2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT) | 979-8-3503-9905-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICEICT57916.2023.10245266

Visual Servo Control with YOLOv5 and Kalman Filter for Apple Sorting

1st Donghua Tan
College of Mechanical and Electrical Engineering, Qingdao University
Qingdao, Shandong, 266071, China
15266728317@163.com

2nd Jirong Wang
College of Mechanical and Electrical Engineering, Qingdao University
Qingdao, Shandong, 266071, China
Weihai Innovation Research Institute of Qingdao University,
Weihai, Shandong, 264200, China
wangjirong@qdu.edu.cn

3rd Hongzhi Tian
College of Mechanical and Electrical Engineering, Qingdao University
Qingdao, Shandong, 266071, China
Weihai Innovation Research Institute of Qingdao University,
Weihai, Shandong, 264200, China,
tianhz2020@qdu.edu.cn

***Abstract*—Real-time sorting of items on an assembly line is a challenging task to accomplish without causing any stoppages. Visual servoing may enabel to be a highly efficient solution. This paper achieved sorting by utilizing the Dobot Magician manipulator and the ZED2i binocular camera as the hardware framework. The system is capable of performing tasks such as targeted identification, tracking, and grasping of various apples, on the moving conveyor belt. This paper studies a new system for sorting immature apples by binocular camera and manipulator, and applies deep learning and visual servo technology. The YOLOv5 algorithm is used to realizes object detection and positioning in 2D image and multi-thread processing of the image to achieve accurate and rapid positioning of different targets. The kinematic model and hand-eye calibration of the manipulator are used to realize the coordinate transformation between the camera and the manipulator, and the real-time detection and tracking of the target are completed. Finally, kalman filter is used to predict the trajectory of the target and complete the sorting action.**

The YoloV5 algorithm realizes the precise positioning and maturity identification of apples in 2D images, and the ZED2i binocular vision realizes the 3D spatial positioning of apples. Use visual servo control to guide the manipulator for real-time and precise sorting. This study provides an effective way to improve the level of agricultural automation, reduce labor costs and ensure fruit quality, and has high practical value and broad application prospects.

***Index Terms*—robot vision, robot kinematics, deep learning, visual servoing, object recognition and tracking, kalman filter**

I. Introduction

With the advent of Industry 4.0, the manipulator combined with machine vision technology has begun to be applied in agriculture [1]. At present, the yield of fruit is gradually increasing, but the fruit industry has not achieved high income. The main reason is that the low efficiency and error-prone manual operation is mainly used for sorting, the post-harvest treatment level of fruit is low, and the added value of high-quality fruit cannot be improved. Therefore, it is more effective to use the manipulator combined with machine vision and visual servo technology to identify, track and sort fruits in complex production environments.

Naoshi Kondo has made a robot for fruit grading system using a visual system composed of color cameras and lighting equipment [2]. The system uses multiple color cameras to check the target in all directions and suck the apple through the sucker. In 2016, M.M.Sofo released a real-time apple sorting system that can distinguish multiple apples based on color, size and composition. The machine achieved the efficiency of sorting 15 apples per second in real time using two channels [3]. At the same time, various visual algorithms for apple discrimination have been continuously improved. For example, for multi-fruit sorting technology, Mathew George proposed a low-cost image acquisition device using fuzzy logic and K-Means image clustering method [4]. Wen and Tao's team introduced a rule-based apple classification system. Nakano used two neural networks to classify the color of 'SanFuji' apples [5].

What is proposed in this paper is a deep learning using YOLOv5 algorithm, combined with visual servo technology to control the manipulator to realize the recognition and sorting of immature apples on the apple conveyor line, and add multi-threading processing and Kalman filtering of visual recognition to optimize the tracking effect of the manipulator [6]. Through this way of apple sorting, a single manipulator can be connected to multiple cameras to complete the sorting of multiple conveyor lines, which increases work efficiency and reduces energy consumption.

II. METHODOLOGY

A. Kinematics analysis of the manipulator

The research method of the manipulator used in this study is the Denavi-Hartenberg method, which is a method to describe the relationship between two joints by quadratic transformation matrix [7]. In order to establish the kinematics model, firstly, a coordinate system of the manipulator is established as shown in Fig.1, where P point is the grasping point of the end effector, and the D-H parameter table is established with the coordinate system as the reference, as shown in TABLE I. Where θ_i represents the rotation angle between the two connecting rods, its direction is along the z_i direction, the angle between x_{i-1} and x_i, α_{i-1} represents the angle between the two rotation axes, its direction is along the x_{i-1} direction,

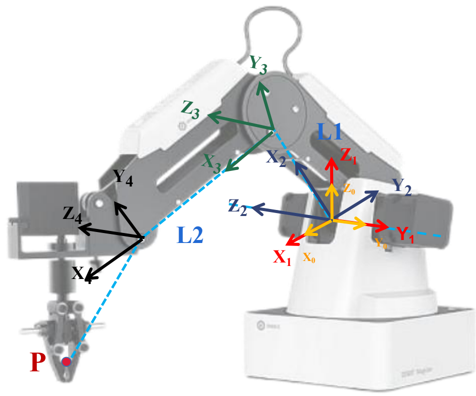

Fig. 1. Real-time apple sorting system

the angle between z_{i-1} and z_i. a_{i-1} represents the length of the connecting rod, and d_i represents the distance between adjacent connecting rods.

TABLE I
D-H PARAMETER TABLE

i	α_{i-1}	a_{i-1}	d_i	θ_i
1	0	0	0	θ_1
2	90°	0	0	θ_2
3	0	135mm	0	θ_3
4	0	147mm	0	θ_4

According to the kinematics of the robot, the relative position relationship between the two links of the manipulator can be expressed by the homogeneous transformation matrix between the two adjacent coordinate systems. The general formula is formula (1). The data in the D-H parameter table are brought into formula 1 to obtain the homogeneous transformation matrix of the end coordinate system $\sum_4$ of the manipulator relative to the base coordinate system $\sum_0$. In the formula (3), $\mathrm{c}_i = \cos\theta_i$, $\mathrm{s}_i = \sin\theta_i$.

$$
{}_{i}^{i-1}\boldsymbol{T} = \begin{bmatrix} c\theta_i & -s\theta_i & 0 & a_{i-1} \\ s\theta_i c\alpha_{i-1} & c\theta_i c\alpha_{i-1} & -s\alpha_{i-1} & -s\alpha_{i-1} d_i \\ s\theta_i s\alpha_{i-1} & c\theta_i s\alpha_{i-1} & c\alpha_{i-1} & c\alpha_{i-1} d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}
$$

$$
{}_{4}^{0}\boldsymbol{T} = {}_{1}^{0}\boldsymbol{T}{}_{2}^{1}\boldsymbol{T}{}_{3}^{2}\boldsymbol{T}{}_{4}^{3}\boldsymbol{T} = \begin{bmatrix} n_x & s_x & a_x & p_x \\ n_y & s_y & a_y & p_y \\ n_z & s_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2}
$$

$$
\begin{cases} n_x = c_1 c_3 \left(c_2 c_4 - s_2 s_4\right) - c_1 s_2 s_3 \left(c_4 + s_4\right) \\ s_x = c_1 s_2 s_3 \left(s_4 - c_4\right) - c_1 c_2 \left(c_3 s_4 + s_3 c_4\right) \\ n_y = s_1 c_2 \left(c_3 c_4 - s_3 s_4\right) - s_1 s_2 \left(s_3 s_4 + c_3 s_4\right) \\ s_y = s_1 s_2 \left(s_3 s_4 - c_3 c_4\right) - s_1 c_2 \left(c_3 s_4 + s_3 c_4\right) \\ n_z = s_2 c_3 c_4 + c_2 s_3 c_4 - s_2 s_3 s_4 + c_2 c_3 s_4 \\ s_z = -s_2 c_3 s_4 - c_2 s_3 s_4 + -s_2 s_3 c_4 + c_2 c_3 c_4 \\ p_x = L_1 c_1 c_2 + L_2 c_1 \left(c_2 c_3 - s_2 s_3\right) \\ p_y = L_1 s_1 c_2 + L_2 s_1 \left(c_2 c_3 - s_2 s_3\right) \\ p_z = L_1 s_2 + L_2 \left(s_2 c_3 + c_2 s_3\right) \\ a_x = s_1, \mathrm{a}_y = -c_1, a_z = 0 \end{cases} \tag{3}
$$

If the position of the target point P is known, then the rotation angle of each joint can be obtained. At this time, $x = p_x$, $y = p_y$, $z = p_z$, the inverse solution of the kinematics of the manipulator is :

$$\theta_1 = Atan2(y, x) \tag{4}$$

$$\cos\theta_3 = \frac{\left(x c_1 + y s_1\right)^2 + z^2 - L_1^2 - L_2^2}{2 L_1 L_2} \tag{5}$$

$$\theta_3 = Atan2(\sin\theta_3, \cos\theta_3) \tag{6}$$

$$\cos\theta_2 = \frac{z L_2 s_3 + \left(x c_1 + y s_1\right)\left(L_2 c_3 + L_1\right)}{\left(L_2 c_3 + L_1\right)^2 + \left(L_2 s_3\right)^2} \tag{7}$$

$$\theta_2 = Atan2(\sin\theta_2, \cos\theta_2) \tag{8}$$

B. The construction of visual system

The visual system is mainly used to identify the immature apples mixed with mature apples on the conveyor belt. The design goal is to be able to complete the identification and classification of the target in time, so that the visual servo control manipulator can track the target in real time. The construction process of all visual systems is shown in Fig.2.

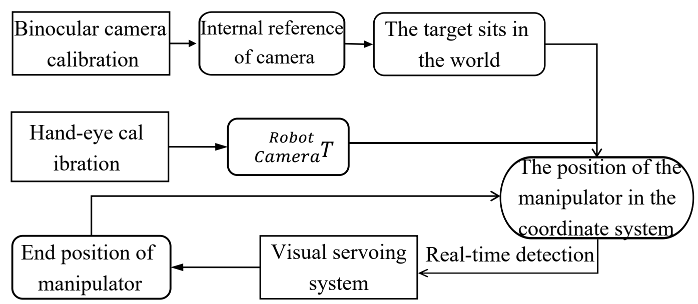

Fig. 2. The flow chart of visual system construction

In order to realize the accurate identification of immature apples, this paper uses the YOLOv5 algorithm to conduct deep learning on the type and maturity of apples. In the training process of the model, 100 pictures were selected as the training set and 30 pictures were used as the verification set. The 740 rounds of iterative training were performed on a laptop equipped with GTX1050 GPU, and the final training results are shown in Fig.3.

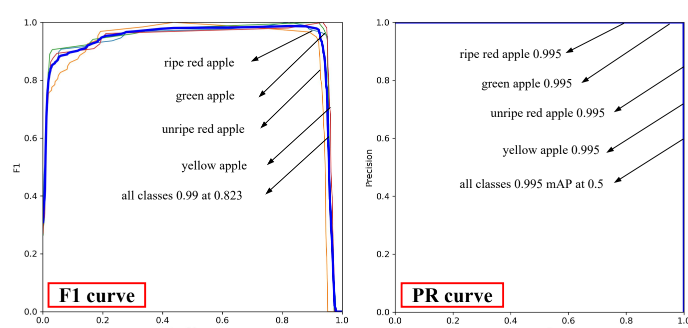

Fig. 3. Results of deep learning

The training direction of the model is to train the maturity of red apples such as Fuji and Gala, and the distinction between red apples and green apples such as green apples and golden crown apples such as golden apples. The maturity of red apples is judged according to the surface color, and surface peel color more than 80 % red is mature. Four labels are set during the training process, and the mAP of all labels reaches 0.995 [8], and the performance of the model is excellent.

In order to detect the effect of model training, the images containing various types of apples are identified and verified in the experiment. The recognition results are shown in Fig.4. It can be seen that the accuracy of distinguishing apple types is high, and the task of tracking, sorting and grasping can be realized. For the occlusion phenomenon that is easy to occur in the recognition process, the recognition accuracy of apples under different occlusion conditions is also verified [9]. The results are shown in Fig.5. It can be seen that when the apple only has a small area in the line of sight, the target can also be identified.

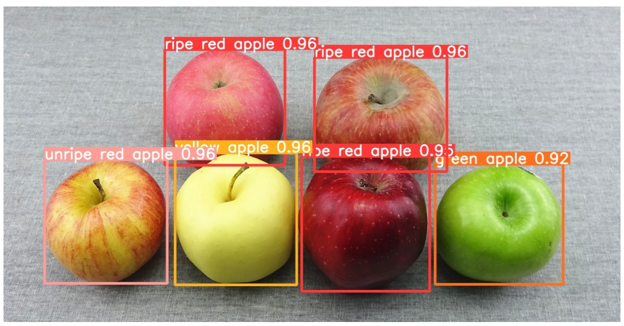

Fig. 4. Identification results of different varieties of apples

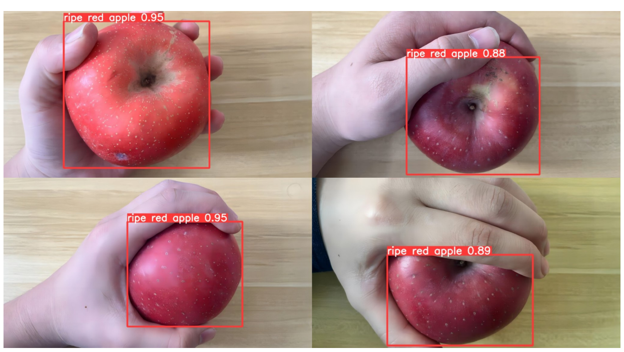

Fig. 5. Identification results of apples under different occlusion degrees

In this experiment, we take the position-based visual servoing(PBVS) as the core of the scheme, and adopt the look-and-move working method [10]. This method is relatively simple and direct to operate. The coordinates of the end of the manipulator are used as feedback signals to compare with the coordinates of the target apple detected by the camera in real time. The difference of coordinates is used to control the manipulator to keep close to the target apple. In the tracking process, Kalman filter is added to realize that apple can still be identified when it is blocked in a short time. At the same time, in order to solve the defects of slow tracking speed and tracking frustration caused by slow recognition speed, multi-thread technology is used to speed up the recognition speed of each frame in the recognition process, and the multi-core computing ability of the computer is fully utilized. The control structure of the visual servo system is shown in Fig.6.

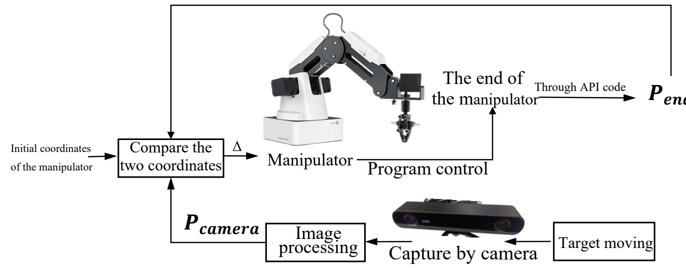

Fig. 6. Visual servoing system structure diagram

After completing the calibration of the camera's internal parameters, the entire visual system is built as shown in Fig.7. In the picture, the ZED2i binocular camera can be seen to draw a depth map from the parallax values measured by the left and right cameras. The distance between the target apple and the camera is measured and the test results are drawn on a color map.

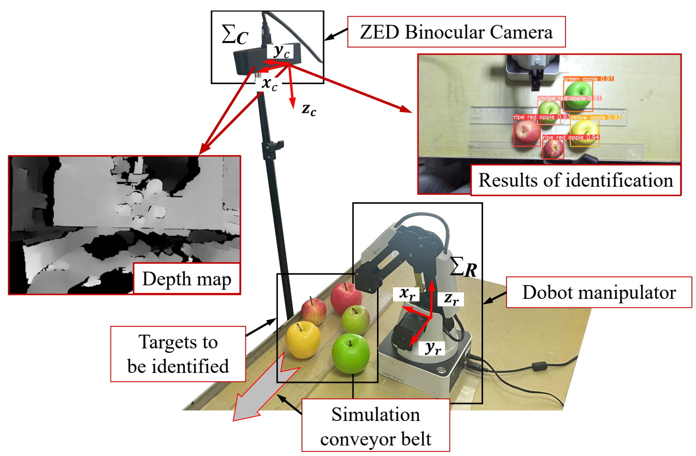

Fig. 7. The construction of visual system

C. Effect of tracking and sorting

In the experiment, the tracking effect of the manipulator on the target apple is shown in Fig.8.

In the process of tracking the target apple by visual servoing, 320 sampling points are set to record the coordinates of the end of the manipulator at this time. Finally, the tracking results of the manipulator are shown in Figure 9 and Figure 10 . In the figure, $\sum_C$ represents the camera coordinate system, $\sum_R$ represents the manipulator coordinate system, ${}^{R}x_E$ represents the position of the end of the manipulator in the x-axis direction of the manipulator coordinate system, ${}^{R}x_a$ represents the actual position of the apple in the x-axis direction, Δx

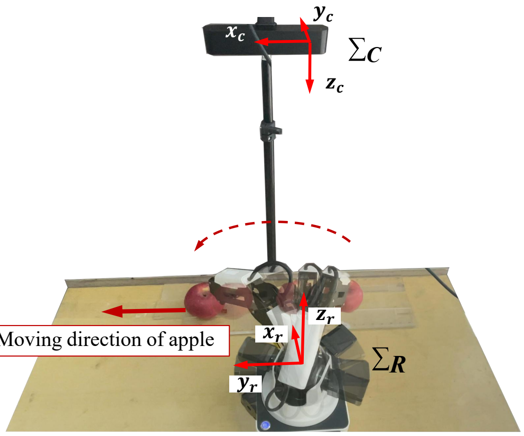

Fig. 8. Tracking process trajectory diagram

represents the error in the x-axis direction, which is the same in the y-axis and z-axis. It can be seen that the average error in the x-axis direction is 1.91 mm, the average error in the y-axis direction is 4.25 mm, and the average error in the z-axis direction is 2.94 mm.

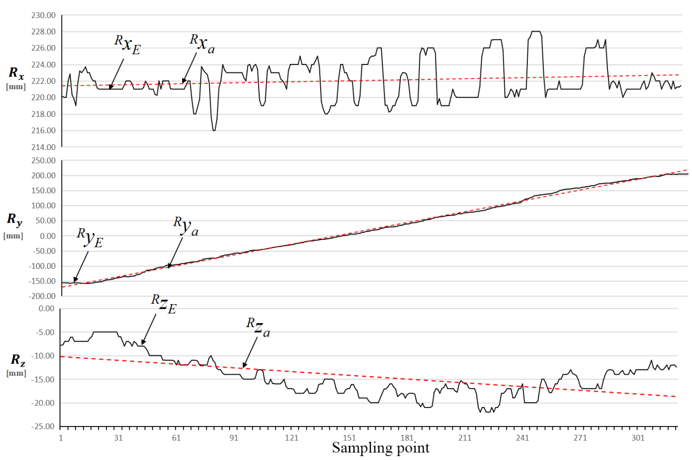

Fig. 9. The pose fitting curve of the manipulator

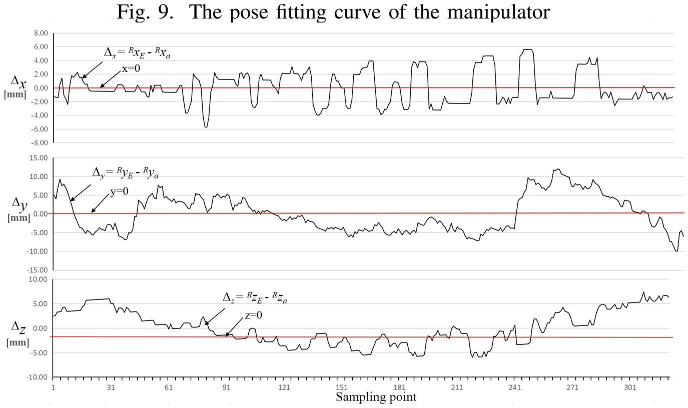

Fig. 10. Tracking error curve of manipulator

III. CONCLUSION

This study utilizes the YOLOv5 algorithm for object detection in visual servo control, and integrates binocular vision to achieve maturity recognition and positioning of apples. A Kalman filter is utilized to predict the trajectory of the apples on the conveyor belt. Compared to traditional automatic conveyor lines and manual sorting methods, this approach boasts higher accuracy, flexibility, and real-time performance. Additionally, multi-threaded processing technology is implemented in the experiment to accelerate target recognition speed, thus further enhancing the system's performance. In conclusion, this methodology is highly effective for sorting apples.

REFERENCES

[1] Y. Lu, "Artificial intelligence: a survey on evolution, models, applications and future trends," *Journal of Management Analytics*, vol. 6, no. 1, pp. 1–29, 2019.

[2] N. Kondo, "Robotization in fruit grading system," *Sensing and instrumentation for food quality and safety*, vol. 3, pp. 81–87, 2009.

[3] M. Sofu, O. Er, M. Kayacan, and B. Cetişli, "Design of an automatic apple sorting system using machine vision," *Computers and Electronics in Agriculture*, vol. 127, pp. 395–405, 2016.

[4] M. George, "Multiple fruit and vegetable sorting system using machine vision," *Int J Adv Technol*, vol. 6, no. 142, p. 2, 2015.

[5] K. Nakano, "Application of neural networks to the color grading of apples," *Computers and electronics in agriculture*, vol. 18, no. 2-3, pp. 105–116, 1997.

[6] S. Chen, "Kalman filter for robot vision: a survey," *IEEE Transactions on industrial electronics*, vol. 59, no. 11, pp. 4409–4420, 2011.

[7] D. Songyi and Xuefeng, "Development of a self-balance dual-arm robot for inspection of high-voltage power transmission lines," in *2012 IEEE International Conference on Mechatronics and Automation*. IEEE, 2012, pp. 2482–2487.

[8] A. M. Roy, J. Bhaduri, T. Kumar, and K. Raj, "Wildect-yolo: An efficient and robust computer vision-based accurate object localization model for automated endangered wildlife detection," *Ecological Informatics*, vol. 75, p. 101919, 2023.

[9] S. Sun, M. Jiang, D. He, Y. Long, and H. Song, "Recognition of green apples in an orchard environment by combining the grabcut model and ncut algorithm," *Biosystems Engineering*, vol. 187, pp. 201–213, 2019.

[10] F. Conticelli and B. Allotta, "Two-level visual control of dynamic look-and-move systems," in *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, vol. 4. IEEE, 2000, pp. 3784–3789.