ROBOMECH Journal

RESEARCH ARTICLE

Open Access

Fruit recognition method for a harvesting robot with RGB-D cameras

Takeshi Yoshida[1*], Takuya Kawahara[2] and Takanori Fukao[3]

Abstract

In this study, we present a recognition method for a fruit-harvesting robot to automate the harvesting of pears and apples on joint V-shaped trellis. It is necessary to recognize the three-dimensional position of the harvesting target for harvesting by the fruit-harvesting robot to insert its end-effector. However, the RGB-D (red, green, blue and depth) camera on the harvesting robot has a problem in that the point cloud obtained in outdoor environments can be inaccurate. Therefore, in this study, we propose an effective method for the harvesting robot to recognize fruits using not only three-dimensional information obtained from the RGB-D camera but also two-dimensional images and information from the camera. Furthermore, we report a method for determining the ripeness of pears using the information on fruit detection. Through experiments, we confirmed that the proposed method satisfies the accuracy required for a harvesting robot to continuously harvest fruits.

Keywords: Harvesting robot, Object detection, Machine learning

Introduction

There is an urgent need to develop labor-saving technologies for fruit production to cope with the significant decrease in the number of growers and aging of the growers in fruit production in Japan. To solve this problem, it is necessary to develop new work machines and robots that can be used for centralized management, such as harvesting.

Machine development for fruit cultivation needs to be done for each fruit type because the shapes are different from one fruit tree to the other. This is a factor that makes it difficult to develop machines for fruit cultivation. Lined dense planting cultivation is a form of cultivation in which fruit trees are arranged in rows. The development of lined dense planting cultivation enables the development of machinery that can be used commonly for several tree species. Lined dense planting cultivation enables numerous work machines to work with a straight flow that they are skilled at [1]. By making the position where the fruit grows flat, lined dense planting cultivation makes it possible for workers or robots to work with higher efficiency. The purpose of this study is to automate harvesting with a fruit harvesting robot for joint V-shaped trellis, which is a type of lined dense planting cultivation. We have been conducting research on this tree shape using a harvesting robot [2]. Fig. 1 shows an example of the joint V-shaped trellis.

For the harvesting robot to harvest the fruit, it first uses a red, green, blue, and depth (RGB-D) camera to detect the positions of the objects to be harvested in three dimensions, and then obtains the positions that the end-effector will eventually reach. After solving the inverse kinematics of the robot arms for the final reaching position, the harvesting robot inserts the end-effectors into the bottom of the fruit, and then harvests by twisting the aiming fruit. Currently, the RGB-D camera has a problem in that the position of the point cloud it acquires outdoors is inaccurate. Therefore, in this study, we propose an effective fruit recognition method that uses not only 3D information acquired by an RGB-D camera but also 2D images and camera models. In addition, pears need

*Correspondence: yoshidat934@naro.affrc.go.jp

[1] Research Center for Agricultural Robotics, National Agriculture and Food Research Organization, Tsukuba, Japan
Full list of author information is available at the end of the article

Fig. 1 Joint V-shaped trellis

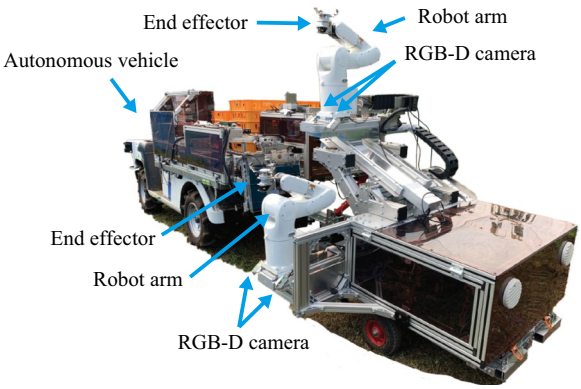


Fig. 2 Components of harvesting robot

Fig. 3 Robot hand for harvesting

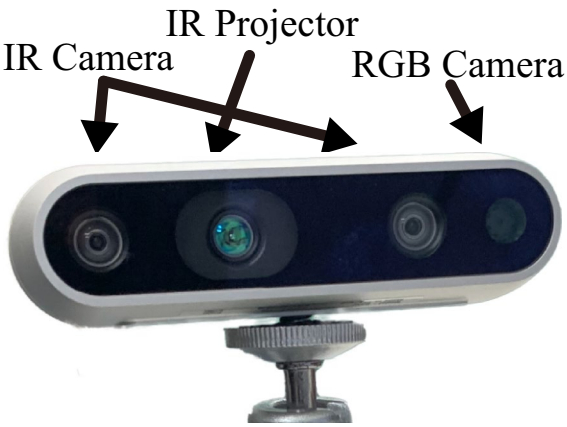


Fig. 4 Intel RealSense D435

to be judged for harvestability based on their ripeness at the time of harvest. In this study, we also report a method to judge whether pears can be harvested or not by using information of fruit recognition.

Harvesting robot

Outline of harvesting robot

Figure 2 shows the components of the harvesting robot used in this study. The robot was designed to reach all fruits meeting the standard using two robot arms and two slide mechanisms. The robot also attempted to speed up the entire harvesting operation by harvesting using two arms simultaneously. The end-effectors at the tips of the robot arms were used to grasp the fruit. Four RGB-D cameras were used to detect the fruits to be harvested and were equipped in different directions to prevent leaves and branches from hiding the fruit. Each robot arm uses data from two cameras mounted on its base. However, the camera data is not integrated, and each camera data is used alternately. The autonomous vehicle at the head of the components moved through the field while towing the fruit-harvesting robot.

Robot hand

The harvesting robot uses a robot hand developed by DENSO Corporation as the end-effector to harvest the fruits (Fig. 3). This end-effector can open and close its fingers and rotate with a single servomotor using a spring and clutch. It also has three silicon fingers to harvest softly. At the time of harvesting, the center of rotation of the hand and the direction of the peduncle of the fruit are aligned, and the peduncle is twisted by rotating the hand. It is necessary to grasp the fruits directly or slightly diagonally below them to twist them effectively.

Camera

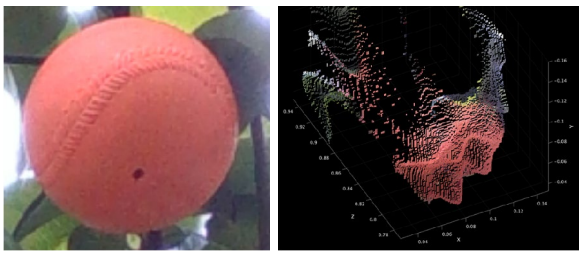
The Intel Realsense D435, shown in Fig. 4 was used to determine the 3D position of the harvest target. D435 is equipped with two infrared cameras, an infrared projector, and an RGB camera. As a method to provide RGB-D information, it uses two infrared cameras as a stereo camera to provide the depth information and an RGB camera to overlay color information onto the depth information. Although it is difficult for normal stereo cameras to provide accurate depth information of areas with few features, D435 uses an infrared projector to improve the accuracy of the depth information by projecting a patterned image info featureless areas. However, there is a problem that the effect of improving accuracy cannot be obtained even if the pattern image is projected because the sunlight cancels most of the infrared patterns in a daytime outdoor environment. Figure 5b shows the point cloud extracted from the entire point cloud taken with D435, with the corresponding region in Fig. 5a. This figure shows that the surface of the point cloud is wavy and inaccurate even if the point cloud of the softball is almost spherical.

Related works

Various studies focus on fruit recognition. These studies can be divided into two categories depending on the sensors they mainly deal with.

The first category mainly deals with 2D images. There are several studies in this category that do not deal with the acquisition of the 3D position of an object. Gao et al. proposed a multi-class apple detection method that considers the use of a harvesting robot. The multi-class apple detection method labels four classes: non-occluded, leaf-occluded, branch/wire-occluded, and fruit-occluded fruits to avoid the robotic end-effector from being damaged by the obstacles [3]. Arad et al. proposed a robot for harvesting sweet pepper fruits in greenhouses [4] [5]. They proposed a Flash-No-Flash controlled illumination acquisition protocol to stabilize the effects of illumination for color-based detection algorithms. Their sweet pepper harvesting robot applies a visual servo that keeps the detected center of the fruit in a predetermined position in the camera image to lower the requirements for camera calibration and 3D coordinates. Yu et al. proposed a method for strawberry fruit target detection based on Mask Region-based Convolutional Neural Network (R-CNN) [6]. In addition, they performed a visual localization method for strawberry picking points by analyzing the shape and edge features of mask images generated from Mask R-CNN. Yu et al. also proposed a localization algorithm to detect the picking point on strawberry stems with Rotational You Only Look Once (R-YOLO), which predicts the rotation of the bounding box of the fruit target [7]. Their harvesting robot measures the distance to the target fruit with a pair of laser beam sensors attached to the head of the fingers of the robot instead of detecting the depth of the target fruit. Fu et al. proposed an algorithm that can detect individual kiwifruits even if they are crowded using two types of lines [8]. The first type is a calyx line that connects together all the calyxes in one cluster. Another type of line is a separating line drawn between two closest contact points between adjacent fruits. Liu et al. proposed a method to detect unevenly red apple fruits that include the green or yellow color with two features [9]. Simple linear iterative clustering (SLIC) is adapted to segment images into superpixel blocks and determine candidate regions with color features. The histogram of oriented gradient (HOG) is adopted to detect fruits in candidate regions and locate the position with the shape feature. Feng et al. proposed an apple fruit recognition algorithm based on multispectral dynamic images [10]. It is based on the fact that the fruit and the leaf can be identified easily because the surface temperature change of the fruit is slower than that of the neighboring leaves. Their proposed algorithm with multi-spectral dynamic images extracts texture features using a gray-level co-occurrence matrix after several pre-processing steps. It then classifies objects by a linearly separable support vector machine. Sa et al. proposed approaches for a vision-based fruit detection system with a field farm dataset, maintaining fast detection and a low burden for ground truth annotation [11]. Their approaches are the novel use of RGB and Near Infra-Red (NIR) multimodal information within early and late fusion networks that provide improvements over a single deep convolutional neural network.

The second category mainly deals with 3D information obtained by RGB-D cameras and stereo cameras. Nguyen et al. proposed an algorithm for the detection and localization of red and bicolored apples on trees in an orchard based on color and range data captured with an RGB-D camera under a light shield blocking



(a) Photo (b) Point cloud

Fig. 5 Example images taken by RealSense D435

direct sunlight [12]. Their algorithm estimates the location and diameter of each fruit by applying the random sample consensus (RANSAC) algorithm to the clustered apple point cloud. Lin et al. proposed a global point cloud descriptor that integrates the shape, angular, and color features of the object of interest [13]. This descriptor is used to distinguish between fruits and non-fruits using a support vector machine. The potential fruits are detected from the clustered point cloud using the M-estimator sample consensus based 3D shape detection algorithm. Lin et al. proposed a vision sensing algorithm that can detect guava fruits on trees and obtain promising 3D pose information with an RGB-D sensor [14]. They applied Euclidean clustering to obtain all of the individual fruits from the fruit point cloud corresponding to segmented fruits on the image and estimated the pose of the fruit relative to its mother branch. Yaguchi et al. proposed a tomato fruit recognition method for a harvesting robot. First, color-based point cloud extraction was applied to a 3D point cloud from a stereo camera. Second, distance-based clustering was applied to separate the candidate point cloud into tomato clusters. Thereafter, the harvesting robot inserts its end-effector into the fruit position, which is decided with sphere fitting using RANSAC. Yoshida et al. proposed a method for detecting cutting points on tomato peduncles using an RGB-D camera mounted on a harvesting robot [15] [16]. In their approach, several types of Region Growing were used to construct a directed acyclic graph. Subsequently, using the Mahalanobis distance defined based on statistical information, they detected appropriate cutting points on the peduncles. Tao et al. proposed an improved 3D descriptor with the fusion of color and 3D geometric features to help a fruit-picking robot's recognition ability [17]. They also proposed a method to automatically recognize apples, branches, and leaves using a support vector machine optimized by a genetic algorithm.

In this study, we combine the properties of both the aforementioned categories in that we effectively use the information in the 2D image against the inaccuracies of the RGB-D camera.

Algorithm for fruit detection to harvest

The harvesting robot needs to be told where to insert its end-effectors to harvest the fruits. However, as explained in the subsection on the camera, the 3D positions acquired outdoors with the RGB-D camera tend to be inaccurate. After recognizing the fruits on the 2D image, the accuracy of the 3D position was improved by fitting the fruit to a sphere in a 3D space. This is based on the assumption that the shape of a fruit is close to a sphere. By adopting the sphere fitting, it is possible to estimate the position where the end-effectors should be inserted, even if the lower part of the fruit is not visible to the RGB-D cameras. In addition, our proposed method simplifies annotation work because the annotation work required for 3D object detection can be performed on a 2D image instead of a point cloud.

Position recognition of fruit on the two-dimensional image

Based on the assumption that the fruit that is the target of harvesting is spherical, the spherical shape of the fruit becomes circular when it is projected onto a 2D image. In this section, after going through the fruit recognition stage, we estimate the circle that can fit the fruit. Our proposed method adopts Mask R-CNN proposed by He et al. as a fruit detection method on an image [18]. The detection accuracy of Mask R-CNN, including the shape of a bounding box, is excellent. We used Detectron 2 [19] as a library providing Mask R-CNN. Fig. 6 shows a result that Mask R-CNN detecting a fruit.

The shape of the circle is obtained by setting the center and the short side of the bounding box as the center and diameter of the circle, respectively. In addition, the point cloud corresponding to the pixel on the binary mask of the fruit is assumed to be the point cloud that constitutes the fruit, and will be used in the next section.

Sphere fitting

The fruit circle obtained in the previous section is a projection of the fruit onto the image plane. From the perspective projection model of the camera, the 3D position of the fruit is assumed to exist in a similar relationship somewhere on the extension of the center of the circle on the image plane from the center of the image, as shown in Fig. 7. Based on this assumption, the location of the fruit

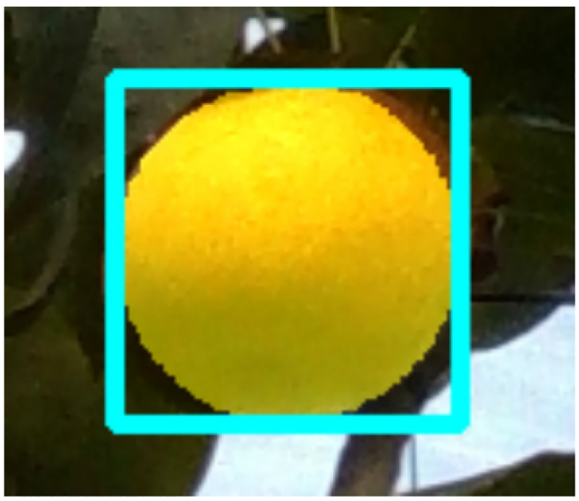


Fig. 6 Example of the detection result of Mask R-CNN

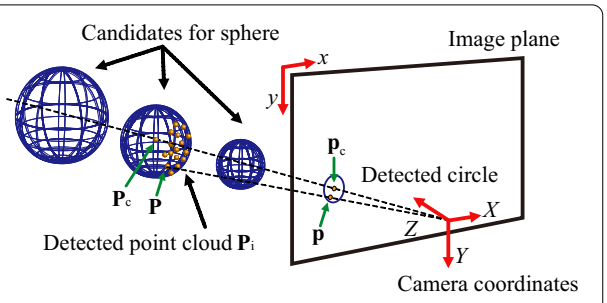


Fig. 7 Candidates for sphere

is identified using the 3D point cloud corresponding to the binary mask obtained by Mask R-CNN as a cue.

The relationship between point $\mathbf{P} = [X\ Y\ Z]^T$, which exists in a 3D space in the camera coordinate, and point $\mathbf{p} = [x\ y]^T$, which is the projection of point $\mathbf{P}$ onto a image plane, can be expressed by Eq. (1), where $\mathbf{K}$ is the intrinsic parameter of the camera and s is the scale factor that makes the third line on the left side 1. In this study, the intrinsic parameter K is obtained by camera calibration.

$$s\begin{bmatrix}\mathbf{p}\\1\end{bmatrix} = \mathbf{KP} \tag{1}$$

By transforming Eq. (1), the relationship between a sphere in a 3D space whose center is $\mathbf{P}_c = [X_c\ Y_c\ Z_c]^T$ and a circle whose center is $\mathbf{p}_c = [x_c\ x_c]^T$, which is a projection of the sphere onto a 2D image, can be expressed by Eq. (2).

$$\mathbf{P}_c = s\mathbf{K}^{-1}\begin{bmatrix}\mathbf{p}_c\\1\end{bmatrix} = s\begin{bmatrix}k_{11} & 0 & k_{13}\\ 0 & k_{22} & k_{23}\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}x_c\\y_c\\1\end{bmatrix} \tag{2}$$

Because the center $\mathbf{p}_c$ of the circle was obtained in the previous section, the center $\mathbf{P}_c$ of the sphere corresponding to the circle is determined only by s when the intrinsic parameter s of the camera is known. Conversely, the relationship between the radius R of the sphere and the radius r of the circle can be expressed by Eq. (3).

$$\mathbf{P}_c + \begin{bmatrix}R\\0\\0\end{bmatrix} = s\mathbf{K}^{-1}\left(\begin{bmatrix}\mathbf{p}_c\\1\end{bmatrix} + \begin{bmatrix}r\\0\\0\end{bmatrix}\right) \tag{3}$$

Eliminating the sphere and circle centers from Eqs. (2) and (3) yields Eq. (4). Because the radius of the circle is known, the radius of the sphere is determined only by s.

$$R = sk_{11}r \tag{4}$$

Substituting Eqs. (2) and (4) into Eq. (5), which is the equation of the sphere, yields Eq. (6).

$$(X - X_c)^2 + (Y - Y_c)^2 + (Z - Z_c)^2 = R^2 \tag{5}$$

$$(X - sS_x)^2 + (Y - sS_y)^2 + (Z - sS_z)^2 = (sS_r)^2 \quad , \tag{6}$$

where each coefficient in Eq. (6) is as follows.

$$S_x = k_{11}x_c + k_{13} \tag{7}$$

$$S_y = k_{22}y_c + k_{23} \tag{8}$$

$$S_z = 1 \tag{9}$$

$$S_r = k_{11}r \tag{10}$$

The objective function of the least squares method to fit a sphere from a 3D point cloud $\mathbf{P}_i = [X_i\ Y_i\ Z_i]^T$ can be expressed by Eq. (11).

$$f(s) = \sum_i e_i^2 \tag{11}$$

where e_i is the following equation.

$$e_i = (X_i - sS_x)^2 + (Y_i - sS_y)^2 + (Z_i - s)^2 - (sS_r)^2 \tag{12}$$

To organize Eq. 12), the coefficients are set as in the following equations.

$$A = S_x^2 + S_y^2 + 1 - S_r^2 \tag{13}$$

$$B_i = S_xX_i + S_yY_i + Z_i \tag{14}$$

$$C_i = X_i^2 + Y_i^2 + Z_i^2 \tag{15}$$

Using these coefficients, Eq. (11) can be transformed into Eq. (16).

$$f(s) = \sum_i \left(As^2 - 2Bs + C\right)^2 \tag{16}$$

The s that minimizes the objective function is the solution to Eq. (17), which is a cubic equations, where RANSAC [20] (Random sample consensus) is used to suppress the effect of a noisy point cloud.

$$\frac{\delta f(s)}{\delta n} = \sum_i \left(A^2s^3 - 3ABs^2 + (AC + 2B^2)s - BC\right) \tag{17}$$

The 3D coordinates and radius of the center of the target sphere are obtained by substituting the obtained s into Eqs. (2) and (4). By assuming that the sphere exists on the extension of the circle detected in the 2D image, instead

of directly fitting the sphere from the point cloud, it is possible to fit the sphere even from an inaccurate point cloud.

Based on the calibration results between the robot arm and the camera, the last step is to transform the coordinates to the robot coordinate system.

Ripeness determination

Extraction of determining region

Kosui and Hosui are pear varieties that are the targets of harvesting in this study. It is important to determine the right time to harvest them because they do not ripen at once and are non-climacteric-type fruits. We propose a method to determine whether harvesting them is possible based on images captured by RGB-D cameras mounted on a harvesting robot. On the other hand, since apples are harvested all at once, ripeness determination is not necessary.

Farmers check the color of the bottom of the pears at harvest time to determine whether they can be harvested because the part of the pear near the bottom depression is less susceptible to discoloration caused by sunburn and the cork layer on the surface of the pear. In this section, we describe how the harvesting robot extracts the color around the bottom of the pear to determine whether it can be harvested or not, similar to the viewpoint used by farmers. By extracting the color near the bottom of the pear, the harvest robot determines ripeness, similar to the viewpoint used by farmers. However, depending on the direction of the fruit, the harvesting robot may observe the entire bottom of the fruit. To solve this problem, we propose a method to obtain information about the entire area around the bottom of the fruit based on the spherical shape obtained in the previous section.

First, to obtain the center of the bottom of the fruit, Faster R-CNN is applied to the region of the bounding box of the fruit obtained with Mask R-CNN as the second step of object detection [21]. Thereafter, based on the sphere information of the fruit obtained in the previous section, the sphere belt corresponding to the region of the bottom of the fruit is obtained.

When the point on the sphere is $\mathbf{X}$, the center of the sphere is $\mathbf{X}_c$, the radius of the sphere is R, and the tangent plane at point $\mathbf{B}$, which corresponds to the peak at the bottom of the fruit, can be expressed by Eq. (18).

$$(\mathbf{B} - \mathbf{X}_c) \cdot (\mathbf{X} - \mathbf{X}_c) = R^2 \tag{18}$$

Here, as shown in Fig. 8, the points $\mathbf{B}_1$ and $\mathbf{B}_2$ from the center of the sphere $\mathbf{X}_c$ to the point $\mathbf{B}$ are represented by equations (19) and (20).

$$\mathbf{B}_1 = n_1(\mathbf{B} - \mathbf{X}_c) + \mathbf{X}_c \tag{19}$$

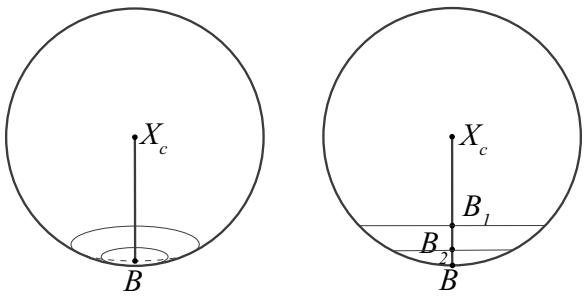


Fig. 8 Sphere belt

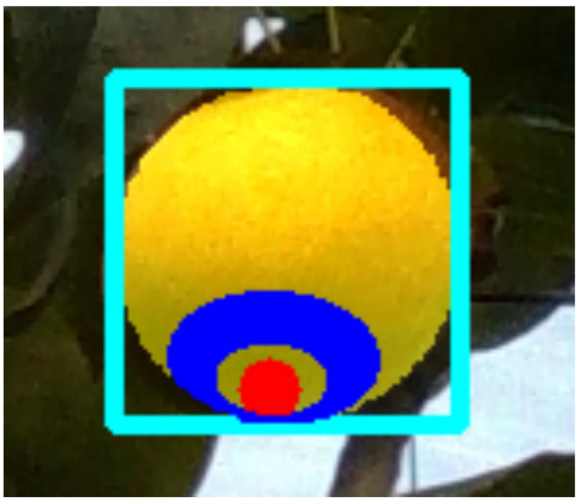


Fig. 9 Detected area for determining ripeness

$$\mathbf{B}_2 = n_2(\mathbf{B} - \mathbf{X}_c) + \mathbf{X}_c \tag{20}$$

The sphere belt of a sphere sandwiched between planes parallel to the tangent plane of the point $\mathbf{B}$ passing through points $\mathbf{B}_1$ and $\mathbf{B}_2$ can be expressed by equations (21) and (22).

$$(\mathbf{B}_1 - \mathbf{X}_c) \cdot (\mathbf{X} - \mathbf{B}_1) > 0 \tag{21}$$

$$(\mathbf{B}_2 - \mathbf{X}_c) \cdot (\mathbf{X} - \mathbf{B}_2) < 0 \tag{22}$$

By reprojecting the points that satisfy equations (21) and (22) from the point cloud that composes the sphere used in the previous section onto the original image, we can extract the region for maturity judgment, as shown in Fig. 9. In this study, n_1 and n_2 were set to 0.8 and 0.95, respectively.

Fig. 10 Fruit sorting system

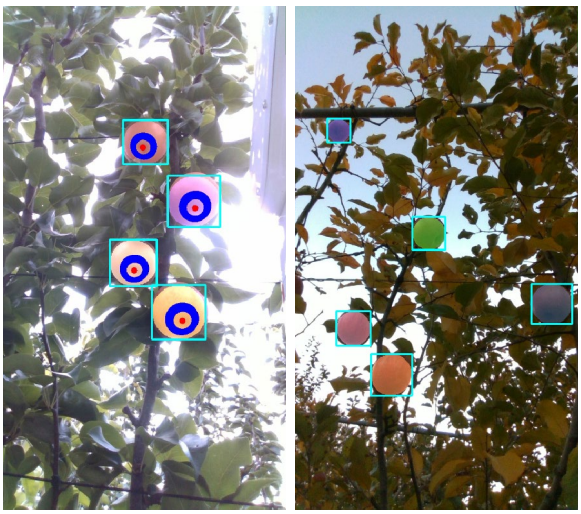
Learning of ripeness determination

We constructed a convolutional neural network (CNN) for regression to determine the ripeness of pears from the RGB data of the sphere belt shown in the previous section. Continuous ripeness data used for training the network were obtained using a fruit sorting system (Fig. 10), which has an internal quality sensor manufactured by Sibuya Seiki Co. On the other hand, the ripeness data obtained by the sorting machine cannot be used as is. In the field, the line between harvestability and non-harvestability is drawn each year based on the ripeness data. To judge whether or not to harvest from the continuous data of ripeness, a classifier using logistic regression was constructed through questionnaires provided to skilled pear cultivators, referring to the eyeballing meeting conducted by pear farmers in the field. The input data used to train the classifier of logistic regression is the ripeness value, and the output data is the harvestability decision. By passing the RGB-D data of the sphere belt of the bottom of the pear through this two step classifier, the harvesting robot can now determine whether the pear can be harvested or not.

Results and Discussion

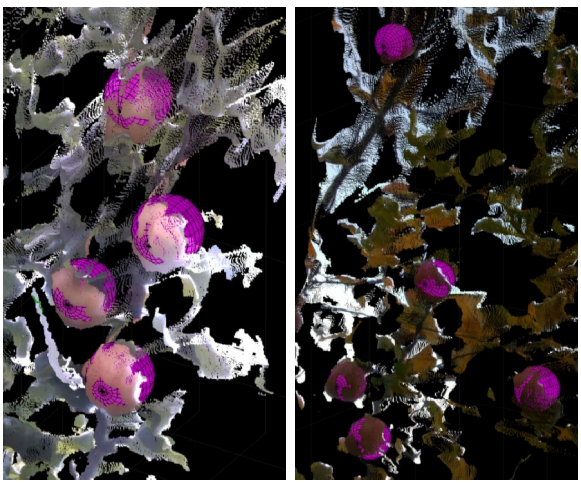
We performed experiments to confirm the effectiveness of our proposed method for the fruit-harvesting robot at Kanagawa Agricultural Technology Center (Pear) and Miyagi Prefectural Institute of Agriculture and Horticulture (Apple). Fig. 11 shows examples of the recognition results on the 2D images. Figure 11a shows that detection of the bottom of the fruits and extraction region to determine their ripeness are applied to the pears. However, the detection of the bottom was not applied to apples because apples are harvested together simultaneously and therefore do not need the ripeness determination.

Figure 12a shows a recognition result in a 3D space based on the detection result of Fig. 11a. In particular,



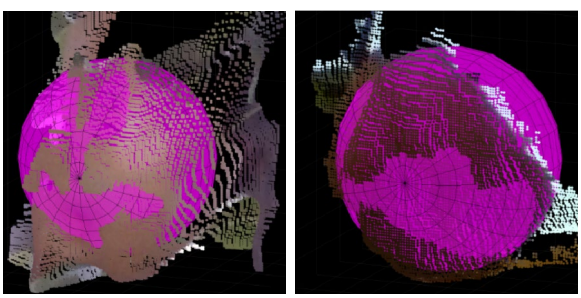
(a) Pear (b) Apple

Fig. 11 Recognition results on the 2D images



(a) Pear (b) Apple

Fig. 12 Results of sphere shape estimation



(a) Pear (b) Apple

Fig. 13 Details of sphere shape estimation

Fig. 14 Example of easy confirmation of harvestability

Fig. 15 Placement of the laser measure to measure the distance between the fruits and camera

as shown in Fig 13a, the original point cloud is highly distorted for the fruit in the highest position. However, the pink sphere estimated by the proposed method is roughly applied to the roundness of the fruit, and it can be observed that the error can be suppressed. Similarly, based on the detection results shown in Fig. 12b, the fruit recognition results in the 3D space are shown in Fig. 13b. Similar to Fig. 13a, in this figure, the pink sphere is fitted to the fruit of the point cloud, and it can be seen that the proposed method does not depend on the type of fruit.

The harvestability of pears was verified by extracting the color around the bottom of 137 fruits using the method described in the previous section. The percentage of correct answers was 87% in comparison to the correct responses provided by skilled workers. Collecting data to determine whether to harvest is labor-intensive compared to fruit recognition; therefore, to improve the accuracy of the data, it is necessary to consider a system that not only increases the amount of data, but also allows for efficient data collection. An example of easy confirmation of harvestability is shown in Fig. 14. The upper fruit is not the fruit at harvest time, but the lower fruit is.

Next, to verify the fruit position using our proposed method, an orange softball simulating a fruit was attached to a tree in the field while changing the location, and the actual measurements by the laser measure shown in Fig. 15 were compared with the estimation results obtained using the proposed method. The distance from the center of the lens, which is the starting point of the laser measure placed as shown in Fig. 15, to the ball marked to be detected at the bottom of the fruit was measured. The softball was used as a target

(a) Case 1 (b) Case 2

Fig. 16 Target fruits for distance measurement

Table 1 Detection results for different peduncle lengths

	Measured value	Estimated value	Error
Case 1	0.913 [m]	0.919 [m]	− 0.006 [m]
Case 2	0.992 [m]	0.961 [m]	0.032 [m]

because the actual fruits were distorted and there was a depression at the bottom of the fruits, making it impossible to measure the exact points.

Table 1 shows the comparison between the proposed method and the actual measurements for the two fruits shown in Fig. 16. It can be observed that each error is small enough for the robot hand palm size.

(c) Scene 1

(d) Scene 2

(e) Scene 3

Fig. 17 Example of continuous harvesting

Fig. 17 shows an example of a harvesting robot moving while harvesting. In this case, 23 out of 25 fruits were harvested. Conversely, the cause of the harvest failure was occlusion by leaves and branches. Excluding state transitions, the average time taken for recognition and harvesting was about 24 seconds. Jetson AGX Xavier from NVIDIA was used for the calculations.

Conclusions

In this study, we proposed a method for estimating the position of fruits in a 3D space such that the fruit harvesting robot cloud perform automatic harvesting. Even when using data from RGB-D cameras where the acquired point cloud is inaccurate owing to its use in an outdoor environment, the proposed method could suppress the inaccuracy of the point cloud using not only the 3D information of RGB-D cameras but also the 2D image information and information about the cameras obtained simultaneously. In addition, by using the 3D information of the fruits obtained in this process, the ripeness of the fruits was also determined. In the experiment, recognition was performed on an actual joint V-shaped trellis, and the effect was confirmed.

Abbreviations

RGB-D: Red, green, blue, and depth; R-CNN: Region-based convolutional neural network; R-YOLO: Rotational you only look once; SLIC: Simple linear iterative clustering; HOG: Histogram of oriented gradient; NIR: Near infraRed; RANSAC: Random sample consensus; CNN: Convolutional neural network.

Acknowledgements

We would like to express our gratitude to Sibuya Seiki Co., Ltd. and other related parties for their cooperation in collecting maturity data.

Author contributions

TY conducted all of the research and experiments. TK and TF conducted a research concept, participated in design adjustment, and drafted a paper draft assistant. All authors read and approved the final manuscript.

Funding

This research was supported by grants from the Project of the Bio-oriented Technology Research Advancement Institution, NARO (the research project for the future agricultural production utilizing artificial intelligence).

Availability of data and materials

Not applicable

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Research Center for Agricultural Robotics, National Agriculture and Food Research Organization, Tsukuba, Japan. [2]Graduate School of Science and Engineering, Ritsumeikan University, Kusatsu, Japan. [3]Graduate School of Information Science and Technology, University of Tokyo, Bunkyo, Japan.

Received: 14 January 2022 Accepted: 19 May 2022
Published online: 28 May 2022

References

1. Kusaba S (2017) Integration of the tree form and machinery. Farm Mechanization 3189:5–9 (**In Japanese**)
2. Onishi Y, Yoshida T, Kurita H, Fukao T, Arihara H, Iwai A (2019) An automated fruit harvesting robot by using deep learning. ROBOMECH J. https://doi.org/10.1186/s40648-019-0141-2
3. Gao F, Fu L, Zhang X, Majeed Y, Li R, Karkee M, Zhang Q (2020) Multi-class fruit-on-plant detection for apple in snap system using faster R-CNN. Comput Electron Agric 176:105634
4. Arad B, Kurtser P, Barnea E, Harel B, Edan Y, Ben-Shahar O (2019) Controlled lighting and illumination-independent target detection for real-time cost-efficient applications. The case study of sweet pepper robotic harvesting. Sensors. https://doi.org/10.3390/s19061390
5. Arad B, Balendonck J, Barth R, Ben-Shahar O, Edan Y, Hellström T, Hemming J, Kurtser P, Ringdahl O, Tielen T, van Tuijl B (2020) Development of a sweet pepper harvesting robot. J Field Robot 37(6):1027–1039

6. Yu Y, Zhang K, Yang L, Zhang D (2019) Fruit detection for strawberry harvesting robot in non-structural environment based on mask-rcnn. Comput Electron Agric 163:104846
7. Yu Y, Zhang K, Liu H, Yang L, Zhang D (2020) Real-time visual localization of the picking points for a ridge-planting strawberry harvesting robot. IEEE Access 8:116556–116568
8. Fu L, Tola E, Al-Mallahi A, Li R, Cui Y (2019) A novel image processing algorithm to separate linearly clustered kiwifruits. Biosyst Eng 183:184–195
9. Liu X, Zhao D, Jia W, Ji W, Sun Y (2019) A detection method for apple fruits based on color and shape features. IEEE Access 7:67923–67933. https://doi.org/10.1109/ACCESS.2019.2918313
10. Feng J, Zeng L, He L (2019) Apple fruit recognition algorithm based on multi-spectral dynamic image analysis. Sensors 19(4):949
11. Sa I, Ge Z, Dayoub F, Upcroft B, Perez T, McCool C (2016) Deepfruits: a fruit detection system using deep neural networks. Sensors. https://doi.org/10.3390/s16081222
12. Nguyen TT, Vandevoorde K, Wouters N, Kayacan E, De Baerdemaeker JG, Saeys W (2016) Detection of red and bicoloured apples on tree with an RGB-D camera. Biosyst Eng 146:33–44
13. Lin G, Tang Y, Zou X, Xiong J, Fang Y (2020) Color-, depth-, and shape-based 3d fruit detection. Precision Agric 21:1–17
14. Lin G, Tang Y, Zou X, Xiong J, Li J (2019) Guava detection and pose estimation using a low-cost RGB-D sensor in the field. Sensors. https://doi.org/10.3390/s19020428
15. Yoshida T, Fukao T, Hasegawa T (2018) Fast detection of tomato peduncle using point cloud with a harvesting robot. J Robot Mechatron 30(2):180–186
16. Yoshida T, Fukao T, Hasegawa T (2020) Cutting point detection using a robot with point clouds for tomato harvesting. J Robot Mechatron 32(2):437–444
17. Tao Y, Zhou J (2017) Automatic apple recognition based on the fusion of color and 3d feature for robotic fruit picking. Comput Electron Agric 142:388–396
18. He K, Gkioxari G, Dollár P, Girshick R (2020) Mask R-CNN. IEEE Trans Pattern Anal Mach Intell 42(2):386–397
19. Wu Y, Kirillov A, Massa F, Lo W-Y, Girshick R (2019) Detectron2. https://github.com/facebookresearch/detectron2. Accessed 14 Jan 2022
20. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24(6):381–395
21. Ren S, He K, Girshick R, Sun J (2017) Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149

Publisher's Note