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

Zhuoling Huang[1], Genki Miyauchi[1,3], Adrian Salazar Gomez[1], Richie Bird[1], Amar Singh Kalsi[1], Chipp Jansen[2], Zeyang Liu[1], Simon Parsons[1,4] and Elizabeth I Sklar[2,5]
[1]Dept of Informatics, King's College London, Strand, WC2R 2LS, UK
[2]Dept of Engineering, King's College London, Strand, WC2R 2LS, UK
[3]Dept of Automatic Control & Systems Engineering, University of Sheffield, S1 3JD, UK
[4]College of Science, University of Lincoln, LN6 7TS, UK
[5]Lincoln Institute for Agri-food Technology, University of Lincoln, LN6 7TS, UK
Contact authors: *zhuoling.huang@kcl.ac.uk, 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.

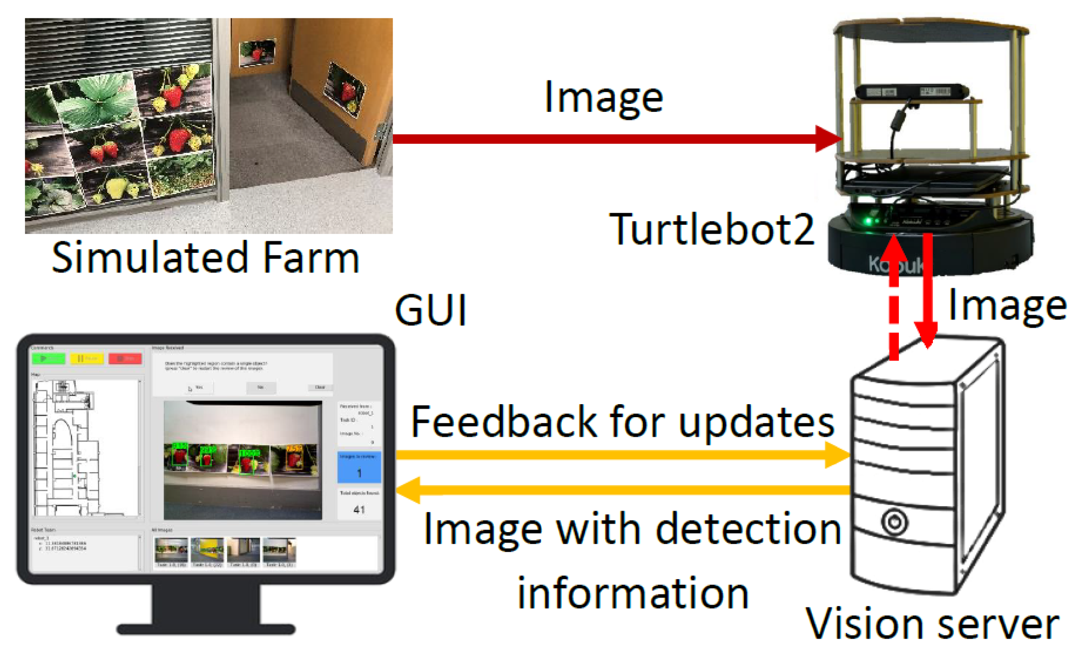
Introduction

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* [6, 9, 7] and *R-CNN* [13, 5, 6, 8, 2] 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.

Experimental System Design

Our experimental system for collaborative strawberry detection includes several components as shown in figure below.



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.

Robot: the robot is a Turtlebot2 holding an Asus Xtion RGB camera, controlled using the MRTeAm framework [12] 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.

GUI: 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.

Vision server: uses these data to identify ripe strawberries with twin RTX2080Ti GPUs. Different methods can be applied to the object identification process. Here, we employed two different deep learning detection methods (details below): *Faster R-CNN* and *YOLOv3*.

- *YOLOv3* is a one-stage detection method, which does the localization and classification at the same time in a single 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).
- *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 [11], but can lead to mismatched goals in feature learning [3]. The Faster R-CNN used in our experiments follows the network structure proposed in [11], while the detection network following the RPN is based on a Fast R-CNN structure [4].

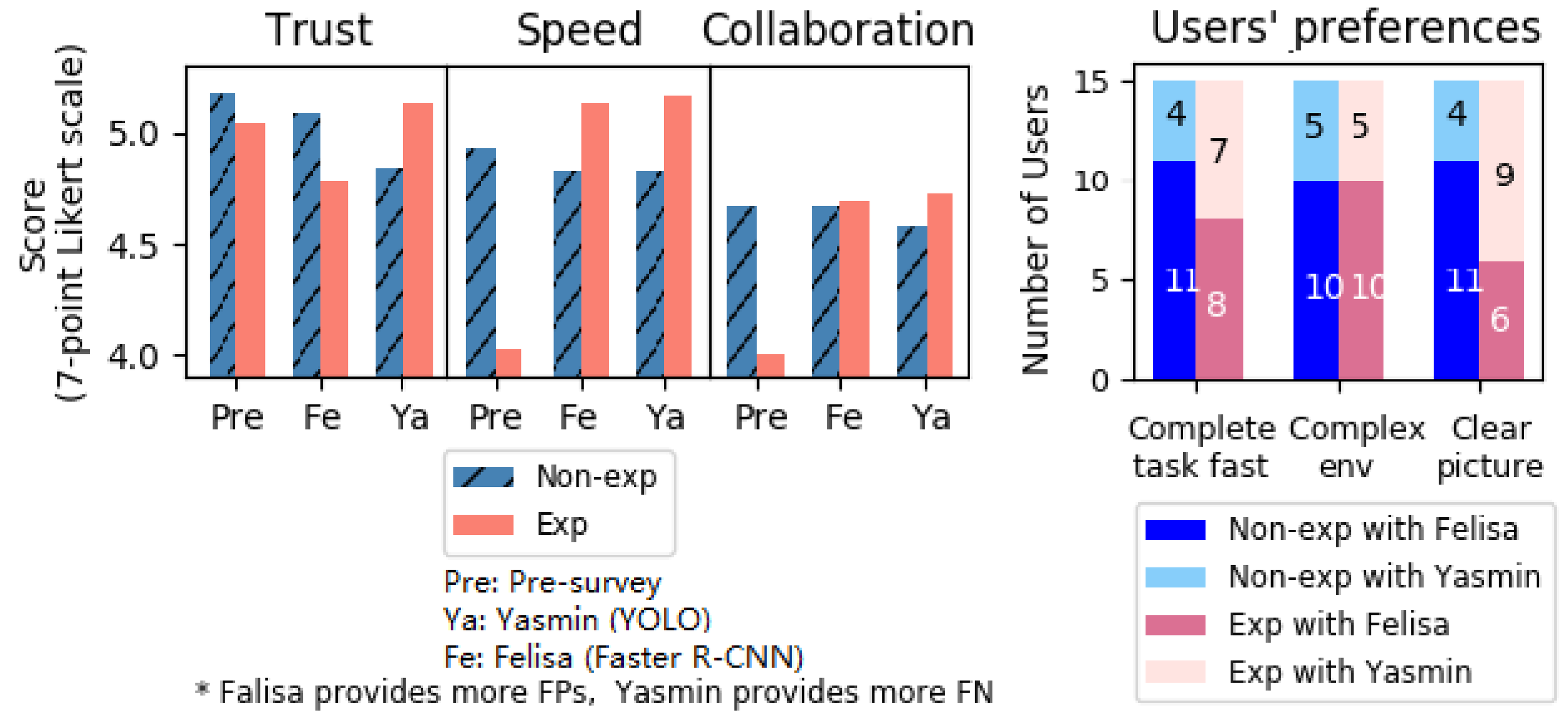
Experiment Design

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 reduce human bias against the named algorithms, when talking about the two robots with the human subjects, we referred to them as 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.

Experiment Results

Comparison of classifiers. According to *objective data*, 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. The trend, with which robot provides more FPs or FNs, follows our experiment design and was noticeable to the users. According to survey *subjective data*, users trust the two robot behaviours equally, but they prefer Felisa (more FPs) for completing this task faster, 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 two groups 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 figure below, 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 FNs) more while non-experienced users trust Felisa (more FPs) 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.



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] Suchet Bargoti and James Underwood. Deep fruit detection in orchards. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3626–3633. IEEE, 2017.

[3] Bowen Cheng, Yunchao Wei, Honghui Shi, Rogerio Feris, Jinjun Xiong, and Thomas Huang. Revisiting RCNN: On awakening the classification power of Faster RCNN. In *Proc of European Conference on Computer Vision (ECCV)*, 2018.

[4] Ross Girshick. Fast R-CNN. In *Proc of IEEE intl conf on computer vision*, 2015.

[5] Michael Halstead, Christopher McCool, Simon Denman, Tristan Perez, and Clinton Fookes. Fruit quantity and ripeness estimation using a robotic vision system. *IEEE Robotics and Automation Letters*, 3(4):2995–3002, 2018.

[6] Hanwen Kang and Chao Chen. Fast implementation of real-time fruit detection in apple orchards using deep learning. *Computers and Electronics in Agriculture*, 168:105108, 2020.

[7] Cuixiao Liang, Juntao Xiong, Zhenhui Zheng, Zhuo Zhong, Zhonghang Li, Shumian Chen, and Zhengang Yang. A visual detection method for nighttime litchi fruits and fruiting stems. *Computers and Electronics in Agriculture*, 169:105192, 2020.

[8] Xiaoyang Liu, Dean Zhao, Weikuan Jia, Wei Ji, Chengzhi Ruan, and Yueping Sun. Cucumber fruits detection in greenhouses based on instance segmentation. *IEEE Access*, 7:139635–139642, 2019.

[9] Yan-Ping Liu, Chang-Hui Yang, Huang Ling, Shingo Mabu, and Takashi Kuremoto. A visual system of citrus picking robot using convolutional neural networks. In *2018 5th International Conference on Systems and Informatics (ICSAI)*, pages 344–349. IEEE, 2018.

[10] Morgan Quigley, Ken Conley, Brian P. Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y. Ng. ROS: an open-source robot operating system. In *ICRA OSS Workshop*, 2009.

[11] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, 2015.

[12] Eric Schneider, Elizabeth I Sklar, Simon Parsons, and A Tuna Özgelen. Auction-based task allocation for multi-robot teams in dynamic environments. In *Towards Autonomous Robotic Systems: 16th Annual Conference (TAROS)*. Springer, 2015.

[13] Yang Yu, Kailiang Zhang, Li Yang, and Dongxing Zhang. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture*, 163:104846, 2019.

Acknowledgements: This research is supported by a China Scholarship Council (CSC) award to the first author.