TWO-STAGE FRAME MATCHING IN VSLAM BASED ON FEATURE EXTRACTION WITH ADAPTIVE THRESHOLD FOR INDOOR TEXTURE-LESS AND STRUCTURE-LESS

Ding Wang,* Jing Wang,** and Huan Wang*

Abstract

Visual simultaneous localization and mapping (VSLAM) is a key algorithm in the navigation of mobile robots to be studied by many researchers. In VSLAM, it is the first and fundamental key technology to determine the common features of the different views of the same object from the acquired environment images. It determines the adaptability of the VSLAM algorithm to difficult environments with poor textures and structures. Therefore, the challenge of the VSLAM algorithm imposed by the poor texture and structure of the indoor environment is indeed the requirement of the good matching technology. To this end, that two-stage frame matching in VSLAM based on feature extraction with adaptive threshold for indoor texture-less and structure-less is proposed in the paper. The ORB algorithm with fast extraction speed and strong real-time performance is improved, adding adaptability to the FAST corner detection algorithm while increasing the scale invariance, so that the algorithm can automatically adjust the threshold to get enough feature points. After comparing the common feature matching methods, the violent matching was selected as the coarse matching, and the improved mismatches elimination method based on the motion smoothing model is proposed. The experiments show the improved feature extraction and matching algorithm Ours can extract enough key points in texture and structure-free environment, refine the local points to achieve better feature matching between current frame and reference frame, and provide a basis for subsequent pose estimation.

Key Words

VSLAM, matching, feature extraction, adaptive threshold, texture-less, structure-less

* Xinjiang College of science & technology, Korla 841000, China; e-mail: {782939172, 3376849488}@qq.com
** North China Institute of Aerospace Engineering, Langfang 065000, China; e-mail: 782939172@qq.com
 Corresponding author: Jing Wang

1. Introduction

With the wide application of mobile robots, visual simultaneous localisation and mapping (VSLAM) algorithm has become research hotspots, which is a key algorithm in navigation of mobile robots of the dynamic environment and objects. The applicational problems which were produced from the original static use of VSLAM into dynamic environment are studied, such as handled dynamic objects with mask loop closing [1], improved the performance of SLAM in dynamic environments using parallel map [2]. A consensus estimation algorithm was proposed to correct the weight distribution of local information to ensure the accuracy of the map under dynamic network conditions [3]. A cubature-extended H∞ filter-based SLAM algorithm (CEH∞F-SLAM) is proposed to possess advantages of both algorithms [4]. VSLAM is proposed based on the RatSLAM method to reduce the higher computational requirements and increase the lower accuracy [5]. It is the first key technology to extract the object feature points and determine the matching of the same object in VSAM. It determines the adaptability of the VSLAM algorithm to difficult environments with poor textures and structures. Therefore, the challenge of the VSLAM algorithm imposed by the poor texture and structure of the indoor environment is the requirement of the target frame matching technology.

To overcome these challenges, von Gioi *et al.* [6] used line segment detection (LSD) to extract the corresponding line features in the untextured environment based on the structural regularity of the indoor environment and represented a 3D line with as a segment with two endpoints. Van Opdenbosch and Steinbach [7] build a small workspace environment map, whose endpoints do not easily disappear from the camera view, based on the system architecture of ORB-SLAM 2, using the LSD. Zhu *et al.* from Shanghai Jiao Tong University studied the line characteristics in the detection environment and improved the original KinectFusion method [8], which is very different from the ordinary feature point detection method. The matching accuracy of the motion estimated by line matching in areas

with few texture and few feature points is higher than the point matching [9]. A graph-based visual SLAM system characterised by Zhou *et al.* [10] compared to the point features, the line features provide much richer information about the environmental structure, making it possible to infer the spatial semantics from the maps. Compared with the point-based SLAM system, it has better reconstruction performance in both indoor and outdoor rich texture environments, and the disadvantage is that it cannot be used normally in an unstructured environment where the texture is not obvious.

The fusion of point and line features is a method that has been more applied in the SLAM field in recent years. Compared with the methods that use only point features or line features, the method of combining point and line features can be widely applicable to low-texture scenes and has better accuracy and robustness. Pumarola *et al.* proposed a real-time single point system with point and line (PL-SLAM) based on the ORB-SLAM 2 framework, allowing the integration of line representation into the SLAM mechanism. When most feature points disappear from the input image, it can still estimate the camera attitude and 3D map by detecting five lines in three consecutive images [11]. The system can run in low-texture scenes, but not in environments where structures are similarly blurred. Ruben *et al.* proposed a stereo SLAM system using points and lines and additionally extended a closed-loop module based on points and lines, which can be built in a low-texture environment. However, although the new closed-loop method increases robustness, the high computational cost makes it difficult for the system to achieve real-time [12]. Lee *et al.* developed a low-cost embedded indoor service robot system based on vanishing point and vanishing line landmarks and proposed a new method for implementing SLAM using forwards monocular vision sensors[13]. This method has good performance in complex scenarios without textured areas, environmental changes, and with moving obstacles, but is poorly applied in structure-free environments. The team of Hunan University proposed a complete high-precision SLAM system based on the combination of point and line information, and proposed a line-based localised point refinement algorithm to eliminate the outlier [14]. The reprojection error optimisation model of points and lines is established, and extended to the local bundle adjustment (BA) [15]–[17], which can be built in non-textured scenes. However, the disadvantage is that the highly redundant point cloud requires a large amount of storage resources, and the occlusion problem of extracting line features is not solved, and it is difficult to achieve in a non-structural environment.

But feathers of point and line is played of distributions of the geometry in environment, and adaptive characters of feature extraction to various kinds of geometry with adaptive threshold were necessary to raise the improve the extract accuracy, and more comprehensive and refined extract must is combined with them to difficult feather extraction to the poor texture and structure of the indoor environment. The violent matching was selected to improve mismatches elimination method based on the motion smoothing model is proposed. Hence, the paper presented such an algorithm based on the concept. Study protocol is presented in Section 2 and experimental analysis is presented in Section 3. The contributions of the paper is to combine adaptive feature extraction to catch the point of feature more with the violent matching to use of existing the point of feature better to improve the performance of the template matching algorithm for the same object in VSAM that determines work effect in harsh environments.

2. Study Protocol and Implementation Method

2.1 Study Protocol

The ORB algorithm has currently been commonly used feature extraction algorithms was proposed in 2011, using the oriented features from accelerated segment test (oFAST) [18] algorithm for feature point extraction, namely, by adding the direction information to the extracted FAST angles. ORB has the advantage of fast feature extraction speed and can basically meet the demand of real-time performance. Therefore, it has too little feature extraction and no scale invariant, so that it can extract enough feature points in texture and structure-free environment.

Common feature matching algorithms include Brute-Force (BF) matching algorithm [19], fast library for approximate nearest neighbors (FLANN) [20] and RANSAC [21] algorithm. It can be seen that the violent matching method tries all the matches as much as possible in both the textured and untextured scenes, getting a large number of matching pairs, and there are also many mismatches in the textured scenes. Compared to the number of the mismatches for violent matches, the FLANN algorithm decreased in both textured and untextured environments. However, the number of matching pairs in the textureless environment is also small, so it is very suitable for feature matching in high-texture scenes than the violent matching FLANN method. Therefore, the RANSAC algorithm is often used to eliminate mismatches in practise. Compared with violent matching and FLANN, this algorithm has no mismatch in either textured or no textured scenes, which has very good robustness. However, the RANSAC algorithm has some limitations. If the point set sample is to be optimal, the number of iterations must be large enough. Especially in the untextured environment, not only is it not easy to ensure the matching accuracy but also has a very large computational cost.

Therefore, to extract enough feature points, the ORB algorithm is improved by enhancing the scale invariant and adaptive setting the corner point determination threshold; to be suitable for no texture and no structure environment, the violent matching device which is easy to implement and has rich matching set is used as coarse matching, and then the improved motion smoothing model algorithm is eliminated to improve the accuracy and reliability of the feature matching algorithm. The proposed study protocol which is combined the improved ORB algorithm with the violent matching is presented as follows.
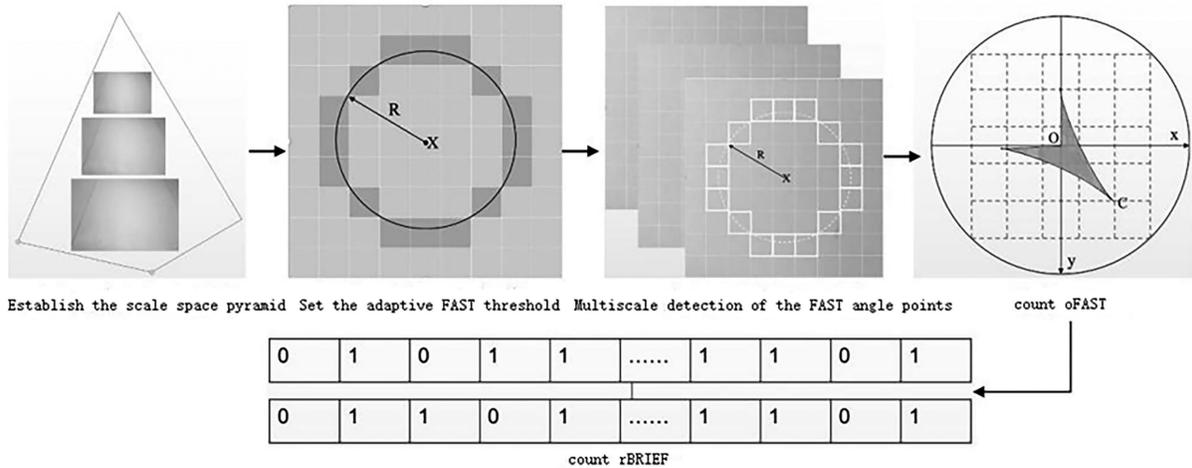
Figure 1. Improved ORB algorithm feature extraction diagram.

2.2 Improvement of the ORB Feature Extraction Algorithm

The improvement of the ORB feature extraction algorithm is mainly in two parts. In the first part, the image scale-in-variance is enhanced by constructing a Gaussian-scale spatial pyramid. Equation (1), the total number of layers of the pyramid is defined as the L layer, set the scale of the input image as $M \times N$, select the smaller value between M and N, and the size of each layer image is determined by down-sampling from the upper layer, and different spatial scale information can be extracted from the same picture.

$$L = \log_2\left[\min\left(M, N\right) - 2\right] \qquad (1)$$

The scale of each layer is calculated as follows:

$$\sigma\left(l\right) = \sigma_0 \times 2^{\frac{l}{2}} \qquad (2)$$

where l represents the l th layer and σ_0 represents the initial scale. For each layer, key points are extracted at multiple scales to increase the scale-invariant properties of the FAST algorithm.

Next, the FAST corner point detection algorithm is improved. Since the minimum contrast threshold of the feature point and its surrounding pixels determines the selection of key points, the lower the threshold, the more key points, so different scenarios require manual threshold adjustment. Based on this method, the FAST detection algorithm is improved, which can set the determination threshold around the corner points without manual adjustment.

The candidate feature point x and its grey value $I(x)$ were selected, centered on x, with 16 pixels $\{I(y_i), i = 1, 2, \ldots, 16\}$ around the search radius $R = 3$. The determination of FAST angles is related to the grey scale change of the image, so a dynamic adjustment strategy is used to handle different scenarios. The mean of the grey change at the corner was calculated and the quantitative relationship between it and the threshold was

balanced by the proportional coefficient k. The formula for calculating each frame threshold is expressed as follows:

$$\varepsilon = k\left(\frac{1}{m}\sum_{i=1}^{m} I_{i_{\max}} - \frac{1}{m}\sum_{i=1}^{m} I_{i_{\min}}\right)/I_{i_{\mathrm{avg}}} \qquad (3)$$

where $I_{i_{\max}}$, $I_{i_{\min}}$, and $I_{i_{\mathrm{avg}}}$ are all statistics, $I_{i_{\max}}$ and $I_{i_{\min}}$ represent the m largest and m smallest grey values on the circle, respectively, $I_{i_{\mathrm{avg}}}$ is the mean of the grey values on the circle, k represents the proportional coefficient, and ε is the threshold.

Select N points in $I(y_i)$, and if all N points satisfy $\|I(y_i) - I(x)\|$, then x is the key point. Through the Formula of (4), the improved FAST algorithm can also adaptively extract enough feature points in the scene without texture and structure.

The ORB algorithm consists of the oFAST algorithm and the rBRIEF algorithm, and the oFAST increases the rotation in variance of the ORB algorithm by adding an additional orientation information,θ. That is, for the key point x, its direction is represented by the vector $\overrightarrow{OC}$, where O is the origin of the coordinate system, and C represents the neighbourhood centre of the key point x. The key point x is then described by the rBRIEF algorithm with a rotation factor rBRIEF. The algorithm first performs Gaussian smoothing of the image, and then selects a certain number of point pairs of D, and generally selects 256 pairs. Then Gaussian sampling is performed near the key point x, and direction θ is obtained by oFAST algorithm. The new point pairs are calculated as follows:

$$D_\theta = \theta D \qquad (4)$$

The final comparison with the values of the new dot pair resulted in 256-bit descriptors. Figure 1 shows the feature extraction graph of the improved ORB algorithm, the Gaussian distribution enhances the scale consistency of the algorithm, the adaptive FAST algorithm guarantees enough key points, and the rotation factor improves the rotation invariance of the algorithm.

2.3 Improved Motion Smoothing Model Mismatches and Elimination

The motion-smoothing model was proposed by Lin *et al.*, as the equation [22]:

$$\lim_{\triangle p \to 0} f_k\left(P + \triangle P\right) - f_k\left(P\right) = 0, P = [x,y]^T \in \mathfrak{R}^2 \quad (5)$$

where P represents the pixel coordinates, $\triangle P$ is the increment of P, $f_k\left(P\right)$ is the smoothing function and $\mathfrak{R}^2$ is the n-dimensional space. Functional values around P tend to be equal, however, the model is unstable due to the random distribution of mismatches.

The two plots are the reference RGB frame I_r and the current RGB frame I_c, respectively. After violence matching, obtain the set $X_{i,c \to r} = \{x_1, x_2, \ldots, x_m\}$ of nearest neighbour feature matching pairs. The number of matched pairs with correctly matched regions $\varphi_i = 2$ is greater than the number of matched pairs with correctly matched regions $\varphi_i = 0$. By determining the support number of the matching pairs within the region, the wrong matches can be eliminated. The number of matching pair supports $F\left(x_i\right)$ can be expressed as (6):

$$F\left(x_i\right) = |X_i| - 1, X_i \subseteq X_{i,c \to r} \quad (6)$$

here X_i represents the sets of matching pairs and x_i within the surrounding neighbourhood.

Two hypotheses are proposed for the improvement of mismatch elimination. First, the selection region is small enough, with both correct and incorrect matches, and the ident distribution of matching pairs. Then, the distribution of the matching pair support number $F\left(x_i\right)$ in the neighbourhood can be approximated as a binomial distribution:

$$F\left(x_i\right) \sim \begin{cases} B\left(Kn, p_t\right), & x_i \in True \\ B\left(Kn, p_f\right), & x_i \in False \end{cases} \quad (7)$$

where x_i represents the ith matched pair, n represents the number of matched pairs, $x_i \in True$ indicates that x_i belongs to the correct matched pair region, and $x_i \in False$ indicates that x_i belongs to the mismatched region. K represents the number of small disjoint regions around x_i, and p_t and p_f indicate the probability of correctly and mismatched pairs, respectively.

Then, the mean value of the distribution is:

$$\begin{cases} E_t\left(x_i\right) = Knp_t, & x_i \in True \\ E_f\left(x_i\right) = Knp_f, & x_i \in False \end{cases} \quad (8)$$

The standard deviation of this distribution can be calculated as:

$$\begin{cases} \sigma_t\left(x_i\right) = \sqrt{Knp_t\left(1 - p_t\right)}, & x_i \in True \\ \sigma_f\left(x_i\right) = \sqrt{Knp_f\left(1 - p_f\right)}, & x_i \in False \end{cases} \quad (9)$$

If the mean of the correctly matched and mismatched distributions separates by sufficiently large values relative to the standard deviation of the distribution, it represents a
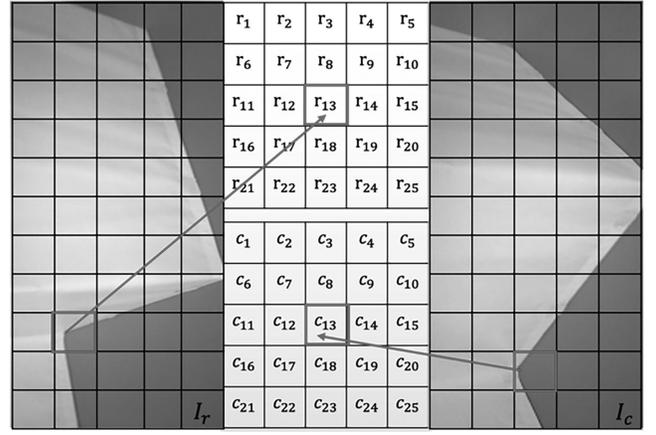
Figure 2. Schematic diagram of cell (r, c).

large interval between the correctly distributed $B\left(Kn, p_t\right)$ and the $B\left(Kn, p_f\right)$ incorrectly distributed, and indicating that the value of $F\left(x_i\right)$ becomes a useful indicator to distinguish true and false matches. Therefore, the design parameter P_x to evaluate the distribution area, which can be expressed as:

$$\begin{aligned} P_x &= \frac{E_t\left(x_i\right) - E_f\left(x_i\right)}{\sigma_t\left(x_i\right) - \sigma_f\left(x_i\right)} \\ &= \frac{\sqrt{Kn}\left(p_t - p_f\right)}{\sqrt{p_t\left(1 - p_t\right)} + \sqrt{p_f\left(1 - p_f\right)}} \end{aligned} \quad (10)$$

From (10), it can be concluded that P_x and $\sqrt{Kn}$ are proportional, that is, the distribution interval is determined by the small area K and the number of matches n which determines the performance of mismatch elimination.

In the actual operation of the algorithm, the two images were divided into G non-repeating grid cells. The grid cell r in the reference RGB frame I_r was matched to the grid cell c in the current RGB frame I_c, and each cell pair (r, c) and surrounding regions were used to evaluate the support number of $F\left(x_i\right)$ matches. According to the empirical value $G = 20 \times 20$ cells corresponding to 10,000 feature points, the mean of the matched pair n is 25. When n is large enough, the constraints is strong but not enough features in scenes without texture. Therefore, to adapt to the environment without texture and structure, the algorithm divides the image more by increasing the number of small regions K and G, and counts the number of matches in more regions. With $G = 30 \times 30$, the image is divided into more cells on the original basis, and the number of small areas K is added to calculate the 55 grids around each cell, so that more matching numbers can be calculated in the disjoint small areas, and the matching accuracy of the algorithm is improved, as shown in Fig. 2.

$\{r_k, c_k\}$ represents the small area of the k pair in the (r, c) neighbourhood meets the I_r and I_c motion smoothing model, and in the neighbourhood of the 5×5 grids around the cell pair (r, c) $F\left(x_i\right)$ can be expressed as:

$$F\left(x_i\right) = \sum_{k=1}^{25} |X_{r_k, c_k}| - 1 \quad (11)$$

Here X_{r_k,c_k} is the number of matching pairs in the $\{r_k, c_k\}$ region.

$F(x_i)$ divide cell pairs into true and false sets $\{T, F\}$, with the following formula:

$$\{r_k, c_k\} \in \begin{cases} \text{True}, \ F(x_i) > 6\sqrt{n_i} \\ \text{False, otherwise} \end{cases} \tag{12}$$

where n_i is the average of the number of features in the 25 cells, and the empirical value of 6 is the parameter for establishing a quantitative relationship between $F(x_i)$ and n_i. Cell matching pairs that satisfy $F(x_i) > 6\sqrt{n_i}$ are correct matches and otherwise mismatches.

In the algorithm, each cell should first be traversed from 1 to G in I_r, and the same cell from 1 to G in I_c, select the cell c with the most matching pairs with the cell r, get (r, c), then calculate the number of matching pairs X_{r_k,c_k}, sort by calculating the support number $F(x_i)$, if $F(x_i) > 6\sqrt{n_i}$, $\{r_k, c_k\}$ for the correct match set, otherwise the false match set, and eliminate. Since many features are located at the edge of the grid, the above steps are iterated three times, and each time the grid moves a half of the cell width in the horizontal, vertical, and diagonal directions, resulting in the refined localised point.

3. Improved Feature Extraction and Matching Experimental Analysis

Experiments were tested on the algorithm by the TUM dataset under the Ubuntu16.04 system. Here choose four challenging scenes with no texture with structure a) fr 3 _ str _ not _ far, b) fr 3 _ str _ not _ near _ wl and no texture with no structure c) fr 3 _ nostr _ not _ far, d) fr 3 _ nostr _ not _ near_wl scenes. Where far and near represent the distance of the camera from the scene, and wl represents the closed-loop loop. Set the ratio threshold ratio = 0.7, and the improved feature extraction and matching method is represented by Ours. Feature extraction and matching in four different scenarios were conducted on SIFT + ratio, SURF + ratio, ORB + ratio, and Ours through OpenCV3.2.0, and the maximum quantitative threshold of key points was set to 5,000 to extract as many key points as possible. Feature extraction and matching experiments are shown in Fig. 3.

In Fig. 3, SIFT + ratio can extract key points in these scenes without texture, but it can extract very little to obtain enough matching pairs. SURF + ratio failed in feature extraction and matching in the four scenes without textured and without textured structure. ORB + ratio extracted a large number of key points, but the ratio method relying on distance discrimination produces considerable mismatch in texture-less scenes and fails to get more correct matching pairs. In contrast, the improved feature extraction and matching algorithm Ours is independent of distance and is able to retain more correct matches in extreme scenarios, significantly improving the matching accuracy.

Since the four scenarios in Fig. 3 are still scenarios, to compare the correct matching ratio and running time of
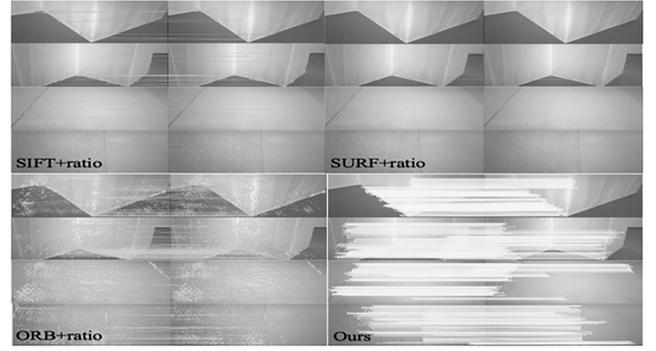
Figure 3. Comparison of feature extraction and matching algorithms in texture-less scene.

the algorithm, continuous video segments in the TUM-fr 3 dataset were selected for performance testing, as shown in Fig. 4.

Figure 4 shows the internal point numerical statistics of the four sequences, in Fig. 4(a) the sequence running time is 27.28 s, in Fig. 4(b) the sequence running time is 36.44 s, in Fig. 4(c) the sequence running time is 15.79 s, and in Fig. 4(d) the sequence running time is 37.74 s. From Figs. 4, SURF + ratio failed in all four scenarios. The SIFT + ratio algorithm extracts a small number of key points in both Fig. 4(a) and (b), and it fails at around 13 s. ORB + ratio has fewer localized points in the four sequences. Under the same threshold constraint, the improved algorithm Ours can get more localised points with strong stability.

Statistics are made on the local points in Fig. 4(a)–(d), respectively, and the correct matching ratio refers to the percentage of the local points after using the RANSAC algorithm, *i.e.*

$$\text{Correct matching ratio} = \frac{N_{\text{RANSAC}}}{N_{\text{inliers}}} \tag{13}$$

The correct matching ratio of Fig. 4 is calculated according to (13), and the results are shown in Table 1. The improved method Ours has a correct matching rate of more than 82% in a textured and structure-free environment, so the method can not only guarantee a sufficient number of localised points but also obtain enough accurate matching pairs.

Although SIFT, SURF, and ORB all use the same ratio test method for local point refinement, they will also have large differences in computation time because different extractors output a different number of key points. The running time of the algorithm was calculated by dividing the total time by the number of local points, *i.e.*

$$\text{Running Time} = \frac{\text{Time}_{\text{total}}}{N_{\text{inliers}}} \tag{14}$$

According to Table 2, the average running time of Ours is 0.03 μs, which is only 0.01 μs slower than that of ORB + ratio. Therefore, the improved algorithm also shows very good performance in running time.
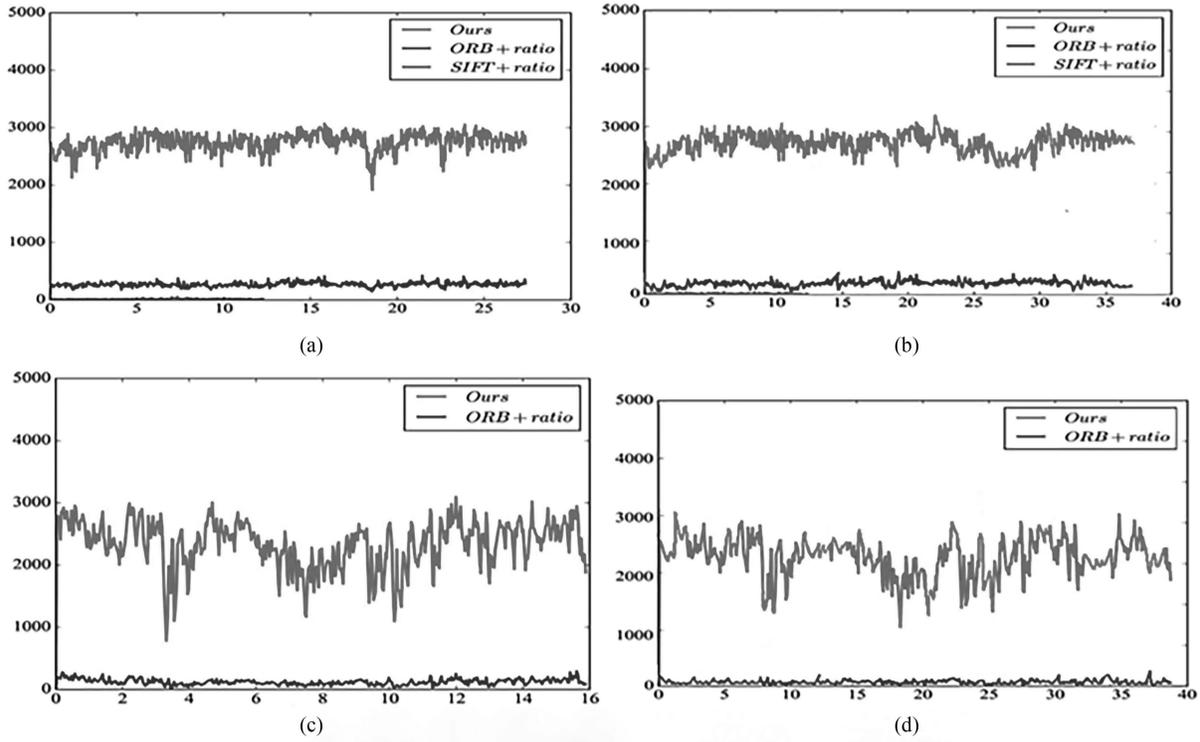
Figure 4. Statistics of inlier numbers: (a) fr3_str_not_far dataset; (b) fr3_str_not_near_wl dataset; (c) fr3_nostr_not_far dataset; and (d) fr3_nostr_not_near_wl dataset.

Table 1
Correct Matching Ratio Compared Among Four Feature Extraction Algorithms (%)

	fr3_str_not_far	fr3_str_not_near_wl	fr3_nostr_not_far	fr3_nostr_not_near_wl
SIFT+ratio	77.13	76.88	—	—
SURF+ratio	—	—	—	—
ORB+ratio	66.41	66.20	56.52	56.44
Ours	86.78	86.33	82.51	82.30

Table 2
Comparison of Average Running Time of Four Algorithms (μs)

	fr3_str_not_far	fr3_str_not_near_wl	fr3_nostr_not_far	fr3_nostr_not_near_wl
SIFT+ratio	1.30	1.31	—	—
SURF+ratio	—	—	—	—
ORB+ratio	0.03	0.03	0.02	0.02
Ours	0.02	0.02	0.03	0.03

4. Conclusion

In indoor environment with texture-less and structure-less, the object view lacks a feature point. And better creative or logical alternative scientific solution is to get more useful information of feather points similar to improving resolution rates and to match every got points one by one. In fact based on this concept, the FAST detection algorithm is improved, which can set the determination threshold around the corner points without manual adjustment. For to make use of the feather points,the violent matching was selected as further to treat them. The experiments show the the correctness and usefulness of the method proposed in the paper. The results of this paper show that the adaptability and combination of the algorithm improves its speed and accuracy. Under the premise that the computing power is relatively satisfied, it is an inevitable direction of the algorithm improvement. The algorithm in this paper has been applied to VSAM system and will be further investigated.

References

[1] B. Han and L. Xu, MLC-SLAM: Mask loop closing for monocular SLAM, *International Journal of Robotics and Automation,* 37(1), 2022, 107–114.

[2] S. Badalkhani, R. Havangi, and M. Farshad, An improved simultaneous location and mapping for environments, *International Journal of Robotics and Automation,* 36(6), 2021, 374–382.

[3] S. Wen, Z. Wang, J. Chen, L. Manfredi, and Y. Tong, CSLAM system consensus estimation in dynamic communication networks, *International Journal of Robotics and Automation,* 37(2), 2022, 227–235.

[4] Z. Qiguang, P. Yingchun, Y. Mei, and C. Weidong, CEH∞F-SLAM: A robust and jacobian-free solution to SLAM problem, *International Journal of Robotics and Automation,* 34(1), 2019, 808–820.

[5] J. Ni, Y. Chen, K. Wang, and S.X. Yang, An improved vision-based SLAM approach inspired from animal spatial cognition, *International Journal of Robotics and Automation,* 34(5), 2019, 491–502.

[6] R.G. von Gioi, J. Jakubowicz, J. Morel, and G. Randall, LSD: A fast line segment detector with a false detection control, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 32(4), 2010, 722–732.

[7] D.Van Opdenbosch, and E. Steinbach, Collaborative visual SLAM using compressed feature exchange, *IEEE Robotics and Automation Letters, 4*(1), 2019, 57–64.

[8] R.A. Newcombe, S. Izadi, and O. Hilliges, KinectFusion: Real-time dense surface mapping and tracking, *Proc. 2011 10th IEEE International Symposium on Mixed and Augmented Reality,* Basel, 2011, 127–136.

[9] X. Zhu, Q. Cao, and Y. Yang, An improved kinectFusion 3D reconstruction algorithm, *Robot, 36*(2), 2014, 129–136.

[10] H. Zhou, D. Zou, and L. Pei, StructSLAM: Visual SLAM with building structure lines, *IEEE Transactions on Vehicular Technology, 64*(4), 2015, 1364–1375.

[11] A. Pumarola, A. Vakhitov, and A. Agudo, PL-SLAM: Real-time monocular visual SLAM with points and lines, *Proc. 2017 IEEE International Conf. on Robotics and Automation,* Singapore, 2017, 4503–4508.

[12] R. Gomez-Ojeda, F. Moreno, and D. Zuñiga-Noël, PL-SLAM: A stereo SLAM system through the combination of points and line segments, *IEEE Transactions on Robotics, 35*(3), 2019, 734–746.

[13] T. Lee, C. Kim, and D.D. Cho, A monocular vision sensor-based efficient SLAM method for indoor service robots, *IEEE Transactions on Industrial Electronics, 66*(1), 2019, 318–328.

[14] Q. Fu, H. Yu, and L. Lai, A robust RGB-D SLAM system with points and lines for low texture indoor environments, *IEEE Sensors Journal, 19*(21), 2019, 9908–9920.

[15] B. Triggs, P.F. Mclauchlan, and R.I. Hartley, Bundle adjustment-a modern synthesis, *Proc. International Workshop on Vision Algorithms: Theory and Practice,* Berlin, Heidelberg, 1999, 298–372.

[16] K. Liu, H. Sun, and P. Ye, Research on bundle adjustment for visual SLAM under large-scale scene, *Proc. 2017 4th International Conf. on Systems and Informatics,* Hangzhou, 2017, 220–224.

[17] A. Eudes, M. Lhuillier, and S. Naudet-Collette, Fast odometry integration in local bundle adjustment-based visual SLAM, *Proc. 20th International Conf. on Pattern Recognition,* Istanbul, 2010, 290–293.

[18] P. Dollár, R. Appel, and S. Belongie, Fast feature pyramids for object detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(8), 2014, 1532–1545.

[19] N.K. Ratha, J.H. Connell, and R.M. Bolle, An analysis of minutiae matching strength, *Proc. International Conf. on Audio-and Video-Based Biometric Person Authentication,* Berlin, Heidelberg, 2001, 223–228.

[20] M. Muja and D.G. Lowe. Fast matching of binary features, *Proc. 2012 Ninth Conf. on Computer and Robot Vision,* Toronto, ON, 2012, 404–410.

[21] S. Choi, T. Kim, and W. Yu, Performance evaluation of RANSAC family, *Journal of Computer Vision, 24*(3), 1997, 271–300.

[22] W. Lin, M.M. Cheng, and J. Lu, Bilateral functions for global motion modeling, *Proc. European Conf. on Computer Vision,* Cham, 2014, 341–356.

Biographies

Ding Wang received the M.S. degree in electric drive and automation from Yanshan University, Qinhuangdao, China. He is currently with the Xinjiang College of Science & Technology, China.

Jing Wang received the M.S. degree in computer application technology from the University of Chinese Academy of Sciences, Beijing, China. She is with the North China Institute of Aerospace Engineering, China.

Huan Wang received the M.S. degree from Beihang University, Beijing, China. He is currently with the Xinjiang College of Science & Technology, China.