Understanding human responses to errors in a collaborative human-robot selective harvesting task

Zhuoling Huang[1], Adrian Salazar Gomez[1], Richie Bird[1], Amar Singh Kalsi[1], Chipp Jansen[1], Zeyang Liu[1]
[1]*King's College London*, UK
zhuoling.huang@kcl.ac.uk

Genki Miyauchi[1,2]
[2]*University of Sheffield*, UK
g.miyauchi@sheffield.ac.uk

Simon Parsons[1,3], Elizabeth I. Sklar[1,3]
[3]*University of Lincoln*, UK
sparsons,esklar@lincoln.ac.uk

***Abstract*—A human-robot approach for the farm of the future motivates robotics researchers to consider ways in which automated devices and intelligent systems can work alongside farmers to address a wide range of highly specialised but often repetitive tasks. The work presented here investigates a collaborative task in which a human and robot share decision making about the readiness of strawberries for harvesting. Preliminary experimental results with two different robot behaviours and two different user groups are compared.**

***Index Terms*—Human-robot collaboration, Agriculture, Computer vision, Machine learning**

I. Introduction

Family farms provide 70% of the food consumed worldwide [1]. However, because young labour is moving out of the agricultural sector, the existing labour force is ageing and the labour pool is not being replenished, the sector faces many difficult challenges. One strategy for addressing some of these challenges is to develop intelligent human-robot solutions in which agriculture involves *human farm workers* collaborating with *robots* to perform a wide variety of tasks. The work presented here introduces a strawberry harvesting system, but the techniques could also apply to other high-value crops. A collaborative human-robot solution to high-value crop harvesting could entail tasks such as identifying which fruits are ready to pick, selecting an appropriate position for a robot manipulator, gently removing the fruit from its stalk and packing the fruit for shipping.

Along this selective harvesting pipeline, detecting the target is, without a doubt, a very important step. Recent research has applied *machine learning (ML)* to detect fruits among which variations of *YOLO* [2]–[4] and *R-CNN* [2], [5]–[8] have been proven to work well. However, none of the existing detection methods can guarantee perfect *precision* (percentage of selected answers that are correct) and *recall* (percentage of all correct answers that are selected) for the detection task—there are always some number of false positive (type I, incorrectly selected) and false negative (type II, missed selection) errors. Here we test the hypothesis that different human users will respond differently to these two different types of errors within the context of selective harvesting.

Supported by China Scholarship Council (CSC).

II. Experimental System Design

Basic structure of the setup. Our experimental system for collaborative strawberry detection includes several components: an emulated farm, a mobile robot, a vision server and a graphical user interface (GUI), as shown in Fig. 1. In the *emulated strawberry farm*, 43 high-resolution images of strawberry plants taken on four different real strawberry farms are presented. Of the images, 35 contain 51 mature strawberries in total, 3 contain green strawberries only, and 5 contain only farm background (i.e., leaves, grass, soil). Full-colour prints of these images are attached to the office corridor walls near our robotics lab at a height appropriate for the robot. The *robot* is a Turtlebot2 holding an Asus Xtion RGB camera, controlled using the MRTeAm framework [9] and Robot Operating System (ROS) [10]. The framework enables the robot to move following a command from the user, and to take images of the environment around it, which it passes to the vision server with twin RTX2080Ti GPUs. The *vision server* then uses these data to identify ripe strawberries.

Different methods can be applied to the object identification process. Here, we employed two different deep learning detection methods (details below): *Faster R-CNN*, a typical two-stage detection network, and *YOLOv3*, a widely used one-stage detection network. The *user interface* receives strawberry identification and ripeness estimates, as well as raw images. The user confirms or corrects the robot's estimate and sends their decision back to the vision server.

Here we are interested in comparing the responses of different user groups to different types of errors produced by different ML methods used to identify ripe fruit. We hope the results can help us to distinguish more suitable users to provide feedback for future reinforcement learning tasks and to choose preferred classifiers for these users.

Neural networks used for identification. Many different object detection methods can be applied to this system. Since one of our aims is to test the overarching hypothesis that people will respond differently to two common types of classification errors (*false positive* versus *false negative*), we chose two widely used methods for our pilot study.

YOLOv3 is a one-stage detection method, which does the localization and classification at the same time in a single

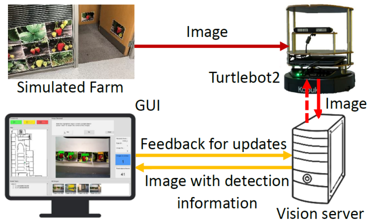

Fig. 1. Components and structure of the setup.

end-to-end network. One stage methods are usually faster and better able to support real-time operation than two-stage detection methods. YOLOv3 predicts bounding boxes and a corresponding "objectness" score for each (best predicted overlap with ground truth). We started with a YOLOv3 model with pre-trained weights from Darknet-53 that had been trained on the COCO data set [11]. We then trained it using 145 images drawn from a set of images taken by the authors at 4 strawberry farms in the UK and China. These images include 1497 ripe strawberries and 2047 unripe strawberries.

Faster R-CNN is a two-stage object detection approach that uses convolutional neural networks (CNNs) with two key components: a Regional Proposal Network (RPN) and a detection network. These share convolutional layers, making the model faster and more efficient [12], but can lead to mismatched goals in feature learning [13]. The Faster R-CNN used in our experiments follows the network structure proposed in [12], while the detection network following the RPN is based on a Fast R-CNN structure [14]. The network was trained on 540 images from our dataset, containing 3738 ripe strawberries, and 3735 unripe strawberries.

III. Experiment Design and Results

Human subjects were instructed to complete two *missions*, each with a different robot: one which employs Faster R-CNN and the other YOLOv3. For this experiment, the major difference is that Faster R-CNN produces more false positives while YOLOv3 produces more false negatives when applied to our experiment setup. To avoid human bias against the named algorithms, when talking about the two robots with the human subjects, we named them Robot ***F**elisa* for the **F**aster R-CNN behaviour and Robot ***Y**asmin* for the **Y**OLOv3 behaviour. The order of missions assigned was randomised. Thirty (30) human subjects participated, primarily postgraduate students, 15 with a background in deep learning (the experienced group), 15 with other backgrounds including other engineering communities, law and linguistics (the non-experienced group). They were asked to complete a pre-survey, then ran two missions (each with a different robot behaviour), completed a survey after each mission and a final survey after both missions. The survey questions were grouped according to four features: perceived *success*, *collaboration*, *trust* and *speed*. The two post-mission surveys have the same questions, but specify the name of each robot rather than just generically "robot". The order of survey questions was randomised. Answers indicate any predisposed user bias.

Comparison of classifiers. According to *objective data* collected by the system, we found that for the time-related measures and the total number of strawberries detected, there were no statistically significant differences between the two classifiers (using Student's t-test). However, there were statistically significant differences in the number of true positive (TP), false positive (FP) and false negative (FN) results. This trend, with Felisa providing more FPs and Yasmin providing more FNs, follows our experiment design and was also noticeable to the users. According to *subjective data* collected from surveys, users trust the two robot behaviours equally, but they prefer Felisa for completing this task faster as well as both in complex environments and environments with clear pictures, since users believe that collaboration is quicker with Felisa.

User background comparison. The objective data shows that there is no significant difference in performance between the experienced group and the non-experienced group when collaborating with either robot behaviour. This means that our system is suitable for users with different backgrounds. However, the subjective data shows different attitudes towards the classifiers from different groups of users. As shown in Fig. 2, the groups of users have different expectations for collaboration and speed, but the two missions reduced the difference between them. Overall, experienced users trust Yasmin more while non-experienced users trust Felisa more. Considering preferences for different tasks, the same difference holds for clear pictures. However, both groups prefer Felisa for complex environments and when speed is prioritised.

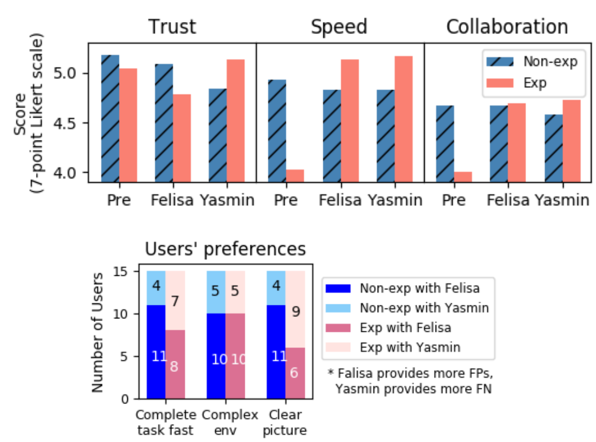

Fig. 2. Subjective data from different user groups. Trust, speed and collaboration are reported on a 7-point Likert scale.

IV. Summary

Our results show that our human-robot collaborative strawberry harvesting system could be successfully used for comparing people's preferences between the two error types when working with robots. Experienced users trust the robot that provides more false negative results and non-experienced users are the opposite. However, the majority of both groups prefer to work with the robot that produces more false positive results when completing tasks in complex environments.

REFERENCES

[1] "Proc of the Conf on Family Farming: A dialogue towards more sustainable and resilient farming in Europe and the world," European Commission, Nov 2013, ec.europa.eu/agriculture/events/family-farming-conference-2013_en.htm.

[2] H. Kang and C. Chen, "Fast implementation of real-time fruit detection in apple orchards using deep learning," *Computers and Electronics in Agriculture*, vol. 168, p. 105108, 2020.

[3] Y.-P. Liu, C.-H. Yang, H. Ling, S. Mabu, and T. Kuremoto, "A visual system of citrus picking robot using convolutional neural networks," in *2018 5th International Conference on Systems and Informatics (ICSAI)*. IEEE, 2018, pp. 344–349.

[4] C. Liang, J. Xiong, Z. Zheng, Z. Zhong, Z. Li, S. Chen, and Z. Yang, "A visual detection method for nighttime litchi fruits and fruiting stems," *Computers and Electronics in Agriculture*, vol. 169, p. 105192, 2020.

[5] Y. Yu, K. Zhang, L. Yang, and D. Zhang, "Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN," *Computers and Electronics in Agriculture*, vol. 163, p. 104846, 2019.

[6] M. Halstead, C. McCool, S. Denman, T. Perez, and C. Fookes, "Fruit quantity and ripeness estimation using a robotic vision system," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 2995–3002, 2018.

[7] X. Liu, D. Zhao, W. Jia, W. Ji, C. Ruan, and Y. Sun, "Cucumber fruits detection in greenhouses based on instance segmentation," *IEEE Access*, vol. 7, pp. 139 635–139 642, 2019.

[8] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 3626–3633.

[9] E. Schneider, E. I. Sklar, S. Parsons, and A. T. Özgelen, "Auction-based task allocation for multi-robot teams in dynamic environments," in *Towards Autonomous Robotic Systems: 16th Annual Conference (TAROS)*. Springer, 2015.

[10] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source robot operating system," in *ICRA OSS Workshop*, 2009.

[11] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015.

[13] B. Cheng, Y. Wei, H. Shi, R. Feris, J. Xiong, and T. Huang, "Revisiting RCNN: On awakening the classification power of Faster RCNN," in *Proc of European Conference on Computer Vision (ECCV)*, 2018.

[14] R. Girshick, "Fast R-CNN," in *Proc of IEEE intl conf on computer vision*, 2015.