Article

Research on Target Ranging Method for Live-Line Working Robots

Guoxiang Hua [1,2], Guo Chen [3], Qingxin Luo [3] and Jiyuan Yan [1,*]

[1]   School of Automation, Wuxi University, Wuxi 214105, China; hgx@cwxu.edu.cn
[2]   School of electrical and Electronic Engineering, North China Electric Power University, Beijing 100096, China
[3]   School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China
[*]   Correspondence: wuxi_yjy@cwxu.edu.cn

Abstract: Due to the operation of live-line working robots at elevated heights for precision tasks, a suitable visual assistance system is essential to determine the position and distance of the robotic arm or gripper relative to the target object. In this study, we propose a method for distance measurement in live-line working robots by integrating the YOLOv5 algorithm with binocular stereo vision. The camera's intrinsic and extrinsic parameters, as well as distortion coefficients, are obtained using the Zhang Zhengyou calibration method. Subsequently, stereo rectification is performed on the images to establish a standardized binocular stereovision model. The Census and Sum of Absolute Differences (SAD) fused stereo matching algorithm is applied to compute the disparity map. We train a dataset of transmission line bolts within the YOLO framework to derive the optimal model. The identified bolts are framed, and the depth distance of the target is ultimately calculated. And through the experimental verification of the bolt positioning, the results show that the method can achieve a relative error of 1% in the proximity of positioning. This approach provides real-time and accurate environmental perception for symmetrical structural live-line working robots, enhancing the stability of these robots.

Keywords: live-line working robot; binocular stereo vision; target ranging; Census algorithm

1. Introduction

By employing live-line working robots, high-voltage transmission line inspection and maintenance can see a substantial improvement in efficiency and a reduction in the labor intensity of workers, compared to traditional inspection methods [1,2]. Robots in operation often employ sensors to detect components and obstacles in the surroundings. However, the accuracy of obstacle recognition by sensors can be affected by factors such as lighting and weather, leading to robot misjudgments and accidents. Therefore, achieving the precise outdoor positioning of targets is a major challenge in the application of live-line working robots.

Common methods for target localization in live-line working robots include the use of laser radar, millimeter-wave radar, and visual cameras. Laser radar can accurately identify obstacles ahead and measure their distance but is costly and involves complex laser signal processing [3]. Millimeter-wave radar is insensitive to target shapes and cannot distinguish information about the type of target. On the other hand, visual cameras, with mature hardware technology and lower costs, can use software algorithms to obtain information about target types and distances.

Currently, inspection robots mainly employ monocular ranging systems for their simplicity, low cost, and ease of development. Cao et al. [4] proposed a monocular ranging algorithm for de-icing robots, deriving a distance expression based on the pinhole imaging model combined with the robot's pitch angle, although it involves multiple parameters and complex calculations. Zhang et al. [5] introduced a method for obstacle recognition

and ranging positioning based on monocular vision, determining obstacle types from captured images, establishing a ranging geometric model based on recognized obstacle feature points and the camera's position, but this method is significantly influenced by the environment. Ye et al. [6] presented a texture-based ground line positioning method suitable for outdoor complex lighting environments, but it has real-time algorithm limitations. Cheng et al. [7] and others proposed a monocular ranging algorithm for visual navigation of line-following robots, utilizing the imaging characteristics of the robot-walked wire in the lens to calculate the distance to obstacles based on the known distance and coordinate difference at the near-critical point, but the algorithm is cumbersome and lacks real-time performance. Relying on monocular cameras for ranging no longer meets the actual operational needs of live-line working robots. Feng et al. [8] utilized a binocular intelligent inspection robot, combined with image processing and analysis techniques, to realize the detection of insulation defects on the surface of cable lines and diameter measurement, but the method could not realize automatic measurement. Xie et al. [9] proposed a transmission line dance reconstruction method based on binocular vision. The transmission line itself is used as a feature to realize the three-dimensional dance reconstruction of the transmission line. This method is not applicable to measurement and localization in close range. Zhou et al. [10] proposed a binocular 3D scanning measurement system, which collects the position information through binocular cameras and performs noise reduction, filtering, and integral measurement through the program to achieve high measurement accuracy, but the preliminary work involved in this method is too cumbersome.

Therefore, this paper proposes a target ranging method based on binocular vision for electric power operation robots. Firstly, the YOLOv5 algorithm is used to realize the recognition and selection of the target, and the binocular camera is calibrated to obtain the internal and external parameters of the camera using Zhang's calibration method, and the aberration correction is completed. Binocular stereo vision is used to complete the target localization, and the Census transform is improved and combined with the SAD algorithm to improve the effect of image matching and target localization accuracy. After experimental verification, the measurement accuracy of this method meets the actual demand, and the cost is low.

2. Binocular Stereo Vision Ranging

2.1. Camera Calibration

Camera calibration involves establishing the relationship between image pixel points and real scene location points. Its objective is to determine the internal, external, and distortion parameters of the camera. This process serves as the foundation for the stereo correction module and subsequent 3D scene applications, making it a crucial step in binocular vision [11,12]. The calibration incorporates four main coordinate systems: the world coordinate system, the camera coordinate system, the image coordinate system, and the pixel coordinate system [13–15]. Real-world 3D coordinates are transformed into 2D coordinates, as illustrated in Figure 1, depicting the relationships within the coordinate systems.

The world coordinate system defines the positions of actual objects in the real world, using X_w, Y_w, and Z_w. Its origin varies based on the scene. The camera coordinate system has its origin at the optical center of the camera, with X_c and Y_c axes parallel to the x axis and y axis, respectively. The Z_c axis aligns with the optical center, forming the camera coordinate system. In the image coordinate system, the origin (*O*) is where the main optical axis intersects the image plane. For the pixel coordinate system, the origin is typically at the image vertex. Figure 2 illustrates the directions of the coordinate axes for both systems. The first three coordinate systems use length units, while the pixel coordinate system uses pixels [16]. If the world coordinates of a point P are (X_w, Y_w, Z_w) and its imaging point in

the pixel coordinate system is $p(u, v)$, the transformation relationship from world to pixel coordinates is as follows:

$$
Z \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}
$$
$$
= \begin{bmatrix} \lambda_x & 0 & u_0 & 0 \\ 0 & \lambda_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}
$$
$$
= M_1 M_2 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}
\tag{1}
$$

where R is the orthogonal rotation matrix and t is the translation matrix: these two matrices are the external parameters of the camera. They describe the relationship between the world coordinate system and the camera coordinate system, as well as the binocular camera's position; f is the focal length of the camera, with the origin (O) of the image coordinate system as the principal point. The point (u_0, v_0) on the pixel coordinate system has different physical dimensions on the horizontal and vertical axes, represented by d_x and d_y, respectively. $\lambda_x = f/d_x, \lambda_y = f/d_y$. The projection matrix (M) combines these parameters. M_1 is the internal parameter, associated with the camera's internal structure, while M_2 is the external parameter, indicating the camera's relative position in physical space. Internal parameters often include distortion functions, which account for distortions and variations in capturing image data within a limited view. These typically encompass radial and tangential distortion.
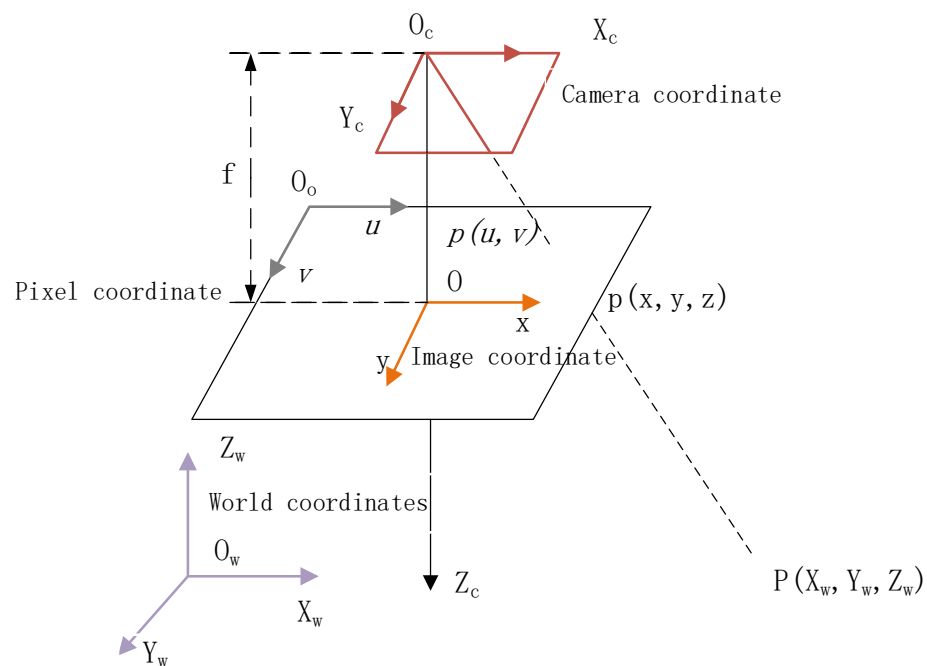

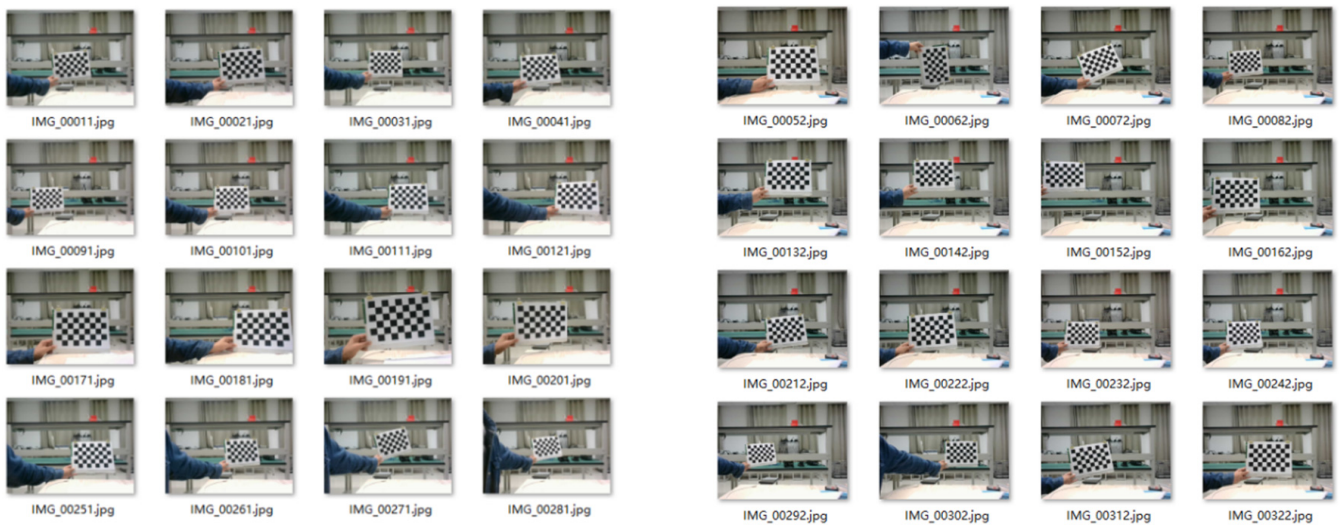**Figure 1.** Coordinate relationship diagram.

Figure 2. Left and right camera calibration images.

Camera calibration is categorized into traditional and automatic methods. Zhengyou Zhang's approach strikes a balance, offering simple yet mature technology. Utilizing a fixed checkerboard grid during image acquisition at various angles and positions, the calibration process establishes equations based on key points, with parameter values determined through maximum likelihood estimation [17–19]. In this paper, Zhengyou Zhang's method, implemented with MATLAB calibration toolbox, is employed for offline calibration to find the camera's internal reference, external reference, and distortion parameters.

In this paper, a 9×6 checkerboard grid, featuring 8×5 corner points and a size of 30 mm $\times$ 30 mm, is used for calibration. Twenty sets of photographs with diverse poses are captured using a binocular camera and segmented (see Figure 2). Employing MATLAB R2022a's Stereo Camera Calibrator toolbox, we import photos, extract corner points, compute world coordinates, and determine the internal reference matrix and distortion coefficients. The left camera serves as the world coordinate system, facilitating the calibration of the right camera's external parameters relative to the left camera, as detailed in Table 1.

Table 1. Internal and external parameters and distortion coefficients of the camera.

	Left Camera			Right Camera		
Internal reference matrix	$\begin{bmatrix} 321.18 & 0 & 320.83 \\ 0 & 322.99 & 259.71 \\ 0 & 0 & 1 \end{bmatrix}$			$\begin{bmatrix} 322.18 & 0 & 329.83 \\ 0 & 321.99 & 248.71 \\ 0 & 0 & 1 \end{bmatrix}$		
External reference matrix	$R = \begin{bmatrix} 0.99999 & 0.00141 & 0.00214 \\ -0.00139 & 0.99999 & 0.01178 \\ -0.00213 & -0.01182 & 0.99996 \end{bmatrix}$			$T = \begin{bmatrix} -118.284 & \cdots & 0.319982 & -0.843788 \end{bmatrix}$		
Distortion coefficients	$\begin{bmatrix} -0.0549 & 0.1909 & -0.0012 & 0.0023 & -0.3401 \end{bmatrix}$			$\begin{bmatrix} -0.0527 & 0.2163 & 0.0013 & 0.0036 & -0.2206 \end{bmatrix}$		

Following camera calibration, the spatial position of each calibration plate in relation to the camera can be back-calculated using the calibrated camera parameters, as illustrated in Figure 3:
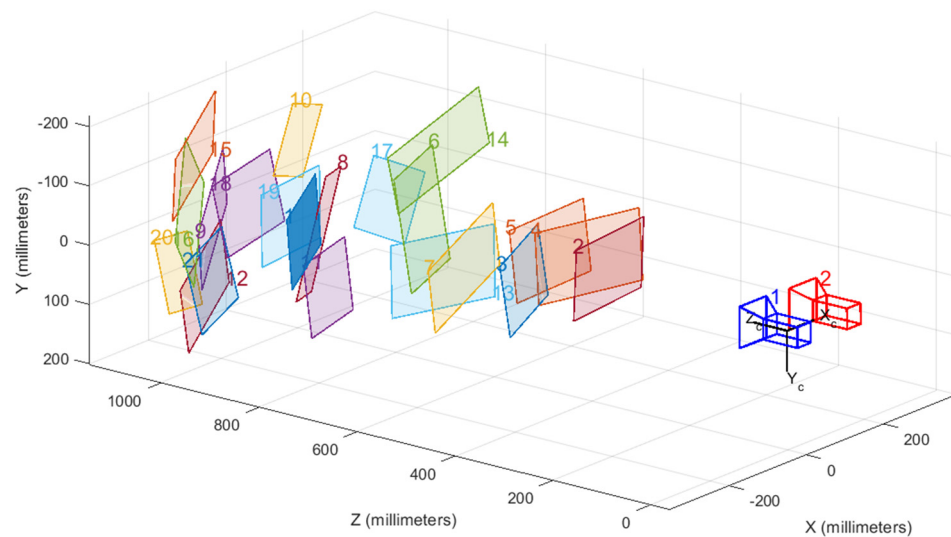
Figure 3. Relative position of calibration plate and camera.

2.2. Stereoscopic Correction

Stereo correction starts with image correction, which is the use of camera distortion parameters to de-distort the image. Then, polar line correction is performed on the image with the aim of making the corresponding pixel points in the left and right images of a horizontally placed binocular camera strictly on the same horizontal line [20]. In stereo correction, the most commonly used method is Bouguet's stereo correction algorithm.

The algorithm's principle involves a polar line correction, depicted in Figure 4. Here, P represents a point in space, with its projection points in the left and right cameras as P_l and P_r, respectively. Binocular stereo correction entails rotating and translating the two small-aperture imaging camera models for calibration, ensuring a single horizontal directional offset post-correction [21]. Since corresponding pixel points in binocular images adhere to a pair of polar geometries, the polar line constraint dictates that a feature point on the imaging plane must have its matching point on the corresponding pair of polar lines in the other imaging plane. Calculating image parallax to find the corresponding point of a pixel point only necessitates a linear search on that line, thereby expediting the calculation and reducing false match rates.
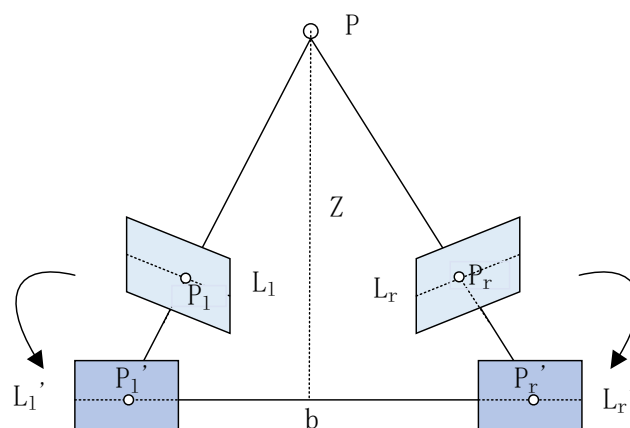

Figure 4. Stereo calibration principle.

To align the optical axes of the binocular camera in parallel, two methods are employed: one involves fixing one camera while adjusting the position of the other, and the second method adjusts both cameras simultaneously. The Bouguet correction algorithm adopts the

latter approach, rotating each camera plane by half, minimizing left–right ghosting errors and maximizing the common field of view. The correction effect is illustrated in Figure 5.
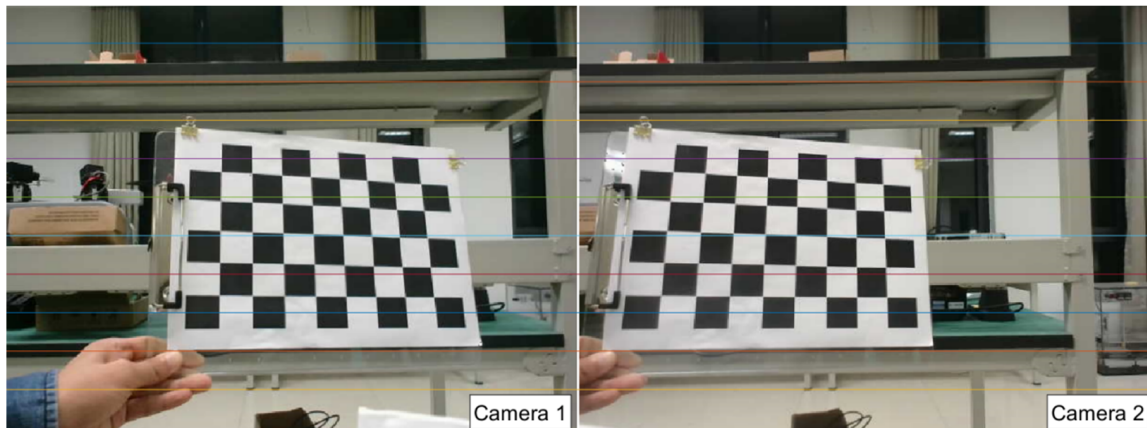

Figure 5. Polar line correction.

2.3. Stereo Matching

Stereo matching is a technique for extracting depth information from a flat 2D image and is a key part of binocular stereo vision ranging technology. By matching the corresponding pixel points in the stereo-corrected binocular image, a parallax map is formed by calculating the difference between the left and right images of these corresponding points in the pixel coordinate system. According to the principle of binocular stereo vision ranging, the most important thing is the acquisition of parallax, and different methods of parallax acquisition correspond to different matching strategies. Stereo matching algorithms are complex and diverse, mainly divided into global matching, local matching, and semi-global matching algorithms. The commonly used ones are the BM algorithm [22], SGBM algorithm [23], GC algorithm [24], etc.

The Census transform is a stereo matching algorithm based on local regions, defining a window in the image and traversing the entire image [25]. The reference pixel is the center of the window, and the gray value of each pixel in the window is compared with the reference pixel's gray value. If the region's gray value is less than or equal to the reference pixel's gray value, it is recorded as 0; if greater, it is recorded as 1. The original neighboring gray value relationships are converted into binary characters, forming a binary string. The comparison of Hamming distance between the reference pixel and the matching pixel yields the matching generation value.

$$C(p,q) = \begin{cases} 0, \cdots f(p) \leq f(q) \\ 1, \cdots f(p) > f(q) \end{cases} \tag{2}$$

The algorithm incorporates the concept of nonparametric transformation [26]. The Census cost calculation process is straightforward in principle, operationally swift, and resilient to illumination changes. However, it exhibits high dependence on the grayscale value of the central pixel point, leading to significant noise in the results [27].

Among local algorithms, the absolute Sum of Absolute Differences (SAD) is a frequently employed similarity measure function due to its efficiency. Nevertheless, it is more sensitive to illumination changes. The expression for SAD is as follows:

$$C_{SAD} = \sum_{(i,j)\in w} |I_L(x+i, y+j) - I_R(x+i+d, y+j+d)| \tag{3}$$

where $I_L(x,y)$ and $I_R(x,y)$ are the pixel grayscale values of (x,y) in the left and right views, respectively, w is the template window size, and all pixel points are traversed by incrementing *i,j*. After calculating the corresponding C_{SAD}, the value of d is added by 1,

and the same operation steps are performed. At the end of the traversal, the point with the smallest C_{SAD} is selected as the matching point, and the d is the corresponding parallax value of that point.

A pixel reference point (x_0, y_0) is chosen in the left view. Subsequently, the matching window in the right view is systematically moved for right-to-left pixel matching, commencing at row y_0 under the polar line constraint. This search step is iterated until reaching the predefined maximum parallax search range [28]. As depicted in Figure 6, the SAD function of the matching window is locally optimal when its value is minimal, and this point is the best matching point $B(x_1, y_0)$; then, the parallax of stereo matching is $d = x_1 - x_0$. After matching a sufficient number of points, the parallax map can be derived.
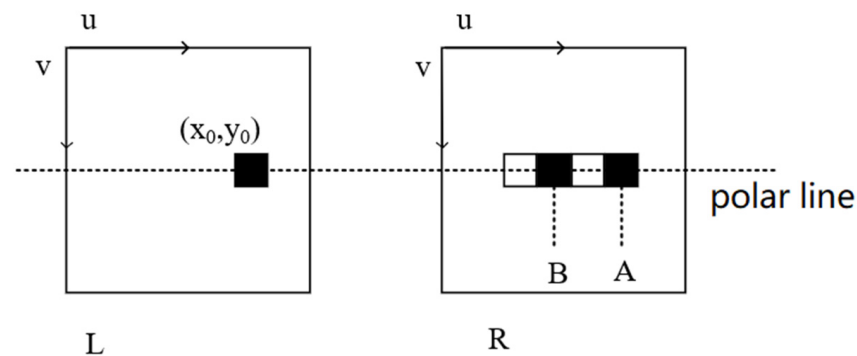

Figure 6. Stereo matching search process.

To improve the efficiency and robustness of matching, a weighted fusion of the two algorithms is performed, as in Figure 7:
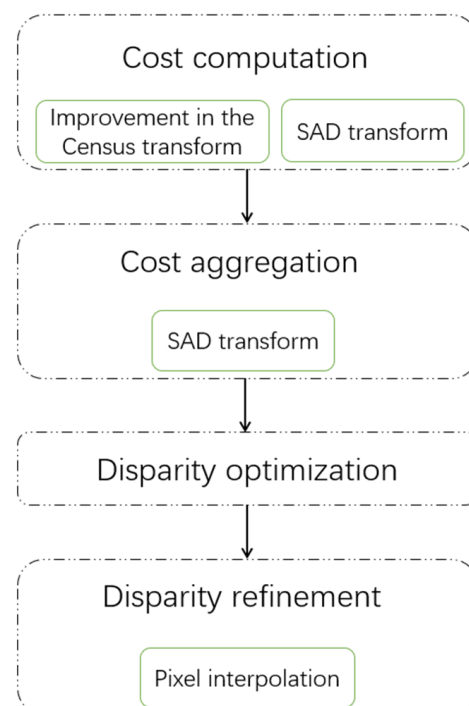

Figure 7. Algorithm flow chart.

1. Matching cost calculation. The left view is used as the reference image, and a central pixel point is selected to create a rectangular window. The right view is taken as the matching image and searched along the polar lines, and the Census and SAD algorithms are fused by a weighted approach with the expression:

$$C(x,y,d) = \alpha \sum_{(i,j) \in w} SAD(x+i, y+i, d) +$$
$$(1-\alpha) \sum_{(i,j) \in w} C(x+i, y+i, d) \tag{4}$$

The algorithm combines the weights of the Census algorithm and the SAD algorithm, adjusting the weight of the Census algorithm (i.e., reducing the value) in the presence of significant lighting changes. For real-time requirements, the weight of the SAD algorithm can be increased.

2. Parallax calculation. When the cost function is determined, the minimum value is taken as the parallax.
3. Parallax optimization. The computed parallax is a discrete value, which can be pixel-accurate by pixel interpolation, and the filled parallax map is processed by weighted median filtering to eliminate the transverse noise in the map and generate the final parallax map.

Figure 8 illustrates the parallax map resulting from the traditional Census algorithm, while Figure 9 displays the parallax map achieved through the improved stereo matching algorithm.

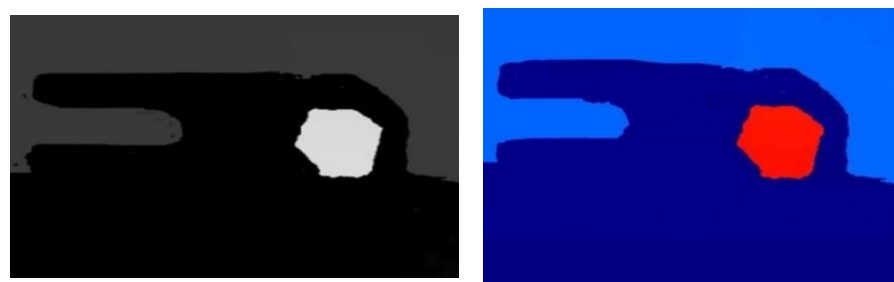**Figure 8.** Parallax diagram of Census algorithm.

Figure 9. Parallax map of fusion algorithm.

The traditional Census algorithm exhibits robustness in scenarios with luminance differences between the left and right views. However, it tends to produce noisier parallax maps in weakly textured or repeated scenes. Conversely, the SAD algorithm boasts higher matching efficiency but is more susceptible to luminance differences and illumination variations. Combining both algorithms through fusion allows for the leveraging of their respective strengths, enhancing the efficiency and reliability of image matching.

We experimentally validate the bolts for ranging using the traditional Census algorithm, the SAD algorithm, and the improved SAD–Census algorithm, mainly comparing them in terms of parallax effect and algorithm running time. In order to quantify the parallax effect, the mis-match rate of the non-obscured region (Nocc) and the error rate of the overall region (All) are calculated, and the average mis-match rate is compared. The experimental results are shown in Table 2. For the weakly textured scene with non-occluded regions, the improved SAD–Census algorithm shows a significant improvement over the

traditional Census algorithm, with the overall matching error rate reduced to 8.3% and the non-occluded mis-matching rate reduced to 6.7%. On the other hand, the timeliness of the image matching algorithm is also verified. The image matching time of the Census algorithm is 3.5 s, and the image matching time of the improved SAD–Census algorithm is reduced to 2.7 s, which improves the efficiency of the algorithm.

Table 2. Comparison of algorithm performance.

	All%	Nocc%	Runtime (s)
SAD	54.1	48.6	1.8
Census	28.8	26.3	3.5
SAD–Census	8.3	6.7	2.7

2.4. Binocular Parallax Ranging

Two laterally parallel cameras synchronously capture images controlled by a computer, with identical parameters and quality for both cameras. A common target point results in corresponding imaging points on the imaging surfaces of the left and right cameras. The positional difference (baseline) between the cameras introduces pixel disparities in the image plane, referred to as parallax. The target distance is then calculated from similar triangles using the parallax information.

Binocular stereo vision is a technique for computing the real distance between a camera and an object using the principle of parallax. Depth information between the camera and the object in the real-world scenario is derived from two-dimensional images captured by two cameras in the same scene at different orientations. Illustrated in Figure 10, the binocular stereo vision ranging system comprises two monocular cameras with parallel Z and Y axes. The optical axis is perpendicular to the image plane, and theoretically, the X-axis extensions of the two cameras coincide.
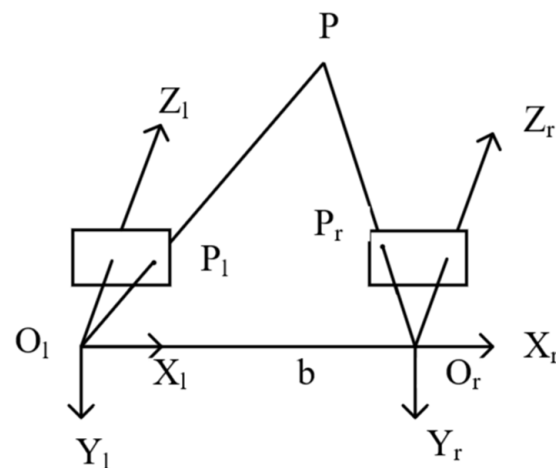

Figure 10. Diagram of binocular camera.

In Figure 11, the optical centers of the two cameras are denoted as O_l and O_r, and the distance between the two cameras is referred to as the baseline, represented by the length '*b*'. The left and right cameras are labeled as *l* and *r*, respectively, where *O* signifies the op-tical center, *I* represents the imaging plane, *P* is any point on the object in space, and P_l and P_r are the projection points of the point onto the imaging surfaces of the left and right cameras. The camera's focal length is denoted as '*f*', and '*z*' represents the depth distance of the target point, which is the desired result to be calculated.
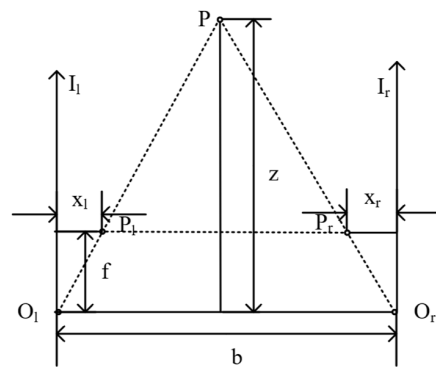
Figure 11. Binocular distance measurement schematic.

According to the triangle similarity principle, $\triangle PO_lO_r \sim \triangle PP_lP_r$ to obtain the following:

$$\frac{z-f}{z} = \frac{P_lP_r}{b} \tag{5}$$

Since $P_lP_r = b - (X_l + X_r)$, $(X_l + X_r) = d$, and d is the parallax of the binocular camera, substituting the above equation, the following can be obtained:

$$\frac{z-f}{z} = \frac{b-d}{b} \tag{6}$$

After transforming the equation, we obtain z:

$$z = \frac{fb}{d} \tag{7}$$

where d is the parallax of the two cameras; the acquisition of parallax is a complex process that requires the stereo matching of binocular images to obtain it. The specific parameters of f and b of the cameras can be obtained through calibration, and the depth distance z can be calculated by combining the d derived from the stereo matching parallax map, whose accuracy is related to the focal length of the camera, the baseline length, and the distance of the object.

As depicted in Figure 12, the parallax increases as the object gets closer and decreases as the object moves farther away. There exists an inverse proportional and nonlinear relationship between distance and parallax. When the parallax is close to 0, a small change in parallax results in a large change in distance, whereas for larger parallax values, a change in parallax induces only a small change in distance. In binocular distance measurement, larger distances are associated with greater errors, while smaller distances exhibit smaller errors. Therefore, the method is well suited for measuring objects at close distances.
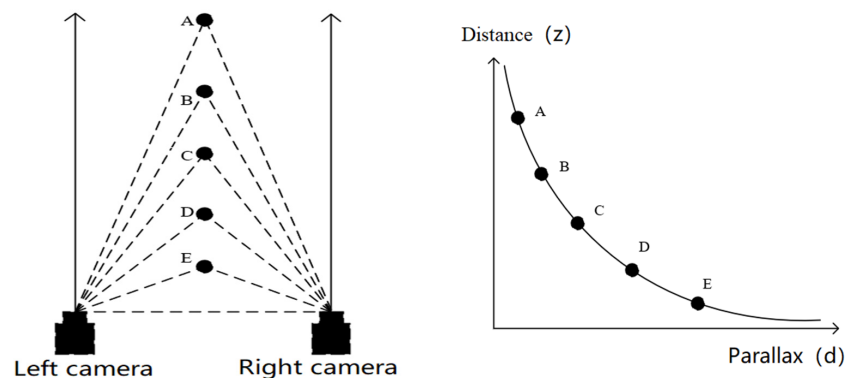

Figure 12. Depth versus parallax.

3. Target Recognition Based on YOLOv5

In recent years, with the advancements in convolutional neural networks and the enhancement of hardware computing power, deep learning algorithms have found extensive applications across the entire spectrum of computer vision. YOLOv5, a single-stage target detection algorithm, has made substantial improvements based on YOLOv4, resulting in significant enhancements in both speed and accuracy. YOLOv5 consists mainly of Input, Backbone, Neck, and Output components. The network structure is illustrated in Figure 13, with particular emphasis on the Backbone and Neck. For this study, YOLOv5s, featuring the smallest network width and depth, is selected as the network model.
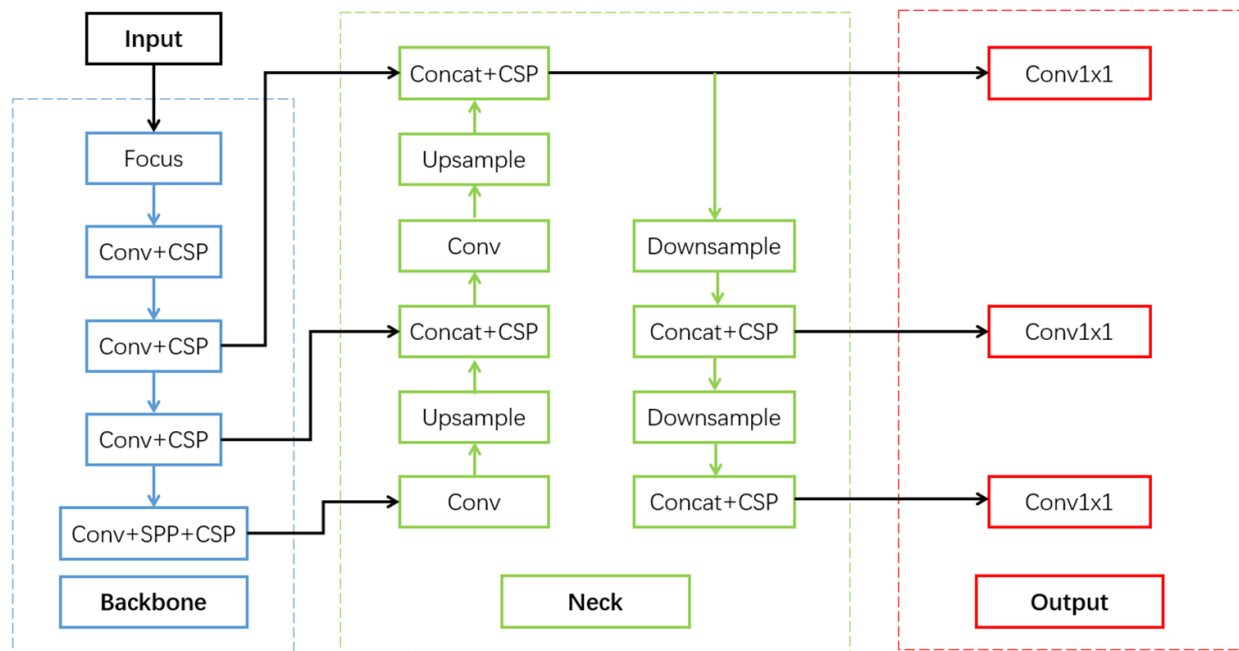

Figure 13. Network structure YOLOv5.

Initially, the original image undergoes adaptive scaling to a 640×640 three-channel image. Mosaic data enhancement is applied, involving random scaling, random cropping, and random scheduling, enriching the dataset and accelerating network training. Post-slicing operations in the Focus network compress the height and width while expanding the channels by four times, resulting in a feature layer of (320, 320, 12). Subsequently, three convolutional normalization and feature extraction operations yield feature layers of (80, 80, 256), (40, 40, 512), and (20, 20, 1024). The Neck performs convolution, upsampling, downsampling, and feature extraction to ultimately generate enhanced feature layers. YoloHead is employed for classification and regression predictions based on these features.

In this paper, we use the original dataset of 600 pieces of bolts of various kinds of gages collected from different scenes and angles, and then the dataset is expanded to 1500 pieces by Mosaic, manually labeled, and converted to the data input format of the Yolo network, with 80% of the data used as the training set of the model and 20% of the data used as the test set. After 300 iterations, the Loss value is stabilized at around 0.03, and the model converges well. In order to verify the accuracy of YOLO v5 on the transmission line fixture bolt recognition model, 300 images in the test set are used for testing, and the test results are shown in Figure 14. After several sets of experiments, the average image recognition rate is 95.89%, and the average recognition time is 17.63 ms, which means that the recognition model can meet the real-time operation requirements of the power line operation robot.
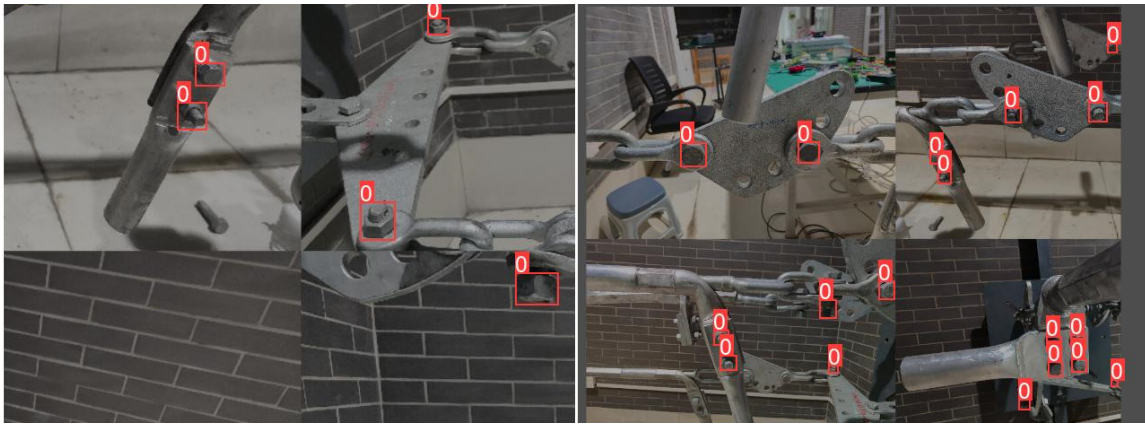
Figure 14. Bolt recognition effect.

4. Experimental Verification

In this paper, a novel transmission line bolt fastening live-line working robot is designed, depicted in Figure 15. In order to ensure the balance of the robot, we have adopted a symmetrical structural design. The robot comprises a drive motor, wire rollers, guided compression wheels, a lifting and lowering sliding table, a base, a multi-degree-of-freedom robotic arm, and an electric screw gun and is equipped with a wide-angle camera and binocular ranging module for visual recognition and localization. The industrial control machine runs on the Windows system, and the PyTorch deep learning framework is configured in Python 3.7, accelerated with CUDNN. The HBV-2V11 binocular camera, manufactured by Huibo Vision Network Technology Co., Ltd., Baoding, China, was chosen to capture the original images. As shown in Figure 16, the binocular camera module's sensor chip is OV4689, featuring a 3.0 mm focal length, a 60 mm baseline, and a resolution set at 1280×480.
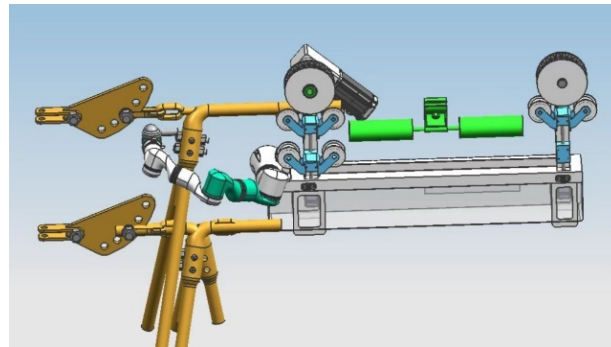

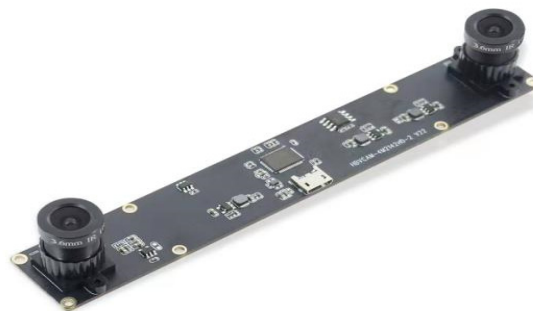**Figure 15.** Bolt fastening robot design.

Figure 16. Binocular camera module.

The robot employs a binocular ranging module to recognize and detect bolts, followed by binocular ranging and positioning. Using the position information, the robotic arm of the robot is controlled to adjust to the suitable operating area. The camera connects to the industrial control machine via the USB port, and the measured position information is communicated by the industrial control machine through the USB serial port. Advanced RISC machine (ARM) real-time communication is utilized, allowing the ARM to control the robot's actions based on the received position information, thus accomplishing the bolt fastening operation task.

The real-time ranging output effects collected by the binocular camera module at different distances during the robot's movement are illustrated in Figure 17.

Figure 17. Bolt ranging effect.

The distance measurement information of the robot at various positions is presented in Table 3. The actual distance ranges from 0.6 to 1.0 m, and the measured distance exhibits variation from the actual values. Notably, when the actual distance is 0.5 m or less, the robot's measured bolt distance aligns more closely with the actual distance, with a relative error within 1%. Through experimental verification, the binocular-vision-based localization method integrated with YOLOv5, as proposed in this paper, proves capable of meeting the operational requirements of the live-line working robot, thereby enhancing the reliability of its operations.

Table 3. Bolt ranging information.

Serial Number	Practical Distance/m	Measuring Distance/m	Relative Error/%
1	1.000	0.970	3.00
2	0.900	0.876	2.67
3	0.800	0.780	2.50
4	0.700	0.684	2.29
5	0.600	0.611	1.83
6	0.500	0.495	1.00
7	0.400	0.402	0.50
8	0.300	0.298	0.67
9	0.200	0.199	0.50
10	0.150	0.149	0.67

The experimental results clearly indicate that the measurement error increases as the target moves farther away from the camera, while the error decreases when the target is

closer. Given that the live-line working robot in this study is employed for bolt fastening, which involves operations in close proximity to the target, a binocular camera with a smaller baseline is deliberately chosen. This selection aligns with the suitability of measuring and localizing in close-proximity scenarios.

5. Conclusions

In this paper, a method of target recognition and localization for live-line robots based on binocular vision is proposed, which combines binocular vision with the YOLOv5 target recognition algorithm and improves the image matching algorithm, improves the effect of parallax map, and identifies the bolts by combining with the YOLOv5 algorithm, so as to complete the real-time target recognition and localization of live-line robots.

The main work is as follows: (1) the Census algorithm is improved by replacing the center value with the pixel mean value and the fixed window with the adaptive window to improve the effect of image matching and enhance the real-time performance of the algorithm; (2) then, weighted fusion with the SAD algorithm is used to overcome the shortcomings of the SAD algorithm which is easily affected by light and noise, and at the same time, the advantages of its simplicity and high efficiency are retained. The bolt is identified and localized by live-line robots. After experimental verification, the method proposed in this paper can efficiently identify the target and complete the localization, and it has a good effect on the localization of the target at a close distance, and the relative error of the localization is less than 1%, which can satisfy the practical application requirements of live-line robots.

The method combines binocular vision technology and deep learning for bolt recognition and localization of live-line robots, which can efficiently and accurately locate the target, improve the intelligent level of bandwidth operation, lay the foundation for the refined operation of live-line robots on transmission lines, and provide assistance for the promotion of the application of the bandwidth operation robot.

Author Contributions: Data Curation, G.H. and G.C.; Formal Analysis, G.H. and Q.L.; Funding Acquisition, G.H. and J.Y.; Investigation, Q.L.; Methodology, G.H. and G.C.; Project Administration, Q.L.; Resources, G.H. and G.C.; Software, G.H., G.C., Q.L. and J.Y.; Supervision, G.H.; Validation, G.H., G.C., Q.L. and J.Y.; Visualization, G.C.; Writing—Original Draft, G.H. and G.C.; Writing—Review and Editing, Q.L. and J.Y. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hu, Y.; Liu, K.; Peng, Y.; Su, Z.M.; Wu, T. Research Status and Development Trend of Live Working Key Technology. *High Volt. Technol.* **2014**, *40*, 1921–1931.
2. Tanveer, M.H.; Fatima, Z.; Zardari, S.; Guerra-Zubiaga, D. An In-Depth Analysis of Domain Adaptation in Computer and Robotic Vision. *Appl. Sci.* **2023**, *13*, 12823. [CrossRef]
3. Kumar, G.A.; Lee, J.H.; Hwang, J.; Park, J.; Youn, S.H.; Kwon, S. LiDAR and Camera Fusion Approach for Object Distance Estimation in Self-Driving Vehicles. *Symmetry* **2020**, *12*, 324. [CrossRef]
4. Cao, W.M. *Research on Visual Control Method for High Voltage Transmission Line Deicing Robot*; Hunan University: Changsha, China, 2014.
5. Zhang, F.; Guo, R.; Lu, S.B.; Li, Z.Y.; Yang, B.; Sun, X.B. Obstacle identification and localization of high-voltage transmission line inspection robot. *China Electr. Power* **2019**, *52*, 111–118.
6. Ye, X.H. *Research on Key Technology of Vision System of Robot Vision System for Crossing Inspection along Ground Line of High Voltage Line Road*; Wuhan University: Wuhan, China, 2019.
7. Cheng, L.; Wu, G.P. Monocular ranging algorithm in visual navigation of high-voltage line patrolling robot. *Optoelectron.-Laser* **2016**, *27*, 941–948. [CrossRef]

8. Feng, S.; Bao, W.; Yang, Y. Defect detection and diameter measurement of cable line based on binocular intelligent inspection robot. *Autom. Instrum.* **2022**, *274*, 253–257.
9. Xie, D.; Sun, T.; Shi, Z. A study on the detection method of transmission line dancer based on three-dimensional reconstruction. *Int. Electron. Meas. Technol.* **2022**, *41*, 96–101.
10. Zhou, M.; Hu, L.; Ou, K. Application of improved dense binocular matching algorithm in 3D reconstruction of transmission line foundation. *Electron. Meas. Technol.* **2022**, *45*, 1–7.
11. Yan, Y.; Zhu, Q.; Lin, Z.; Chen, Q. Camera Calibration in Binocular Stereo Vision of Moving Robot. In Proceedings of the 2006 6th World Congress on Intelligent Control and Automation, Dalian, China, 21–23 June 2006; pp. 9257–9261.
12. Tadic, V.; Toth, A.; Vizvari, Z.; Klincsik, M.; Sari, Z.; Sarcevic, P.; Sarosi, J.; Biro, I. Perspectives of RealSense and ZED Depth Sensors for Robotic Vision Applications. *Machines* **2022**, *10*, 183. [CrossRef]
13. Tang, Y.P.; Pang, C.J.; Zhou, Z.S.; Chen, Y.Y. Binocular omni-directional vision sensor and epipolar rectification in its omni-directional images. *J. Zhejiang Univ. Technol.* **2011**, *1*, 20.
14. Wang, D.; Sun, H.; Lu, W.; Zhao, W.; Liu, Y.; Chai, P.; Han, Y. A novel binocular vision system for accurate 3-D reconstruction in large-scale scene based on improved calibration and stereo matching methods. *Multimed. Tools Appl.* **2022**, *81*, 26265–26281. [CrossRef]
15. Meng, C.; Bao, H.; Ma, Y.; Xu, X.; Li, Y. Visual Meterstick: Preceding Vehicle Ranging Using Monocular Vision Based on the Fitting Method. *Symmetry* **2019**, *11*, 1081. [CrossRef]
16. Deng, C.; Liu, D.; Zhang, H.; Li, J.; Shi, B. Semi-Global Stereo Matching Algorithm Based on Multi-Scale Information Fusion. *Appl. Sci.* **2023**, *13*, 1027. [CrossRef]
17. Liu, H.; Zhao, Y.; Wang, W. A Matching Algorithm Based on the Fusion of SAD and Census. In Proceedings of the 3rd International Conference on Wireless Communications and Applications(ICWCA 2019), Haikou, China, 21–23 December 2019; p. 7.
18. Turski, J. The Conformal Camera in Modeling Active Binocular Vision. *Symmetry* **2016**, *8*, 88. [CrossRef]
19. Yu, Y.; Fan, S.; Li, L.; Wang, T.; Li, L. Automatic Targetless Monocular Camera and LiDAR External Parameter Calibration Method for Mobile Robots. *Remote Sens.* **2023**, *15*, 5560. [CrossRef]
20. Zhang, X.; Zhao, Y.; Wang, H.; Zhai, H.; Sun, H.; Zheng, N. End-to-end learning of self-rectification and self-supervised disparity prediction for stereo vision. *Neurocomputing* **2022**, *494*, 308–319. [CrossRef]
21. Cai, L.; Zhou, C.; Wang, Y.; Wang, H.; Liu, B. Binocular Vision-Based Pole-Shaped Obstacle Detection and Ranging Study. *Appl. Sci.* **2023**, *13*, 12617. [CrossRef]
22. Lei, X.C. Research on Denoising of Brain MRI of Alzheimer's Disease Based on BM3D Algorithm. *Int. J. Health Syst. Transl. Med.* **2021**, *1*, 33–43.
23. Li, Z.; Chu, T.; Dan, N.; Yang, W.; Zhang, H. Research on a random sampling method for bulk grain based on the M-Unet and SGBM algorithms. *J. Stored Prod. Res.* **2023**, *104*, 102200. [CrossRef]
24. Lu, B.; Sun, L.; Yu, L.; Dong, X. An improved graph cut algorithm in stereo matching. *Displays* **2021**, *69*, 102052. [CrossRef]
25. Zhou, Z.; Pang, M. Stereo Matching Algorithm of Multi-Feature Fusion Based on Improved Census Transform. *Electronics* **2023**, *12*, 4594. [CrossRef]
26. Li, Y.; Wang, C.Y.; Wang, Y.H. Fusion stereo matching algorithm of adaptive SAD with Super-Pixel segmentation constrains and census. *Opt. Technol.* **2022**, *48*, 478–485. [CrossRef]
27. Wang, Y.; Gu, M.; Zhu, Y.; Chen, G.; Xu, Z.; Guo, Y. Improvement of ad-census algorithm based on stereo vision. *Sensors* **2022**, *22*, 6933. [CrossRef]
28. Qi, J.; Liu, L. The stereo matching algorithm based on an improved adaptive support window. *IET Image Process.* **2022**, *16*, 2803–2816. [CrossRef]