

Algorithm Design and Integration for a Robotic Apple Harvesting System

Kaixiang Zhang*, Kyle Lammers*, Pengyu Chu, Nathan Dickinson, Zhaojian Li**, and Renfu Lu

Abstract—Due to labor shortage and rising labor cost for the apple industry, there is an urgent need for the development of robotic systems to efficiently and autonomously harvest apples. In this paper, we present a system overview and algorithm design of our recently developed robotic apple harvester prototype. Our robotic system is enabled by the close integration of several core modules, including visual perception, planning, and control. This paper covers the main methods and advancements in deep learning-based multi-view fruit detection and localization, unified picking and dropping planning, and dexterous manipulation control. Indoor and field experiments were conducted to evaluate the performance of the developed system, which achieved an average picking rate of 3.6 seconds per apple. This is a significant improvement over other reported apple harvesting robots with a picking rate in the range of 7-10 seconds per apple. The current prototype shows promising performance towards further development of efficient and automated apple harvesting technology. Finally, limitations of the current system and future work are discussed.

I. INTRODUCTION

The apple industry relies heavily on manual labor. For instance, in the United States alone, it is estimated that the seasonal labor force needed for apple harvesting is more than 10 million worker hours each year, attributing to about 15% of the total production costs [1]. The growing labor shortage and increased labor cost have thus become major concerns for the long-term sustainability and profitability of the apple industry. In the meantime, the past decade has seen great transitions in apple production systems; traditional unstructured orchards have been replaced with high-density orchard systems where trees are smaller and more uniformly structured (i.e., v-trellis, vertical fruiting wall, etc.). These modern tree structures can greatly facilitate orchard automation, and thus there has been a renewed interest in pursuing robotic harvesting as a promising solution to reduce the harvesting cost and dependence on manual labor.

Over the past few years, several robotic systems have been designed to autonomously harvest different horticultural crops, including sweet pepper [2], strawberry [3], apple

[4], and kiwifruit [5]. For apple harvesting, the automation system designs can be mainly grouped into two categories. The first category is the *shake-and-catch* harvesting [6], where vibrations are applied to the tree trunk and/or branches to detach the fruits. Although the shake-and-catch harvesting systems are efficient in detaching fruits from trees, they often result in a high rate of apple bruising that is not acceptable for fresh market. The other category is the *fruit-by-fruit harvesting* where manipulators are used to pick fruits in a controlled manner, and thus can substantially reduce fruit damage. However, designing such systems with high picking efficiency and practical viability presents a great challenge.

So far, several fruit-by-fruit robotic apple harvesting systems have been developed [4], [7]–[10]. For instance, Baeton et al. combines a 7 degree-of-freedom (DOF) industrial manipulator with a vacuum activated, funnel shaped gripper for apple detachment, and the harvesting cycle time is 8-10 s/fruit [7]. In [4], both hardware and software designs of an apple harvester are presented. Field tests conducted on a v-trellis orchard show that this system is able to pick 84% of 150 apples attempted with the overall harvesting time being 7.6 s/fruit. In [8], Hohimer et al. developed a harvesting robot based on a pneumatic soft-robotic end-effector, and the average time that the system takes from apple detachment to transported to storage bin is 7.3 s/fruit. Despite the aforementioned progresses, the low picking efficiencies of existing systems are still unsatisfactory for their practical use in the real orchard environment [11].

Towards the goal of developing a practically and economically viable robotic harvesting system, we have been developing an efficient automated apple harvesting system over the past three years. Tests in orchard field and indoor simulated orchard environment demonstrated a promising picking rate of ~ 3.6 s/fruit, a significant improvement over the existing systems reviewed above. While the mechanical and preliminary control designs have been reported in [9], this paper presents the algorithm design and integration of the developed system, focusing on three major modules – perception, planning, and control – where several advancements have been made. First, we develop a deep learning-based multi-view fruit detection and localization framework by fusing two RGB-D sensors facing different angles. Compared to the single camera-based algorithm we developed earlier [12], this multi-view fusion offers enhanced performance in both detection and localization. Second, a unified planning algorithm that simultaneously optimizes picking sequence and dropping spots is developed, which significantly improves harvesting efficiency. Lastly, a

This work is supported by the USDA-ARS inhouse project 5050-43640-003-00D.

Kaixiang Zhang, Kyle Lammers, Pengyu Chu, and Zhaojian Li are with the Department of Mechanical Engineering, Michigan State University, East Lansing, MI 48824, USA (e-mail: zhangk64@msu.edu; lammer18@msu.edu; chupengy@msu.edu; lizhaoj1@egr.msu.edu).

Nathan Dickinson is with the Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA (e-mail: dicki112@msu.edu).

Renfu Lu is with the United States Department of Agriculture Agricultural Research Service, East Lansing, MI 48824, USA (e-mail: renfu.lu@usda.gov).

* The authors contribute equally to this paper.

** Zhaojian Li is the corresponding author.

computationally-efficient nonlinear controller is synthesized to enable accurate and smooth manipulator movement. Experiments in both an indoor simulated orchard environment and a real orchard field were conducted to illustrate the performance of the integrated system.

The remainder of this paper is organized as follows. Section II provides an overview of the developed robotic apple harvesting system. The algorithm designs of perception, planning and control modules are detailed in Section III, and experiment results are discussed in Section IV. Finally, conclusions are drawn in Section V.

II. SYSTEM OVERVIEW

The developed robotic apple harvesting system is shown in Fig. 1, which is comprised of four primary hardware components: a perception module consisting of two Intel RealSense D435i RGB-D cameras, a 3-DOF manipulator, a vacuum-based end-effector, and a dropping module. All components are affixed to a Segway mobility platform for ease of movement in orchard. The RGB-D cameras, the manipulator, and all communication devices are connected to an industrial computer (Xeon E2176G CPU and 64 GB RAM) resided in the mobility platform. The robot operating system (ROS) is used to fully integrate the entire software and facilitate the communication and control of different components.

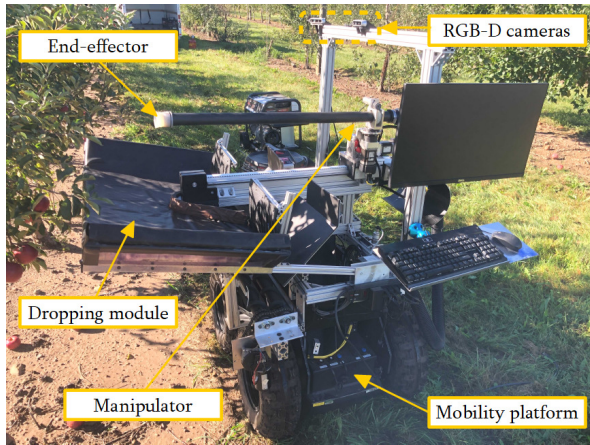


Fig. 1. The developed robotic apple harvesting prototype.

A. Hardware Design

For automated apple harvesting, the first and foremost task is orchard perception, which detects and localizes the fruits to guide robotic manipulations. Different from existing works (e.g., [7], [10]) that attach the camera to the manipulator or the end-effector, the RGB-D cameras are installed on the Segway mobility platform to provide a global view of the scene, facilitating the use of multiple manipulator arms planned in our future versions. Moreover, the multi-camera setup is introduced to provide multi-view sensing from different perspectives, which is intended to achieve enhanced perception accuracy and robustness through sensor

fusion to alleviate the impact of occlusions and challenging lighting conditions.

To efficiently approach the target fruits, a 3-DOF manipulator with simple and compact mechanical structure is designed and assembled. Specifically, the manipulator is comprised of one prismatic joint and two revolute joints. The two revolute joints are linked using an L -shaped aluminum plate, which creates a pan-and-tilt module. The prismatic joint is assembled as the base of the pan-and-tilt module to extend the depth of the manipulator's workspace. A hollow aluminum link is installed on the pan-and-tilt module to ensure that the end-effector can reach the apple locations, and it also acts as a vacuum tube for grasping fruits in the harvesting process. Instead of relying on a hybrid pneumatic/motor actuation mechanism in our previous design [9], all the joints of the current manipulator are driven by servo motors, which not only reduces actuation complexity but also facilitates integrated control scheme design.

In our system, a vacuum-based end-effector is designed to grasp and detach fruits. A soft silicone vacuum cup is attached to the front end of the aluminum tube. The vacuum cup with a special geometric configuration has shown satisfactory performance in conforming to various apple contours [13]. Meanwhile, the rear end of the aluminum tube is connected to an electric powered wet/dry vacuum via a flexible and expandable tube. The vacuum-based end-effector can reduce potential damage to fruits. Moreover, if the manipulator does not reach the apple accurately, the vacuum-based end-effector can tolerate some approaching inaccuracies since it can attract the fruit within a certain distance (about 1.5 cm in our current prototype) when adequate vacuum flow is provided.

For ease of collecting and transporting picked apples, a dropping module is assembled and affixed to the mobility platform. The base of the dropping module is a rectangular aluminum plate with a foam cushion covering. The manipulator can stop at any spots above the dropping module and then release the harvested fruit, thus reducing the harvesting cycle time. After the apple have fallen on the sloped surface of the dropping module, it rolls down to the rear end of the dropping module where a screw-driven conveyer is installed to transport the apple to a bin [14].

B. Software Design

The software suite is designed and integrated in the ROS framework. Different software components are primarily communicated via custom messages sent through ROS actions and services. Fig. 2 shows the main algorithm flow of the software system during an apple harvesting cycle. The algorithm structure mainly consists of three modules: perception, planning, and control. The logic flow of the apple harvesting cycle is detailed in the following.

At the beginning of each harvesting cycle, the RGB-D cameras are triggered to acquire images. With the obtained image information, the perception algorithm (Section III-A) is used to detect and localize the fruits within the system's workspace. A list of 3D apple locations are then

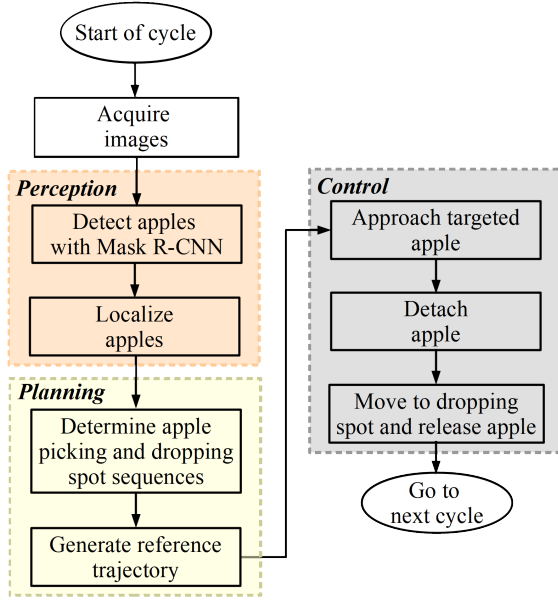


Fig. 2. Algorithm flowchart in an apple harvesting cycle.

generated and subsequently transformed into the 3D positions expressed in the coordinate frame of the manipulator. Based on the apple location list, the planning algorithm is utilized to optimize the apple picking sequence and its corresponding dropping spots (Section III-B). The detected apples will be chosen as the targeted fruits by following the planned picking sequence, and a reference trajectory will be generated to guide the motion of the manipulator. The target apple location and its corresponding reference trajectory are passed onto the control module, which then actuates the manipulator to follow the reference trajectory to reach the fruit. Once the fruit is successfully attached to the end-effector (detected by a pressure sensor mounted inside the tube), the rotation mechanism is triggered to rotate the whole aluminum tube by a certain angle to detach the apple (Section III-C). Finally, the manipulator returns to a dropping spot and releases the fruit. It is apparent that the software design of our robotic system requires multi-disciplinary advances to enable various synergistic functionalities and coordination for achieving reliable automated apple harvesting. The next section describes each of the software components in more details.

III. ALGORITHM DESIGN AND INTEGRATION

In this section, we describe our software components on perception, planning, and control in details.

A. Multi-view fusion for robust detection and localization

One of the key tasks in robotic apple harvesting is fruit detection and localization, where the former is to segment apples from the background areas whereas the latter subsequently calculates the 3D positions of the detected apples. In our preliminary work, a network with Mask R-CNN backbone and a suppression end was developed in [12] using a single RGB-D camera. In this new version, we extend the perception system to systematically fuse two RGB-D cameras to enhance the detection performance. This is motivated

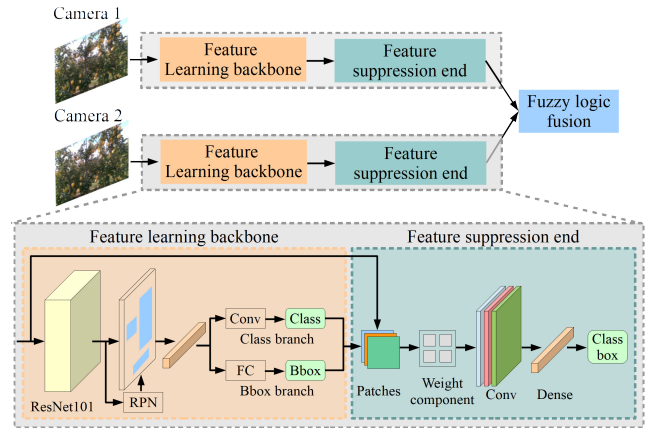


Fig. 3. Apple detection structure based on two-camera setup.

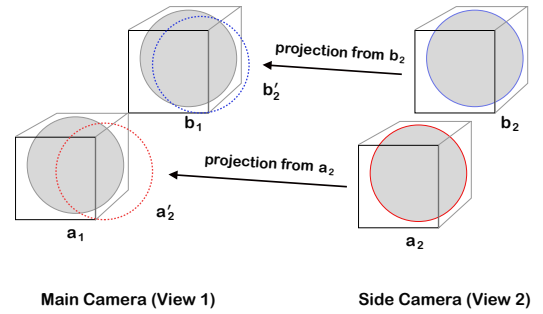


Fig. 4. Apple matching unit of the fusion scheme: assume we have two apples a and b which are identified as a_1, b_1 in the main camera (View 1) and are identified as a_2, b_2 in the side camera (View 2). Based on the extrinsic calibration of the two cameras, a_2, b_2 in the side camera are first transformed into a'_2, b'_2 in the main camera. Then we match a_1 with a'_2 and match b_1 with b'_2 based on the overlap.

to address the two major challenges in orchard perception as identified in our previous filed tests: leaf/branch occlusion and varying lighting conditions. Exploiting multiple cameras from different views can alleviate the impact of occlusion and challenging lighting conditions as the two cameras can provide complementary views for enhanced performance.

The network architecture of the proposed detection approach is shown in Fig. 3. The raw images captured by the two RGB-D cameras from different perspectives are fed into identical but separate deep learning network which consists of two components: a feature learning backbone and a feature suppression end. The feature learning backbone adopts the classical backbone designed in Mask R-CNN [15] to extract apple features and generate region proposals. Since the feature learning backbone might generate wrong inference features, the image patches inside the bounding boxes are then passed to a feature suppression end to remove some misclassified candidates. Once the images from the cameras are processed by the deep learning network, the bounding boxes of apple candidates are obtained. This suppression Mask R-CNN design has been reported in [12]. For the multi-view object detection, the key task is to associate the identical objects from two views. To merge the detection results from the two camera channels, we further design a fusion scheme

that consists of an apple matching unit and a fuzzy logic unit. In the matching unit, we match the bounding boxes from the two camera frames based on the extrinsic calibration of the two cameras. As showed in Fig. 4, the bounding boxes and apple positions detected from the side camera are transformed into the corresponding ones represented under the coordinate frame of the main camera. Based on the overlap relationship, we build associations for apples between two cameras. For the matched and unmatched bounding boxes, a fuzzy logic unit is utilized to combine the detection results to further enhance the accuracy of the labeled candidates. As shown in Fig. 5, we design expert rules for the fuzzy logic unit. A triangular fuzzy membership function [16] is applied to process the crisp input, and the results are distilled to determine the final detection confidence. Fig. 6 shows an example of the detection results. It is clear that apples with high occlusion can be identified more effectively by introducing the multi-view fusion detection mechanism. Moreover, F1-score is used to quantitatively evaluate the detection performance. Specifically, all detection outcomes are divided into four types: true positive (TP), false positive (FP), true negative (TN), and false negative (FN), based on the relation between the true class and predicted class. Then precision (P) and recall (R) are defined as follows:

$$P = \frac{TP}{TP + FP},$$

$$R = \frac{TP}{TP + FN}.$$

The F1-score is the harmonic mean of the precision and recall, which is defined as follows:

$$F1 = \frac{2P \cdot R}{P + R}.$$

The detection algorithm with two-camera setup achieves an F1-score of 93.92%, while the one with a single camera achieves an F1-score of 90.5%. The code on the network and fuzzy logic implementation is open sourced (<https://github.com/pengyuchu/DualCamFusion>).

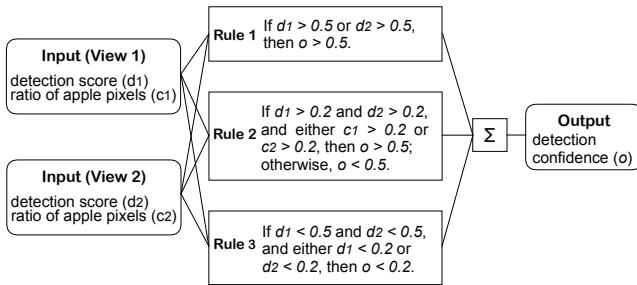


Fig. 5. Fuzzy logic unit: the inputs are detection score (d_1, d_2) and the ratio of apple pixels (c_1, c_2) from both view 1 and view 2 with a limited range of 0 – 1. We use fuzzy reasoning to evaluate all three rules in parallel and then the results of the rules are combined and distilled to the detection confidence o .

After detecting the fruits, apple localization is performed by employing the depth information provided by the RGB-D camera. More precisely, for each bounding box, the image



Fig. 6. A comparison example between multi-view fusion detection and single-view detection. The blue bounding boxes represent the targets detected with the single-view detection algorithm. The yellow bounding boxes denote additional detection by introducing multi-view fusion mechanism.

pixels of the detected apple are extracted to generate a range matrix by utilizing the disparity map. We then calculate the mean value of the range matrix and regard it as the apple's depth range. Combining the depth range with the center of the bounding box pixels, back-projection [17] is used to calculate the 3D position of the apple. This process is conducted for each of the bounding boxes to obtain positions of all detected apples in the image.

B. Unified picking/dropping and motion planning

In our system, there are two levels of planning tasks. At a high level, we need to plan the apple picking sequence and dropping spots based on the list of detected apple locations provided by the perception module. At a lower level, we need to generate reference trajectory for a selected target apple for the manipulation control.

The high-level harvesting sequence planning is necessary as it plays a crucial role in reducing the harvesting cycle time. Different from existing works [4], [18] that only focus on optimizing the fruit picking sequence, we take both the apple picking and dropping spot sequences into consideration. This flexibility is enabled by our dropping module design where the end effector does not need to return all the way to the home position to release the picked fruits. As shown in Fig. 1, the dropping module allows the end effector to release the detached apple in a large area, offering additional flexibility and optimization freedom for improved harvesting efficiency. More specifically, given N detected apples with $p_i \in \mathbb{R}^3$, $i = 1, \dots, N$, denoting their 3D positions expressed in the manipulator frame. The apple picking sequence and the dropping spot sequence are defined as follows:

- The apple picking sequence $S = \{s_1, \dots, s_N\}$ is a permutation of $\{1, \dots, N\}$, which determines the sequence of picking apples with the corresponding position sequence $\{p_{s_1}, \dots, p_{s_N}\}$ that the manipulator will follow to travel through. Each apple only will be visited once.

- The dropping spot sequence $S_d = \{\bar{p}_{s_1}, \dots, \bar{p}_{s_{N-1}}\}$ is a list of ordered 3D positions where the manipulator will stop by and release the harvested fruit. As discussed in Section II-A, the dropping module provides a specific domain (which is denoted by \mathbb{P}) for the manipulator to release the fruit, and hence the dropping spots \bar{p}_{s_i} should be generated from this domain, i.e., $\bar{p}_{s_i} \in \mathbb{P}$.

In the planning phase, we consider that the manipulator will start from its home position p_0 and approach the detected apples by following the sequence S defined above. For $i = 1, \dots, N-1$, once the apple located at p_{s_i} is harvested, the manipulator will move to the position \bar{p}_{s_i} to release the fruit and then heads to the next apple located at $p_{s_{i+1}}$. In particular, if the last apple in the picking sequence is harvested, the manipulator will return back to the home position p_0 for fruit release. According to the above description, it can be concluded that the manipulator's maneuver satisfies the following sequence:

$$p_0 \rightarrow p_{s_1} \rightarrow \bar{p}_{s_1} \rightarrow \dots \rightarrow p_{s_{N-1}} \rightarrow \bar{p}_{s_{N-1}} \rightarrow p_{s_N} \rightarrow p_0. \quad (1)$$

The planning objective is to determine the picking sequence S and its corresponding dropping spot sequence S_d by optimizing the travel cost along the maneuver sequence (1). We use Euclidean distance to define the travel cost, as follows:

$$g = \|p_{s_1} - p_0\| + \sum_{i=1}^{N-1} g_{s_i, s_{i+1}} + \|p_0 - p_{s_N}\|, \quad (2)$$

where $g_{s_i, s_{i+1}} = \|\bar{p}_{s_i} - p_{s_i}\| + \|p_{s_{i+1}} - \bar{p}_{s_i}\| \in \mathbb{R}$ ($i = 1, \dots, N-1$) is the travel cost between two adjoining apples in S . Given p_{s_i} and $p_{s_{i+1}}$, the optimal dropping spot $\bar{p}_{s_i}^*$ can be determined by solving the following problem:

$$\begin{aligned} \min_{\bar{p}_{s_i}} g_{s_i, s_{i+1}}(\bar{p}_{s_i}) &= \|\bar{p}_{s_i} - p_{s_i}\| + \|p_{s_{i+1}} - \bar{p}_{s_i}\|, \\ \text{s.t. } \bar{p}_{s_i} &\in \bar{\mathbb{P}}. \end{aligned} \quad (3)$$

Let $g_{s_i, s_{i+1}}^* = \|\bar{p}_{s_i}^* - p_{s_i}\| + \|p_{s_{i+1}} - \bar{p}_{s_i}^*\| \in \mathbb{R}$ be the minimal value of $g_{s_i, s_{i+1}}$. Then, the optimization problem over the picking sequence S is formulated as

$$\begin{aligned} \min_S g(S) &= \|p_{s_1} - p_0\| + \sum_{i=1}^{N-1} g_{s_i, s_{i+1}}^* + \|p_0 - p_{s_N}\|, \\ \text{s.t. } s_i &\in \{1, \dots, N\}, i = 1, \dots, N, \\ s_i &\neq s_j, \text{ for any } i \neq j \text{ and } i, j = 1, \dots, N. \end{aligned} \quad (4)$$

To determine the apple picking sequence and the dropping spot sequence. We first calculate the minimal travel cost between any two apple positions via (3). With the obtained minimal travel cost for any pair of two apples, the optimization problem (4) can be reformulated as a travel salesman problem (TSP). The nearest neighbor algorithm [19] is utilized to address the TSP, and then the apple picking sequence S and the dropping spot sequence S_d can be determined.

Based on the sequences S and S_d , the apple location p_{s_i} and the dropping spot \bar{p}_{s_i} will be assigned in turn as the targeted position $p_d = [x_d, y_d, z_d]^\top$ where the manipulator needs to reach. In our implementation, the planning

algorithm described above will be performed whenever the perception system sends an updated list of detected apple locations. To facilitate the manipulation control, given a targeted position p_d (e.g., top of the picking list), we use the quintic function [20] to generate a corresponding reference trajectory $p_r(t) = [x_r(t), y_r(t), z_r(t)]^\top$. This reference trajectory is a function of time with its terminus being the target position p_d . The introduction of the quintic function-based reference trajectory p_r brings the following advantages: First, the reference trajectory is continuously differentiable and its terminal velocity and acceleration are zero, which is conducive to ensuring that the end-effector approaches the desired position along a smooth path. Second, by adjusting function parameters, the velocity profile of the reference trajectory can be modified, and thus the end-effector can reach the desired position within a specific time interval.

C. Efficient nonlinear control for accurate reference tracking

Given a target apple and the generated reference trajectory using the planning algorithm discussed above, we next introduce the control algorithm that drives the manipulator to follow the reference trajectory. As shown in Fig. 2, one key requirement of the control module is to adjust the manipulator to approach the detected fruits or dropping spots with high accuracy and flexibility. To achieve this goal, a motion control strategy is developed by fully exploiting the mechanical structure of the developed 3-DOF manipulator.

The kinematic description of the 3-DOF manipulator is shown in Fig. 7. Denote $p = [x, y, z]^\top \in \mathbb{R}^3$ as the position of the end-effector. Based on the Denavit–Hartenberg convention [20] and the kinematical diagram presented in Fig. 7, the forward kinematics function of the manipulator can be derived, as follows:

$$\begin{aligned} x &= d_{x3} \cos(\theta) \cos(\varphi) + d_{x2} \cos(\theta) + d_{z2} \sin(\theta) + d_{x1} + D, \\ y &= d_{x3} \sin(\varphi) + d_{y2} + d_{y1}, \\ z &= -d_{x3} \sin(\theta) \cos(\varphi) - d_{x2} \sin(\theta) + d_{z2} \cos(\theta) + d_{z1}, \end{aligned} \quad (5)$$

where $d_{x1}, d_{x2}, d_{x3}, d_{y1}, d_{y2}, d_{z1}, d_{z2} \in \mathbb{R}$ are the link lengths, and $[\varphi, \theta, D]^\top \in \mathbb{R}^3$ are the joint variables.

As described in Section III-B, the planning module will provide the reference trajectory $p_r(t) = [x_r(t), y_r(t), z_r(t)]^\top$ for the targeted position p_d . The objective of the manipulation control is to regulate the end-effector to follow the reference trajectory p_r and finally approach the target position p_d . The revolute joint parameters

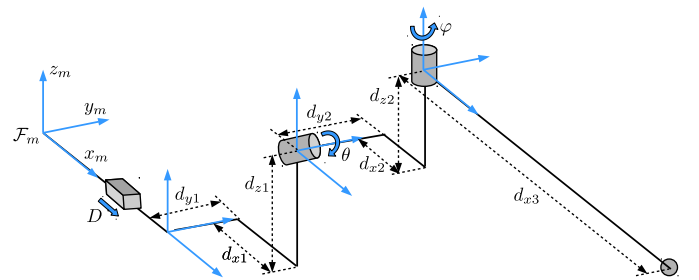


Fig. 7. Kinematical description of the 3-DOF manipulator.

φ , θ and prismatic joint parameter D are all driven by electrical motors, and the velocity-based control scheme is employed to generate explicit speed command to smoothly adjust the joints based on real-time position feedback. Specifically, based on (5), the time derivative of $[x, y, z]^\top$ can be calculated as

$$\begin{aligned}\dot{x} &= -d_{x3}(\sin(\theta)\cos(\varphi)\omega_\theta + \cos(\theta)\sin(\varphi)\omega_\varphi) \\ &\quad - d_{x2}\sin(\theta)\omega_\theta + d_{z2}\cos(\theta)\omega_\theta + v_D, \\ \dot{y} &= d_{x3}\cos(\varphi)\omega_\varphi, \\ \dot{z} &= -d_{x3}(\cos(\theta)\cos(\varphi)\omega_\theta - \sin(\theta)\sin(\varphi)\omega_\varphi) \\ &\quad - d_{x2}\cos(\theta)\omega_\theta - d_{z2}\sin(\theta)\omega_\theta,\end{aligned}\quad (6)$$

where ω_φ , $\omega_\theta \in \mathbb{R}$ are the angular velocity of the revolute joints φ and θ , respectively, and $v_D \in \mathbb{R}$ is the linear velocity of the prismatic joint D . Furthermore, the error signals $[e_x, e_y, e_z]^\top \in \mathbb{R}^3$ are constructed as

$$\begin{aligned}e_x &= x - x_r, \\ e_y &= y - y_r, \\ e_z &= z - z_r.\end{aligned}\quad (7)$$

Based on (6), (7), and by virtue of Lyapunov-based control techniques [21], the velocity controller is designed as

$$\begin{aligned}\omega_\varphi &= \frac{-k_y e_y + \dot{y}_r}{d_{x3} \cos(\varphi)}, \\ \omega_\theta &= \frac{k_z e_z + d_{x3} \sin(\theta) \sin(\varphi) \omega_\varphi - \dot{z}_r}{d_{x3} \cos(\theta) \cos(\varphi) + d_{x2} \cos(\theta) + d_{z2} \sin(\theta)}, \\ v_D &= -k_x e_x + d_{x3}(\sin(\theta)\cos(\varphi)\omega_\theta + \cos(\theta)\sin(\varphi)\omega_\varphi) \\ &\quad + d_{x2}\sin(\theta)\omega_\theta - d_{z2}\cos(\theta)\omega_\theta + \dot{x}_r,\end{aligned}\quad (8)$$

where $k_x, k_y, k_z \in \mathbb{R}^+$ are positive constant gains. The velocity controller (8) can ensure that the end-effector position tracks the reference trajectory $[x_r, y_r, z_r]^\top$ asymptotically, and rigorous stability analysis can be conducted based on the Lyapunov-based control techniques [21].

IV. EXPERIMENT AND RESULTS

In this section, both indoor and field experiments are presented to demonstrate the performance of the developed system. The indoor tests with artificial trees are focused on validating the planning and control schemes, and the integrated system is further evaluated in the orchard field.

A. Indoor test validation

Indoor tests were carried out in a simulated orchard environment, which consists of artificial trees hanging real apples, a flat-panel lighting system, and a Qualisys motion capture system.

We first validate the performance of the planning algorithm that is used to generate picking sequence and dropping spots based on the apple locations provided by the perception module. Three trials with different apple configurations are carried out for thorough validation. These three trials each have 5 apples, 7 apples, and 9 apples randomly hung on the artificial tree, and Fig. 8 depicts the scenario of Trial 1 for reference. The perception algorithm is used to detect

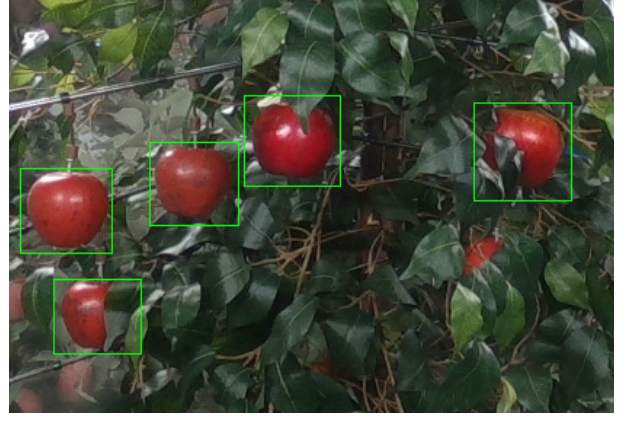


Fig. 8. The scenario of Trial 1 for planning evaluation, where green bounding boxes represent detected apples.

TABLE I
COMPARISON OF TRAVEL DISTANCE AND TRAVEL TIME BETWEEN THE PROPOSED PLANNING SCHEME AND NON-PLANNING CASE

Trial	Fruit number	Travel distance NP (P)	Travel time NP (P)
1	5	4.12 (3.18) [m]	12.55 (10.70) [s]
2	7	5.61 (4.28) [m]	17.11 (14.51) [s]
3	9	6.81 (4.86) [m]	21.07 (17.65) [s]

where NP = non-planning case and P = proposed planning scheme.

and localize these fruits, and then the planning module is triggered to determine the picking sequence and dropping spots. The travel cost defined in (2) is calculated with the obtained picking sequence and dropping spots. We also actuate the manipulator to reach the apple positions and the dropping spots by following the planning results, and the total travel time (i.e., the maneuver time of the manipulator) is recorded. Moreover, to better demonstrate the effectiveness of the planning algorithm, the non-planning case is introduced for comparison. In the non-planning case, we do not optimize the dropping spots and consider that once the manipulator reaches a detected apple, it will always move to the home position to release the fruit. The same testing scenarios are used to obtain the travel cost and travel time under the non-planning case. The results are summarized in Table I. It is clear that the proposed planning algorithm can significantly reduce the travel cost by optimizing the picking sequence and dropping spots. Furthermore, the travel time is an intuitive indicator to show the effect of the planning module in reducing the harvesting cycle time. Compared to the non-planning case, the proposed planning algorithm can efficiently reduce the travel time.

To thoroughly validate the control performance, the manipulator are driven to multiple target positions, and then the Qualisys motion capture system is utilized to evaluate the motion accuracy. Specifically, as discussed in Section III-C, the developed manipulator includes three joints, i.e., $[\varphi, \theta, D]^\top$. The desired joint values are selected from the following sets: $\varphi_d, \theta_d \in \{-20^\circ, -10^\circ, 10^\circ, 20^\circ\}$ and $D_d \in \{0.1\text{m}, 0.3\text{m}, 0.5\text{m}\}$, and the corresponding target positions can be computed based on (5). A total of 48

TABLE II

AVERAGE ABSOLUTE ERROR BETWEEN TARGET POSITIONS AND MANIPULATOR FINAL POSITIONS WITH THE DESIGNED CONTROLLER

	Nonlinear controller (8)
x -axis error [cm]	0.3905
y -axis error [cm]	0.3324
z -axis error [cm]	0.2742
Distance error [cm]	0.6566

target positions are generated, which are evenly distributed in the workspace of the manipulator. Furthermore, a spherical marker is attached to the end-effector, ensuring that the Qualisys motion capture system can measure the end-effector position precisely through marker identification. The manipulator is actuated from the home position to each of the target positions, and the final position of the end-effector is recorded. Based on 48 pairs of final position records and the corresponding target positions, the average errors along the x -axis, y -axis, z -axis and the average distance errors are calculated. The controller designed in (8) is tested, and the results are shown in Table II. It can be seen that the average distance error is less than 1 cm, indicating that the proposed control scheme can achieve satisfactory performance.

B. Field test and validation

To further evaluate the performance of the integrated prototype, field experiments are conducted in the Horticultural Teaching and Research Center of Michigan State University during the 2021 harvest season¹.

In the field test, the robotic apple harvesting system is run autonomously and continuously to harvest fruits within its workspace with fully integrated perception, planning, and control functionalities. On average, the duration for the manipulator to approach a target apple or move back to its corresponding dropping spot ranges between 0.75 s and 1.4 s. Detaching and releasing the fruit roughly take 1 s and 0.5 s, respectively. For the successfully harvested apples, the average cycle time is approximately 3.6 s, including software algorithm processing and hardware execution. Compared to our previous prototype [9] and the existing literature [4], [7], [8], [10] which took 7-10 seconds to harvest an apple, the current robotic apple harvesting prototype clearly has made significant advancement in terms of harvesting efficiency, thanks to the simple yet efficient mechanism as well as the integrated algorithm design.

However, there is still a considerable gap in achieving a satisfactory picking rate. In this field test, a total of 142 apples were attempted and 74 of them were picked successfully with the picking rate being 52.1%. We studied the failed cases and identified the following major causes. First, we found that the depth measurement in the RealSense RGB-D camera is susceptible to varying lighting conditions and cases when fruit is partially obscured by foliage or branches, and sometimes provides inaccurate depth information with an error as large as 10 cm. Second, unlike the

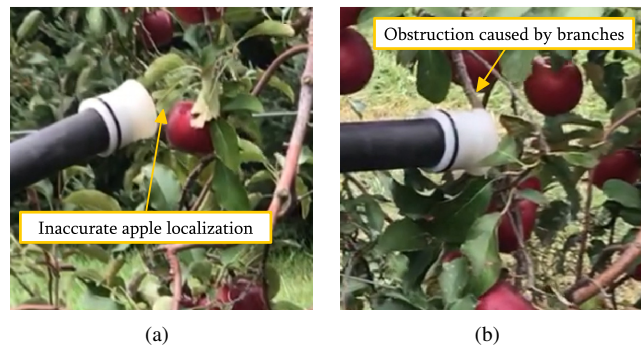


Fig. 9. Examples of failed apple harvesting: (a) due to inaccurate localization; and (b) due to obstruction caused by branches.

v-trellis structured orchard used in [4] where most of the apples are well exposed, the orchard where we conducted the experiment does not have a well-structured fruiting system and a high percentage of fruits are being occluded by leaves and branches, which create challenges for the robotic system to approach the target fruits. Third, there are also many occurrences during which the end-effector has failed to detach target fruits due to inadequate vacuum power. Fig. 9 illustrates two failed harvesting scenarios.

These findings provide useful insights towards further improvement of our system. The two-camera setup enhances fruit detection and localization in the indoor environment but appears to have limited contribution to improving the accuracy of fruit localization in the field. Fusing additional sensing modalities such as LiDAR could be a good way to achieve robust fruit localization and is currently under investigation. We also need to design an object segmentation algorithm to identify the obstacles (trunks, branches) and develop a path planning scheme to avoid obstruction. Finally, the vacuum system and fruit detachment strategy should also be further improved for reliable fruit picking.

V. CONCLUSION

The algorithm design and integration for a newly-developed robotic apple harvesting prototype was introduced in the paper. The algorithm component is comprised of three core modules: perception, planning, and control. Indoor and field experiments demonstrated that the developed algorithm component can synergistically work with the hardware component to achieve the primary apple harvesting functionalities, offering a promising picking cycle time of 3.6 seconds. Guided by lessons learned from these experiments, future work will include improving fruit localization accuracy and robustness, developing object segmentation algorithms for obstacle detection, and designing optimal path planning scheme for obstacle avoidance.

REFERENCES

- [1] R. K. Gallardo and S. P. Galinato, *2012 Cost Estimates of Establishing, Producing, and Packing Red Delicious Apples in Washington*. Washington State University Extension, 2012.
- [2] C. Lehnert, A. English, C. McCool, A. W. Tow, and T. Perez, "Autonomous sweet pepper harvesting for protected cropping systems," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 872–879, 2017.

¹A video of the robotic apple harvester demonstrating the field tests is available at https://www.youtube.com/watch?v=_6-5qbZp1Zo

- [3] Y. Xiong, Y. Ge, L. Grimstad, and P. J. From, "An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation," *J. Field Robot.*, vol. 37, no. 2, pp. 202–224, 2020.
- [4] A. Silwal, J. R. Davidson, M. Karkee, C. Mo, Q. Zhang, and K. Lewis, "Design, integration, and field evaluation of a robotic apple harvester," *J. Field Robot.*, vol. 34, no. 6, pp. 1140–1159, 2017.
- [5] H. Williams, C. Ting, M. Nejati, M. H. Jones, N. Penhall, J. Lim, M. Seabright, J. Bell, H. S. Ahn, A. Scarfe *et al.*, "Improvements to and large-scale evaluation of a robotic kiwifruit harvester," *J. Field Robot.*, vol. 37, no. 2, pp. 187–201, 2020.
- [6] X. Zhang, L. He, M. Karkee, M. D. Whiting, and Q. Zhang, "Field evaluation of targeted shake-and-catch harvesting technologies for fresh market apple," *Trans. ASABE*, vol. 63, no. 6, pp. 1759–1771, 2020.
- [7] J. Baeten, K. Donné, S. Boedrij, W. Beckers, and E. Claesen, "Autonomous fruit picking machine: A robotic apple harvester," in *Field and service robotics*. Springer, 2008, pp. 531–539.
- [8] C. J. Hohimer, H. Wang, S. Bhusal, J. Miller, C. Mo, and M. Karkee, "Design and field evaluation of a robotic apple harvesting system with a 3D-printed soft-robotic end-effector," *Trans. ASABE*, vol. 62, no. 2, pp. 405–414, 2019.
- [9] K. Zhang, K. Lammers, P. Chu, Z. Li, and R. Lu, "System design and control of an apple harvesting robot," *Mechatronics*, vol. 79, p. 102644, 2021.
- [10] D. M. Bulanon, C. Burr, M. DeVlieg, T. Braddock, and B. Allen, "Development of a visual servo system for robotic fruit harvesting," *AgriEngineering*, vol. 3, no. 4, pp. 840–852, 2021.
- [11] R. Lu, Z. Zhang, and A. K. Pothula, "Innovative technology for apple harvest and in-field sorting," *Fruit Q. J.*, vol. 25, no. 2, pp. 11–14, 2017.
- [12] P. Chu, Z. Li, K. Lammers, R. Lu, and X. Liu, "DeepApple: Deep learning-based apple detection using a suppression Mask R-CNN," *Pattern Recognit. Lett.*, vol. 147, pp. 206–211, 2021.
- [13] R. Lu, N. Dickinson, K. Lammers, K. Zhang, P. Chu, and Z. Li, "Design and evaluation of end effectors for a vacuum-based robotic apple harvester," *Trans. ASABE*, in press, 2022.
- [14] Z. Zhang, Y. Lu, and R. Lu, "Development and evaluation of an apple infield grading and sorting system," *Postharvest Biol. Technol.*, vol. 180, p. 111588, 2021.
- [15] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 2961–2969.
- [16] W. Pedrycz, "Why triangular membership functions?" *Fuzzy Sets Syst.*, vol. 64, no. 1, pp. 21–30, 1994.
- [17] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- [18] Y. Edan, T. Flash, U. M. Peiper, I. Shmulevich, and Y. Sarig, "Near-minimum-time task planning for fruit-picking robots," *IEEE Trans. Robot. Autom.*, vol. 7, no. 1, pp. 48–56, 1991.
- [19] G. Gutin and A. P. Punnen, *The Traveling Salesman Problem and its Variations*. Springer Science & Business Media, 2006.
- [20] B. Siciliano, L. Sciavicco, L. Villani, and G. Oriolo, *Robotics: Modelling, Planning and Control*. London: Springer-Verlag, 2010.
- [21] H. K. Khalil, *Nonlinear Systems*. Upper Saddle River, NJ: Prentice-Hall, 2002.