

A Design of Robot System for Rapidly Sorting Express Carton with Mechanical Arm Based on Computer Vision Technology

Yuzhang Li*

School of mechanical engineering, City university of Hongkong, Hongkong, China

* Corresponding Author Email: yuzhangli2-c@my.cityu.edu.hk

Abstract. With the rapid development of logistics and express industry, more and more cartons need to be sorted rapidly, which also consumes a lot of human labor. Manipulators are currently used in industrial and medical fields, with high requirements for accuracy, repeatability, and operation stability. With the continued expansion of the social economy and the advancement of science and technology, computer vision technology has become increasingly widely used in a variety of industries. Using computer vision technology in an automation system can improve picture processing accuracy and allow precise recognition. A prototype robot system for picking express cartons is designed, which combines object detection technology based on computer vision and trajectory planning methods in UR3 manipulator grabbing. This design aims to simulate the car with mechanical arm identify the express carton, grab and place it in the storage after moves to the preset grabbing point. Building simulation environment in gazebo, simulating the grabbing and placing process in gazebo by using ROS1 to analysis the kinematics and dynamics control of the manipulator system. The simulation results verify the feasibility of the architecture design and the accuracy of the dynamic controlling system design. Different size express cartons are recognized by using Open CV. Using MoveIt to achieve collision detection and smoother trajectory planning. This prototype design can also provide a reference to future space intelligent robot.

Keywords: Robot System; Mechanical Arm; Computer Vision.

1. Introduction

The classic definition of robot is an actuated mechanism that can be programmed in two or more axes, has some autonomy, and moves through its environment to carry out specific tasks. A control system and an interface for that control system are both components of a robot. After a lengthy period of evolution during the electrical and digital ages, robots are now entering a new era of intelligence.

Conventional automated technologies such as servomotors have been replaced by high-technology intelligent robots integrated with computer vision functions. Intelligent robot is a machine system with vastly enhanced perception, decision-making, and performance in comparison to conventional robot. Intelligent manufacturing and industrial upgrading rely heavily on industrial robots and the usage of robots throughout the world is expanding year after year. Intelligent robots can be broken down into three categories, based on their intended use case: industrial, service, and specialized [1], [2].

Multi degree of freedom robot manipulators have been widely utilized in the industrial commerce context for more complex or more precise operations than human operators can perform. Automatically searching for the ideal path from the current pose to the desired pose, also known as trajectory planning. A robot's implementation of tasks with high-quality performance is facilitated by its ability to find a successful path [3].

In contrast to AI technologies such as neural networks, which have been widely utilized since the 1940s, computer vision technology did not become widespread until the late 1960s and early 1970s. Initially addressing themes such as simulating the human visual system through a variety of understandings of camera models. In later years, computer vision emerged in all sectors.

Vision-based automated manipulator sorting and grasping technology is a new field of machine vision technology with mechanical motion methodology. It interprets and analyzes the image data obtained by the camera to determine the position of the target object, and then directs the industrial robot to perform the appropriate tracking and grasping actions. For robots, machine vision provides

its sophisticated computing system and processing system, simulating bio-visual imaging and information processing, allowing the robot arm to perform more anthropomorphic and flexible operations while simultaneously identifying, comparing, and processing scenes, to generate execution instructions, and then carry out the action. Robots' adaptability and operability are contingent upon machine vision.

The rapid expansion of the express industry has resulted in an abundance of express packaging waste. An efficient sorting machine is required to facilitate reducing waste of human resources. Computer vision analyzes the image data obtained by camera to determine the position of the express carton, and then direct the robot arm to perform the tracking and grabbing actions. Through efficient intelligent robot system, it can reduce labor costs and reduce carbon emissions. Aiming at these problems mentioned above, this paper designs a robotic system for rapid sorting express cartons using robotic arms based on computer vision technology. This design aims to simulate the car with mechanical arm identify the express carton, grab and place it in the storage after moves to the preset grabbing point.

2. System Overall Design

The overall process of the system designed in this paper is shown in Fig. 1. First, the YOLO algorithm is used to identify the carton to be grabbed, and position information is obtained through a depth camera. At the same time, the camera position and posture are calculated, as well as the position and posture of the mechanical claw end actuator are also obtained. Then, through Eye-in-hand calibration method for coordinate system transformation and analysis of grasping attitude, a collision detection method combining cylindrical envelope and reference frame perimeter boxes method is designed. Finally, RRT algorithm is used for path planning, shortening planning time, and smoothing path curves.

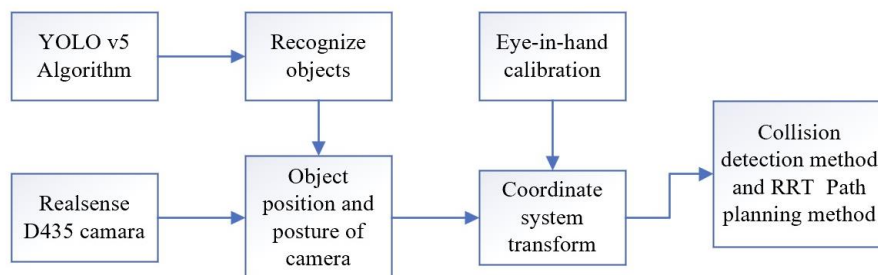


Fig.1 System architecture and composition

In the subsequent research process, this design can be combined with the simultaneously mapping technology SLAM and autonomous navigation function of the car to achieve autonomous navigation of the vehicle equipped with a 6-degree of freedom mechanical arm to avoid obstacles and navigate to the designated grasping location in the logistics factory.

3. Methodology

3.1. Target Detection Technology for Grabbing Robots

This design provides a technique for sorting express cartons using the vision-based manipulator-enhanced YOLOv5 algorithm and designs a methodology with real-time processing capability to address conventional industrial robotic arms' high failure proportion and poor feedback efficiency during the conveyor logistics sorting process [4]. Locating the relevant item in a picture and simultaneously calculating its location information are two essential steps in the process of object detection [5]. Even though the two-stage object detection algorithm has a decent detection rate and accuracy, the detection speed is sluggish to meet the industry's authentic and quick-speed standards [6, 7]. Consequently, the YOLOv5 convolutional neural network model, which represents the one-

stage technique, is used for this design. As a consequence of this, the YOLOv5 convolutional neural network (CNN) model is utilized for the creation of this system. This model is representative of the one-stage method.

3.2. Eye-in-hand Calibration

In response to a rise in the complexity and variety of occupations that are done by machine and manipulators, "hand-eye" cooperative process has garnered considerable interest. In terms of determining robot and camera parameters, many issues remain [8, 9]. By constructing a statistical equation, the manipulation hand-eye cooperation process precisely achieves an accurate solution to the homogeneous transition matrix. The flow chart of hand-eye calibration is shown as follow.

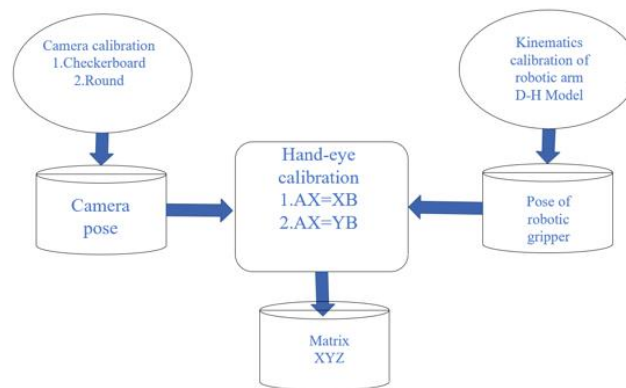


Fig.2 The flow chart of hand-eye cooperation process

To improve the accuracy of the robot's optical service, researchers have developed a hand-eye cooperation process to obtain a transition matrix that is homogeneous between the camera and the manipulator [9]. Hand-eye cooperation process can be classified into different types based on the installation location of the camera. Eye-to-hand indicates that the depth camera remains stationary while the manipulator moves. The alternative approach involves a camera that is attached to the front section of the clamp for the manipulator and moves along with it. Since eye-in-hand cooperation process has a broader application range. This design chooses eye-in-hand as its target detection technology. This paper chooses eye-in-hand as a technical component of the technical framework.

3.2.1 Camera Calibration

Estimating the pose of the camera is a very important area of research in the science of computer vision. The multi-coordinates of feature points and image coordinates are the most influential elements on the accuracy of camera pose estimate during camera calibration. According to the concept of tag generation, there are primarily two types of calibration plates shown in Fig.3. Conventional labels mostly consist of checkerboard calibration board and circle grid calibration board. The common calibration board is shown in the following figure [10].

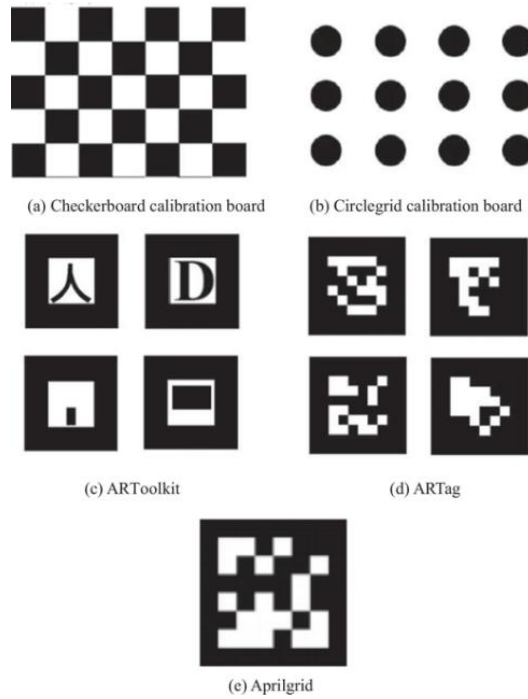


Fig.3 Several planar patterns marked systems

The mathematical model of camera imaging can usually be represented using a pinhole imaging model. The following image depicts the imaging principle that operates within the camera.

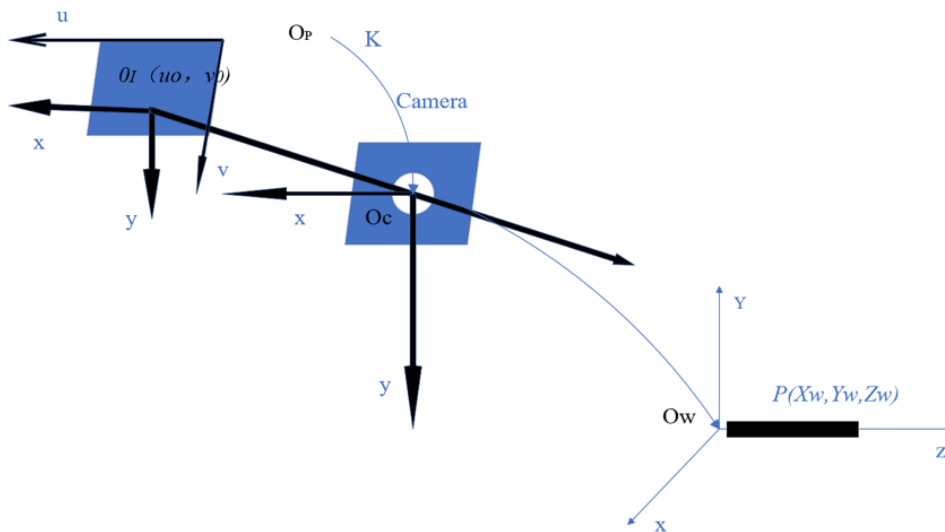


Fig.4 Pinhole model of the camera

As can be seen in the Fig.4, $\{O_w\}$, $\{O_c\}$, $\{O_i\}$, and $\{O_p\}$ each denote a different reference frame: the camera reference frame, the image reference frame, the calibration plate reference frame, and the pixel reference frame, respectively. There is an equation (1) that can be used to express the picture architecture of the camera.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} s = \begin{bmatrix} \alpha & 0 & u_0 & 0 \\ 0 & \beta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = KN \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (1)$$

Where the coordinate of the point in the pixel coordinate system is denoted by $(u, v, 1)^T$, and the homogeneous coordinate of the point in the world coordinate system is denoted by

$(X_\omega, Y_\omega, Z_\omega, 1)^T$. K refers to the camera's internal settings and parameters. Both f and k denote the focal length after it has been adjusted. The primary point coordinates are written out as u_0 and v_0 . The camera's external settings are represented by the matrices labelled N , which are each 4 by 4.

3.2.2 Robot Calibrations

The robot's repetitive position accuracy and absolute location accuracy define the position of its end-effector. Thus, the robot needs to have its calibration done in order to improve the real-time precision it possesses. The discrepancy between the estimated and real end-effector positions is what needs to be adjusted in order for aligning to achieve this goal [11]. The kinematics equation (2) of the robot can be used to provide a representation of this error.

$$f_{min} = \|T_{na} - T_{1,e}\|_2 \quad (2)$$

$$T_{1,e} = T_{0,1}T_{1,2}T_{2,3}T_{3,4}T_{4,e}$$

Where f_{min} denotes the minimum distance between the end-nominal effector's attitude and its actual pose; The T_{na} matrix denotes the supposedly homogeneous transformation from the robot arm's base frame to the end-effector frame. $T_{1,e}$ is the matrix that depicts the predicted homogenous translation from the frame structure of the robotic arm to the frame of the end-effector. The equation that best summarizes the model of the robot's direct kinematics is as follows:

$${}^{i-1}_i T = R_z(\theta_i) \times T_z(d_i) \times T_x(a_i) \times R_x(\alpha_i)$$

$$= \begin{bmatrix} c\theta_i & -sca_i & s\theta_i ca_i & ca_i \theta_i \\ s\theta_i & c\theta_i ca_i & -c\theta_i a_i & a_i c\theta_i \\ 0 & sa_i & ca_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

As can be seen, the four parameters θ_i , a_i , d_i and α_i , respectively describe the joint angle, the length of the link, the offset of the link, and the twist of the link. When the pose of the camera and the pose of the pose of robotic arm's end-effector are obtained, establishing a statistical equation of hand-eye calibration and obtaining the homogenous transformation matrix from the camera frame to the end-effector frame is the subsequent stage. According to the mathematical model's structure, the existing hand-eye electrical measurement are classified as follows: AX=XB, AX=YB, reprojection error-based calibration [12].

In this design, an online hand-eye calibration based on CharUco board which means combines the advantages of Chessboard and Aruco marker technologies is designed. Initially, after the robot has reached the faculty position, a calibration picture of the ChArUco board is taken by the camera. Secondly, it is determined if the calibrating panel's location satisfies the conditions by recognizing the ChArUco board. If the calibration board image position is not satisfied, the robot will automatically modify the robot's posture depending on the feedback ChArUco board's location and repeat the procedures above, save the data, modify the robot's position, and proceed to the following instructional step. Solve the hand-eye calibration matrix if the collected calibration data satisfy the specifications. The process is shown as following:

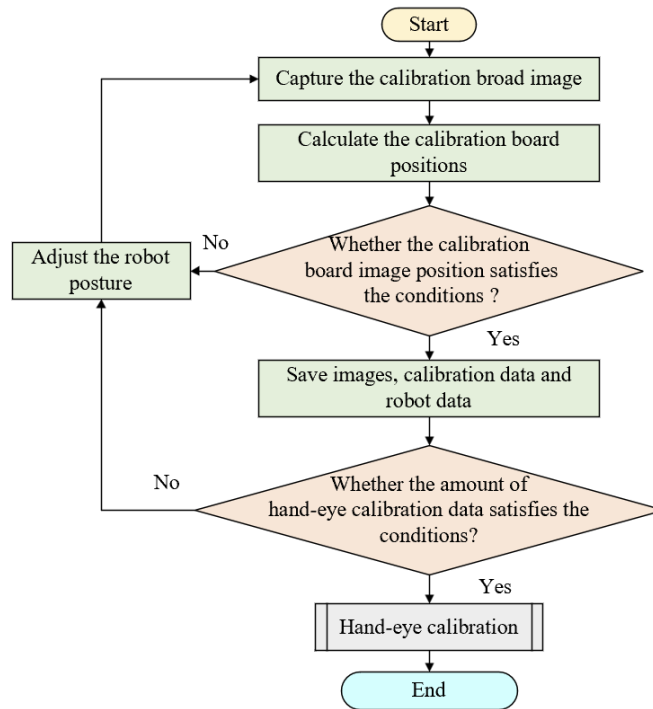


Fig.5 Hand-eye calibration procedure based on ChArUco board.

3.3. Research on Combination of Collision Detection and Motion Planning

Many studies have been conducted on collision detection techniques. Methods of collision detection vary according to robot type, operating scenario, and real-world demands. The boundary box technique, which simplifies the amount of computation used for blocking robots and the surrounding environment, is utilized to reduce superfluous computation. A cylinder is the simplest shape that most closely approximates the linkages of a robot manipulator. As a consequence of this, a cylinder enclosure may be used for the purpose of obstacle detection. The issue of collision between the robotic arm's links is transformed into determine whether the distance between two cylinders meets the minimum distance requirement to avoid collisions. The difficulty has been reduced. Utilizing the cylinder enclosure method necessitates determining the volume, location, and orientation of the enclosed sphere that replaces the actual joints, links and other objects. The spatial line segment formula describing the area where C_i is located will be represented as follows:

$$\frac{x-x_i}{x_{i+1}-x_i} = \frac{y-y_i}{y_{i+1}-y_i} = \frac{z-z_i}{z_{i+1}-z_i} = t, t \in (0,1) \quad (4)$$

Where (x_i, y_i, z_i) and $(x_{i+1}, y_{i+1}, z_{i+1})$ represent the end face center coordinates of the link envelope cylinder and t represents the proportional constant. The parametric representation of the equation is:

$$\begin{cases} X_i(t) = tx_{i+1} + (1-f)x_i \\ Y_i(t) = ty_{i+1} + (1-f)y_i \\ Z_i(f) = t_{i+1} + (1-t)z_i \end{cases} \quad (5)$$

Simplify, the vertical spatial length between the center line of circular cylinder is shown as follows:

$$d_{ij} = \sqrt{(X_i(t) - X_j(m))^2 + (Y_i(t) - Y_j(m))^2 + (Z_i(t) - Z_j(m))^2} \quad (6)$$

When $d_{ij} > r_i + r_j$, the distance meets the distance requirements for collision avoidance. When $d_{ij} < r_i + r_j$, this distance does not meet the distance requirements for collision prevention. Many studies have been conducted on trajectory planning. The Rapidly-Exploring Random Tree's fundamental stage is to build certain branch from the path's initial point and arbitrarily grow the leaf nodes until the goal point is reached. A path is finally obtained when an expanding connected nodes reaches or approaches the target. Owing to the high unpredictability of the path of Rapidly Exploring Random Tree, the outcome is frequently significantly more costly than the best approach. In the entirely randomly investigation case, the correlation between the area's extent and its length from the objective can be conducted as

$$\lim_{n \rightarrow \infty} A_{probe} = \pi l_{dis}^2 \tag{7}$$

, Where n represents the number of investigations, l_{dis} represents the distance between the starting and ending sites, and A_{probe} represents the region containing the exploration sites. It is evident that the strategy of boosting the amounts of trees can significantly decrease the number of searches by reducing the searching length of a sole tree.

As a result, the goal of this paper's algorithm improvement is to expand the root of the tree. In the case of the straight path, average sampling is used to sample five places throughout the whole path, and collision detection is conducted at each point. If there is no collision at any of the five places, it is assumed that the path is free of collisions [13]. The generation and selection of roots is show in Fig. 6.

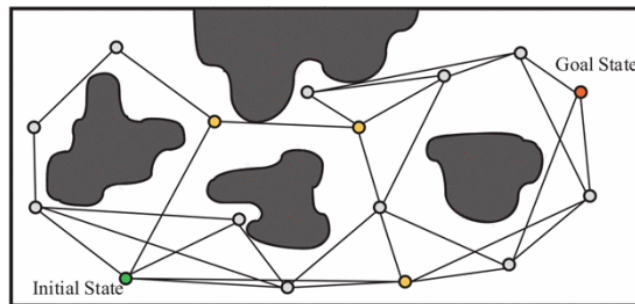


Fig. 6 Origination and choosing of roots

At the start, a specific number of elements are originated at random in the working space, and these are the potential root growth points. After joining the points collision, collision detection is performed on each edge, leaving only unobstructed edges [14].

As all of the randomized roots are chosen, they all begin to grow at the same time, as illustrated in Fig. 7. Finally, connecting block paths to form a complete path, which is shown in Fig. 8.

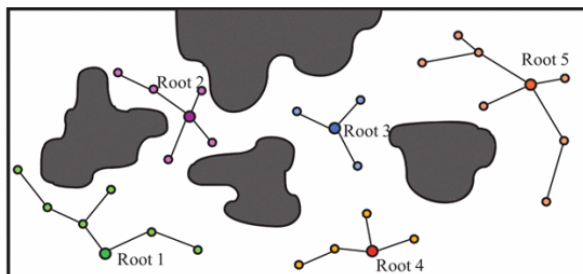


Fig.7 Tree grows simultaneously

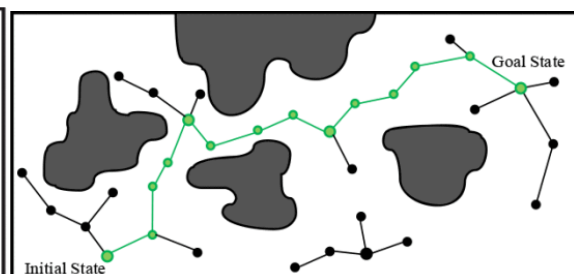


Fig.8 Path generating

4. Experiment and Analysis

4.1. Introduction to Simulation Environment

The robot operating system (ROS) offers a unified controlling platform for robotic devices and robot systems. ROS is a framework developed specifically for manipulator robot arm and moving

robots in the field of research with its open-source, simple constructure and compatibility. The development of ROS has gone through two versions and ROS1 is more compatible and stable. The ROS1 will be installed on the Ubuntu 20.04 noetic system.

With the physical simulation engine Gazebo, it is possible to simulate the robot's working conditions in the real world, thereby reducing the complexity of the work and increasing its efficiency. RViz is one of the most important and fundamental function of ROS. RViz is capable of completing the visualization of 3D data and executing the coordinate system establishment and sensor coordinate transformation in complex systems. Taking inspiration from the Arm Navigation architecture in ROS, Moviet! incorporated kinematics, motion planning, three-dimensional perception, and a control interface to provide the fundamental function of moving an arm in virtual environments. Moveit can quickly solve forward and reverse kinematics for robotic arms. It also has a function of collision detection to provide a collision-free track for grabbing.

4.2. Simulation Experiment Results and Analysis

The simulation environment consists of an operating table in a simulated logistics sorting scenario, a UR3 robot arm on a base with a depth camera above the jaws of this arm, which is shown in Fig. 9. Blocks represent logistics cartons, and the logistics cartons will be placed in the classification box. Run simulation environment in gazebo is shown as follow. The model file of the robotic arm with 6 degrees of freedom is shown in Fig.10.

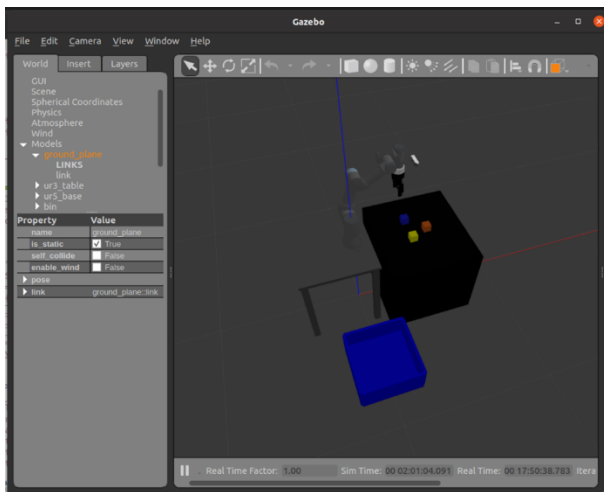


Fig.9 simulation environment

```

1 ?xml version="1.0"?
2 <robot xmlns:xacro="http://wiki.ros.org/xacro">
3
4 <!--
5 Author: Felix Messner
6 -->
7
8 <xacro:include filename="$(find ur_description)/urdf/ur.transmission.xacro" />
9 <xacro:include filename="$(find ur_description)/urdf/ur.gazebo.xacro" />
10
11 <xacro:macro name="cylinder_inertial" params="radius length mass *origin">
12 <inertial>
13 <mass value="$(mass)" />
14 <xacro:insert_block name="origin" />
15 <inertia ixx="$(0.833333 * mass * (3 * radius * radius + length * length))" ixy="0.0" iyz="0.0"
16 iyy="$(0.833333 * mass * (3 * radius * radius + length * length))" izx="0.0"
17 izy="0.0" izz="$(0.5 * mass * radius * radius)" />
18 </inertial>
19 </xacro:macro>
20
21 <xacro:macro name="ur3_robot" params="prefix joint_limited">
22 shoulder_pan_lower_limit:=${pi} shoulder_pan_upper_limit:=${pi}
23 shoulder_lift_lower_limit:=${pi} shoulder_lift_upper_limit:=${pi}
24 elbow_joint_lower_limit:=${pi} elbow_joint_upper_limit:=${pi}
25 wrist_1_lower_limit:=${pi} wrist_1_upper_limit:=${pi}
26 wrist_2_lower_limit:=${pi} wrist_2_upper_limit:=${pi}
27 wrist_3_lower_limit:=${pi} wrist_3_upper_limit:=${pi}
28 transmission_hw_interface:hardware_interface/PositionJointInterface
29 safety_limits:=${false} safety_pos_margin:=${0.15}
30 safety_k_position:=${20}
31
32 <!-- Inertia parameters -->
33 <xacro:property name="base_mass" value="2.0" /> <!-- This mass might be incorrect -->
34 <xacro:property name="shoulder_mass" value="2.0" />
35 <xacro:property name="upper_arm_mass" value="3.5" />
36 <xacro:property name="forearm_mass" value="1.26" />
37 <xacro:property name="wrist_1_mass" value="0.8" />
38 <xacro:property name="wrist_2_mass" value="0.8" />
39 <xacro:property name="wrist_3_mass" value="0.35" />
40

```

Fig.10 UR3 robotic arm model file

Using Rviz to complete the visualization of 3D data of robot arm. The operation process is shown in the following figure.

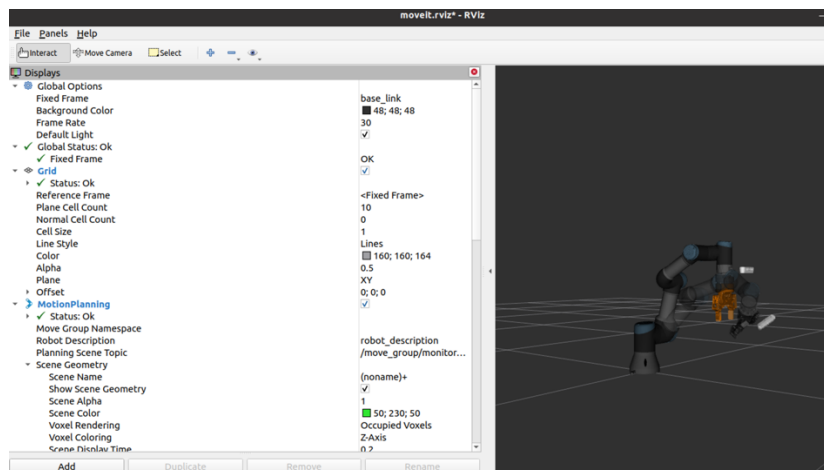


Fig.11 The interface of UR3

Utilizing Open CV to recognize box and running grabbing program, observing 3D images of robot arm using Rviz. The process of grasping and placing is shown in the following figure.

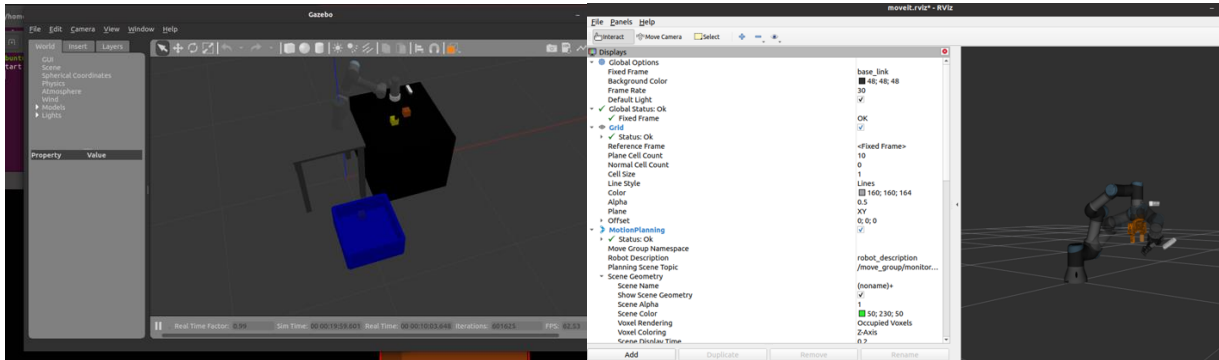


Fig.12 process of grabbing

Forward and reverse kinematics of the robot arm calculated using Moveit. After sorting process is completed, the robotic arm returns to its starting position, as shown in Fig.13.

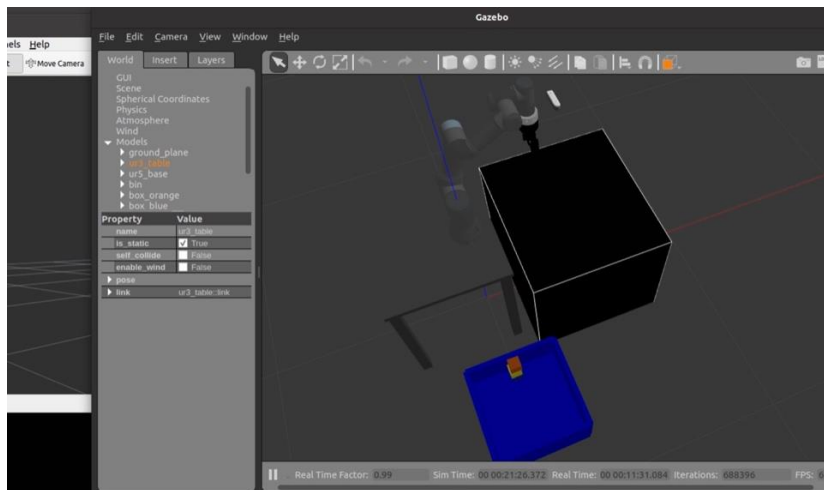


Fig.13 Sorting process is completed

The simulation results verify the feasibility of the mechanical arm structure design of the robot and the accuracy of the control system design. Express cartons are recognized by using Open CV. Using MoveIt to achieve collision detection and smoother trajectory planning.

5. Conclusion

The robotic arm designed in this article is a new application of machine vision technology with robot motion technology. This design provides a technique for sorting express cartons using the vision-based manipulator-enhanced YOLOv5 algorithm and designs a methodology with real-time processing capability to address conventional industrial robotic arms' high failure proportion and poor feedback efficiency during the conveyor logistics sorting process. The accuracy of the visual-grabbing serving is improved by employing the hand-eye cooperation process, which yields a transition matrix that is homogeneous between the camera and the manipulator. The designed robot integrates collision detection and RRT path trajectory planning algorithms, achieving smoother grabbing and collision-free grabbing. Simulating the grabbing process with ROS in gazebo. Express cartons are recognized by using Open CV. Using MoveIt to achieve collision detection and smoother trajectory planning. The simulation results verify the feasibility of the mechanical arm structure design of the robot and the accuracy of the control system design. The designed robotic limb can facilitate rapid sorting and significantly reduce labor costs.

References

- [1] Myers B A. A brief history of human-computer interaction technology [J]. *interactions*, 1998, 5(2): 44-54.
- [2] Li C, Yang J, Chang S. Review on key technologies of space intelligent grasping robot [J]. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 2022, 44(2): 64.
- [3] Chen Kuiye; Shi Xuhua; Research on robotic arm grasping method based on ensemble learning [A]; Summary of the 31st China Process Control Conference (CPCC 2020) [C]; 2020
- [4] Silwal A, Davidson J R, Karkee M, et al. Design, integration, and field evaluation of a robotic apple harvester [J]. *Journal of Field Robotics*, 2017, 34(6): 1140-1159.
- [5] Li Y, Zhang Y. Application research of computer vision technology in automation [C]//2020 International Conference on Computer Information and Big Data Applications (CIBDA). IEEE, 2020: 374-377.
- [6] Ma Z, Zeng Y, Zhang L, et al. The Workpiece Sorting Method Based on Improved YOLOv5 For Vision Robotic Arm [C]//2022 IEEE International Conference on Mechatronics and Automation (ICMA). IEEE, 2022: 481-486.
- [7] Brosnan T, Sun D W. Inspection and grading of agricultural and food products by computer vision systems—a review [J]. *Computers and electronics in agriculture*, 2002, 36(2-3): 193-213.
- [8] Shaoqing Ren et al., "Faster r-cnn: Towards real-time object detection with region proposal networks", *Advances in neural information processing systems*, vol. 28, 2015.
- [9] Wei Liu et al., "Ssd: Single shot multibox detector", *European conference on computer vision*, pp. 21-37, 2016.
- [10] Jiang, J., Luo, X., Luo, Q. et al. An overview of hand-eye calibration. *Int J Adv Manuf Technol* 119, 77–97 (2022).
- [11] Liu C, Wang X, Jiang H, et al. Inverse Kinematics Solution of Manipulator Based on IPSO-BPNN[C]//2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI). IEEE, 2022: 175-179.
- [12] Shen H, Xie W F, Tang J, et al. Adaptive Manipulability-Based Path Planning Strategy for Industrial Robot Manipulators [J]. *IEEE/ASME Transactions on Mechatronics*, 2023
- [13] Zhao Y, Li H. POSITIONING AND GRABBING TECHNOLOGY OF INDUSTRIAL ROBOT BASED ON VISION [J]. *Academic Journal of Manufacturing Engineering*, 2019, 17(3).
- [14] Kakani V, Nguyen V H, Kumar B P, et al. A critical review on computer vision and artificial intelligence in food industry [J]. *Journal of Agriculture and Food Research*, 2020, 2: 100033.