Hindawi
Mobile Information Systems
Volume 2021, Article ID 6617286, 11 pages
https://doi.org/10.1155/2021/6617286

Research Article

A 3D Machine Vision-Enabled Intelligent Robot Architecture

Yanjun Zhang,[1,2] Jianxin zhao,[1,2] and Heyong Han[1,2]

[1]*College of Mechanical Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, China*
[2]*Research Center for Basic Hydraulic Components and Intelligent Manufacturing Engineering of Major Equipment, Taiyuan 030024, China*

Correspondence should be addressed to Heyong Han; hhy@tyust.edu.cn

Received 15 December 2020; Revised 7 February 2021; Accepted 23 February 2021; Published 4 March 2021

Academic Editor: Mian Ahmad Jan

Copyright © 2021 Yanjun Zhang et al. 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.

In this paper, the principle of camera imaging is studied, and the transformation model of camera calibration is analyzed. Based on Zhang Zhengyou's camera calibration method, an automatic calibration method for monocular and binocular cameras is developed on a multichannel vision platform. The automatic calibration of camera parameters using human-machine interface of the host computer is realized. Based on the principle of binocular vision, a feasible three-dimensional positioning method for binocular target points is proposed and evaluated to provide binocular three-dimensional positioning of target in simple environment. Based on the designed multichannel vision platform, image acquisition, preprocessing, image display, monocular and binocular automatic calibration, and binocular three-dimensional positioning experiments are conducted. Moreover, the positioning error is analyzed, and the effectiveness of the binocular vision module is verified to justify the robustness of our approach.

1. Introduction

Internet of Robotic Things (IoRT) allows intelligent physical devices to monitor various events, gather data from multiple sources, and fuse them to extract useful features [1]. IoRT uses local and distributed intelligence for determining the best course of action. Based on this intelligence, various actions are taken and objects of the real-world are manipulated. IoRT allows the Internet of Things (IoT) and robotic devices to intersect by providing advanced robotic features. The concept of IoRT has made the realization of advanced and novel applications and has felt its presence in industries, businesses, and various other sectors. IoRT works based on the concept of computer vision. Though computer vision research has made considerable progress, there are also many controversies. According to the school of three-dimensional reconstruction led by Marr, vision is a procedure for information processing to obtain the description of external world, i.e., to recover three-dimensional scenes from image features such as edge points, contours, lines, curves, textures, grey levels, and colors. The features are obtained from either single or multiple two-dimensional images, and then used to recognize, locate, and analyze the three-dimensional scenes [2, 3]. In fact, the inverse process of image restoration often leads to unsolvability or instability of Horn's optical flow field constraint equation due to some ill-posed problems in the scene. Many works have been carried out around this problem, but so far, no real successful practical system has been developed.

More than 70% of the external information received by human beings is obtained through vision. As an imitation product, the realization and generalization of the vision system of home service robot is particularly important [4, 5]. First of all, the visual sensor has the advantages of low cost, abundant information, and high reliability. Secondly, the environment of intelligent service robots is usually an unstructured indoor environment. Vision can provide rich environmental information for robots, so as to achieve navigation, recognition, control, and other tasks. As a general-purpose robot vision platform, the system should have abundant and universal communication interfaces, so as to facilitate the exchange of information with various robot control hosts [6–8]. The system should have powerful processing functions, because the amount of visual information data is extremely large, and the requirement of video processing for processors is very high. Besides, the system

should also have good software and hardware expansion capabilities to facilitate the expansion of the system functions.

At the end of 1980s and the beginning of 1990s, a behavioural and application-oriented purposivism school emerged. It puts forward the concepts of active vision and qualitative vision. It says that vision is purposeful and pursues the purpose more actively. Some objects in the scene need quantitative analysis, while others need only qualitative description [9–11]. They accomplish their tasks together according to a certain purpose. This viewpoint has been criticized by the "reconstruction" school, which holds that "purposivism" uses specific tasks as restrictions to avoid problems and deal with facts. Secondly, the active sensing technology for acquiring scene data has many problems, such as single function and large noise impact, which lead to its limitations on the use of occasions and objects. On the other hand, from the initial remote-control manipulator to the intelligent robot with certain intelligence, the robot has gone through half a century, forming a unique information, energy, and material transformation feature system of the robot.

Machine vision has attained significant attention in recent times. In fact, machine vision is used as a unit of robot to feel the external information, its theory and methods, which cannot be separated from the large-scale system requirements of robots to discuss [12–15]. In this paper, we address all these issues and concerns. The major contributions of this paper are as follows:

(1) We start the discussion with the architecture of general-purpose intelligent robots and aim at different viewpoints in machine vision research in this context. After analyzing the generality and particularity, this paper puts forward a general research scheme of hierarchical modular structure of machine vision, and a vision integrated sensor supporting this system.

(2) Using experimental results, we justify the validity and feasibility of designing a 3D machine vision-enabled intelligent robot architecture.

The rest of the paper is organized as follows. In Section 2, we present the platform required for designing a robot vision system. In Section 3, we discuss the realization of binocular camera 3D vision location. In Section 4, the experimental process and result analysis are discussed. Finally, the paper is concluded and future research directions are provided in Section 5.

2. Platform Design of the Robot Vision System

Images can be classified into two categories according to their description properties: intrinsic and nonintrinsic. The former refers to the images that represent the objective characteristics of the scene and the scene that have nothing to do with the nature of the observer and the collector, for example, the most commonly used depth images, where each pixel value represents the distance between the pixel and the camera. The physical quantity of the latter is related not only to the scene, but also to the nature of the observer and the collector as well as the surrounding environment [16–18]. A typical representation is the grey image. Each pixel reflects the light radiation received and reflected by the observer. To accomplish the vision task mentioned above is to solve the problem of restoring intrinsic features from nonintrinsic images. The imaging process of noneigenvalue image is a degenerate transformation process, in which a lot of physical information about scene is mixed up in noneigenvalue image, but these pieces of information are not completely lost. There are always a lot of redundant information in the image, and various processing techniques (such as distance transformation) can be used to eliminate the degradation and "restoration of" intrinsic characteristics. In order to restore the scene, it is necessary to directly collect intrinsic images or nonintrinsic images containing stereo information. In this way, certain image acquisition equipment (imaging device) and a certain image acquisition mode (imaging mode) are needed. In the following subsections, we first discuss designing an image acquisition device followed by system architecture.

2.1. Design of Image Acquisition Device. The imaging devices are stimulated by external radiation (visible light radiation, etc.), which is converted into electrical signals, and then this signal response is converted into digital signals that can be processed with a computer. There are two functions of the image acquisition device. One is to accept the excitation, and the other is the conversion of analog signal. There are many kinds of image acquisition devices, usually consisting of video source and video acquisition. Of course, digital technology has made it possible to integrate these two functions into a packaged device such as digital camera on hardware [19]. Considering the maturity, advantages, and environmental adaptability of technology, the structure composed of camera and acquisition card is still adopted in mobile robot vision system at present (the first half of Figure 1). The hardware of the image capture card will bring greater flexibility and performance improvement. Generally, video cards integrate user conversion, signal amplification, filtering, and other processing of signals. With the progress of technology and the development of manufacturing industry, some cards also integrate functions that are usually completed by software, such as directly compressing into a certain data format (such as MPEG4) or even matrix operation, which will greatly speed up data processing.

At present, the charging source device is mainly used in video devices. It is a kind of solid-state camera device which works by charge storage, transmission, and readout. It has the advantages of low reading noise, large dynamic range, and high response sensitivity. With the continuous progress of technology, its resolution is gradually improved, and color restoration is becoming more realistic. It is a high cost-effective imaging device at present. The working principle of CCD is that the object reflects light and propagates to the lens. The lens focuses on the CCD chip. According to the intensity of the light, the CCD accumulates corresponding charges and discharges periodically to produce electrical

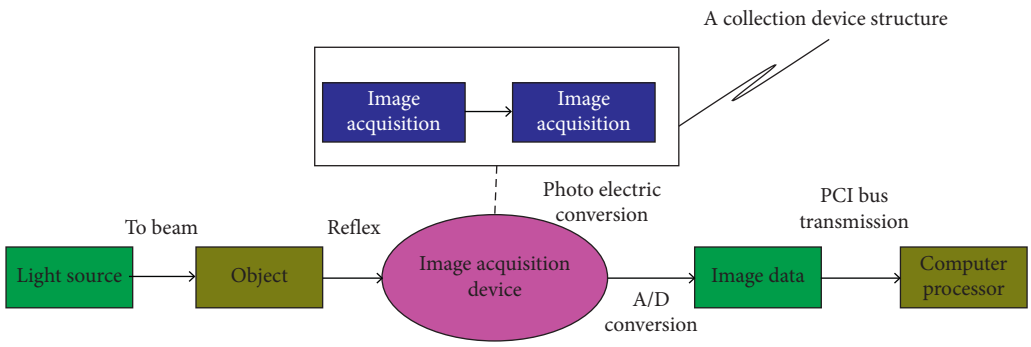

Figure 1: Image acquisition process.

signals representing a picture [20]. After filtering and amplification, a standard compound video signal is generated through the output terminal of the camera. CCD is the main sensing component of the camera. It has the characteristics of high sensitivity, small distortion, long life, antivibration, antimagnetic field, small volume, and no residual shadow. CCD can change light into charge, store and transfer charge, and take out the stored charge to change voltage. Therefore, it is an ideal camera element. It is a new device instead of traditional camera tube sensor. Because of the richness and importance of color information, we use color CCD as image collector. The image acquisition process is shown in Figure 1.

The acquisition card is the process of importing video data or mixed data of video and audio output from other data sources into the computer, transforming them into digital signals by processing, and storing these pieces of information. This analog-to-digital conversion is carried out by the acquisition chip or related software on the video acquisition card, so that the acquisition card can be divided into real-time acquisition card and nonreal-time acquisition card. According to the data sources, it can be divided into three categories: digital acquisition, analog acquisition, and digital-analog acquisition. Considering that more than one video source is needed to construct stereo vision, a multichannel acquisition card will be more feasible. As mentioned above, data compression will be extremely important in this context. Therefore, the acquisition card has the following functions: hardware compression, 1 card multiplex, and real-time analog acquisition.

2.2. System Structure. In our proposed approach, imaging is determined mainly by the position and movement of the three sources, the video source, and the scenery. The simplest part is monocular imaging, which uses a video source to take a picture from the scene in a fixed position. Stereo vision is a method of perceiving distance by binocular cues. Two video sources are used to get images of the same scene in one location. Parallax generated between two images is used to help find the distance between video sources and objects so as to realize the perception of three-dimensional information. In principle, it is based on triangulation. Therefore, the specific means of implementation are not limited to the use of two cameras as the source of video. For example, it is also possible to use one camera to take images of the same scene at multiple locations, and to use more than two cameras to image the same scene from different locations.

Let us analyze and compare the commonly used stereoscopic imaging methods. First of all, the general principle of binocular vision is expounded. Figure 2 is a hint of the simplest case.

In this figure, C_L and C_R are the lens centers of the left and right cameras, the distance between them is b (called baseline), and the focal length between the cameras is f. Here, P is the point on the surface of the object. The projection points on the two cameras are rulers, where the distance between P and the central line of the lens of the two cameras is d, the vertical line from C_L and C_R to the horizon is A_L and A_R, the vertical line from P to the horizon is B, and the intersection point between the extension line and the central line of the lens is C. In this way, it is easy to get the distance between the object point and the camera from two pairs of triangles: ΔPP_RB and ΔPC_RC, $\Delta P_LA_LC_L$ and ΔP_LBP.

$$d = f\frac{a + l_a}{b} = \frac{bf}{l_a - l_b}. \quad (1)$$

Obviously, d is related to b and $l_a - l_b$, one of which is called parallax formed by P in the two viewing planes. For a visual system, b and f are known, so the key to achieve ranging is to obtain parallax, that is, to achieve the correspondence of a point in space between the projection points on the left and right images, which is the process of image registration.

The principle of binocular imaging is very simple, but image registration is still a difficult problem. Stereo matching is the most difficult and important step in stereo vision. When a three-dimensional scene is projected into a two-dimensional image, the image of the same scene will be very different in different viewpoints, and many changes and unpredictable factors in the scene, such as illumination conditions, noise, geometric shape of the scene, and the characteristics of the camera, are integrated into a single image color/gray value. Conversely, it is a morbid problem to determine the three-dimensional information from the single image color/gray value, which has not been solved well up to now. Binocular registration has a large search space for corresponding points, which is often complex. Therefore, a trinocular registration method is proposed, that is, taking three images of an object at the same time, so as to increase

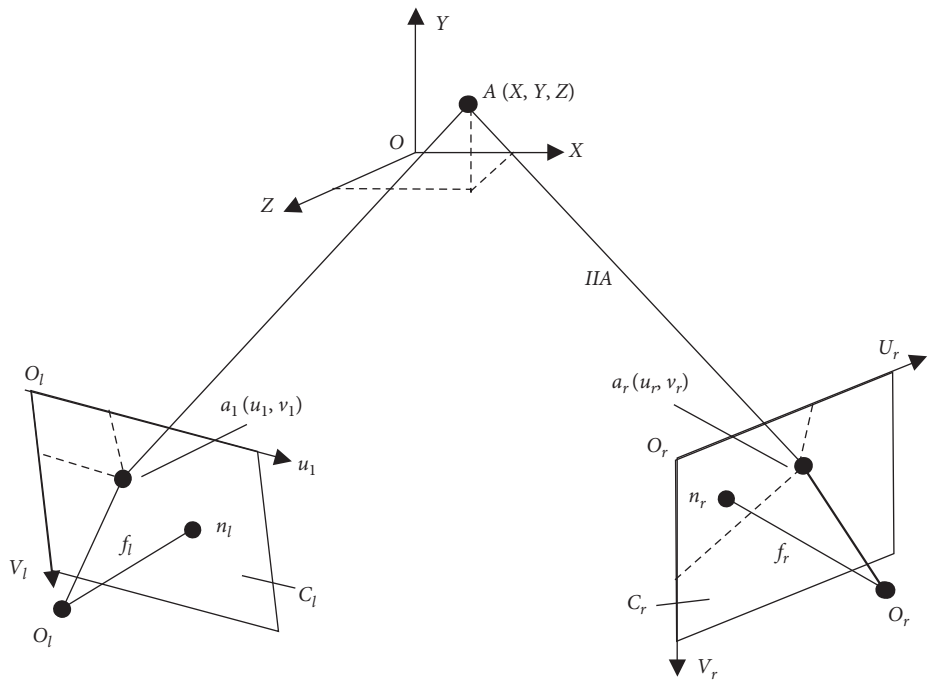

Figure 2: Geometric principle of binocular vision parallax ranging.

the constraints of registration and simplify the registration problem.

Next, we discuss the light displacement imaging and structural light imaging. The basic principle of the former is that the brightness of the same scene is different under different illumination conditions, so the surface orientation of the object can be obtained by light moving image, which is different from several other techniques, so that the absolute depth information cannot be obtained, but the surface shape description of the object can be obtained. Since our mobile robot is moving in an unknown environment, it will be difficult to control the mobile light source, so we do not consider this technique. The latter is to explain the surface shape of the scenery by the projection mode collected. This method can fix the light source and the acquisition and rotate the scenery. It can also fix the scenery and move the light source and the acquisition around the scenery together. With the progress of science and technology, especially the emergence of various new sensors, more means of acquiring distance appear. If one eye in binocular vision is replaced by a laser source, the light source produces a series of point or line lasers to illuminate the object surface, and the illumination part is recorded by the light sensitive camera, the depth information can be obtained directly. However, other information on the surface of the object will be ignored. The schematic diagram of robot vision information acquisition system is shown in Figure 3.

Trinocular and multiocular imaging, which involves the calibration of multiple cameras, is mostly used in three-dimensional reconstruction in virtual reality human-computer interaction and other fields. For mobile robots, the uncertainty of their environment will bring many unpredictable factors. At the same time, the processing of some environmental information also needs some real-time requirements (such as obstacle avoidance). Therefore, too complex imaging methods will have many limitations in practical applications. Therefore, the general mobile robot system adopts binocular imaging in visual information. However, for unknown environments, the task facing robots is not only to ensure that they can safely avoid obstacles and move smoothly, but also to obtain as much environmental information as possible in order to realize the process of recognizing the unknown. The real objective environment is three-dimensional, and the reconstruction of three-dimensional environment has naturally become an important aspect of cognition. Therefore, we also attach enough importance to this function in our system. Obviously, the existing imaging methods alone cannot meet the needs of vision system for mobile robots in unknown environments. The combination of multiple imaging methods will be a worthy consideration. Considering the above-mentioned imaging methods and system requirements, we decided to construct a visual platform by combining binocular imaging and monocular imaging, in order to make full use of the advantages of various imaging methods. Redundant design methods will lead to the avoidance of using a single method as far as possible, while information fusion and flexible combination of imaging methods will bring higher reliability, flexibility, and fault tolerance.

3. Realization of Binocular Camera 3D Vision Location

In this section, we discuss the 3D vision location of binocular camera. First, we discuss the realization of binocular camera

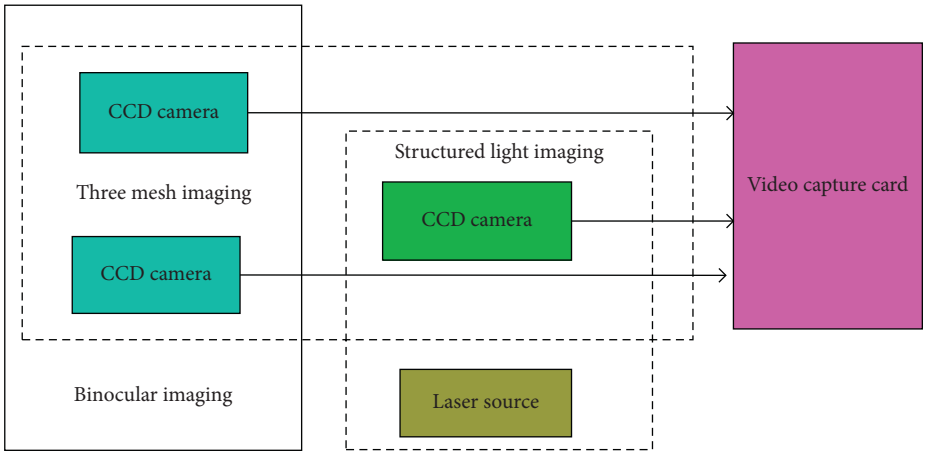

Figure 3: Schematic diagram of robot vision information acquisition system.

calibration in Section 3.1 followed by binocular 3D positioning of target points in Section 3.2.

3.1. Realization of Binocular Camera Calibration. Binocular vision refers to the simultaneous imaging of the overlapping parts of the field by two cameras. The matching algorithm is used to obtain the image position of the same spatial point in two cameras. Based on the matching algorithm, the three-dimensional location of the scene is realized according to the internal parameters of the two cameras and the spatial position and attitude relationship between the two cameras. The task of dual target determination is to determine the internal parameters of two cameras and the relationship between the relative position and attitude of two cameras. The internal parameters of the two cameras can be achieved by performing single target calibration algorithm for them, and the spatial relative position and attitude relationship between them can be obtained by transforming the external parameters of the two cameras for the same position calibration target.

When the target is set at a certain position, the transformation matrix from the world coordinate system to the left-eye camera (i.e., the external reference matrix of the left-eye camera) is $[R_1]$, and the transformation matrix from the world coordinate system to the right-eye camera (i.e., the external reference matrix of the right-eye camera) is $[R_2]$. Then, the relationship between (x, y, z) points in the world coordinate system and (x_1, y_1, z_1) coordinates in the right camera coordinate system and (x_2, y_2, z_2) coordinates in the left camera coordinate system satisfies the following formulas:

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = [R_2] \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \tag{2}$$

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = [R_1] \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \tag{3}$$

Both sides can get the same $[R_2]^{-1}$:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [R_2]^{-1} \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}. \tag{4}$$

The coordinate transformation relation of the same point in the left and right cameras can be obtained as follows:

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = [R_1][R_2] \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}. \tag{5}$$

That is, the position and attitude transformation matrix of left and right cameras is $[R_1][R_2]^{-1}$.

The process of binocular automatic calibration software is shown in Figure 4.

3.2. Realization of Binocular 3D Positioning of Target Points. According to the analysis of camera imaging transformation and the principle of three-dimensional positioning of target points, the focus and target points of two cameras can be transformed into the same coordinate system through a series of coordinate transformation. In this paper, we choose to convert them all into the camera coordinate system of the left-eye camera, and then use these four points to establish two equations of imaging light. By solving the equations composed of these two equations, we can get the coordinates of the target point in this coordinate system. However, due to the following reasons, this method is almost infeasible in practical applications:

(1) Firstly, there is a certain gap between the practical camera and the simplified camera model, and the camera that meets the ideal camera model actually does not exist

(2) Secondly, in the process of camera calibration, because of the simplification of camera model, the discretization of digital imaging, and the calculation errors introduced in the process of numerical

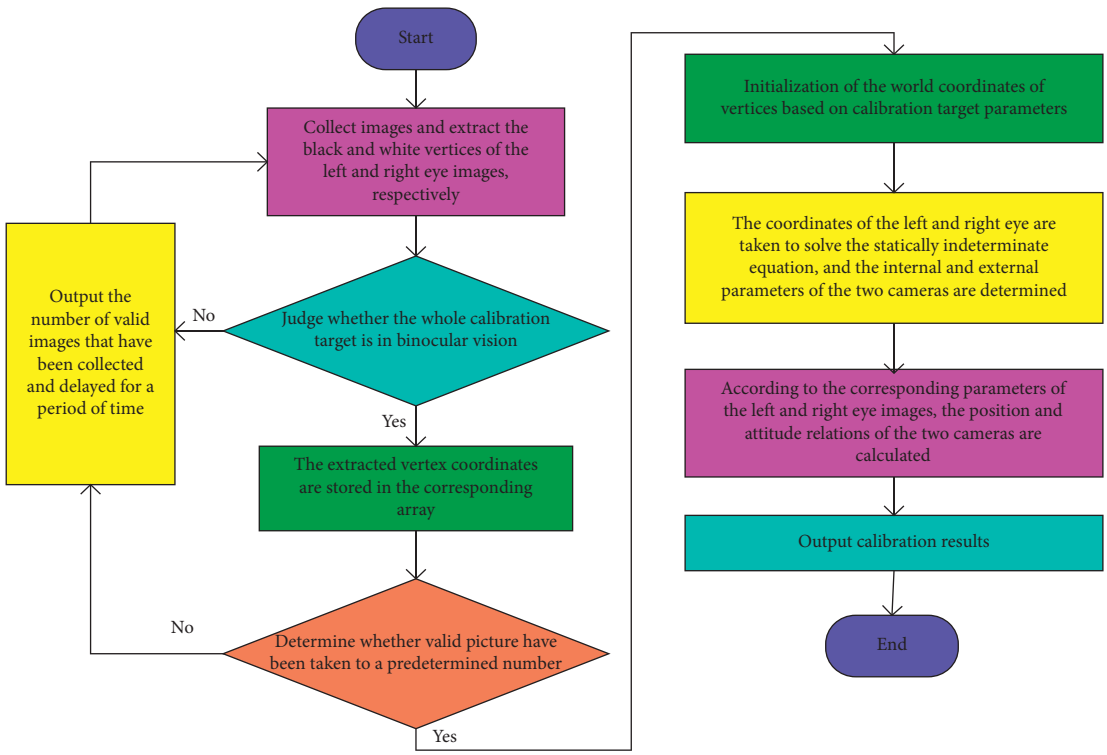

Figure 4: Software flow chart of binocular automatic calibration algorithm.

solution, it is almost impossible for the two transformed imaging rays to be intersecting lines

(3) Finally, even if two imaging rays intersect, it is a relatively difficult problem to solve directly the equations whose parameters change at any time by digital method

Based on the above reasons, in order to achieve the three-dimensional positioning of the target point, this paper adopts the method of restoring two imaging rays, using the method of solving the points in the common perpendicular line segment of the spatial heterogeneous straight line to approximately replace the position of the target point, and gives the length of the common perpendicular line segment as the credible error range of the calculation. In the process of solving the problem, the general equation of the imaging light of the left camera is set as follows: the imaging light of the left camera cannot pass through the origin of the camera coordinate in the XY plane of the camera coordinate system at the same time, and the imaging light of the left camera must pass through the origin of the camera coordinate.

$$\frac{x}{a_0} = \frac{y}{b_0} = \frac{z}{1}. \tag{6}$$

Similarly, the general equation for imaging light on the right camera is set as

$$\frac{x - x_1}{a_1} = \frac{y - y_1}{b_1} = \frac{z - z_1}{1}. \tag{7}$$

The parameters of $a_0, b_0, a_1, b_1, x_1, y_1, z_1$ in the general formula are determined by the coordinates of image points and focus points converted to the coordinates of the left-eye camera.

According to the theory of analytic geometry, the algorithm for solving the length of common vertical line and its midpoint coordinates of the two straight lines can be realized by writing C language code. The software flow of binocular 3D positioning algorithm for target point is shown in Figure 5.

In this paper, the principle of camera imaging is studied, and the transformation model of camera calibration is analyzed. Based on Zhang Zhengyou's camera calibration method, an automatic calibration method for monocular and binocular cameras is developed on a multichannel vision platform, which realizes the automatic calibration of camera parameters under the call of human-machine interface of the host computer. Based on the principle of binocular vision, a feasible three-dimensional positioning method for binocular target points is proposed, and the algorithm implementation program is compiled to provide a basis for the experiment of binocular three-dimensional positioning of target in simple environment.

4. Experimental Process and Result Analysis

In this section, first we discuss image acquisition and preprocessing followed by calibration of monocular and binocular cameras.

4.1. Image Acquisition and Preprocessing. The multichannel vision platform designed in this paper supports four-channel video input at most. The experiment of image acquisition

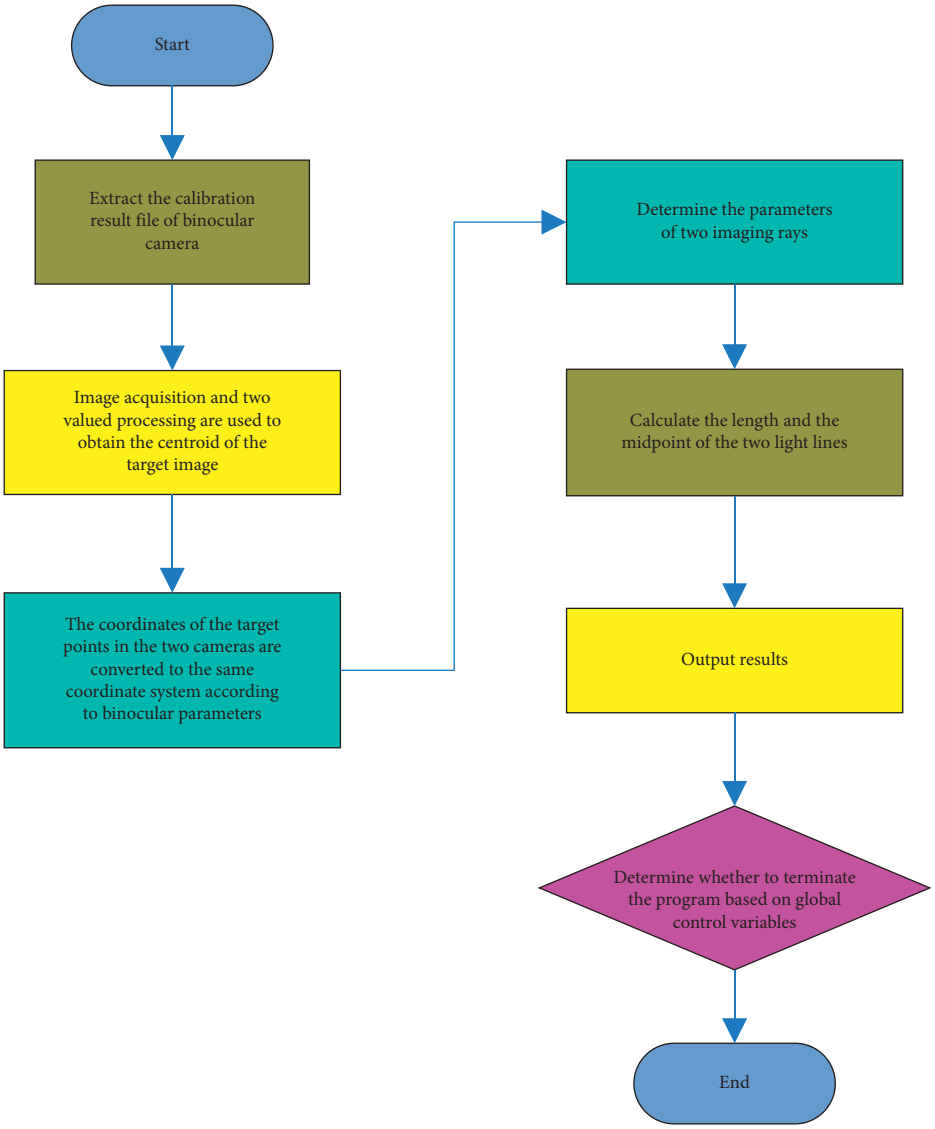

Figure 5: Software flow chart of binocular 3D location algorithm for target points.

and preprocessing is to verify the effectiveness of the system's input-output interface control and the ability of the system's dual-core architecture to coordinate work. The software structure used in the experiment of single-channel image acquisition and preprocessing as shown in Figure 6 includes three parts: PC user interface program, embedded system management program on ARM, and preprocessing algorithm program on DSP.

The execution process of the program is that when a preprocessing algorithm instruction is pressed on the PC user interface, the PC sends the corresponding control command string to the embedded system through the serial port. After receiving the command string, the embedded system management program running on the ARM processor starts the video capture threads by modifying the global control variables in turn. Control the CCD camera to capture the image and store the image in the corresponding memory space, start the algorithm call thread, call the corresponding preprocessing algorithm of the DSP processor to execute the program, realize the corresponding preprocessing of the collected image, and control the video back end to display the processed image to the TV. After that, the communication control thread monitors the instruction of the network port and serial port, decides whether to transmit the processed image to the host computer according to the instruction state of the network port, changes the preprocessing algorithm, or terminates the processing program according to the instruction of the serial port.

4.2. Automatic Calibration of Monocular and Binocular Cameras. As shown in Figure 7, the calibration algorithm of monocular and binocular camera automatic calibration test runs on PC, and the embedded system runs in the image acquisition mode. It receives the instructions of the host computer to provide single- or multichannel image information for the host computer and displays the video information collected by the camera to the TV.

In the PC user interface, press the single target or binocular calibration button to start the calibration program. Then,

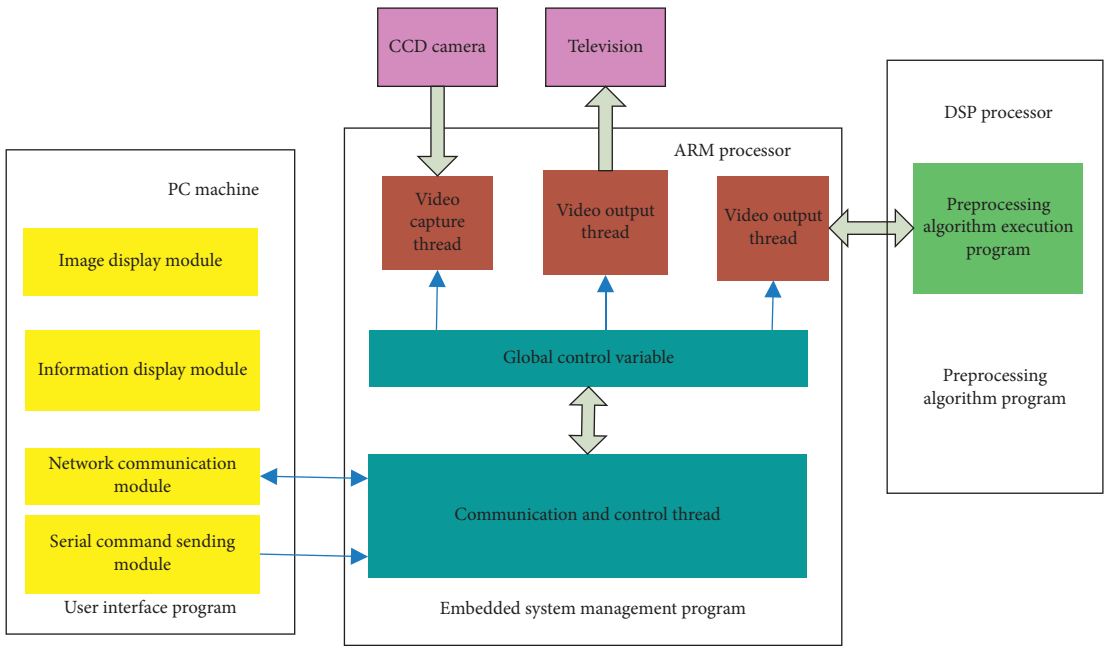

Figure 6: Single-channel image acquisition and preprocessing experimental system structure diagram.

according to the prompt information displayed on the PC user interface, the position and attitude of the calibration target can be changed so that the system can acquire 10 valid calibration images to complete the calibration under the condition of ensuring the whole calibration target in the field of vision. The calibration results are displayed on the man-machine interface. At the same time, the system stores the calibration results in document form and saves 10 calibration images in JPEG format. In order to evaluate the results of automatic calibration algorithm, 10 calibration images used in automatic calibration are calibrated with the calibration toolbox of MATLAB. The comparison of calibration results is shown in Table 1.

Compared with the calibration results of the MATLAB toolbox, the error of focal length calibration is about 5%, and the deviation of image center is basically zero. The results of automatic binocular calibration are shown in Table 2 and Figure 8.

Experiments show that the automatic calibration method is feasible. The validity of the calibration value needs further verification by three-dimensional positioning experiment.

4.3. Binocular 3D Positioning Experiment. As shown in Figure 9, the binocular three-dimensional positioning experimental system is similar to the image preprocessing experimental system, which consists of three parts. The three-dimensional positioning algorithm runs on the DSP, and the processing results are displayed on the PC user interface through network communication.

Binocular localization algorithm is sensitive to errors in Z-direction (along the optical axis). The comparison between the location output and the measured data is shown in Table 3 and Figure 10.

The reliability of the proposed binocular three-dimensional positioning method is verified by positioning experiments, and the validity of the results of binocular automatic calibration is indirectly proved. The errors between positioning results and actual measurements in the range of 2 m are less than 10 cm. The larger errors mainly occur in the range of Z coordinates less than 50 cm and greater than 150 cm. The main sources of errors are two aspects. On the one hand, there is a gap between the mathematical model established by the positioning method and the actual camera imaging model. The positioning model algorithm itself is sensitive to the Z-direction error, which results in a larger error when the Z coordinate is greater than 150 cm. On the other hand, because the errors caused by image distortion during camera imaging are not taken into account in the ranging model, the tangential distortion of image has a greater impact on the results when Z coordinates are smaller, which results in larger errors when Z coordinates are less than 50 cm.

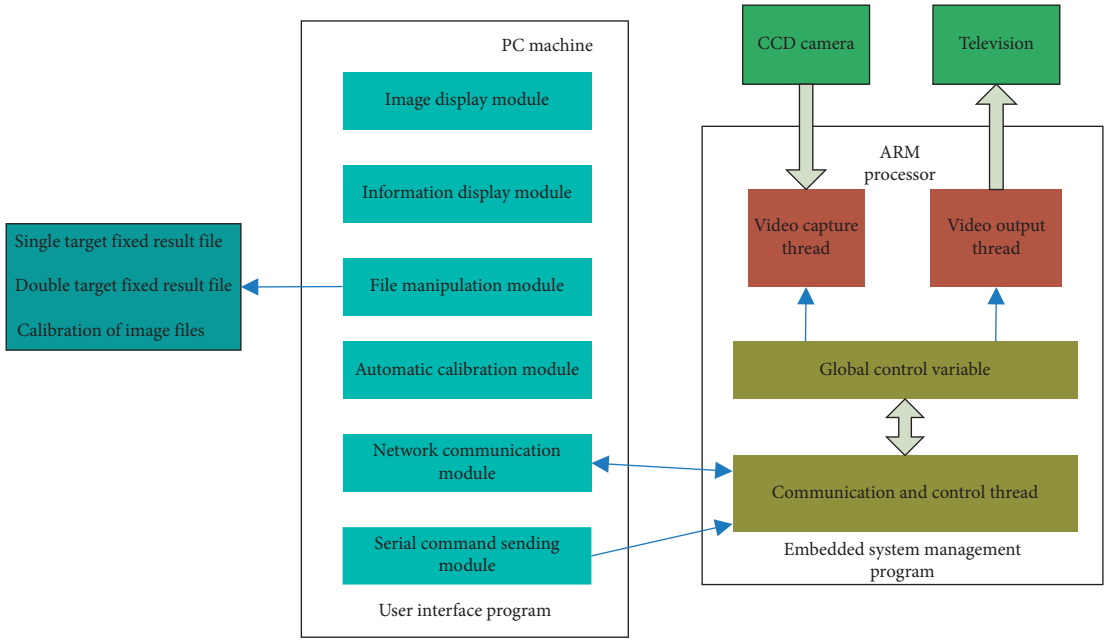

Figure 7: Structure chart of automatic calibration system for single and binocular vision.

Table 1: Comparison between automatic calibration results of monocular and calibration results of MATLAB toolbox.

Calibration parameters	f_x	f_y	C_x	C_y
Automatic calibration results	798.48	878.67	360.6	240.6
MATLAB calibration results	813.85	886.79	360.6	240.1

Table 2: Automatic calibration results of binocular.

Calibration parameters	f_x	f_y	C_x	C_y
Left-eye camera	394.22	435.16	180.6	153.1
Right-eye camera	394.22	427.62	180.6	153.1

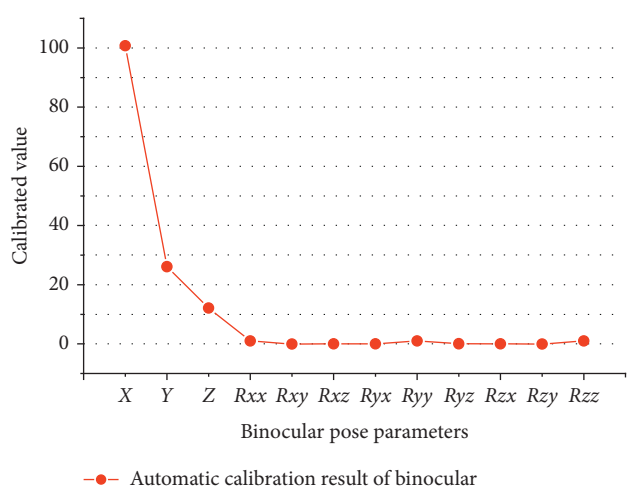

Figure 8: Automatic calibration results of binocular.

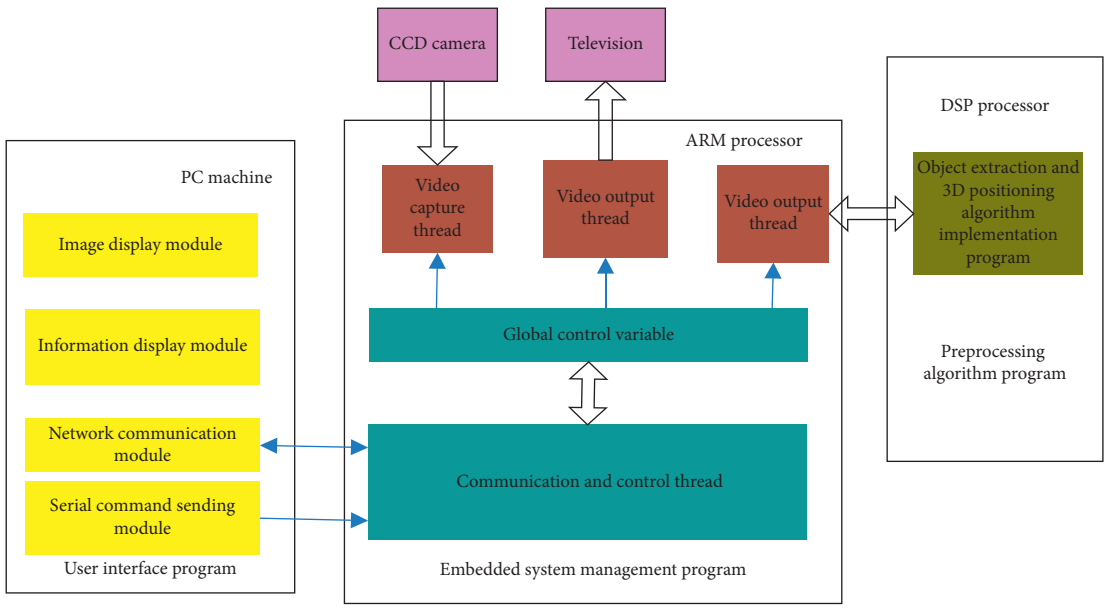

Figure 9: System structure diagram of binocular 3D positioning experiment.

Table 3: Comparison table of Z-axis coordinate experiment results of binocular 3D positioning (unit: cm).

Experimental value	25	50	75	100	125	150	175	200
Location experiment results	24.63	54.8	71.36	99.97	127.42	164.7	180.73	208.75
Location error estimation	4.14	2.36	1.46	1.28	1.17	4.27	7.76	12.57
Measured error	3.63	3.79	0.36	0.25	1.42	3.7	4.73	7.73

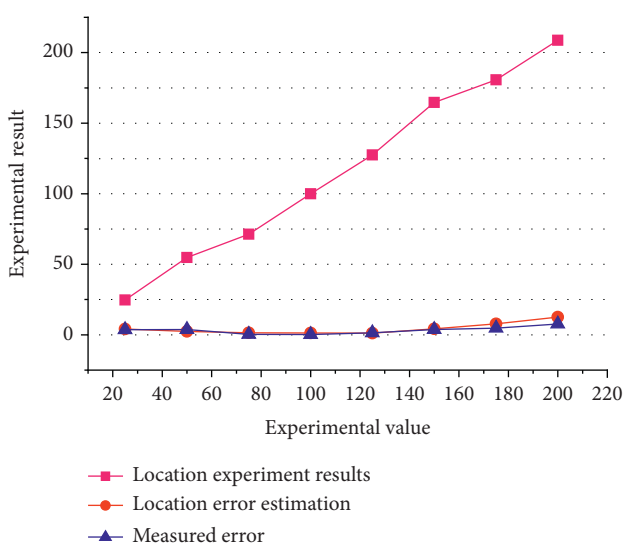

Figure 10: Comparison table of Z-axis coordinate experiment results of binocular 3D positioning.

5. Conclusion

Using the human intelligence system, and combining the theory and methods of artificial intelligence and computer vision, this paper proposes an intelligent reconfigurable mobile robot vision system and studies the appropriate structure on the hardware and software levels. On the basis of the existing work, the function expansion method is adopted to gradually improve each module so that it can achieve the desired effect with the support of a large system structure. The imaging principle of the camera is studied, and the transformation model of the camera calibration is analyzed. Based on the Zhang Zhengyou camera calibration method, the automatic calibration method of the single-binocular camera is developed on the multichannel vision platform, and the camera automatic parameters under the human-machine interface are realized. A feasible three-dimensional positioning method for binocular target points is proposed based on the principle of binocular vision, and the algorithmic implementation is provided for the binocular three-dimensional positioning experiment of the target in a simple environment. Finally, image acquisition, preprocessing, image display, single and double eye automatic calibration, and binocular three-dimensional positioning experiment were examined and evaluated via the experimental results. The positioning error was analyzed to verify the effectiveness of the binocular vision module.

Data Availability

All data can be obtained upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the College of Mechanical Engineering, Taiyuan University of Science and Technology, Shanxi.

References

[1] P. P. Ray, "Internet of robotic things: concept, technologies, and challenges," *IEEE Access*, vol. 4, pp. 9489–9500, 2016.

[2] B. Tovar, F. Cohen, L. Bobadilla et al., "Combinatorial filters: sensor beams, obstacles, and possible paths," *Acm Transactions on Sensor Networks*, vol. 10, no. 3, pp. 317–332, 2013.

[3] P. Leica, J. M. Toibero, F. Roberti, and R. Carelli, "Switched control to robot-human bilateral interaction for guiding people," *Journal of Intelligent & Robotic Systems*, vol. 77, no. 1, pp. 73–93, 2015.

[4] M. Birem and F. berry, "DreamCam: a modular FPGA-based smart camera architecture," *Journal of Systems Architecture*, vol. 60, no. 6, pp. 519–527, 2014.

[5] C. Astua, R. Barber, J. Crespo, and A. Jardon, "Object detection techniques applied on mobile robot semantic navigation," *Sensors*, vol. 14, no. 4, p. 6734, 2014.

[6] W. Song, G. Wang, J. Xiao, G. Wang, and Y. Hong, "Research on multi-robot open architecture of an intelligent CNC system based on parameter-driven technology," *Robotics and Computer-Integrated Manufacturing*, vol. 28, no. 3, pp. 326–333, 2012.

[7] Y. Zhou and W. K. Huang, "The study on robot control using underwater sensor network," *Advanced Materials Research*, vol. 490-495, pp. 813–817, 2012.

[8] Y. Zhi-Gang and A. Yi, "Research of intelligent glue-coating robot based on visual servo," *Shanxi Electronic Technology*, vol. 24, pp. 2165–2171, 2012.

[9] M. D. Truong, N. N. Yu Feng, and L. Ricky, "Intelligent autonomous navigation system for the wheeled mobile robot," *Advanced Materials Research*, vol. 383-390, 2012.

[10] H. Sun, Z. Shi, J. Wu et al., "Research on multi-functional intelligent boiler slag-cleaning robot at the coal-fired power stations," *Lecture Notes in Electrical Engineering*, vol. 143, no. 11, pp. 601–608, 2012.

[11] T. Krajnik, N. Matias, J. Faigl et al., "A practical multirobot localization system," *Journal of Intelligent & Robotic Systems*, vol. 76, no. 3-4, pp. 539–562, 2014.

[12] H. Zhao, L. Xiong, Y. Liu et al., "Dynamic environment sensing using an intelligent vehicle," *Contemporary Issues in Early Childhood*, vol. 6, no. 2, pp. 63–74, 2012.

[13] X. W. . Wang, "Three-dimensional vision applications in GTAW process modeling and control," *International Journal of Advanced Manufacturing Technology*, vol. 80, no. 9-12, pp. 1601–1611, 2015.

[14] H. Mousazadeh, "A technical review on navigation systems of agricultural autonomous off-road vehicles," *Journal of Terramechanics*, vol. 50, no. 3, pp. 211–232, 2013.

[15] A. Araújo, D. Portugal, M. S. Couceiro, and R. P. Rocha, "Integrating arduino-based educational mobile robots in ROS," *Journal of Intelligent & Robotic Systems*, vol. 77, no. 2, pp. 281–298, 2015.

[16] M. Carol, F. M. Iván, C. Pascual, and S.-L. Jose Luis, "A hierarchical tracking strategy for vision-based applications on-board UAVs," *Journal of Intelligent & Robotic Systems*, vol. 72, no. 3-4, pp. 517–539, 2013.

[17] F. Flacco, T. Kroeger, A. D. Luca et al., "A depth space approach for evaluating distance to objects," *Journal of Intelligent & Robotic Systems*, vol. 80, no. 1, pp. 7–22, 2015.

[18] C. Rasmussen, K. Yuvraj, R. Vallett et al., "Towards functional labeling of utility vehicle point clouds for humanoid driving," . *Intelligent Service Robotics*, vol. 7, no. 3, pp. 133–143, 2013.

[19] S. Mohammed, C. Ramos, R. S. Chang et al., "Advanced sensor technology and applications in industrial control system," *International Journal of Distributed Sensor Networks*, vol. 2013, no. 3, 2 pages, 2013.

[20] M. Uhart, O. Patrouix, and Y. Aoustin, "Improving manufacturing of aeronautical parts with an enhanced industrial Robotised Fibre Placement Cell using an external force-vision scheme," *International Journal on Interactive Design & Manufacturing*, vol. 10, no. 1, pp. 1–21, 2015.