Received November 27, 2020, accepted December 28, 2020, date of publication January 18, 2021, date of current version February 1, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3052240

Towards an Efficient Tomato Harvesting Robot: 3D Perception, Manipulation, and End-Effector

JONGPYO JUN, JEONGIN KIM, JAEHWI SEOL, JEONGEUN KIM, (Student Member, IEEE), AND HYOUNG IL SON, (Senior Member, IEEE)
Department of Rural and Biosystems Engineering, Chonnam National University, Gwangju 61186, South Korea

Corresponding author: Hyoung Il Son (hison@jnu.ac.kr)

This work was supported by the Technology Innovation Program funded by the Ministry of Trade, Industry & Energy (MOTIE) under Grant 10060070.

ABSTRACT Fruit and vegetable harvesting robots have been widely studied and developed in recent years. However, despite extensive research commercial tomato harvesting robots still remain a challenge. In this paper, we propose an efficient tomato harvesting robot that combines the principle of 3D perception, Manipulation, and an End-effector. For this robot, tomatoes are detected based on deep learning, after which 3D coordinates of the target crop are extracted and motion control of the manipulator based on 3D coordination. In addition, a suction pad featuring the kirigami pattern, which is a part of the suction gripper, was developed to grip individual tomatoes in clusters. A scissor-shaped cutting module with an assist unit, which is used to overcome structural limitations and implement effective cutting, was also desinged and tested. The proposed tomato harvesting robot was validated and evaluated on a laboratory testbed basd on the performance of each component. Therefore, in this study, we propose and verify a new robot design for the effective harvesting of tomatoes.

INDEX TERMS Harvesting robot, end-effector, 3D perception, tractional cutting unit.

I. INTRODUCTION

Smart farms offer an advantage in terms of being able to provide stable supplies of produce throughout the year, and because they can manage crops more efficiently, research on smart farms is ongoing worldwide. Some of the crops cultivated in smart farms, such as tomatoes, are manually harvest; however, the available labor is insufficient because of the declining population of agriculture workers. Therefore, a number of researchers have begun investigating methods to harvest fresh fruit using agricultural robots in a greenhouse [1], [2]. To reduce the labor required, many attempts have been made to apply harvesting robots to the field, but it was reported in [3] that there were was not commercialized.

In agriculture, crops and fruits are grown in an unknown and unstructured environment. Because the physical properties of a crop, such as surface strength, weight, and size, differ depending on the type of crop and its shape, and damage to the fruit tree or crop plant can be reduced only with consideration of the characteristics of the crop. Because of the irregular characteristics of crops, most harvesting robots have been developed for specific crops. Therefore, several researchers have begun investigating methods to harvest fresh fruit, such as tomatoes, using agricultural robots in greenhouses. However, performing fast, accurate and intact harvesting of fruit grown in greenhouses is still a problem for robots [4].

For robots to be able to perform effective harvesting they should have the ability to detect and locate target crops, via deep learning-based fruit recognition and 3D perception of the access to target crops. When the 3D position of the target crop is acquired, the coordinates are then used for the motion control of the manipulator.

However, for tomatoes in particular, because the plants grow in clusters, the detection of target crops is more difficult. Furthermore, tomatoes possess weak surface strength and slippery surfaces, making it considerably difficult to grasp commercial-grade fruits. Careful gripping of the fruit is critical, because the value of a tomato decreases when a small sharp object blemishes its appearance. In addition, clusters of tomatoes usually grow in unpredictable directions, making it very likely for the fruit to be damaged by the end-effector during the harvesting process. Therefore,

The associate editor coordinating the review of this manuscript and approving it for publication was Luigi Biagiotti.

developing an end-effector that considers the growth environment and physical properties of tomatoes is essential to preventing crop damage during harvesting. Because there should be no damage to the target crop and surrounding crops, it is necessary for a robot to be able to accurately estimate the 3D coordinates of the fruit and access to the target crops.

In this paper, we propose a harvesting robot that combines the elements of each if these challenges for the effective harvesting of tomatoes. In this robot, 3D perception based on deep learning detects the tomato fruit and approaches the fruit through motion control of the manipulator. With the end-effector, the harvesting is completed using a soft material gripper module and a scissor type cutting module.

II. RELATED WORK

Fruit harvesting offers important opportunities in the field of agricultural robotics and has received significant attention from researchers in recent decades. Several robots have been developed for harvesting fruits and vegetables such as apples, sweet peppers, cucumbers, kiwi fruit, strawberries, and tomatoes. Research related to harvesting robots can be classified into fruit perception [5]–[12], manipulation [13]–[15], end-effectors [7], [16]–[18].

A. FRUIT PERCEPTION

Fruit perception is defined as image processing and sensor-based determination of status and location of the status and location of a fruit tree. Because the presence of many undesirable factors, such as non-uniform lighting, unstructured fields, occlusion, and other unpredictable factors, in actual environments, it is considerably difficult to determine the exact state and location of a fruit tree [5]. Nevertheless, research efforts continue to attempt to solve these problems. For example, a machine vision system based on color thresholding methods- [6], [7] employed to distinguish between ripe strawberries and other strawberries and plants. As reported in [8], for strawberry detection, image processing based on color threshold values is a method that is often applied in research. Also color, depth, shape information is used to detect spherical or cylindrical properties of fruit.

Fluctuating illumination and weather conditions, complicated environments, and dense fruits make it difficult to apply robotics to agriculture. With the development of machine learning, many studies focusing of the application of deep learning to agriculture have been conducted. Deep learning has been used for leaf classification [9], yield estimation via machine learning [10], and fruit detection in orchards using Faster R-CNN [12]. However, the speeds of these methods are notably low, thus making these methods unsuitable for real-time detection with high image resolution in actual harvesting scenarios. On the other hand, the You Only Look Once (YOLO) [11] method enables high-speed detection with high accuracy; thus, it is widely used for real-time detection.

It is used for leaf classification [9], yield estimation using machine learning [10], fruit detection in orchards using Faster R-CNN [12]. However the speed of these method is so slow that it is not proper to real-time detection with high image resolution in harvesting condition. The You Only Look Once (YOLO) [11] method not only provides high accuracy detection but also has fast speed. It is widely used for real-time detection.

B. MANIPULATION

Because of unknown and unstructured environments, such as the presence of clusters of fruits and canopies, manipulation is considered one of the major challenges in the development of harvesting robots [13]. Harvesting from clusters is difficult because surrounding fruits, leaves, stems, and other obstacles are difficult to isolate from the target during detection and manipulation.

For harvesting fruits or vegetables without damage to the surrounding environment, a number of studies focus on the path planning and motion planning of manipulators. Based on the detected fruit, the end-effector of the robotic arm must move to a position where it can harvest fruit, while avoiding obstacles. However this problem is further complicated by the necessity of controlling both the position and orientation of the end-effector.

Manipulator motion planning is performed using visual servoing to maintain a predetermined position while moving to the image center coordinates of the detected fruit [14]. To avoid and harvest obstacles in the work environment, the manipulator path is planned by modeling of the surrounding obstacles. Different motion plans are established according to the pose of the fruit and the end-effector [15].

C. END-EFFECTOR

End-effectors for harvesting have been developed according to the different methods used to harvest existing crops. In general, a blade is used to harvest the fruit. It was developed mainly for the purpose of separating fruits from stems, specifically as a scissors-type end-effector [7], [16]. In [17], the fruit was separated from the stem through the rotation of the end effector's infinite rotation joint, such that a blade is not required. Harvesting methods that use a blade necessitate accuracy in estimating the stem whereas harvesting mechanisms using rotation do not consider the stem direction and are robust against estimation errors. For paprika, another harvesting method was developed, wherein the stem and fruit are separated using a high-temperature arc generated by a connection with an electrode wire [18]. Compared to cutting with scissors, this approach has an advantage in terms of delaying or reducing infections such as viral diseases.

III. HARVESTING ROBOT SYSTEM

The harvesting robot system proposed in this paper was designed and developed based on the following concepts: perception, manipulation, and end-effector. The harvesting robot system is shown in Fig. 1. Fruit detection, motion control, cutting modules and grasping modules for efficient tomato harvesting robots are described in detail in the subsequent subsections. The flow-chart for the overall operation of the

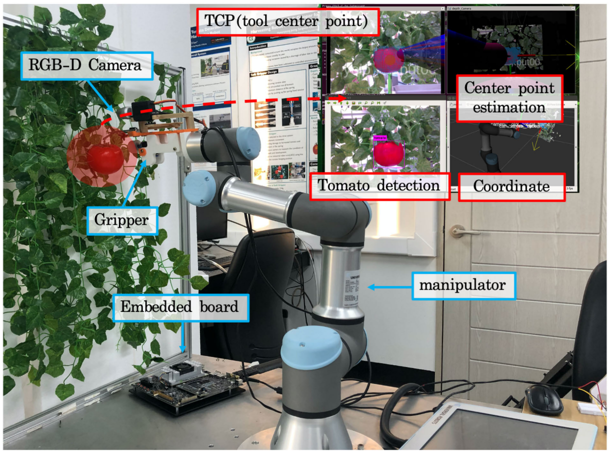

FIGURE 1. Harvesting robot system setup comprising UR3 arm, Intel Realsense D435 RGB-D camera, and an end-effector.

harvesting robot, show in Fig. 2, provides an overview of the sub-steps, including motion control, of the harvesting process.

A. HARDWARE SETUP

The harvesting robot system consists of a 6-DOF manipulator(UR3), a custom end-effector and embedded board (Jetson TX2), and RGB-D camera(Intel Realsense D435). The RGB-D camera is attached to the end-effector, and is used to transmit the pose data of the detected tomato to the embedded board via USB communication. The Embedded software environment of the harvesting robot system consists of an Dual-Core NVIDIA Densor 2 64-Bit CPU and Quad-Core ARM®Cortex®-A57 MPCore, a 8G 128 bit LPDDR4 Memory 1866 MHx RAM, 256-core NVIDIA Pascal GPU architecture with 245 NVIDIA CUDA cores, Jetpack SDK 4.2, Ubuntu 18.04 LTS 64-bit, and CUDA vesrion 10.2.

B. SOFTWARE SETUP

The software system is defined based on the Robot Operating System (ROS) framework. The system comprises customized subsystems shaped as a node, as shown in Fig. 3. Various open software libraries were tested for each function implementation and linkage, and motion control was enabled based on a selection of appropriate package that were linked with the robot's controller.

Software is written in Python and C++ mostly, running on ROS melodic on Ubuntu 18.04. Most programs, such as detection and planning is executed on the embedded board mounted on the system. The complicated function is running on the board's GPU and CUDA 10.2 is installed to accelerated it. The system has customized subsystems shaped as node shown as Fig 3.

Various open software libraries tested for each function implementation and linkage, and motion control is enabled by selecting appropriate packages and linking them with the robot's controller. The motion control software ROS Moveit! package was used to make a appropriate order to move manipulator and solve kinematics.

C. TOMATO DETECTION

Fruit detection is a central function of the proposed tomato-harvesting system. To detect a particular fruit in the real world, robust data regarding the fruit are required. From the viewpoint of a vision camera, fluctuating illumination and weather conditions of the surroundings make it difficult to detect the fruit [14], because the same color can appear as differently colored pixels depending on the environment. To solve this problem and obtain robust results, it is necessary to use a deep-learning model.

1) YOLOV3 MODEL

YOLO is not only suitable for real-time detection with any other CNN model but also features high accuracy [19]. In comparison to R-CNN, YOLO describes the detection work as a single regression problem. YOLOV3 [20] was developed from YOLO [11] and YOLOV2 [21]. Compared with the network used in YOLO and YOLOV2, the network used in YOLOV3 is Darknet-53 composed of 53 convolutional layers. Its run time is shorter, and its accuracy is high.

The YOLO detection model is presented in Fig. 4. The network divides the input image as a training set into $\mathcal{S} \times \mathcal{S}$ grids. If the center of an target ground truth lies in a grid,

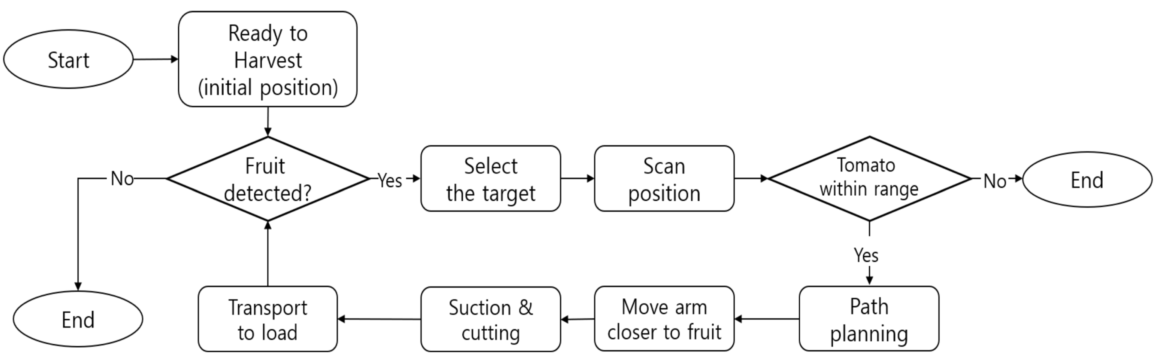

FIGURE 2. The flowchart of harvesting.

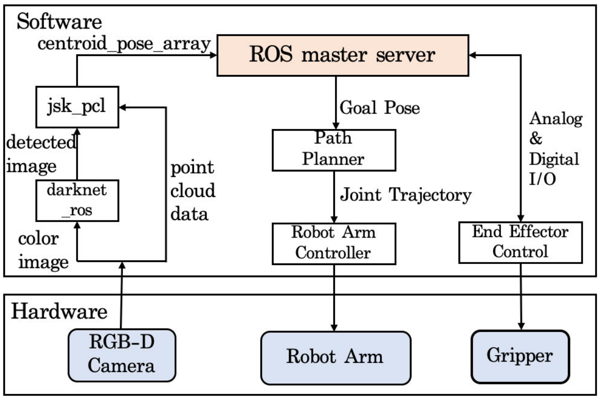

FIGURE 3. System architecture illustrating software nodes connected by ROS: darknet_ros node detects the tomato using the YOLOV3 deep learning model; Jsk_pcl node constructs centroid_pose_array with the detected image and point cloud data; harvesting construction is operated by path planner, robot arm controller, and end-effector control nodes.

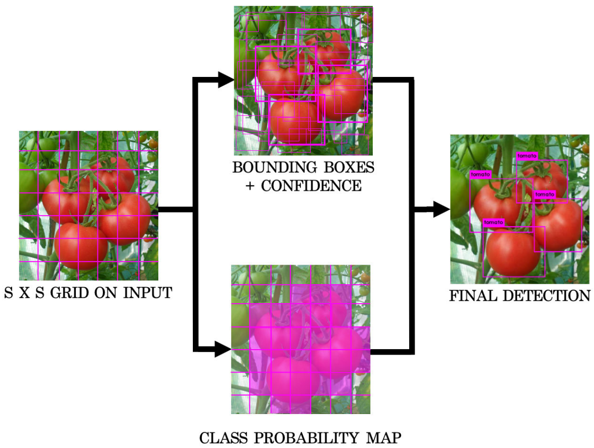

FIGURE 4. YOLO object detector pipeline.

the grid cell is responsible for detecting the target. Each grid cell predicts $\mathcal{B}$ bounding boxes and confidence scores, as well as $\mathcal{C}$ class conditional probabilities. Confidence is defined as follows:

$$confidence = p_r(\text{ Object }) \times IoU_{pred}^{truth}, p_r(\text{ Object }) \in \{0, 1\}$$

If the target is in the grid, $p_r(\text{Object}) = 1$; otherwise, it is 0. IoU_{pred}^{truth} represents the juncture between the predicted box and the ground truth. Confidence denotes whether the objects are in the grid and the accuracy of the predicted bounding box if it contains objects [11].

2) DATASET

In this study, the image data were acquired from a tomato farm located in Gimje-si, Jeollabuk-do, Republic of Korea, using an RGB-D camera with a resolution of 1920 × 1080. The dataset was acquired from a facility horticulture farm in June 2019. From the tomato farm, 770 images were collected among them 70 images were randomly selected for the test dataset. The tomatoes were labeled with bounding boxes in advance using marking boundary software Yolo_mark. After training the model 4000 times in portion to the number of classes, we obtained weights for which the loss was close to zero.

3) YOLOV3 RESULTS

Tomatoes that were detected above the 0.95 threshold were identified by a bounding box. We evaluated performance using the area under the precision recall curve [22]. The precision and- recall equations are as follows:

$$Precision = \frac{TP}{TP + FP}, \quad (1)$$

$$Recall = \frac{TP}{TP + FN}, \quad (2)$$

where TP is the number of true positives (correct detections), FP is the number of false positives(wrong detections), and FN is the number of false negatives(mis-detections). AP is the value of the average precision for class. The results in Table 1 show the performance of the learned deep-learning model.

TABLE 1. Performance of learned deep-learning model.

Precision	Recall	mAP
0.80	0.91	0.9082

D. POSE DETECTION

For its end effector to reach the fruit, it has to view the scene as a three-dimensional(3D) world to determine the pose of the fruit. A pose is composed of a position vector and an orientation vector [23]. An RGB-D camera is used to obtain the pose of the target. It generates point cloud data that indicate the distance between the target and the camera. The point cloud data are an assemble of mass from wasted time and distance data from the object and are obtained using Lidar sensor or an RGB-D sensor. We used an open-source point cloud library to easily handle the point cloud data, enabling the robot to perceive the 3D world. An algorithm that estimates features is then incorporated [24]. By using this library, we combine the target image detected by YOLOV3 with point cloud data. Fig. 5 describes the method of obtaining the pose array of a detected tomato. When tomato is perceived by the YOLOV3 model, the center of the detected box is denoted as a tool center point(TCP) coordinate. Using point cloud data, we can determine a centroid pose array for the detected fruit.

E. MANIPULATION

After the tomatoes are detected, the target and sequence of the harvesting process are determined. To harvest the tomatoes according to the determined method, the manipulator containing the developed gripper module needs to be moved to an operable position. Fig. 6 shows the mapping process between the end-effector attached to the manipulator and the tomatoes targeted for harvesting.

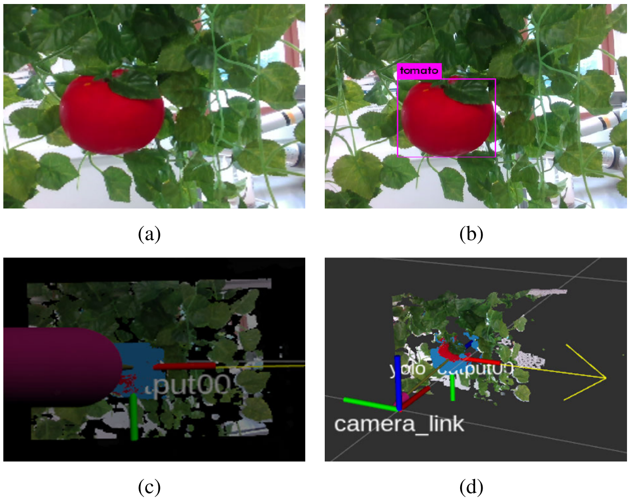

FIGURE 5. **Method of estimating fruit 3D coordination: (a) Original color image; (b) Detecting fruit using YOLOV3 model; (c) Point cloud data; (d) 3D coordinate of fruit.**

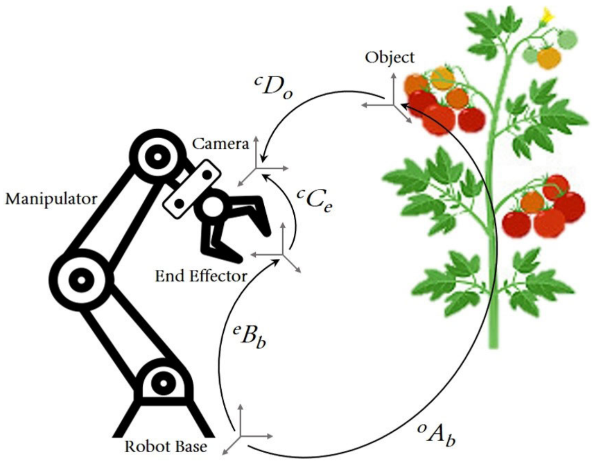

FIGURE 6. **Mapping for motion control.**

The end-effector has an RGB-D camera attached to it for the detection of tomatoes. Because the interrelationship between the end effector and camera affects the success rate of harvesting, hand-eye transformation is essential. Hand-eye cameras and end-effectors are characterize by rigid transformation relationships that include rotation and translation, which enable hand-eye calibration through calculations. Calibration effectively reduces errors in the positioning system.

We performed it based on the position and posture where the end-effector can be moved for cutting up to the center point of the tomato. The relationship for calibration between the robot base and object is modeled [25] in Eq (3):

$$\mathbf{DA} = \mathbf{CB} \tag{3}$$

where $\mathbf{D}$ is a simple representation of ^cD_o for the transformation from the object coordinate frame to the camera coordinate frame. In addition, $\mathbf{A}$ is ^oA_b for the transformation from the robot base coordinate frame to the object coordinate frame, and $\mathbf{C}$ is ^cC_e for the transformation from the end-effector coordinate frame to the camera coordinate frame. Finally, $\mathbf{B}$ is ^eB_b for the transformation of the robot base frame to the end-effector coordinate frame. Each matrix is expressed in Eq. (4) based on the decomposition of the rotational part $\mathbf{R}$ of dimensions 3×3 and the translation vector $\mathbf{t}$ of dimensions 3×1:

$$\begin{bmatrix} \mathbf{R}_D & \mathbf{t}_D \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_A & \mathbf{t}_A \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_C & \mathbf{t}_C \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_B & \mathbf{t}_B \\ \mathbf{0}^T & 1 \end{bmatrix} \tag{4}$$

As a result, the object coordinate can be estimated using Eq. (5) for the relationship and condition of the variables:

$$\mathbf{A}_i = \mathbf{C}\mathbf{B}_i\mathbf{D}^{-1} \tag{5}$$

Furthermore, the robot arm should be able to move in a specific posture to the converted position coordinates. The specific posture was set based on optimum state the robots can solve kinematics. And it should be able to relocate continuously along the generated path. By changing the joint angle of each motor, the robot moves to a specific position, and motion control is achieved as a the result.

F. END-EFFECTOR

1) CUTTING MODULE DESIGN

Cutting from the pedicel is one of the most necessary tasks in tomato harvesting. By removing the pedicel as close to the fruit surface as possible, a robot can prevent damage to the tomatoes during the transfer process and preserve the value of the fruit as a commodity. For this purpose, we used a pair of scissors, which is the tool used in traditional harvesting, as a cutting tool of our robot. Basically, the principle of the lever was applied to the scissor mechanism, and this can be expressed as shown in Fig. 7.

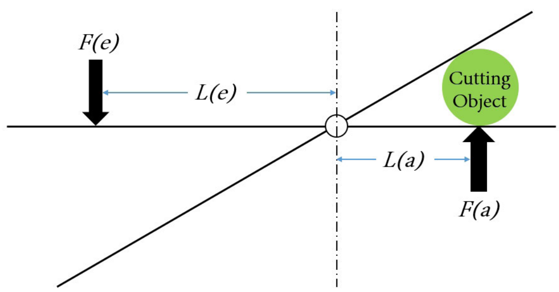

FIGURE 7. **The mechanism of scissor structure.**

The force required to cut the pedicel can be calculated based on these mechanisms [26]. Fig. 7 can be expressed as Eq. (6) by moment parallelism, and the required acting force based on the relationship of each variable is calculated:

$$F(e)L(e) = F(a)L(a) \tag{6}$$

However, the required force varies depending on the position of the cutting object. The required force for the change in variables can be expressed based on Eq. (6) to Eq. (7):

$$F(e)_{min} \geq \frac{F(r)_{max}L(a)}{L(e)} \tag{7}$$

where the acting force $F(a)$ that actuates at the cutting position $L(a)$ can only cut it the stalk if it is greater than the cutting resistance $F(r)$ of the pedicel. Additionally, the minimum acting force must be greater than the maximum cutting resistance for stable cutting. Based on this relationship, the minimum exerted force $F(e)_{min}$ for cutting at the point $L(e)$ is defined.

According to this the relationship, for the cutting process to be efficient, it must be performed as close to the rotating shaft as possible. However, depending on the surface characteristics of the cutting target, the pedicle may not be easily cut and may be pushed out. In particular, this is likely to occur with hard or tough objects such as tomato stems. Because of the possibility of slippage occurring during the cutting process, a momentary large force will also be necessary. Despite this measures, it is possible that the pedicle is not cut completely. Fig. 8 shows a cutting module applying the tractional cutting unit (TCU) developed to ensure clear-cutting performance in the scissor structure.

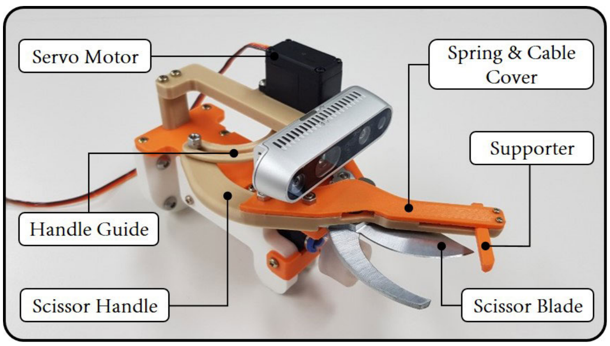

FIGURE 8. The tractional cutting unit for scissors.

The operational process of the TCU with the scissors is depicted in detail in Fig. 9. First, the winch part and scissor blade begin to rotate via the motor drive, and simultaneously, the support moves along the central axis. Second, the pedicel is towed in the direction of the central axis, which prevents it from being pushed out of the scissor blade. With the scissor blades completely overlapping, the pedicel is then cut and separated. Finally, after complete cutting, the support moves to its initial position by means of the reverse rotation of the motor and the elasticity of the spring.

The rotation of the scissors and the operation of the TCU can be performed simultaneously using one motor. When the scissors blade rotates for cutting, the supporter transmits the rotational force as traction force through the cable connected to the handle guide. The pedicel is prevented from being pushed out from the cutting area by the supporter, and complete cutting is performed as the scissor blades overlap. This mechanism can compensate for problems that can occur during the cutting process and enable efficient cutting for harvesting robots that use scissors.

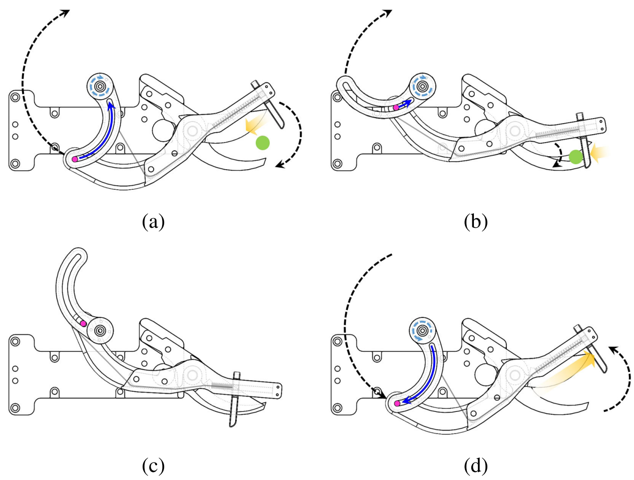

FIGURE 9. Detailed image of the TCU operation: (a) The rotating scissor blade and the moving support; (b) The towing pedicel; (c) The overlapping scissor blades and cutting pedicel; (d) The moving support to its initial position.

In addition, the connecting structure of the handle and guide was selected to deliver sufficient cutting power to the scissors. handle and guide are connected at a position as far as possible from the rotation axis of the scissor blades. Through this design, a greater acting force and torque are generated Fig. 9. Furthermore, to transmit a relatively high torque from the motor, the structure was designed such that the center of motor rotation is approached as closely as possible.

Applying the motor was applied with the same specifications, it was not easy for the previously designed prototype, with a general rotating structure, to cut a pedicel with a diameter of 3.5mm. However, when the modified structure design was applied, even stems with diameters of up to 6mm were successfully cut.

2) GRIPPER MODULE DESIGN

The design of the end-effector for tomato harvesting is one of the main factors that determine the performance of the harvesting robot. In particular, the surfaces of tomatoes are slippery and moist, making it difficult to grip the fruits. To minimize damage to the crops suction grippers with, a soft material are used to pick the fruit. When not under excessive pressure, grasping through suction is one of the methods that can minimize damage to cause by negligence.

The proposed suction gripper creates a pressure difference between the inner and outer surfaces, thereby enabling the grasping of objects such as tomatoes. The capacity of the suction pad depends on the time required to create a vacuum between the object and the suction gripper [27]. A soft material suction pad has an advantage in terms of flexibly adapting to the surface, but may require structural changes when heavy objects need to be lifted. Therefore, our goal was to study the design form of the suction pad to enable it to grip non-structured objects, specifically tomatoes, more smoothly. However, soft elastic materials can bend, twist, compress, or stretch, which makes their modeling difficult. Therefore, a gripper that utilizes a single suction pad, rather than a large number of suction pads, to improve its suction capacity should be constructed. The suction pad is adaptable for contact with the round surface of the tomato and is constructed with a kirigami structure as shown in Fig. 10.

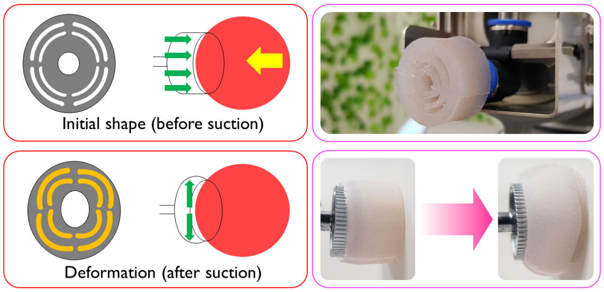

FIGURE 10. **Kirigami-based suction pad.**

In theory, a suction pad with an extremely small, flat circular opening would fit any surface because it could approximate an infinite planar surface [28]. The suction cup holding force is directly proportional to the opening surface area, expressed by

$$F = PA \tag{8}$$

where F is the holding force, P is the vacuum pressure, and A is the suction pad opening surface area.

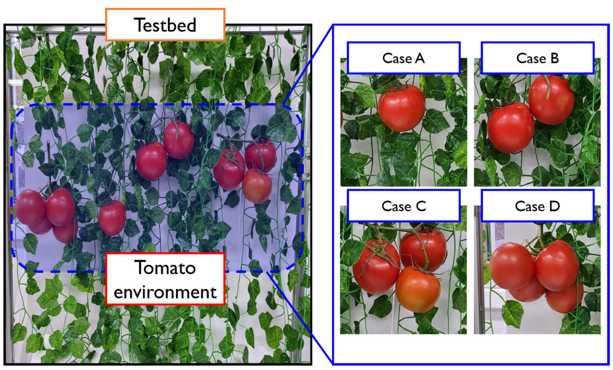

FIGURE 11. **Testbed setup according to growing type: case A: isolated ripe tomato; case B: two ripe tomatoes; case C: three ripe tomatoes; case D: four ripe tomates.**

IV. EXPERIMENT

A. EXPERIMENTAL SETUP

The experiment was conducted on an testbed in the laboratory, using the setup described in Chapter 3. The testbed was constructed in an environment similar to that of an actual tomato farm in order to evaluate the performance of the harvesting robot in practical harvesting scenarios as shown in Fig. 11. Typically, tomatoes grow from stems, forming clusters containing up to four tomatoes. Thus, the stems used in the experiment carried clusters containing one to four tomatoes. Although it is not possible to determine the distributions of all tomatoes by a specific type, the tomatoes that for a cluster can be classified. The experiments can be divided into cases where there are only 1, 2, 3, or 4 in a cluster. The greater the number of tomatoes in a cluster, the more difficult it is to harvest them. Additionally, if a tomato has a pedicel attached to it, the difficult involved on harvesting is further increase. It was found that the numbers of tomatoes in the clusters affected the experimental results.

B. EXPERIMENTAL RESULT

The harvesting cycle includes all the sequences that operate to successfully harvest the fruit. To evaluate the performance of the robot, two main performance measures, i.e., success rate and cycle time, were determined. The total cycle time can be obtained via the addition of the amounts of time consumed for each subtask. For a successful harvest, and time consumed in each step: perception, manipulation, and harvesting is logged. The amounts of time recorded for the subtasks and the total cycle time are listed in Table 2. Manipulation divided the trajectory sequence into sub-steps to avoid obstacles, and additional time was required due to the path planning involved in each step. After approaching the fruit, the process of detaching the fruit was regarded as harvesting. The total cycle time was 5.9 seconds, and in the testbed, the fruit was located close to the robot arm, so the harvesting speed was relatively fast.

In general, harvest success indicates that fruit is harvested without damage. Therefore, one of the important points when using harvesting robots is to harvest target fruits and surrounding crops without damage. The experiment result performed on the testbed to verify the performance of the proposed harvesting robot are shown in Fig. 12 and Table 3. The performance was indicated by defining a score according to the damage occurring on the surface of tomatoes during the harvesting process. If the damage does not occur, it is expressed as O and defined as a 100 score. Among the damages, weak damage on the surface were marked with $\triangle$ and defined as 50 score, and those that gave strong damage such as tearing were expressed as $\times$, and defined as 0 score.

According to experiment result, the tendency to create damage increased as the number of clusters increased. In the case of A, the tomato was alone, so there were no obstacle such as surrounding fruits, which did not cause damages during the harvest process. However case B, C, and D, as the number of tomatoes in the cluster increases, neighboring tomatoes other than the target tomato can be considered as obstacles. Non-target tomatoes can be dynamically swing during the harvesting process. Such swings can cause damage by generating the movement of the target crop, and even if the target crop is harvested without damage, it can cause damage to non-target crops.

To further improve the performance of the robot, the following problems are addressed in the discussion and considered as challenges to be overcome in the future: inaccurate fruit localization due to errors in the detection of fruit;

TABLE 2. For a successful harvest, and time consumed in each step.

	Perception		Manipulation			Harvesting
	Fruit detection	Pose detection	Go to initial	Planning	Reach fruit	
Time(s)	0.2	0.014	0.75	0.94	2.47	1.5
Total cycle time(s)	5.874					

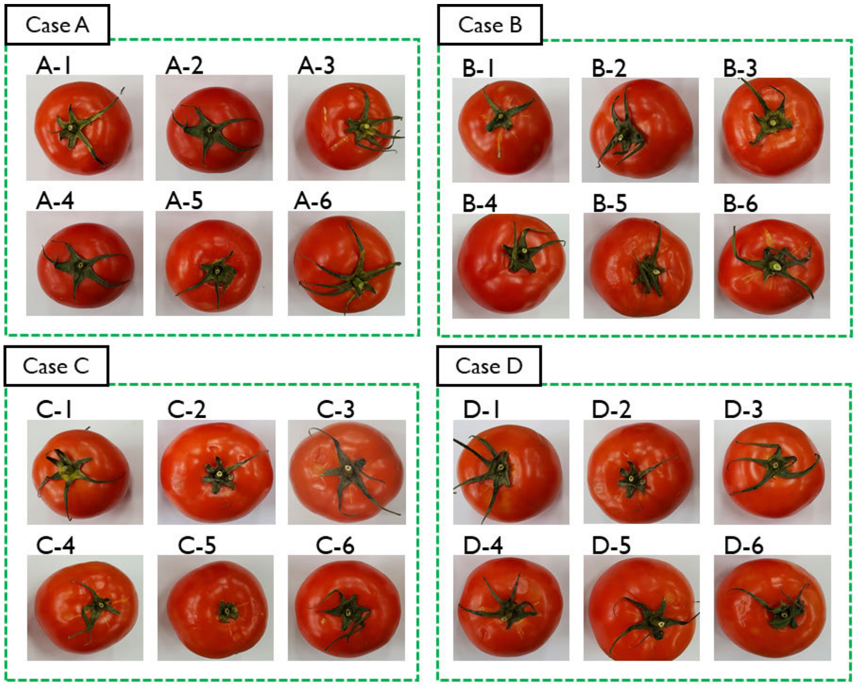

FIGURE 12. Harvesting result according to the number of clusters.

TABLE 3. Harvesting score according to the number of clusters.

	Case A	Case B	Case C	Case D
Harvest score	O	O	O	O
	O	O	×	×
	O	O	×	O
	O	O	O	△
	O	△	△	×
	O	O	O	×
Average	100	91.67	58.33	41.67

incorrect path planning; unsolved kinematics of the manipulator for moving toward detected tomatoes; and obstruction to target tomatoes.

V. DISCUSSION

In this paper, we proposed combination of 3D perception, manipulation, and gripper modules for realizing an efficient tomato harvesting robot. To evaluate the proposed harvesting robot, experiments were conducted in a laboratory testbed environment for each substep. However, the system has a limitation in terms of not being able to apply several variables that exist in actual environments, such as stem localization and obstacles. In the future, we need to address the following challenges:

A. STEM DETECTION

The detection procedure of the proposed system is composed of only one step. This detection procedure only produces the TCP, i.e., the center point of the bounding box generated by the YOLOV3 model. The end- effector then moves to the target point. This method causes the success rate to be dependent on the size and pose of the detected tomato. To successfully harvest the fruit, it must be cut precisely, and therefore stem detection is essential. In addition, the manipulator is continually moved following the detected pose of the tomato stem. Also working continuous work, the picking and cutting action affect the position and pose of the other tomato by vibration and small crash. To effectively follow the changes of target fruit and cut it, visual servoing for the pose control of the robot arms should be explored in future research [29].

B. REINFORCEMENT LEARNING-BASED PATH PLANNING

To control the motion of the manipulator, open source software was used, for the application of Cartesian path planning and solution of complicated kinematics. However, for some specific poses, the manipulator can not solve the problem; therefore, movement is blocked, and an appropriate trajectory is not produced. In order to create an appropriate trajectory, we need to determine the optimal harvesting sequence that can be harvested more easily, and path planning that has changed accordingly is a way to shorten the harvesting time. The planning trajectory is randomly generated, and thus, either surrounding obstacles hinder movement or the the harvesting time is changed depending on the produced trajectory. In future research, to successfully reach the target point, the motion planning algorithm must be improved via reinforcement learning. For the robots to execute reinforcement learning based on earned rewards from several repeated several trials and errors, the manipulator should be able to determine the optimal trajectory by itself [30]. Also the manipulators decide the order to target fruit according to produced optimal trajectory based on reinforcement learning. This technique consumes less time and can enables the robots to avoid obstacles.

C. END-EFFECTOR

We developed a scissor-type cutting module in which a tractional cutting unit was utilized to ensure cutting performance. The proposed cutting module of end-effector has advantage in cluster. When a movement occurs in a crop other than the target, target tomato also occurs movement. The TCU in the cutting module minimized the movement of the stem, making cutting smooth. However, the cutting scissors had a sharp tip, which caused damage to the fruit on approach. Therefore, another cutting mechanism must be developed. The structure of the cutting module should not feature protrusions, and the structure of the surrounding environment should be developed in a form that is relatively less affected by the system.

VI. CONCLUSION

In this paper, we propose an efficient tomato-harvesting robot that combines the principles of 3D perception, manipulation, and an end-effector. With this robot, deep-learning-based detection and 3D perception are performed considering tomatoes as the target. Motion control of the manipulator was implemented based on 3D perception, whereas the developed end-effector comprised two parts: a grasping module and a cutting module. The grasping module grips tomatoes in a cluster and is based on a suction gripper using soft robotics. The suction gripper allows suction pads, which were based on the kirigami pattern, to grip unstructured shapes more easily. The cutting module, which has the shape of scissors, is equipped with a tractional cutting unit to overcome structural limitations and improve cutting. The proposed tomato-harvesting robot was evaluated and verified using a laboratory testbed. Although the proposed harvesting robot did not exhibit high performance, it could be sufficiently improved if further research is conducted, as explained in the discussion.

REFERENCES

[1] V. Bloch, A. Degani, and A. Bechar, "A methodology of orchard architecture design for an optimal harvesting robot," *Biosyst. Eng.*, vol. 166, pp. 126–137, Feb. 2018.

[2] W. Lili, Z. Bo, F. Jinwei, H. Xiaoan, W. Shu, L. Yashuo, Q. Zhou, and W. Chongfeng, "Development of a tomato harvesting robot used in greenhouse," *Int. J. Agricult. Biol. Eng.*, vol. 10, no. 4, pp. 140–149, 2017.

[3] C. W. Bac, E. J. van Henten, J. Hemming, and Y. Edan, "Harvesting robots for high-value crops: State-of-the-art review and challenges ahead," *J. Field Robot.*, vol. 31, no. 6, pp. 888–911, Nov. 2014.

[4] B. Zhang, J. Zhou, Y. Meng, N. Zhang, B. Gu, Z. Yan, and S. I. Idris, "Comparative study of mechanical damage caused by a two-finger tomato gripper with different robotic grasping patterns for harvesting robots," *Biosyst. Eng.*, vol. 171, pp. 245–257, Jul. 2018.

[5] Y. Zhao, L. Gong, Y. Huang, and C. Liu, "A review of key techniques of vision-based control for harvesting robot," *Comput. Electron. Agricult.*, vol. 127, pp. 311–323, Sep. 2016.

[6] S. Hayashi, S. Yamamoto, S. Saito, Y. Ochiai, J. Kamata, M. Kurita, and K. Yamamoto, "Field operation of a movable strawberry-harvesting robot using a travel platform," *Jpn. Agricult. Res. Quart., JARQ*, vol. 48, no. 3, pp. 307–316, 2014.

[7] Y. Xiong, C. Peng, L. Grimstad, P. J. From, and V. Isler, "Development and field evaluation of a strawberry harvesting robot with a cable-driven gripper," *Comput. Electron. Agricult.*, vol. 157, pp. 392–402, Feb. 2019.

[8] S. Hayashi, K. Shigematsu, S. Yamamoto, K. Kobayashi, Y. Kohno, J. Kamata, and M. Kurita, "Evaluation of a strawberry-harvesting robot in a field test," *Biosystems Eng.*, vol. 105, no. 2, pp. 160–171, Feb. 2010.

[9] J. I. Arribas, G. V. Sánchez-Ferrero, G. Ruiz-Ruiz, and J. Gómez-Gil, "Leaf classification in sunflower crops by computer vision and neural networks," *Comput. Electron. Agricult.*, vol. 78, no. 1, pp. 9–18, Aug. 2011.

[10] K. Yamamoto, W. Guo, Y. Yoshioka, and S. Ninomiya, "On plant detection of intact tomato fruits using image analysis and machine learning methods," *Sensors*, vol. 14, no. 7, pp. 12191–12206, 2014.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[12] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 3626–3633.

[13] 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. 1, pp. 1140–1159, Sep. 2017.

[14] B. Arad, J. Balendonck, R. Barth, O. Ben-Shahar, Y. Edan, T. Hellström, J. Hemming, P. Kurtser, O. Ringdahl, T. Tielen, and B. Tuijl, "Development of a sweet pepper harvesting robot," *J. Field Robot.*, vol. 37, no. 6, pp. 1027–1039, Sep. 2020.

[15] H. Kang, H. Zhou, and C. Chen, "Visual perception and modeling for autonomous apple harvesting," *IEEE Access*, vol. 8, pp. 62151–62163, 2020.

[16] C. W. Bac, J. Hemming, B. A. J. van Tuijl, R. Barth, E. Wais, and E. J. van Henten, "Performance evaluation of a harvesting robot for sweet pepper," *J. Field Robot.*, vol. 34, no. 6, pp. 1123–1139, Sep. 2017.

[17] H. Yaguchi, K. Nagahama, T. Hasegawa, and M. Inaba, "Development of an autonomous tomato harvesting robot with rotational plucking gripper," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 652–657.

[18] S. Bachche and K. Oka, "Performance testing of thermal cutting systems for sweet pepper harvesting robot in greenhouse horticulture," *J. Syst. Des. Dyn.*, vol. 7, no. 1, pp. 36–51, 2013.

[19] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Comput. Electron. Agricult.*, vol. 157, pp. 417–426, Feb. 2019.

[20] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: http://arxiv.org/abs/1804.02767

[21] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.

[22] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[23] R. Paul, "Manipulator Cartesian path control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-9, no. 11, pp. 702–711, Nov. 1979.

[24] R. B. Rusu and S. Cousins, "3D is here: Point cloud library (PCL)," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 1–4.

[25] A. Tabb and K. M. Ahmad Yousef, "Solving the robot-world hand-eye(s) calibration problem with iterative methods," *Mach. Vis. Appl.*, vol. 28, nos. 5–6, pp. 569–590, Aug. 2017.

[26] B. Jia, A. Zhu, S. X. Yang, and G. S. Mittal, "Integrated gripper and cutter in a mobile robotic system for harvesting greenhouse products," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2009, pp. 1778–1783.

[27] J. Shintake, V. Cacucciolo, D. Floreano, and H. Shea, "Soft robotic grippers," *Adv. Mater.*, vol. 30, no. 29, Jul. 2018, Art. no. 1707035.
[28] Z. Zhakypov, F. Heremans, A. Billard, and J. Paik, "An origami-inspired reconfigurable suction gripper for picking objects with variable shape and size," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 2894–2901, Oct. 2018.
[29] J. Gamba, P. J. From, and A. C. Leite, "A visual servoing approach for robotic fruit harvesting in the presence of parametric uncertainties," in *Proc. CBA Volumes, 22nd Congresso Brasileiro de Automática*, 2018.
[30] M. Kim, D.-K. Han, J.-H. Park, and J.-S. Kim, "Motion planning of robot manipulators for a smoother path using a twin delayed deep deterministic policy gradient with hindsight experience replay," *Appl. Sci.*, vol. 10, no. 2, p. 575, Jan. 2020.

JONGPYO JUN received the B.S. degree from the Department of Mechanical Engineering, Sunchon National University, in 2008, and the M.S. degree from the Department of Mechanical Engineering, Chonnam National University, in 2015. He is currently pursuing the Ph.D. degree with the Department of Rural and Biosystems Engineering, Chonnam National University. His research interests include mechanical engineering, design, and harvesting robot.

JEONGIN KIM is currently pursuing the integrated B.S.-M.S. degree in biosystems engineering with Chonnam National University, South Korea. Her research interests include manipulator control, reinforcement learning, and perception.

JAEHWI SEOL received the B.S. degree from the Department of Rural and Biosystems Engineering, Chonnam National University, South Korea, in 2020. He is currently pursuing the M.S. degree in biosystems engineering with Chonnam National University. His research interests include field robotics, supervisory control, and discrete event systems.

JEONGEUN KIM (Student Member, IEEE) received the B.S. degree from the Department of Rural and Biosystems Engineering, Chonnam National University, South Korea, in 2019. She is currently pursuing the M.S. degree in biosystems engineering with Chonnam National University. Her research interests include agricultural robotics, field robotics, multi-robot task allocation, and machine learning.

HYOUNG IL SON (Senior Member, IEEE) received the B.S. and M.S. degrees from the Department of Mechanical Engineering, Pusan National University, South Korea, in 1998 and 2000, respectively, and the Ph.D. degree from the Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 2010. He also had several appointments both academia and industry as a Senior Researcher with LG Electronics, Pyungtaek, South Korea, from 2003 to 2005, Samsung Electronics, Cheonan, South Korea, from 2005 to 2009, a Research Associate with the Institute of Industrial Science, The University of Tokyo, Tokyo, Japan, in 2010, and a Research Scientist with the Max Planck Institute for Biological Cybernetics, Tübingen, Germany, from 2010 to 2012. From 2012 to 2015, he lead the Telerobotics Group, Central Research Institute, Samsung Heavy Industries, Daejeon, South Korea, as a Principal Researcher. He joined the Faculty of the Department of Rural and Biosystems Engineering, Chonnam National University, Gwangju, South Korea, in 2015, where he is currently an Associate Professor. His research interests include field robotics, agricultural robotics, haptics, teleoperation, and discrete event and hybrid systems.