The 11th Asian-Australasian Conference on Precision Agriculture (ACPA 11)
October 14-16, 2025, Chiayi, Taiwan

YOLO STRAWBERRY MATURITY CLASSIFICATION AND HARVEST PRIORITY WITH 3D CAMERA

Jihyeon Yang[1], Ki-su Park[2], Woo-joo Choi[2], Seong-Hwan Lee[2], Yeong-Jin Kim[2], Dokyun Jung[2], Myongkyoon Yang[1, 3*]

[1]Department of Bioindustrial Machinery Engineering, Jeonbuk National University, Jeonju 54896, Republic of Korea.
[2]Department of Agricultural Machinery Engineering, Jeonbuk National University, Jeonju 54896, Republic of Korea.
[3]Institute of Agricultural Machinery & ICT Convergence, Jeonbuk National University, Jeonju 54896, Republic of Korea.

*Corresponding Author: yangmk@jbnu.ac.kr

Abstract

Accurate harvesting timing is essential to improve crop quality and productivity, and recent advances in agricultural automation have led to the emergence of fruit maturity classification and harvest optimization algorithms for agricultural robots as major technical challenges. This study proposes a pipeline for strawberry object detection, maturity classification, distance estimation, and harvest priority. We train a YOLOv8 detector on an open RGB dataset, and estimate the camera-fruit distance by sampling depth from the center of the bounding box of RGB-D frames aligned on RealSense D435i. Then, we transform the 3D position of each target into world coordinates using the camera pose in the GCS estimated by AprilTag. We designate the in-screen harvest area ROI as the initial harvestable area, and assign a tracker-based persistence ID to the ROI target to prevent duplicate counting and maintain identity consistency. To support its operation on a mobile platform, WorldFrame 3D gating maintains identity between frames and viewpoints, mitigating transient outliers through median depth sampling and light time smoothing. All observations, including ID, class label, reliability, and 3D position, are recorded in CSV in real time for traceability and downstream analysis. This design translates frame-by-frame detection into a viable target, providing maturity and reachability clues for robotic manipulation and path planning. We evaluate the pipeline in the designed environment, and performance evaluation of detection accuracy, maturity classification performance, and distance estimation across operational scope, along with disruption of ROI policy and filtering choices. This approach goes beyond simple detection by enabling fast harvesting of marketable mature fruits, the architecture is sensor- and crop-agnostic, and proposes a general foundation for autonomous harvesting systems with clear paths to multi-camera settings and alternative anchoring via visual SLAM.

Keywords: 3D Depth Camera, GCS, Prioritization Algorithm, Strawberry, YOLOv8

INTRODUCTION

Accurate harvest-timing decisions and precise 3D localization directly affect strawberry quality and productivity. However, most previous studies have focused on improving maturity classification or detection accuracy, with limited connection to field-ready applications and harvest automation. To address this gap, we propose a real-time algorithm that integrates harvest-priority scoring, a global coordinate system (GCS) for world-frame localization, and tracker-based duplicate prevention on top of detection and classification, with linkage to path planning.

MATERIALS AND METHODS

Real-Time RGB-D Imaging Pipeline

RGB-D frames were collected with the Intel RealSense D435i (640×480, 30 fps). Color-depth matching was performed by rs.align, and depth values were converted in meters according to the device depth_scale. The world coordinate system was defined based on AprilTag, and when a tag was observed, the camera pose was continuously updated through this, and the RGB-D odometry or SLAM was temporarily used in the non-visible section and the drift was corrected during re-observation. Strawberry detection, ripe/unripe classification was performed simultaneously with a single YOLOv8 model, and only center-in detection with the center point included in the ROI was selected as a harvest candidate by defining the ROI in the screen. For the candidate, the 3D points (X, Y, Z) of the camera coordinate system were restored through the deproject of RealSense and converted into world coordinates with the current camera pose. All observations are managed by a global DB, and major items (time, id, class, dist, conf, counted) were recorded as real-time CSV.

RESULTS & DISCUSSION

This study designed a pipeline through one flow to detect, classify maturity, estimate distance, consistency of world coordinates, and prevent duplication. YOLOv8 acquired precision, recall, mAP50, and mAP50–95 of 90%, 83%, 90%, and 74%, respectively; the ROI policy reduced off-path detections, and world-coordinate 3D matching maintained ID stability on revision. Using ROI, unnecessary detections off the robot's driving route were reduced, and 3D matching based on world-coordinate maintained ID maintenance the ID stable even when revisiting objects. These results suggest a foundation framework for the proposed pipeline to be applicable in a real agricultural environment.

In particular, it functioned as a practical algorithm to maximize the robot's work efficiency by assigning harvest priority beyond simply limiting the detection area using ROI. In addition, the 3D object tracking and management function using world coordinates decisively contributed to increasing the accuracy of yield prediction by preventing time point changes or duplicate coefficients.

CONCLUSIONS

This study aimed to design a real-time 3D object recognition pipeline for agricultural robots and verify its effectiveness. To this end, by combining the Intel RealSense camera and the YOLOv8 model, the entire process from strawberry detection, maturity classification, 3D location estimation, and duplication prevention was implemented as one integrated flow. As a result of the experiment, the proposed pipeline achieved a high detection accuracy (mAP50) of 90% and successfully secured efficiency and data consistency in real work environments through ROI and object management based on world coordinates. Finally, this study is significant in that it presents a practical and basic methodology that can be applied to an actual unmanned harvesting system by unifying the complex recognition process. In the future, it is expected that the completeness of this system can be further improved through research that increases the robustness of unpredictable external environments such as lighting changes, severe masking, and various harvest environments.

ACKNOWLEDGEMENTS

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry(IPET) and Korea Smart Farm R&D Foundation(KosFarm) through Smart Farm Innovation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs(MAFRA) and Ministry of Science and ICT(MSIT), Rural Development Administration(RDA)(RS-2024-00399854).

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1C1C2005959).

REFERENCES

Bac, C. W., E. J. van Henten, J. Hemming, and Y. Edan. 2014. Real-time CNN-based computer vision system for open-field strawberry harvesting robot. *Journal of Field Robotics* 31(6): 888–911.

Kukartsev, V. V., Ageev, R. A., Borodulin, A. S., Gantimurov, A. P., & Kleshko, I. I. (2024). Deep learning for object detection in images: Development and evaluation of the YOLOv8 model using Ultralytics and Roboflow libraries. In R. Silhavy & P. Silhavy (Eds.), *Software Engineering Methods Design and Application* (Lecture Notes in Networks and Systems, Vol. 1118, pp. 629–637). Cham: Springer Nature Switzerland.

Lemsalu, M., V. Bloch, J. Backman, and M. Pastell. 2022. Real-time CNN-based computer vision system for open-field strawberry harvesting robot. *IFAC-PapersOnLine* 55(32): 24–29.