2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)
March 29–31, 2018, Xiamen, China

Design and Implementation of an Automatic Peach-Harvesting Robot System

Yongjia Yu and Zengpeng Sun
State Key Laboratory of Management and Control for Complex Systems Institute of Automation
Chinese Academy of Sciences, Beijing, 100109, China
University of Chinese
Academy of Sciences, Beijing, 100049, China
Email: yuyongjia2015@ia.ac.cn;
Email: sunzengpeng2016@ia.ac.cn

Xiaoguang Zhao, Jiang Bian, and Xiaolong Hui
State Key Laboratory of Management and Control for Complex Systems Institute of Automation
Chinese Academy of Sciences, Beijing, 100109, China
University of Chinese
Academy of Sciences, Beijing, 100049, China
Email: xiaoguang.zhao@ia.ac.cn;
Email: bianjiang2015@ia.ac.cn;
Email: huixiaolong2015@ia.ac.cn;

***Abstract*—In this paper, we present a novel autonomous peach-harvesting approach. The robot can autonomously move around the tree and detect the peaches accurately by a well-designed fully connected neural network, adapting to different illumination. Next, the robot generates a score for each peach, taking into account the confidence of the classification and the size of the bounding box, and then pick one of the most suitable peach. In order to track the peach more robustly and quickly, we choose the KCF tracking algorithm. Finally, the robot picks the peach through a 6-DOF manipulator. In addition, we propose a monocular high-precision ranging method based on structured light. The cost of our method is lower than the traditional ranging method, which enables the robot to obtain the accurate spatial position of the peach.**

Keywords—robot; harvesting; peach; manipulator; detection; measurement; segmentation

I. INTRODUCTION

From 2001 to 2016, the scale of Chinese peach industry grew rapidly. The peach cultivation area and yield ranked first in the world.In 2013, Chinese total yield of peaches was about 12 million tons, accounting for 61.59% of the total output of peach in the world. In China, peaches are harvested carefully by human labor. Some peach trees are tall, and so the harvesting work must be conducted using pairs of steps. This makes harvesting dangerous and inefficient. Moreover, with the decrease of rural population, there will be a rise in labor costs and labor shortage; In recent decades, many advanced laboratories in the world have begun to study the automatic harvesting of peaches by intelligent robots, and it's shown that it could increase the efficiency of production, and improve the peach quality.Therefore, it's essential to devote effort into researching robotic harvesting.

Research on peach-harvesting robots has already been conducted in 1980s [1]. Later, Monta et al. (1998) [2], Arima and Kondo (1999) [3], Hayashi et al. (2001) [4], Wei et al(2013) [5], and Christopher el al(2016) [6] reported research prototypes of harvesting robots for tomatoes, cucumbers, eggplants, apples, and sweet-peppers grown in greenhouses. However, autonomous harvesting robots have not yet been commercially applied in horticultural practice. Actually, the successful harvesting robots in introduction haven't yet satisfy the price and performance of requirements in horticultural practice. The first challenge for peach harvesting robot is to build a robot vision system that is as sensitive as the human eye to identify and locate the peach. There are 3 types of sensing solutions in the above mentioned papers:

1) A 3-D vision system that has two laser diodes was manufactured [7]. One of them emits a red beam and the other emits an infrared beam. Position sensitive devices (PSDs) are used to detect the reflected light. The 3-D shape of the object was measured by laser beam and the red peaches are distinguished from other objects by the difference in the spectral-reflection characteristics between the red and infrared laser beams.

2) Multi-spectral imaging is widely used in robot system [8]. We can use the spectral information presenting in both visible and near-infrared wavelengths. Absorption rates along spectral bands are unique for different objects. So we can select different wavelengths to distinguish plant parts.

3) Use industrial camera to capture RGB color images, and then identify and locate objects in the picture [9].

The advantages of the first two methods are that they can avoid sunlight interference and have less noises. However the cost is too high and the recognition accuracy is low. The strength of the third method is lower sensor cost, which brings about wider range of applications in production.

We will face two challenges when we build this robotic system. The first challenge is how to identify the peach in the picture. There are many algorithms for recognition, such as classical color segmentation and contour recognition [10]. These methods run faster but are more easily affected by illumination.In addition, the use of deep learning Faster Region-based Convolutional Neural Network (Faster R-CNN)

978-1-5386-4362-4/18/$31.00 © 2018 IEEE

for target detection [11], although the recognition rate is improved, but the need for more powerful computers and GPU. Another challenge is how to control the manipulator to pick peach.Commonly, a robot equipped with a multi-DOF manipulator is widely adopted. A vacuum is used to suck the peach into the sucking pipe of the end-effector [12] or a gripper on the end-effector is utilized to pick peach [13].

In this paper, we propose a peach recognition method based on color region proposal, and then use a well-designed fully connected neural network to classify the picture into 2 categories. We find a tradeoff between speed and accuracy. Not only does the recognition accuracy is ensured, but also the requirements of hardware performance are not high.Finally, we apply a 6-DOF manipulator with a gripper to robotic system and propose a low-cost method based on structured light to measure the distance to peach which greatly improve the measurement precision.

The rest of the paper is organized as follows. The hardware and system architecture of robot platform is introduced in Section II. Proposed methods for peach detection and measurement are described in detail in Section III. Section IV will focus on the detail process of our robot system. Relevant experiments and results are further provided in Section V. Finally, in Section VI, we conclude the paper with an outline of future work.

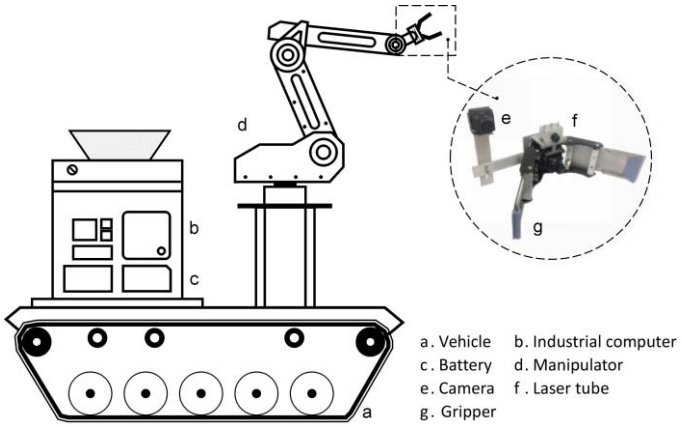

Fig. 1. Peach-harvesting robot platform: a tracked vehicle, an industrial computer, a manipulator, 3 cameras, a gripper and some other sensor.

II. Hardware Platform And System Architecture

The harvesting robot mainly consists of the following parts (shown in Fig. 1): a tracked vehicle as a mobile platform. It is equipped with a power supply system, including voltage conversion module. Battery duration is about 6 hours; We use a low power consumption computer, equipped with an Intel Atom Processor and without GPU. Considering the link type manipulator can move to more positions in the three-dimensional space and can be more flexible to avoid obstacles, we use a 6DOF Z20 manipulator; In order to adapt to the different heights of the peach tree, the manipulator is equipped with a lifting platform below. We have installed 3 cameras on the robot with resolution 480×640 pixels. Two of them are assembled on the end-effector which is used for peach identification and location. The other one is arranged on the body to avoid obstacles; We placed a red laser transmitter in the end-effector, for measuring the distance of peach by an auxiliary camera; In addition, the end effector is equipped with a gripper for peach picking.

III. Image Process Algorithms

The robot vision system consists of a CCD camera and an industrial computer equipped with an Intel Atom processor. The CCD camera can be used to capture the raw images of peach. We programmed at QT5.8 (with Opencv2.4.9) development platform, the corresponding code will be announced at GitHub in the near future.

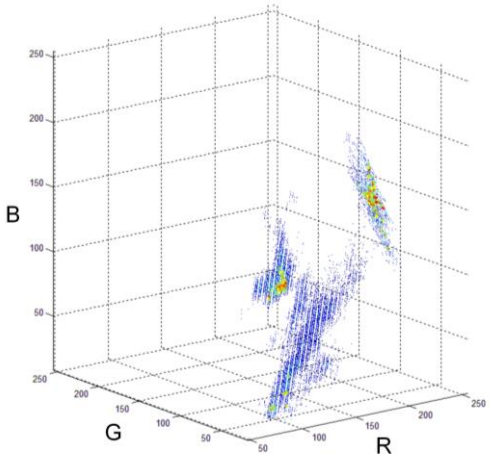

Fig. 2. The distribution of peach color vectors in RGB space.These scatter points in the graph represent the distribution density of the color vectors.

A. Peach Image Segmentation

The object detection algorithm can be roughly divided into two categories. The two stage method (first generation region proposal and then classification) and one stage method (directly get the location and types of bounding box(bbox)),we choose the former. The former object detection algorithms including traditional methods (such as HOG+SVM [14]) and deep learning methods (such as Faster R-CNN, SSD[15], YOLO[16]), the traditional methods mostly depend on the sliding windows, limited to low efficiency and high time complexity. The latter needs to rely on the support of GPU; So we put forward an efficient and accurate method of regional proposal according to the peach color.

The peach color is very distinctive, in contrast with other colors in the background, so this is a powerful feature for pixel level segmentation; First, we collected more than 300 of photos in different illumination environment. Then we made a statistical analysis of every pixel of peach in RGB color space. In order to prevent the interference caused by outliers, we only counted the pixel (RGB color vector) whose probability is greater than a threshold value. The process result is shown in Fig. 2(b).

The above statistical colors are labeled as positive samples and the rest of the colors are as negative samples, we use a fully connected neural network to classify the color. When a

positive sample is input, the neural network outputs 1, and when a negative sample is input, the neural network outputs 0. It can be concisely expressed as Eq. (1).

$$Net(\mathbf{c}) = \begin{cases} 1 & \mathbf{c} \in P \\ 0 & \mathbf{c} \in N \end{cases} \tag{1}$$

where $\boldsymbol{c}$ is the color vector, P represents positive samples, and N represents negative samples.

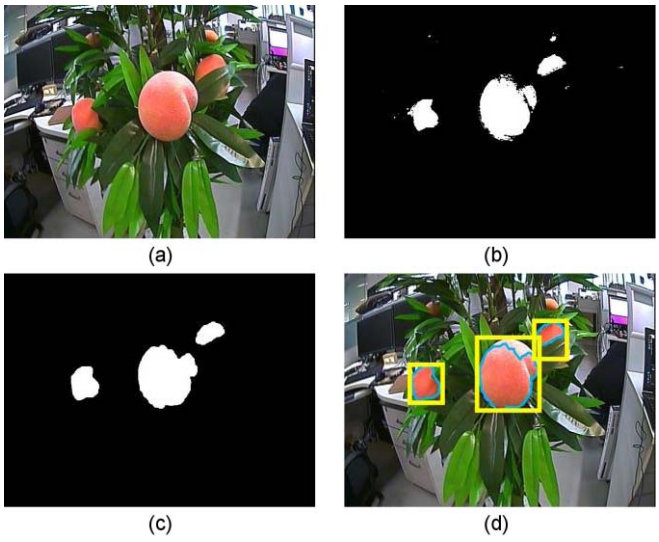

Fig. 3. (a) is the raw image, (b) is the color segmentation mask of raw image, (c) is the segmentation mask after morphology operation, the yellow bounding boxes on the (d) represent the ROIs.

B. ROI Generation

Due to illumination changes and camera noises, the color of objects in the background are similar to those in the peach, resulting in that the many pixels in the image are misclassified and many spots appear in the segmentation mask. In order to solve this problem, Erosion and dilation operations are applied to process the segmentation mask of image. It can be concisely expressed Eq. (2).

$$IMG_1 = ((IMG_0 \Theta E_{r1}) \oplus E_{r2}) \Theta E_{r3} \tag{2}$$

The IMG_0 represents the mask before processing. The IMG_1 represents the processed image. E_r represents the element of the morphological processing. r represents the radius(size) of the kernel element. Θ represents the erosion operation. $\oplus$ represents the dilation operation. The process result is show in Fig. 2(c).

Next, it's necessary to generate a series of region of Interests (ROIs) from segmentation masks to the neural network for classification. The iterative methods could help us to find all the independent connected domain from the segmentation mask. The method of finding the connected domain can be expressed by the formula:

$$X_k = (X_{k-1} \oplus \mathrm{E}) \cap \mathrm{A} \qquad \mathrm{k} = 1,2,3,... \tag{3}$$

where X_k is the connected domain obtained by the k-th iteration, B is a suitable structuring element, A represents a set of the same size as the image size. After several iterations, if $X_k = X_{k-1}$, then the algorithm converges. The process of ROI generation illustrated in Fig. 2.

C. Neural Network Classification

After region proposal it will generate a series of ROIs from the image, and the next step is to classify these ROIs. We adopt a well-designed fully connected neural network. In addition, we will compare this algorithm with the traditional fully connected neural network. It will be described in section 5.

First of all, the scope of the ROI is enlarged to 1.1 and 1.5 times, with more background information. Because the peach is surrounded by leave and the background is relatively simple without enlargement. If ROIs can be combined with the appropriate background information, the samples will effectively improve the robustness of object recognition;

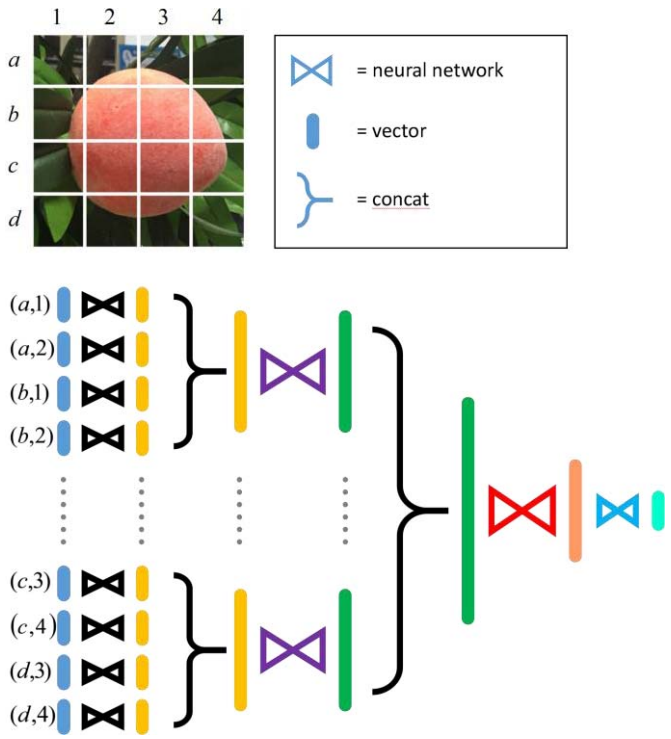

Fig. 4. The ROI is resized to 32×32 pixels, and divided into 16 blocks, then flatten it to a short vector. Each short vector will be connected to a neural network which shares weights.

The structure of our neural network are shown in Fig. 4. First, all the ROIs are resized to 32 pixels×32 pixels size. Next, we cut the image into 16 square blocks, the size of one block is 8 pixels×8 pixels. These blocks are then encoded by a single-layer neural network sharing the same weight to generate 16 vectors with length 36. The adjacent four vectors are combined into a long vector. The long vector is represented by a larger perception field in the image. Finally, the long vectors are encoded and classified by neurons in the back layer. The neural networks in the

same layer sharing the same weights, and have the same size of perception field, as shown in Figure 4, the same color of the neurons are the same weight sharing and have the same size of the perception field.

The classification loss function of the neural network is the cross entropy loss function, the formula is as follows:

$$L = -\frac{1}{N}\sum_{x}\left[y \ln a + (1-y)\ln(1-a)\right] \tag{4}$$

where n is the total number of items of training data, the sum is over all training inputs a, and y are the corresponding desired outputs.Compared with the traditional neural network, our 'semi' neural network architecture is similar in accuracy, but the parameters are reduced by 80% and the consumption time is reduced by almost 50%.

D. Peach Selection

Because there may be a lot of peach detected in an image, it's necessary to choose the most suitable one to pick; We have two principles below to determine whether the peach is suitable for picking.

1) Distance from the robot is close. The robot can easily pick the peach as close as possible to the center of the image.

2) Peach classification confidence is relatively high. To prevent the robot picking other things., we will give each peach in the image a score, as follows:

$$score = \frac{W_{ROI} \cdot H_{ROI}}{k} \cdot e^{C} \tag{5}$$

where W is the width of the ROI, H is the height of the ROI, C is the confidence of the peach, and k is a constant parameter.

The coordinates of the peach in the image are $x_i=(x_i,y_i)$, and the centroid of all peaches can be expressed as:

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=0}^{N} score_i \cdot \mathbf{x}_i \tag{6}$$

The closest peach from the centroid is the most suitable one to pick, whose distance can be expressed as:

$$\arg\min_{i}\left(\mathbf{x}_i - \bar{\mathbf{x}}\right) \tag{7}$$

E. Peach Distance Measurement

In order to pick the peaches more accurately on the peach trees, the robot need to measure the distance of the peaches. Especially when the distance from gripper to peach is very close, the precise measurement of peach is fairly meaningful. It can greatly improve the success rate of picking. At present, the common distance measurement method is

1) Laser ranging method. The disadvantage is that the laser ranging equipment cost is high;

2) Ultrasonic distance measurement. Due to the peach is spherical and there are many leaves around, so this method is not suitable;

3) Binocular range method, but this method is not suitable when the distance is close;

So we propose a high-precision ranging method based on structured light. Because the above three methods have disadvantages, so we propose a high-precision measurement method based on structured light.

As shown in the Fig. 5, the laser device in the center of the gripper will emit a vertical laser light on the peach and then the wide-angle camera captures an image from side to detect the position of the laser in the image. Assuming that the pixel coordinates of the laser in the image are (x_i,y_i) i=1,2,...,N, the coordinates of the peach relative to the center of the gripper can be expressed as:

$$(u,v) = \left(\frac{1}{N}\sum_{i=1}^{N} f_h(x_i,y_i), \frac{1}{N}\sum_{i=1}^{N} f_d(x_i,y_i)\right) \tag{8}$$

where f_u, f_v are the mapping functions from the image pixel position to the spatial position. u and v represent the spatial position. x and y represent the image pixel position.

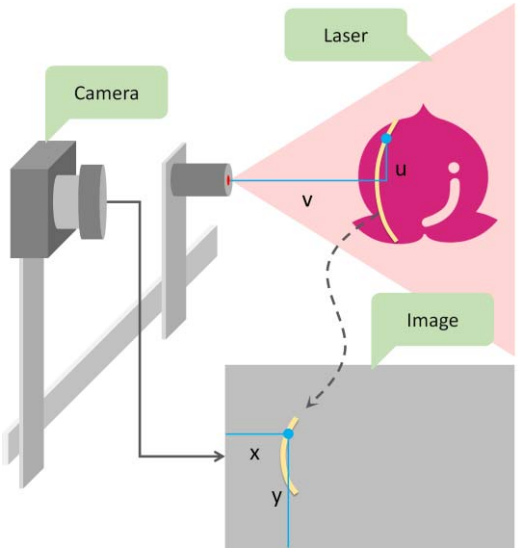

Fig. 5. Laser tube in the gripper emits a laser beam vertically on the peach, and the camera on the side capture a picture. Next, the robot will calculate the distance of peach based on the position of the laser in the picture.

We need to make a calibration plate marked with distance scale. Then we place the laser tube horizontally and keep the position of camera fixed. Laser tube in the gripper emits a laser beam vertically on the calibration plate. Next, we move the calibration plate to change its distance to camera and capture a few pictures of the calibration plate. After that, the pixel coordinates of laser in pictures and the corresponding spatial coordinates are manually marked. Suppose we have a total of K group training samples $(x_i,y_i)\xrightarrow{f}(u_i,v_i)$ i=1,2,...,K. We assume that the mapping function is a cubic polynomial function. So the mapping function can be expressed as:

$$\begin{cases} u = f_u(x,y) = \boldsymbol{\alpha}^T \mathbf{x} \\ v = f_v(x,y) = \boldsymbol{\beta}^T \mathbf{x} \end{cases} \tag{8}$$

Where $\mathbf{x}$, α , β denote as:

$$\begin{cases} \mathbf{x} = \left(1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3\right)^T \\ \boldsymbol{\alpha} = \left(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9, \alpha_{10}\right)^T \\ \boldsymbol{\beta} = \left(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}\right)^T \end{cases} \tag{9}$$

In order to reduce the structural risk and the over-fitting problem, we add the L2 regularization factor to the loss function. The loss function is expressed as follows:

$$\begin{aligned} L_u &= \sum_{i=0}^{K} \boldsymbol{\alpha}^T \mathbf{x}_i + \left\| \boldsymbol{\alpha}^T \right\|^2 \\ L_v &= \sum_{i=0}^{K} \boldsymbol{\beta}^T \mathbf{x}_i + \left\| \boldsymbol{\beta}^T \right\|^2 \end{aligned} \tag{10}$$

IV. The Detail Process

Fig. 6 shows the flow chart of the harvest experiment. First, the robot captures the picture through the camera in the gripper, then it detects the peach in the image. Next, the robot will choose a peach which is most suitable for picking according to the centroid method mentioned in 3.1.3. Then the KCF algorithm is used to track the target. Next, robot can calculate the offset distance between bounding box and the image center, and then use the PID algorithm to control the movement of manipulator to the peach. If the size of bounding box is greater than a certain threshold, the robot can measure precisely the distance to peach by the side camera and laser tube. Finally, the robot grabs the peach by manipulator.

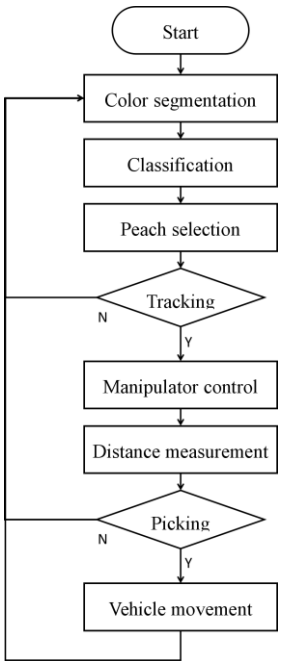

Fig. 6. Flowchart of proposed approach. The whole process can be divided into 3 parts. Part I is target detection and tracking. Part II is peach distance measurement. Part III is manipulator and vehicle movement control.

V. Experiment And Result

A. Detection Experiment

Experiments were conducted on the image dataset to test the performance of the proposed method. Our dataset has a total of 59,136 pictures, of which 29238 are positive samples, 29898 are negative samples. The dataset are collected in different illumination environment and various background. The size of the picture is resize to 32 × 32 pixels.

Here, we compared our method with some other methods (such as HOG-SVM, PCA-SVM and the traditional fully connected neural network) on the same platform. The results of the experiment are shown in Table I. It can be seen that our method has a high accuracy rate and has a 92.8% reduction in neural network parameters compared to traditional fully connected neural networks. The amount of calculation of forward propagation is reduced by almost 73.5% and run time is also reduced by nearly 60%.

TABLE I. The Comparison of Different Detection Algorithms

	OURS	***HOG-SVM***	***PCA-SVM***	***FC***
The average of recognition rate(%)	90.58%	87.76%	83.5%	90.71%
Running time (ms)	71ms	160ms	120ms	110ms
Params	5.6×10^5	-	-	3.2×10^4

B. Measurement Experiment

When the distance from gripper to peach is between 15 to 35 cm, we can use the side camera to accurately measure the distance. We conducted a total of 30 experiments. As shown in Table II, the measuring error range of experimental results are about ±5mm, and such an error in our task can be accept.

TABLE II. Distance Measurement Result and Measuring Error

Ground Truth(mm)	150	180	220	250	290	320	350
measuring result(mm)	155.2	183.1	221.8	254.2	288.1	317.5	345.3
Error(mm)	+5.2	+3.1	+1.8	+4.2	-1.9	-2.5	-4.7

C. Peach Harvesting Experiment

We have done a lot of experiments indoors and outdoors, and achieved the satisfactory experimental results, the robot has been able to successfully harvest the peach, but sometimes is affected by the sunlight, especially in low illumination environment in which the color of peach becomes very deep. It’s difficult to distinguish peaches from the color of the surrounding objects, which will lead to high probability of missed detection. We took a set of photos of robot in the experiment, and it is shown in the Fig. 7.

Fig. 7. All the images below shows the entire experimental workflow in a lab environment. The robot detected the fruit and picked it successfully.

VI. Conclution And Future Work

In this paper, we present a novel autonomous peach-harvesting approach. The robot can detect the peach accurately in a variety of light conditions and achieve robust tracking through the KCF algorithm. The target detection speed is about 14 fps, the average recognition rate is more than 90%, the target tracking speed is about 40 fps, which fully meets the requirements of our project. The robot can automatically plan an optimal path around the tree and find the most suitable peach to pick. Finally, the robot picks the peach through a 6-DOF manipulator. In addition, we propose a monocular high-precision ranging method based on structured light. In the range of 15 to 30 cm, it can achieve high-precision visual ranging and the error does not exceed 6mm.

In the future, we will continue to improve the peach detection algorithm. We will apply a more advanced deep learning method to our system to enhance the robustness of target detection algorithm. In addition, we want to find a better mapping function for peach distance measurement in the future.

References

[1] N. Kawamura, K. NAMIKAWA, T. FUJIURA, and M. URA, "Study on agricultural robot (part 1)," Journal of the Japanese Society of Agricultural Machinery, vol. 46, no. 3, pp. 353 – 358, 1984.

[2] M. Monta, N. Kondo, K. Ting, G. Giacomelli, D. Mears, Y. Kim, and P. Ling, "Harvesting end-effector for inverted single truss tomato [ly-copersicon esculentum] production systems," Journal of the Japanese Society of Agricultural Machinery (Japan), 1998.

[3] S. Arima and N. Kondo, "Cucumber harvesting robot and plant training system," Journal of Robotics and Mechatronics, vol. 11, pp.208 – 212, 1999.

[4] S. Hayashi, K. Ganno, Y. Ishii, and I. Tanaka, "Basic harvesting experiment of eggplant using a robotic system," in Proceedings of the 6th International Symposium on Fruit, Nut and Vegetable Production Engineering. Institut für Agrartechnik Bornim eV (ATB), Potsdam,Land Brandenburg,Federal Republic of Germany, 2001, pp. 221 – 226.

[5] J. Wei, F.-y. Cheng, D. Zhao, Y. Tao, S. Ding, and J. Lü, "Obstacle avoidance method of apple harvesting robot manipulator," Transactions of the Chinese Society for Agricultural Machinery, vol. 44, no. 11,pp. 253 – 259, 2013.

[6] C. Lehnert, I. Sa, C. McCool, B. Upcroft, and T. Perez, "Sweet pepper pose detection and grasping for automated crop harvesting," in Robotics and Automation (ICRA), 2016 IEEE International Conference on. IEEE, 2016, pp. 2428 – 2434.

[7] K. Tanigaki, T. Fujiura, A. Akase, and J. Imagawa, "Cherry-harvesting robot," Computers and electronics in agriculture, vol. 63, no. 1, pp.65 – 72, 2008.

[8] C. Bac, J. Hemming, and E. Van Henten, "Robust pixel-based classification of obstacles for robotic harvesting of sweet-pepper," Computers and electronics in agriculture, vol. 96, pp. 148 – 162, 2013.

[9] Y. Song, C. Glasbey, G. Horgan, G. Polder, J. Dieleman, and G. Van der Heijden, "Automatic fruit recognition and counting from multiple images," Biosystems Engineering, vol. 118, pp. 203 – 215,2014.

[10] W. Ji, D. Zhao, F. Cheng, B. Xu, Y. Zhang, and J. Wang, "Automatic recognition vision system guided for apple harvesting robot," Computers & Electrical Engineering, vol. 38, no. 5, pp. 1186−1195,2012.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances inneural information processing systems, 2015, pp. 91 – 99.

[12] S. Arima, N. Kondo, and M. Monta, "Strawberry harvesting robot on table-top culture," in 2004 ASAE Annual Meeting. American Society of Agricultural and Biological Engineers, 2004, p. 1.

[13] S. Mehta and T. Burks, "Vision-based control of robotic manipulator for citrus harvesting," Computers and Electronics in Agriculture, vol.102, pp. 146 – 158, 2014.

[14] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Visual object detection with deformable part models," Communications of the ACM, vol. 56, no. 9, pp. 97 – 105, 2013.

[15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu,and A. C. Berg, "Ssd: Single shot multibox detector," in European conference on computer vision. Springer, 2016, pp. 21 – 37.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779 – 788.