

# Robot Manipulation of Tomato Fruits using a Commercial Soft Gripper

Riccardo Monica, Dario Lodi Rizzini  
Department of Engineering and Architecture and  
Center for Energy and Environment (CIDEA)  
University of Parma  
Parma, Italy  
{riccardo.monica,dario.lodirizzini}@unipr.it

Stefano Caselli  
Department of Veterinary Science and  
Center for Energy and Environment (CIDEA)  
University of Parma  
Parma, Italy  
stefano.caselli@unipr.it

**Abstract**—This paper presents a robot system for manipulation and picking of tomato fruits guided by computer vision. A CNN algorithm trained on a custom image dataset is used for fruit detection and the depth camera enables position estimation. The robot manipulator is equipped with a commercial soft gripper for picking fragile objects. Three planning procedures have been proposed to successfully reach and grasp the tomatoes. Experiments on simulated crops have compared the effectiveness of the proposed procedures.

**Index Terms**—autonomous robot manipulation, soft object grasping

## I. INTRODUCTION

In recent years, automation and robotics have been increasingly applied to agriculture to increase precision and efficiency in crop production, to achieve sustainability goals, and to reduce manual labor following similar trends in industrial automation. Distributed sensor networks for monitoring [1], variable rate irrigation [2], robot phenotyping [3] are just few examples of applications. Crop collection for harvesting, pruning or sampling is technologically a rather challenging task. Retail sales of fruits and vegetables like tomatoes, strawberries and apples require delivery of intact products. Thus, automatized fruit picking cannot be performed according to traditional mass harvesting techniques, but requires careful and non-destructive manipulation. Cultivation is usually arranged in ordered orchards in the field or in rows inside greenhouses. However, the unpredictable distribution of leaves, branches and canopies, the fragility and irregular shape of fruits, the variable weather and light conditions make such task challenging. Successful picking strongly relies on robust perception and effective motion planning.

Several perception systems have been developed with the purpose of picking fruits often based on depth or stereo cameras for detection and pose estimation [4]. Convolutional neural networks (CNN) have become the standard algorithm for generic object detection, since it is easily adaptable to specific applications. YOLO [5], [6], Faster R-CNN [7] and Mask R-CNN [8] are among the most common algorithms used in agriculture. These networks are available with generic training and can be further trained to recognize novel object categories on specific datasets. The position or the pose of

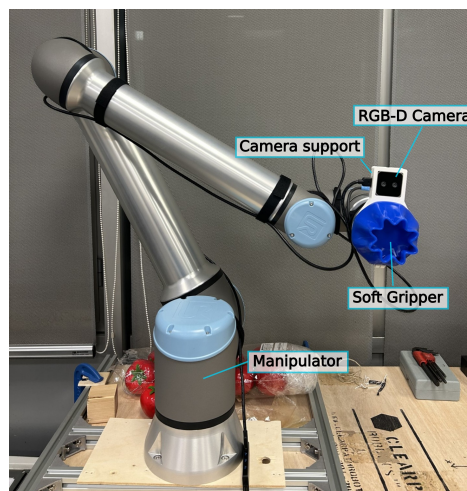


Fig. 1. The components of the proposed robot manipulation system: the manipulator UR10e, the soft gripper OnRobot SG-a-H, and the eye-in-hand camera Intel Realsense D405.

the fruits is estimated through processing of the point cloud corresponding to the detected region [9].

Several end-effectors have been designed and used for fruit grasping [10]. The large majority of solutions is represented by finger-based grippers or pneumatic suction cups. Jun *et al* [11] present a tomato harvesting robot using a YOLO-based detector and equipped with a custom suction cup end-effector. This end-effector is completed with a scissor-like module for separating the fruit from the stem. While this device has proven effective, it is rather unsuitable to be mounted on a mobile robot with limited space and autonomy. A similar scissor-like gripper is also used by the mobile manipulation system presented in [12] designed to collect leaf samples directly from crops. There are also examples of soft finger-based grippers like [13], but are used for fruits like apples which are less fragile than tomatoes. The majority of solutions rely on highly customized end-effectors.

In this paper, we illustrate a robot system for manipulation and picking of tomato fruits. The fruits are detected by an eye-in-hand RGB-D camera using Faster R-CNN algorithm. The algorithm is trained on a dataset of tomato pictures acquired

in fields. The depth camera has a limited optimal range for accurate position estimation of the target. An important distinguishing feature of the proposed manipulation is the adoption of a commercial soft gripper instead of a customized one. The OnRobot SG-a-H envelopes the target fruit through a silicon cup, whose opening diameter is controlled. The execution of the task requires careful planning to cope with the range limitation of the sensor and to reach the target through approaching direction suitable for picking. Three manipulation procedures have been devised and tested. They differ in the adoption of multiple observations of the target fruit and in the approaching direction either in the line-of-sight of the camera or from the bottom. Experiments have been carried out in a laboratory setup simulating tomato crops. Target re-observation and reaching procedure are both important for successful execution of the task.

## II. SETUP OF ROBOT MANIPULATION SYSTEM

The robot manipulation system used in this work consists of a collaborative robot manipulator, a soft gripper and a depth camera. Figure 1 illustrates all the components. The collaborative manipulator is a *Universal Robot UR10e*. Although rather voluminous to be mounted on standard mobile robots, this specific model has been selected to guarantee reasonable reachability and collection of fruit samples.

The soft gripper *OnRobot SG-a-H* is a commercial soft gripper designed for fragile, delicate and irregularly shaped objects like fruits and vegetables. The far end of the tool is a silicon-molded cup, whose opening diameter is controlled by a motor. The diameter is in the range between 17 mm and 74 mm so that its opening is sufficiently large for tomatoes. This soft gripper is designed to grasp objects by lightly pushing them against a surface, e.g. against a conveyor belt, and then shrinking the cup. Direct crop harvesting cannot rely on such assumptions. Branches and leaves have irregular structure and the fruit is connected to the crop by a stalk. Our challenge is to perform this task using a commercial gripper instead of a custom one.

The system is equipped with the depth camera *Intel RealSense D405*. The sensor is mounted just above the soft gripper very close to the silicon cup, since this stereo camera is designed for closer inspection. Indeed, the recommended operation range of the camera is between 70 mm and 500 mm, even though it can operate up to 1 m. The depth error for distant points is greater than 1.4%. The sensor is used for detection and pose estimation of tomato fruits in proximity of our manipulation system.

In the final application, the presented robot manipulation system will be mounted on a mobile robot equipped with additional sensors like LIDARs and depth cameras with complementary field-of-view and range.

## III. TOMATO FRUIT PERCEPTION

Perception enables manipulation of tomato fruits by detecting their presence in the scene and by estimating their location. Agriculture fields are rather challenging environments due

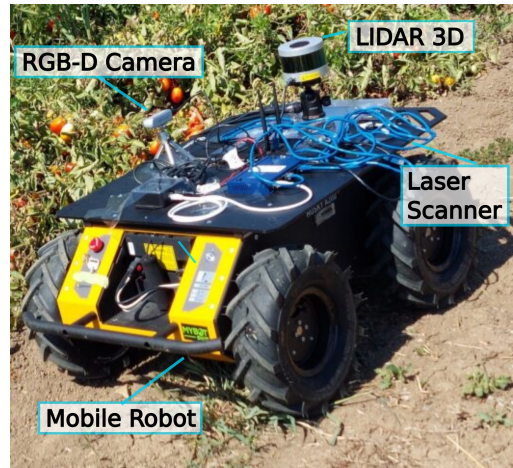


Fig. 2. The mobile robot used in acquisition field tomato dataset used to train the fruit detector.

to variable and non-uniform light, unstructured crops with unequal branches and canopy, occlusion, and irregular shape, color and distribution of fruits. Depending on the ripeness tomato color ranges from green to red. Fruits often appear in clusters springing from the same sprout and are close to each other. The task of locating the tomato fruits in proximity of the proposed manipulation system is based on the RGB-D camera, which provides both color and depth images. The RGB image is suitable for detection of target objects whereas the depth image enables evaluation of their position. These two subtasks are addressed separately as described in the following.

### A. Tomato Detection Algorithm

Tomato detection is the task of finding the region of interest (ROI) corresponding to each fruit instance in the image. As briefly illustrated in section II, the final robot system is equipped with multiple cameras observing tomato crops at different ranges and conditions. For example, a medium range RGB-D camera can be used to locate the spots with tomato fruits whereas the eye-in-hand camera enables close observation required for accurate location and manipulation. However, it would be convenient to develop a unique detection algorithm to handle both the scenarios.

Deep learning enables robust and adaptable identification of general objects and, specifically, of fruits and crop parts in agriculture. Algorithms like Mask R-CNN enable pixel-wise recognition of object shapes and could be used for recognition of fruit parts, but they also entail burdensome annotation and datasets with specialized observation condition. Simple bounding boxes allow sufficiently accurate estimation of target object position.

Faster R-CNN is adopted as a trade-off among detection accuracy, execution time and deployment simplicity. The network is initially pre-trained with the weights *COCO\_V1*<sup>1</sup> and then configured to classify the two classes *tomato* and

<sup>1</sup>[https://download.pytorch.org/models/fasterrcnn\\_resnet50\\_fpn\\_coco-258fb6c6.pth](https://download.pytorch.org/models/fasterrcnn_resnet50_fpn_coco-258fb6c6.pth)

*background.* We collected a novel datasets in the field to fine-tune the last layers of pre-trained network while keeping the weights of the first layers.

The dataset has been acquired in a field of growing tomatoes for sauce production (so called “tomato for industry”) using a mobile robot Clearpath Husky equipped with a *Intel Realsense D435* camera and other range sensors displayed in Figure 2. The camera is on the left side of the robot directed downward in the direction of the tomato crops. Two sessions of image acquisition took place in *Stuard Experimental Farm* (latitude 44.808813, longitude 10.273078) respectively on July 26th and on September 5th 2023. The output of first session consists of images of green or yellow colored tomatoes surrounded by copious leaves and branches. The second one comprises images of ripe tomatoes and crops at the end of the season and very close to harvesting and a single image comprises multiple orchards. Figure 3 shows two examples of both unripe and ripe tomatoes. A training set of 170 images with resolution  $640 \times 460$  has been used. The large number of fruits in each frame and occlusion make annotation difficult for the operator. The precision and recall of the algorithm measured on a test set of 20 images are respectively 96% and 74%. The limited value of recall is at least partially motivated by labeling errors.

The proposed detection algorithm allows satisfactory detection of foreground fruits in images encompassing a large portion of the field. In the manipulation and grasping task tomatoes are observed using an eye-in-hand camera at a shorter range as illustrated in Section II. Performance of Faster CNN is even better in this context, since closer tomatoes occupy larger portions of the image. This issue is discussed in section V.

### B. Position Estimation

The position of the detected fruits is evaluated using the depth image or the point cloud corresponding to their ROIs. RGB-D cameras provide optical and depth images with direct correspondence between their pixels. These data are used to detect the location of the tomato with respect to the camera frame. Some of the ranges in each bounding box are either invalid or not part of the fruit, but the outlier measurements can be easily filtered out through consensus criterion.

Tomato fruits are characterized by strong central symmetry and by lack of strong features enabling assessment of their orientation. In principle, the stem and the top of tomato could be recognized in close-range images through further image analysis. However, evaluation of a complete reference frame is unstable and too dependent on occlusion and observation conditions. The implemented algorithm computes the position of the center of tomato fruit as the mean value of the points in the bounding box corrected through the approximate radius. Least-square fitting of the measured points into a sphere model has been tried to improve the estimation accuracy, but at long range the outcome is rather sensitive to measurement errors. The computation of the center position of each tomato has proven sufficiently accurate for manipulation and grasp. Figure 4 illustrates an example of detection and position estimation tomato fruit in the planner representation.

## IV. GRASP AND MANIPULATION PLANNING

The task of planning fruit manipulation and grasp is addressed through several software components integrated as nodes through Robot Operating System (ROS) framework. The main component of the system is the *task coordinator* node organizing the data flow from and to the other nodes. It receives a list with the estimated positions of the tomatoes detected in the scene by the *tomato detection* node and selects the target object to be manipulated. Such selection occurs once the object position has been observed multiple times with stable position and the confidence value is greater than a given threshold. Moreover, the task coordinator computes the configuration of the end-effector required either to re-observe the target fruit from a better viewpoint or to grasp it. Some manipulation strategies illustrated in the following require a second observation. The *MoveIt!* planner is used to compute a collision-free trajectory reaching the desired configuration of the manipulator. Figures 4 and 5 present respectively an example of the camera viewpoint and the screenshot of the planner with the fixed obstacles and the point cloud. The task coordinator also sends motion commands to move the robot manipulator as well as to open and close the soft gripper using the *ur\_ros\_rtde* driver [14].

### A. Manipulation Procedures

The accuracy of the target object position depends also on its distance from the eye-in-hand camera. Without prior information the first observation of the scene is performed with the robot in an initial *home configuration*. Tomato detection and position estimation rely on the RGB images and range measurements acquired there. The fruit to be picked is selected according to the confidence value returned by the CNN detector, its distance from the sensor and the stability of its estimated position. The initial viewpoint is chosen to encompass a large portion of crops and is often too distant from the chosen target. Indeed, the target object is often out of the best operational range of the eye-in-hand depth camera. One possible way to better assess its position is to repeat observation from a closer viewpoint computed according to such initial guess.

Another issue is grasping the fruit using the commercial soft gripper presented in Section II. Special grippers designed for fruits often rely on artificial digits closing on contact points. In our system, in order to achieve a stable grasp the object must be fully enveloped by the silicon-molded cup. Line-of-sight (LoS) side approach to the fruit has the advantage of following the line-of-sight of the camera, which is free from leaves and branches of crops. However, the tomato is easily pushed by the approaching end-effector due to the fruit stem or to slight position estimation inaccuracy and may escape the cup. If approached from bottom, the object tends to fall in the gripper cup.

Thus, we propose three different procedures for observation and manipulation of tomato fruits.

- 1) *Line-of-sight single-view approach* (L1VA). The first procedure is direct reaching of fruit after observing



Fig. 3. Examples of images from the tomato datasets of unripe (left) and ripe (right) fruits. The outcome of the trained Faster CNN are shown with true positives (green), false negatives (blue) and false positives (red).

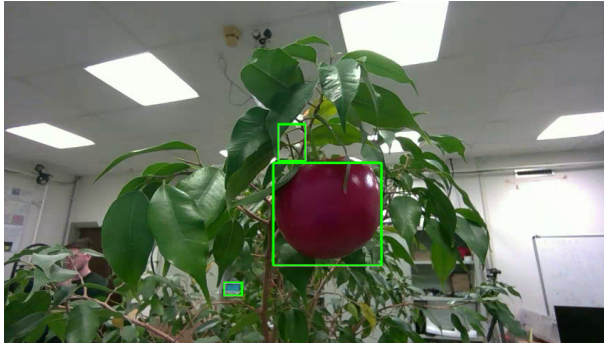


Fig. 4. A screenshot displaying the image observed by the eye-in-hand camera with the bounding boxes of candidate target objects (the image also displays ROIs with low confidence rate).

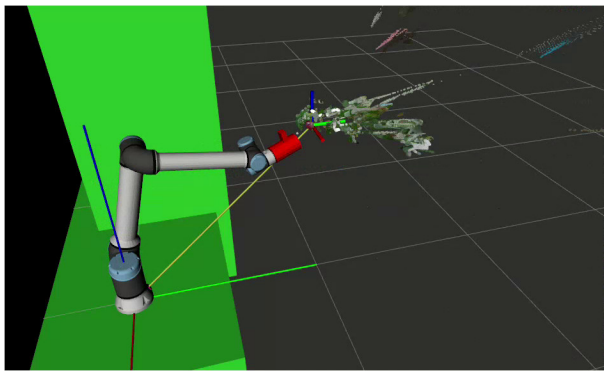


Fig. 5. A screenshot displaying the view of the planner *MoveIt!* planner with the loaded scene, the point cloud with an instance of tomato fruit and the detection window.

the target from the home configuration. The approach direction follows the LoS from the eye-in-hand camera to the fruit.

- 2) *Line-of-sight two-views approach* (L2VA). The target fruit is observed from the home configuration like in L1VA, but the initial assessment of its position is used

to compute a second closer viewpoint.

- 3) *Bottom two-views approach* (B2VA). This procedure performs two times observation like L2VA, but this time the second viewpoint is from the bottom of the fruit. The target is also approached from the bottom for grasping.

In each viewpoint, the position of a target is considered acquired after 6 measurements are provided by the tomato detector. The repeated observation enables filtering of false targets. The procedures L2VA and B2VA are illustrated in Figure 6.

### B. Observation and Grasp Poses

The manipulation procedures illustrated before require to compute and move the gripper or camera frames in proper poses. Let  $\{B\}$ ,  $\{W\}$ ,  $\{G\}$  and  $\{C\}$  be respectively the robot base frame, the robot wrist frame, the soft gripper frame and the camera frame. The relative poses of the gripper  ${}^W_C \mathbf{T}$  and the camera  ${}^W_C \mathbf{T}$  with respect to the wrist are known from the calibration. The position of the target object with respect to the camera  ${}^C \mathbf{t}_T$  are estimated by the detection algorithm. Since  $\{W\}$ ,  $\{G\}$  and  $\{C\}$  changes their pose with respect to the base  $\{B\}$  during the execution of the task, an integer index subscript  $i = 0, 1, 2$  is used to label the frames at the different phases of the task: 0 for the initial frame, 1 for the second observation, 2 for the grasp.

The initial observation returns the position of the target with respect to the camera  ${}^{C_0} \mathbf{t}_T$ , which can be transformed in base frame as  ${}^B \mathbf{t}_T = {}^B_{C_0} \mathbf{T} {}^{C_0} \mathbf{t}_T$ . The two-views approaches require to compute the new pose of the camera  $\{C_1\}$  that in turn defines the transformation to  $\{W_1\}$  and  $\{G_1\}$ .

Observation and grasp poses are defined using two primitives,  $\text{dir}(\mathbf{a}, \mathbf{b})$  and  $\text{from2vec}(\mathbf{a}, \mathbf{b})$ . Given two position vectors  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ , the primitive  $\text{dir}(\mathbf{a}, \mathbf{b})$  returns the unit direction from position  $\mathbf{a}$  to  $\mathbf{b}$  as

$$\text{dir}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{b} - \mathbf{a}}{\|\mathbf{b} - \mathbf{a}\|} \quad (1)$$

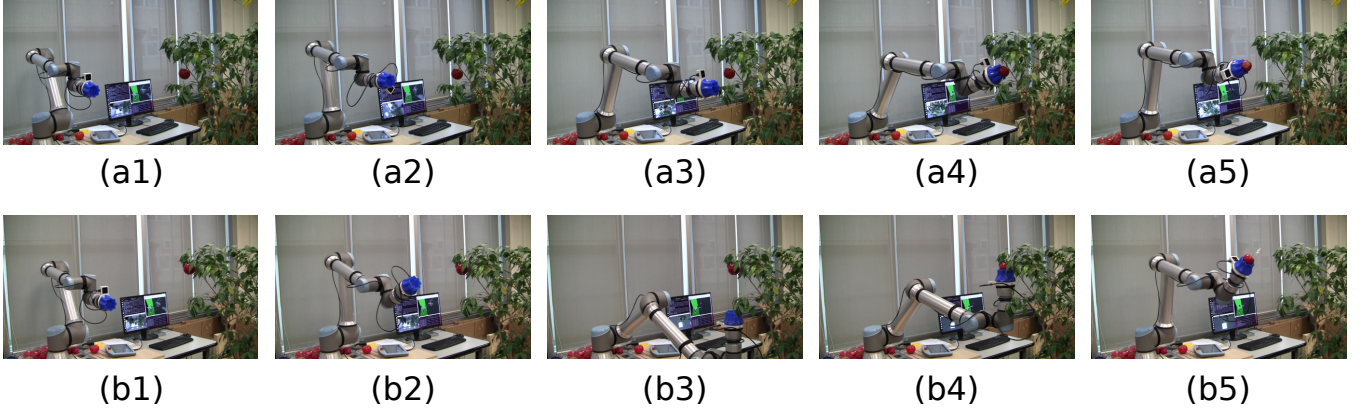


Fig. 6. Examples of the double-viewpoint observation and manipulation of tomato fruits according to procedures (a) L2VA and (b) B2VA. Some snapshots illustrating the phases of the two procedures: (1) observation of target from home configuration, (2) motion to the planned viewpoint, (3) second observation of target from the second viewpoint, (4) grasp and (5) collection of the target object.



Fig. 7. The four scenarios labeled *A*, *B*, *C* and *D* used to assess the position accuracy obtained with the eye-in-hand RGB-D camera.

Given two unit vectors  $\|\mathbf{a}\| = \|\mathbf{b}\| = 1$ , primitive  $\text{from2vec}(\mathbf{a}, \mathbf{b})$  returns a rotation matrix with axis  $\mathbf{a} \times \mathbf{b}$  (normalized) and angle  $\cos^{-1}(\mathbf{a} \cdot \mathbf{b})$ . Using these primitives, computation of the new camera pose  $\{C_1\}$  is defined differently for L2VA and B2VA:

- The L2VA computes the viewpoint pose  ${}^B_{C_1}\mathbf{T}$  such that its origin is on the LoS between the initial camera  $C_0$  position and the target  $T$  and the optical axis  ${}^B\hat{\mathbf{z}}_{C_1}$  is toward the target. Given the LoS direction  ${}^B\mathbf{v}_{los} = \text{dir}({}^B\mathbf{t}_{C_0}, {}^B\mathbf{t}_T)$ , the frame origin and orientation are

$${}^B\mathbf{t}_{C_1} = {}^B\mathbf{t}_T - d_{obs} {}^B\mathbf{v}_{los} \quad (2)$$

$${}^B_{C_1}\mathbf{R} = {}^B_{C_0}\mathbf{R} \text{from2vec}({}^B\hat{\mathbf{z}}_{C_0}, {}^B\mathbf{v}_{los}) \quad (3)$$

The parameter  $d_{obs}$  is the observation distance from the target. The homogeneous transformation matrix  ${}^B_{C_1}\mathbf{T}$  is obtained by composing translation  ${}^B\mathbf{t}_{C_1}$  and rotation  ${}^B_{C_1}\mathbf{R}$ .

- The B2VA computes the pose  ${}^B_{C_1}\mathbf{T}$  such that in this second viewpoint the optical axis direction is oriented along a given bottom direction  ${}^B\mathbf{v}_{bot}$ , e.g.  ${}^B\mathbf{v}_{bot} = [0, 0, 1]^\top$ . Hence, the position and orientation of the new frame is

$${}^B\mathbf{t}_{C_1} = {}^B\mathbf{t}_T - d_{obs} {}^B\mathbf{v}_{bot} \quad (4)$$

$${}^B_{C_1}\mathbf{R} = {}^B_{C_0}\mathbf{R} \text{from2vec}({}^B\hat{\mathbf{z}}_{C_0}, {}^B\mathbf{v}_{bot}) \quad (5)$$

The planner requires the manipulator wrist  ${}^B_{W_1}\mathbf{T}$  that is straightforwardly obtained from the camera transformation  ${}^B_{C_1}\mathbf{T}$ .

TABLE I  
POSITION ERROR

scene	object num.	position error [mm]	standard deviation [mm]
A	1	16.3	0.6
B	2	30.7	1.0
C	3	40.6	2.4
D	4	28.2	2.1

The grasping pose  ${}^B_{G_2}$  is computed in a similar way. The main difference is that the pose is referred to the gripper frame  $\{G_2\}$ . The position and orientation of the grasp pose are

$${}^B\mathbf{t}_{G_2} = {}^B\mathbf{t}_T - d_{grasp} \mathbf{v}_{grasp} \quad (6)$$

$${}^B_{G_2}\mathbf{R} = {}^B_{C_1}\mathbf{R} \text{from2vec}({}^B\hat{\mathbf{z}}_{C_1}, {}^B\mathbf{v}_{grasp}) \quad (7)$$

The vector  $\mathbf{v}_{grasp}$  is equal to  $\text{dir}({}^B\mathbf{t}_{C_1}, {}^B\mathbf{t}_T)$  in the case of L2VA and to  ${}^B\mathbf{v}_{bot}$  in the case of B2VA. The parameter  $d_{grasp}$  is a tolerance distance between the target and the palm of the gripper and is set to a value close to zero.

## V. EXPERIMENTS

The proposed manipulation and grasp system has been tested in the laboratory setup illustrated in section II. Artificial tomato fruits made of polystyrene material have been used to simulate the fruits due to convenience and limited availability of vegetables when experiments have been executed. These objects have contact properties rather similar to real tomatoes since the material is soft, fragile and slippery. They have been fixed to the plant through a yielding attachment (a piece of

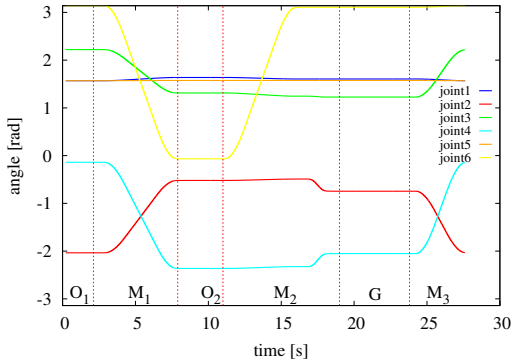


Fig. 8. Evolution of manipulator joint values during a trial of B2VA manipulation. The vertical dashed lines separates the different phases: first and second observations  $O_1$  and  $O_2$ , grasping  $G$  and the robot motions  $M_i$  between them.

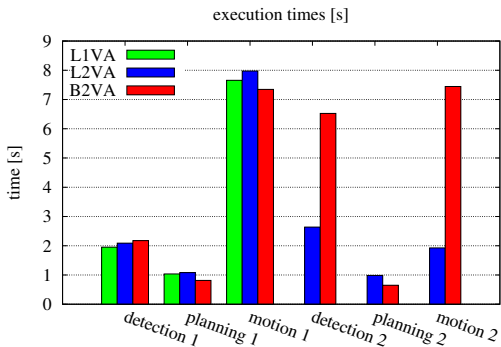


Fig. 9. The average times for detection, planning and robot motion in the first and second repetition using the procedures L1VA, L2VA and B2VA.

adhesive tape). The simulated fruits, henceafter referred simply as fruits, have been fixed to a plant with leaves, branches and vegetables. Figures 6 and 7 illustrate the scenario where experiments have been executed.

#### A. Position Assessment Experiments

Experiments have been designed to assess the accuracy of the algorithm that evaluates the position of the fruits using the eye-in-hand camera. The same Faster R-CNN algorithm trained using the dataset collected in the field is used also in the laboratory setup without additional training. Detection performance has qualitatively proven rather accurate in the novel setup. Only few occasional false positive are sometimes observed for distant and small sized objects in the background, but they can be filtered out easily.

The four scenes shown in Figure 7 have been staged. The scenes are labeled with letters from A to D and have a different number of tomatoes from 1 to 4 distributed on the plant. Each scene is observed from three different viewpoints. The offset between the center of the objects and robot flange has been manually removed.

The average position errors and the corresponding standard deviations for each scene are reported in Table I. The error

TABLE II  
NUMBER OF TRIALS (NUM) AND NUMBER OF SUCCESSES (HIT)

procedure	center		lateral	
	num	hit	num	hit
L1VA	12	10	12	0
L2VA	12	12	12	12
B2VA	12	12	12	12

values are between 16 mm and 40 mm. The larger errors are obtained in the scenes where some fruits are distant from the camera and partially occluded. Some of the bounding boxes, e.g. in Figure 7 A and C, only partially overlap with the image. The standard deviation is rather small and comparable with the accuracy of the depth camera. Thus, the significant position error is possibly the outcome of ground truth inaccuracy or partial observation. The better performance for closer objects suggests that close observation reduces inaccuracies.

#### B. Grasp Experiments

A second set of experiments has been designed to evaluate the effectiveness of the three procedures L1VA, L2VA and B2VA described in section IV-A. The first procedure reaches the object from LoS side after a single observation, whereas the other ones require another observation and approach the object respectively from the LoS or from the bottom. For each procedure, 24 trials have been performed: 12 with the target object in central position and 12 with target object located on the left or right of the eye-in-hand camera in home configuration. The distance between the camera and the fruit varies from 35 cm to 60 cm. Figure 8 shows an example of joint values over the time when executing a two-view manipulation.

Table II presents the number of successful grasping trials. With the single view procedure L1VA the manipulator successfully picks tomatoes only when in central position and always fail otherwise. Conversely, the second observation performed in L2VA and B2VA significantly increases the success rate.

The execution times of detection, trajectory planning and robot motion are shown in Figure 9. The two-view procedures L2VA and B2VA require a repetition of the three phases after the second observation. Their execution time is on average longer than the single view procedure L1VA. In particular, the bottom approach strategy requires longer time for detection due to failures occurring when the camera is oriented toward the light source. Planning computation does not start until the target fruit is consecutively observed multiple times (fixed to 6).

## VI. CONCLUSION

This paper has presented a robot system for manipulation and picking of tomato fruits. Fruit perception achieves precise detection in agriculture settings with multiple instances as well as in manipulation experiments whereas position estimation is sufficiently accurate for picking tasks. Three planning procedures have been proposed to overcome the limitations

of the range camera and to effectively pick the fragile target objects using a commercial soft gripper. A second close-up observation has proven essential for successful execution of the task. The approach direction toward the line-of-sight guarantee a path free from occlusion, but it also make object envelopment more difficult. Lifting the object from the bottom has complementary advantages and drawbacks.

We expect to integrate the developed manipulation system on a mobile robot and to use it for fruit collection in the field using real tomatoes. Future picking procedures will combine observation and motion to increase grasping success rate and to address other issues like detaching fruits from robust stems.

#### ACKNOWLEDGMENTS

Research carried out within Agritech National Research Center, funded by NextGenerationEU (PNRR, Mission 4, Component 2, Investment 1.4 – D.D. 1032 17/06/2022, Code CN00000022, CUP D93C22000420001).

#### REFERENCES

- [1] T. Ojha, S. Misra, and N. Singh Raghuwanshi, "Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges," *Computers and Electronics in Agriculture*, vol. 118, pp. 66–84, 2015.
- [2] E. Penzotti, D. Lodi Rizzini, and S. Caselli, "A planning strategy for sprinkler-based variable rate irrigation," *Computers and Electronics in Agriculture (COMPAG)*, vol. 212, no. 108126108126, 2023, DOI 10.1016/j.compag.2023.108126, EID 2-s2.0-85132779189.
- [3] T. Mueller-Sim, M. Jenkins, J. Abel, and G. Kantor, "The robotanist: A ground-based agricultural robot for high-throughput crop phenotyping," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2017, pp. 3634–3639.
- [4] Y. Tang, M. Chen, C. Wang, L. Luo, J. Li, G. Lian, and X. Zou, "Recognition and localization methods for vision-based fruit picking robots: A review," *Frontiers in Plant Science*, vol. 11, pp. 1–17, 2020.
- [5] G. Liu, J. Nouaze, P. Touko Mbouembe, and J. Kim, "YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3," *Sensors*, vol. 20, no. 7, 2020.
- [6] M. Sozzi, S. Cantalamessa, A. Cogato, A. Kayad, and F. Marinello, "Automatic bunch detection in white grape varieties using YOLOv3, YOLOv4, and YOLOv5 deep learning algorithms," *Agronomy*, vol. 12, no. 2, 2022.
- [7] Z. Wang, Y. Ling, X. Wang, D. Meng, L. Nie, G. An, and X. Wang, "An improved faster R-CNN model for multi-object tomato maturity detection in complex scenarios," *Ecological Informatics*, vol. 72, p. 101886, 2022.
- [8] Y.-P. Huang, T.-H. Wang, and H. Basanta, "Using fuzzy mask R-CNN model to automatically identify tomato ripeness," *IEEE Access*, vol. 8, pp. 207 672–207 682, 2020.
- [9] X. Du, Z. Meng, Z. Ma, w. Lu, and H. Cheng, "Tomato 3d pose detection algorithm based on keypoint detection and point cloud processing," *Computers and Electronics in Agriculture*, vol. 212, pp. 1–12, 2023.
- [10] B. Zhang, Y. Xie, J. Zhou, K. Wang, and z. Zhang, "State-of-the-art robotic grippers, grasping and control strategies, as well as their applications in agricultural robots: A review," *Computers and Electronics in Agriculture*, vol. 177, pp. 1–12, 2020.
- [11] J. Jun, J. Kim, J. Seol, J. Kim, and H.-I. Son, "Towards an efficient tomato harvesting robot: 3d perception, manipulation, and end-effector," *IEEE Access*, vol. 9, pp. 17 631–17 640, 2021.
- [12] A. Dechemi, D. Chatziparaschis, J. Chen, M. Campbell, A. Shamshirgaran, C. Mucchiani, A. Roy-Chowdhury, S. Carpin, and K. Karydis, "Robotic assessment of a crop's need for watering: Automating a time-consuming task to support sustainable agriculture," *IEEE Robotics & Automation Magazine*, vol. 30, no. 4, pp. 52–67, 2023.
- [13] X. Wang, H. Kang, H. Zhou, W. Au, W. Wang, and C. Chen, "Development and evaluation of a robust soft robotic gripper for apple harvesting," *Computers and Electronics in Agriculture*, vol. 204, p. 107552, 2023.
- [14] A. Saccuti, "ur\_ros\_rtde," 2024. [Online]. Available: [https://github.com/SuperDiodo/ur\\_ros\\_rtde](https://github.com/SuperDiodo/ur_ros_rtde)