

Research Article

Study on Optimization Method of Visual Odometry Based on Feature Matching

Haili Song ¹, Zengxu Zhao ¹, Bin Ren,¹ Zhanpu Xue ², Junliang Li,¹ and Hao Zhang²

¹School of Mechanical Engineering, Shijiazhuang Tiedao University, Shijiazhuang 050043, China

²School of Mechanical Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China

Correspondence should be addressed to Zengxu Zhao; zhaozx_2015@stdu.edu.cn

Received 29 August 2022; Revised 27 October 2022; Accepted 7 November 2022; Published 17 November 2022

Academic Editor: Hao Gao

Copyright © 2022 Haili Song et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The mismatching of image features affects the calculation of the fundamental matrix and then leads to poor estimation accuracy of SLAM visual odometry. Aiming at the above problems, a visual odometry optimization method based on feature matching is proposed. Firstly, the initial matching set is roughly filtered by the minimum distance threshold method, and then, the relative transformation relationship between images is calculated by the RANSAC algorithm. If it conforms to the transformation relationship, it is an interior point. The iteration result with most interior points is the correct matching result. Then, the homography transformation between images is calculated, and the fundamental matrix is calculated by it. The interior points are determined by epipolar geometric constraints, and the fundamental matrix with the most interior points is obtained. Finally, the effects of the visual odometry optimization algorithm are verified by the TUM data set from two aspects: feature matching and fundamental matrix calculation. The experimental results show that an improved feature matching algorithm can effectively remove mismatched feature points while improving the operation efficiency. At the same time, the accuracy of feature point matching is increased by 15.8%. The fundamental matrix estimation algorithm not only improves the calculation accuracy of the fundamental matrix but also increases the interior point rate by 11.9%. A theoretical basis for improving the accuracy estimation of visual odometry will be provided.

1. Introduction

VO is an important part of SLAM, which mainly includes monocular VO, stereoVO, and 3D VO [1]. VO obtains image information through the camera and then calculates pose based on the constraint relationship of image feature points between frames, and to estimate object motion [2], it is mainly used in the fields of autonomous navigation, 3D reconstruction, and autonomous driving of mobile robots. VO can provide key pose information for robot localization, mapping, and navigation. However, when the sensor is affected by factors such as environment, illumination, and moving objects [3], the feature vectors of image feature points will change greatly, resulting in mismatches in the matching candidate set, thus affecting the robot pose estimation. Therefore, how to eliminate mismatched feature points is a key issue in VO and an important prerequisite for

pose estimation. Mismatch elimination methods can be divided into three categories: resampling methods [4], graph theory-based, and block theory-based methods [5]. Among them, the related resampling techniques include epipolar constraints [6] and homography constraints [7]. The RANSAC algorithm used in [8] may lead to mismatches due to its randomness and hypotheses, and scholars have studied it. Ma et al. proposed a matching optimization algorithm called pixel shift clustering RANSAC, and it can eliminate the residual mismatches in matching results [9]. Zhai et al. proposed an improvement of RANSAC to an iterative form, which can reduce the blind sample process and ensure fast convergence speed [10]. Liu et al. used optimized neighborhood topology consistency and RANSAC to determine interior points, which improves matching robustness [11]. Meng et al. proposed a robust feature point matching algorithm named spatial order constraints bilateral-neighbor

vote [12], and it can remove outliers for a set of matches between two images and have strong robustness. In addition, some scholars reduce the mismatch by improving the robustness of feature points. The robustness is improved by increasing the number of feature matches in extreme scenes by image denoising [13]. Emam et al. used a Harris corner detector with nonmaximum suppression to detect keypoints in missing regions to evenly distribute detected keypoints [14], thereby improving the matching rate. In order to obtain better feature point quality and matching efficiency, Xie et al. proposed the idea of combining adaptive histogram equalization with the ORB algorithm [15]. Fan et al. proposed visualized local structure generation-Siamese attention network (VLSG-SANet), which eliminates mismatches through dynamic visual similarity evaluation [16].

The fundamental matrix [17] is estimated after establishing the feature point matching relationship between images. Correctly calculating the fundamental matrix is beneficial to accurately estimate the camera pose [18]. Fundamental matrix estimation algorithms can be divided into linear methods [19], iterative methods, and robust methods. As a linear method, the 8-point method [20] has poor adaptability to complex situations although its calculation is simple. Bugarin et al. proposed a single-step method that can solve both steps of the eight-point algorithm [21], and it had fewer iterations. The iterative algorithm [22] estimates the fundamental matrix by minimizing the objective function. Geometric error minimization [23], point to epipolar distance minimization [24], and epipolar geometric estimation [25] can improve the calculation accuracy of the fundamental matrix. The linear method and iterative method are suitable for data sets without mismatching points. However, robust estimation methods can still obtain higher estimation accuracy when matching point sets generate outliers, such as RANSAC, least median squares method, M estimation method [26–28]. Tatar proposed a soft decision optimization method to estimate the fundamental matrix [29], and a soft decision objective function is developed to remove outliers in the candidate correspondence set. In order to reduce the computational complexity, Ke et al. proposed to combine the singular value decomposition method with the interleaving parameterization method [22]. The robust algorithm can also be combined with a neural network. Shao et al. proposed a semantic filter based on a faster region-based convolutional neural network to address the outlier problem in RANSAC-based fundamental matrix computation [26]. Yang et al. used the improved convolutional block attention module to ensure the estimation of an accurate fundamental matrix which is rank-2 with 7 degrees of freedom and scale invariance [30]. Although the robust algorithm has a certain anti-interference performance, when the outlier ratio exceeds a certain range, the solution accuracy will be affected. Therefore, how to remove mismatched feature points is the focus of research.

Aiming at the problem that environmental interference causes the mismatch of image features, affects the calculation accuracy of the basic matrix, and leads to the low accuracy of VO estimation, a VO method based on feature matching is innovatively analyzed. Firstly, the minimum distance

threshold method is used to roughly filter the initial matching set, and then, RANSAC is used to calculate the model Q . It is a correct match if it conforms to the model, and the mismatched feature points can be filtered out. Secondly, the homography transformation between images is calculated, and outliers are filtered through epipolar geometric constraints. The fundamental matrix can be calculated by the homography matrix. The experimental results show that the algorithm in this paper improves the correct matching rate of feature points and the calculation accuracy of the basic matrix, which lays a foundation for improving the estimation accuracy of VO.

2. Methods

2.1. Basic Theory. The RANSAC algorithm calculates the best model through iteration. Use random data to calculate model parameters, and then estimate other points in the dataset based on the model. The error within the threshold is an interior point; otherwise, it is an exterior point. Repeat the above steps and save model parameters corresponding to the maximum number of interior points as a final result. The RANSAC algorithm can be used to eliminate false matches, and the principle of epipolar geometry is the basis of calculating the fundamental matrix. As shown in Figure 1, I_1 and I_2 represent the imaging planes of the previous frame and the current frame, respectively. O_1 and O_2 are the corresponding optical centers of the camera. L_1 and L_2 are epipolar lines. p_1 and p_2 are matching feature points. The extreme points are e_1 and e_2 .

When P is not in the space plane and feature points are correctly matched, the normalized plane coordinates and fundamental matrix F satisfy equation (1). If the feature points cannot fall on the epipolar line due to factors such as mismatching, calculate the distance from the feature points to their respective epipolar lines. When the distance is greater than the threshold, the point is considered to be an outlier.

$$\mathbf{p}_2^T \mathbf{F} \mathbf{p}_1 = 0. \quad (1)$$

If the feature points in the scene all fall on the same plane, motion estimation can be performed through homography. The homography matrix between two images can be expressed as (2). The homography transformation is the mapping of a coordinate point from one image plane to another.

$$\mathbf{H} = \mathbf{K}(\mathbf{R} + \mathbf{t}\mathbf{n}_d^T)\mathbf{K}^{-1}, \quad (2)$$

where \mathbf{n} is the unit normal vector, d is the distance from the coordinate origin to the plane, \mathbf{K} is the camera internal parameter matrix, \mathbf{R} and \mathbf{t} are rotation matrix and translation vector respectively, \mathbf{H} is the homography matrix, and \mathbf{p}_1 and \mathbf{p}_2 are homogeneous coordinates of p_1 and p_2 on the image.

2.2. VO Optimization Algorithm. In traditional visual SLAM, image information is collected by sensors and then input into the VO. Initial pose optimization is performed through

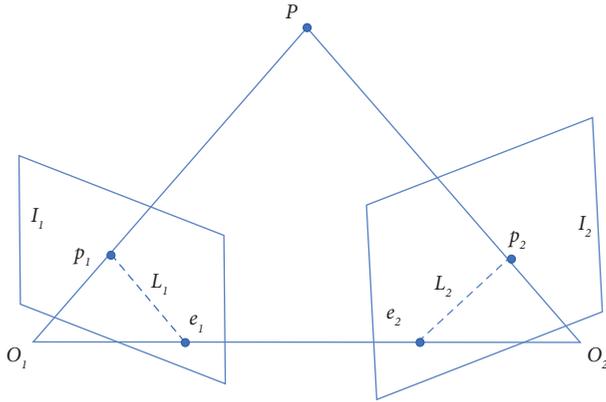


FIGURE 1: Epipolar geometry.

ORB feature matching, and camera motion is estimated based on adjacent image information. At the same time, closed-loop detection helps to correct the pose of the mobile robot, and the pose information is input to the back-end for optimization. Finally, a map is constructed to facilitate the positioning and navigation of the robot.

This paper optimizes two parts based on traditional visual SLAM: feature matching and basic matrix calculation. The optimization framework of VO based on visual SLAM is shown in Figure 2, and red part is VO optimization framework. Aiming at the problem of image feature mismatching, a feature-matching optimization algorithm is proposed. Firstly, the image features are extracted, and the feature points are roughly filtered out by the minimum distance method. Then, the RANSAC algorithm is used to further filter out the mismatched feature points. Aiming at the problem of poor fundamental matrix estimation accuracy, an optimized fundamental matrix estimation algorithm is proposed. Firstly, the homography transformation is calculated by the remaining feature points optimized by feature matching. Then, the fundamental matrix is calculated based on epipolar geometric constraints, so as to improve the accuracy and efficiency of fundamental matrix estimation.

The basic steps of VO include feature extraction, feature matching, coordinate transformation, and motion estimation. The research on the optimization algorithm of monocular VO consists of two parts: feature matching and basic matrix calculation. In terms of feature matching, the robust algorithm RANSAC has a certain antinoise performance. However, too high ratio of outliers will still lead to more mismatches, which will affect the subsequent calculation of the fundamental matrix. At the same time, RANSAC needs to set the threshold artificially when judging the internal and external points. Therefore, a feature-matching method based on the minimum threshold method is proposed. In terms of fundamental matrix calculation, RANSAC estimates the fundamental matrix by randomly selecting feature points as interior points. The accuracy depends on the ratio of interior points, which is easy to fall into the local optimal solution. As a result, the estimation accuracy of the visual odometer is low, which affects subsequent positioning and mapping.

Therefore, it is necessary to study the optimization calculation method of the fundamental matrix. To sum up, feature matching based on the minimum threshold method is used to reduce mismatched feature points. Then, the remaining point set is used to estimate the camera pose through the optimized fundamental matrix calculation method based on homography transformation. In this way, the initial pose optimization is realized, and the overall estimation performance of VO will be improved.

2.2.1. Feature Matching Based on Minimum Threshold.

The threshold scale setting of RANSAC is highly dependent on experience. In order to further improve the matching accuracy, an optimized feature matching method is proposed. The minimum threshold method [31] is used to roughly filter out the false matching and then used RANSAC to calculate the corresponding model of matching points. By verifying the correctness of the model, the mismatched feature points can be precisely filtered out. The motion model of transformation between images is shown in equation (3). The relative transformation matrix H has 8 degrees of freedom, which can be calculated and solved by 4 pairs of corresponding feature points.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \quad (3)$$

(x, y) is the feature point of $K-1$ frame image; (x', y') is the feature point corresponding to (x, y) of K frame image. The minimum number of iterations K of the RANSAC algorithm satisfies

$$K = \frac{\ln(1 - P_m)}{\ln(1 - \eta^m)}, \quad (4)$$

where m is the minimum data required to calculate the model parameters; P_m is the confidence level, indicating the probability that at least one of the selected data is an interior point. η represents the probability that the selected data is an outer point.

The minimum threshold method needs to select a feature point in the image and then perform distance testing with the feature points in adjacent images in turn. Finally, return the closest feature point. When matching pairs satisfy (5), the matching is correct. Otherwise, the matching is eliminated.

$$d_i < \alpha d_{\min}, \quad (5)$$

d_i is the distance of eigenvectors in the i matching pair; d_{\min} is the minimum distance of matching pairs in the initial matching set; α is the set threshold.

The algorithm achieves the effect of eliminating mismatches by setting different thresholds. The principle is simple, and the operation speed is fast. However, when the threshold is not set correctly, the correct matching is easy to be eliminated. It will also result in the consequence that the number of matching after elimination does not meet

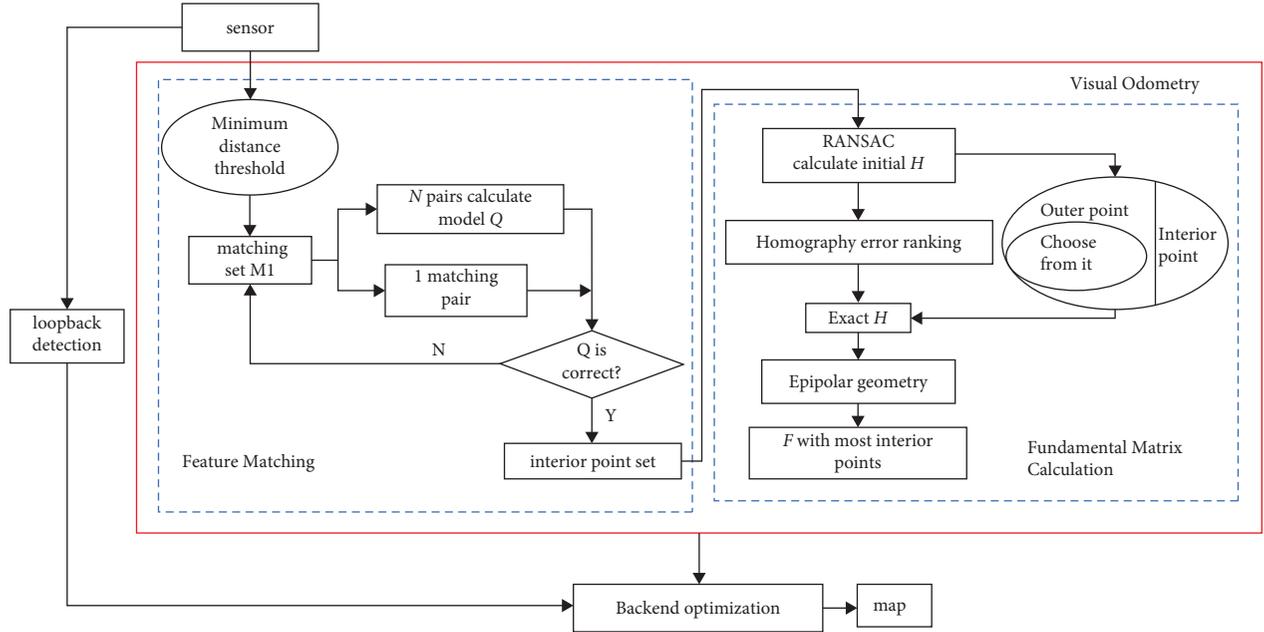


FIGURE 2: Optimization framework of VO based on visual SLAM.

application requirements. Therefore, RANSAC is combined with a minimum threshold. The mismatch elimination algorithm based on RANSAC is shown in Figure 2. Firstly, the initial matching set M is obtained by using the similarity of feature vectors. Then, the minimum distance threshold method with a threshold of eight is used to roughly eliminate the feature matching. When the matching pair of the initial matching set satisfies equation (5), it is the correct matching. Otherwise, the matching is eliminated, and then, the matching set M_1 is obtained. Randomly select $N + 1$ pairs of matching pairs from the matching set M_1 , where N pairs of matching pairs are used to calculate the relative transformation matrix to obtain the model Q (equation (3)). Then, the accuracy of model Q is verified by using the remaining one pair of matching. If the imported formula (2) does not hold, the model is not accurate. Then, eliminate the model and repeat the above steps. On the contrary, the model is accurate, and the number of inliers that fit the model is counted. The iterative result with the largest number of inner points is the correct matching result.

2.2.2. Fundamental Matrix Optimization Method. The feature matching based on the minimum threshold method can reduce mismatched feature points. However, when using a residual point set for camera pose estimation, the “randomness” and “easy to fall into the local optimal solution” of the RANSAC algorithm will lead to the poor calculation accuracy of the basic matrix. In order to improve the overall performance of VO, a fundamental matrix estimation algorithm combining homography transformation and epipolar geometric constraints is proposed. By improving the estimation accuracy of the basic matrix between different images, the optimization of camera pose estimation is realized, thereby improving the overall performance of the VO.

The projected coordinates of the point P located on the plane in the space on different images should satisfy both homography and epipolar geometric constraints. According to the homography transformation relationship, the formulas are as follows:

$$\begin{cases} (u_2 \ v_2 \ 1) \times \begin{bmatrix} h_7 & h_8 & h_9 \\ 0 & 0 & 0 \\ -h_1 & -h_2 & -h_3 \end{bmatrix} \times \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = 0, \\ (u_2 \ v_2 \ 1) \times \begin{bmatrix} 0 & 0 & 0 \\ h_7 & h_8 & h_9 \\ -h_4 & -h_5 & -h_6 \end{bmatrix} \times \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = 0, \end{cases} \quad (6)$$

where $P_1 = (u_1, v_1, 1)$ and $P_2 = (u_2, v_2, 1)$ are the homogeneous coordinate form of p_1 and p_2 on the image. After finishing, the formula is abbreviated as (7) where M_1 and M_2 are both matrices consisting of parameters in H .

$$\begin{cases} P_2^T M_1 P_1 = 0, \\ P_2^T M_2 P_1 = 0. \end{cases} \quad (7)$$

The coefficient relationship between fundamental matrix F and homography matrix H is established as (8) and (9). The coordinates of the projection imaging points of P on the images I_1 and I_2 are substituted into two formulas to solve the coefficients a and b . Thus, the value of the fundamental matrix F is calculated according to equation (10).

$$P_2^T (aM_1 + bM_2) P_1 = 0, \quad (8)$$

$$ah_{13} + bh_{23} = -1, \quad (9)$$

$$F = aM_1 + bM_2. \quad (10)$$

The fundamental matrix optimization algorithm is shown in Figure 2, and the algorithm steps are as follows:

Step 1: The feature points in the interior point set obtained by the optimized feature matching algorithm are normalized.

Step 2: Use RANSAC to obtain the homography matrix H between images.

Step3: The error existing in homography transformation is the distance between pixel coordinates obtained by homography transformation of interior points in image I_1 and the matching feature points in image I_2 . The errors are sorted, and inner points with a smaller error are brought into equation (2). The exact homography matrix H is obtained by calculation.

Step 4: Select feature points in the collection of outer points for calculating the initial H and bring them into equations (8) and (9) to solve the coefficients a and b . And then calculate F from equation (10).

Step 5: The epipolar geometric constraint is used to filter outer points. The distance from corresponding feature point to epipolar line is obtained by equation (12) and compared with the set threshold. If the error exceeds threshold, it is an outlier.

Step 6: Loop above steps until the fundamental matrix F with the most interior points is found.

3. Experimental Results and Analysis

In order to verify the effectiveness and robustness of the VO optimization algorithm proposed in this paper, the effects were analyzed from two aspects: feature matching and fundamental matrix calculation. The experiment selected the data `rgbd_dataset_freiburg1_room` in the TUM dataset [32]. The experimental platform adopted Ubuntu18.04 system, CPU Intel-i5, 8 GB memory, and Opencv4.5.2 open-source library.

In terms of feature matching, in order to verify the effects of the VO optimization algorithm to remove the mismatched feature points, a comparative experiment of the improved feature matching algorithm was carried out. In order to compare the error matching elimination effects using the RANSAC algorithm and improved feature matching algorithm, use $K-1$ and K frames of `rgbd_dataset_freiburg1_room` data in the TUM dataset to perform a mismatch elimination experiment as shown in Figure 3. The matching results of the initial matching algorithm, minimum threshold method, RANSAC, and improved feature matching algorithm are shown in Figure 4.

The dots at both ends of the line are shown in Figure 4, representing the feature points for feature matching between images. The straight lines represent the corresponding relationship of feature points after image matching, where parallel lines represent correct matching and the intersecting lines represent wrong matching. The initial matching results in Figure 4(a) indicate that the features between images will produce false matches due to noise and other reasons. The minimum threshold method only reduces false matching to a certain extent, and the logarithm of correct matching is

affected by the threshold setting in Figure 4(b). If the RANSAC algorithm shown in Figure 4(c) is used, the number of mismatches is reduced compared to the threshold method. However, relying on artificially set thresholds results in unstable computations. The number of matching pairs of the minimum threshold method increases with the increase of the threshold. When the threshold is set to 11, the matching pairs of the minimum threshold method are 285. There are many mismatched feature points, and the running time is long. When the threshold is set to 9, the remaining matching pairs are less. Therefore, the threshold is set to 10 in this paper. As a result, the fusion method of minimum distance threshold and random sampling consistency is used to eliminate false matching. As shown in Figure 4(d), the mismatching pairs are reduced.

The performance comparison based on picture Church [33] between this feature matching algorithm and other matching algorithms is shown in Table 1. In order to further verify the running speed of the algorithm in this paper, five different environment images of bear, computer, desk, floor, and building in TUM are used as test images to compare the running time of the algorithm, as shown in Figure 5. In this paper, the accuracy [34] is taken as the evaluation index of algorithm accuracy. Set the slope of the matching line segment [39] between two horizontal matching feature points on the matching image as the standard slope, and then compare the slope of other matching line segments with it. If the slopes are equal, it is a valid match, and the correct matching logarithm is obtained; otherwise, it is a wrong match. The definition of accuracy is as follows (11), and the slope formula is as follows (12):

$$\text{Accuracy} = \frac{\text{Correct matching pairs}}{\text{Matching pairs}}, \quad (11)$$

$$k = \frac{y_2 - y_1}{x_2 - x_1}, \quad (12)$$

where (x_1, y_1) and (x_2, y_2) are the matching feature points obtained from the interframe images filtered by the feature matching optimization algorithm.

The number of matching pairs filtered by the algorithm in this paper has decreased, but more correct matching pairs are retained than in other kinds of algorithms. Table 1 shows that the running speed of the algorithm in this paper is 0.046 s, and that of the RANSAC algorithm is 0.121 s. It can be seen that the algorithm in this paper runs faster. At the same time, the accuracy of the algorithm in this paper is 91%, which is 14% higher than that of RANSAC. At the same time, compared with TSAC [34], FLANN [35], GMS [36], KNN [37], LPM [38], the accuracy of the algorithm in this paper is increased by 7%, 26%, 8%, 30%, and 10%, respectively. In Figure 5, the horizontal axis shows the images of five different environment categories, and the vertical axis shows the running time. It can be seen from Figure 5 that the running time of the improved feature-matching algorithm is lower than other algorithms in different categories of images. Therefore, the operation efficiency of the improved feature-matching algorithm based on the minimum threshold proposed in this paper is better than the traditional



FIGURE 3: Images used in the experiment.



FIGURE 4: Comparison of different algorithms for removing mismatches. (a) Feature point initial matching between scene images. (b) Removing mismatch by a minimum distance threshold method. (c) RANSAC removes mismatches. (d) Improved feature matching algorithm.

TABLE 1: Comparison of matching performance of different algorithms.

Algorithm	Matching pairs	Correct matching pairs	Accuracy (%)	Running time (s)
OURS	54	49	0.91	0.046
RANSAC [8]	49	38	0.77	0.121
TSAC [34]	25	21	0.84	0.054
FLANN [35]	170	110	0.65	0.075
GMS [36]	47	39	0.83	0.048
KNN [37]	18	11	0.61	0.063
LPM [38]	85	69	0.81	0.050

RANSAC algorithm. It can effectively perform false matching screening, and the screening results obtained have a high accuracy rate.

In terms of basic matrix calculation, in order to verify the camera pose estimation accuracy of the VO optimization algorithm proposed in this paper, the basic matrix experiment was carried out. Experiments are carried out on different scenes of `rgb_dataset_freiburg1_room` to obtain the epipolar geometric relationship, as shown in Figure 6.

It can be known from the epipolar geometry that all epipolar lines intersect at the epipolar point, which can be used to verify whether the estimated fundamental matrix is correct. As shown in Figure 6, epipolar lines of adjacent frame images intersect at one point, which is the epipolar point. It shows that the calculation result of the fundamental matrix is correct. The matching corresponding points are represented by circles. The corresponding fundamental matrix is estimated by the method in this paper, and the

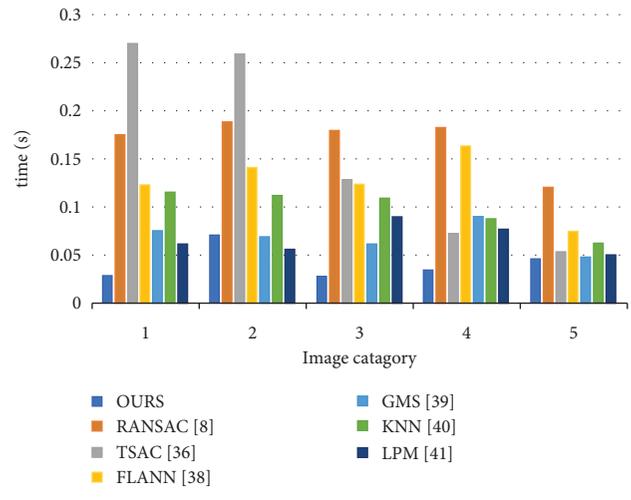
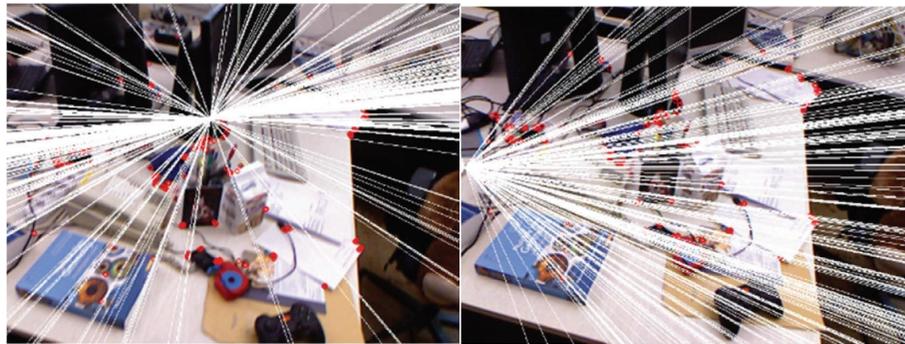
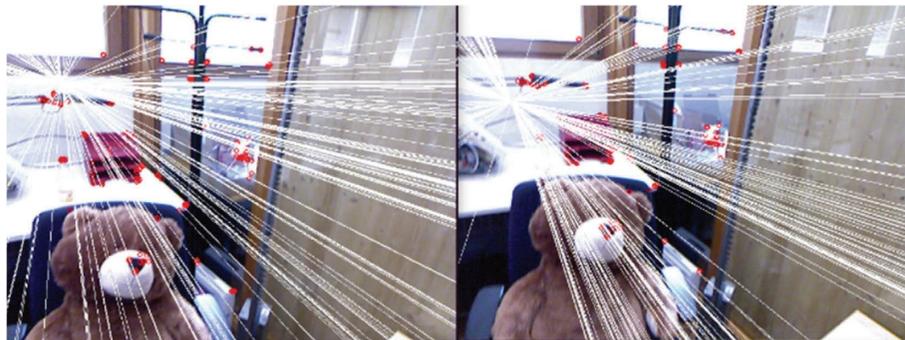


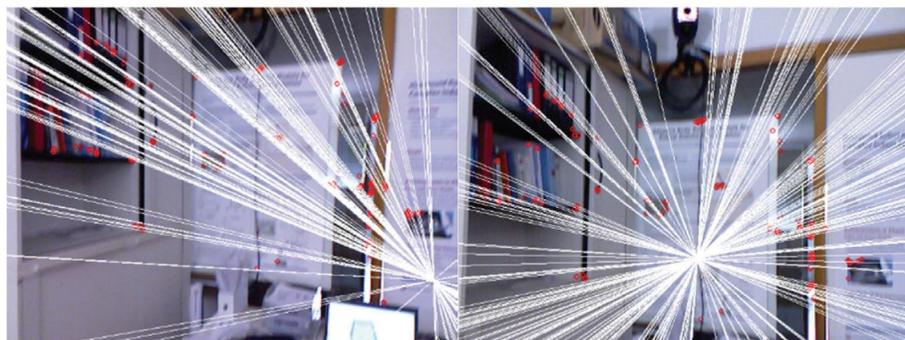
FIGURE 5: Comparison of algorithm running time based on different types of images.



(a)



(b)



(c)

FIGURE 6: Continued.

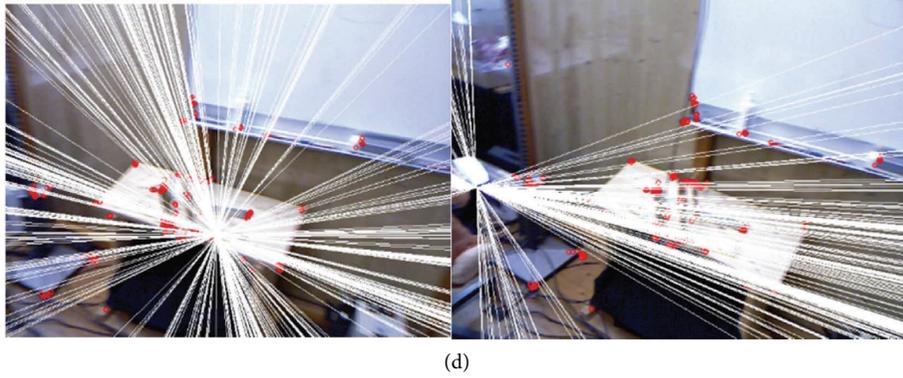


FIGURE 6: The epipolar geometric relationship estimated by our method for different scenes. (a) Computer, (b) bear, (c) blackboard, (d) desk.

TABLE 2: Interior point rate of each algorithm based on different scenes.

Scenes	RANSAC [8]	Reference [41]	Reference [42]	Algorithm in this paper (%)
Computer	74	78	86	89
Bear	71	72	80	85
Blackboard	69	72	76	92
Desk	80	82	88	91

corresponding epipolar line which is straight line in figures is drawn. At the same time, feature points basically fall on the estimated corresponding epipolar lines, which accurately represents obtained epipolar geometric relationship. It shows the accuracy of the algorithm in this paper.

The method in this paper is compared with the algorithm based on RANSAC to calculate the fundamental matrix. Interior point rate is used as an evaluation error for the estimation accuracy of the fundamental matrix, and the evaluation is performed on all feature points. It is the probability that the distance from point to epipolar line is less than 1 pixel [40]. Firstly, calculate the distance from the point \mathbf{p}'_i to the epipolar line $L'_i = \mathbf{F}\mathbf{p}_i$, as shown in equation (13). The average distance from matching points to epipolar line is further calculated, as shown in (14).

$$D = \frac{\mathbf{p}'_i{}^T \mathbf{F}\mathbf{p}_i}{\sqrt{(\mathbf{F}\mathbf{p}_i)_1^2 + (\mathbf{F}\mathbf{p}_i)_2^2}} \quad (13)$$

$$d = \frac{|\mathbf{p}'_i{}^T \mathbf{F}\mathbf{p}_i|}{2} \left(\frac{1}{\sqrt{(\mathbf{F}\mathbf{p}_i)_1^2 + (\mathbf{F}\mathbf{p}_i)_2^2}} + \frac{1}{\sqrt{(\mathbf{F}^T \mathbf{p}'_i)_1^2 + (\mathbf{F}^T \mathbf{p}'_i)_2^2}} \right), \quad (14)$$

where $d(\mathbf{p}_i, \mathbf{F}\mathbf{p}'_i)$ and $d(\mathbf{p}'_i, \mathbf{F}^T \mathbf{p}_i)$ are the geometric distance between the matching point pair \mathbf{p}_i (\mathbf{p}'_i) and the corresponding epipolar line in the image, respectively. The interior point rate of each algorithm is shown in Table 2.

In the comparative experiment, the basic matrix is calculated based on traditional RANSAC [8], reference [41], and reference [42], respectively. For four different scenes, it can be seen interior point rate of the algorithm in this paper has been improved by about 15.75%, 13.25%, and 6.75% compared with the above three algorithms. Experiments show that the estimation accuracy of the basic matrix is

improved by using the optimization fundamental matrix algorithm based on homography transformation and epipolar geometric constraints.

Therefore, combining the experiments of feature matching and fundamental matrix calculation, the VO optimization method proposed in this paper can obtain reliable object motion estimation. The performance of VO will be improved.

4. Conclusions

The sensor is affected by environment, texture repetition, and illumination, resulting in the difficulty of image matching. Due to the influence of mismatching on the calculation of the fundamental matrix, the estimation accuracy of VO is poor. An optimization method of VO based on feature matching is proposed. Firstly, the feature matching based on the minimum threshold is used to reduce the number of mismatched feature points. Then, the camera pose is estimated by the fundamental matrix optimization algorithm based on the remaining point set to realize the initial pose optimization. The VO estimation performance is improved by the above algorithm. The following conclusions were drawn through the above research:

- (1) Integrating the minimum threshold method with the RANSAC algorithm can remove mismatched feature points. The homography matrix is calculated by the remaining point set. It combines epipolar geometric constraints to calculate a precise fundamental matrix, so as to realize object motion estimation.
- (2) The results show that the accuracy of image feature points matching increased by 15.8%, and the rate of interior points increased by 11.9%. Thus, the problem of feature points mismatching in RANSAC has

been solved, and the estimation accuracy of the fundamental matrix has been improved.

- (3) Reliable motion estimation and the robustness of VO are innovatively analyzed in this paper. It will also provide a basis for subsequent theoretical research on VO.

Abbreviations

VO:	Visual odometry
SLAM:	Simultaneous localization and map building
RANSAC:	Random sample consistency
ORB:	Oriented FAST and rotated BRIEF
TSAC:	Triangular topology probability sampling consensus
FLANN:	Fast approximate nearest neighbor search library
GMS:	Grid-based motion statistics
KNN:	K-nearest neighbors
LPM:	Locality preserving matching.

Data Availability

The data supporting the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The research was supported by the Youth Program of Hebei Province Department of Education (No. QN2019232) and Natural Science Foundation of Hebei Province (No. E2019210299).

References

- [1] F. Fraundorfer and D. Scaramuzza, "Visual odometry: Part II—matching, robustness, and applications," *IEEE Robotics and Automation Magazine*, vol. 19, no. 2, pp. 78–90, 2012.
- [2] D. Scaramuzza and F. Fraundorfer, "Visual odometry: Part I: the first 30 Years and fundamentals," *IEEE Robotics and Automation Magazine*, vol. 18, no. 4, pp. 80–92, 2011.
- [3] G. I. Kwon, Y. H. Seo, and H. S. Yang, "Spatiotemporal 3D motion vector filtering method for robust visual odometry," *Electronics Letters*, vol. 49, no. 5, pp. 341–343, 2013.
- [4] Y. Shuai, J. Rui, W. Han, and S. S. Ge, "Road constrained monocular visual localization using Gaussian-Gaussian cloud model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 12, pp. 1–8, 2017.
- [5] X. Yang, S. Li, K. Liang, F. Nie, and L. Lin, "Structured graph optimization for joint spectral embedding and clustering," *Neurocomputing*, vol. 503, pp. 62–72, 2022.
- [6] J. Bentolila and J. M. Francos, "Conic epipolar constraints from affine correspondences," *Computer Vision and Image Understanding*, vol. 122, pp. 105–114, 2014.
- [7] P. Chen and D. Suter, "Rank constraints for homographies over two views: revisiting the rank four constraint," *International Journal of Computer Vision*, vol. 81, no. 2, pp. 205–225, 2009.
- [8] D. Nistér, "Preemptive RANSAC for live structure and motion estimation," *Machine Vision and Applications*, vol. 16, no. 5, pp. 321–329, 2005.
- [9] S. Ma, P. Guo, H. You, P. He, G. Li, and H. Li, "An image matching optimization algorithm based on pixel shift clustering RANSAC," *Information Sciences*, vol. 562, pp. 452–474, 2021.
- [10] M. Zhai, S. Fu, and Z. L. Jing, "Homography estimation from planar contours in image sequence," *Optical Engineering*, vol. 49, no. 3, Article ID 037202, 2010.
- [11] Y. Liu, Y. Li, L. Dai et al., "Robust feature matching via advanced neighborhood topology consensus," *Neurocomputing*, vol. 421, pp. 273–284, 2021.
- [12] F. Meng, X. Li, and J. Pei, "A feature point matching based on spatial order constraints bilateral-neighbor vote," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4160–4171, 2015.
- [13] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
- [14] M. Emam, Q. Han, and H. Zhang, "Two stage keypoint detection scheme for region duplication forgery detection in digital images," *Journal of Forensic Sciences*, vol. 63, no. 1, pp. 102–111, 2018.
- [15] Y. G. Xie, Q. Wang, Y. X. Chang, and X. Y. Zhang, "Fast target recognition based on improved ORB feature," *Fast Target Recognition Based on Improved ORB Feature Applied Sciences*, vol. 12, no. 2, p. 786, 2022.
- [16] X. Fan, L. Xing, J. Chen et al., "VLSG-SANet: a feature matching algorithm for remote sensing image registration," *Knowledge-Based Systems*, vol. 255, Article ID 109609, 2022.
- [17] J. Mallon and P. F. Whelan, "Projective rectification from the fundamental matrix," *Image and Vision Computing*, vol. 23, no. 7, pp. 643–650, 2005.
- [18] H. Li, J. Zhao, J. C. Bazin, and Y. H. Liu, "Robust estimation of absolute camera pose via intersection constraint and flow consensus," *IEEE Transactions on Image Processing*, vol. 29, pp. 6615–6629, 2020.
- [19] H. Kiskinov, E. Madamlieva, M. Veselinova, and A. Zahariev, "Existence of absolutely continuous fundamental matrix of linear fractional system with distributed delays," *Mathematics*, vol. 9, no. 2, p. 150, 2021.
- [20] W. Chojnacki and M. J. Brooks, "On the consistency of the normalized eight-point algorithm," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 1, pp. 19–27, 2007.
- [21] F. Bugarin, A. Bartoli, D. Henrion, J. B. Lasserre, J. J. Orteu, and T. Sentenac, "Rank-constrained fundamental matrix estimation by polynomial global optimization versus the eight-point algorithm," *Journal of Mathematical Imaging and Vision*, vol. 53, no. 1, pp. 42–60, 2015.
- [22] F. Ke, H. Liu, G. Sun, D. Zhao, W. Xu, and W. Feng, "A hybrid algorithm for calculating the global solution of the fundamental matrix problem," *Optik*, vol. 204, Article ID 164183, 2020.
- [23] G. H. Lee, B. Li, M. Pollefeys, and F. Fraundorfer, "Minimal solutions for the multi-camera pose estimation problem," *The International Journal of Robotics Research*, vol. 34, no. 7, pp. 837–848, 2015.
- [24] G. L. Mariottini, G. Oriolo, and D. Prattichizzo, "Image-based visual servoing for nonholonomic mobile robots using epipolar geometry," *IEEE Transactions on Robotics*, vol. 23, no. 1, pp. 87–100, 2007.

- [25] Y. Goldman, E. Rivlin, and I. Shimshoni, "Robust epipolar geometry estimation using noisy pose priors," *Image and Vision Computing*, vol. 67, pp. 16–28, 2017.
- [26] C. Shao, C. Zhang, Z. Fang, and G. Yang, "A deep learning-based semantic filter for ransac-based fundamental matrix calculation and the ORB-slam system," *IEEE Access*, vol. 8, pp. 3212–3223, 2020.
- [27] Y. Kayhan and S. Gunay, "A new approach to least median of squares and regression through the origin," *Communications in Statistics - Theory and Methods*, vol. 37, no. 5, pp. 773–781, 2008.
- [28] M. Chichignoud and J. Lederer, "A robust, adaptive M-estimator for pointwise estimation in heteroscedastic regression," *Bernoulli*, vol. 20, no. 3, pp. 1560–1599, 2014.
- [29] N. Tatar and H. Arefi, "Stereo rectification of pushbroom satellite images by robustly estimating the fundamental matrix," *International Journal of Remote Sensing*, vol. 40, no. 23, pp. 1–20, 2019.
- [30] R. Yang, J. Zhang, and B. Li, "Estimating the Fundamental Matrix Based on the End-To-End Convolutional Network," *Applied Intelligence*, vol. 52, pp. 1–12, 2022.
- [31] G. Shao and N. Sang, "Max–min distance analysis by making a uniform distribution of class centers for dimensionality reduction," *Neurocomputing*, vol. 143, pp. 208–221, 2014.
- [32] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [33] G. Wang, H. T. Tsui, Z. Hu, and F. Wu, "Camera calibration and 3D reconstruction from a single view based on scene constraints," *Image and Vision Computing*, vol. 23, no. 3, pp. 311–323, 2005.
- [34] Z. X. He, C. T. Shen, Q. Y. Wang, X. Zhao, and H. Jiang, "Mismatching removal for feature-point matching based on triangular topology probability sampling consensus," *Remote Sensing*, vol. 14, no. 3, p. 706, 2022.
- [35] M. M. Esmaeili, R. K. Ward, and M. Fatourehchi, "A fast approximate nearest neighbor search algorithm in the hamming space," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 12, pp. 2481–2488, 2012.
- [36] J. W. Bian, W. Y. Lin, Y. Liu, T. D. Nguyen, and M. M. Cheng, "GMS: grid-based motion Statistics for fast, Ultra-robust Feature Correspondence," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 6, Springer US, Honolulu, HI, USA, July 2017.
- [37] P. Franti, O. Virmajoki, and V. Hautamaki, "Fast agglomerative clustering using a k-nearest neighbor graph," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1875–1881, 2006.
- [38] J. Ma, J. Zhao, H. Guo, J. Jiang, and H. Zhou, "Locality Preserving Matching," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Toronto, Ontario, Canada, July 2017.
- [39] Z. Zheng, H. Zheng, Y. Ma et al., "Feedback unilateral grid-based clustering feature matching for remote sensing image registration," *Remote Sensing*, vol. 11, no. 12, p. 1418, 2019.
- [40] X. f Feng and B. Fang, "Algorithm for epipolar geometry and correcting monocular stereo vision based on a plane mirror," *Optik*, vol. 226, Article ID 165890, 2021.
- [41] S. D. Wei and S. H. Lai, "Fast template matching based on normalized cross correlation with adaptive multilevel winner update," *IEEE Transactions on Image Processing*, vol. 17, no. 11, pp. 2227–2235, 2008.
- [42] W. Beiyi, Z. Xiaohong, and W. Weibing, "Feature matching method based on SURF and fast library for approximate nearest neighbor search," *Integrated Ferroelectrics*, vol. 218, no. 1, pp. 147–154, 2021.