Design and Implementation of Intelligent Truck Based on Azure Kinect

Jingfang Wei[1], Kaiyang Xu[1], Zilin Hu[2], Yuji Iwahori[3], Haibin Wu[1](✉), and Aili Wang[1]

[1] Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application, Harbin University of Science and Technology, Harbin 150080, China
woo@hrbust.edu.cn

[2] City University of Hong Kong, Hong Kong 610200, China

[3] Computer Science, Chubu University, Kasugai 487-8501, Aichi, Japan

Abstract. In recent years, due to the impact of COVID-19, the market prospect of non-contact handling has improved and the development potential is huge. This paper designs an intelligent truck based on Azure Kinect, which can save manpower and improve efficiency, and greatly reduce the infection risk of medical staff and community workers. The target object is visually recognized by Azure Kinect to obtain the center of mass of the target, and the GPS and Kalman filter are used to achieve accurate positioning. The 4-DOF robot arm is selected to grasp and transport the target object, so as to complete the non-contact handling work. In this paper, different shapes of objects are tested. The experiment shows that the system can accurately complete the positioning function, and the accuracy rate is 95.56%. The target object recognition is combined with the depth information to determine the distance, and the spatial coordinates of the object centroid are obtained in real time. The accuracy rate can reach 94.48%, and the target objects of different shapes can be recognized. When the target object is grasped by the robot arm, it can be grasped accurately according to the depth information, and the grasping rate reaches 92.67%.

Keywords: Automatic Collection · Template Matching · RGB-D · Precise Positioning

1 Introduction

Transportation has always been an indispensable work in various industries. The transportation vehicle can not only solve the manpower problem, but also improve the work efficiency. With the continuous development of intelligent trucker technology, the market is far from saturated and the development potential is huge. Especially in the post epidemic era so far, as far as the reported sources of infection in China are concerned, food contact infection and express contact infection are the main ways of infection at present. Therefore, non-contact handling is the market demand that arises as the times require. With the normalization of nucleic acid testing, how to learn to fight a protracted

X. Jiang (Ed.): MLICOM 2022, LNICST 481, pp. 50–62, 2023.
https://doi.org/10.1007/978-3-031-30237-4_5

war against the epidemic is a very critical issue. Sufficient sampling materials are the premise to ensure the rapid and orderly work of sampling points, and also the work that wastes the most mobile manpower. In order to ensure the safety of medical personnel and community workers, it is often necessary to carry out multiple alcohol disinfection and sealed packaging of materials, which is time-consuming and laborious, as shown in Fig. 1. At this time, non-contact handling can just solve this problem. Therefore, this paper designs an intelligent truck based on Azure Kinect.

a) Nucleic acid sampling tube transport

b)Non contact distribution of fruits and vegetables

Fig. 1. Handling application scenario

In terms of the realization of the transport vehicle, Zou Jun designed a transport robot, which uses UWB and ultrasonic sensors to collect environmental information, and selects fuzzy algorithm to achieve effective obstacle avoidance [1]. Zhang Jun designed a navigation and positioning system by using GPS technology and established a navigation and positioning correction system, so that the transport robot can realize the autonomous following function [2].

In addition, Kalman filter technology has developed rapidly in the application of GPS. Zhu Zhongxiang used dual fuzzy Kalman filter for GPS/IMU/MV sensor information fusion to navigate autonomous tractors [3]. Choi et al. proposed that a vision sensor should be installed above the robot arm to sense the three-dimensional environment information, and obtain the position to be reached by the robot, so as to complete the identification and grasping of the target object [4]. Rosenberger et al. proposed a human-computer interaction method using real-time robot vision to realize the capture and control of target objects through object detector, fast capture selection algorithm and RGB-D camera [5]. In recent years, the target detection algorithm has also been widely opened and applied. Zhang Zhanpeng's team proposed an improved training network to increase the utilization rate of training data and achieve a 94% success rate [6]. Guo Di of Tsinghua University realized the detection of object grasping order in the complex environment of overlapping objects [7].

Azure Kinect provides developers SDKs with four kinds of sensors including depth, vision, sound and direction, through one million pixel TOF depth camera and 12 million pixel RGB high-definition camera. The color camera obtains high-definition RGB images and video information. The infrared emitter projects pulse light to objects within a certain range, receives pulse light through the depth camera, and creates a depth image within a certain range [8]. Therefore, we designed an intelligent truck based on Azure Kinect for the three-dimensional information of the target object can be accurately obtained. Firstly, GPS is combined with Kalman filter algorithm to realize real-time accurate positioning of the target and assist the accurate movement of the car. The Azure Kinect camera acquires depth information, combines template matching algorithm to intelligently identify and determine the distance of the nucleic acid sampling tube, and marks the spatial position of its centroid. The third function is to realize the accurate grasping and placement of the target object through the robot arm module.

2 System Software Design

First, the location of the captured target is obtained through Bluetooth communication, and the positioning is updated in real time by combining Kalman filter algorithm, and the data is transmitted to the GPS module, so that the GPS module can search for the target. Use Azure Kinect to intelligently identify and determine the distance of the target object through template matching algorithm and depth information, draw a rectangular box and mark the centroid, and obtain the centroid space coordinates in real time. Drive the robot arm to grasp the target object accurately. The functional flow chart is shown in Fig. 2.

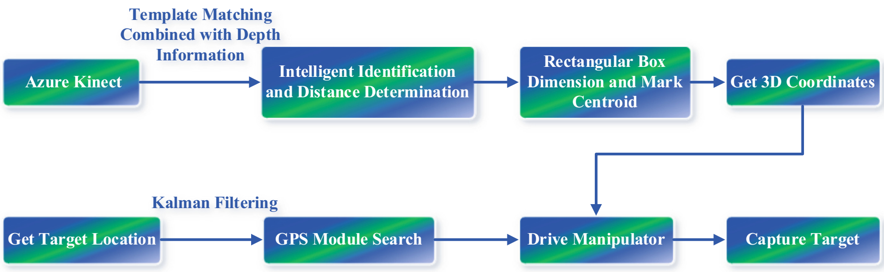


Fig. 2. The functional flow chart of software system

2.1 Accurate Positioning Model Based on Kalman Filter Algorithm

The communication between Arduino and GPS module mainly includes positioning, navigation and path planning of GPS module. In order to improve the positioning accuracy, Kalman filter algorithm is adopted to make the optimal estimation of the state of the system, which carries out an iterative cycle, and the minimum value of the mean square error can be obtained, that is, the optimal estimation value. It is assumed that $\hat{X}_{k-1}$ is the optimal estimation value of the state quantity X_{k-1} is obtained when the filtering algorithm operates at the $k-1$ time. With the aid of the input data u_{k-1} and the measurement data y_k, the optimal estimation value X_k of the Kalman filter for the system state X_k at the kth time is derived.

The prediction process is as follows:

$$\hat{X}_{k,k-1} = \Phi_{k,k-1}\hat{X}_{k-1} + Bu_{k-1} \tag{1}$$

$$P_{k,k-1} = \Phi_{k,k-1}P_{k-1}\Phi^{T}_{k,k-1} + Q_{k-1} \tag{2}$$

where $\hat{X}_{k-1}$ is the state vector at time $k-1$, $\hat{X}_{k,k-1}$ is the state vector at time k, B is the additional transition matrix, and P is the covariance matrix of the state vector.

The update process is as follows:

$$K_k = P_{k,k-1}H_k^T\left(H_kP_{k,k-1}H_k^T + R_k\right)^{-1} \tag{3}$$

$$\hat{X}_k = \hat{X}_{k,k-1} + K_k\left(Z_k - H_k\hat{X}_{k,k-1}\right) \tag{4}$$

$$P_k = (I - K_kH_k)P_{k,k-1} \tag{5}$$

This process represents the correction of the actual value to the predicted value, K is the gain matrix and I is the unit diagonal matrix.

2.2 Intelligent Recognition and Distance Measurement Model Based on Template Matching Algorithm and Depth Information

In real life, object detection can be used for 3D matching, image understanding, image compression, etc. At present, two new research ideas are gradually emerging, namely, significant target detection in RGB and RGB-D image. Compared with traditional RGB image data, depth image can also complete functions in three dimensions that cannot be completed in two dimensions [9]. Due to the accuracy problem of Kinect camera, image noise is unavoidable when obtaining data. Noise has a considerable impact on image recognition, especially on video image processing. Frequent noise will cause unstable recognition results. Therefore, the image is denoised before image recognition [10]. One of the methods that can be considered is median filtering, which is beneficial to retain edge information and remove speckle noise and salt and pepper noise [11]. At the same time, we can analyze a single target image or multiple targets with different perspectives at the same time [12]. The RGB-D image classification process is shown in Fig. 3. The target background segmentation calculation method based on depth is similar to the method based on color.

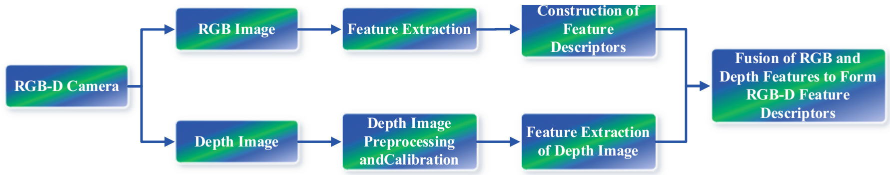


Fig. 3. RGB-D image classification method

$$P(x_i^t \in f_g) = \frac{P\left(x_i^t \in f_g|\mathrm{depth}(x_i^t), h_{fd}^{(t-1)}\right)}{P\left(x_i^t \in f_g|\mathrm{depth}(x_i^t), h_{fd}^{(t-1)}\right) + P\left(x_i^t \in b_g|\mathrm{depth}(x_i^t), h_{bd}^{(t-w1)}\right)} \quad (6)$$

The depth histograms of target and background in the depth image sequence frame $t-1$ are expressed as $h_{fd}^{(t-1)}$ and $h_{bd}^{(t-1)}$ respectively. The Formula 6 calculates the possibility that each pixel in the depth image sequence frame belongs to the target.

In the target detection, this paper uses the image processing technology to process the collected color image, in order to reduce the detection error, and finally converts the collected image into a binary black-and-white image, reducing the impact of environmental factors such as light and background on the detection results during the image acquisition process [13]. This algorithm is template matching, which is a key part of the image processing process. The process of finding and locating the target position in another image according to the known target template map is called template matching. See formula (7) for the calculation method.

$$R(x, y) = \sum_{x',y'} \left(T(x', y') - I(x + x', y + y')\right)^2 \quad (7)$$

where $T(x', y')$ is the template image matrix and $I(x, y)$ is the original image matrix.

2.3 Robot Arm Program Design of Grab Control

By adjusting the movement trajectory of the robot arm and the closing angle of the robot claw, the robot claw can grasp the target object. The specific flow chart is shown in Fig. 4.

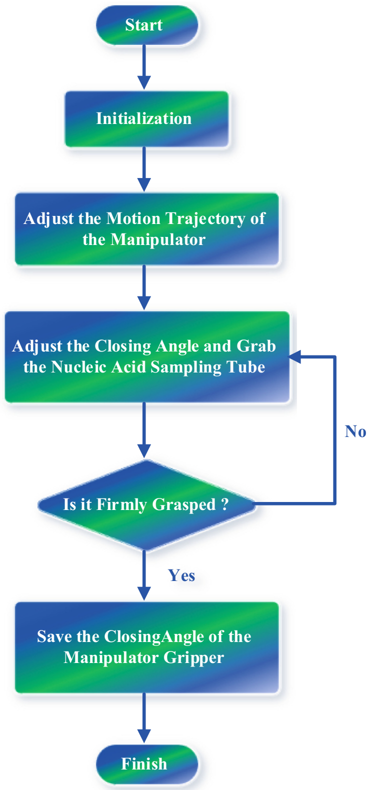


Fig. 4. The flow chart of Mechanical arm module

2.4 Function Programming of Intelligent Search System

The GPS module and the Bluetooth module are used to guide the smart car, so that the smart car can reach the position of the target object. When programming, according to the modular programming idea, each module generates independent sub functions for the main program to call. The program flow chart is shown in Fig. 5.

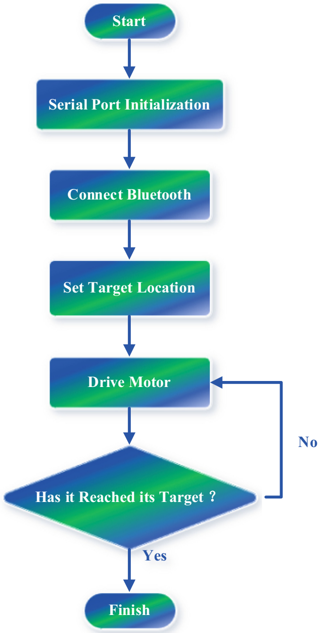


Fig. 5. GPS precise positioning flow chart

2.5 Design Intelligent Identification and Distance Measurement

The Azure Kinect camera needs to be calibrated during the application process to obtain the real position information of the measured object in the three-dimensional space. The flow is shown in Fig. 6.

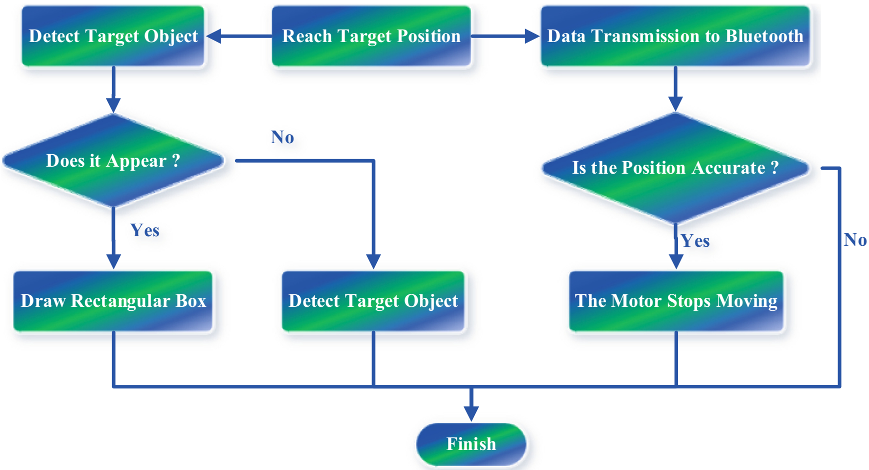


Fig. 6. The flow chart of intelligent identification and distance measurement

3 Hardware Design of the System

The overall scheme design flow chart is shown in Fig. 7.

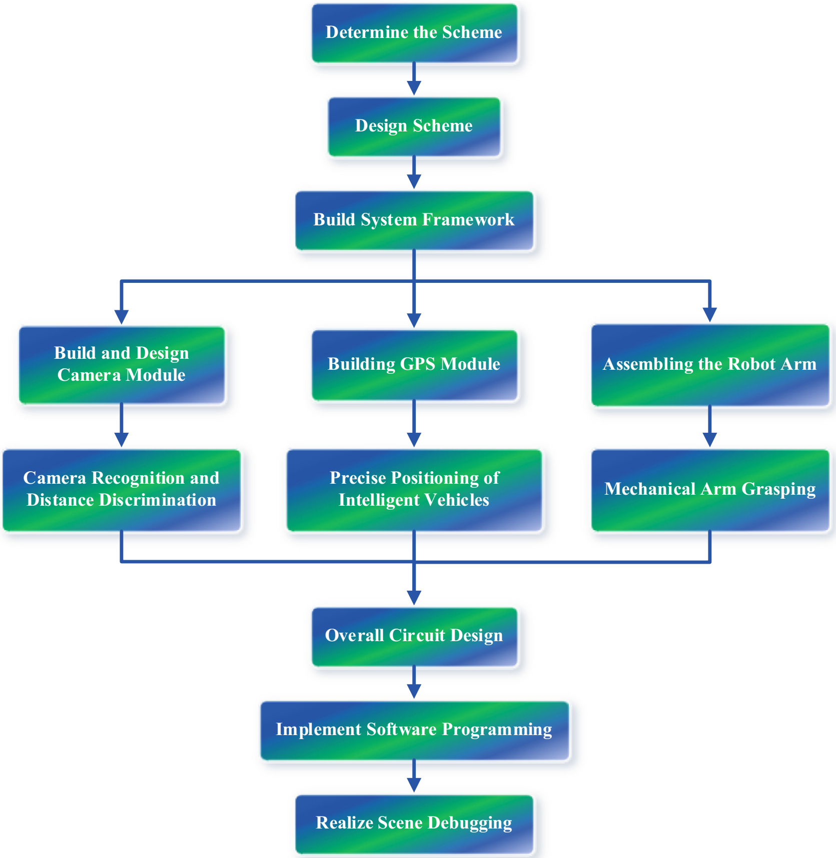


Fig. 7. Overall design scheme of intelligent truck based on Azure Kinect.

The main control module uses Arduino Mega2560 as the control chip, which has 54 channels of digital input and output and is suitable for the design requiring a large number of IO interfaces. It can supply power through three modes, and can automatically select the power supply mode. The GPS module adopts the ATGM332D-5N series, which has 32 tracking channels. Its advantages are high sensitivity, low power consumption and low cost. It is suitable for vehicle navigation, handheld positioning and wearable devices, which can directly replace the U-blox NEO series.

This article uses Azure Kinect DK, which was launched by Microsoft in February 2019, as the camera module to obtain data. Compared with previous products, Azure Kinect DK supports higher resolution for both RGB camera and depth camera. It is more compact and lightweight, and more suitable for vehicle applications. The robot arm module is mainly used to grasp the target object. The opening of the claw can reach 225 mm, the length of the robot arm is 401 mm, and the weight of the robot arm is 460 g. It is applicable to all kinds of bionic robot joints, so it can be used in multiple scenarios under the same system. The hc06 Bluetooth module is adopted in this paper, which can be downward compatible with even numbered versions. The host can automatically search the slave devices for pairing, and the adapter can support pairing communication with multiple modules.

4 System Commissioning

The hardware peripheral circuit is composed of various data interfaces of Arduino Mega2560, and the above functions are realized by corresponding software programs, as shown in Fig. 8.

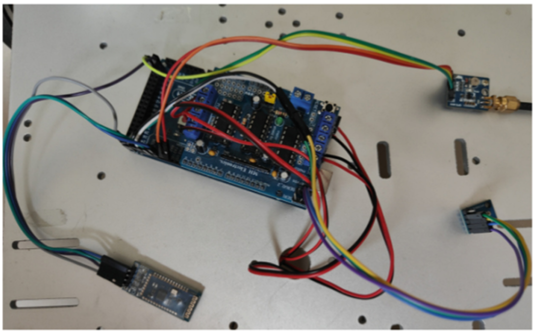


Fig. 8. Design physical drawing.

At the same time, The truck needs to be equipped with computer, and considering its load-bearing and balance, this paper chooses to assemble an aluminum alloy intelligent truck. The upper layer of the car is placed with a computer and a robot arm, and the lower layer is placed with Arduino mega2560 and Azure Kinect, as shown in Fig. 9.

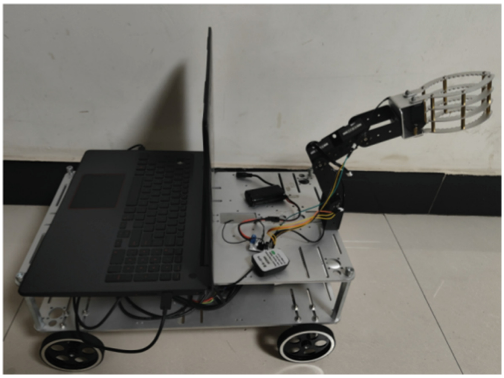


Fig. 9. Physical image of intelligent truck.

4.1 GPS Precise Positioning Test

When the GPS module is used to locate the car, the received data is analyzed, mainly to extract the longitude and latitude in the GPS module. The specific method is to separate the data according to the comma between the data, extract the longitude and latitude data, and then convert the data to obtain the results. Results as shown in Fig. 10, the positioning function can be accurately completed with an accuracy rate of 95.56%.

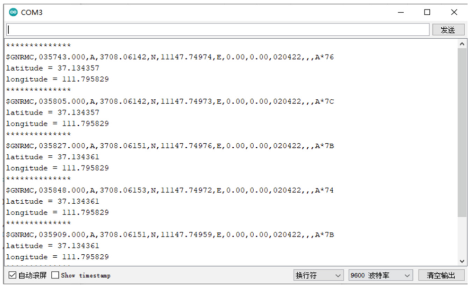


Fig. 10. GPS module of data reception.

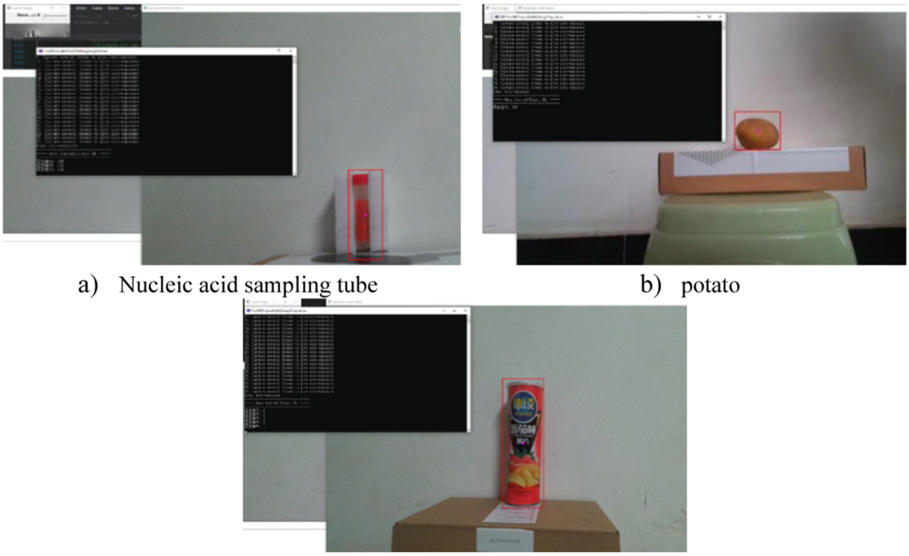
4.2 Experiments of Target Recognition and Distance Measurement

In this paper, we use nucleic acid sampling tube, potato and potato bucket to simulate three scenarios. The results show that when the target object appears in the picture, it can be recognized. After many experiments, it is found that the recognition rate can reach 94.48%, which meets the actual needs. The recognition results are shown in Fig. 11, and the obtained centroid space coordinates are shown in Table 1.

4.3 Grasping Accuracy of Manipulator

When the target object is grasped by the robot module, the design of the control method of the robot arm must meet the physical characteristics of the robot arm and the constraint conditions. Otherwise, the robot arm will be seriously damaged and even threaten the personal safety of the user. The grasping control using the robot arm is shown in Fig. 12.

In the experiment, the robot arm is used to grasp the target object, and the rotation angle of the steering gear is recorded, that is, the closing angle of the mechanical claw. After many experiments, when the closing angle of the robot arm is 30°, it can firmly grasp the target object, and the grasping rate reaches 92.67%.



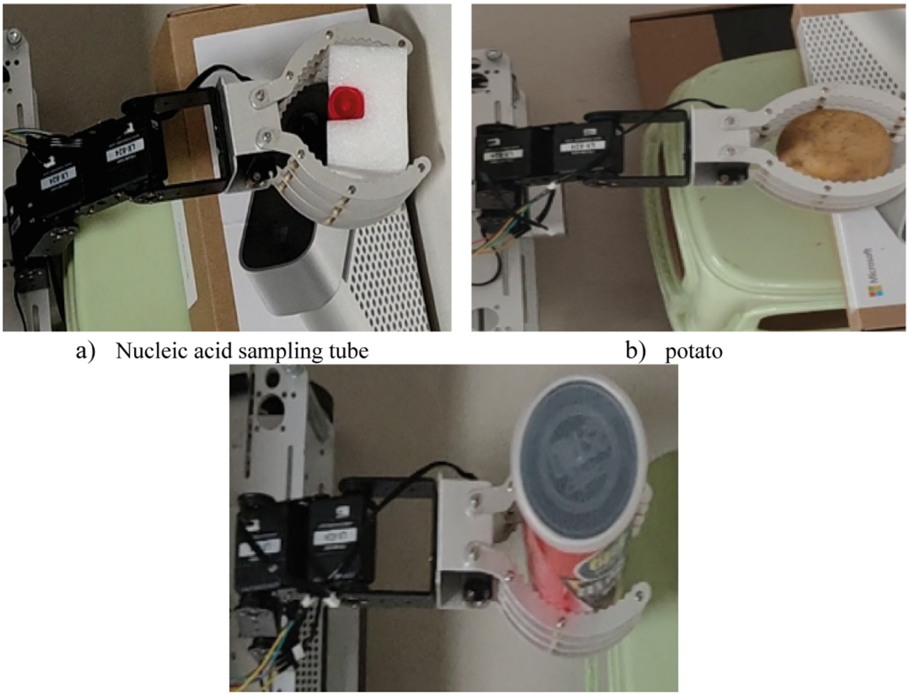
a) Nucleic acid sampling tube b) potato

c) Crisps bucket

Fig. 11. Intelligent recognition and distance measurement experiment.

Table 1. Centroid space coordinates.

Number	Nucleic acid sampling tube	Potato	Crisps bucket
1	(313,243,177)	(275,211,167)	(293,223,170)
2	(314,243,179)	(273,210,166)	(295,225,166)
3	(314,243,178)	(275,213,168)	(292,224,169)
4	(314,243,179)	(274,211,169)	(291,221,171)
5	(317,243,178)	(276,212,165)	(296,220,173)
6	(313,243,181)	(277,210,167)	(295,226,165)
7	(315,243,179)	(275,216,164)	(297,220,172)
8	(314,242,177)	(276,210,167)	(295,219,170)
9	(314,241,178)	(271,211,166)	(296,223,171)
10	(312,241,179)	(273,213,167)	(295,221,170)



a) Nucleic acid sampling tube b) potato

c) Crisps bucket

Fig. 12. Grasping control experiment of manipulator.

5 Conclusions

This paper has completed the design of an intelligent truck based on Azure Kinect. Using GPS and Kalman filter algorithm to obtain the position of the target object, and the accuracy rate is 95.56%. Azure Kinect camera obtains depth information, combines template matching algorithm to intelligently identify and determine the distance of the target object, and marks the spatial position of its centroid, with an accuracy rate of 94.48%. When the target object is grasped by the robot arm, it can be grasped accurately according to the different depth information, and the grasping rate reaches 92.67%. It can complete contactless material transfer, improve work efficiency, reduce the burden of staff, reduce the risk of staff infection, and meet the needs of epidemic prevention.

References

1. Zou, J.: Research and development of intelligent robot tracking system. Huazhong University of Science and Technology (2018)
2. Zhang, J.: Research on autonomous tracking control system of tracked agricultural mobile robot. Zhejiang University of Technology (2017)
3. Liu, Y.: Application of Kalman filter in GPS dynamic positioning. Huazhong University of Science and Technology (2009)

4. Choi, C., Taguchi, Y., et al.: Voting-based pose estimation for robotic assembly using a 3D sensor. In: IEEE International Conference on Robotics and Automation, pp. 1724–1731. IEEE (2012)
5. Rosenberger, P., Cosgun, A., Newbury, R., et al.: Object-independent human-to-robot handovers using real time robotic vision. J. IEEE Robot. Autom. Lett. **1**, 17–23 (2016)
6. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: International Conference on Artificial Intelligence and Statistics, pp. 315–323 (2011)
7. Guo, D.: Research on target detection and grasping planning for robot operation. Tsinghua University (2016)
8. Yang, P.: Research on gesture recognition based on Kinect and its application in virtual simulation system. Nanjing University (2018)
9. Kuang, H., Cai, S., Ma, X., et al.: An effective skeleton extraction method based on Kinect depth image. In: International Conference on Measuring Technology & Mechatronics Automation. IEEE (2018)
10. Ding, Y.: Research on identification and positioning methods of automatic fruit picking. Shenyang University of Technology (2020)
11. Fu, Y.: Research on humanoid robot motion and algorithm driven by depth camera data. Zhejiang Normal University (2019)
12. Kato, H., Harada, T.: Learning view priors for single-view 3D reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, State of California, pp. 9778–9787(2019)
13. Yang, M.: Human motion recognition and tracking robot based on image processing. Chang'an University (2019)