Design and Development of Harvesting Mechanism for a Tomato Harvesting Robot

Ying-Ti Weng and Kuang-Wen Hsieh

Corresponding Author’s e-mail address, wong_yingti@smail.nchu.edu.tw
Department of Bio-Industrial Mechatronics Engineering, National Chung Hsing University, Taiwan (R.O.C)

Keywords: Tomatoes, Image recognition, Stereo vision, harvesting robot

ABSTRACT
The labor shortage resulting from the decreasing agricultural workforce in recent years has driven the advancement of smart farming practices. This experiment focuses on the development of image recognition technology and harvesting robots for beef tomato cultivation in greenhouses. The end-effector used in this study for tomato harvesting combines ripeness and pedicel recognition, stereo vision-based 3D positioning, and pedicel angle computation, and has undergone testing in beef tomato harvesting. To enhance the success rate of tomato harvesting, two methods were employed: multiple fruit harvesting and single fruit harvesting. Image recognition determined the position and angles of the pedicels or peduncles, while the cutting and grasping of tomatoes were performed by the end-effector developed in this study. The image recognition system and end-effector were integrated into a smart harvesting system, and their effectiveness was evaluated through harvest success rate tests. The success rates for pedicel angle positioning and peduncle angle positioning, both based on image recognition, were 92.06% and 85.7% respectively. A total of 163 integrated tests were conducted, with an accuracy rate of 66.3% for single fruit positioning and 82.4% for multiple fruit positioning. The end-effector successfully harvested tomatoes 121 times, achieving an overall success rate of 77.7%: 81.4% for multiple fruit harvesting and 74.2% for single fruit harvesting. On average, each beef tomato harvesting process took approximately 27 seconds.

1 Introduction
In recent years, the agricultural industry has faced labor shortages due to population aging, leading to research in agricultural mechanization for harvesting. Previous studies have explored methods for measuring ripe tomato fruits in natural environments, classifying them based on separation, adjacency, overlap, and leaf occlusion in images [1]. Additionally, Qingkuan et al. developed a machine vision-based agricultural implement guidance system in 2015[2], utilizing the HIS color model to process images and identifying crop guidance lines based on their features, ultimately controlling the implement for inter-row weeding purposes. Building upon these studies, this research has developed an automated harvesting robot suitable for greenhouse-grown beef tomatoes. The robot utilizes a stereo camera to capture images, locates the positions of beef tomatoes, determines their suitability for harvesting based on ripeness, calculates the three-dimensional positions of targeted beef tomatoes, and connects the smart harvesting system to the robot arm to initiate the harvesting process. The end effector integrates gripping and cutting functions to prevent damage to tomatoes during harvesting and improve efficiency. Considering the varying ripeness of tomatoes, the harvesting robot can be configured for either multiple-fruit or single-fruit harvesting modes, with the former targeting the pedicels and the latter focusing on the pedicel angles.

1) The practice of mechanized agriculture is increasing due to the shrinking population of agricultural workers.
2) A stereo camera was used to judge the ripeness and 3D positions of beef tomatoes.
3) The end effector was designed with a combination of gripper and cutter to increase the success rate of harvesting.
4) The target of harvesting was pedicels in the multifruit harvesting mode and peduncles in the single-fruit harvesting mode.

2 Experiment
Figure 1 presents the beef tomato harvesting robot. Integrating image capture, object detection, and stereo vision technologies, this robot captures images, analyzes these images with an object detection model, calculates the 3D coordinates of beef tomatoes based on the analysis results, and converts them into corresponding coordinates in the robot arm coordinate system to complete harvesting tasks. The variety of beef tomato used in this study was TMB-768. As shown in Fig. 2, the end effector of the proposed robot is equipped with a pneumatic parallel gripper that has block-shaped fingers, angled fingertips, and knives on the inside of the fingers. The end effector weighs 706 g in total and operates at the pressure of 0.4 MPa with a horizontal stroke of 36 mm.

1. Detecting ripeness: The Mask region-based convolutional neural network was used to classify beef tomatoes into seven types according to their ripeness.

2. 3D positioning with a depth camera: The 3D positions of ripe beef tomatoes were calculated.
3. Calculating the angles of pedicels and peduncles: A precision camera was employed for the calculations.

4. Measuring physical properties: The lengths and widths of the peduncles and pedicels, the weights of individual beef tomatoes, and the weights of individual tomato bunches were measured.

5. Designing the end effector: The end effector is a parallel gripper that has block-shaped fingers, angled fingertips, and knives on the inside of the fingers; weighs 706 g in total; and operates at the pressure of 0.4 MPa.

6. Evaluating the performance of the end effector: The performance was determined by testing its gripping performance as well as carrying capacity in the multifruit and single-fruit harvesting modes.

7. On-site testing: The system’s success rate in field harvesting was determined.

Smart harvesting system


Fig. 2 Structure of the end effector: parallel gripper, block-shaped fingers, knives, and angled fingertips.

2.1 Training Dataset and Model Development for Image Detection

During the initial phase of training, the desired objects for detection in this study were manually labeled as samples. These labeled data, along with the original images, were used to train a neural network. The algorithm compared the machine’s predicted results with human judgments, calculating errors and adjusting internal parameters accordingly. The collected data were divided into a training dataset and a validation dataset.

For this study, 825 images captured by a depth camera were used for training and validation. The training dataset consisted of 65 images, while the validation dataset comprised 50 images. To augment the training data, the images were horizontally flipped, effectively doubling the dataset size.

2.2 Coordinate Transformation between Camera and Robotic Arm

The study is divided into two parts. The first part involves combining the relative position and depth information of the tomatoes calculated by the depth camera with the relative position of the depth camera and the high-precision camera’s origin. The second part involves merging the relative position of the tomatoes calculated by the high-precision camera with the relative position of the high-precision camera and the robotic arm’s origin. The coordinate systems of the depth camera and the high-precision camera are shown in Figure 3, while the coordinate systems of the high-precision camera and the robotic arm are shown in Figure 4.

Fig. 3 Coordinate Transformation between Depth Camera and High-Precision Camera

Fig. 4 Coordinate Transformation between Depth Camera and Robotic Arm

3 Results

1. Detecting ripeness: The system was verified with a precision of 97% in model-based testing, as shown in...
2. 3D positioning with a depth camera: The average error between the calibrated predictions and the actual positions was 1.67 cm.
3. Calculating the angles of pedicels and peduncles with a precision camera: The percentages of pedicels and peduncles that were correctly identified were 92.06% and 87.5%, respectively.
4. Investigating the physical properties of beef tomatoes: The maximum downward force was 14.56 N in the multifruit harvesting mode and 3.18 N in the single-fruit harvesting mode. A cutting force of up to 70 N was required to cut the pedicels and peduncles.
5. Evaluating the performance of the end effector: The theoretical gripping force of the end effector is 278.1 N. Fig. 3 shows the actual operation of the end effector in harvesting. The end effector withstood a minimum downward force of 18.7 N in the multifruit harvesting mode and 22.4 N in the single-fruit harvesting mode.
6. On-site testing: The smart harvesting system achieved a successful harvesting rate of 77.7%.

4. Conclusions
1. The system was verified with a 97% precision in model-based testing. On-site tests are subject to environment factors.
2. The robot arm used in this study is a homemade five-axis robot arm and has a limited range of motion for harvesting operations.
3. The end effector had a carrying capacity and cutting force greater than those required, as indicated in the physical property measurements, and would thus be applicable in field harvesting.
4. The successful harvesting rate may be improved by modifying the design of the gripper’s fingers.

5. The smart harvesting system achieved a 77.7% successful harvesting rate in the field test.

References