Automated Mosquito Dissection – Vision

Critical Review of

Multi-mosquito object detection and 2d pose estimation for automation of PfSPZ malaria vaccine production

Alan Lai
alai13@jhu.edu

Mentors

Dr. Russell Taylor
rht@jhu.edu

Balazs Vagvolgyi
balazs@jhu.edu

2020/03/04

Project Goals

The goal of our project is to create a ROS-integrated computer vision system for mosquito detection and keypoint identification, in order to guide an automated mosquito dissection robotic system for live malaria vaccine production. The target keypoints on the mosquito to be identified with the vision system include the proboscis, the location where the robot grasps and manipulates the mosquito, and the neck, which the robot aligns with the decapitation blade to maximize vaccine yield from the salivary glands. This project is in conjunction with Sanaria Inc. (Rockville, MD).

Paper Selection and Relevance

The selected paper, by Wu et al., is Multi-mosquito object detection and 2d pose estimation for automation of PfSPZ malaria vaccine production. The paper, written by members of our lab, aims to solve a problem that aligns almost exactly with our project goal, which is to create a ROS integrated computer vision system to support an automatic Mosquito Microdissection System (MMS) for production of a malaria vaccine. Though both the paper and our project aims to solve the same problem, there are a few differences, including the stages at which the vision algorithms are to be used to support robotic manipulation, and also the general location and orientation of the mosquitoes when vision support is required. Despite the differences, the author's implementation of both a deep learning and an image processing workflow for mosquito orientation and pose estimation, along with a comparison of these two methods remain extremely relevant to our project, providing inspiration and reference for our own computer vision algorithms.

Key Results

There are three key contributions made by the selected paper

1. The paper provides validation of the approach of using deep learning (DL), specifically Mask R-CNN and DeeperCut, for mosquito detection and pose estimation.
2. DL has been shown to outperform image processing approaches in terms of mosquito detection and pose estimation, but falls short in terms of processing speed.
3. DL approach has also been shown to be more robust in terms of varying luminance and scale compared to image processing approaches, which depends on heuristics

Altogether, these contributions help validate our exploration of using DL approaches to solve for mosquito pose estimation and orientation classification in our project.

Background Information

Malaria is a global problem, having caused over 400,000 deaths and over $12 billion USD of loss in 2017 worldwide. Despite the obvious need for a solution, no effective malaria vaccine currently exists in the market. In an effort to address this issue, Sanaria, a biotechnology company based in Rockville MD, has developed a live malaria vaccine that has been shown to be 100% effective in clinical trials.

Because of the nature of live vaccines, Sanaria's Plasmodium falciparum (Pf) sporozoite (SPZ) based vaccine (PfSPZ) has to be extracted from mosquito salivary glands. Initially, this was done manually, using hand tools and syringes, but Schrum et al. were able to develop a semi-automated Mosquito Microdissection System (saMMS) that greatly improved workflow[2]. Despite the advancements, Sanaria's goal of being able to disseminate the vaccine around the world requires a fully automated system in order to meet the demand. Hence an automatic MMS has been proposed by Phalen et al.,

with the use of robots to manipulate the mosquitoes and extract the salivary glands[3]. To support robot manipulation of mosquitoes, a computer vision system is required to identify keypoints on the mosquito. Some keypoints, including the proboscis, where gripping of the mosquito by the robot is done, and the neck, where the decapitation point is to be, are shown on the figure on the right. To tackle this problem, the authors implemented a deep learning workflow and an image processing workflow separately to identify these key points, validate these approaches, and compare the efficacy of using deep learning methods as compared to traditional image processing methods.

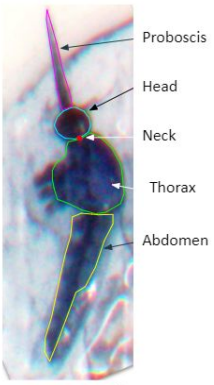

Figure 1. Parts of a mosquito

Experimental Methods

Dataset

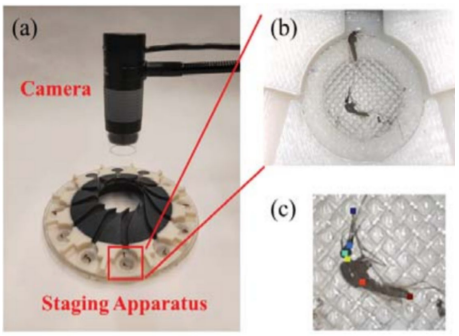

Figure 2. Dataset. (a) camera staging apparatus for MMS (b) image taken from camera (c) keypoints labeled on mosquito (Wu)

The dataset used for the experiment was collected by the authors, by placing the mosquitoes into the expected staging apparatus for the MMS either manually or automatically via water transfer, which is the expected method with which the mosquitoes will be put onto the staging apparatus. What the authors did well was to account for the possibility of varying luminance and water flow conditions, and hence collected images at varying luminances, scales, and water flow. The 1460 images collected were then manually labeled. The staging apparatus is shown on the left in figure 2a, with the camera view shown in figure 2b, and mosquito with labeled keypoints in figure 2c. The labeled keypoints shown in figure 2c include the proboscis tip, proboscis end, head, neck, thorax, and abdomen tip. Bounding box labeling of non-clustered and clustered regions for mosquitoes was also done, shown in the middle image of figure 3, with red squares as clustered and blue as non-clustered.

Deep Learning

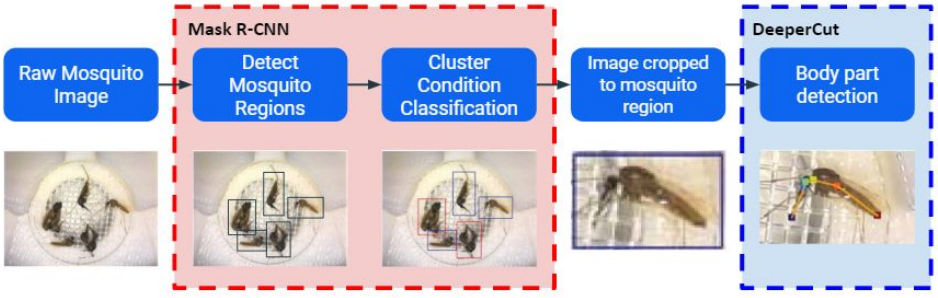

Figure 3. Process workflow for deep learning approach (Mosquito images from Wu)

For deep learning, mosquito detection and cluster classification was done using the Mask R-CNN network, and body part/keypoint detection via the DeeperCut network. Mask R-CNN was chosen as the region proposal network as it has been demonstrated to have state-of-the-art accuracy on object detection benchmarks, and DeeperCut was chosen as it has been shown to be effective in animal body part detection. Due to the limited size of the dataset, transfer learning was performed, with the Mask R-CNN pretrained on the COCO dataset, and DeeperCut with ImageNet dataset. With regards to the workflow, shown in figure 3, the camera images are first passed into Mask R-CNN, where regions containing mosquitoes are identified. The same network also classifies these regions into either non-clusters (blue) or clusters (red). Clusters occur when two or more mosquitoes appear in the same region; these areas are ignored by the robot as untangling mosquitoes is a difficult task, and these

mosquitoes will be washed back into the water storage, where hopefully they will untangle. Subsequent to finding the non-clusters, the image is cropped to those regions, and these regions are fed into the DeeperCut network, where the six keypoints (proboscis tip, proboscis base, center of head, neck, thorax, and abdomen tip) are subsequently identified and labeled by the network.

Image Processing

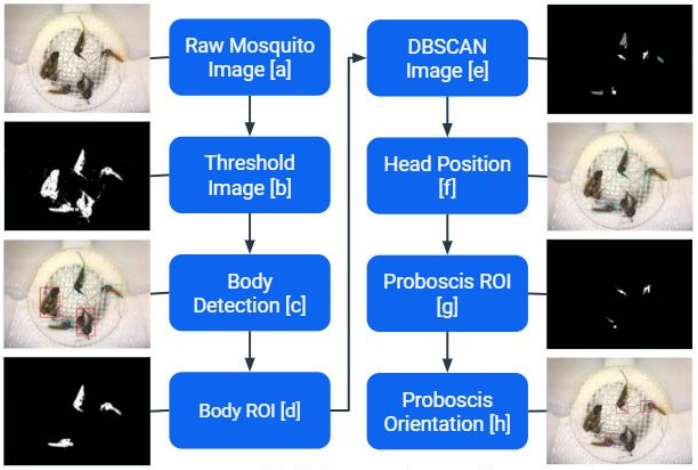

Figure 4. Process workflow for image processing approach (Images from Wu)

With regards to image processing, a multistage approach is proposed. The raw images are first thresholded to get a binary image (4b), which is then put through a watershed algorithm to separate different clusters from each other, allowing for body detection. Like in the DL approach, the authors are only interested in non-clustered regions, and this was obtained by comparing the aspect ratios of the proposed regions to determine which regions correspond to non-clusters (4c). Once the body ROI has been identified, body removal is done, where the thorax is eroded from the image, and DBSCAN is used to cluster the head regions (4e). Hough transform is then performed to detect the head position (4f). Once the head position is obtained, heuristically the proboscis has to be attached to the head, and hence Hough Line Transform is used in the region around the detected head (4g) to determine the orientation and endpoints of the proboscis (4h).

Results

Deep Learning

For the evaluation of bounding box regression for mosquito detection, an IoU > 0.75 is defined as a true positive. At this threshold:

- Precision for non-clustered and clustered mosquito classes are 0.86, 0.88 respectively
- Average precision (AP) for non-clustered and clustered classes are 0.85, 0.82 respectively
- Mean average precision (mAP) is 0.84

For the evaluation of body part/keypoint detection, the root mean squared error (RMSE) measure is used. On the image scales reported by the author, 1 pixel corresponds to 0.05 mm. We see in figure 5 that the average RMSE of all body parts is around 3.1 pixels, while the neck is more accurate at 2 pixels. However, the proboscis tip, vital for the robot to grip and manipulate the mosquito, has a higher RMSE of 7.1 pixels. The authors claimed that this was due to the similarity of morphology of the leg and proboscis, causing misidentification of mosquito legs as the proboscis. However, the authors assured that despite the higher RMSE for the proboscis tip, 7.1 pixels is translates to around 0.35 mm, and as the robot gripper in the MMS has a tolerance error of 0.5 mm, they are confident that the error is within bounds and will not be an obstacle against the use of this algorithm for the MMS.

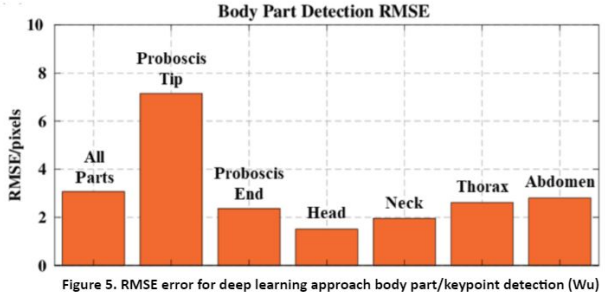

Figure 5. RMSE error for deep learning approach body part/keypoint detection (Wu)

Comparison between DL and Image Processing

From figure 6, we see that the deep learning approach outperforms the image processing approach for both mosquito bounding box identification and body part detection. At an IoU > 0.75, DL is able to achieve greater mAP and recall, and has a lower RMSE for head detection. In evaluating the image processing approach, the authors identified the dependence of the image processing algorithm on luminance and scale conditions. Hence the results shown in figure 6 are of images with a constant luminance and scale, and the authors have correspondingly noted the robustness of the DL approach with respect to the variance of luminance and scale. The one thing that the image processing algorithm excels at is processing speed, as it was able to process and label images eight times as fast as the DL approach.

Approach	**Deep Learning**	**Image Processing**
Detection mAP (IoU>0.5)	0.97	0.80
Detection Recall	0.97	0.90
Head Position RMSE	1.61 pixels	2.70 pixels
Proboscis Orientation Error	14.3°	24.7°
Processing Speed	2.5 fps	20 fps

Figure 6. Comparison between DL and Image Processing Approaches (Wu)

Paper Assessment

The authors were able to validate the use of deep learning approaches, specifically with Mask R-CNN in bounding box regression and DeeperCut in keypoint detection, for mosquitoes. The results also suggest that the DL algorithm has sufficient accuracy in providing computer vision support for the proposed MMS to operate. Finally, it is clear from the paper that the DL approach outperforms the proposed image processing approach.

One of the strengths of the paper was the comprehensiveness of the dataset that was collected. Though there were only 1460 images in total, the dataset images included a wide variety of possible conditions, including varying luminance, scale, and even water flow conditions. This made the DL algorithm much more robust to these variations, which the author discussed when comparing the DL and image processing approaches. However, the author could have added more detail about the dataset in the paper by specifying, for example, how many of the images were manual placements vs water transferred, or the degree of luminance, for the reader to get a better understanding of the dataset. Since the dataset is always a crucial portion of deep learning, having this additional detail would have conveyed a better understanding of the deep learning techniques used to the reader.

The comprehensiveness of the entire workflow of both the DL and image processing techniques was impressive. The authors were able to create models that took raw images and returned accurate keypoint detection for both methods essentially from scratch, and the evaluation of the deep learning model was comprehensive. The comparison between DL and image processing approaches was also a strength of the paper, being able to not only show that DL approaches are accurate and precise enough to support robotic workflow for the MMS, but also demonstrate that DL is able to outperform and is more robust than alternative image processing approaches. This further lends credence into the validity of using DL approaches to tackle the problem.

However, despite these strengths, there were still some limitations of the study. One of these includes the lack of justification or comparison between different neural networks. Though Mask R-CNN may be state-of-art for object detection benchmarks, there are many other networks that can perform just as well, such as Cascade R-CNN. Though Mask R-CNN may outperform them for certain datasets, the blackbox nature of neural networks means that different networks will perform differently on different datasets, and hence exploration with different neural networks would have made the choice of using

Mask R-CNN more robust. The same can be said for DeeperCut. A comparison of the accuracy for different networks would have increased support for these design choices..

Another limitation is the lack of reported results for the image processing approach. There are no results for the image processing approach for RMSE of detected keypoints save for the head position. The lack of these results for the image processing approach then makes the comparison incomplete, as the accuracy and precision of the image processing approach is not fully characterized. This lack of focus suggests that the image processing approach was created solely for the use of comparison with the DL approach, and hence suggests that the authors may have spent less time on creating a viable image processing algorithm, detracting from the comparison results with the DL approach.

Finally, though the results are able to capture succinctly the performance of the approaches, the authors could have extended the evaluation by testing these algorithms on the automated MMS or the robotic system. This would provide concrete results as to whether these approaches would solve the proposed problem, which would further justify and validate their approaches, or highlight areas of improvement if these approaches were not successful. Results such as RMSE can be effective in comparing approaches on paper, but if both approaches have a 100% success rate when integrated with the robotic system (as the error is within sufficient margins), then image processing would actually be superior due to its faster processing speed. Hence the comparison would have been more robust if evaluation metrics that pertain more to the problem context were explored in addition to these standard evaluation metrics.

Despite the limitations, this paper has laid important groundwork for and validated the use of DL approaches for mosquito detection and keypoint detection. Our current project of creating a computer vision system to support the MMS has been inspired by this paper, which has also set the benchmarks for accuracy and precision that our system must at least match, and hopefully exceed. The success of using Mask R-CNN and DeeperCut has inspired me to incorporate these networks with my own DL model, while the shortcomings of the image processing techniques have guided our image processing design decisions. The strengths of this paper, including the comprehensive dataset and comparison with image processing techniques has also made us more aware of these factors and hence have taken this into account. The limitations of this paper have also highlighted the pitfalls to avoid, such as a lack of focus on pure image processing techniques, and inspired us to do better by exploring more neural networks and more image processing approaches.

References

1. Wu, H., Mu, J., Da, T., Xu, M., Taylor, R. H., Iordachita, I., & Chirikjian, G. S. (2019). Multi-mosquito object detection and 2d pose estimation for automation of PfSPZ malaria vaccine production. In 2019 IEEE 15th International Conference on Automation Science and Engineering, CASE 2019 (pp. 411-417). [8842953] (IEEE International Conference on Automation Science and Engineering; Vol. 2019-August). IEEE Computer Society. https://doi.org/10.1109/COASE.2019.8842953
2. M. Schrum, A. Canezin, S. Chakravarty, M. Laskowski, S. Comert, Y. Sevimli, G. S. Chirikjian, S. L. Hoffman, and R. H. Taylor, “An efficient production process for extracting salivary glands from mosquitoes,” arXiv:1903.02532, 2019.
3. H. Phalen, P. Vagdargi, M. Pozin, G. S. Chirikjian, I. Iordachita, R. H. Taylor, "Mosquito pick-and-place: Automating a key step in pfspz-based malaria vaccine production", Accepted to the 2019 15th IEEE International Conference on Automation Science and Engineering (CASE 2019).