energies

MDPI

Article

Research on Positioning Method of Coal Mine Mining Equipment Based on Monocular Vision

Rui Yu [1,2], Xinqiu Fang [1,*], Chengjun Hu [3,4], Xiuyu Yang [2], Xuhui Zhang [5], Chao Zhang [5], Wenjuan Yang [5], Qinghua Mao [5] and Jicheng Wan [5]

1 School of Mines, China University of Mining and Technology, Xuzhou 221116, China
2 Wangjialing Coal Mine, China Coal Husain Group Co., Ltd., Yuncheng 043300, China
3 College of Energy and Mining, China University of Mining and Technology (Beijing), Beijing 100083, China
4 China Coal Underground Engineering Intelligent Research Institute Co., Ltd., Tianjin 300000, China
5 College of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, China
* Correspondence: xinqiufang@cumt.edu.cn; Tel.:+86-139-1204-0883

Abstract: In view of the insufficient characteristics and depth acquisition difficulties encountered in the process of uniocular vision measurement, the posture measurement scheme of tunneling equipment based on uniocular vision was proposed in this study. The positioning process of coal mine tunneling equipment based on monocular vision was proposed to extract the environmental features and match the features, and the RANSAC algorithm was used to eliminate the pair of mismatching points. This was done to solve the optimized matching pair and realize the pose estimation of the camera. The pose solution model based on the triangulation depth calculation method was proposed, and the PNP solution method was adopted based on the three-dimensional spatial point coordinates so as to improve the visual measurement accuracy and stability and lay the foundation for the 3D reconstruction of the roadway. This was done to simulate the downhole environment to build an experimental verification platform for monocular visual positioning. The experimental results showed that the position measurement accuracy of the uniocular visual roadheader positioning method within 60 mm and 1.3° could realize the accurate registration of the point cloud in the global coordinate system. The time required for the whole monocular visual positioning was only 179 ms, so it had good real-time performance.

Keywords: monocular vision; 3D reconstruction; equipment position; intelligent tunneling

Citation: Yu, R.; Fang, X.; Hu, C.; Yang, X.; Zhang, X.; Zhang, C.; Yang, W.; Mao, Q.; Wan, J. Research on Positioning Method of Coal Mine Mining Equipment Based on Monocular Vision. *Energies* **2022**, *15*, 8068. https://doi.org/10.3390/en15218068

Academic Editor: Sergey Zhironkin

Received: 20 September 2022
Accepted: 27 October 2022
Published: 30 October 2022

1. Introduction

The global positioning of underground mine tunneling equipment is the key to realizing directional navigation, automatic cutting and collaborative control between equipment [1]. With the characteristics of limited underground operation space, numerous equipment and insufficient light [2], it is difficult to realize roadway visualization and equipment positioning, mainly due to the accurate detection and equipment positioning of the underground environment [3].

In view of the positioning problem of tunneling equipment, many research institutions and scholars have carried out fruitful research. Wu Miao China university's mining team [4], with the help of boring machine fuselage posture measurement, by placing a three-target prism on the tunneling robot, using the whole station of three prism point observation constituting three spatial features and inputting the observed feature points into a computer for further processing, realized the boring machine in the roadway space pose solution. Zhang Xuhui of Xi'an University of Science and Technology used the whole station and the inertial guide for joint positioning and fused the measurement data of the whole station and the Kalman filter to realize the accurate positioning of the roadheader [5]. However, due to the simultaneous observation of the three-target prism, it is difficult to achieve measurement in real time using the measurement method based on the whole station, and due to the

existence of large-scale underground station moving problems, it has not been widely used. Fu Shishen et al. put forward an initial autonomous positioning guidance method based on UWB positioning [6], and Liu Chao et al. put forward a hybrid algorithm of UWB, ranging and the TSOA principle to calculate the 3D spatial pose of the roadheader [7]. The above pose measurement method has deficiencies such as time accumulation error and unstable measurement accuracy. Reid P.B. [8] of Queensland Mining Technology Center in Australia applied inertial navigation to the offset detection of hydraulic supports in a fully mechanized mining face, providing a basis for the straightening of the scraper conveyor. The Australian research organization used inertial navigation to obtain the dynamic pose data of fully mechanized mining equipment and designed the LASC system [9] to measure the pose of fully mechanized mining equipment. KHONZI [10] applied inertial navigation to the autonomous positioning of underground equipment, but there is a large cumulative error in the real-time positioning of the equipment using inertial navigation.

The sensors adopted in visual measurement systems mainly include monocular vision, binocular vision and multiocular vision. By collecting the measured object or environmental information, the real-time solution of the equipment position is realized by using post-stage image processing technologies. With the help of a cross laser, Du Yuxin and others [11] established the features and the fuselage position so as to realize the measurement of a roadheader position in a tunneling roadway. Yang Wenjuan [12] set up two parallel laser lines behind the roadway of Xi'an University of Science and Technology and proposed a roadheader position solution model based on two points and three lines to realize the accurate measurement of the roadheader position. A tunneling-robot-positioning method based on binocular vision obtains roadway environmental features through a binocular vision sensor, establishes a relative relationship model between the tunneling robot and the roadway through collected roadway environmental image recognition information and calculates the relative roadway position of the tunneling robot [13]. Aiming at the problem of point cloud registration, Professor Besl of Stanford University in the United States proposed an iterative closest-point algorithm, namely the ICP algorithm [14]. Since then, there have been many improved algorithms based on ICP to achieve accurate point-cloud matching. For example, Zhang [15] used a k–d tree to accelerate the corresponding point search, which had good real-time performance. Senin Bouaziz et al. [16] used sparse representation to improve the real-time performance and accuracy of ICP.

With the continuous progress of computer technology, intelligent visual perception technology has been preliminarily applied to personnel monitoring and video recognition in underground coal mines. The visual SLAM method [17,18] has shown great potential in the field of robot localization. Therefore, this paper proposed a method of tunneling equipment positioning and environmental reconstruction based on monocular vision. The main contributions of this paper are as follows: (1) The positioning process of underground mine tunneling equipment based on uniocular vision was proposed. (2) The improved BRIEF (Binary Robust Independent Elementary Features) descriptor method was used to realize the feature point matching on two frames of images in a complex environment, and the location of the tunneling equipment was solved according to the method based on depth recovery. (3) The ORB (Oriented Fast and Rotated Brief) feature point [19] extraction method was adopted to realize the environmental 3D sparse-point cloud map construction.

The rest of this paper is arranged as follows: Section 2 describes the positioning process of tunnel equipment. The improved FAST (accelerated segmented test feature) corner extraction method combined with the improved BRIEF descriptor method is used to match the environmental feature points. RANSAC algorithm is used to remove the wrong matching points. The pole geometry method and triangulation depth recovery method are used to solve the pose of the tunneling robot. In Section 3, monocular visual feature extraction and three-dimensional reconstruction of roadway are verified through experiments, and then pose measurement is realized. In Section 4, this paper and the experiment are summarized.

2. Methods

2.1. Positioning Process of Tunneling Equipment Based on Monocular Vision

The underground tunneling face of a coal mine has the characteristics of low illumination, long and narrow channels and numerous pieces of equipment, which poses challenges to the precise positioning of tunneling equipment. Therefore, with the coal mine roadway environment as the feature, the tunneling equipment positioning method with uniocular vision as the core was constructed. The monocular camera and the tunneling equipment were firmly connected. By collecting the environmental images of the underground coal mine roadway, introducing the in-camera parameters with calibration in advance, the image preprocessing method was used to realize image enhancement, filtering and distortion correction, and then the ORB feature extraction method was used to extract the feature points, and the RANSAC algorithm was used to eliminate the mismatched points [20]. Finally, the pose estimation between adjacent frames was realized according to the principle of polar geometry constraints. The three-dimensional spatial coordinates of feature points were calculated by the triangulation depth calculation method. Then, the position information of the tunneling equipment was calculated in real time. By solving the pose of the tunneling robot, the 3D point-cloud registration between image sequences was realized, and sparse 3D space points were recovered from the collected images. Positioning process of coal mine heading equipment based on uniocular vision is shown in Figure 1.

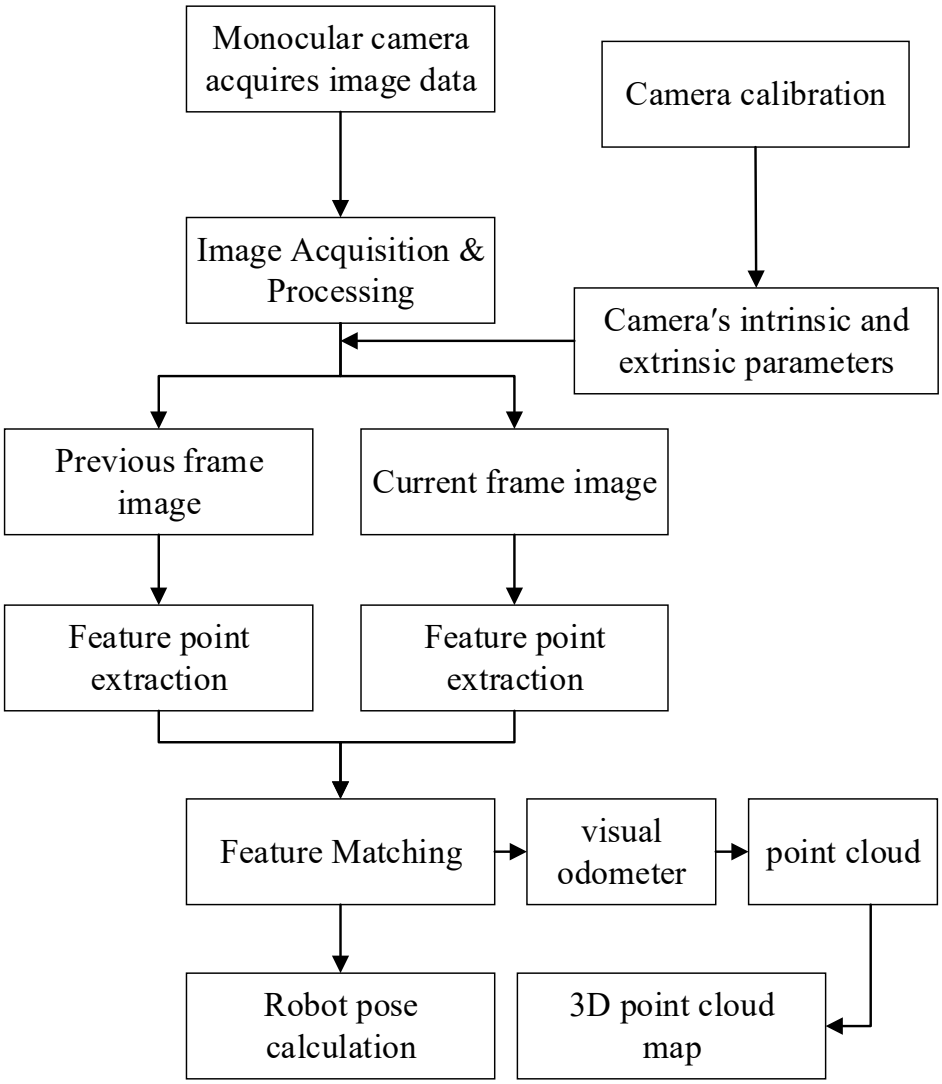

Figure 1. Positioning process of coal mine heading equipment based on uniocular vision.

2.2. Image Feature Extraction and Matching Method in Complex Background

2.2.1. Image Preprocessing

Due to the unstructured and complex environment of an underground coal mine and the hardware differences between the monocular cameras themselves, it was easy to produce image noise and image distortion in the image acquisition process. Further image processing was needed to realize the image feature extraction and calculation, so the image acquisition and pre-processing link was particularly critical. Image preprocessing mainly includes camera de-distortion and an image filtering process. In the process of image acquisition, a pinhole camera model will produce camera distortion, specifically divided into radial distortion, eccentric distortion and thin prism distortion. By establishing a mathematical model, camera lens distortion implements a camera lens distortion correction, while the interference of random signals around the image will cause noise on the image, so the filtering effect has a direct impact on the effectiveness and reliability of the visual measurement system to ensure the extraction and calculation of spot features.

Distortion Correction

The flaw on the camera lens was the main cause of the radial distortion. Due to its particularity, the closer to the outside, the more diffuse or concentrated inside. Let the tangential distortion be δ_{xr}, δ_{yr}, thus the mathematical model is shown in Equation (1):

$$\begin{cases} \delta_{xr} = x(k_1(x+y)^2 + k_2(x+y)^4 + \cdots \\ \delta_{yr} = y(k_1(x+y)^2 + k_2(x+y)^4 + \cdots \end{cases} \tag{1}$$

where the coefficient $k_1, k_2 \ldots$ represents the radial distortion coefficient.

Both radial and tangential distortion were included, mainly because the camera's optical center and geometric center did not coincide. Let the eccentric distortion be A, thus the mathematical model is shown in Equation (2):

$$\begin{cases} \delta_{xd} = p_1(3x^2 + y^2) + 2p_2xy + \cdots \\ \delta_{yd} = 2p_1xy + p_2(3x^2 + y^2) + \cdots \end{cases} \tag{2}$$

An effective camera model with all distortion types could not be established, so we considered integrating nonlinear models in the camera calibration process. Too many nonlinear parameters will cause instability in the solution, so at most, two distortion factors were considered, and the improvement method of fusion radial distortion and eccentric distortion was adopted.

For the above lens distortion analysis, the correction model combining the two models is:

$$\begin{cases} \delta_x(x,y) = \delta_{xr} + \delta_{xd} \\ \delta_y(x,y) = \delta_{yr} + \delta_{yd} \end{cases} \tag{3}$$

Figure 2 shows the image before and after the distortion correction. It can be seen that the monocular camera had a more obvious pillow-shaped distortion, which could better recover the real scene after the correction.

(**a**) (**b**)

Figure 2. Aberration correction contrast. (**a**) Before distortion correction; (**b**) after distortion correction.

Gaussian Filtering

Gaussian filtering builds mathematical models to transform the energy of the image data and filter out the low-energy part of the noise. The 3 × 3 Gaussian filter template is shown in Equation (4):

$$\frac{1}{16} \times \begin{bmatrix} 1\,2\,1 \\ 2\,4\,2 \\ 1\,2\,1 \end{bmatrix} \tag{4}$$

Based on the selected kernel size k and the size of the weight coefficient, the weighted average is calculated by the method of each pixel multiplied by the associated weight with the weighted average instead of the current pixel value. The weight coefficients are obtained from a one-dimensional Gaussian function.

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{5}$$

In Equation (5), σ represents the standard deviation and x represents the pixels in the image. For 2D images, the 2D Gaussian function is established as:

$$G(\mu, v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\mu^2 + v^2)}{2\sigma^2}\right) \tag{6}$$

In Equation (6), A represents the standard deviation, B represents the pixel abscissa in the image and C represents the pixel ordinate in the image.

By making $\sigma = 5$ and the Gaussian filter core $k = 3$, the acquired image can be filtered to compare the filtered image with the pre-filtered image. The comparison results are shown in Figure 3.

(**a**) (**b**)

Figure 3. Before and after Gaussian filtering comparison. (**a**) Before Gaussian filtering; (**b**) after Gaussian filtering.

2.2.2. Feature Extraction

Due to the difficulty of estimating the camera motion directly from the matrix changes, selecting the appropriate feature extraction algorithm was crucial for image matching and the pose solution calculation. The feature points included corners, blocks and edges, etc. The ORB feature extraction [21] belonged to diagonal feature extraction, and the feature point description was performed by improving the FAST feature extraction and the BRIEF descriptor algorithm. The experiments showed that ORB performed better than the BRIEF and SURF methods. By analyzing the feature extraction method, the ORB feature extraction principle could realize fast feature extraction and description, so the ORB feature extraction method was adopted.

Improved FAST Angle-Point Extraction

A pixel point p was first selected in the image and its pixel brightness was H_p. The judgment threshold for corner points was further set at 20% of H_p. Then, the pixel p was selected as the center point according to the setting threshold, and 16 pixels were selected on a neighborhood circle with a radius of 3 pixels.

According to the above selection principle, the feature point was judged. If the brightness T of N points on the neighborhood circle met $T > H_p + 20\%H_p$ or $T < H_p + 20\%H_p$, the p point could be regarded as the corner point, while it was not the characteristic corner point. The selection was cyclic according to the above selection principle, and the image was traversed until the FAST angles that met the requirements were extracted.

Further, the improved FAST method in ORB achieved scale invariance by constructing image pyramids and detecting corners at each layer of the pyramid and realized its rotation invariance with the help of the grayscale centroid-of-mass method.

The specific implementation process of the grayscale centroid method is as follows:

The moment of defining an image block in an image block B is:

$$m_{pq} = \sum_{x,y \in B} x^p y^q I(x,y), p, q \in \{0,1\} \tag{7}$$

In the formula, x, y represents the pixel coordinates, and I (x, y) represents the gray value of the pixel coordinates.

The image block center of mass is further obtained by moment calculation:

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \tag{8}$$

The direction vector $\overrightarrow{OC}$ is obtained from the geometric center and center of mass C of the image block B and can be used to represent the feature point direction:

$$\theta = \arctan\left(\frac{m_{01}}{m_{10}} \right) \tag{9}$$

The above improved FAST algorithm can describe the FAST, which improves the robustness of the algorithm.

Improving on the BRIEF Descriptor

After obtaining the feature points, N pairs of points are selected around the key points in the form of BRIEF. The key point P is selected as the center, and the radius is d, and n points are randomly selected in a certain neighborhood range for comparison. The selection conditions are as follows:

Suppose A, B, H_a, H_b are the gray scale of A, B,

$$F(P(A,B)) = \begin{cases} 0 \; H_a > H_b \\ 1 \; H_a \leq H_b \end{cases} \tag{10}$$

The n-point pair (x_i, y_i) is defined as the $2 \times n$ matrix S according to the above selection conditions:

$$S = \begin{bmatrix} x_1 & x_2 & \cdots x_n \\ y_1 & y_2 & \cdots y_n \end{bmatrix} \tag{11}$$

The obtained θ pair S matrix rotation is used to obtain:

$$S_\theta = R_\theta S \tag{12}$$

In the formula, R_θ is the rotation matrix of the angle θ.

Thus, ORB feature point extraction and descriptor establishment are realized. Through the established descriptor of feature extraction, the robot position acquisition was then realized by the position solution algorithm. However, because the perspective changed greatly or had similar regions, the mismatching situation was prone in the process of feature matching, so the mismatching situation needed to be eliminated.

2.2.3. Mismatch Elimination of the RANSAC Algorithm

The feature matching process is a particularly critical step in the visual SLAM algorithm, but there is a mismatching situation, so the RANSAC algorithm was used to eliminate the mismatching. The specific steps of the RANSAC algorithm are as follows:

1. Data from s samples are randomly drawn from the dataset, multiple models are fit and the most suitable unistress matrix $H_{3\times3}$ is calculated according to the principle of the most matrix data points.

$$s \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} h_{11}\, h_{12}\, h_{13} \\ h_{21}\, h_{22}\, h_{32} \\ h_{31}\, h_{32}\, h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{13}$$

As in Equation (13), make h_{33} = 1 and normalize the matrix $H_{3\times3}$. Since there are eight unknowns in the matrix, it should be solved by at least four sets of matching points in the extracted image.

2. Calculate the projection errors of all the data models H in the dataset and set the threshold value. If the error is less than the threshold, it meets the requirements, and it is added to the inner point set I, and the data greater than the threshold value are removed. The projection error function is:

$$\sum_{i=1}^{n} \left(x_i' - \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}} \right)^2 + \left(y_i' - \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}} \right)^2 \tag{14}$$

3. If the number of inner points is insufficient, it is necessary to reselect the 4 pairs of matching points in the image and calculate the model H until the number of inner point sets exceeds the selected optimal inner point set threshold;
4. If the number of elements of the inner point set I is greater than the optimal inner point set condition, the number of inner points is recorded, and the iteration number k is updated;
5. If the number of iterations is greater than the maximum number K, the iteration ends; otherwise resume steps 1−5. The maximum number of iterations, K, is equal to:

$$K = \frac{\log(1 - P)}{\log 1 - w^m} \tag{15}$$

In Equation (15), the confidence degree of P is made to be 0.0995, w the "inner point" ratio and m the minimum number of samples required to calculate the model.

2.3. Pose Position Solution Based on Depth Recovery

2.3.1. Position and Pose Measurement of Roadheader Robot

According to the extraction matching point for the camera pose estimation, as shown in Figure 4 below, good feature points were paired from two pictures according to the feature points in the two-dimensional image correspondence, the camera image between the frame pose was restored, and the camera posture solution was preliminarily completed in order to realize the uniocular visual pose measurement based on the feature point depth.

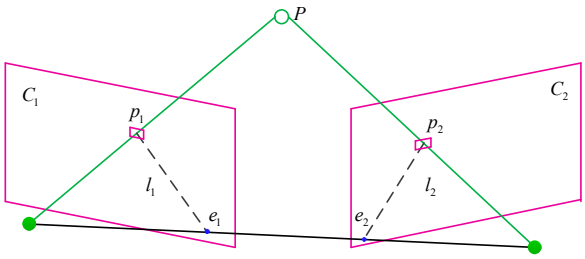

Figure 4. Constraints on the polar geometry.

Assuming the two frames of images are C_1 and C_2 during the camera motion, the motion from images C_1 to C_2 is R t, and the two cameras are centered at O_1 and O_2, respectively. p_1 and p_1 are the corresponding feature points obtained by matching the P points between the two images, respectively, and O_1p_1 and O_2p_2 are connected to obtain the rays $\overrightarrow{O_1p_1}$ and $\overrightarrow{O_2p_2}$ intersecting at the point P.

The plane composed of the three points of O_1, O_2 and P is called the polar plane, and the intersection between O_1, O_2 and the image plane C_1, C_2 is e_1 and e_2, respectively. e_1 and e_2 are the poles, and O_1O_2 is the baseline.

The intersection lines l_1 and l_2 between the pole plane and the two image planes C_1 and C_2 are the pole line.

The coordinates of this spatial point P can be expressed as:

$$P = [x, y, z]^T \tag{16}$$

According to the camera imaging model, the two-pixel $p_1 \cdot p_2$ locations are as follows:

$$s_1 p_1 = k_2 p, s_2 p_2 = k_2(RP + t) \tag{17}$$

In Equation (17), k_2 is the in-camera parameter, R, and t is the rotation matrix, and the translation matrix of the camera motion multiplied by the non-zero constant can be reduced to:

$$p_1 = k_2 p, p_2 = k(RP + t) \tag{18}$$

Let the pixel point normalized plane coordinate $x_1 = k_2^{-1} p_1, x_2 = k_2^{-1} p_2$, bring x_1, x_2 into Formula (18), and multiply by $x_2^T t^\wedge$ on both sides:

$$x_2^T t^\wedge x_2 = x_2^T t^\wedge R x_1 \tag{19}$$

In Equation (19), the $t^\wedge x_2$ vector is perpendicular to t and x_2, so it is 0, then:

$$x_2^T t^\wedge R x_1 = 0 \tag{20}$$

Bring p_1 and p_2 in again with:

$$p_2^T k_2^{-T} t^\wedge R k_2^{-1} p_1 = 0 \tag{21}$$

In the formula, p_1, p_2 are the corresponding feature points obtained by matching P points between two images, and k_2 is the in-camera parameter.

The above Equation (20) and Equation (21) are collectively referred to as the polar constraint, indicating the geometric meaning of the O_1, P and O_2 three common planes, and by including the rotation matrix R and translation matrix t, the formula can be obtained:

$$x_2^T E x_1 = p_2^T F p_1 = 0 \tag{22}$$

$E = t\hat{}R$ is the essential matrix and A is the basic matrix, so the camera pose solution calculation can be simplified as follows: solve the essential matrix E or the basic matrix F according to the pixel position of the pairing point, and then solve the camera pose R and t combined with the solved E or F substitution Formula (22).

According to the principle of monocular visual measurement analysis, one cannot obtain pixel depth information through only one single picture, so this needs to be conducted by triangulation. According to the frame between postures, to extract the roadway characteristic point depth estimation, triangulation between two perspective angles and the realization of the point distance calculation is required, thus the measurement principle is as shown in Figure 5:

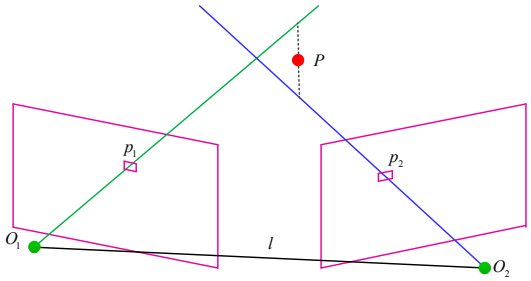

Figure 5. Schematic representation of the triangulation depth recovery.

According to the two frames in Figure 5, the two images are C_1 and C_2, with the left being C_1 for the previous moment and the right being moment C_2, p_1 and p_2, respectively, show the P point in the pixel coordinates in the previous and later point, and O_1 and O_2 are the center of the camera, and the two-image camera posture transformation matrix R and t, and the calibration in the camera parameters for k_2 are obtained according to the visual odometer. According to the definition of polar geometry, the two points p_1 and p_2 of the normalized plane can satisfy the following relationship:

$$d_1 p_1 = d_2 R p_2 + t \tag{23}$$

where d_1 and d_2 represent the camera depth value of the point P at the previous moment and the second moment, respectively. Since R and t are solved according to the visual odometer, Equation (23) is written in a matrix form:

$$\begin{bmatrix} -R \cdot p_1 & p_2 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = t \tag{24}$$

Further, Formula (24) is written as a system of linear equations such as:

$$Ax = b \tag{25}$$

Among $A = [-R \cdot p_1\ p_2]$, $x = [d_1\ d_2]^T$, $b = t$.

The optimization function can be built to solve the depth:

$$f(x) = \|Ax - b\| \tag{26}$$

According to the above formula, the x result can be obtained as follows:

$$x = \left(A^T A\right)^{-1} A^T b \tag{27}$$

The matrix $A^T A$ is a non-singular matrix, and the feature point depth solution of R and t implemented for the extracted roadway images is obtained by the combination of the visual odometer calculation. In addition, when A is a singular matrix, the LM algorithm is used to transform A into a non-singular matrix before solving it. After the depth value solution is realized, the tunneling robot pose solution based on the depth recovery can be realized through the 3D–2D method.

2.3.2. Position and Pose Measurement of Roadheader Robot

The monocular visual odometer solves the camera pose by matching the correspondence between the features, namely, the rotation matrix R and the translation matrix t. The general solution methods are 2D–2D as well as 2D–3D. Since the spatial coordinates of the feature points are found according to the method of triangulation depth, the 3D–2D method, namely, the PNP algorithm, was used to use the matching relationship between the recovered feature points' 3D-point coordinates and the image feature points to optimize and minimize the camera attitude according to the reprojection error.

The spatial coordinates of the same feature point X_i^k and X_i^{k-1} in the two frame pictures at the ck and ck-1 moment camera frame where the rotation matrix is R and the flat matrix is t. According to the motion relationship of different frame images, the following equation is established:

$$X_i^k = R * X_i^{k-1} + t \tag{28}$$

For the feature points recovered by triangulation, X_i^k is normalized, where $\overline{X}_i^k$ is the normalized coordinate of X_i^k and z_i^k is the modular length of X_i^k in the z-axis direction, and thus this is obtained according to the above formula:

$$X_i^k = R * X_i^{k-1} + t \tag{29}$$

The transformation matrices R and t are decomposed into three row vectors, R_h and t_h, where $h \in \{1, 2, 3\}$, and this is brought into the upper formula, eliminating z_i^k by the elimination method and obtaining the following system of equations:

$$\begin{cases} \left(R_1 - \overline{X}_i^k R_1\right) X_i^{k-1} + t_1 - \overline{X}_i^k t_3 = 0 \\ \left(R_2 - \overline{X}_i^k R_3\right) X_i^{k-1} + t_1 - \overline{X}_i^k t_3 = 0 \end{cases} \tag{30}$$

According to the above Equation (30), through at least six pairs of matching points, the linear solution of the pose transformation matrix T can be realized, that is, the direct linear transformation method can be realized.

3. Results and Discussion

For the roadway environment characteristics texture, the tunneling robot tunnel environment perception is difficult. Through the study of visual feature extraction, using the ORB feature point extraction method of 3D sparse point cloud map construction, the 3D reconstruction method based on single eye visual SLAM was verified. For the corridor environment simulation roadway, the experimental verification platform was set to evaluate the effectiveness and real-time performance of the method.

Based on the uniocular vision tunneling robot 3D reconstruction experiment, in the corridor control robot for continuous motion, using the monocular visual image feature extraction matching method and 3D reconstruction method of the tunneling robot in the running image processing and solution, the function points, by calculating the global coordinate system of the 3D point cloud and the results in the visual interface, the uniocular

vision roadway 3D reconstruction effect was completed and the sparse 3D point map was evaluated. In order to verify the reliability of the monocular visual feature extraction and 3D reconstruction, the 3D reconstruction experiment of the tunneling robot was carried out in a corridor, and the picture information of the roadway was collected immediately during the movement process, and the feature point matching and point cloud registration were completed. As shown in Figure 6, some pictures of feature points in the roadway were extracted during the robot's movement.

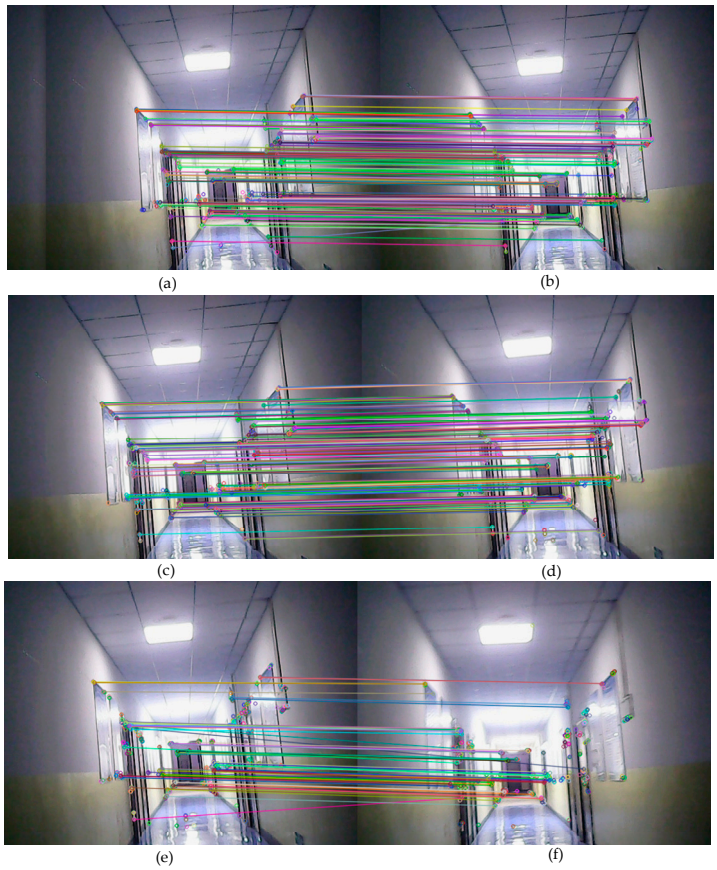

Figure 6. The matching corridor and corridor characteristic points. (**a**) First frame picture, (**b**) Second frame picture, (**c**) Third frame picture, (**d**) Fourth frame picture, (**e**) Fifth frame picture, (**f**) Sixth frame picture.

Figure 6 shows the matching process of four adjacent frames. It can be seen that in the dark corridor environment, the roadway feature extraction effect obtained by the ORB feature extraction method is represented in color in the figure. After mismatching and elimination, fewer feature points were extracted to meet the requirements of the ORB feature extraction. Based on the feature point extraction, the feature point matching between each picture frame was realized, and the average matching time of each of the two frames was 179 ms, which met the real-time requirements of feature extraction and matching.

Figure 7 shows the process of the 3D sparse-point cloud-map-mapping process for the monocular visual SLAM realized with the help of the ORB feature extraction.

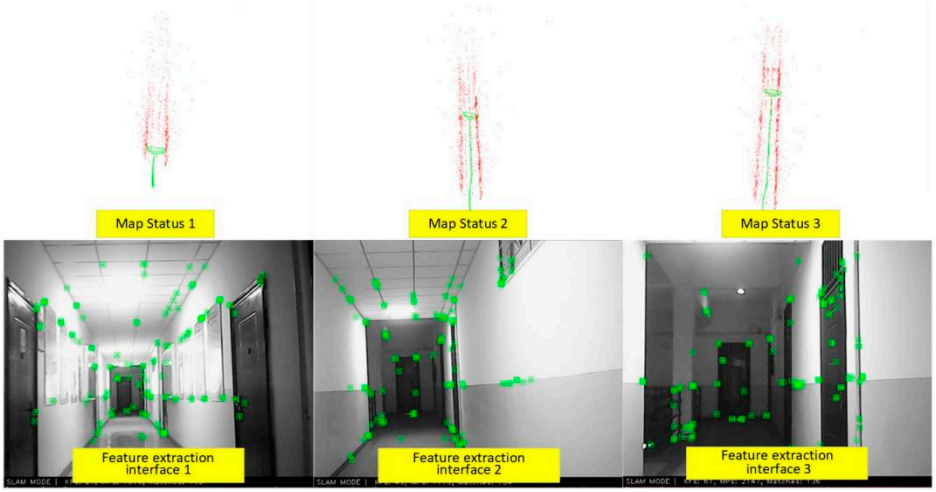

Figure 7. Three-dimensional sparse-point cloud map construction process.

The green part of the feature extraction interface is the extracted and matched feature points of the corridor and corridor, the green part of the map state interface is the track of the robot and the red points are the three-dimensional sparse roadway points under construction. It can be seen that the visual feature points extracted on each frame picture were relatively rich, and the robot's movement track was relatively stable with good stability, which is not easy to obtain in feature loss.

Based on the above feature extraction and matching process, the optimized corridor 3D sparse-point cloud map was obtained. Figure 8a shows the top view of the corridor, and Figure 8b shows the construction of the corridor. It can be seen that the trend was roughly consistent with the direction of the corridor, and the roadway was constructed using the uniocular visual SLAM method based on ORB. Due to the small amount of feature extraction and mismatching elimination, the sparse 3D reconstruction of the tunnel was thus realized.

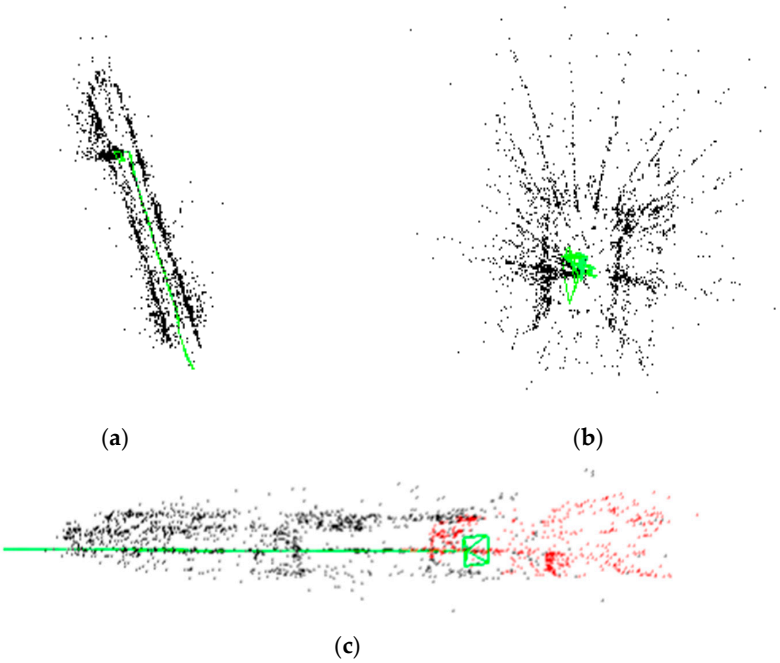

Figure 8. Three-dimensional sparse reconstruction. (**a**) Top view of the corridor; (**b**) face view of the corridor; (**c**) side view of the corridor.

Figure 8a–c shows the constructed three views of the sparse-point cloud for the three-dimensional roadway realized by monocular vision, respectively. The black point cloud shows the roadway feature points extracted and registered, the green line is the robot track, the green box is the real-time pose of the robot and the red points are the feature points under construction. The global map of the roadway composed of sparse point cloud could be found by overlooking, from a side view and face view, which was roughly in line with the roadway trend, and the three-dimensional roadway reconstruction based on the sparse-point cloud was initially realized.

Using the monocular visual SLAM method based on ORB feature extraction, the position measurement accuracy was within ±60 mm, and the angle measurement error was within ±1.3°. The robot position and the boundary perception of the point cloud in the global coordinate system in the system was obtained from the real-time acquisition of the ORB-based feature image reconstruction time, which was 179 ms.

4. Conclusions

According to the process of monocular visual SLAM, this paper introduced in detail many processes, such as image preprocessing, feature extraction, feature matching and mismatching elimination, pose solution calculation and sparse 3D point-cloud reconstruction. This paper also simulated the experimental verification platform of the downhole environment and set up a platform for the verification of the monocular visual positioning in the laboratory.

The experimental results showed that the position measurement accuracy of the unioc-ular visual roadheader positioning method was within ±60 mm and the angle measurement error was within ±1.3°, and that it could realize the accurate registration of the point cloud under the global coordinate system. The required time of the whole monocular visual positioning was only 179 ms, so it had good real-time performance. However, there was still a certain gap between the environmental simulation and the real excavation face in the experimental verification stage. Coal walls on both sides of an underground coal mine roadway are not very smooth, and many large pieces of equipment would also have a certain influence on the field of view of the monocular camera, so this will be continuously improved in further research.

Author Contributions: Conceptualization, R.Y., X.F., C.H. and X.Z.; data curation, R.Y.; methodology, C.Z. and W.Y.; project administration, R.Y., C.H. and X.Z.; software, C.Z., W.Y. and J.W.; supervision, X.Z.; validation, R.Y., C.H. and X.Y.; visualization, Q.M.; writing—original draft, R.Y., C.H., C.Z.; writing—review & editing, X.Z. and Q.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Founds of China (Grant No. 51974228), the National Natural Science Founds of China (Grant No.52104166).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data and code used or analyzed in this study are available from the corresponding author on reasonable request.

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

References

1. Wang, G.; Wang, H.; Ren, H.; Zhao, G.; Pang, Y.; Du, Y.; Zhang, J.; Hou, G. 2025 scenarios and development path of intelligent coal mine. *J. China Coal Soc.* **2018**, *43*, 295–305. [CrossRef]
2. Jia, L.; Yu, Y.; Li, Z.-P.; Qin, S.-N.; Guo, J.-R.; Zhang, Y.-Q.; Wang, J.-C.; Zhang, J.-C.; Fan, B.-G.; Jin, Y. Study on the Hg0 removal characteristics and synergistic mechanism of iron-based modified biochar doped with multiple metals. *Bioresour Technol.* **2021**, *332*, 125086. [CrossRef] [PubMed]
3. Wang, G.; Ren, H.; Zhao, G.; Zhang, D.; Wen, Z.; Meng, L.; Gong, S. Research and practice on intelligent coal mine construction(primary stage). *Coal Sci. Technol.* **2019**, *47*, 1–36. [CrossRef]

4. Wang, S.Y.; Mu, J.; Wu, M. Spatial pose kinematics model and simulation of boom-type roadheader. *J. China Coal Soc.* **2015**, *40*, 2617–2622. [CrossRef]
5. Zhang, X.H.; Liu, B.X.; Zhang, C. Roadheader positioning method combining total station and strapdown inertial navigation system. *J. Mine Autom.* **2020**, *46*, 1–7. [CrossRef]
6. Fu, S.C.; Li, Y.M.; Zong, K.; Zhang, M.J.; Wu, M. Accuracy analysis of UWB pose detection system for roadheader. *Chin. J. Sci. Instrum.* **2017**, *38*, 1978–1987. [CrossRef]
7. Liu, C.; Fu, S.C.; Cheng, L.; Liu, D.; Sheng, Y.; Wu, M. Pose detection method based on hybrid algorithm of TSOA positioning principle for roadheader. *J. China Coal Soc.* **2019**, *44*, 1255–1264. [CrossRef]
8. Reid, P.B.; Dunn, M.T.; Reid, D.C.; Ralston, J.C. Real-World Automation: New Capabilities for Underground Longwall Mining. In Proceedings of the Australasian Conference on Robotics and Automation 2010, Queensland University of Technology, Brisbane, Australia, 1 December 2010; pp. 1–8.
9. Amo, I.A.D.; Letamendia, A.; Diaux, J. Nuevas resistencias comunicativas: La rebelión de los ACARP. *Rev. Lat. De Comun. Soc.* **2014**, *69*, 307–329.
10. Hlophe, K.; du Plessis, F. Implementation of an autonomous un-derground localization system. In Proceedings of the 2013 6th Robotics and Mechatronics Conference (RobMech), Durban, South Africa, 30–31 October 2013; pp. 1–34.
11. Du, Y.X.; Liu, T.; Tong, M.M.; Dong, H.B.; Zhou, L.L. Pose measurement system of boom-type roadheader based on machine vision. *J. China Coal Soc.* **2016**, *41*, 2897–2906. [CrossRef]
12. Yang, W.; Zhang, X.; Ma, H.; Zhang, G. *Laser Beams-Based Localization Methods for Boom-Type Roadheader Using Underground Camera Non-Uniform Blur Model*; IEEE Access: New York, NY, USA, 2020; Volume 8, pp. 190327–190341. [CrossRef]
13. Xu, X.S.; Zhang, X.H.; Mao, Q.H.; Zheng, J.K.; Wang, M. Localization and orientation method of roadheader robot based on binocular vision. *J. Xi'an Univ. Sci. Technol.* **2020**, *40*, 781–789. [CrossRef]
14. Kataoka, R.; Suzuki, R.; Ji, Y.; Fujii, H.; Kono, H.; Umeda, K. ICP-based SLAM Using Li DAR Intensity and Near-infrared Data. In Proceedings of the 2021 IEEE/SICE International Symposium on System Integration (SII), Iwaki, Japan, 11–14 January 2021; pp. 100–104.
15. Senin, N.; Colosimo, B.M.; Pacella, M. Point set augmentation through fitting for enhanced ICP registration of point clouds in multisensor coordinate metrology. *Robot. Comput.-Integr. Manuf. Robot.* **2013**, *29*, 39–52. [CrossRef]
16. Bouaziz, S.; Tagliasacchi, A.; Pauly, M. Sparse Iterative Closest Point. *Comput. Graph. Forum* **2013**, *32*, 113–123. [CrossRef]
17. He, M.; Zhu, C.; Huang, Q.; Ren, B.; Liu, J. A review of monocular visual odometry. *Vis. Comput.* **2019**, *36*, 1053–1065. [CrossRef]
18. Guan, X.J.; Wang, X.L. Review of Vision-Based Navigation Technique. *Aero Weapon.* **2014**, *5*, 3–8+14. [CrossRef]
19. Geng, C.; Yang, J.; Lin, J.; Yu, T.; Shi, K. An Improved ORB Feature Extraction Algorithm. *J. Phys. Conf. Ser.* **2020**, *1616*, 012026. [CrossRef]
20. Wu, M. Research on optimization of image fast feature point matching algorithm. *EURASIP J. Image Video Process.* **2018**, *2018*, 106. [CrossRef]
21. Wang, Z.H. Research on Key Technology of Target Location Based on Image Matching. Master's Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2020. [CrossRef]