

Research Article

Target Positioning and Grasping of NAO Robot Based on Monocular Stereo Vision

Hongbin Chen ^{1,2}

¹Automation School, Beijing University of Posts and Telecommunications, Beijing 100876, China

²School of Information Engineering, Beijing Institute of Graphic and Communication, Beijing 100260, China

Correspondence should be addressed to Hongbin Chen; chb@bigc.edu.cn

Received 27 August 2021; Revised 26 November 2021; Accepted 6 December 2021; Published 6 January 2022

Academic Editor: Sang-Bing Tsai

Copyright © 2022 Hongbin Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous advancement of science and technology and the rapid development of robotics, it has become an inevitable trend for domestic robots to enter thousands of households. In order to solve the inconvenience problem of the elderly and people with special needs, because the elderly and other people in need may need the help of domestic robots due to inconvenient legs and feet, the research of the robot target position based on monocular stereo vision and the understanding of the robot NAO are very important. Research and experiments are carried out on the target recognition and positioning in the process of NAO robot grasping. This paper proposes a recognition algorithm corresponding to quantitative component statistical information. First, extract the area of interest that contains the purpose from the image. After that, to eliminate interference in various fields and achieve target recognition, the robot cameras have almost no common field of view and can only use one camera at the same time. Therefore, this article uses the monocular vision principle to locate the target, and the detection algorithm is based on the structure of the robot head material, establishes the relationship between the height change of the machine head and the tilt angle, and improves the monocular vision NAO robot detection algorithm. According to experiments, the accuracy of the robot at close range can be controlled below 1 cm. This article completes the robot's grasping and transmission of the target. About 80% of the external information that humans can perceive comes from vision. In addition, there are advantages such as high efficiency and good stability.

1. Introduction

1.1. Background. With the development of robot-related technologies, robots are widely used in production and life: smart home service robots, industrial robots, and other medical robots and entertainment robots. The emergence of smart home service robots, on the one hand, meets the needs of the market; on the other hand, according to the change of science and technology from the information age to the intelligent age, new products are born in new fields. Grasping the goal in the home environment is an action that smart robots must often complete one. The recognition and positioning of the target object are a prerequisite for the robot to successfully understand the target object. Although smart home service robots are developing rapidly, most

home service robots cannot be used well to help the elderly and those in need with services, and they are still in their infancy. Regarding the grasping task of the intelligent room service robot, it cannot be understood without target recognition and location. Therefore, it is necessary to carry out the identification and positioning of the target for the purpose of grabbing.

1.2. Significance. It is very important to apply vision to intelligent robots so that the robot can obtain external information through visual sensors. The integration of applied technology represents the country's high-tech and industrial modernization level. Compared with the research and development of robots made by foreigners, the research of imitating

human robots in my country is relatively backward. There is still a big gap with overseas. Aldebaran robots represent the pioneering development of world-class technology in the past five years. In July 2007, the robot NAO replaced Aibo. Sony's robot dog was selected as the "standard platform group" by the World Cup Organizing Committee. NAO is a biped humanoid robot that is used throughout the global education market. The 58 cm tall NAO has the same natural body language as humans and is able to listen, see, speak, and interact with people or interact with other NAOs. Since then, the NAO robot has stood on the stage of history. Many students and researchers at home and abroad have conducted a lot of research on this platform. Most of them are concentrated in the RoboCup football match of the RoboCup World Cup. NAO robots have relatively few applications in domestic service. Based on the same movement mode as similar appearance features, the NAO robot is used as a platform to study the behavior of imitating human robots, helping people to bring objects into their daily lives. It has a very important meaning. Life can be applied to future home service robots, such as caring for the elderly, doing simple things for the disabled, and carrying luggage.

1.3. Related Work. The human auditory system of the machine is a way of interacting with the outside through sound. Compared with the visual interaction of robots, due to the diffractive nature of sound, immediate vision is not required. In the case of visual impairment and visual differences, the way of interaction through sound has its unique advantages. Cui et al. propose a dual-prism-based monocular stereo vision system that can shoot different perspectives of the same object in a single shot. Compared with the traditional two-camera system, it has several advantages. In order to measure the position or restore the shape of the object, a valid image pair must first be captured, which depends on the field of view of the system. In this article, we propose a general method for establishing a practical dual-prism-based monocular stereo vision system. The relationship between system parameters and object distance is analyzed in detail. The standard of parameter optimization and the process of system setting are introduced, but the site of the system will change a lot [1]. Wu et al. pointed out that, based on the binocular vision theory, a swing monocular stereo vision measurement method was proposed and a suitable measurement system was established. Determine the field of view of each part of the system. The swing arm drives the camera to complete shooting at two different positions. Choose Zhang Zhengyou method for system calibration. Canny operator and Hough operator are used to complete the contour detection, and the matching points of the image are obtained. The least square method is used to obtain the coordinates of the intersection of the rectangles, and the space circle optimization method is used to obtain the diameter of the circle, so as to realize the measurement of the experimental sample and restore the 3D model. The results show that the detection accuracy of the system is above 0.6%, but there are also large errors [2]. Bai et al. pointed out that the single-eye vision pedestrian detection cannot obtain depth information, so the detection efficiency

and accuracy are limited. First, in the image preprocessing, the optimization analysis of disparity information is proposed to simplify the expression of dynamic programming Stixel-world in complex scenes based on stereo vision. Then, in the pedestrian target detection stage, the influence of the block size on the traditional HOG feature detection effect is analyzed, and the Fisher criterion is used to obtain multiple HOG features suitable for the road environment. The multi-HOG function is integrated with the LUV color channel function. Finally, hik-SVM is used for pedestrian target classification. Experimental results show that the improved Stixel-world image preprocessing algorithm greatly reduces the calculation time and reduces the candidate area for pedestrian detection. The target detection algorithm based on feature fusion and hik-SVM has good real-time and robustness, but the detection and the accuracy are not particularly high [3].

1.4. Main Content. The research of this subject is based on the NAO robot platform to realize the task of grasping and transferring the target of the service robot, recognize the target location in the robot coordinate system, describe the grasping path of the robot, and control the gripping behavior of the arm. Finally, according to the principle of dynamics, the robot will track the direction of the human sound source and move towards the human after receiving the successful feedback information. The main research content is the target recognition and position recognition of the single vision mobile robot system, including the recognition and detection of color targets, camera calibration, monocular vision ranging, and robot mobile control strategy [4]. The article will elaborate on the core ideas and characteristics of these related contents and put forward their own improvement ideas for the specific experimental system of this article. In this paper, on the experimental platform of the NAO mobile robot, the target recognition and positioning of the mobile robot system based on monocular vision are realized. Based on the background of the above problems, this thesis conducts research on the target object recognition and positioning and robot grasping system on the NAO robot. Establish an integrated system based on the monocular vision detection system and the robot motion control system that can accurately realize the robot's grasping tasks.

2. Positioning and Grabbing Method Based on Monocular Vision

2.1. Introduction to NAO Robot Software. NAOqi OS is the operating system of NAO V5 [5, 6]. It is an embedded GNU\Linux development kit specially developed for Aldebaran robots. NAOqi is the main program running on the robot. It completes the overall control of the NAO robot. It can run directly on the robot or on a personal computer so that the program can be tested on the NAO virtual machine.

2.2. Threshold Segmentation Method in Target Detection. In actual situations, sometimes only a part of the target area in the result obtained by background subtraction is what we

need. Therefore, we need to use a certain method to extract the area we want. Threshold segmentation is a good method. Threshold segmentation is a very simple, effective, and commonly used segmentation algorithm [7]. The so-called threshold segmentation is to segment the region with a certain range of gray values in the image, which is mainly used when there is a large gray value difference between the background and the object. The mathematical expression of threshold segmentation is as follows:

$$F = \{(r, c) \in R | g_{\min} \leq S \leq g_{\max}\}. \quad (1)$$

The operation process is to output the points in the image area R whose gray value is between g_{\min} and g_{\max} to the area S .

Threshold segmentation is divided into three situations:

- (1) Fixed threshold segmentation: the minimum threshold g_{\min} and maximum threshold g_{\max} can remain unchanged without being adjusted. This method is suitable for stable external conditions and there is a very significant grayscale difference between the segmented object and the background.
- (2) Automatic threshold segmentation: this algorithm is based on the grayscale histogram of the image. This method is more suitable when the grayscale histogram of the image has obvious double peaks. Figure 1 shows the collected object image, Figure 2 shows the grayscale histogram of the image, the first peak corresponds to the background image, the second peak corresponds to the foreground image, and the minimum value between the two peaks is the threshold. However, in practical applications, if the illumination is uneven, causing the histograms to be scattered and there is no obvious double peak, this method will fail.
- (3) Dynamic threshold segmentation in a local area: the area of interest is always much brighter or darker than the background image. According to this rule, dynamic threshold segmentation is to compare the image with its local background.

The dynamic threshold segmentation of bright objects is processed as follows:

$$F = \{(r, c) \in R | S_{r,c} - g_{r,c} \geq g_{\text{diff}}\}. \quad (2)$$

The dynamic threshold segmentation processing for dark objects is as follows:

$$F = \{(r, c) \in R | S_{r,f} - g_{r,f} \leq -g_{\text{diff}}\}. \quad (3)$$

The method used in this paper is the dynamic threshold segmentation method. The threshold segmentation of Figure 1 is performed, and the result is shown in Figure 2.

It can be seen from Figure 2 that the area obtained by dynamic threshold segmentation usually contains some unwanted areas. Since the gray value of these areas is very close to the gray value of the target area, this is the case when performing threshold segmentation. Part of the area is selected. In this case, if we want to get the result we want, we need to adjust

the shape of the segmented area. This is the application of morphology. These operations mainly include corrosion, expansion, opening operation, and closing operation.

2.3. Camera Calibration

2.3.1. Camera Parameters. Camera parameters are divided into internal parameters and external parameters. The most common fourth internal parameter model [8, 9] is the x -axis magnification factor $1/d_x$ in the x -axis direction, the y -axis magnification factor $1/d_y$, the intersection of the camera's optical axis and the imaging plane, that is, the physical coordinates of the image (u_0, v_0) . The external parameters of the camera include the rotation matrix R , and the translation vector t . If the coordinates of point m in the camera coordinate system are (U, V) , then there is formula (4) to convert the coordinates (x_c, y_c, z_c) of the point M in the camera coordinate system to coordinates (u, v) in the pixel coordinate system. The global coordinate system and the camera coordinate system are transformed by the rotation matrix R and the vector t , and the coordinates in the global coordinate system of m are set to (X_w, Y_w, Z_w) . Let $F_x = f/DX$, $F_y = f/dy$. Combine (5) and (6) to obtain the conversion relationship equation (7) between the coordinates of a point in the world coordinate system and the coordinates in the image pixel of the point.

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}, \quad (4)$$

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}, \quad (5)$$

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}. \quad (6)$$

Let $M_{\text{in}} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$, $M = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$ get the following formula:

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M_{\text{in}} M_{\text{out}} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}. \quad (7)$$

The task of camera calibration is to find M_{in} and M_{out} . From the above relationship, we can see that, for the situation where the camera's internal and external parameters

have been obtained, a point (x_w, y_w, z_w) in the world coordinate system is known and then the above mapping relationship is used. The pixel coordinates (u, v) of the projection point of the point on the camera imaging plane can be solved, but according to the process of projection imaging, it can be known that when projecting from the three-dimensional plane to the two-dimensional, the ray is where the point in the world coordinate system and the optical center line are located. All the points on the projection are a point on the imaging plane; that is to say, when the above coordinate solution process is reversed from the pixel coordinate system, according to the known mapping relationship, the corresponding point in the world coordinate system cannot be obtained. Therefore, other constraints need to be added to determine the world coordinates of the corresponding point in the pixel coordinate system (u, v) in the world coordinate system.

2.3.2. The Distortion Model of the Camera. The small hole imaging model of the camera [10] is an ideal camera model. It is impossible for the actual lens to maintain such an ideal state during the manufacturing process, which causes the deformation of the actual object in the imaging plane. The distortion model of the camera is a modeling of the imaging error of the camera lens. The distortion of the camera mainly includes radial distortion, centrifugal distortion, and thin prism distortion.

(1) Radial Distortion. Radial distortion will cause errors in the radial direction of the camera during the projection process. It is a type of distortion that is symmetrical about the main optical axis of the camera. It is caused by the error in the curvature of the shape of the camera lens during production. The size of the error is determined by the radial displacement of the imaging point and the point in the ideal model, as shown in Figure 3. When the mapping point [7] is at a certain distance from the imaging center of the camera, the distortion becomes large. d_1 is the radial displacement, d_2 is the tangential displacement, and O is the imaging center.

Then its radial distortion can be expressed by E_{xr} and E_{yr} after ignoring the higher order. Its mathematical expression is as

$$\begin{cases} E_{xr} = k_1x(x^2 + y^2) + k_2x(x^2 + y^2)^2 + k_3x(x^2 + y^2)^3, \\ E_{yr} = k_1y(x^2 + y^2) + k_2y(x^2 + y^2)^2 + k_3y(x^2 + y^2)^3. \end{cases} \quad (8)$$

(2) Centrifugal Distortion. The centrifugal distortion is caused by the shift of the optical center. With the change of the focal length of the lens, it changes slightly. It includes both radial and tangential distortion. It is mathematically modeled to get formula (9), where p_1 and p_2 are distortion coefficients.

$$\begin{cases} E_{xd} = p_1(3x^2 + y^2) + 2p_2xy, \\ E_{yd} = p_2(3x^2 + y^2) + 2p_1xy. \end{cases} \quad (9)$$

2.3.3. Thin Prism Distortion. The thin prism error of the camera [11] is caused by the installation error of the camera. This kind of distortion also includes the already directed and tangential directions, and the distortion will become smaller as the focal length of the lens becomes larger. In the case of ignoring higher order, its mathematical model is as

$$\begin{cases} E_{xp} = s_1(x^2 + y^2), \\ E_{yp} = s_2(x^2 + y^2). \end{cases} \quad (10)$$

The final imaging of the camera is the result of the combined effects of the above distortions, but under normal circumstances, the high-order factors in the distortion have little effect on the distortion. Therefore, when considering the comprehensive modeling, the high-order factors can be ignored and the distortion of the camera can be expressed as

$$\begin{cases} E_x = k_1x(x^2 + y^2) + k_2x(x^2 + y^2)^2 + k_3x(x^2 + y^2)^3 \\ \quad + p_1(3x^2 + y^2) + 2p_2xy + s_1(x^2 + y^2) \\ E_y = k_1y(x^2 + y^2) + k_2y(x^2 + y^2)^2 + k_3y(x^2 + y^2)^3 \\ \quad + p_2(3x^2 + y^2) + 2p_1xy + s_2(x^2 + y^2). \end{cases} \quad (11)$$

If it is assumed that the physical coordinates of the ideal imaging are (x_i, y_i) , and the following conversion relationship exists:

$$\begin{cases} x_i = x \left[1 + k_1(x^2 + y^2) + k_2(x^2 + y^2)^2 + k_3(x^2 + y^2)^3 \right] \\ \quad + p_1(3x^2 + y^2) + 2p_2xy + s_1(x^2 + y^2) \\ y_i = y \left[1 + k_1(x^2 + y^2) + k_2(x^2 + y^2)^2 + k_3(x^2 + y^2)^3 \right] \\ \quad + p_2(3x^2 + y^2) + 2p_1xy + s_2(x^2 + y^2). \end{cases} \quad (12)$$

Under normal circumstances, since radial distortion has the greatest influence on imaging, other distortions can be ignored relatively speaking, so only the influence of radial distortion is considered. Let $x_2 + y_2 = r_2$ to express the coordinate after distortion correction as

$$\begin{cases} x_i = x(1 + k_1r^2 + k_2r^4 + k_3r^6), \\ y_i = y(1 + k_1r^2 + k_2r^4 + k_3r^6). \end{cases} \quad (13)$$

(1) Camera Calibration Method. When referring to the camera parameters above, the calibration of the camera is to obtain the values of M_{in} and M_{out} . Assuming that there are coordinates of a point in the three-dimensional space, according to the above model of camera hole imaging [12], it can be known that the projection process of the internal and external parameter matrix can be mapped to a point on the camera imaging plane. The internal parameters in the camera projection include some parameters related to the camera and the lens body. They are fixed for a vision system, while the external parameters of the camera can be changed

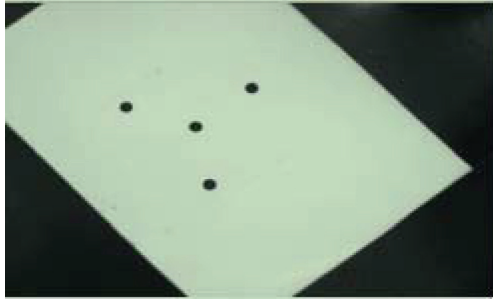


FIGURE 1: Dynamic threshold segmentation.

for the vision system. When the world coordinates are set, when the system moves, the external parameters of the camera will also be changed or updated. The process of obtaining these parameters mentioned above is the calibration process of the camera. For a machine vision system, the calibration of the camera is an indispensable and extremely important part. Whether it is a monocular system measurement system, a binocular positioning system, or a structured light positioning system, the calibration accuracy and its errors will be correct. Vision applications have a direct impact, even the decisive condition that determines the success of vision applications or whether they have application value.

The research of camera calibration methods has gone through a long history. The calibration block is assumed to be a rigid body with completely unchanged shape and size. Through the collection of the collected calibration block upper corner straight line, combined with the mutual constraint relationship between them, the solution equation is obtained, and then a certain method of nonlinear optimization for the solution equation is adopted. Obtain or solve the internal and external parameters of the camera. The method of calibrating fast calibration cameras has a wide range of applications, including various camera models, and it is easy to obtain higher calibration accuracy at a price. The disadvantage is that the calibration process assumes that the calibration block is a rigid body with completely unchanged size and shape, but the actual calibration process will be due to the temperature in the environment, manufacturing errors, and other factors, and the errors of the calibration block itself will be transmitted to the calibration process and reflected in the calibration results.

The Tsai two-step method and the Zhang Zhengyou calibration method are frequently used methods by researchers of vision applications. This method has become a classical calibration method. The linear fitting method is the least square method. For solving the radial distortion, this method uses three variables optimization method and obtains the camera internal parameters at the same time. Compared with the classic calibration method, Zhang Zhengyou's calibration method uses a checkerboard instead of a calibration block with a certain accuracy and known geometric parameters. In the first step of the solution, Zhang's calibration method also uses the linear method. The value obtained by the linear method is used as the initial value of the external parameter iteration.

After a multistep iterative process, the camera parameters with higher accuracy are solved. Zhang Zhengyou's calibration method has many advantages in solving camera parameters and distortion parameters. It is superior to other calibration methods in terms of solving speed and number of parameters. It is also one of the most commonly used calibration methods at present. Based on the above analysis, this article uses the Zhang Zhengyou calibration method to perform camera calibration. The calibration process requires the use of a calibration board to shoot several sets of checkerboard images at different angles and positions for calibration. The checkerboard image used is shown in Figure 4. Specify the number of rows and columns of the input checkerboard and the actual width and height of each checkerboard, and then perform calibration. The main process of calibration includes input image, corner extraction, equation construction, parameter calculation, least square method of parameter estimation, maximum likelihood method of parameter optimization, distortion parameter calculation, distortion correction, and input correction image [13, 14]. Compared with the direct linear calibration method, it has the advantages of simple operation and high calibration accuracy. The internal parameters of the camera and the external parameters of the reference coordinate system calibrated in a certain experiment in this paper are shown in Table 1.

3. Robot Motion Control Integration and Experiment

3.1. Reflex Manipulator Gripper. As shown in Figure 5, in the experiment, we fixed the Reflex manipulator claw at the end of the manipulator arm to grab the target object. The reflexive manipulator [15] is an intelligent manipulator based on Linux Ubuntu system, which is driven by four mx-28 servo motors and controlled by the robot operating system. It consists of three fingers with three degrees of bending (one finger for each finger) and one two finger coupling rotation degree of freedom. At the proximal end of the joint, there are 14 takfile sensors on the encoder grabbing surface, and each finger has a fingertip IMU. Table 2 shows the angle information that each joint can rotate.

3.2. Activating the Gripper

- (1) Use the following roslaunch command to start the hand grab: `roslaunch reflex_driver2 reflex_takktile2_driver.launch`. View real-time data of hand grabbing `rostopic echo/reflex_takktile2/hand_state.rosopic echo/reflex_takktile2/hand_state/motor.rosopic echo/reflex_takktile2/hand_state/finger`.
- (2) Calibrate the hand grip [16]: calibration means to set the current hand grip state to 0 state. Calibrate the sensor value of the hand and the value of the finger joints. Calibrate the sensor `rosservice call/reflex_takktile2/calibrate_tactile`. Calibrate the finger `rosservice call/reflex_takktile2/calibrate_finger`.

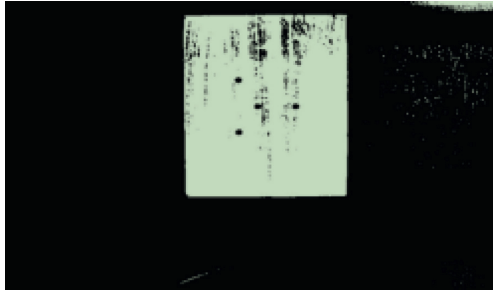


FIGURE 2: Grayscale histogram.

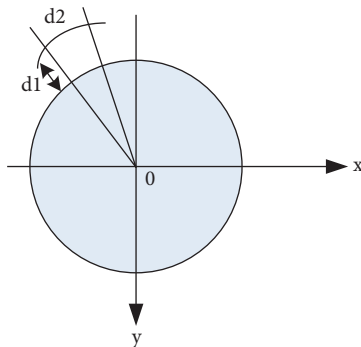


FIGURE 3: The distortion of the camera.

- (3) Basic opening and closing control of the claw: create a new package and enter in the workspace `src:catkin_create_pkg reflex_control reflex_driver2 reflex_msgs2 std_msgs roscpp rospy` and then `catkin_make`. This creates a new package called `reflex_control`.

The code for simple control of the manipulator claw has been written. `example-control.cpp` is mainly used to control the hand grip and realize the opening and closing of the hand grip by assigning values to four motors. Add the `example-control.cpp` file to the `src` of the newly created package, and add corresponding execution statements at the end of `cmakelist.txt`: `add_executable(example-control src/example-control.cpp)` and `target_link_libraries(example-control ${catkin_LIBRARIES})`. Execute `roscpp run reflex_demo first_example` to control the switch to close.

- (4) Robot D-H parameters: first, use the standard D-H modeling method [17, 18] to model the positive kinematics of the manipulator, and the D-H parameter table obtained is shown in Table 3.

3.3. Research on the Capture of the Target. In the experiment, three frames of trajectory were inserted in the NAO robot [19] dedicated software *Choregraphe* by using key frames to obtain the angle corresponding to the trajectory to complete the planning of the grasping trajectory. By fixing the grasping posture of the end effector, there is only one translation between the robot hand and the target position,

the unique value of d_1 , d_2 , d_3 , d_4 can be solved, and the corresponding angle of each joint of the robot can be controlled to complete the grasp. Take the process as shown in Figure 6.

Considering the camera depth, the measurement range of the camera [20] is 500 mm–4500 mm, and due to the limitation of joints, the robot will have some inaccessible space, so the object cannot be placed too close to the robot or the camera. First of all, we select five independent experimental areas on the experimental platform, which are all within the reach of the robot and the field of vision of the camera.

After the robot obtains the three-dimensional coordinate position of the target object in the robot base coordinate system, it can drive the robot to grab the object. According to the 5 areas selected in the experimental platform, in the experiment, we randomly select a few of the 9 types of objects and place them in the 5 areas, or place all of them in the area. There may be overlaps between the objects [21]. In the experiment, we let the robot grab the target object 6 times in each area; that is, the total number of grabbing experiments is $9 * 9 * 6 = 486$ times.

4. Positioning and Analysis of Crawling Results

4.1. Positioning Experiment Results. Comparing the errors in Figure 7, it can be seen that the closer the robot to the target, the smaller the ranging error. The reason is that the closer the robot camera is to the target, the closer the target contour extracted by the recognition algorithm [22] is to the actual contour. When the target is farther away, the extracted contour of the target object tends to have a larger error, which leads to the fact that the center position of the target obtained is different from the actual contour. The value deviation is large. Therefore, this paper adopts a combination of long-distance and short-distance positioning.

4.2. Results of the Object Grasping Experiment. Figure 8(a) shows the experimental success rate of 9 objects in 6 different regions. The highest experimental success rate of region 5 is 96%. Because it is in the middle position, the robot is easy to grasp and the image obtained by the camera is clearer. The lowest success rate in area 3 is 91%, because the robot in this area is prone to encounter the limit positions of joints when performing inverse kinematics calculations. Figure 8(b) shows the grasping success rate of 9 objects. Among them, the grasping rate of bases of different shapes is generally higher, and the grasping rate of the triangular prism is the lowest at 82%. The main reason for the low crawling success rate is that the triangular prism is small and there is less information that can be used for detection, and it is difficult to find the best crawling point. Therefore, the number of crawling experiments failed is relatively large. Based on the results of all grabbing experiments, the overall success rate of the grabbing system is about 92%. Experiments show that the system can detect the target object through monocular stereo vision, map the pixel coordinate value and depth value to the robot base

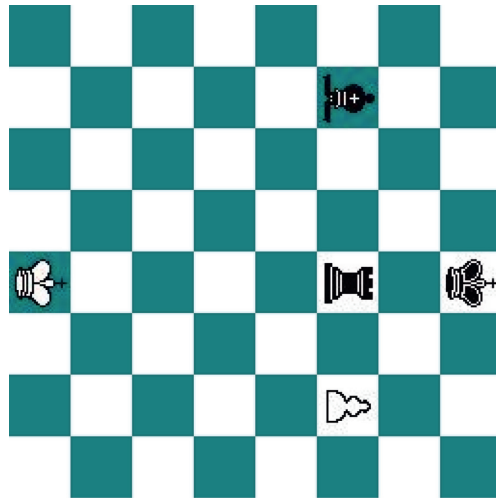


FIGURE 4: Checkerboard used in this article.

TABLE 1: Camera parameter calibration results.

Camera internal parameters	Numerical value	Camera external parameters	Numerical value
f (mm)	16.017	t_1	-98.740
dX (μm)	2.200	t_2	-103.010
dY (μm)	2.200	t_3	837.187
u_0	1204.590	r_1	2.210
v_0	1189.870	r_2	2.240
K_1	0.065	r_3	0.070
K_2	-6.470		
K_3	89.609		



FIGURE 5: Reflex robotic claw.

TABLE 2: Joint information.

Joint	Angle range (degrees)	Angle range (radians)
LShoulderPitch	-119.5 to 119.5	-2.0857 to 2.0857
LShoulderRoll	-18 to 76	-0.3142 to 1.3265
LElbowYaw	-119.5 to 119.5	-2.0857 to 2.0857
LElbowRoll	-88.5 to -2	-1.5446 to -0.0349
LWristYaw	-104.5 to 104.5	-1.8238 to 1.8238
LHand	Open and close	Open and close

TABLE 3: D-H parameters of Kuka iiwa.

Joint i	Joint angle q_i	Connecting rod corner a_i	Connecting rod offset d_i (m)	Link length a_i (mm)	Joint range l
1	q_1	$-\pi/2$	0.34	0	$-170\sim+170$ F02D
2	q_2	$-\pi/2$	0	0	$-120\sim+120$
3	q_3	$-\pi/2$	0.4	0	$-170\sim+170$
4	Q_4	$-\pi/2$	0	0	$-120\sim+120$
5	Q_5	$-\pi/2$	0.4	0	$-170\sim+170$
6	Q_6	$-\pi/2$	0	0	$-120\sim+120$
7	Q_7	0	0.125	0	$-175\sim+175$

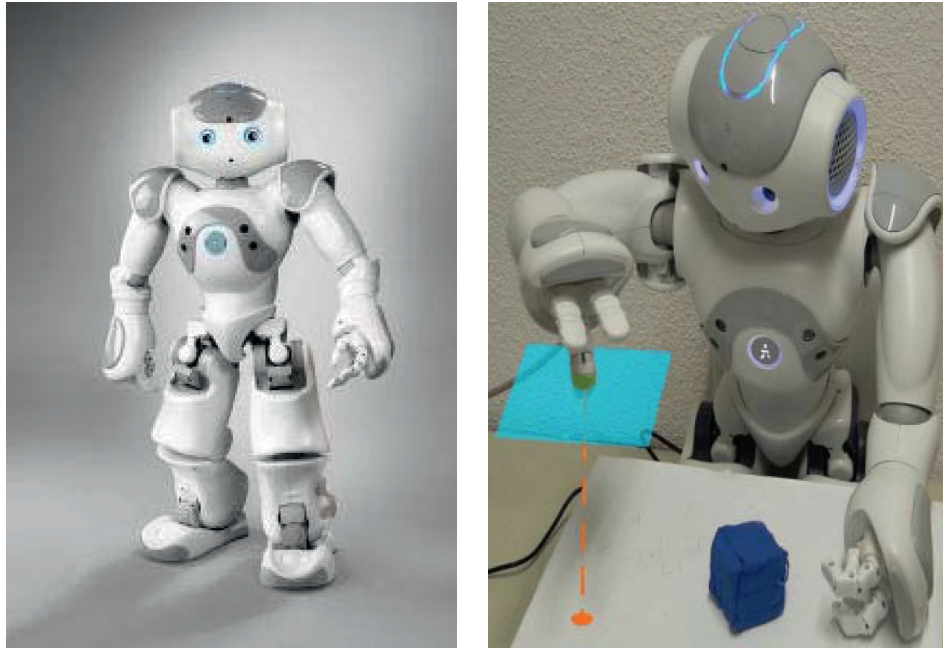


FIGURE 6: Robot grasping process.

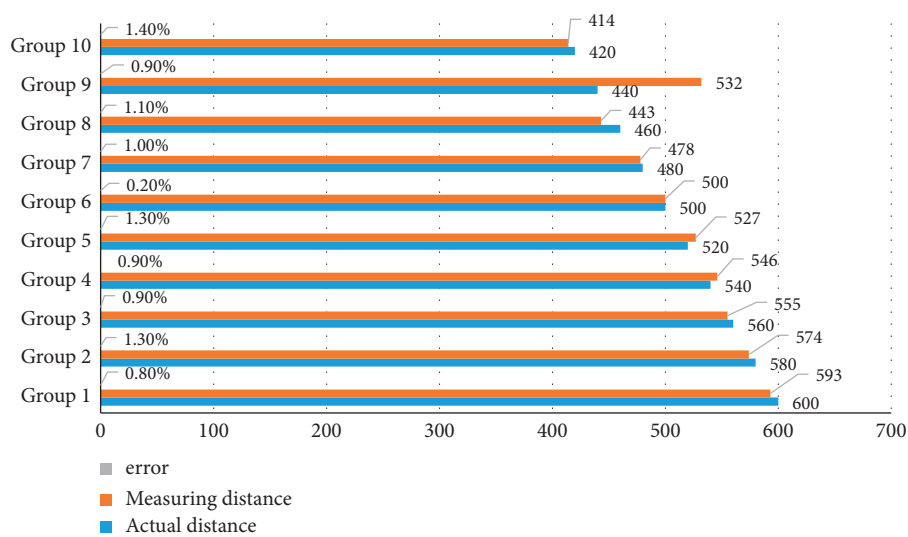


FIGURE 7: Positioning experiment results.

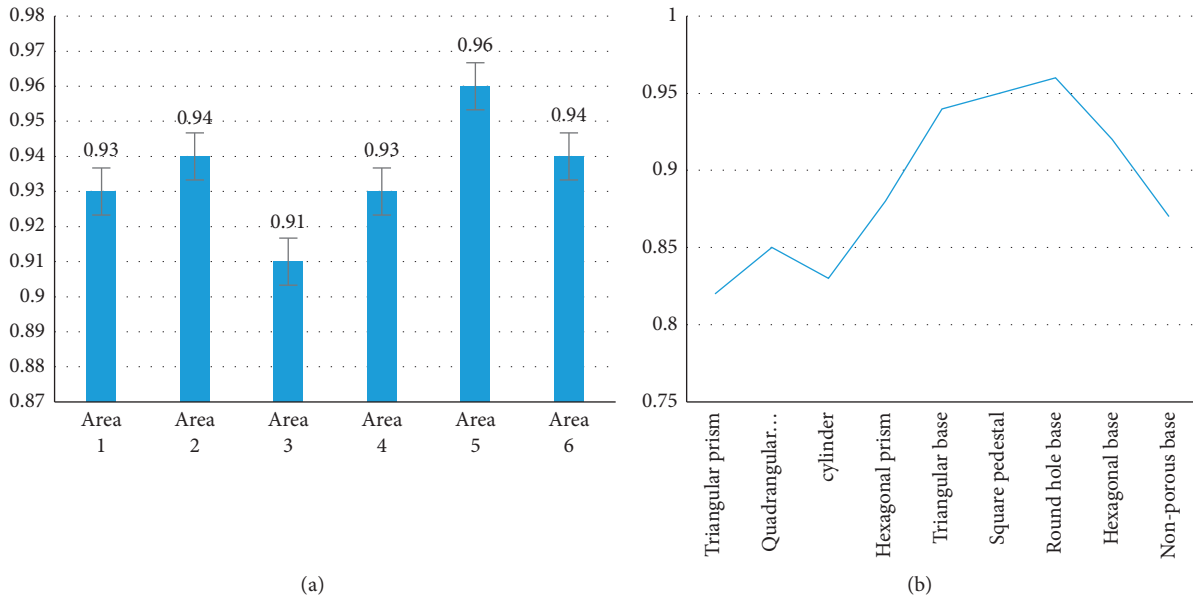


FIGURE 8: Experimental results. (a) The success rate of object grabbing in six areas. (b) 9 objects grabbing success rate.

coordinate system, get the three-dimensional coordinate [23] position of the target object, and drive the robot to realize the grasping task.

5. Conclusions

This paper conducts an experimental study on the hot research issues involved in the process of helping people pick up objects by home service robots. Based on the NAO robot platform, it has completed its tasks of identifying, grasping, and transmitting a given target, including the recognition of the target, obtaining the accurate position of the target in the robot coordinate system, and implementing the grasping behavior. With the continuous deepening of robotics research and increasing market demand, robots [24] will enter every small family in the not-too-distant future. In this paper, the NAO robot platform based on monocular stereo vision is used to locate and grasp the target, and the target recognition and localization are studied experimentally. Some deficiencies also appeared in the experiment process, which deserves further in-depth study. Exploration and research should mainly include the following aspects: the matching recognition method based on quantitative component statistics proposed in this paper can eliminate the interference of certain targets, but in the actual complex home environment, especially when the target is blocked, it cannot be accurately recognized. Need to study deeper recognition algorithms, such as machine learning methods. When the NAO robot is moving towards the target, it needs to locate the target many times to achieve the accurate capture of the target in the next step. The main reason is that the robot has a large walking error. Because the friction factor of the ground in different environments is different, the walking error produced is also different. There is no unified measurement standard. This part of the problem needs further research.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this article.

References

- [1] X. Cui, J. Zheng, S. Chen, H. Guo, and Y. Zhao, "Biprism-based monocular stereovision system parameter optimization," *Journal of Electronic Imaging*, vol. 27, no. 3, pp. 033020.1–033020.9, 2018.
- [2] E. Wu, S. Tian, S. Xu, and M. Jiang, "Research on single-camera stereo vision measurement technology," *Guangxue Jishu/Optical Technique*, vol. 44, no. 1, pp. 1–5, 2018.
- [3] Z. Bai, P. Wang, and Z. Li, "Stereo visual pedestrian detection based on Stixel-world and feature fusion," *Chinese Journal of Scientific Instrument*, vol. 38, no. 11, pp. 2822–2829, 2017.
- [4] W. P. Sanberg, G. Dubbelman, and P. With, "ASTEROIDS: a stixel tracking extrapolation-based relevant obstacle impact detection system," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 1, 2020.
- [5] Y. Zhang, Y. Wang, Y. Qian, and S. Liu, "Improved 3D reconstruction method based on the Scheimpflug principle," *Optical Review*, vol. 27, no. 3, pp. 1–7, 2020.
- [6] M. M. Ali, H. Liu, N. Stoll, and K. Thurow, "Grasping and placing operation for labware transportation in life science laboratories using mobile robots," *Advances in Science, Technology and Engineering Systems Journal*, vol. 2, no. 3, pp. 1227–1237, 2017.
- [7] M. Costanzo, S. Stelter, C. Natale et al., "Manipulation planning and control for shelf replenishment," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, p. 1, 2020.
- [8] A. Gardecki, M. Podpora, and A. Kawala-Janik, "Implementation of an external laser scanner into control system of

- the NAO robot,” *IFAC-PapersOnLine*, vol. 51, no. 6, pp. 231–237, 2018.
- [9] A. M. Pinto, A. P. Moreira, and P. G. Costa, “WirelessSyncroVision: wireless synchronization for industrial stereoscopic systems,” *International Journal of Advanced Manufacturing Technology*, vol. 82, no. 5–8, pp. 909–919, 2016.
- [10] D. R. Parhi and P. B. Kumar, “Smart navigation of humanoid robots using DAYKUN-BIP virtual target displacement and petri-net strategy,” *Robotica*, vol. 37, no. 4, pp. 626–640, 2019.
- [11] E. I. Ershov, V. N. Karnaukhov, and M. G. Mozerov, “Stereo vision algorithms applicability investigation for motion parallax of monocular camera case,” *Journal of Communications Technology and Electronics*, vol. 61, no. 6, pp. 695–704, 2016.
- [12] Q. Liu, J. Z. Hui, and L. Luo, “Target identification and location algorithm based on SURF-BRISK operator,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 30, no. 6, pp. 1655016.1–1655016.11, 2016.
- [13] Wei-Ping, W. X. Ma, P. X. Li, and Cao, “Binocular vision object positioning method for robots based on coarse-fine stereo matching,” *International Journal of Automation and Computing*, vol. 17, no. 4, pp. 86–95, 2020.
- [14] P. K. Donepudi, “Reinforcement learning for robotic grasping and manipulation: a review,” *Asia Pacific Journal of Energy and Environment*, vol. 7, no. 2, pp. 69–78, 2020.
- [15] D. Zhong, Y. Yang, R. Liu, and J. Han, “Study and application of monocular vision-based assembly robot,” *Hsi-An Chiao Tung Ta Hsueh/Journal of Xi’an Jiaotong University*, vol. 52, no. 5, pp. 81–87, 2018.
- [16] H. Masatoshi, “Research on identifications of rubble characteristics with camera images and grasping forces for rescue robots,” *Journal of Japan Society for Design Engineering*, vol. 52, no. 1, pp. 31–40, 2017.
- [17] A. H. Rajpar, A. E. Eladwi, I. Ali, and M. B. Ali Bashir, “Reconfigurable articulated robot using android mobile device,” *Journal of Robotics*, vol. 2021, Article ID 6695198, 8 pages, 2021.
- [18] J. Q. Lin and B. L. Han, “Research on the orientation of hydraulic four-legged robot in the field,” *Beijing Ligong Daxue Xuebao/Transaction of Beijing Institute of Technology*, vol. 38, no. 7, pp. 676–682, 2018.
- [19] J. Nassour, F. H. Hamker, and G. Cheng, “High-performance perpendicularly-enfolded-textile actuators for soft wearable robots: design and realization,” *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 3, p. 1, 2020.
- [20] A. Sujiwo, E. Takeuchi, L. Y. Morales et al., “Robust and accurate monocular vision-based localization in outdoor environments of real-world robot challenge,” *Journal of Robotics and Mechatronics*, vol. 29, no. 4, pp. 685–696, 2017.
- [21] D. Yukiyasu, “Technology trends and future about picking and manipulation by robots,” *Journal of the Robotics Society of Japan*, vol. 35, no. 1, pp. 13–16, 2017.
- [22] J. Lai, Y. He, and X. Li, “Design of manipulator target positioning system based on monocular vision,” *Guangxue Jishu/Optical Technique*, vol. 45, no. 1, pp. 6–11, 2019.
- [23] Z. Lei, Z. Huayan, X. Ning, Z. Tianwei, and B. Guibin, “Gait planning and control method for humanoid robot using improved target positioning,” *Science China (Information Sciences)*, vol. 63, no. 7, pp. 125–127, 2020.
- [24] Z. Jin, M. Zhao, R. Zhang, and J. Wei, “Optimized aggregated channel features pedestrian detection algorithm based on binocular vision,” *Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/Journal of Tianjin University Science and Technology*, vol. 49, no. 12, pp. 1225–1230, 2016.