ROBOTIC ARM TOMATO HARVESTING SYSTEM AND NEXT BEST VIEW ALGORITHM DEVELOPMENT

Chia-Jung Yu, Sakir Kanmis, Ping-Lang Yen*

Department of Biomechatronics Engineering National Taiwan University, Taiwan.

*Ping-Lang Yen: plyen@ntu.edu.tw

ABSTRACT

Replacing human labor with robots is a trend for future agriculture due to its efficiency and consistency. However, in automatic fruit harvesting tasks, leaf occlusion and the dynamic orientation of fruit make it difficult for robots to directly observe the picking point. To address this problem, this research focuses on tomato harvesting, and proposes a next-best-view (NBV) algorithm based on two main structures: "tomato pose prediction" and a "target-hit-gain function". The goal of the research is to develop a next-best-view algorithm that can efficiently and accurately predict the optimal viewpoint which (1) can observe the target without occlusion and (2) adjusts to the tomato's orientation by maximizing the visible surface based on the current scene.

Results show that the system can successfully select the optimal view even when only one fifth or the bottom of the tomato are observed in the initial view, and can converge in 2–3 iterations, significantly reducing unnecessary robot movement.

Keywords: next best view algorithm, tomato harvesting, automation, agricultural robot, ROS2

INTRODUCTION

Similar algorithms have been used in object reconstruction. However, these cases often lack consideration of occlusion. Other cases are unsuitable for adjusting with dynamic tomato orientation due to calculating ray gain only from the observed views without performing model registration or pose prediction [3]. Our method, instead, uses ICP registration to predict the orientation in advance and select the best view based on target-hit-gain function (which scores candidate views by the number of rays that successfully hit the target tomato without being occluded). This allows the system to address with both leaf occlusion and target orientation at the same time.

In summary the contribution of this result are:

1. Proposed a next-best-view system based on ICP registration and target-hit-gain function that complete both occlusion avoidance and orientation adjustment tasks
2. Developed a next-best-view system specifically designed for the tomato harvesting scenario that predicts the optimal viewpoint where the camera can avoid the leaf occlusion and observed the largest area of the target tomato based on initial view.
3. Showing the developed NBV system can successfully predict and guide the robot arm camera to the optimal view, where the grasping point could be clearly observed, even when the initial occlusion rate is over 70%.

MATERIALS AND METHODS

Registration-and-target-hit-gain-based Next Best View Algorithm Overview

The proposed NBV algorithm consists of four segments: object detection by YOLOv8, tomato reconstruction by ICP registration, candidate view generation, and target-hit-gain calculation using ray casting method in Octomap and SDF (sign-distance-field) method.

Experiment Setup

In the experiment, the next-best-view algorithm is tested in two occlusion cases: straight tomato with occlusion on right and lied tomato with occlusion on the bottom. Each case is tested for different initial occlusion percentage conditions (defined at the following formula) around 20%, 50%, 80% for multiple tests. The following figure shows the initial scene of the two occlusion cases.

Figure 1: Case1-Left, Case2-right

Coverage	28%	43%	72%	86%
Initial Scene				

actual coverage	46% Ca	60%	81%
Initial Scene			

RESULTS & DISCUSSION

The results are shown in table1, in Case 1, the NBV system only required an average of 2.4 iterations, with most successful rates ranging between 75% and 85%. Although Test ID6 achieved only a 49% success rate after the third iteration, a satisfactory viewpoint had already been obtained during the second iteration. Remarkably, even when over 71% of the target tomato were occluded by leaf, the system was able to identify the optimal viewpoint within 4 iterations, with the 3rd and 4th iterations involving only minor adjustments in camera positioning. This demonstrates the robustness of the system under challenging initial conditions.

Table2 presents the results of Case 2, where the successful rate was consistently reached approximately 90%, notably higher than that in Case 1. However, when the occlusion rate exceeded 80%, the system occasionally failed to detect the tomato, highlighting a critical boundary condition for future improvement.

Table1: Result for case1-Left

TEST ID	Actual Initial Coverage Rate	Test	number of Iteration in movement	Final Occlusion Rate	Final Orientation Rate	Successful rate	Note
1	30.56%	Test1	1	1%	76%	75%	
2		Test2	2	3%	77%	75%	
3		Test3	3	11%	78%	70%	
4	42.42%	Test1	2	1%	86%	85%	
5		Test2	2	11%	92%	81%	
6	48.48%	Test1	3	45%	91%	49%	Got to red pixel rate for 0.82% at iteration 2 (successful rate 93%), but continue to calculate nbv in iteration 3 and end up in red pixel rate 0.48% (successful rate 55%)
7	71.59%	Test1	4	0%	89%	89%	
8	84.85%	Test1	x	x	x	0%	Fail to detect the tomato

Table 2 Result for case2-Right

TEST ID	Actual Initial Coverage Rate	Test	number of Iteration in movement	Final Occlusion Rate	Final Orientation Rate	Successful rate	Note
1	46.59%	Test1	3	0%	90%	90%	
2		Test2	4	2%	100%	98%	The Orientation red pixel rate is served as the Red rate at 30 cm facing directly to the tomato largest area
3	61.36%	Test1	1	3%	99%	96%	
4		Test2	2	24%	92%	70%	
5	81.82%	Test1	x	x	x	0%	Fail to detect the tomato

CONCLUSIONS

In both of the cases, the proposed next-best-view algorithm successfully demonstrates strong capability in finding the optimal viewpoint without leaf occlusion and adjusting with the dynamic tomato orientations within only 2–3 iterations, significantly reducing the occlusion rate from over 71% to 0%. This minimizes unnecessary robot movement when finding the grasping point. The system effectively reduces leaf occlusion while preserving alignment with the tomato's orientation, enabling the camera to capture grasping points that are otherwise invisible from the initial view. These results confirm that the NBV strategy provides a practical and efficient solution for robotic perception in dynamic agricultural environments.

REFERENCES

[1] McGreavy, C., Kunze, L., \& Hawes, N. (2016). Next best view planning for object recognition in Mobile ... Next Best View Planning for Object Recognition in Mobile Robotics. https://ceur-ws.org/Vol-1782/paper_6.pdf

[2] Breyer, M., Ott, L., Siegwart, R., \& Chung, J. J. (2022). Closed-Loop Next-Best-View Planning for Target-Driven Grasping. https://doi.org/https://arxiv.org/pdf/2207.10543.pdf

[3] Yi, T., Zhang, D., Luo, L., & Luo, J. (2024). View planning for grape harvesting based on active vision strategy under occlusion. IEEE Robotics and Automation Letters, 9(3), 2535–2542. https://doi.org/10.1109/LRA.2024.3357397

[4] M. H. Putra, Z. M. Yussof,K. C. Lim, S. I. Salim. (2018). Convolutional Neural Network for Person and Car Detection using YOLO Framework. View of convolutional neural network for person and car detection using Yolo Framework. https://jtec.utem.edu.my/jtec/article/view/3599/2491

[5] Kurtser, Polina \& Lowry, Stephanie. (2023). RGB-D datasets for robotic perception in site-specific agricultural operations-A survey. Computers and Electronics in Agriculture. 212. 108035. 10.1016/j.compag.2023.108035.