# Estimation of Image-Based End-Effector Approach Angles for Tomato Harvesting Robots

**Kizuna Yoshinaga**
*Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan*

**Hikaru Sato**
*Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan*

**Kazuo Ishii**
*Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan*

**Shinsuke Yasukawa**
*Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, 808-0196, Japan*
*Email: kizuna.yoshinaga187@mail.kyutech.jp, sato.hikaru292@mail.kyutech.jp, ishii@brain.kyutech.ac.jp, s-yasukawa@brain.kyutech.ac.jp*

**Abstract**

We propose a method to estimate a suitable approach angle for the end-effector of a tomato harvesting robot based on image data. Agricultural harvesting robots often face obstacles such as other fruits or stems around the target crop. Additionally, it is important to approach the target from a direction appropriate for harvesting, considering the shape of the end-effector. The proposed method uses a deep learning-based instance segmentation model to extract regions of fruits and stems, and estimates the suitable approach angle based on their positional relationships. We demonstrated the usefulness of the proposed method using an image dataset acquired in an actual tomato greenhouse.

*Keywords*: Image processing, Robot vision, Agricultural robot, Harvesting robot

## 1. Introduction

In recent years, aging workers and lack of successors have become major issues in the agricultural sector due to the declining birthrate and aging population. To solve these problems, research and development of smart agriculture using AI and robots has been actively conducted since around the 1970s. However, there are few practical examples of agricultural robots in the field. Two factors are considered to be responsible for this: technological factors due to differences in the complexity and diversity of crops, tasks, and environments, and economic factors such as high cost [1].

There have been many research and development efforts on robots for the selective harvesting of high-value crops, such as strawberries, tomatoes, and bell peppers, which ripen heterogeneously and require careful handling to harvest only the ripe produce while avoiding damage to surrounding plants and unripe fruits[2], [3], [4].

,In our research group, Fujinaga et al. have developed a tomato harvesting robot with a 3-axis orthogonal arm [5], [6]. This robot is characterized by its simple structure. However, this simple structure has a disadvantage that it cannot harvest tomatoes by avoiding unripe fruits and stems in front of the harvesting target. Therefore, our group has developed a new robot with a horizontally articulated arm to avoid obstacles and to adapt the harvesting approach to the target.

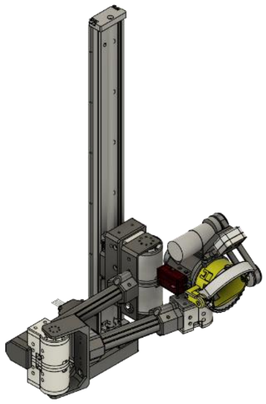
For harvesting that takes advantage of the high degree of freedom of the arm, a method for determining the target posture of the robot's end-effector to the posture of the fruit is required. To determine the target posture of the robot end-effector, it is necessary to extract 3D posture information of the tomato from the image information. The following is an example of a previous study on estimating the 3D posture of a harvesting target. Zhang et al. used a method called tomato pose method (TPM) to estimate the 3D position of a tomato bunch [7]. TPM is based on object detection of a tomato bunch and detection of 11 key points defined as the main stem, stem, and fruit in the bunch. Xiaoqiang et al. proposed a method for detecting tomato posture called the TPD algorithm [8]. The TPD algorithm uses the YOLO-lmk model, which is a modification of YOLOv5s, point cloud segmentation, and a point cloud processing module to achieve fruit and key point detection.

In this paper, we propose and evaluate an image processing method for estimating the direction of tip entry, which is necessary for a robot with a horizontally articulated arm.

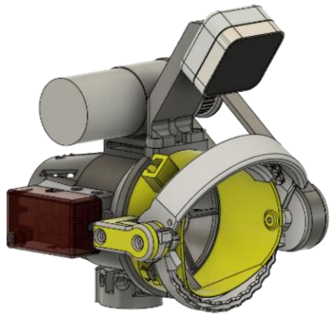
## 2. Proposed Method

### 2.1. *Harvest strategy*

We describe the configuration and specifications of our group's articulated robot to explain the target environment and requirements of the proposed method. Fig. 1(a) shows the 3D CAD model of the robotic arm used in this study. The horizontally articulated arm is mounted on a linear actuator that moves in the direction



(a) Appearance of the robot arm

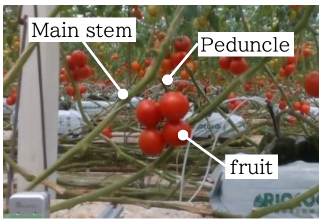


(b) Appearance of the end-effector

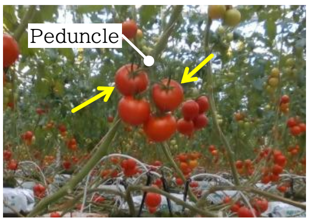
Fig.1 3D CAD model

of gravity. This arm allows the tomatoes to be harvested not only from the front, but also from the side. Fig. 1(b) shows an end-effector with a harvesting mechanism. The suction-cut harvesting mechanism is designed to harvest one fruit per operation. A fan mounted behind the end-effector rotates to draw the fruit in, while blades cut the stems. The proposed image processing method estimates the optimal entry posture for the end-effector.

Fig. 2 shows an example of the appearance of a tomato that is the target of the proposed image processing method. The figure also labels the parts of the tomato as defined in this study. In this study, the thick stem of a tomato cluster is called the "main stem" and the thin stem extending from the main stem to the fruit is called the "peduncle". This study focuses on tomatoes located near the robot and without occlusion, as shown in the figure. In the case of the tomato shown in Fig. 2(a), it can be imagined that any tomato can be harvested because the fruit can be sucked by a frontal approach. However, in the case of the tomato indicated by the arrow in Fig. 2(b), the end-effector interferes with the peduncle when approaching from the front, and the fruit cannot be sucked properly, resulting in a failure of harvest. In this case, the robot can harvest the tomatoes by approaching from the right or left to avoid the peduncles. The proposed method aims to estimate the appropriate approach posture of the end-effector according to the orientation of the fruit by image processing.




(a) Example of a preferred frontal harvest approach




(b) Example of a preferred harvesting approach from the left or right

Fig. 2 Example images of camera view the tomatoes that are the subject of the proposed method.

## 2.2. *Image Processing Algorithm*

After an overview of the proposed method, each process is described in detail. Fig. 3 shows the processing flow of the proposed method.

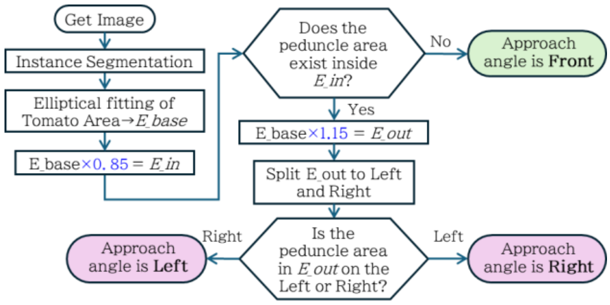



Fig. 3 Process flow of the proposed method.

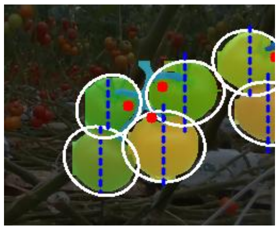
(a) Original image

(b) Instance Segmentation image

(c) Elliptical fitting of tomato area

(d) Check the Peduncle area inside E_in

(e) Check the Peduncle area inside E_out



(f) Which the Peduncle area in E_out is on the right or left

Fig. 4 Output of each processing step.

Fig. 4 shows the output of each processing step. First, an RGB-D camera captures an image of a tomato, and the RGB component (Fig. 4(c)) is segmented into the main stem, peduncle, and fruit regions (Fig. 4(b)). Hereafter, the analysis focuses on the tomato bunches within the square box in Fig. 4(a). Next, the fruit region is approximated by an ellipse by performing ellipse fitting on the fruit region (Fig. 4(c)). The size of the ellipses is changed as needed (Fig. 4(c)-(e)), and the positional relationship between the fruit and the peduncle is obtained based on the criteria of whether and where the peduncle region exists inside these ellipses(Fig. 4(f)).

Based on the acquired positional relationships, the appropriate end-effector entry posture is estimated.

Instance segmentation (IS) is used to segment the main stem, peduncle, and fruit. IS is an image processing task that uses deep learning to predict pixel-by-pixel class labels for objects in an image, and then identify them individually. We adopted IS for the proposed method for the following three reasons. First, it can flexibly adapt to changes in visibility and illumination, Second, the applied images are suitable for analyzing thin objects, and third, it is necessary to recognize individual fruits for harvesting. In this study, the YOLOv8 segmentation model was used as the training model for IS [9].

After segmentation, consider the ellipse E_in (Fig. 4(d)), which is a 15% reduction of the ellipse generated by ellipse fitting (Fig. 4(c)) for the tomato fruit region. If the peduncle region does not exist inside this *E_in*, the peduncle is judged not to be in the front, and the harvest approach is determined to be front. The reason for reducing the original ellipse (=*E_base*) is as follows. Since the generated ellipse is only an approximation, the ellipse often contains areas near the edges that are not tomato areas. This region may include the surrounding peduncle region, which is not directly related to the target tomato, and using the ellipse *E_base* as it is in the processing may affect the estimation results.

If the peduncle region is included in E_in, this means that the peduncle is located in front of the tomato fruit. Therefore, an ellipse E_out (Fig. 4(d)) is defined by expanding E_base by 15%. By calculating the intersection region between E_out and the peduncle region (Fig. 4(e)) within E_out, we can determine the position of the peduncle around the fruit in relation to the fruit (Fig. 4(f)). Finally, the direction in which the intersection region does not exist is determined to be the end-effector's entry direction.

## 3. Experiments and Results

To demonstrate the effectiveness of the proposed method, verification is performed using images from the image dataset used during IS training. The evaluation is based on accuracy. The true value was set by the author's own visual inspection using the following concept. Among the tomatoes in the image used, the tomato to be evaluated is determined, and the approach direction of the end effector that the author himself judges to be appropriate is chosen from the three directions: right, front, and left as true value. The accuracy is calculated by comparing this true value with the estimated value by the proposed method. In addition, the tomatoes to be evaluated were classified into three categories according to the growth direction of the stalk, as shown in Table 1, and 15 tomatoes were selected for each category for evaluation and verification.

Table 1. Categorization of tomatoes to be evaluated.

| | | |
|---|---|---|
| *Ⅰ* | When the peduncles extend toward the front of the image | Fruit head and peduncle are visible. |
| *Ⅱ* | When the peduncles extend sideways of the image | Fruit head and peduncle may or may not be visible in some cases |
| *Ⅲ* | When the peduncles extend toward the back of the image | Fruit head and peduncle are not visible. |

Table 2 shows the accuracy and average accuracy for categories *Ⅰ* to *Ⅲ*. When viewed by category, the highest values were *Ⅲ*, followed by *Ⅰ* and *Ⅱ*.

Table 2. Accuracy by Category and Overall Average.

| | |
|---|---|
| *Ⅰ* | 0.800 |
| *Ⅱ* | 0.600 |
| *Ⅲ* | 0.933 |
| *Ave.* | 0.777 |

The results are analyzed separately for Categories *I*, *II*, and *III*. The estimation accuracy of the proposed method depends heavily on the recognition accuracy of IS. First, In Category *I*, the peduncle is fully visible, enabling accurate recognition, while in Category *III*, the tomato fruit is easily recognized.

Category *III* achieves the highest accuracy because the recognition of tomato fruits, which have simple shapes, is easier than that of peduncles. In Category *II*, the peduncle often extends sideways, and occlusion by the fruit frequently makes parts of the peduncle invisible. This occlusion reduces the recognition accuracy of the peduncle, which subsequently decreases the estimation accuracy. Factors such as the shooting angle, which affects the visibility of peduncles, and the resolution and shooting distance of images, which influence feature extraction, significantly affect recognition accuracy.

## 4. Conclusion

We proposed an image processing method for estimating the direction of entry of the end-effector based on tomato recognition and posture estimation of tomato fruit for implementation in a tomato harvesting robot with a horizontally articulated arm. We confirmed that the proposed method can estimate the direction of entry of the end-effector with an average accuracy of 0.777. In this method, it was analyzed that the visibility of the fruit peduncle, depending on the orientation of the tomato, significantly affects the performance of the pose estimation. To solve this problem, we plan to examine strategies for shooting at close range and re-photographing from different viewpoints, and to develop an IS model that can handle these strategies. Demonstrating the effectiveness of this method through harvesting trials using actual equipment remains a future challenge.

## 5. Acknowledgements

This work was supported by JSPS KAKENHI (grant number: 23H03476).

## References


1. Zhou, H., Wang, X., Au, W., Kang, H., & Chen, C., “Intelligent robots for fruit harvesting: Recent developments and future challenges”, Precision Agriculture, Vol. 41, No. 5, 1856-1907, 2022.
2. Rajendran, V., Debnath, B., Mghames, S., Mandil, W., Parsa, S., Parsons, S., & Ghalamzan-E, A., "Towards autonomous selective harvesting: A review of robot perception, robot design, motion planning and control," Journal of Field Robotics, Vol. 41, No. 7, pp. 2247-2279, 2024.
3. Arad, B., Balendonck, J., Barth, R., Ben-Shahar, O., Edan, Y., Hellström, T., Hemming J., Kurtser, P., Ringdahl O., Tielen, T., & van Tuijl, B., "Development of a sweet pepper harvesting robot," Journal of Field Robotics, Vol. 37, No. 6, pp. 1027-1039, 2020.
4. Xiong, Y., Ge, Y., Grimstad, L., & From, P. J., "An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation," Journal of Field Robotics, Vol. 37, No. 2, pp. 202-224, 2020.
5. Fujinaga, T., Yasukawa, S., & Ishii, K., "Development and evaluation of a tomato fruit suction cutting Device," 2021 IEEE/SICE International Symposium on System Integration (SII), 628-633, 2021.
6. Fujinaga, T., Yasukawa, S., Li, B., & Ishii, K., " Image Mosaicing Using Multi-Modal Images for Generation of Tomato Growth State Map," Journal of Robotics and Mechatronics, Vol. 30, No. 2, 187-197, 2018.
7. Zhang, F., Gao, J., Zhou, H., Zhang, J., Zou, K., & Yuan, T., "Three-dimensional pose detection method based on keypoints detection network for tomato bunch," Computers and Electronics in Agriculture, Vol. 195, 106824, 2022.
8. Du, X., Meng Z., Ma, Z., Lu, W., Cheng, H., "Tomato 3D pose detection algorithm based on keypoint detection and point cloud processing," Computers and Electronics in Agriculture, Vol. 212, 108056. 2023.
9. Jocher, G., Chaurasia, A., & Qiu, J., "Yolo by ultralytics," Code Repository, 2023.


## Authors Introduction

Mr. Kizuna Yoshinaga

He received his Bachelor of Engineering degree in 2023 from National Institute of Technology, Sasebo College, Advanced Engineering Course Department of Complex Engineering. Currently enrolled in the Master's Program of the Graduate School of Kyushu Institute of Technology.

Mr. Hikaru Sato

He received his Bachelor of Engineering degree in 2023 from National Institute of Technology, Tsuruoka College, Advanced Engineering Course Department of Production Systems Engineering. Currently enrolled in the Master's Program of the Graduate School of Kyushu Institute of Technology.

Dr. Kazuo Ishii

He is currently a Professor at the Department of Human Intelligence Systems of Kyushu Institute of Technology, Japan. He obtained his M. S. degree in 1993 and his D. Eng. degree in 1996 at The University of Tokyo. His research interests are in the fields of Underwater Robotics, Field Robotics, Neural Networks and Intelligent Systems.

Dr. Shinsuke Yasukawa

He received his Ph.D. degree in electrical, electronic, and information engineering from Osaka University, Japan, in January 2017. Currently, he is an Associate Professor in the Department of Human Intelligence Systems, Kyushu Institute of Technology, Japan. His research interests include information processing in biological systems and their applications in robotics.