IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 18, 2020, accepted July 10, 2020, date of publication July 15, 2020, date of current version July 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009387

Coal Mine Rescue Robots Based on Binocular Vision: A Review of the State of the Art

GUODONG ZHAI, WENTAO ZHANG, WENYUAN HU, AND ZHENDONG JI
School of Mechanical Electronic & Information Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China

Corresponding author: Guodong Zhai (zgd@cumtb.edu.cn)

This work was supported in part by the National Training Program of Innovation and Entrepreneurship for Undergraduates under Grant 201911413045, in part by the Fundamental Research Funds for the Central Universities under Grant 2020YJSJD06, and in part by the Graduation Project (Entrepreneurship) of Practical Training Plan for Cross-Cultivation of High-Level Talents in Beijing Colleges and Universities under Grant 2019110750145 and Grant 2019110750146.

ABSTRACT Rescue work after a coal mine accident is fraught with challenges and dangers. Considering the safety of rescue workers and the urgency of a rescue mission, it is necessary to use coal mine rescue robots to perform the tasks of environmental detection and rescue. As a key part of the robot sensing system, a visual sensor can provide much information about a scene. Among vision sensor types, binocular vision has the advantages of being noncontact and passive, and it is the key technology for a robot to acquire obstacle information and reconstruct a three-dimensional scene. Therefore, coal mine rescue robots based on binocular vision have become a popular research topic in the field of mine safety. First, the research status of camera calibration and stereo vision matching for binocular vision is systematically introduced in this paper. Second, the latest research progress on coal mine rescue robots based on binocular vision is reviewed from the perspective of technological applications and development. Finally, the technical challenges and future development trends of binocular vision in coal mine rescue robots are described.

INDEX TERMS Coal mine rescue robots, binocular vision, camera calibration, stereo vision matching.

I. INTRODUCTION

Coal's share of primary energy is 27.2% according to the British Petroleum (BP) Statistical Review of World Energy released in 2019 [1]. Coal is still an extremely vital part of global energy patterns, and its mining methods are generally divided into open-pit mining and underground mining, the latter of which is considered one of the most dangerous jobs in the world [2]–[4]. Taking China as an example, according to statistics from the National Coal Mine Safety Administration, there were 198 coal mine accidents in 2018, with 303 deaths [5]. Emergency rescue must be carried out after a mine accident. However, compared with ground rescue, underground coal mine rescue faces particular challenges [6]. As shown in Fig. 1, according to the relevant research of Zhang *et al.* [7], the top three major or especially major types of coal mine accidents in China from 2001 to 2018 were gas accidents, fire accidents and flooding accidents. Additionally, there are many secondary disasters, including the above types of hazards, at accident sites. Therefore, to reduce the risk of secondary disasters, a variety of protective measures must be taken before rescuers can first enter a mine to carry out rescue work. Considering the existence of an optimal rescue window, an accident scene is in urgent need of robots that can replace the rescuers to explore the mine environment and carry out rescue work. A coal mine rescue robot (CMRR) can bring an appropriate amount of emergency supplies and various sensors to the accident scene, and the rescue command center can then formulate an efficient rescue strategy based on the information provided by the robot; this guarantees the smooth progress of various rescue operations as well as possible. The application of CMRRs not only reduces the threat to rescue worker safety but also improves the rescue efficiency [8], [9].

At present, most CMRRs are in the experimental research stage, and only a few CMRRs are involved in mine rescue. For instance, a CMRR from the Western Australia Water Service Company was allowed to participate in a search-and-rescue after a mine accident at New Zealand's Pike River coal mine on November 19, 2010 [10]. As shown in Fig. 2 (a), this robot was equipped with cameras, gas detection sensors, and other equipment and was remotely controlled by fiber optics. Similarly, on February 25, 2016, a coal mine accident occurred in the Northland Coal Mine near the city of Vorkuta in the Komi

The associate editor coordinating the review of this manuscript and approving it for publication was Abdel-Hamid Soliman.

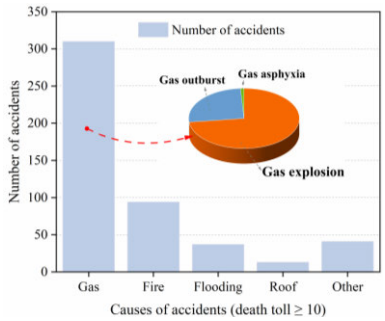

FIGURE 1. Statistics on the causes of major or especially major coal mine accidents in China from 2001 to 2018.

FIGURE 2. The application of CMRRs after mine accidents. (a) Robot from the Western Australia Water Service Company. (b) The Leader robot [10], [11].

Republic of Russia. The Leader robot was transported to the scene of the accident and entered the mine to explore in place of rescuers [11]; the robot is shown in Fig. 2 (b). This robot was loaded with cameras and other detection equipment that can operate in the unstructured environment of a coal mine.

In the unstructured environment of a coal mine, a robot must have the ability to perform autonomous obstacle avoidance when carrying out environmental detection and rescue, so it is necessary to conduct accurate three-dimensional (3D) reconstruction of the field environment. In general, vision provides most of the information about the environment; especially for industrial robots in harsh conditions, vision sensors are critical [12], [13]. In the field of robot vision, binocular vision has the two advantages of being noncontact and passive, and great progress has been made in its theory and practice. Binocular vision have become the main component of the vision modules of robot sensing systems; Fig. 3 shows the 3D model of a CMRR with vision sensors, named MSRBOTS, developed by a research team at the Beijing Institute of Technology, and it obtains information on the scene of a coal mine through two cameras to achieve the correct positioning [6]. Although binocular vision has been preliminarily applied to the sensing system of CMRRs, the existing technical level still needs to be improved, especially in the areas of camera calibration accuracy, image correction algorithms, stereo vision matching effects and 3D reconstruction models. This paper is founded on the research on CMRRs based on binocular vision over the past ten years, and it not only enables relevant researchers to understand the current research status but also provides references for future investigation.

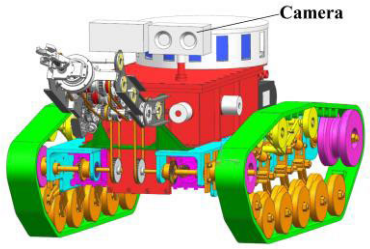

FIGURE 3. MSRBOTS, a CMRR equipped with vision sensors [6].

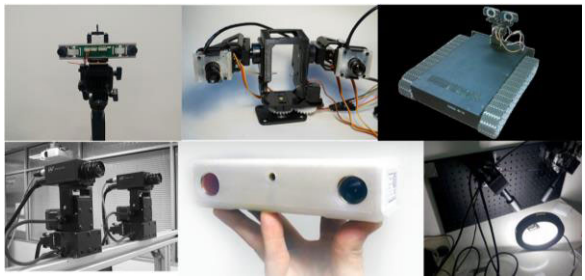

FIGURE 4. Binocular cameras or systems [18]–[23].

The remainder of the paper is organized as follows: Section 2 introduces the research progress on the key technologies of binocular vision, including camera calibration and stereo vision matching. Section 3 reviews the specific application of binocular vision in CMRRs. Section 4 describes the technical challenges and development trends of binocular vision in CMRRs. Finally, the conclusion is given in section 5.

II. OVERVIEW OF THE KEY TECHNOLOGIES OF BINOCULAR VISION

At the scene of an accident, the binocular vision module of a robot sensing system collects image information, conducts image analysis and processing, and performs obstacle recognition and 3D reconstruction. The concept of binocular vision originated from the basic vision theory established by Marr and Poggio [14]. Subsequently, Grimson [15] further supplemented and modified the theory.

A. THE PRINCIPLE OF BINOCULAR VISION

The basic principle of binocular vision can be expressed as follows: An object in a certain area is first imaged by two cameras, then the parallax map is obtained after a series of image processing steps, and finally, the depth information is recovered by triangulation and a 3D image is constructed [16]–[18]. Fig. 4 shows a series of common binocular cameras or systems, which can be used to perceive the external environment and perform the 3D reconstruction.

The principle of binocular ranging is shown in Fig. 5, in which part (a) is a linear model and P is a target point in a 3D space. Due to errors in camera manufacturing and installation, the two optical axes are generally not parallel, and their optical parameters are also different in actual use. The ranging principle in the actual process is shown in part (b).

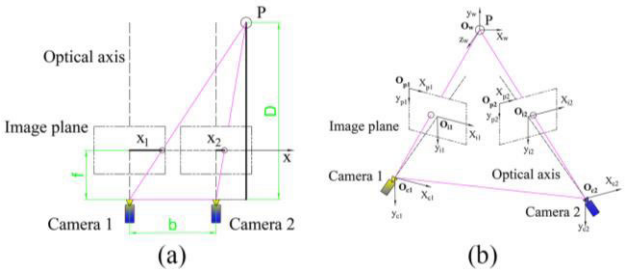

FIGURE 5. The principle of binocular vision ranging. (a) Linear model. (b) Nonlinear model.

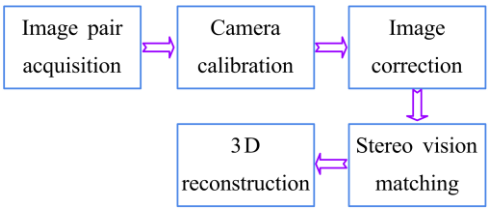

FIGURE 6. Binocular vision implementation process.

By obtaining the mapping relationship between the camera coordinate system O_c, image plane coordinate system O_i, image pixel coordinate system O_p and world coordinate system O_w, a nonlinear model can be transformed into a linear model.

The key to performing 3D reconstruction is to obtain depth information. The depth expression obtained from an analysis of Fig. 5 (a) is as follows [24]:

$$D = \frac{fb}{d}, \tag{1}$$

$$d = x_1 - x_2, \tag{2}$$

where D is the depth from the camera to the target point, f is the focal length of the camera, b is the baseline distance between the two cameras, d is the binocular parallax, and x_1 and x_2 are the abscissa values of the target point P corresponding to camera 1 and camera 2 on the image plane, respectively.

The camera lens causes distortion, including centrifugal distortion, radial distortion and thin prism distortion [25]. Therefore, to improve the accuracy of binocular vision, distortion correction is required. The general implementation of binocular vision is shown in Fig. 6. According to Eq. (1), the prerequisite of calculating depth D is obtaining f, b, and d in advance. The values of f and b can be obtained by camera calibration, and d can be obtained by stereo vision matching. In summary, camera calibration and stereo vision matching are the key steps to achieve binocular vision, and the calibration accuracy and matching effect have a decisive influence on 3D reconstruction [26], [27].

B. OVERVIEW OF CAMERA CALIBRATION

The purpose of camera calibration is to obtain the intrinsic and extrinsic parameters of the camera by using the features of a specific object or scene and to thereby create a mapping from the stereo space to a plane image. The intrinsic parameters include the focal length, principal point position, scale factor, lens distortion, etc. The extrinsic parameters include the translation matrix and rotation matrix [28]. Generally, calibration methods are divided into the three types below.

1) TRADITIONAL CALIBRATION METHODS

A reference object with a known shape and size is used as the target of the camera; after image processing and mathematical transformation, the intrinsic and external parameters of the camera can be obtained. For instance, Zhang [29] proposed a flexible calibration method. The camera needed only to shoot a printed checkerboard from multiple angles, and the parameters could then be obtained by linear calculations. The results showed that the mean value and sample deviation of the calibration parameters were 0.334 and 0.04, respectively, so the algorithm was reliable. However, this method considered only the radial distortion of the lens; if a lens has a large distortion, the internal parameters need to be initialized manually. Abdel-Aziz and Karara [30] proposed a calibration method using a direct linear transformation. Based on the linear constraint equation, the camera coordinate system and the spatial coordinate system were transformed linearly, and the camera parameters could be obtained by solving the linear equation. The operation time of this method was only 22.64% of that of the traditional collinearity method, so it was faster. However, this study neglected the nonlinear distortions of the camera, so the calibration precision was not ideal. Additionally, there are Weng's iteration method [31], Tsai's two-step method [32] and so on. It should be emphasized that the reference objects required by this calibration method are generally easy to identify and extract the features, and the image distortion amplitude is small.

2) CALIBRATION METHODS BASED ON ACTIVE VISION

By precisely controlling the motion of the camera workbench, including pure translations, pure rotations and so on, the camera can capture images from many different angles, and then the camera parameters can be obtained by using the constraint relationships between the camera motion parameters and the image captured. For instance, a calibration method based on plane motion was proposed by Zhu *et al.* [33]. This method considered the radial distortion, tangential distortion, and thin lens distortion of the camera. Experiments showed that after distortion correction and global optimization, the standard deviations of the image point error in the x and z directions were 0.002761 mm and 0.004012 mm, respectively, and the calibration accuracy reached 0.005 mm. This method provides an effective solution for camera calibration after distortion. Furthermore, camera calibration methods based on pure rotational motion and projective reconstruction also have a certain range of applications [34], [35].

3) SELF-CALIBRATION METHODS

Camera parameters can be obtained under the condition that the scene is unknown from only the constraint relationships obtained in an image sequence. For instance, Deng *et al.* [36] proposed a calibration method based on differential

TABLE 1. Ave and Std of the results of three calibration algorithms.

Algorithm	80 groups		800 groups	
	Ave	Std	Ave	Std
DEPSO	0.192335	0.165329	0.177186	0.162637
DE	0.280495	0.262143	0.265371	0.253308
PSO	0.345165	0.323514	0.331645	0.308450

TABLE 2. Advantages and disadvantages of camera calibration methods.

Method	Advantages	Disadvantages
Traditional calibration	High accuracy.	Complicated process.
Active vision calibration	Linear solution and strong robustness.	High-precision calibration devices and high-cost calibration tables are required.
Self-calibration	Has good flexibility and is not limited by reference objects.	Poor algorithm stability and robustness.

evolution and particle swarm optimization (DEPSO). As shown in Table 1 [36], the results of 10, 80 and 800 groups showed that, compared with other algorithms, the proposed algorithm had the smallest average value (Ave) and standard deviation (Std), so the proposed calibration method performed best.

Moreover, a calibration method based on image quality assessment was proposed by Fayyaz and Joo [37]. Its main characteristic was that the quality of the image was evaluated in advance before calibration, and the number of frames with high quality was obtained. The experimental results showed that the self-calibration accuracy reached 93.2%, and the image had almost no distortion, while the traditional self-calibration method only had 7.7% of the image to obtain a good perspective effect. It is worth noting that camera self-calibration is a nonlinear method, which requires a nonlinear optimization solution, so it relies heavily on the initial value.

To summarize, the advantages and disadvantages of the three calibration methods are shown in Table 2.

C. OVERVIEW OF STEREO VISION MATCHING

Stereo vision matching is based on a matching algorithm, the purpose of which is to search for a point in the captured image corresponding to a certain point on the target in the 3D space. Scharstein and Szeliski [38] first proposed the four classical steps of the matching algorithm, as shown in Fig. 7. It should be noted that not all matching algorithms need to go through these four steps, and some do not follow the diagram sequence. The purpose of the matching cost calculation is to measure the correlation between the pixel to be matched and a candidate pixel; generally, the smaller the cost, the greater the correlation. The purpose of cost aggregation is to establish a connection between adjacent pixels so that the cost can accurately reflect the correlation between pixels. The aim of the parallax calculation is to determine the optimal parallax value of each pixel. The aim of parallax optimization is to improve the quality of the parallax map [39]. It should be emphasized that cost aggregation is the most significant step in this process, and it directly affects the accuracy of

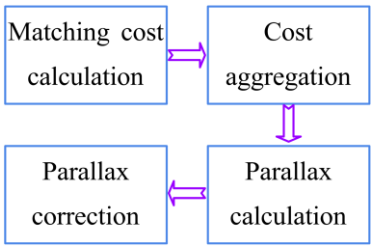

FIGURE 7. Stereo vision matching implementation process.

matching. The selection of a matching algorithm is based on the following steps: First the matching model is constructed, second, the matching primitive is selected, and finally, the most suitable algorithm is selected in accordance with the characteristics of the matching primitive. Generally, the commonly used matching algorithms can be roughly divided into the three types below.

A global matching algorithm comprehensively considers the information of the entire region to be matched and then calculates the parallax by establishing and solving a global energy function. At present, the commonly used global matching algorithms include algorithms based on image segmentation, algorithms based on confidence, and algorithms based on partial differential equations. Many researchers have conducted extensive and in-depth research on these algorithms. For instance, Ma *et al.* [40] proposed an algorithm based on image segmentation in which the image was segmented by simple linear iterative clusters. An image test with the Middlebury dataset showed that the mismatch rate of the whole region was 5.275%, the mismatch rate of shaded regions was 7.370%, and the mismatch rate of unshaded regions was 2.315%; the comprehensive average mismatch rate was 4.990%. Thus, the matching precision of the algorithm was high, and the matching effect in occluded areas was better. Kim *et al.* [41] proposed a deep network structure for estimating confidence that can significantly improve the application of a matching algorithm. Experimental results showed that the average mismatch score with one/three on the KITTI 2015 dataset was 9.67/7.09, so this method had strong robustness. Kim and Hilton [42] proposed a parallax estimation method based on partial differential equations. With this method, a relatively smooth parallax map could be obtained by using a new energy function. Experiments showed that the average depth error of 3D scene reconstruction was 0.2 cm and the standard deviation was 15.2 cm, so this method could effectively perform smooth reconstruction of the depth map.

A local matching algorithm relies on the constraint information of the local window to determine matches by comparing local features. First, the support area of the pixel needs to be calculated, and the size, shape, and weight of the area should meet certain requirements. Second, the parallax value in the support area needs to be a weighted average. Finally, a parallax map can be obtained with high accuracy. According to the characteristics of the matched primitives, local matching algorithms are mainly divided into region-based algorithms, feature-based algorithms, and phase-based

TABLE 3. Results of region matching.

Algorithm	Proportion of regions matched (left, right)	
	Office	House
Proposed	(67.21,69.49)	(65.00,74.29)
GI	(62.30,64.41)	(52.50,60.00)

TABLE 4. Relative pose errors between the two algorithms.

Relative pose errors	Proposed	Direct visual odometry
Translation error (%)	0.8061	0.8168
Rotation error (°/m)	0.0051	0.0053

algorithms. Many scholars have carried out in-depth research on these algorithms. For instance, a region-based matching algorithm was proposed by Ansari *et al.* [43]. As shown in Table 3 [43], a matching experiment with two regions was conducted to compare this algorithm with the traditional matching method based on a grayscale image (GI); the proposed algorithm improved the proportion of regions matched by 13% and 5% in matching the house and office pairs, respectively. The region-based matching algorithm usually yields a dense parallax map; it is easy to implement, but it is greatly affected by radioactive distortion.

Krombach *et al.* [44] integrated a feature-based matching algorithm into the visual odometer method and proposed a feature-based visual odometer method to achieve real-time tracking of a target. As shown in Table 4 [44], compared with the direct visual odometer method, the translation and rotation errors were small, so this method had a strong ability to perform map reconstruction.

Moreover, an absolute-difference census algorithm based on phase was proposed by Zhang *et al.* [45], which can balance the goals of efficiency and accuracy. Phase-based matching can reduce the influence of radiation distortion and geometric distortion to some extent, but it has some defects, such as phase singularity and phase winding.

A semiglobal matching algorithm is improved based on the dynamic programming method and it is first proposed by Hirschmüller [46]. In this algorithm, mutual information is used to calculate the initial matching cost, and the problem of mismatches caused by illumination variation is solved. The global energy function used in this algorithm can be used to penalize different depth changes and achieves the effect of smoothing constraints. To ensure uniqueness constraints, the algorithm performs consistency checks on the left and right parallax maps. However, the algorithm cannot make full use of the pixels around the matching point to carry out effective matching for cost aggregation. Currently, many improved semiglobal matching algorithms have been developed by researchers. For instance, Chai and Yang [47] added a minimum spanning tree to the cost aggregation step, which satisfactorily addressed the pixel matching problem around the points to be matched. A matching experiment on the Middlebury dataset showed that the matching error rate of the traditional semiglobal matching algorithm was 8.56%, while that of the proposed semiglobal matching algorithm was 8.44%, so the matching accuracy was improved.

TABLE 5. Advantages and disadvantages of the matching algorithms.

Algorithm	Advantages	Disadvantages
Global matching	Discontinuous and weakly textured areas can be better handled.	Less efficient.
Local matching	Low time complexity and faster operation.	Multi-intermittent and weakly textured areas are not processed well.
Semiglobal matching	Good balance of accuracy and efficiency.	Does not work well with occlusions and weakly textured areas.

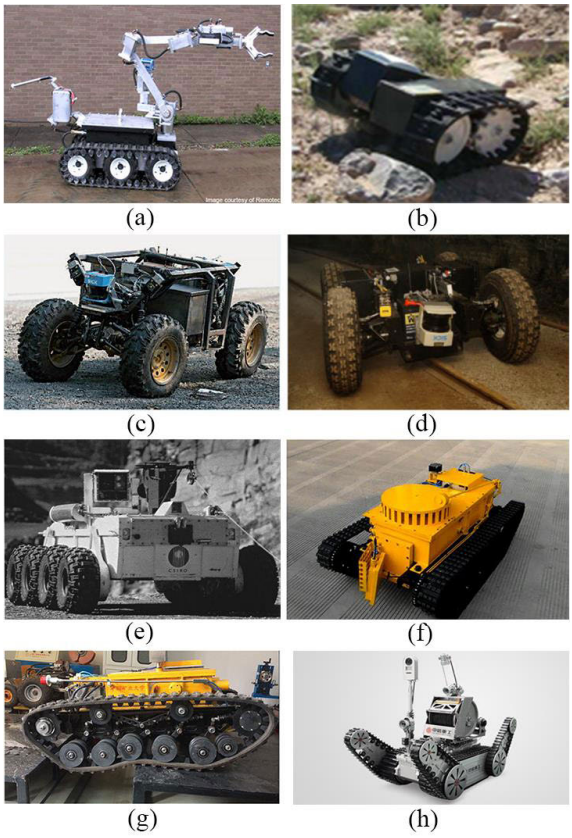

FIGURE 8. CMRRs. (a) The ANDROS Wolverine robot. (b) The Ratler robot. (c) The Groundhog robot. (d) The Cave Crawler robot. (e) The Numbat robot. (f) The CUMT-V (A) robot. (g) The CUMT-V (B) robot. (h) The KRZ-I robot [49], [50], [51], [52], [53], [56], [58].

Ni *et al.* [48] used a semiglobal matching algorithm based on a second-order smoothing constraint to reduce the influence of weak texture regions on matching. Weak texture image tests on the KITTI dataset and Middlebury dataset showed that the average mismatch rate of the algorithm was 4.39% and 23.24%, respectively, indicating that the algorithm had a good application effect in weak texture regions.

The advantages and disadvantages of the three types of matching algorithms are shown in Table 5.

III. RESEARCH STATUS OF CMRRs BASED ON BINOCULAR VISION

Extensive research on different types of CMRRs has been performed by many scholars from different countries.

The ANDROS Wolverine robot was developed by the Remotec Company in conjunction with the Mine Safety and Health Administration (MSHA). As shown in Fig. 8 (a) [49], this robot used a motor-driven crawler to walk and was equipped with a sensing system, a navigation system, a voice system, a communication system, etc.; the site information was transmitted to the rescue center through optical cables. However, it weighed more than 544 kilograms, so it was not very flexible [10]. The Ratler robot was designed by the Sandia National Laboratory and the MSHA in the United States. As shown in Fig. 8 (b) [50], this robot was equipped with an infrared camera and gas inspection sensors, which can be used for coal mine environment detection. Moreover, it was remote-controlled, with a distance of approximately 76 m. The Groundhog robot was designed at Carnegie Mellon University for search tasks in underground mines. As shown in Fig. 8 (c) [51], it used an electrohydraulic drive to control the movement and steering of four wheels, and it carried a variety of sensing equipment such as a laser scanner and night vision camera, which can reconstruct 3D maps of abandoned coal mines. The Cave Crawler robot was also developed at Carnegie Mellon University. As shown in Fig. 8 (d) [52], this robot was equipped with four wheels to achieve independent steering. Compared with the Groundhog robot performance, the flexibility and stability were better. The Numbat robot was developed by the Australian Commonwealth Scientific and Industrial Research Organization. As shown in Fig. 8 (e) [53], this robot was loaded with differential eight-wheel walking mechanisms on both sides of the fuselage, its maximum speed can reach 2 kilometers per hour, and its power was provided by a chrome-nickel battery with an endurance time of 8 hours. Furthermore, it was controlled remotely by ground personnel via fiberoptic cables [54]. Although this robot is not yet operational, it has long been an option in coal mine rescue. Within the framework of a project at the European Union to develop a mine emergency rescue and transportation system, a team at Universidad Carlos III de Madrid in Spain designed the TeleRescuer robot. Its vision system was equipped with ordinary cameras and a thermal camera. The ordinary cameras were used to obtain environmental information and perceive the depth of the scene; the thermal camera was used to detect the temperature information of the equipment and check for equipment failure. Moreover, this robot could carry multiple gas sensors to detect CO_2, CH_4, CO, etc. [55]. The CMRR series CUMT-I, CUMT-II, CUMT-III, and CUMT-IV were developed by China University of Mining and Technology. This series of robots carried a variety of sensors to collect environmental information and could carry emergency supplies to the scene of an accident; they had good application potential, but there were still some problems. Subsequently, the upgraded CUMT-V (A) and CUMT-V (B) crawler robots were developed, as shown in Fig. 8(f) and (g) [56]. In these robots, an explosion-proof multidrive crawler device and slide-block low-speed clutch were used for the motion system, an environment perception module with good interaction was used for the perception system, an optical fiber release device and relay release device were used for the communication system, and a distributed Ethernet structure was used for the control system [56]. A multicrawler snake-shaped CMRR was developed by the Harbin Institute of Technology, and the principle of lightness was used in the design of the robot. Compared with a typical snake robot, it was lighter in weight and smaller in size, and it could achieve 45° pitch and yaw motion [57]. Notably, CITIC Heavy Industry Kaicheng Intelligent Equipment Co., Ltd. is the only specialized robot manufacturing enterprise in China that has obtained a mining safety certification and an explosion-proof safety certification. The company has developed several safety-certified robots for mine detection and rescue, such as ZRXJ127, KRXJ38, KRZ-I, and KQR48. KRZ-I, shown in Fig. 8 (h) [58], carried a camera and a variety of gas sensors. The camera was equipped with infrared lighting, which enabled it to capture clear images under zero illumination. In addition, the robot had an advanced embedded design, with posture alarm and obstacle avoidance alarm functions, and it could carry an appropriate amount of rescue materials to an accident scene while ensuring its own safety.

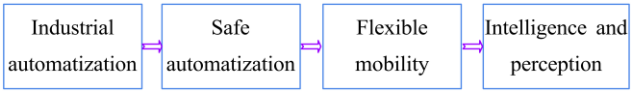

FIGURE 9. Four stages of robot development.

Robotics has experienced four milestone revolutions, shown in Fig. 9. Among them, vision is the basis of the information acquisition and autonomous navigation that support a robot's intelligent perception behavior. In particular, in the unstructured environment of a coal mine, a visual sensor provides most of the information about the environment. Thus, the vision of a CMRR has become the main actor in environmental detection, and it is essential. Common visual sensing techniques include light coding, time of flight, structured light, and binocular vision. Light coding is suitable for close measurement only. Time of flight and structured light are active vision methods. The former has a wide range of applications but a low resolution and high-power consumption. The latter has a higher resolution and lower power consumption but is greatly affected by nonsystem errors and has a greater software complexity. In contrast, binocular vision can restore depth information and enable 3D reconstruction with only one pair of cameras. Although the operative range of binocular vision is limited by the baseline distance, it has high ranging accuracy and low power consumption. In consequence, a CMRR equipped with a binocular camera is the focus of current coal mine safety and intelligence fields, and research mainly focuses on explosion-proof designs, obstacle recognition, stereo vision matching, intelligent algorithms, ultrasonic-assisted ranging, virtual reality simulation, and other topics.

A. EXPLOSION-PROOF DESIGN

After a coal mine accident, explosive gas in the coal seam will continue to emerge and mix with coal dust. To avoid

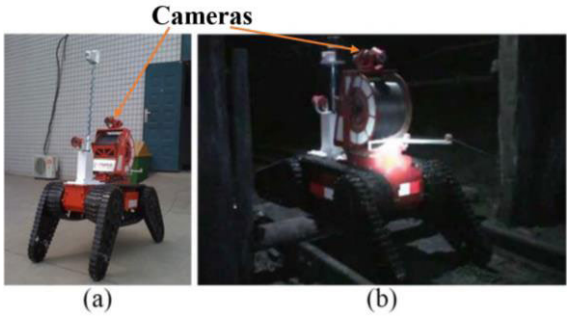

FIGURE 10. CMRRs. (a) MINBOT-II robot. (b) Test in a coal mine [59], [60].

secondary explosion accidents, CMRRs must have explosion-proof performance and need to pass a standard certification. Most countries have strict regulations governing electrical explosion-proof standards, such as the NEC500 electrical specification in the United States and the GB3836 series of standards in China. At present, commonly used explosion-proof technologies include flameproof and intrinsically safe technologies. The former puts the electrical equipment in an explosion-proof box to isolate it from the environment, and the latter prevents explosions by limiting the maximum amount of energy released by electrical equipment. A binocular camera is part of the electrical equipment and is exposed to the coal mine environment, so it must be designed based on the intrinsically safe strategy. For instance, Wang *et al.* [59] developed an explosion-proof CMRR called MINBOT-II, as shown in Fig. 10 (a) [59]. The size of the robot was $1500 \times 589 \times 747$ mm (arm deployed), the mass was 130 kg, and the distance controlled by the optical fiber was 1000 m. The robot was composed of a mechanical system, a communication system, a power supply system, a human-robot interaction system, and a sensing system; the sensing system was used to monitor the environment. The electric explosion-proof design is shown in Fig. 11 [59], the cameras were an integral part of the sensing system, and their electrical design was based on the principle of intrinsic safety; they were mainly used to collect information on the coal mine and the attitude of the robot. It was notable that four cameras were able to provide a full detection range. Moreover, the intrinsically safe devices realized physical isolation with the devices in an explosion-proof box through the isolation circuit. In addition, the signal transmission and energy supply between the two devices were realized by the intrinsically safe CPU and intrinsically safe power in the isolation circuit. Similarly, Lu *et al.* [60] also designed cameras in the sensing system of a CMRR based on the principle of intrinsic safety; a test in a coal mine is shown in Fig. 10 (b) [60]. The cameras used to obtain real-time images of the coal mine were lightweight and small in size, and were powered by a 12 volt intrinsically safe power supply.

B. OBSTACLE RECOGNITION AND FEATURE EXTRACTION

A CMRR relies on binocular vision to recognize obstacles. First, it needs to extract the features of the obstacles,

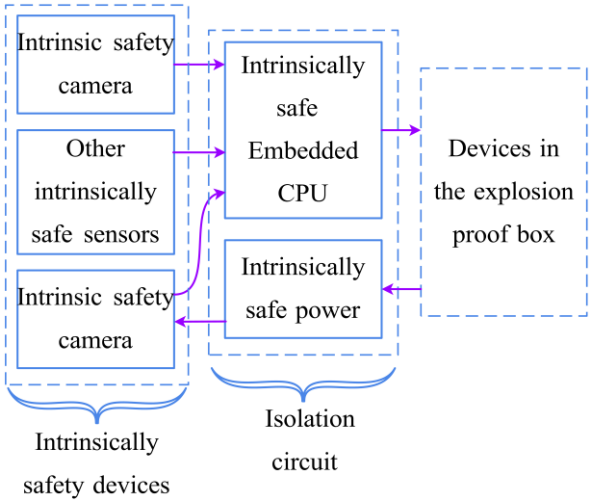

FIGURE 11. Principles of electrical explosion protection design [59].

including the features of the edge and the shape. Common feature extraction algorithms include Moravec extraction, Harris extraction, the features from accelerated segment test (FAST) [61], scale-invariant feature transform (SIFT) [62], and speeded-up robust features (SURF) [63]. Moravec extraction is too direction-dependent, and its edge feature extraction effect is poor. Harris extraction achieves corner point extraction at the pixel level and is rotation invariant, but it is not scale-invariant. The computational efficiency and repeatability of FAST are high, but it is greatly affected by the threshold value and by noise, and it is also not scale-invariant. SIFT's scalability and discrimination are strong, but its computational complexity is high and its runtime is long. SURF is an optimization of SIFT, and its computing speed is better. After obtaining the information of feature points, the robot can identify obstacles by analyzing the topological relationships among feature points. For instance, Niu *et al.* [64] developed a coal mine detection robot for search and rescue, as shown in Fig. 12 [64]. As a part of the information perception and interaction module, the binocular vision system was mainly composed of a Sony camera and a midinfrared (MIR) camera, and it was used for obstacle recognition. The feature points of the image were extracted by the Harris method, where the corner was judged by the response function $F(x, y)$. $F(x, y)$ can be expressed as [64]:

$$F(x, y) = Det(\boldsymbol{M}) - \mathrm{k} \times Trace^2(\boldsymbol{M}) \tag{3}$$

$$\boldsymbol{M} = \begin{bmatrix} I_x^2 & I_{xy} \\ I_{xy} & I_y^2 \end{bmatrix}, \tag{4}$$

where M is the partial derivative matrix, I is the grayscale of the image, and k is an empirical constant.

According to Eqs. (3) and (4), since the corner response function adopted involves only the first derivative and mixed second partial derivative in the x and y directions, a Gaussian smoothing filter must be used for noise reduction when the image has much noise.

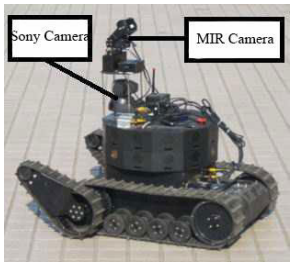

FIGURE 12. Coal mine detection robot [64].

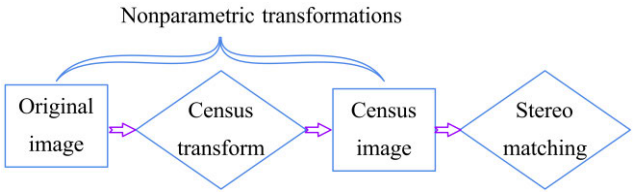

FIGURE 13. Process of a stereo matching algorithm based on census transformation.

C. IMPROVEMENT OF STEREO VISION MATCHING

The core problem of binocular vision is how to improve the positioning accuracy and obtain the target position correctly. The key is to find the homonymic point on the consistent spatial position of the two images obtained by the cameras, that is, to perform stereo vision matching. The traditional matching algorithm is not effective in solving problems such as weak texture regions, depth discontinuity regions and occluded regions. Thus, a more intelligent and adaptable matching algorithm is needed in special environments. Generally, the robustness, accuracy, and real-time performance of the stereo vision matching algorithm affect the operational effect of the vision system. Many scholars have done much work in this area.

He *et al.* [65] investigated an improved census algorithm and applied it to the stereo matching process of a vision system for a CMRR. The census algorithm was a nonparametric transformation, as shown in Fig. 13. Before stereo matching, the original image needed to be converted into a census image; i.e., the grayscale was defined as a string element. In the experimental stage, the improved census algorithm and the matching algorithm based on the sum of absolute differences (SAD) were tested on the Cyclone IV EP4CE115F23C7/FPGA platform, and the resource consumption, real-time performance and accuracy were compared. Partial results are shown in Table 6 [65]; clearly, the improved census algorithm was superior to the SAD algorithm in terms of resource consumption and matching accuracy. Therefore, the proposed improved census algorithm had better performance and was able to adapt to complex coal mine scenes in order to carry out more accurate 3D reconstruction.

Furthermore, He and Ma [66] proposed an improved SIFT algorithm considering the real-time performance and accuracy of the matching. First, the SIFT operator was simplified and the initial matching was performed. Then, the random-sample consensus algorithm and line constraints were used for quadratic matching to eliminate mismatched pairs. The experimental data are shown in Table 7 [66]; compared with the traditional SIFT algorithm, the improved SIFT algorithm has a greater advantage in terms of the number of misplaced points, the correct match rate and the match time. Fig. 14 shows the effect of using the improved SIFT algorithm for binocular matching [66]. The algorithm has the disadvantage of matching fewer feature points, and it may ignore some of the texture information of the target.

TABLE 6. Comparison results of two matching algorithms.

Algorithm	Total registers (resource consumption)	Match window (real-time performance)	Mismatching rate for 4 pixels (accuracy)
Proposed	10352	5×5	7.18%
SAD	40584	5×5	12.16%

TABLE 7. Comparison results of two matching algorithms.

Algorithm	Traditional SIFT	Proposed
Total feature points	269	50
Matched points	257	25
Misplaced points	15	1
Correct match rate	94.1%	96%
Match time (s)	10.5	6.4

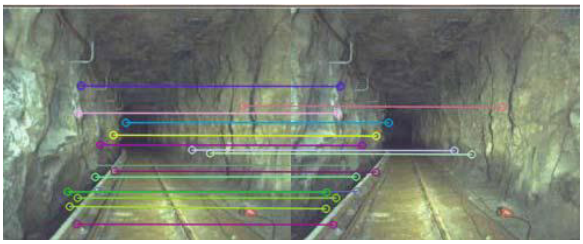

FIGURE 14. Stereo matching effect for a roadway image [66].

D. VISUAL PERCEPTION BASED ON INTELLIGENT ALGORITHMS

The vision sensor of a CMRR has a certain lag and uncertainty in obstacle recognition, while intelligent algorithms such as neural networks and fuzzy logic have a strong ability of deep learning, image understanding, and generalization in obstacle recognition and tracking, which can make up for the shortcomings of traditional binocular vision. At present, binocular vision technology based on an intelligent algorithm is a research hotspot in the field of robot vision. For instance, Shang *et al.* [67] designed a binocular vision system for a CMRR based on a fuzzy neural network. As shown in Fig. 15, first, the CMRR used binocular vision to collect images of the roadway and obstacles and extracted the associated features. Then, the feature information was quantized as inputs, and the weights of a neural network were iterated until convergence. Finally, obstacle avoidance decisions were made based on the identification results. Note that the fuzzy neural network was

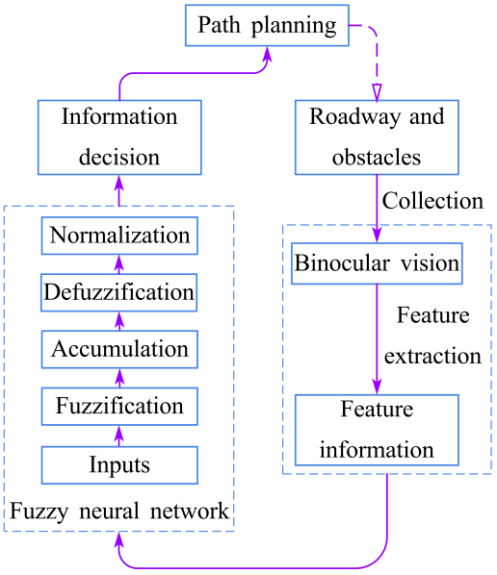

FIGURE 15. Process of binocular vision recognition based on a fuzzy neural network.

a five-layer forward structure based on a back propagation algorithm and was based on a heuristic process. By building a test platform for binocular vision, 200 types of obstacle images were simulated in MATLAB; the results showed that the accuracy of navigation decision making could reach 99.28% and the square root error of the algorithm was 7.20×10^{-4}. This revealed that the robot could recognize obstacles in the roadway relatively accurately.

E. INTEGRATED POSITIONING

In the unstructured environment with strong interference and high dynamic, it is difficult for a single navigation source to meet the application requirements, and integrated positioning is an effective navigation strategy. Specifically, in the coal mine roadway, visual positioning is greatly affected by lighting changes, while ultrasonic positioning is not affected and has no cumulative error. Thus, a binocular camera assists ultrasonic ranging when acquiring obstacle features, and this can effectively reduce the positioning error divergence caused by a single vision sensor. For instance, He and Ma [25] proposed an obstacle avoidance strategy combining ultrasonic ranging and binocular vision ranging for path planning in a CMRR. The robot consisted of a mechanical system, a control system, and a visual system. As shown in Fig. 16 [25], at the heart of the vision system was Point Gray's BumbleBee2 binocular camera. The main feature of this camera is that distortion and position deviation can be automatically corrected; a high-speed 1394 interface was grafted to it, and the internal data conversion rate reached one million 3D points per second. An HC-SR04 ultrasonic module was adopted with a working voltage of 5 volts and a frequency of 40,000 Hertz. In the integrated positioning strategy, the long-distance information was measured by an ultrasonic module, and the height and width characteristics of the obstacle was acquired by a binocular vision image processing module. After the two

FIGURE 16. Underground inspection intelligent robot [25].

FIGURE 17. Simulation result in virtual reality. (a) Roadway with moisture. (b) Roadway with coal dust distribution [68].

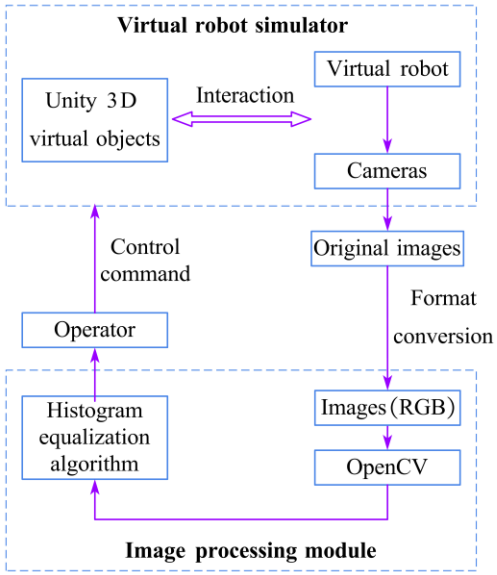

FIGURE 18. Simulation process of the visual enhancement algorithm based on Unity 3D platform.

parts of information were analyzed, an obstacle avoidance strategy was formed, which in turn controlled the robot to execute obstacle avoidance actions. Ultrasonic ranging has a lower cost and more stable functionality, but its accuracy is generally at the centimeter level; moreover, it is easily disturbed by coal dust migration, so the probability of error reporting is relatively high.

F. VISUAL ALGORITHM SIMULATION BASED ON VIRTUAL REALITY

An algorithm used in the vision of a CMRR must be tested and debugged many times before it can meet the application requirements. However, field testing for a coal mine requires

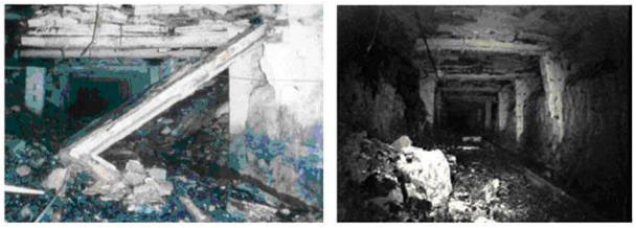

FIGURE 19. The unstructured environment of a coal mine after an accident [52], [59].

a great deal of technical support and funding, and the testing cycle is very long. The development of virtual reality technology provides support for solving this problem. Virtual reality has the characteristics of immersion and interactivity, which can be used to test robot vision algorithms. Current popular virtual reality platforms include Virtools, Unreal Engine, and Unity3D, etc. Unity3D is a professional game engine that uses C# as the scripting language, and it can render a virtual reality world that includes roadways, obstacles, tools, railway tracks, coal dust, water vapor, and trapped people. Moreover, the Unity3D platform can also simulate the motion control of a robot from the perspective of virtual cameras. For instance, Wang *et al.* [68] established an algorithm test platform for the vision system of a CMRR by using virtual reality technology. The Unity3D platform was used for environment modeling. The virtual environment consisted of virtual objects and a virtual robot; Fig. 17 (a) and (b) show a simulated roadway scene with moisture and coal dust distribution, respectively [68]. Additionally, OpenCV was encapsulated into the Unity3D platform so that the vision algorithm in OpenCV could be called directly. As shown in Fig. 18, the virtual cameras were used to collect images of the virtual world. To eliminate the influence of coal dust and other interference factors, the histogram equalization algorithm in OpenCV was used to enhance the image after format conversion. As determined by the processing results, the command was sent to the simulator by the operator to control a virtual robot movement. The results showed that the gray distribution of the image acquired by the virtual cameras was more uniform after processing with the histogram equalization algorithm. Therefore, the visual function of a CMRR can be tested by virtual reality technology.

IV. TECHNICAL CHALLENGES AND DEVELOPMENT TRENDS OF BINOCULAR VISION FOR CMRRs

A. TECHNICAL CHALLENGES

A coal mine environment after an accident is shown in Fig. 19 [52], [59]. Its nonstructural features bring challenges to the binocular vision perception of the CMRR, which are primarily manifested as described below.

1) THE BALANCE BETWEEN THE SPATIAL CONSTRAINTS OF THE ROADWAY AND THE BASELINE DISTANCE OF THE BINOCULAR CAMERA

A wider roadway is less favorable for the support, so the feasible area of a coal mine is very narrow. To enable a CMRR to pass along the roadway smoothly, its structure should be compact and its size should be reduced as much as possible. For instance, the ANDROS Wolverine robot is large, and it is inconvenient to move along a roadway, so its rescue ability is not good [10]. Consequently, in the installation of the binocular camera, space limitations must be considered. From Eq. (1), we can see that the baseline distance b directly affects the target depth D. Generally, the accuracy of the depth measurement increases with the increase in the baseline distance, but the accuracy of the binocular parallax decreases with the increase in the baseline distance, so determining the appropriate baseline distance is critical for the quality of the depth map. Therefore, the difficulty lies in balancing the relationship between spatial installation requirements and the baseline distance according to the coal mine environment.

2) THE EFFECTS OF MOTION BLUR ON THE ROBUSTNESS OF THE BINOCULAR VISION SYSTEM

After a coal mine accident, supporting equipment, transportation equipment, water supply and drainage equipment, etc. are typically severely damaged. In addition, scattered mechanical parts, stones, coal, etc. are distributed on the ground, so the terrain of the coal mine becomes extremely complicated [56]. In the process of moving forward, the binocular camera of a CMRR will vibrate and shake, resulting in different degrees of motion blur in the images collected by the camera. Motion blur reduces the feature information of the image, increases the difficulty of feature extraction, reduces the matching rate, and causes algorithm failure in serious cases, which greatly restricts the robustness of the binocular vision system. Therefore, it is a technical challenge to improve the image stabilization and vibration reduction to reduce the impact of motion blur on the robustness of the binocular vision system.

3) THE CALIBRATION AND COORDINATION OF THE MULTI-DEGREE-OF-FREEDOM BINOCULAR CAMERA

Generally, after a camera is calibrated, its position is fixed, which improves the imaging stability but limits the visual range of the CMRR. Given the urgency of a rescue mission, a binocular camera needs to look at different areas. By increasing the number of rotational and pitch degrees of freedom of the camera, it can explore multiple areas. However, when the camera position changes, its external parameters also change, and multiple calibrations will cause a lag in the acquired information. Consequently, the difficulty lies in calibrating a multi-degree-of-freedom binocular camera online in real time. Additionally, a multi-degree-of-freedom binocular camera can achieve independent motion, so there is a technical bottleneck in effectively coordinating the scene information acquired by the binocular camera, completing the automatic splicing and fusion of 3D point clouds in multiple fields of view, and ultimately meeting the requirements of convergence control.

4) THE UTILIZATION OF THE AUXILIARY MATCHING CRITERION AND THE REDUCTION OF MATCHING DIFFICULTY

Stereo vision matching is the most important and difficult step in the process of binocular vision perception of the environment for a CMRR [65]. Low illumination and moving obstacles in the roadway seriously affect the effectiveness of stereo vision matching, leading to the nonlinear distortion and gray distortion of the image. Therefore, in a coal mine environment, using the environmental features as auxiliary matching criteria to provide additional constraint information for stereo vision matching and to reduce the matching difficulty is a technical problem that restricts the efficient visual exploration of CMRRs.

B. DEVELOPMENT TRENDS

Future research is moving towards high precision, large depth, a wide field of vision, high reliability, and self-learning. This can mainly be seen in the topics described below.

1) MULTISENSOR INFORMATION FUSION

In the face of the complexity and uncertainty of the unstructured environment of a coal mine, a single binocular camera has the problems of a low target signal-to-noise ratio and a low ranging accuracy, which make it difficult to fully meet the working needs of CMRRs. One way to solve these problems is to carry multiple ranging sensors at the same time and achieve information complementarity through sensor information fusion technology. In engineering applications, sensor information fusion refers to the integration and processing of sensor information from different sources and modes according to certain algorithms and strategies in order to describe the sensed objects accurately and reasonably [69]. Information fusion can enhance the availability of data and reduce fuzziness; additionally, it can increase the coverage of time and space [70].Therefore, based on a binocular camera, CMRRs can be equipped with auxiliary sensors such as lidar sensors, infrared sensors or inertial measurement units [71]–[75]. Then, the multidimensional information can be integrated, which can effectively improve the robustness, adaptability, and fault tolerance of the robot's perception system, more accurately obtain the key information of the accident site, and provide a good foundation for the path planning and autonomous decision-making of the robot. The conceptual framework of multisensor information fusion is shown in Fig. 20; after preprocessing and parameterization of the image information acquired by multiple sensors, information fusion can be performed. Typical information fusion methods include neural networks, Bayesian estimation, Kalman filtering, etc.

In addition, downhole light is relatively dim, and there is a large amount of coal dust in the air during migration movements, so a single image feature acquired by a binocular camera cannot completely describe the entire scene in the field of view. Image fusion technology can combine the features of multiple images describing the same scene into a composite image with more information [76]. The use of image fusion technology can significantly reduce the output redundancy and greatly improve the effective information.

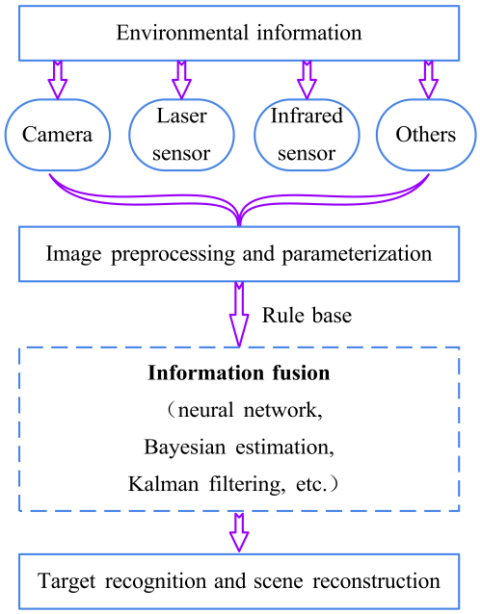

FIGURE 20. Multisensor information fusion framework.

2) EXPAND THE FIELD OF VIEW WITH A FISHEYE LENS

A CMRR needs to construct an environmental reference in a large area in the shortest possible time through the perception of visual sensors. An ordinary binocular camera uses a conventional lens, and its field-of-view angle is small. A fisheye lens is a type of ultrawide-angle lens, which has a wide perspective and can obtain the information of a whole airspace and time domain, so it is widely used in many fields such as autonomous navigation, augmented reality, engineering measurement, and real-time monitoring [77], [78]. For instance, Cao *et al.* [79] proposed a wide-range scene reconstruction method based on fisheye lens imaging. It was used to reconstruct a lunar terrain, and simulation experiments showed that the fisheye camera could recover 3D scene information over a larger range than an ordinary camera. Similarly, if a fisheye lens is mounted on a binocular vision system, a downhole image with a large field of view can be acquired, which can provide more visual information for the robot's autonomous obstacle avoidance and navigation process. The imaging model of a fisheye lens is shown in Fig. 21. Its imaging surface is spherical, and the mapping between the object point and the projected point is nonlinear, so it has the inherent defect of a large distortion [80], [81]. Therefore, it is necessary to use a more intelligent algorithm to correct the distortion and compensate for the inclination of the image when using a fisheye lens than when using a regular lens. Many scholars have done in-depth research in this area. For instance, Lee *et al.* [82] proposed an image-based distortion parameter estimation algorithm. This method is relatively

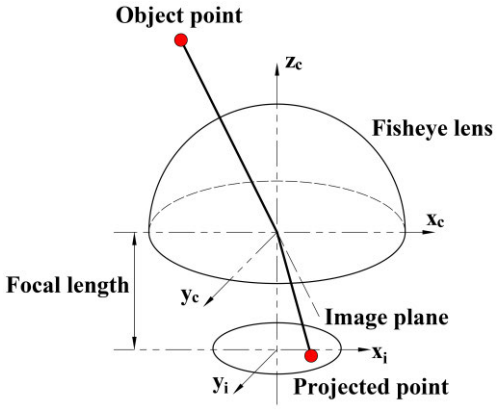

FIGURE 21. Fisheye lens imaging model.

intelligent and has a wide application range. Zhang *et al.* [83] proposed a correction method based on an ellipsoid function model. This method can obtain satisfactory correction results by adjusting the variables in the function model.

Additionally, it is worth mentioning that a fisheye lens and an ordinary lens can be used in combination to achieve complementary advantages. A binocular fisheye camera can obtain a wide range of scene information, while an ordinary binocular camera can obtain clear image features. If the depth information obtained from the two is integrated into a unified scene, large-scale and high-precision 3D reconstruction can be achieved.

3) DEPTH ESTIMATION BASED ON BINOCULAR DISPARITY AND MOTION DISPARITY

When our line of sight moves laterally, the difference in the speed and direction of an object in our field of vision will cause motion parallax in the retina. As a kind of depth clue, motion parallax has a higher intensity than other monocular clues [84]. Binocular parallax and motion parallax have their own advantages. The research findings of Mansour *et al.* [85] indicate that the motion parallax of robots outperforms binocular parallax for distant features of a target and absolute depth estimation error, while binocular parallax is more accurate for near ranges. Therefore, a perception system can use binocular parallax clues and motion parallax clues at the same time. A robot with both types of clue perception can perform depth estimation in a wide range; additionally, when binocular parallax clues are unavailable, motion parallax clues can provide depth information alone, which can greatly improve the reliability of the perception system. Furthermore, according to the research findings of Bradshaw and Rogers [86], in calculating the depth information of the binocular disparity and motion disparity, they are not simply linear superimpositions but interact nonlinearly. Therefore, future research in this area should focus on two points. One is eliminating the parallax compression and interference caused by the joint action of the two clues and then combining them organically for underground exploration. The other is intelligently adjusting the primary and secondary effect of the two clues according to the scene features.

4) SELF-PERCEPTION AND SELF-LEARNING BASED ON ACTIVE VISION

Biological studies indicate that human eye movements can be divided into autokinetic and forced movements. The former is controlled by human consciousness, and the latter is not controlled [87]. The research findings of Lonini *et al.* [88] revealed that active vision can guide the flow of visual information in the cycle of perception and movement, thereby guiding efficient self-learning activities. Lelais *et al.* [89] proposed a perception model based on active binocular-motion vision. Through the application of active coding techniques, this model can effectively encode the visual signals generated by the motion of objects and can improve the coding efficiency autonomously. Ramakrishnan *et al.* [90] emphasized that robots must adapt to a changing environment and actively collect information in search-and-rescue missions. To summarize, an active binocular vision control system based on bionics principles should be constructed by simulating the biological control network of human eye movement. This could effectively solve the problem of the poor real-time performance of the perception system and ensure that the rescue robot can quickly and accurately direct its sight at a target; additionally, when an obstacle or camera produces unpredictable motion, such a system could perform active tracking and motion compensation, enhancing the interactivity between the robot and the environment. Fuzzy logic can combine human fuzzy consciousness with machine vision for active perception, so active binocular vision based on fuzzy logic control should be the focus of the visual field of CMRRs in the future.

V. CONCLUSION

The use of CMRRs to carry out the task of detection and rescue in mines is very important. Binocular vision can help robots to perform obstacle recognition and 3D scene reconstruction, so the development of its key technologies is closely related to CMRRs. This paper primarily does the following: First, the research status of camera calibration and stereo vision matching in binocular vision is introduced. Second, the latest research progress for CMRRs based on binocular vision is reviewed, including explosion-proof designs, obstacle recognition, stereo vision matching, intelligent algorithms, ultrasonic-assisted ranging, virtual reality simulation. Finally, several technical challenges and future development trends in binocular vision applied to future CMRRs are proposed.

REFERENCES

[1] (2019). *BP Statistical Review of World Energy 2019*. Accessed: Jan. 1, 2020. [Online]. Available: https://www.bp.com/content/dam/bp/business-sites/en/global/co-rporate/pdfs/energy-economics/statistical-review/bp-stats-review-2019-full-report.pdf

[2] Q. Liu, X. Li, and X. Meng, "Effectiveness research on the multi-player evolutionary game of coal-mine safety regulation in China based on system dynamics," *Saf. Sci.*, vol. 111, pp. 224–233, Jan. 2019, doi: 10.1016/j.ssci.2018.07.014.

[3] Q. Liu, X. Meng, M. Hassall, and X. Li, "Accident-causing mechanism in coal mines based on hazards and polarized management," *Saf. Sci.*, vol. 85, pp. 276–281, Jun. 2016, doi: 10.1016/j.ssci.2016.01.012.

[4] S. Han, H. Chen, R. Long, H. Qi, and X. Cui, "Evaluation of the derivative environment in coal mine safety production systems: Case study in China," *J. Cleaner Prod.*, vol. 143, pp. 377–387, Feb. 2017, doi: 10.1016/j.jclepro.2016.12.096.

[5] Y. Shi, J. Chen, J. Hao, J. Bi, M. Qi, and X. Wang, "Statistical analysis of coal mine accidents of China in 2018," in *Proc. IEEE Prognostics Syst. Health Manage. Conf. (PHM)*, Oct. 2019, doi: 10.1109/PHM-Qingdao46334.2019.8942991.

[6] J. Zhao, J. Gao, F. Zhao, and Y. Liu, "A search-and-rescue robot system for remotely sensing the underground coal mine environment," *Sensors*, vol. 17, no. 10, p. 2426, Oct. 2017, doi: 10.3390/s17102426.

[7] J. Zhang, J. Fu, H. Hao, G. Fu, F. Nie, and W. Zhang, "Root causes of coal mine accidents: Characteristics of safety culture deficiencies based on accident statistics," *Process Saf. Environ. Protection*, vol. 136, pp. 78–91, Apr. 2020, doi: 10.1016/j.psep.2020.01.024.

[8] A. H. Reddy, B. Kalyan, and C. S. N. Murthy, "Mine rescue robot system—A review," *Procedia Earth Planet. Sci.*, vol. 11, pp. 457–462, Jan. 2015, doi: 10.1016/j.proeps.2015.06.045.

[9] W. Shang, X. Cao, H. Ma, H. Zang, and P. Wei, "Kinect-based vision system of mine rescue robot for low illuminous environment," *J. Sensors*, vol. 2016, pp. 1–9, Oct. 2016, doi: 10.1155/2016/8252015.

[10] Y. Wang, P. Tian, Y. Zhou, and Q. Chen, "The encountered problems and solutions in the development of coal mine rescue robot," *J. Robot.*, vol. 2018, pp. 1–11, Oct. 2018, doi: 10.1155/2018/8471503.

[11] Y. Wang, Y. W. Li, P. Tian, and Y. Zhou, "Analysis and prospect on development course of colliery rescue robots," *Mining Process. Equip.*, vol. 46, no. 05, pp. 1–10, 2018, doi: 10.16816/j.cnki.ksjx.2018.05.001.

[12] K. Hashimoto, "A review on vision-based control of robot manipulators," *Adv. Robot., Int. J. Robot. Soc. Jpn.*, vol. 17, no. 10, pp. 969–991, 2003.

[13] L. Perez, I. Rodriguez, N. Rodriguez, R. Usamentiaga, and D. F. Garcia, "Robot guidance using machine vision techniques in industrial environments: A comparative review," *Sensors*, vol. 16, no. 3, p. 26, Mar. 2016, doi: 10.3390/s16030335.

[14] D. Marr and T. Poggio, "Cooperative computation of stereo disparity," *Science*, vol. 194, no. 4262, pp. 283–287, Oct. 1976, doi: 10.1126/science.968482.

[15] W. E. Grimson, "A computer implementation of a theory of human stereo vision," *Phil. Trans. Roy. Soc. London. Ser. B, Biol. Sci.*, vol. 292, no. 1058, pp. 217–253, May 1981, doi: 10.1098/rstb.1981.0031.

[16] J. Hofmann, J. Korinth, and A. Koch, "A scalable high-performance hardware architecture for real-time stereo vision by semi-global matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2016, pp. 845–853, doi: 10.1109/CVPRW.2016.110.

[17] S. Yang, Y. Gao, Z. Liu, and G. Zhang, "A calibration method for binocular stereo vision sensor with short-baseline based on 3D flexible control field," *Opt. Lasers Eng.*, vol. 124, Jan. 2020, Art. no. 105817, doi: 10.1016/j.optlaseng.2019.105817.

[18] S. Zhang, B. Li, F. Ren, and R. Dong, "High-precision measurement of binocular telecentric vision system with novel calibration and matching methods," *IEEE Access*, vol. 7, pp. 54682–54692, 2019, doi: 10.1109/access.2019.2913181.

[19] H. S. Yin, "SLAM-based self-calibration of a binocular stereo vision rig in real-time," *Sensors*, vol. 20, no. 3, p. 21, Feb. 2020, doi: 10.3390/s20030621.

[20] S. Song and W. Zhang, "LTF robot: Binocular robot with laser-point tracking and focusing function," in *Proc. Int. Conf. Intell. Robot. Appl.*, vol. 11740, 2019, pp. 39–48, doi: 10.1007/978-3-030-27526-6_4.

[21] M. Samadi, M. F. Othman, and S. H. M. Amin, "Stereo vision based robots: Fast and robust obstacle detection method," in *Proc. 9th Asian Control Conf. (ASCC)*, Jun. 2013, pp. 1–5.

[22] X. Zhang and A. L. Phuan Tay, "A binocular vision system with attentive saccade and spatial variant vergence control," *Cybern. Syst.*, vol. 42, no. 1, pp. 45–63, Jan. 2011, doi: 10.1080/01969722.2010.532643.

[23] L. Hu and C. Shen, "A study of visual servo system based on binocular camera," in *Proc. Int. Conf. Electr. Eng. Autom. Control (ICEEAC)*, vol. 367, Oct. 2015, pp. 1105–1112, doi: 10.1007/978-3-662-48768-6_123.

[24] P. Hu, X. Hao, J. Li, C. Cheng, and A. Wang, "Design and implementation of binocular vision system with an adjustable baseline and high synchronization," in *Proc. IEEE 3rd Int. Conf. Image, Vis. Comput. (ICIVC)*, Jun. 2018, pp. 566–570, doi: 10.1109/ICIVC.2018.8492907.

[25] K. He and X. Ma, "Research on avoidance obstacle strategy of coal underground inspection robot based on binocular vision," in *Proc. 29th Chin. Control Decision Conf. (CCDC)*, May 2017, pp. 6732–6737, doi: 10.1109/CCDC.2017.7978390.

[26] A.-S. Poulin-Girard, S. Thibault, and D. Laurendeau, "Influence of camera calibration conditions on the accuracy of 3D reconstruction," *Opt. Express*, vol. 24, no. 3, pp. 2678–2686, 2016, doi: 10.1364/OE.24.002678.

[27] Z. Liu, Q. Wu, S. N. Wu, and X. Pan, "Flexible and accurate camera calibration using grid spherical images," *Opt. Express*, vol. 25, no. 13, pp. 15269–15285, Aug. 2017, doi: 10.1364/OE.25.015269.

[28] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000, doi: 10.1109/34.888718.

[29] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 666–673, doi: 10.1109/ICCV.1999.791289.

[30] Y. Abdel-Aziz, H. Karara, and M. Hauck, "Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry," *Photogram. Eng. Remote Sens.*, vol. 81, no. 2, pp. 103–107, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0099111215303086

[31] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 10, pp. 965–980, Oct. 1992, doi: 10.1109/34.159901.

[32] R. Y. Tsai, "Efficient and accurate camera calibration technique for 3D machine vision," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 1986, pp. 364–374.

[33] J. Zhu, X. F. Li, and Y. X. Xu, "Camera calibration technique based on active vision," *Acta Optica Sinica*, vol. 30, no. 5, pp. 1297–1303, May 2010, doi: 10.3788/AOS20103005.1297.

[34] C. L. Li, J. J. Lu, and L. Z. Ma, "Improved rotation-based self-calibration with a strategy of rotational angles," *Opt. Eng.*, vol. 48, no. 9, p. 11, Sep. 2009, doi: 10.1117/1.3231505.

[35] X. Wu, D. Pi, and X. Jiang, "Linear camera self-calibration with varying intrinsic parameters based on projective reconstruction," in *Proc. IEEE Int. Conf. Ind. Technol.*, Dec. 2006, p. 484, doi: 10.1109/ICIT.2006.372609.

[36] L. Deng, G. Lu, Y. Shao, M. Fei, and H. Hu, "A novel camera calibration technique based on differential evolution particle swarm optimization algorithm," *Neurocomputing*, vol. 174, pp. 456–465, Jan. 2016, doi: 10.1016/j.neucom.2015.03.119.

[37] R. Fayyaz and R. E. Joo, "Accurate camera self-calibration based on image quality assessment," *J. Inf. Technol. Appl. Manage.*, vol. 25, no. 2, pp. 41–52, 2018, doi: 10.21219/jitam.2018.25.2.041.

[38] D. Scharstein and R. Szeliski, "," *Int. J. Comput. Vis.*, vol. 47, nos. 1–3, pp. 7–42, 2002, doi: 10.1023/a:1014573219977.

[39] L. Ma, J. Li, J. Ma, and H. Zhang, "A modified census transform based on the neighborhood information for stereo matching algorithm," in *Proc. 7th Int. Conf. Image Graph.*, Jul. 2013, pp. 533–538, doi: 10.1109/ICIG.2013.113.

[40] R. Ma, F. Zhu, Q. Wu, R. Lu, and J. Wei, "Dense stereo matching algorithm based on image segmentation," *Acta Optica Sinica*, vol. 39, no. 3, Mar. 2019, Art. no. 0315001, doi: 10.3788/AOS201939.0315001.

[41] S. Kim, D. Min, S. Kim, and K. Sohn, "Unified confidence estimation networks for robust stereo matching," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1299–1313, Mar. 2019, doi: 10.1109/tip.2018.2878325.

[42] H. Kim and A. Hilton, "PDE-based disparity estimation with occlusion and texture handling for accurate depth recovery from a stereo image pair," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 4061–4064, doi: 10.1109/ICIP.2010.5654067.

[43] M. El Ansari, L. Masmoudi, and A. Bensrhair, "A new regions matching for color stereo images," *Pattern Recognit. Lett.*, vol. 28, no. 13, pp. 1679–1687, Oct. 2007, doi: 10.1016/j.patrec.2007.04.011.

[44] N. Krombach, D. Droeschel, S. Houben, and S. Behnke, "Feature-based visual odometry prior for real-time semi-dense stereo SLAM," *Robot. Auto. Syst.*, vol. 109, pp. 38–58, Nov. 2018, doi: 10.1016/j.robot.2018.08.002.

[45] S. M. Zhang, M. X. Wu, Y. X. Wu, G. X. Wu, and F. Liu, "Fixed window aggregation AD-census algorithm for phase-based stereo matching," *Appl. Opt.*, vol. 58, no. 32, pp. 8950–8958, Nov. 2019, doi: 10.1364/ao.58.008950.

[46] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008, doi: 10.1109/tpami.2007.1166.

[47] Y. Chai and F. Yang, "Semi-global stereo matching algorithm based on minimum spanning tree," in *Proc. 2nd IEEE Adv. Inf. Manage., Commun., Electron. Autom. Control Conf. (IMCEC)*, May 2018, pp. 2181–2185, doi: 10.1109/IMCEC.2018.8469306.

[48] J. Ni, Q. Li, Y. Liu, and Y. Zhou, "Second-order semi-global stereo matching algorithm based on slanted plane iterative optimization," *IEEE Access*, vol. 6, pp. 61735–61747, 2018, doi: 10.1109/access.2018.2876420.

[49] Mining Technology. *Rescue Robots: The Search for Investment*. Accessed: Apr. 22, 2020. [Online]. Available: https://www.mining-technology.com/features/f-eature111373

[50] J. Zhao, G. Liu, Y. Liu, and Y. Zhu, "Research on the application of a marsupial robot for coal mine rescue," in *Proc. Int. Conf. Intell. Robot. Appl.*, Oct. 2008, pp. 1127–1136.

[51] S. Thrun, S. Thayer, W. Whittaker, C. Baker, W. Burgard, D. Ferguson, D. Hannel, M. Montemerlo, A. Morris, Z. Omohundro, C. Reverte, and W. Whittaker, "Autonomous exploration and mapping of abandoned mines," *IEEE Robot. Autom. Mag.*, vol. 11, no. 4, pp. 79–91, Dec. 2004, doi: 10.1109/mra.2004.1371614.

[52] A. Morris *et al*., "Recent developments in subterranean robotics," *J. Field Robot.*, vol. 23, no. 1, pp. 35–57, Jan. 2006, doi: 10.1002/rob.20106.

[53] J. C. Ralston, D. W. Hainsworth, D. C. Reid, D. L. Anderson, and R. J. McPhee, "Recent advances in remote coal mining machine sensing, guidance, and teleoperation," *Robotica*, vol. 19, no. 5, pp. 513–526, Sep. 2001, doi: 10.1017/s0263574701003447.

[54] R. Murphy, J. Kravitz, S. Stover, and R. Shoureshi, "Mobile robots in mine rescue and recovery," *IEEE Robot. Autom. Mag.*, vol. 16, no. 2, pp. 91–103, Jun. 2009, doi: 10.1109/MRA.2009.932521.

[55] B. G. Guzman, T. M. Cortes, A. R. Lopez, and A. G. Armada, "Design of a communication, vision and sensory system for a rescuer robot in coal mine areas," in *Proc. Int. Conf. Wireless Netw. Mobile Commun. (WINCOM)*, Nov. 2017, pp. 499–509, doi: 10.1109/wincom.2017.8238150.

[56] Y. T. Li, "Development and applications of rescue robots for explosion accidents in coal mines," *J. Field Robot.*, vol. 37, pp. 466–489, Oct. 2019, doi: 10.1002/rob.21920.

[57] Y. W. Li, S. R. Ge, X. Wang, and H. B. Wang, "Steps and stairs-climbing capability analysis of six-tracks robot with four swing arms," in *Proc. Int. Conf. Adv. Des. Manuf. Eng. (ADME)*, vols. 397–400, Jul. 2013, pp. 1459–1468, doi: 10.4028/www.scientific.net/AMM.397-400.1459.

[58] Specialized Robot. *KRZ 1 Disaster Area Detection Robot*. Accessed: Apr. 26, 2020. [Online]. Available: http://www.citichickc.com/productinfo/795587.html

[59] W. Wang, W. Dong, Y. Su, D. Wu, and Z. Du, "Development of Search-and-rescue robots for underground coal mine applications," *J. Field Robot.*, vol. 31, no. 3, pp. 386–407, May 2014, doi: 10.1002/rob.21501.

[60] W. T. Lu, G. H. Lou, C. M. Yang, and S. S. Zhang, "Research and application of coal mine rescue and detecting robots," in *Proc. Int. Symp. Vehicle, Mech., Electr. Eng. (ISVMEE)*, vols. 494–495, Dec. 2014, pp. 1042–1045, doi: 10.4028/www.scientific.net/AMM.494-495.1042.

[61] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105–119, Jan. 2010, doi: 10.1109/tpami.2008.275.

[62] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/b:Visi.0000029664.99615.94.

[63] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008, doi: 10.1016/j.cviu.2007.09.014.

[64] Z. G. Niu, L. J. Li, and T. Wang, "3D reconstruction based on binocular stereo vision of robot," in *Proc. Int. Conf. Mater. Products Manuf. Technol. (ICMPMT)*, vol. 2011, pp. 645–648, doi: 10.4028/www.scientific.net/AMR.338.645.

[65] S. He, Z. Tong, G. Ma, M. Fan, L. Lingzhou, and S. Tang, "Research on stereo vision matching algorithm for rescue robot," in *Proc. Int. Conf. Robot. Autom. Sci. (ICRAS)*, Aug. 2017, pp. 35–38, doi: 10.1109/ICRAS.2017.8071912.

[66] H. Kai and M. Xianmin, "Real-time monitoring for the mining robot based on an improved SIFT matching algorithm," in *Proc. 10th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2017, pp. 1–5.

[67] C. C. Shang and H. W. Ma, "Coal mine robot binocular vision recognition system based on fuzzy neural network," in *Proc. Int. Conf. Electr., Autom. Mech. Eng.*, 2015, pp. 95–98.

[68] Y. Wang, P. Tian, B. Zheng, Y. Zhou, Y. Li, and X. Wu, "Special robot vision algorithm test platform in virtual reality environment," in *Proc. Int. Conf. Internet Things*, Jul. 2019, pp. 416–421, doi: 10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00089.

[69] X. Zhao, Q. Luo, and B. Han, "Survey on robot multi-sensor information fusion technology," in *Proc. 7th World Congr. Intell. Control Autom.*, 2008, p. 5019, doi: 10.1109/WCICA.2008.4593742.

[70] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Corrigendum to–multisensor data fusion: A review of the state-of-the-art' [Information fusion 14 (1) (2013) 28–44]," *Inf. Fusion*, vol. 14, no. 4, p. 562, Oct. 2013, doi: 10.1016/j.inffus.2012.10.004.

[71] K. Park, S. Kim, and K. Sohn, "High-precision depth estimation using uncalibrated LiDAR and stereo fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 321–335, Jan. 2020, doi: 10.1109/tits.2019.2891788.

[72] T. Wang and X. Q. Guan, "Research on obstacle avoidance of mobile robot based on multi-sensor fusion," in *Proc. Int. Conf. Cyber Secur. Intell. Anal. (CSIA)*, Feb. 2019, pp. 760–770, doi: 10.1007/978-3-030-15235-2_104.

[73] Q. Wang, Y. Zhang, W. Shi, and M. Nie, "Laser ranging-assisted binocular visual sensor tracking system," *Sensors*, vol. 20, no. 3, p. 688, Jan. 2020, doi: 10.3390/s20030688.

[74] J.-C. Piao and S.-D. Kim, "Real-time visual–inertial SLAM based on adaptive keyframe selection for mobile AR applications," *IEEE Trans. Multimedia*, vol. 21, no. 11, pp. 2827–2836, Nov. 2019, doi: 10.1109/tmm.2019.2913324.

[75] J. Kelly and G. S. Sukhatme, "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration," *Int. J. Robot. Res.*, vol. 36, nos. 13–14, p. 1619, Dec. 2017, doi: 10.1177/0278364910382802.

[76] A. Ardeshir Goshtasby and S. Nikolov, "Image fusion: Advances in the state of the art," *Inf. Fusion*, vol. 8, no. 2, pp. 114–118, Apr. 2007, doi: 10.1016/j.inffus.2006.04.001.

[77] C. Cai, R. Qiao, H. Meng, and F. Wang, "A novel measurement system based on binocular fisheye vision and its application in dynamic environment," *IEEE Access*, vol. 7, pp. 156443–156451, 2019, doi: 10.1109/access.2019.2949172.

[78] X. Mai, X. Yuwen, Y. Wang, and L. Chen, "Teleoperation master-slave robot based on binocular vision," in *Proc. 11th Int. Workshop Human Friendly Robot. (HFR)*, Nov. 2018, pp. 72–77, doi: 10.1109/HFR.2018.8633522.

[79] T. Cao, Z.-Y. Xiang, X.-J. Gong, and J.-L. Liu, "Terrain reconstruction of lunar surface based on binocular fisheye camera," in *Proc. 11th World Congr. Intell. Control Autom.*, Jun. 2014, pp. 2463–2468, doi: 10.1109/WCICA.2014.7053109.

[80] S. Zhao, B. Zhang, L. Li, and Z. Cao, "Fisheye lens camera system calibration and localization error analysis," in *Proc. Int. Conf. Comput. Des. Appl.*, Jun. 2010, pp. 4441–4445, doi: 10.1109/ICCDA.2010.5540898.

[81] H. Kim, D. Kim, and J. Paik, "Region-based adaptive automatic focal-length estimation method for fisheye-lens distortion correction," in *Proc. Int. Conf. Electron., Inf. Commun. (ICEIC)*, Jan. 2014, pp. 1–2, doi: 10.1109/ELINFOCOM.2014.6914374.

[82] M. Lee, H. Kim, and J. Paik, "Correction of barrel distortion in fisheye lens images using image-based estimation of distortion parameters," *IEEE Access*, vol. 7, pp. 45723–45733, 2019, doi: 10.1109/access.2019.2908451.

[83] B. Zhang, J. Wang, J. Li, and X. Wang, "Fisheye lens distortion correction based on an ellipsoidal function model," in *Proc. Int. Conf. Ind. Informat. Comput. Technol., Intell. Technol., Ind. Inf. Integr.*, Dec. 2015, pp. 217–221, doi: 10.1109/ICIICII.2015.136.

[84] P. Kellnhofer, P. Didyk, T. Ritschel, B. Masia, K. Myszkowski, and H. P. Seidel, "Motion parallax in stereo 3D: Model and applications," *ACM Trans. Graph.*, vol. 35, no. 6, p. 12, Nov. 2016, doi: 10.1145/2980179.2980230.

[85] M. Mansour, P. Davidson, O. Stepanov, and R. Piche, "Relative importance of binocular disparity and motion parallax for depth estimation: A computer vision approach," *Remote Sens.*, vol. 11, no. 17, p. 16, Sep. 2019, doi: 10.3390/rs11171990.

[86] M. F. Bradshaw and B. J. Rogers, "The interaction of binocular disparity and motion parallax in the computation of depth," *Vis. Res.*, vol. 36, no. 21, pp. 3457–3468, 1996, doi: 10.1016/0042-6989(96)00072-7.

[87] Y. Song and X. Zhang, "An active binocular integrated system for intelligent robot vision," in *Proc. IEEE Int. Conf. Intell. Secur. Informat.*, Jun. 2012, pp. 48–53, doi: 10.1109/ISI.2012.6284090.

[88] L. Lonini, Y. Zhao, P. Chandrashekhariah, B. E. Shi, and J. Triesch, "Autonomous learning of active multi-scale binocular vision," in *Proc. IEEE 3rd Joint Int. Conf. Develop. Learn. Epigenetic Robot. (ICDL)*, Aug. 2013, pp. 1–6, doi: 10.1109/DevLrn.2013.6652541.

[89] A. Lelais, J. Mahn, V. Narayan, C. Zhang, B. E. Shi, and J. Triesch, "Autonomous development of active binocular and motion vision through active efficient coding," *Frontiers Neurorobot.*, vol. 13, p. 14, Jul. 2019, doi: 10.3389/fnbot.2019.00049.

[90] S. K. Ramakrishnan, D. Jayaraman, and K. Grauman, "Emergence of exploratory look-around behaviors through active observation completion," *Sci. Robot.*, vol. 4, no. 30, p. 12, May 2019, doi: 10.1126/scirobotics.aaw6326.

GUODONG ZHAI received the B.S. degree in mechanical design and manufacture and the M.S. and Ph.D. degrees from the China University of Mining and Technology-Beijing, Beijing, China, in 1996, 2002, and 2011, respectively. He has been working at the School of Mechanical Electronic and Information Engineering, China University of Mining and Technology at Beijing, since 1996. His research interests include the mine dust removal equipment design, human–machine environment safety, mine rescue robot research, and so on.

WENTAO ZHANG received the B.S. degree in mechanical engineering from Nanjing Agricultural University, Nanjing, China, in 2019. He is currently pursuing the master's degree with the China University of Mining and Technology-Beijing, Beijing, China. His current research interests include robot design and control, machine vision, mine safety, and so on.

WENYUAN HU received the B.S. degree in mechanical engineering from the China University of Mining and Technology-Beijing, Beijing, China, in 2020. His current research interests include robot design, machine vision, and so on.

ZHENDONG JI received the B.S. degree in mechanical engineering from the China University of Mining and Technology-Beijing, Beijing, China, in 2020. His current research interests include robot design, mechanical manufacturing, and so on.