheide, "An efficient visual fiducial localisation system," *ACM SIGAPP Applied Computing Review*, vol. 17, no. 3, pp. 28–37, 2017.