

Retraction

Retracted: Robot Obstacle Recognition and Target Tracking Based on Binocular Vision

Advances in Multimedia

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

Copyright © 2023 Advances in Multimedia. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] W. Sun, "Robot Obstacle Recognition and Target Tracking Based on Binocular Vision," *Advances in Multimedia*, vol. 2022, Article ID 9022038, 12 pages, 2022.

Research Article

Robot Obstacle Recognition and Target Tracking Based on Binocular Vision

Wencheng Sun 

School of Smart Health, Chongqing College of Electronic Engineering, Chongqing 401331, China

Correspondence should be addressed to Wencheng Sun; sunwencheng@cqcet.edu.cn

Received 2 June 2022; Accepted 18 July 2022; Published 18 August 2022

Academic Editor: Qiangyi Li

Copyright © 2022 Wencheng Sun. 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.

To further improve the perception ability of binocular vision sensor in getting rich environment information and scene depth information, a research method of a robot obstacle recognition and target tracking based on binocular vision was proposed. The method focused on target recognition and obstacle recognition of binocular stereo vision. The system based on obstacles of visual identification was set up. Through the analysis of the Bouguet mathematical model algorithm based on OpenCV, the binocular stereo correction was carried out and the obstacle recognition system was calibrated and corrected. Through the experimental data, it was found that the average error of the obstacle recognition and target tracking algorithm based on binocular vision could be controlled within 50 mm within the range of 2100 mm. The average time of obstacle recognition was 0.096 s and the average time consumption of the whole system was 0.466 s, indicating that the robot obstacle recognition and target tracking system based on binocular vision could meet the accuracy and real-timeness requirements of obstacle recognition and detection.

1. Introduction

With the development of science and technology and changes in social development needs, mobile intelligent technology has been widely used in industries such as industry, medical care, agriculture, and breeding. The emergence of the mobile robot technology makes the program preset in advance work automatically in accordance with the system. The applicable scope of the mobile robot device was very broad. The operation flexibility is strong, and it can replace humans to complete difficult work tasks even in many dangerous and high-risk work areas. With the continuous development and progress of technology, mobile intelligent robot technology can have a certain cognition of the surrounding environment. In the application process, through the intervention of tactile sensors, visual sensors, and other devices, the robot can achieve a better cognition of the surrounding environment [1, 2]. Vision sensors installed in robots can be divided into monocular vision, binocular vision, and multi-eye vision. Most research has focused on the multi-eye view. Compared with monocular vision,

multi-eye vision is more of sound vision, which can obtain three-dimensional integrated data of objects and create a real space, to achieve the purpose of product perception and target tracking. How to improve the effectiveness of impact assessment and the accuracy of target tracking is researched.

2. Literature Review

Binocular stereo vision was investigated earlier in many countries, and the technology had been applied to various mobile robots to complete various visual tasks. The EL-E family robot program, developed by a college, is equipped with a stereo vision system and full vision. The user could guide the robot to approach and grasp the target object through the matching green laser rod, which covered almost all daily objects, including water cups, remote controls, and books [3, 4]. Since the robot was guided to grasp by humans using laser pointer, it did not need to carry out the corresponding target recognition function. The operation method and principle were relatively simple. The target was separated from the obtained environmental image, and the three-

dimensional coordinates of its centroid and the rotation angle of the scene plane were calculated to carry out the accurate grasp. The NAO robot developed by a company was equipped with a variety of sensors such as vision, infrared, and pressure. A binocular stereo vision sensor was installed in the head of the NAO robot. The object was recognized by edge detection and color segmentation of the image, and a three-dimensional model of the object was established. The Cosero robot, developed by a university, was equipped with a GRB-D sensor, namely, Kinect 3D sensor, which contained an infrared point cloud projection, a color camera, and an infrared camera to obtain both depth and color values of the scene. The Cosero robot runs on the platform, making shapes to get three combinations of average size and plane angle. et al. analyzed LASP-1 expression in RCC tissues by immunohistochemistry and Western blot analysis. Cell proliferation, migration, invasion, and gene expression were detected by CCK-8 method, Transwell method, and Western blotting. The results showed that LASP-1 was highly expressed in RCC, and its expression level t was positively correlated with lymph node metastasis and tumor, lymph node, and metastasis (TNM) stage. Knockdown of LASP-1 expression significantly inhibited RCC cell proliferation, increased apoptosis rate, and inhibited RCC cell invasion and migration by inhibiting epithelial-mesenchymal transition. In conclusion, LASP-1 promotes RCC progression and metastasis and is a promising therapeutic target for RCC [5]. Nour et al. discuss recent developments in cytoskeletal actin dynamics in liver fibrosis, including how the cellular microenvironment affects HSC function and the molecular mechanisms that regulate actin-induced increased collagen expression typical of activated HSCs [6].

The research on binocular stereo vision in China was a little later than that in other countries, but with the unremitting research and hard work of Chinese researchers, remarkable achievements were made in binocular stereo vision. In view of the lack of intelligence in robot operation, a binocular vision system was built. The template matching technology was used to classify and recognize the feature points accurately. The pose of the target object was obtained through 3D reconstruction. A binocular stereo vision target recognition and capture method based on SURF+ RANSAC was proposed. The two methods were combined to match the image. The affine transformation was used to obtain key point information. Distance measurement and location were carried out according to 3D reconstruction. To solve the problem of target contour center mismatch in the two images obtained by the left and right cameras, the SURF algorithm was used to identify the target and GrabCut algorithm was automatically initialized to extract the target contour. Target location was carried out by template matching method combined with the principle of 3D reconstruction.

3. Binocular Stereo Vision System Based on Mobile Robot

3.1. Construction of Binocular Stereo Vision System

3.1.1. *Hardware Platform.* The binocular camera ZED generation is selected for the research. It is a depth camera

TABLE 1: Main parameters of ZED camera.

Performance indicators	Parameters
Baseline spacing	120 mm
Focal length	2.8 mm
Depth range	0.5 m–25 m
Weight	170 g
Operating temperature	0°C–45°C
Specification size	175 × 30 × 33 mm
View scope	Max 90°(H) × 6°(V) × 100°(D)

based on the RGB binocular stereo vision principle, which can be used both indoors and outdoors. The maximum depth range can reach 25 m. The highest of the resolution of depth map is 4416x1242@15FPS, which can also be adjusted according to the frame rate. The highest frame rate is 1344x376@100FPS. The maximum coverage field angle is 100°. The specific parameters are shown in Table 1.

3.1.2. *Software Platform.* C++ language is created using Visual Studio 2013. The software pole diagram of the binocular system is shown in Figure 1, which is divided into four parts.

- (1) *Image Acquisition.* It includes communication with the upper computer through USB 3.0 interface, setting camera parameters, and obtaining image frames.
- (2) *Target Segmentation.* The target area is segmented first, and then, the precise target segmentation is carried out.
- (3) *Target Recognition.* Firstly, feature extraction is carried out. Then, the classifier trained is carried out. Finally, target recognition is carried out.
- (4) *Target Positioning.* It includes polar line correction, stereo matching, and target pose estimation [7, 8].

3.2. *Camera Model.* Camera imaging transformation includes the transformation of the target in four different coordinate systems, including the world coordinate system, camera coordinate system, image pixel coordinate system, and image physical coordinate system. Firstly, it is necessary to make clear the transformation relationship between these four coordinate systems.

- (1) World coordinate system is expressed as follows:

$$O_w - X_w Y_w Z_w. \quad (1)$$

It is the absolute coordinate system used to represent the real world, which is the reference coordinate system for objects in the real world.

- (2) Camera coordinate system is expressed as follows:

$$O_c - X_c Y_c Z_c. \quad (2)$$

It takes the camera light spot as the origin and Figure 2 camera optical axis Z_c as the axis.

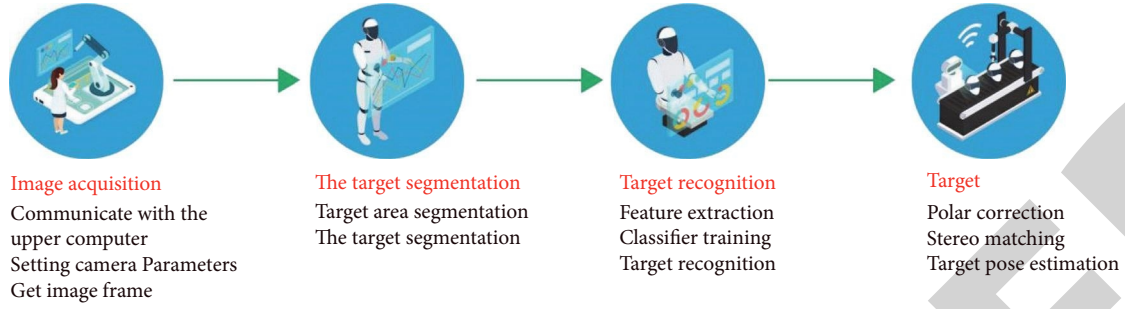


FIGURE 1: Software framework diagram of binocular system.

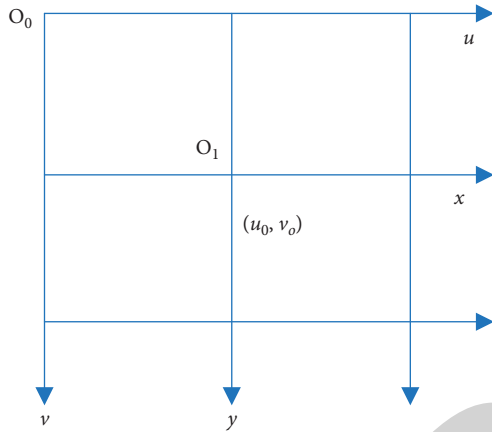


FIGURE 2: Image pixel coordinate system and image physical coordinate system.

- (3) Image pixel coordinate system is expressed as O_0uv . The origin O_0 represents the pixel point in the upper left corner of the digital image. The basis of the u -axis and v -axis is pixels, where the u -axis represents pixel lines and the v -axis represents pixel lines.
- (4) Image physical coordinate system is expressed as O_1xy , which plays a role in connecting camera coordinate system and image pixel coordinate system in the process of camera calibration. The coordinate origin O_1 is the projection center (u_0, v_0) . x -axis is parallel to u -axis. y -axis is parallel to v -axis. To obtain the internal and external parameters of the camera, it is necessary to know the corresponding relationship between coordinate systems O_1xy and O_0uv . The coordinate system is established as shown in Figure 2.

The relationship between the pixel coordinate system and the physical interaction of the image can be represented by the following model:

$$\begin{cases} u = u_0 + \frac{x}{dx} \\ v = v_0 + \frac{y}{dy} \end{cases} \quad (3)$$

In formula (3), (x, y) represents the coordinates of any point in the material coordinate system. (u, v) represents the

corresponding coordinates in the pixel coordinate system. dx and dy are the actual dimensions of a single pixel point. To facilitate transformation between coordinates, formula (3) is transformed into matrix form.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (4)$$

Changing the relationship between the camera coordination and the global control system is an external measure of the camera. Global control system, camera control system, pixel control system, and human image integration are established, as shown in Figure 3.

The relationship of the camera control integration with the global control system shows that

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ O^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M_2 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (5)$$

In formula (5), R is the rotation matrix of order 3. T is the 3×1 translation vector.

$$O^T = (0, 0, 0) \quad (6)$$

M_2 is 4×4 matrix, in Figure 4, to simplify the conversion relationship.

- (1) As shown in [9, 10], in the ideal conditions of physics, the camera is simplified as a particle, that is, a point approaching infinitesimal. When a beam of light shines on a plane with only one particle, the light cannot be transmitted to the other side through parallel lines. At this time, the light can only be fully projected to the imaging plane through the particle. Suppose a point $P(X_c, Y_c, Z_c)$ in the camera coordinate system is projected onto the imaging plane by the camera's optical center, forming an inverted image $P_i(x, y)$. The imaging model can be obtained through similar triangles:

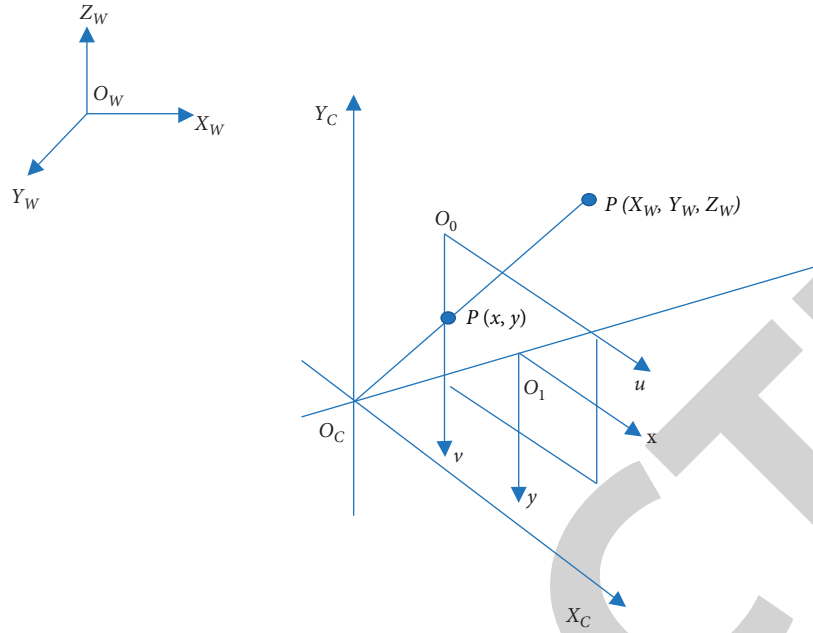


FIGURE 3: Correspondence between the four coordinate systems.

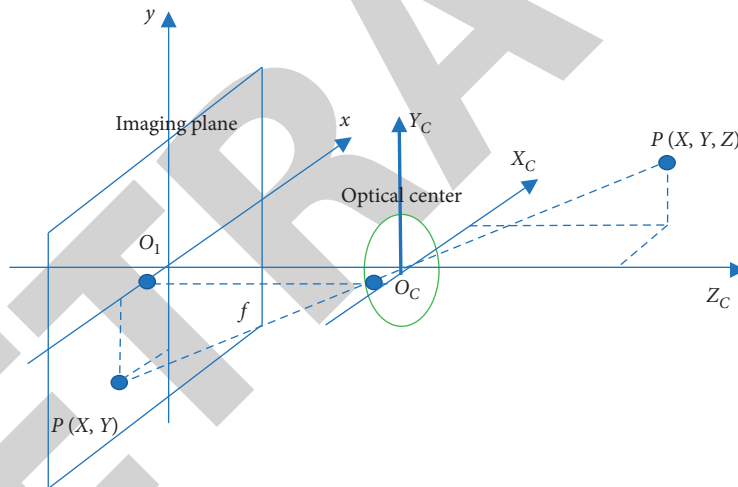


FIGURE 4: Pinhole imaging model.

$$\begin{cases} x = f \frac{X_c}{Z_c}, \\ y = f \frac{Y_c}{Z_c}. \end{cases} \quad (7)$$

In formula (7), f is the focal length of the camera. The above equation can be simplified into a matrix as follows:

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}. \quad (8)$$

Formula (5), formula (6), and formula (7) are combined to obtain the linear model of the camera.

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ O^Y & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \quad (9)$$

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & T \\ O^Y & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M_1 M_2 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \quad (10)$$

In formula (9) and formula (10), $f_x = f/d_x$ is the scale factor of focal length f on u -axis. $f_y = f/d_y$ is the scale factor on v axis. M_1 is the internal parameter matrix of the camera, which is determined jointly by $f, d_x, d_y, u_0,$ and v_0 . It is the inherent parameter, which does not change with environmental changes. M_2 is the external camera parameter matrix, including rotation matrix R and translation vector T , which changes with position changes.

- (2) *Nonlinear Model.* The linear model of the camera is affected by artificial conditions such as the production level of the product and integration error, which leads to problems in the quality of the lens or inaccurate measurement of the focal length of the lens. In this case, the linear model of the camera will have the corresponding distortion and the camera model is a nonlinear model [11, 12]. Distortion includes radial distortion, tangential distortion, and thin prism distortion. Thin prisms have almost no distractions in ZED cameras, so they can be ignored when talking about distractions. Radial distortion and tangential distortion are shown in Figure 5.

The radial error of one pixel is called radial distortion. The distance from the pixel point to the optical center is positively affected by the degree of image distortion. The greater the distance, the greater the change. The radial distortion mode is set:

$$\begin{cases} \delta_x(x, y) = x(k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots), \\ \delta_y(x, y) = y(k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots). \end{cases} \quad (11)$$

In formula (11),

$$r = \sqrt{(x^2 + y^2)}. \quad (12)$$

$k_1, k_2, k_3 \dots$ are the radial distortion parameters. Under normal circumstances, the distortion is relatively small. Parameters can be set to k_1 and k_2 .

The position deviation of x - or y -axis caused by artificial problems of the lens or the camera not parallel to the imaging plane is called tangential distortion. The tangential distortion model is established.

$$\begin{cases} \delta_x(x, y) = 2p_2 xy + p_1(3x^2 + y^2), \\ \delta_y(x, y) = 2p_1 xy + p_2(x^2 + 3y^2). \end{cases} \quad (13)$$

In formula (13), p_1 and p_2 are tangential distortion parameters. In combination with formula (11) and formula (13), the nonlinear model of the camera is as follows:

$$\begin{cases} \delta_x(x, y) = x(k_1(x^2 + y^2) + k_2(x^2 + y^2)^2) + (2p_2 xy + p_1(3x^2 + y^2)), \\ \delta_y(x, y) = y(k_1(x^2 + y^2) + k_2(x^2 + y^2)^2) + (2p_1 xy + p_2(x^2 + 3y^2)). \end{cases} \quad (14)$$

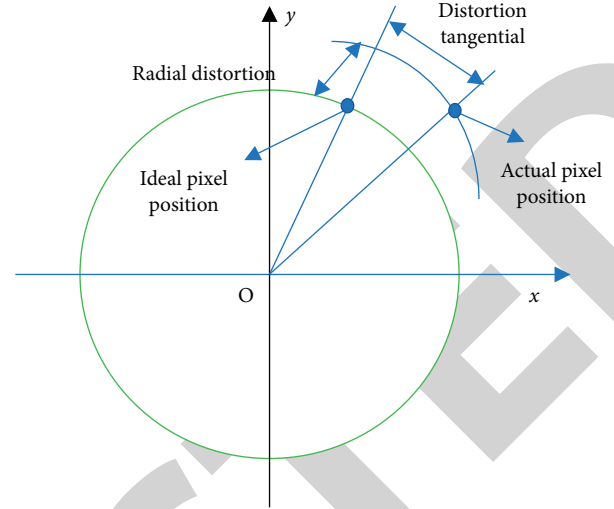


FIGURE 5: Camera distortion model.

4. Binocular Camera Calibration and Stereo Correction

4.1. Binocular Camera Calibration Based on Control Variable Method. Since the focal length determines the accuracy of the focal depth D , this section uses the differential control method to study the influence of the size of the checkerboard calibration plate and the distance of the calibration scale on the error calibration and obtains a high-precision focal length f .

4.1.1. Experiment a. The calibration distance is set as 1600 mm, and checkerboard calibration plates of different specifications are drawn [13, 14]. The most immediate purpose of a dual-target fixed camera is to compute interior and exterior measurements for further integration from the center of the image to the triangle. Due to the binocular calibration completed by OpenCV, the results of multiple calibration may be unstable and even serious errors may occur, resulting in large deviations. Therefore, the identification system uses the MATLAB calibration toolbox with high accuracy to calibrate the binocular camera. The flow chart of the calibration process is shown in Figure 6.

4.1.2. Experiment b. The specification of checkerboard calibration plate is set as 26 mm \times 26 mm. 12 groups of different calibration distances are set, with an increase of 100 mm from 800 mm to 1900 mm. Other conditions are consistent with Experiment a [15, 16].

4.2. Experimental Results and Analysis. The calibration error is the reverse projection error. The ideal value is zero; generally, less than 0.5 pixel can be used. Average calibration errors of each time can be obtained through the MATLAB calibration toolbox. The average calibration errors of experiments a and b are shown in Tables 2 and 3.

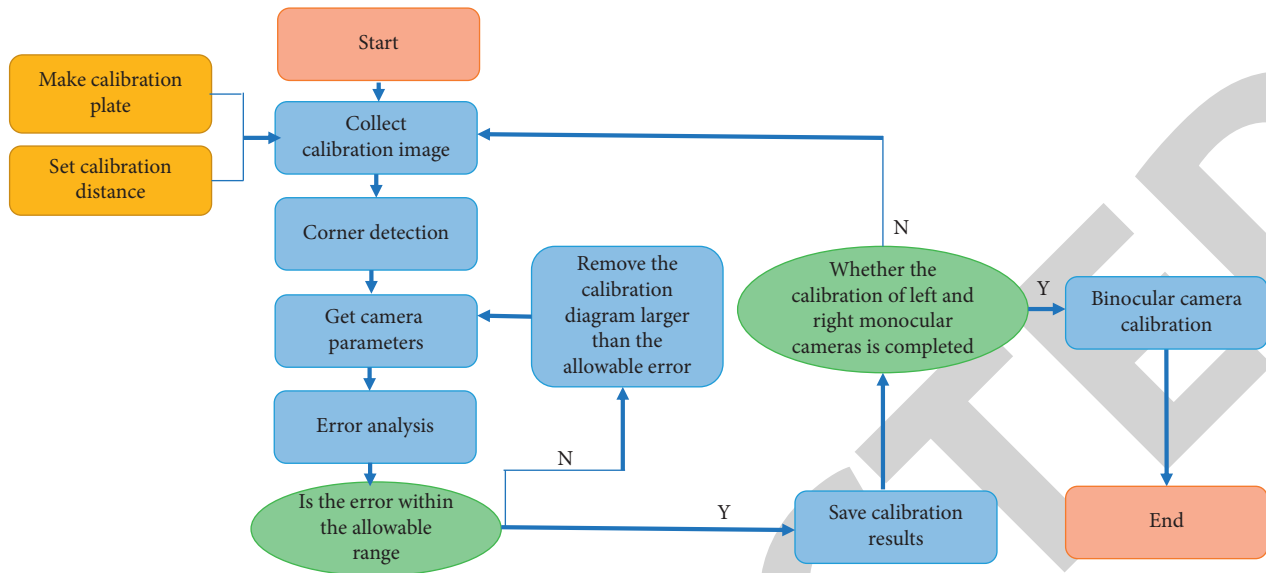


FIGURE 6: Flow chart of binocular camera calibration.

TABLE 2: Average calibration error of Experiment a.

Specification of calibration plate (mm)	Average calibration error (px)
8	1.23844
12	1.01541
14	0.82457
16	0.74556
18	0.64574
20	0.45255
22	0.38452
24	0.26634
26	0.18542
28	0.22524
30	0.25785
40	0.68345

TABLE 3: Average calibration error of Experiment b.

Specification of calibration plate (mm)	Average calibration error (px)
800	0.36541
900	0.34982
1000	0.28364
1100	0.24956
1200	0.22357
1300	0.19352
1400	0.22143
1500	0.26451
1600	0.29452
1700	0.33156
1800	0.34478
1900	0.38542

According to Experiment a, the average calibration error of the checkerboard calibration plate with the specification of 26 mm × 26 mm is the smallest and that of the checkerboard calibration plate with the specification of 8 mm × 8 mm is the largest. The average calibration error of the calibration plate with the specification of 8 mm × 8 mm

TABLE 4: Results of binocular calibration.

Parameters	Results of binocular calibration	
	Left camera	Right camera
f_x	2558.685	556.4126
f_y	2556.491	557.5483
C_x	1020.17	1021.959
C_y	856.3736	857.8058
K_1	-0.05144	0.08049
K_2	0.11352	0.548
P_1	-0.0039	0.00314
P_2	-0.00307	0.00325
K_3	0	0
Average error	0.16866	0.16327
External parameters	R	[-001944 -0.01157 -0.0165801]
	T	[-150.95861 -0.00125 0.00208]

increases from 26 mm × 26 mm, and the average calibration error decreases. In the process of 26 mm × 26 mm to 40 mm × 40 mm growth, the average calibration error shows a downward trend [17, 18]. According to test b, when the calibration distance is 1300 mm, the average calibration error is the lowest. When the calibration distance is 1900 mm, the average calibration error is the largest. When the calibration distance increases from 800 mm to 1300 mm, the average calibration error shows a downward trend. When the calibration distance increases from 1300 mm to 1900 mm, the average calibration error shows a downward trend. Therefore, a checkerboard calibration board with a specification of 26 mm × 26 mm was selected in the research and the calibration distance was 1300 mm. The internal and external parameters of the left and right cameras after the binocular calibration are shown in Table 4.

As can be seen from Table 4, by setting the checkerboard calibration plate of $26\text{ mm} \times 26\text{ mm}$ and the calibration distance of 1300 mm , the difference between f_x and f_y obtained by the binocular calibration test is very small, which can make the accuracy of f higher. The rotation vector is as follows:

$$R = [-0.01944 \quad -0.01157 \quad -0.016580]. \quad (15)$$

The items in the above formula are approximate to 0, indicating that there is almost no relative rotation between left and right cameras. The translation vector is as follows:

$$T = [-150.95861 \quad -0.00125 \quad -0.00208]. \quad (16)$$

The first item -150.95861 represents the distance between the left and right x -axis of the camera, that is, the distance from the base, which is consistent with the actual value of 150 mm . Therefore, binocular vision can be considered a visual impairment. To obtain the optimum viscosity of the binocular surface, high-pressure operation can be carried out. MATLAB 2016a calibration toolbox is used to calibrate the left and right cameras, respectively. The control variable method in physics is introduced to conduct comparative tests by setting different calibration plate specifications and calibration distance. Finally, the calibration distance is set as 1300 mm and the calibration plate specifications are positioned as $26\text{ mm} \times 26\text{ mm}$. The system achieves the optimal parameters and obtains the camera's internal and external parameters with high accuracy, which provides a guarantee for calculating the obstacle distance. Due to camera distortion, manual installation, and other reasons, it is difficult to directly achieve the ideal binocular vision system in real operation. Therefore, after completing binocular camera calibration, the Bouguet stereo correction algorithm based on OpenCV is adopted to complete stereo correction. By analyzing the binocular calibration results and effect pictures after stereo correction, it can be seen that the system has basically reached the ideal binocular vision system.

4.3. Binocular Stereo Matching Algorithm

4.3.1. Binocular Stereo Matching Process. The section aims to design an effective stereo matching algorithm in the intelligent parking environment [19, 20]. The steps of stereo matching are generally divided into four steps, as shown in Figure 7.

Calculating the matching cost is to collect the corresponding points in the left and right views under different parallax and calculate their similarity. Traditional matching cost methods include normalized correlation coefficient (NCC), square of gray difference (SD), and absolute value of gray difference (AD).

Comparing values together creates connections between adjacent pixels and adjusts matrix values according to certain rules so that nominal values can affect pixel correlation. The common matching cost aggregation methods include dynamic programming, path clustering, and scan line clustering. Parallax calculation is to use the cost matrix after cost aggregation to select the optimal parallax value. A

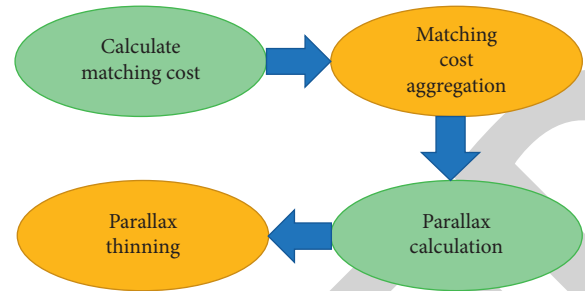


FIGURE 7: Steps of stereo matching.

winner-take-all approach is usually used to complete parallax calculation [21, 22].

The common parallax optimization methods include left-right check, median filter, and intensity consistent. Since the parallax value obtained by the winner-take-all algorithm is the whole lower pixel accuracy, the unary quadratic curve fitting method is usually used to obtain higher sub-pixel accuracy.

4.3.2. Stereo Matching Classification. Based on the classification using the stereo hybrid algorithm, the hybrid algorithm is divided into two categories according to the role of its time limit: the international integration algorithm and the local matching algorithm, as shown in Figure 8. Global comparison algorithms generally include intelligent algorithms such as algorithm matching based on reliable propagation, dynamic programming methods, matching algorithms, and graph cut methods. It obtains a high matching accuracy at the expense of time and through a large number of calculations, which will not be able to meet the intelligent garage parking robot obstacle recognition system for real-time requirements. Basic matching algorithms can be subdivided into combination-based algorithms, feature-based matching algorithms, and phase-based matching algorithms based on different terms. It has the characteristics of low computational complexity and good real-time performance. Due to the real-time requirements of the stereo synthesis algorithm in the research, the local stereo algorithm is studied, and a stereo synthesis algorithm that can meet the needs of the scene is established, intelligent parking.

By comparing the above three local matching algorithms, the application environment of the system is the intelligent parking garage environment with uneven light and weak light, so the feature-based matching algorithm was selected to solve the problem. The traditional feature matching algorithms SIFT and SURF were investigated and an improved algorithm was proposed to adapt to the application environment of the system, so that stereo matching could meet the real-time and precision requirements of the system.

5. Experiment and Analysis of Obstacle Recognition System

In the research, the parking robot obstacle recognition system needs to realize not only the distance detection of obstacles in front of the parking path to ensure the safety of

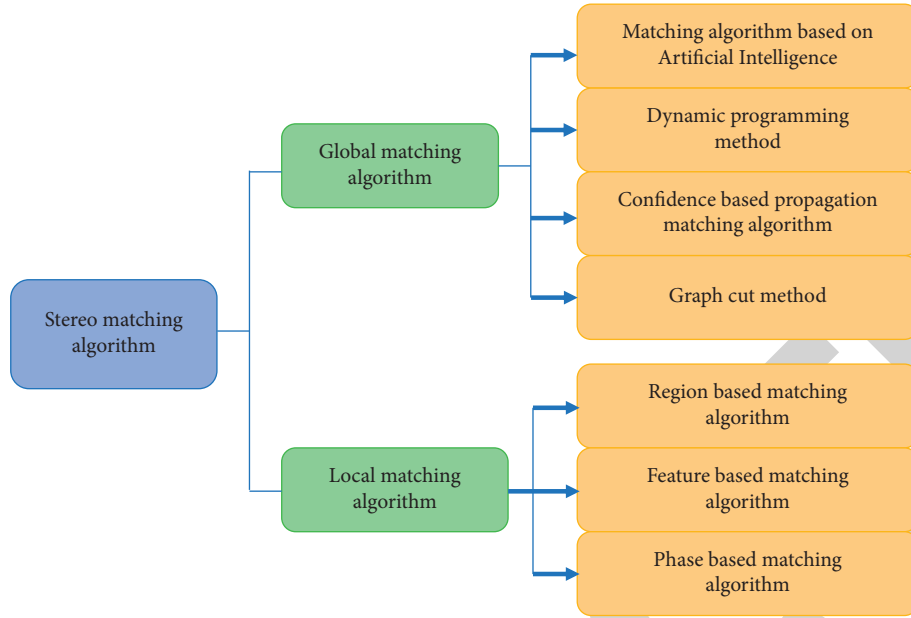


FIGURE 8: Classification of stereo matching algorithms.

the parking process. In addition, biological and nonbiological types of obstacles should be detected. If it is detected that creatures (people, cats, and dogs) enter by mistake, the progress should be stopped and the garage management personnel should be notified to intervene. If it is nonbiological type, autonomous obstacle avoidance should be notified. Therefore, obstacle recognition methods were explored in the section and YOLO convolutional neural network was selected as the obstacle recognition module of the system. Experiment verification and results analysis on the obstacle recognition system were carried out [23, 24].

5.1. Obstacle Category Detection Research Based on YOLO Convolutional Neural Network. The old product search process can be divided into three parts: selecting regions (sliding window), removing features (HAAR, HOG, etc.), and classifiers (SVM, AdaBoost, etc.). The main problems are as follows. On the other hand, the feature robustness of manual extraction is poor. Since entering the era of deep learning, object category detection has achieved unprecedented development, and the most eye-catching two directions are as follows: region proposal-based deep learning object detection algorithm represented by RCNN (RCNN, fast-RCNN, etc.). They are two-staged. Firstly, region proposal is generated using the selective search method. Then, the data are classified and regressed in region proposal—represented by deep learning target detection algorithm (YOLO, SSD, etc.), which only uses a CNN to directly predict the categories and locations of different targets.

The YOLO convolutional neural network model consists of a convolution process, a pooling layer, and all layers. It directly applies the image as input and returns the location

and category of the target box in the output layer. Similar to GoogleNet, the network has 24 convolutional layers and 2 fully connected layers. The steps of YOLO convolutional neural network for recognition purposes are as follows. First, the convolution procedure is trained on Pascal VOC data. The RPN is then designed to use the convolution process and all layers. Finally, the class target and target image location are estimated.

The YOLO convolutional neural network algorithm firstly divides the input image into $S \times S$ grid cells and then predicts B bounding boxes for each cell. Each box contains five predicted values, namely, x, y, w, h , and confidence. (x, y) represents the position of the center of the detection border relative to the grid. w and h indicate the width and height of the detection frame, respectively. Each grid predicts the probability of C hypothetical categories and finally outputs a feature graph with size of $S \times S \times (B \times 5 + C)$. Pascal VOC training data are used, such as training 20 categories; $S=7$, $B=2$, and $C=20$, and then, the output characteristic graph is $7 \times 7 \times 30$.

The confidence of each prediction box containing the target is defined as follows:

$$\Pr(\text{Object}) \times \text{IOU}_{\text{pred}}^{\text{truth}} \quad (17)$$

If the prediction box contains objects, then $P(\text{Object}) = 1$, and otherwise, $P(\text{Object}) = 0$; the probability value of defining the prediction condition for each grid is as follows:

$$C = \Pr(\text{Class}_i | \text{Object}). \quad (18)$$

It represents the probability that the grid has a group, and the relationship between the two numbers is shown as follows:

TABLE 5: Analysis of obstacle recognition performance of YOLO convolutional neural network.

Group	Target category (from left to right)	Confidence level (%)	x	y	w	h	Time-consuming (s)
(a)	Dog	100	357	29	354	409	0.089
(b)	Bicycle	100	302	499	533	408	0.088
	Car	98	3	210	191	70	
(c)	Car	38	165	220	54	49	0.095
	Person	99	342	29	121	321	
(d)	Car	94	581	170	90	114	0.116
	Car	90	0	114	67	87	
	Car	96	12	109	143	69	
	Car	99	325	127	332	238	
	Car	65	571	123	42	49	
	Car	99	587	117	176	156	

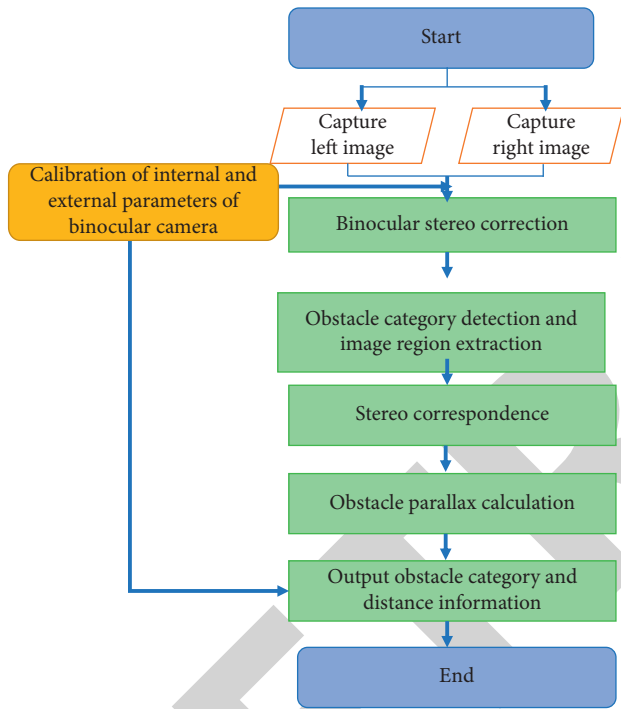


FIGURE 9: Flow chart of obstacle identification system algorithm.

$$\Pr(\text{Class}_i|\text{Object}) \times \Pr(\text{Object}) \times \text{IOU}_{\text{pred}}^{\text{truth}}. \quad (19)$$

After getting the credible value of each prediction box, the threshold is set. Predictions with low scores are filtered. The prediction box is done by NMS. The final examination is obtained.

YOLO convolutional neural network uses error average as loss function. It divides the image into $S \times S$ grids. Each grid has $(B \times 5 + C)$ dimension data, including $(B \times 4)$ dimension position information of detection frame, $(B \times 1)$ dimension confidence of detection frame, and C dimension category probability. To solve the problem that different classification errors are treated equally and the training object center does not fall on the grid, leading to the training instability, in YOLO convolutional neural network, the positioning error is designed with the weight of 5. The confidence error is designed with the weight of 0.5 for the grid that does not

contain the center of the training object. The classification error and the confidence error of the grid that contain the center of the training object remain unchanged. The length and width errors of the detection frame are more sensitive than the positioning errors. So, the network \sqrt{w} is replaced by w and \sqrt{h} is replaced by h . The loss function of YOLO convolutional neural network is as follows.

$$\begin{aligned} \text{Loss} = & \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^B \left[(x_i - \hat{x}_j)^2 + (y_i - \hat{y}_j)^2 \right] +, \\ & \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right) + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right) \right]^2 +, \\ & \sum_{i=0}^{s^2} \sum_{j=0}^B I_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} (C_i - \hat{C}_i)^2 +, \\ & \sum_{i=0}^{s^2} I_{ij}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2. \end{aligned} \quad (20)$$

As the application environment is garage, the data set provided by the authorities is rich in categories, with high-quality images and large number of training samples, which can meet the detection of basic obstacles, including biological categories such as human, cat, and dog and nonbiological categories such as cars, bikes, and bottles. Therefore, its own data set is not retrained, but the weight file provided by YOLO (containing 20 target categories) is used to conduct 4 groups of obstacle category detection experiments [25].

The target detection method based on YOLO convolutional neural network can complete the detection of obstacle categories in the parking garage environment. The performance analysis of the detection of four groups of object categories of single target recognition and multiple target recognition is shown in Table 5.

It can be seen from Table 5 that the weight files provided by the authorities can be used to complete the classification detection of common obstacles in the parking garage environment, with high detection accuracy and time-consuming at the millisecond level, which can fully meet the real-time requirements of obstacle category detection of parking robots.

TABLE 6: Ranging results of four matching algorithms.

Actual distance (mm)	SIFT algorithm	SURF algorithm	FAST algorithm	Algorithm in the research
1000	1012.9	1011.8	1021.5	1009.7
1100	1115.4	1114.2	1126.2	1112.5
1200	1219.5	1218.1	1230.1	1216.9
1300	1322.3	1321.2	1333.2	1319.6
1400	1426.3	1425.7	1437.9	1423.5
1500	1531.6	1530.1	1542.8	1527.2
1600	1637.1	1636.5	1648.7	1630.9
1700	1746.1	1747.8	1752.6	1734.5
1800	1852.4	1853.9	1862.2	1839.1
1900	1968.6	1969.2	1979.5	1943.5
2000	2096.2	2097.5	2115.8	2046.9
2100	2211.5	2212.4	2254.9	2149.8

TABLE 7: Average system time.

Name	Time-consuming (s)
Stereo matching module	0.256
Obstacle-type detection module	0.096
Obstacle identification system	0.466

5.2. Experiment and Result Analysis of Obstacle Recognition System

5.2.1. Overall Algorithm Design of the System. Figure 9 shows the flow chart of algorithm realization of the whole system. Firstly, the binocular camera is used to collect the parking garage image pairs. The control variable method in physics is introduced, and the binocular camera calibration is completed through MATLAB toolbox. The calibration parameters are combined with the Bouguet stereo correction algorithm to complete stereo correction. Then, YOLO convolutional neural network is used to complete obstacle recognition and rectangular frame is used to circle the image range of obstacle. The improved SURF stereo matching algorithm is used to extract the feature points of the left and right images. Finally, combining the YOLO circle convolution neural network image coordinates (x, y) and an obstruction of left and right images for matching feature points, the parallax of all the feature points in the rectangular frame is calculated. The maximum value for parallax is chosen. The binocular ranging method is used to calculate the minimum depth of obstacles in order to understand the output of obstacle category and distance information.

5.2.2. Experimental Results and Analysis. As the application environment of this system is underground parking garage, aiming at the weak and uneven illumination in this environment, the effectiveness of the obstacle identification system in the research is verified, including the ranging accuracy of the system, the accuracy of obstacle detection, and real-time. In the research, the obstacle is set as a person who enters the parking garage by mistake, and different distances from 1000 mm to 2100 mm are set for testing, with an interval of 100 mm for each group. The system test results use four

matching algorithms to calculate the distance of the obstacle. The ranging results are shown in Table 6, and the average time of the system is shown in Table 7.

As can be seen from Table 7, the conceptual algorithm is much more accurate compared with the local problems based on SIFT, SURF, and FAST algorithms. As the distance increases, the data edge interference starts to decrease when the impact distance exceeds 1600 mm. The number of similar points is reduced by improving the SURF algorithm. However, the difference in reality has not changed much, and the energy is very strong. As can be seen from Table 7, the total time consumption and the impact experience of each part are in milliseconds, which can be satisfied in real time.

6. Conclusion

The main work of the research is as follows:

- (1) *Construction of Obstacle Visual Recognition System.* By investigating the principle and mathematical model of binocular vision system, combined with the application environment of the system, the software and hardware platforms of the system were built. On the basis of analyzing the general steps of the realization of binocular vision system, the workflow of the system was designed. In the hardware design, the CCD camera with high imaging quality and strong anti-interference ability was used as the image acquisition equipment. The computer with high configuration was used as the image processing equipment to ensure the real-time image processing. In the software design, the development platform and the key technology required for image processing were mainly introduced. To facilitate the later transplantation, the research mainly adopted the open-source function library OpenCV 3.3.0 as the tool of image acquisition and processing and finally pointed out the parameters that may affect the system ranging accuracy.
- (2) *Calibration and Correction of Binocular Vision System.* By investigating the main camera calibration methods and comparing their advantages and disadvantages, Zhang Zhengyou's checkerboard calibration method was selected as the binocular

calibration method in the research. By comparing the advantages and disadvantages of OpenCV and MATLAB calibration, MATLAB calibration toolbox was selected for calibration. Since focal length F may be affected by calibration distance and calibration plate specification, the control variable method was introduced to investigate calibration plates with different calibration distance and specification. Experiment verification showed that when the calibration plate size was $26\text{ mm} \times 26\text{ mm}$ and the calibration distance was 1300 mm , the calibration average error was minimum and the optimal parameters—focal length f and internal and external parameters of left and right cameras—were obtained. Finally, the Bouguet algorithm based on OpenCV was used to complete stereo correction of left and right image pairs, so that the image plane was parallel and the row was aligned, which laid a foundation for improving stereo matching algorithm.

- (3) *Research on Obstacle Detection and System Experiment.* Since the traditional object detection algorithm had the problems of few object detection types, low accuracy, and high time complexity, the YOLO convolutional neural network based on deep learning was introduced to improve the detection types and accuracy of obstacles. YOLO convolutional neural networks were adopted, which could realize real-time detection of the obstacles. The image area where the obstacle was located was framed. With the help of left and right stereo matching algorithm, feature points matching and parallax calculation were carried out. Finally, by using the triangulation principle, the distance between the obstacles was calculated and obstacle categories and distance detection were realized. The system had strong expansibility, which could be trained according to the actual needs of their own sample data set, so as to apply to different scenarios. By comparing and evaluating the research results of existing targets and discovery plans on the basis of in-depth research, the YOLO convolutional experimental results show that the problem discovery-level vision module based on the YOLO convolutional neural network can complete the real-time detection and quality of problems. Performance: the vision platform was built to identify the accuracy of various interfaces and propose the real-time performance of vision algorithms. The experimental results show that based on the combination of optical technology and YOLO convolutional neural network, the detection of group problems and distances can be realized within the range of 2100 mm , and the error position is within 5 mm . The overall algorithm took less than 500 ms , which could meet the requirements of the system.

Data Availability

The labeled data set used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by the Youth Project of Science and Technology Research Program of Chongqing Education Commission of China (No. KJQN202103103).

References

- [1] F. Yang and Y. Rao, "Vision-based intelligent vehicle road recognition and obstacle detection method," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 07, Article ID 2050020, 2020.
- [2] D. Ruan, W. Zhang, and D. Qian, "Feature-based autonomous target recognition and grasping of industrial robots," *Personal and Ubiquitous Computing*, no. 1, pp. 1–13, 2021.
- [3] G. Zhai, W. Zhang, W. Hu, and Z. Ji, "Coal Mine rescue Robots Based on Binocular Vision: A Review of the State of the Art," *IEEE Access*, vol. 8, pp. 130561–130575, 2020.
- [4] J. H. Kim, S. S. Lee, H. H. Kang, and S. H. You, "Multitarget tracking with lidar and stereo vision based on probabilistic data association fir filter," *Journal of Electrical Engineering & Technology*, vol. 17, no. 1, pp. 641–654, 2021.
- [5] S. Fanghao, W. Zhao, L. Zhang et al., "Oncolytic adenovirus targeting LASP-1 inhibited renal cell cancer progression," *IEEE Transactions on Dependable and Secure Computing*, vol. 44, no. 4, pp. 639–647, 2020.
- [6] H. Nour, D.C. Rockey, Z. Shi, and S. Zengdun, "The cellular microenvironment and cytoskeletal actin dynamics in liver fibrogenesis," *Journal of Intelligent and Fuzzy Systems*, vol. 46, no. 9, pp. 2003–2007, 2022.
- [7] L. S. Liu, J. F. Lin, J. X. Yao et al., "Path planning for smart car based on dijkstra algorithm and dynamic window approach," *Wireless Communications and Mobile Computing*, vol. 2021, no. 4, pp. 1–12, 2021.
- [8] X. Zhao, H. Wu, Z. Xu, and H. Min, "Omni-directional Obstacle Detection for Vehicles Based on Depth Camera," *IEEE Access*, vol. 8, pp. 93733–93748, 2020.
- [9] Y. Mao, M. Chen, X. Wei, and B. Chen, "Obstacle Recognition and Avoidance for Uavs under Resource-Constrained Environments," *IEEE Access*, vol. 8, pp. 169408–169422, 2020.
- [10] S. Wu, R. Li, Y. Shi, and Q. Liu, "Vision-based Target Detection and Tracking System for a Quadcopter," *IEEE Access*, vol. 9, pp. 62043–62054, 2021.
- [11] Y. Ben and K. Cengiz, "Research on visual orientation guidance of industrial robot based on cad model under binocular vision," *Computer-Aided Design and Applications*, vol. 19, no. S2, pp. 52–63, 2021.
- [12] J. H. Ko, "Robot vision system based on 3d depth map and object recognition," *Journal of the Institute of Electronics and Information Engineers*, vol. 57, no. 3, pp. 101–105, 2020.
- [13] A. Sharma and R. Kumar, "A Framework for Pre-computed Multi- Constrained Quickest QoS Path Algorithm," *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 9, pp. 73–77, 2017.
- [14] T. Qin, Y. Yang, B. Wen et al., "Research on human gait prediction and recognition algorithm of lower limb-assisted exoskeleton robot," *Intelligent Service Robotics*, vol. 14, no. 3, pp. 445–457, 2021.
- [15] D. Selva, B. Nagaraj, D. Pelusi, R. Arunkumar, and A. Nair, "Intelligent network intrusion prevention feature collection

- and classification algorithms,” *Algorithms*, vol. 14, no. 8, p. 224, 2021.
- [16] T. Callens, T. van der Have, S. V. Rossom, J. De Schutter, and E. Aertbelien, “A framework for recognition and prediction of human motions in human-robot collaboration using probabilistic motion models,” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5151–5158, 2020.
- [17] T. Smith, Y. Chen, N. Hewitt, B. Hu, and Y. Gu, “Socially aware robot obstacle avoidance considering human intention and preferences,” *International Journal of Social Robotics*, vol. 1, pp. 1–18, 2021.
- [18] Z. Xu, X. Zhou, H. Wu, X. Li, and S. Li, “Motion planning of manipulators for simultaneous obstacle avoidance and target tracking: an rnn approach with guaranteed performance,” *IEEE Transactions on Industrial Electronics*, vol. 69, pp. 3887–3897, 2021.
- [19] X. Liu, C. Ma, and C. Yang, “Power station flue gas desulfurization system based on automatic online monitoring platform,” *Journal of Digital Information Management*, vol. 13, no. 06, pp. 480–488, 2015.
- [20] R. Huang, “Framework for a smart adult education environment,” *World Transactions on Engineering and Technology Education*, vol. 13, no. 4, pp. 637–641, 2015.
- [21] Y. F. Golubev and V.V. Koryanov, “Overcoming obstacles in the form of a stack of rough cylinders by an insectomorphic robot,” *Journal of Computer and Systems Sciences International*, vol. 61, no. 2, pp. 253–269, 2022.
- [22] X. Chen, Z. Huang, Y. Sun, Y. Zhong, R. Gu, and L. Bai, “Online on-road motion planning based on hybrid potential field model for car-like robot,” *Journal of Intelligent and Robotic Systems*, vol. 105, no. 1, p. 7, 2022.
- [23] H. Kong, C. Yang, G. Li, and S. L. Dai, “A Sema-Shared Control System with No-Target Obstacle Avoidance for Omnidirectional mobile Robots,” *IEEE Access*, vol. 8, pp. 26030–26040, 2020.
- [24] J. Gu, W. Wang, R. Yin, C. V. Truong, B. P. Ganthia, and B. P. Ganthia, “Complex circuit simulation and nonlinear characteristics analysis of GaN power switching device,” *Nonlinear Engineering*, vol. 10, no. 1, pp. 555–562, 2021.
- [25] M. Neloy, M. Das, P. Barua, A. Pathak, and S. U. Rahat, “An intelligent obstacle and edge recognition system using bug algorithm,” *American Scientific Research Journal for Engineering, Technology, and Sciences*, vol. 64, no. 1, pp. 133–143, 2020.