



# EMC+GD\_C: circle-based enhanced motion consistency and guided diffusion feature matching for 3D reconstruction

Zhenjiao Cai<sup>1</sup> · Sulan Zhang<sup>1</sup> · Jifu Zhang<sup>1</sup> · Xiaoming Li<sup>1</sup> · Lihua Hu<sup>1</sup> · Jianghui Cai<sup>1</sup>

Received: 4 December 2023 / Accepted: 17 April 2024  
© The Author(s) 2024

## Abstract

Robust matching, especially the number, precision and distribution of feature point matching, directly affects the effect of 3D reconstruction. However, the existing methods rarely consider these three aspects comprehensively to improve the quality of feature matching, which in turn affects the effect of 3D reconstruction. Therefore, to effectively improve the quality of 3D reconstruction, we propose a circle-based enhanced motion consistency and guided diffusion feature matching algorithm for 3D reconstruction named EMC+GD\_C. Firstly, a circle-based neighborhood division method is proposed, which increases the number of initial matching points. Secondly, to improve the precision of feature point matching, on the one hand, we put forward the idea of enhancing motion consistency, reducing the mismatch of high similarity feature points by enhancing the judgment conditions of true and false matching points; on the other hand, we combine the RANSAC optimization method to filter out the outliers and further improve the precision of feature point matching. Finally, a novel guided diffusion idea combining guided matching and motion consistency is proposed, which expands the distribution range of feature point matching and improves the stability of 3D models. Experiments on 8 sets of 908 pairs of images in the public 3D reconstruction datasets demonstrate that our method can achieve better matching performance and show stronger stability in 3D reconstruction. Specifically, EMC+GD\_C achieves an average improvement of 24.07% compared to SIFT-based ratio test, 9.18% to GMS and 1.94% to EMC+GD\_G in feature matching precision.

**Keywords** 3D reconstruction · Feature matching · Circle-based neighborhood · Enhance motion consistency · Guided diffusion

## Introduction

Three-dimensional (3D) reconstruction [1–4] is an important research content in the field of computer vision, which has

been widely used in medical systems [5], autonomous navigation [6], aeronautical and remote sensing measurements [7], and industrial automation [8]. Because the number, precision, and distribution range of feature matching directly affect the effect of 3D reconstruction, feature matching becomes the key step in 3D reconstruction, which is mainly manifested as: (1) the reconstructed 3D model is prone to point cloud sparsity and reconstruction interruption, when the number of feature point matching is limited; (2) if the error rate of feature point matching is too high, noise point cloud and model distortion are likely to occur in 3D reconstruction; (3) when the matching distribution of feature points is concentrated, it will easily lead to instability of the 3D reconstruction model. Therefore, highly robust feature matching methods play an important role in improving the quality of 3D reconstruction.

To improve the performance of feature point matching, previous work mainly focused on improving the distinguishing ability of descriptors, and significant effects have been

---

✉ Sulan Zhang  
zhsulan@126.com

✉ Jianghui Cai  
jjianghui@tyust.edu.cn

Zhenjiao Cai  
caizj224@163.com

Jifu Zhang  
jifuzh@sina.com

Xiaoming Li  
lixm351@163.com

Lihua Hu  
sxyhlh@126.com

<sup>1</sup> School of Computer Science and Technology, Taiyuan University of Science and Technology, Shanxi 030024, China

achieved, such as SIFT [9, 10], ORB [11], SURF [12], and A-SIFT [13], etc. Although these methods have been relatively mature, they still have the following disadvantages: (1) the number of feature point matches obtained by the nearest neighbor is large but the error rate is high; (2) feature point matching in the form of nearest neighbor and second nearest neighbor ratio [9] can improve the quality of feature point matching, but fewer matching pairs are obtained [14]; (3) these methods lack consideration of the distribution of matching feature points [15]; (4) these methods are only improved from the perspective of the descriptors [12, 16], and it is difficult to distinguish true and false matches. It can be seen that these matching methods still have many shortcomings in the number, precision, and distribution of feature point matching, and the establishment of a highly robust feature matching model is still a research topic worth challenging.

Motion consistency [17, 18] generally follows the principle of “similar features sharing consistent motions”. By incorporating smoothness constraints into feature point matching, it can effectively distinguish true and false matches and increase the number of correct matches. At present, many researchers have done a lot of research on feature point matching algorithms based on motion consistency [19–29]. Among them, The Grid-based Motion Statistics (GMS) [23, 24] is an efficient grid-based motion statistic. It converts the motion smoothing constraint into a statistic that eliminates false matches, which can improve the quality of matches while increasing the number of feature matches. However, it divides the neighborhood by an equally divided grid, which tends to cause the absence of correct neighborhoods and thus reduces the number of feature point matches. Considering the rotation invariance of circles, we propose a circle-based neighborhood partitioning method. Specifically, we take each matching point as the neighborhood center, and take the circle region within a certain range as the neighborhood of the matching point, thereby increasing the number of feature matches. In addition, the GMS [23, 24] only uses a single threshold  $\alpha$  to judge true and false matching, which is easy to lead to sensitive matching results and strict requirements for the value of  $\alpha$ , and the value of  $\alpha$  is different for different matching scenarios, making it difficult to choose the optimal  $\alpha$ . Therefore, to solve this problem, we present a new idea of Enhancing Motion Consistency (EMC), that is, based on motion consistency, a threshold  $\beta$  ( $\alpha < \beta$ ) is added to strengthen the discriminative conditions for distinguishing true and false matches, thereby removing false matches caused by repetitive textures with high similarity and reducing the proportion of false matches. Moreover, to further improve the precision of feature point matching, we use Random Sample Consensus (RANSAC) [30, 31] to eliminate outliers, which greatly alleviate the problems of sparse

point cloud, reconstruction interruption, noisy point cloud, and model distortion of the 3D reconstruction model.

Furthermore, to avoid the concentrated distribution of matching points in the local area of the image and improve the stability of the 3D model, we propose a Guided Diffusion (GD) idea. It mainly includes two steps of guided matching and motion consistency. In the step of guided matching, the distribution range of feature point matching is expanded by conducting epipolar geometric guided matching on the high-precision matching set. In the step of motion consistency, a small-range motion consistency constraint is applied to the diffusion matching set, which eliminates the false matching caused by the weak point-line epipolar constraint relationship. We verify the effectiveness of the proposed method by conducting experiments on multiple datasets and comparing it with existing methods. Specifically, our EMC+GD\_C achieves an average improvement of 9.18% compared to GMS, 1.94% to EMC+GD\_G, and 24.07% to the SIFT-based ratio test in feature matching precision.

In summary, to improve the reconstruction effect of 3D reconstruction, we propose a circle-based enhanced motion consistency and guided diffusion feature matching algorithm named EMC+GD\_C for 3D reconstruction from three aspects of the number, precision, and distribution of feature points matching. More concretely, the contributions of this study can be summarized as following:

(1) Propose a circle-based neighborhood division method. Instead of the existing methods of dividing the neighborhood, we combine the characteristics of the rotation invariance of the circle, and take the circle neighborhood within a certain range of each matching point as our neighborhood. This method can effectively avoid the problem of missing neighborhood caused by image rotation and increase the number of feature point matching.

(2) An EMC strategy is proposed to determine whether the correspondences in circular neighborhood pairs satisfy the motion consistency or not, which can effectively improve the precision of the initial matching set.

(3) Propose a novel GD method that fuses guided matching and motion consistency constraints, which effectively expands the distribution range of feature point matching.

(4) Apply our EMC+GD\_C algorithm to 3D reconstruction. It not only effectively alleviates the problems of point cloud sparseness, reconstruction interruption, point cloud noise, and model distortion during 3D reconstruction, but also improves the stability of the 3D model.

The remainder of the paper is organized as follows. Some related work is discussed in Sect. [Related work](#). Section [EMC+GD\\_C](#) introduces the algorithm in detail. In Sect. [Experiments](#), we analyze and discuss the experimental results. The conclusion is drawn in Sect. [Conclusions and limitations](#).

## Related work

As a key process of recovering 3D models from images, feature [32, 33] matching has been studied by more and more researchers from the perspective of 3D reconstruction [7, 34]. For example, Hu et al. [35] used the SIFT matching algorithm to find matching pixel points in two corresponding digital images and used the midpoint on the common perpendicular of a non-coplanar line to estimate 3D points. SIFT-based point matching is robust to obvious translation, rotation, and scaling, but the number of correctly matched points usually decreases sharply with the increase of the angle between the viewpoints. To solve this problem, Stumpf et al. [36] used the affine invariant extension (A-SIFT) of the SIFT detector to provide more reliable matching and obtain a more accurate 3D model than SIFT. Liu et al. [37] used the ORB feature detection algorithm to extract image features and realized a fast 3D reconstruction method for indoor, simple, and small-scale static environments, which has good accuracy, robustness, real-time, and flexibility. However, these algorithms that improve feature point matching from the perspective of descriptors are difficult to distinguish between true and false matches, and even if they can filter out some outliers, they will affect the number of correct matches. Therefore, in the process of 3D reconstruction, point cloud sparsity and reconstruction interruption are easy to occur.

In recent years, many researchers have conducted a lot of research on feature point matching algorithms based on motion consistency. For example, Maier et al. [20] proposed a guided matching method based on statistical optical flow, which constrained the search space by using spatial statistics for matching and filtering corresponding small subsets. Although this method has better performance in terms of processing time, it may filter out very small dynamic objects during the statistical optical flow process, thus reducing the robustness of the algorithm. Wang et al. [21] used Density Maximization (DM) and defined a good local smooth neighborhood to avoid noise, which significantly improved the precision of matching, and can handle outliers and many-to-many object matching at the same time. However, it is limited to sparse feature matching, which makes the calculation cost higher and the implementation complicated. Lin et al. [14] used Bilateral Functions (BF) to reduce the false filtering of correct matches. They calculated the global matching consistency function and enabled motion discontinuity on the object boundary. But it runs slowly and is easily affected by the repetitive structure, which reduces the matching precision. To solve this problem, Lin et al. [22] improved based on BF, they combined BF with RANSAC to form a wide baseline matcher that achieves high precision and recall rates in challenging scenarios. These techniques all used the consistency constraint of matching distribution to separate true and false matches. Although they can alleviate

the interruption of reconstruction, the complicated formulas lead to complex smoothness constraints, which are difficult to understand and precise. Yang et al. [26] proposed a dynamic-scale grid structure into the mismatch removal stage to reduce the time complexity of neighborhood construction, and proposed a Gaussian-based weighted scoring strategy to combine the descriptor matching stability with geometric consistency. GMS [23, 24] used a statistical matching constraint, which is simpler and easier to understand. It adopted the idea of an equipartition grid to judge true and false matching, and encapsulated the consistency of motion in the grid, which improved the quality of matching while increasing the number of feature matching. However, when there are highly similar repeated structures, it will cause a large number of persistent mismatches, which are manifested as sparse and noisy point clouds in 3D reconstruction.

## EMC+GD\_C

Our algorithm framework is shown in Fig. 1, which is mainly divided into two steps: High-precision matching and GD-based uniform matching.

Step 1. High-precision matching. Based on the neighborhood division of circles and using the EMC idea, a certain number of initial matching set named IMS is obtained, and then RANSAC is implemented on the IMS to further improve the feature matching precision, thus obtaining a Dependable Matching Set named DMS.

Step 2. GD-based uniform matching. The idea of guided matching and motion consistency is applied to GD, and the

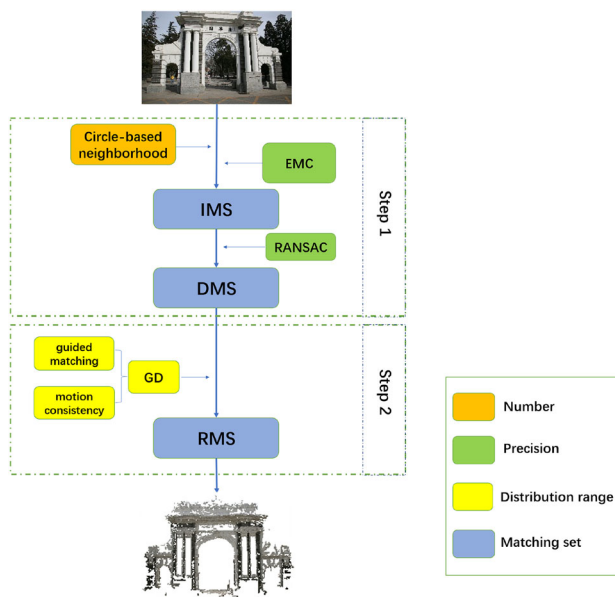
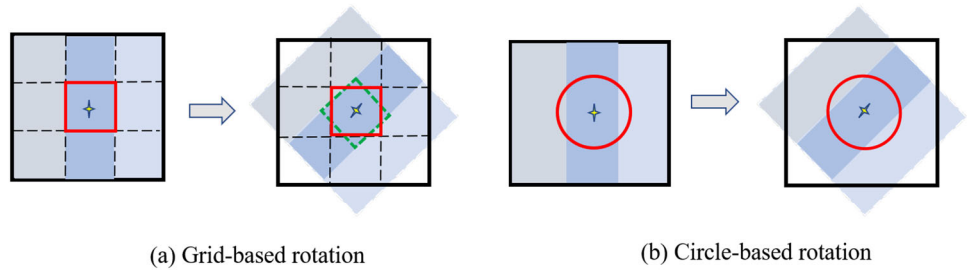


Fig. 1 Algorithm framework

**Fig. 2** Neighborhood comparison chart. The red regions in the figure are the neighborhood actually divided. The green dashed box is the correct neighborhood



Robust Matching Set named RMS with uniform distribution is obtained.

### High precision matching

#### Circle-based neighborhood division

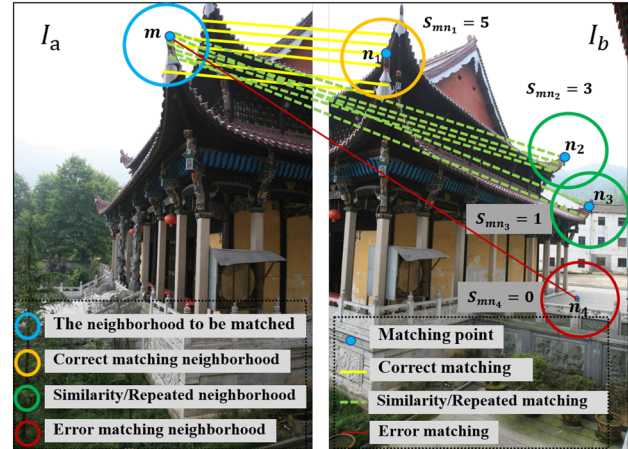
For the adjacent image pairs of two 3D scenes to be matched, a fast and concise ORB [11] algorithm with high computational efficiency was used to detect feature points and descriptors, and after brute force matching, a Brute Force Matching Set named BFMS is obtained, but it contains a lot of mismatches. The idea of motion consistency can be used to make a good judgment of true and false matches. Among them, GMS [23, 24] is the most successful matching algorithm for applying motion consistency. It uses an even grid to divide the neighborhood, but when the image is rotated, GMS [23, 24] only considers several discrete values, and when other rotation angles occur, it is easy to cause the loss of some feature points, as shown in Fig. 2a.

Considering the rotation invariance of the circle, we divide the neighborhood by drawing circles, and only divide the neighborhood of each matching point to avoid the loss of feature points and local neighborhood, as shown in Fig. 2b. Specifically, by traversing each matching feature point in BFMS, and using each matching feature point as the center to find all the feature points within a certain range (neighborhood radius  $r$ ), the feature points in this range are the matching pairs in the corresponding neighborhood. In this paper, we set  $r$  as the radius of the circumcircle of the nine grids in GMS [23, 24] and normalized it to  $r = 0.1$ .

#### EMC-based initial matching

To determine the neighborhood score of each neighborhood, after dividing the neighborhood, we score the matching pairs in each neighborhood  $j$ , as shown in Fig. 3. Assuming that the total number of matching pairs in each neighborhood is  $P_j$ , then the neighborhood score  $S_{mn}$  of each matching pair  $\{m_j, n_j\}$  in the neighborhood can be expressed as:

$$S_{mn} = |P_j| - 1 \tag{1}$$



**Fig. 3** Schematic diagram of neighborhood score

where  $-1$  denotes removing the original matching pair and  $j$  represents the index value of each neighborhood.

Since the number of features in each neighborhood is different, we also take the feature reference value  $\tilde{N}$  in the neighborhood into account, and according to the neighborhood score  $S_{mn}$  of each matching pair  $\{m_j, n_j\}$  in the neighborhood, the distribution of matching pairs  $D_{mn}$  is divided into high support (H) and low support (L):

$$D_{mn} = \begin{cases} H, & \text{if } S_{mn} > \alpha \times \tilde{N} \\ L, & \text{otherwise} \end{cases} \tag{2}$$

where  $S_{mn}$  is the neighborhood score of the matched pair  $\{m_j, n_j\}$ , and  $\tilde{N}$  is the feature reference value in the neighborhood. By reference to the GMS [23, 24] algorithm, we make  $\tilde{N} = |N|^{\frac{1}{2}}$ , where  $N$  is the total number of features in the neighborhood.  $\alpha$  is a threshold parameter.

In the algorithm of GMS [23, 24], only the threshold  $\alpha$  is used to judge the true and false matching, that is, the matching pair belonging to H is the correct matching pair (C), and the matching pair belonging to L is the wrong matching pair (E), which is easy to cause the matching result to be sensitive and the value of  $\alpha$  is strictly required. Especially in some special scenes, such as the repetitive texture regions with high similarity, it is not enough to distinguish between correct and

incorrect matching only by  $\alpha$ . Because repeated textures have similar motion and neighborhoods to the correct matching pairs, if  $\alpha$  is set too small, they will be mistaken for correct.

As shown in Fig. 3, the neighborhood scores of similar matching points  $n_2$  and  $n_3$  are:  $S_{mn_2} = 3$ ,  $S_{mn_3} = 1$ , when a small threshold  $\alpha$  is used, they are largely divided into the correct matching area. Therefore, to avoid the shortcomings of the single threshold  $\alpha$  being sensitive to the matching results and strict requirements on the threshold  $\alpha$ , we added a threshold  $\beta$  ( $\alpha < \beta$ ), giving an idea of EMC to remove false matches caused by high similarity repetitive textures, and improve the precision of feature matching. The specific definition is as follows:

**Definition 1** Enhanced Motion Consistency named EMC: Based on dividing the matching distribution  $D_{mn}$  into H and L in formula (2), the matching conditions of true and false are further subdivided, that is, H is divided into C, similar or repeated matching (R) and L are divided into E, which improves the standard conditions for correct matching. The calculation formula is:

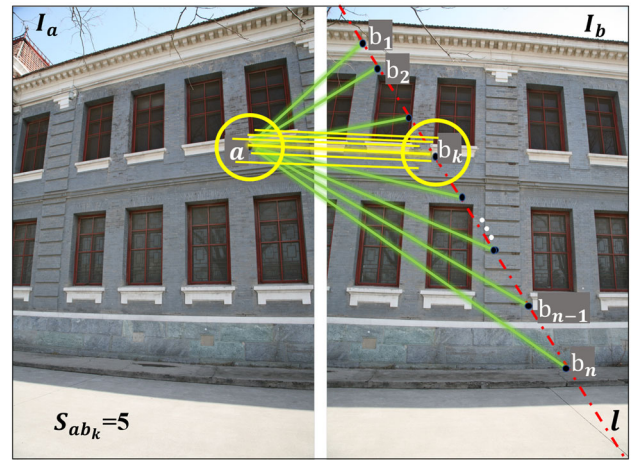
$$D_{mn} = \begin{cases} H = \begin{cases} C, & \text{if } S_{mn} > \beta \times \tilde{N} \\ R, & \text{if } \alpha \times \tilde{N} < S_{mn} \leq \beta \times \tilde{N}, \alpha < \beta \end{cases} \\ L = E, & \text{if } S_{mn} \leq \alpha \times \tilde{N} \end{cases} \quad (3)$$

where  $C$  is the correct matching,  $R$  is the similar or repeated matching,  $E$  is the error matching,  $S_{mn}$  is the neighborhood score of the matched pair  $\{m_j, n_j\}$ ,  $\tilde{N}$  is the feature reference value in the neighborhood, where  $\tilde{N} = |N|^{\frac{1}{2}}$ ,  $N$  is the total number of features in the neighborhood,  $\beta$  is a threshold parameter and  $\alpha < \beta$ . At this time, the set that correctly matches  $C$  is obtained and denoted as IMS:

$$IMS = set\{C\} \quad (4)$$

**RANSAC initial matching optimization**

For image feature point matching, RANSAC is used to fit the model from the matching correspondence and remove outliers [22, 24]. Although the process of obtaining the IMS in this paper also attempts to remove the wrong correspondence, it cannot fit the model as an estimator based on RANSAC. Therefore, the IMS provides a higher-quality corresponding hypothesis for RANSAC. By implementing the RANSAC outlier removal scheme for them, the precision of feature point matching is further improved, and then a DMS is obtained. Since the scenes used for 3D reconstruction are generally large parallax scenes, we fit the DMS to the more precise fundamental matrix model  $F1$ . The fitted accurate model helps to better implement the guided matching, and the more correct matches are obtained, the smaller the error of the matching pair after diffusion.



**Fig. 4** Schematic diagram of motion consistency constraint. Any feature point  $a$  in the image can be constrained by the epipolar constraint to obtain several corresponding points  $b_1 \dots b_n$  on the epipolar line  $l$ , which will cause error matching (green line). In this paper, we apply the motion consistency in a small range to the epipolar constraint (yellow circle in the figure) and get the correct matching point  $b_k$

**GD-based uniform matching**

Although the precision of DMS is high, on one hand, expanding the threshold  $\beta$  would remove the part of the correct matching within a small range; on the other hand, due to the influence of EMC and neighborhood division, similar features are usually located in the same neighborhood, making it easy for correct matching pairs to be concentrated and distributed in the part of the image. To obtain more matches, we combine guided matching and motion consistency constraints, and propose a guided diffusion idea named GD, which is mainly divided into two steps: Guided matching and Motion consistency constraint.

**Guided matching**

Applying the idea of guided matching [38], based on DMS, more matching pairs can theoretically be guided by the model  $F1$ . As a result, we implement guided matching for all BFMS, retain all matching pairs that meet the conditions, and obtain a large number of corresponding points, namely the guided matching set, which is specifically defined as follows:

**Definition 2** Guided Matching Set named GuideMS: By judging whether BFMS conforms to model  $F1$ , that is, verifying whether the distance from BFMS to model  $F1$  is within a certain error range, all matching pairs that meet the conditions are GuideMS. The calculation formula defined as:

$$GuideMS = |D(BFMS, F1)| < t \quad (5)$$

where  $D(x, y)$  is the distance from  $x$  to  $y$ ,  $|\cdot|$  is the absolute value operation,  $t$  is the distance parameter, and based on experience, we set  $t = 10$ .

### Motion consistency constraint

Since the fundamental matrix is a weak point-to-line epipolar constraint relationship (see Fig. 4), there are a large number of false matches in GuideMS (see Fig. 5b). These false matches do not always satisfy motion consistency, therefore, based on GuideMS, we perform a small range of motion consistency constraints, and then obtain the matching result after diffusion, that is, the RMS, which is specifically defined as definition 3:

**Definition 3** Robust Matching Set named RMS: The GuideMS obtained in formula (5) is judged for the motion consistency in a small range  $\gamma$ , and all matching pairs that meet the conditions are retained as the robust matching set RMS, and described by (6):

$$RMS = GuideMS \leftarrow S_{mn} > \gamma \times \tilde{N} \quad (6)$$

where  $\leftarrow$  is the verification operation,  $S_{mn}$  is the neighborhood score of the matching pair  $\{m_j, n_j\}$ ,  $\tilde{N}$  is the feature reference value in the neighborhood,  $\gamma$  is the threshold parameter.

As shown in Fig. 5a, the matching pair before diffusion has higher precision, but it is usually only concentrated in a certain area of the image. After guided matching, we obtained GuideMS, a guided matching set widely scattered on the image. However, due to the point-to-line epipolar constraint, the error rate of the matching pair is very high, as shown in Fig. 5b. After a small range of motion consistency constraints, we improved the precision of feature matching while ensuring the dispersion of matching, as shown in Fig. 5c (red area).

The EMC+GD\_C algorithm in this paper is shown in Algorithm 1.

## Experiments

To verify the effectiveness of our method, we perform it on the datasets of Strecha et al. [39] and the 3D reconstruction datasets provided by the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA) [40]. The introduction of the dataset is shown in Table 1. We compare the SIFT-based ratio test (SIFT [9, 10]+ratio test) and GMS [23, 24] algorithm respectively, in particular, in order to see the advantages of the circle-based neighborhood partition method in this paper, we compare the grid-based enhanced

**Algorithm 1** EMC+GD\_C: Circle-based Enhanced Motion Consistency and Guided Diffusion Feature Matching for 3D Reconstruction.

**Input:** The image pair to be matched:  $I_a, I_b$ .

**Output:** F2 and RMS.

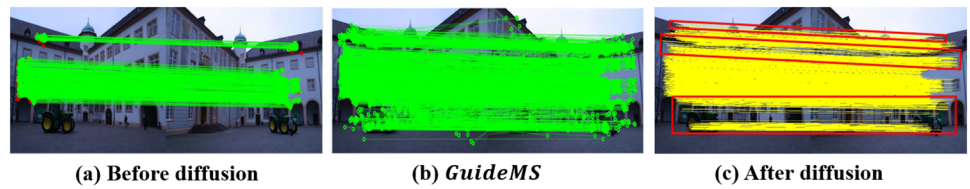
1. Detect feature points and descriptors;
2. Implement brute force matching to obtain the BFMS;
3. Neighborhood Division: Take each matching point as the center and  $r$  as the radius to draw a circle as the neighborhood of each matching point;
4. Traverse the neighborhood of each matching point;
5. **for**  $g = 1 : n$  **do**
6. Calculate the neighborhood score  $S_{mn}$  of each matching pair  $\{m_j, n_j\}$  in the neighborhood by formula (1);
7. **if**  $S_{mn} > \alpha \times \tilde{N}$  **then**
8. Divide the matching distribution:  $D_{mn} = H = C$  or  $R$
9. **if**  $S_{mn} > \beta \times \tilde{N}$ ,  $\beta > \alpha$  **then**
10. Divide the matching distribution:  $D_{mn} = C$
11. **end**
12. **else**
13. Divide the matching distribution:  $D_{mn} = L = E$
14. **end**
15. **end**
16. Keep all the correct matches named C, and get the IMS by formula (4);
17. Perform the RANSAC filtering interior point operation on the IMS to obtain the DMS and the model  $F1$ ;
18. Implement guided matching on BFMS from formula (5) and obtain the GuideMS;
19. Implement the motion consistency constraint within a small range  $\gamma$  for GuideMS to obtain the RMS in formula (6);
20. Fit RMS to a high-precision model  $F2$  for pose estimation in 3D reconstruction.

motion consistency and guided diffusion feature matching algorithm named EMC+GD\_G. That is, we keep our algorithm steps, and only replace the circle-based neighborhood division method in this paper with the grid-based neighborhood division method in GMS [23, 24]. Eight experiments are conducted: (1) evaluation IMS; (2) evaluation DMS; (3) evaluation RMS; (4) feature matching precision ( $P$ ) comparison; (5) feature matching number comparison; (6) feature matching time comparison; (7) quantitative evaluation; (8) integrate our method into the VisualSFM system [41] and use Yasutaka Furukawa's PMVS/CMVS toolchain for dense reconstruction. The experimental environment is a computer with an Intel (R) Core (TM) i9-9900X processor, Dell GeForce RTX 2080TI graphics card.

### Evaluate IMS

In this section, we evaluate the IMS. The experimental results of some image pairs are shown in Fig. 6, where the yellow box represents the overlapping areas of the image pairs, and the red box represents the mismatch caused by similar or repetitive structures. The IMS obtained by SIFT [9, 10] after nearest neighbor matching (the default threshold is 1.5) is

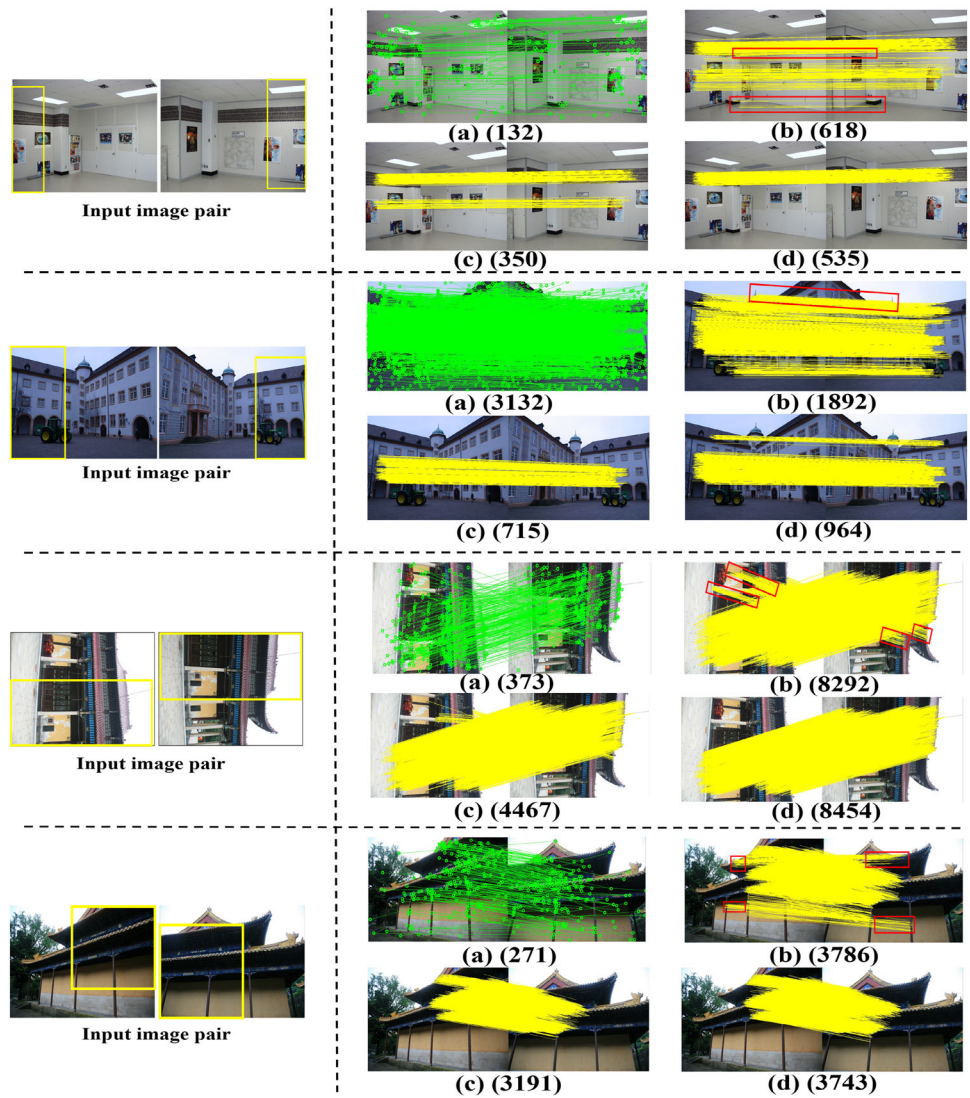
**Fig. 5** Comparison chart before and after diffusion



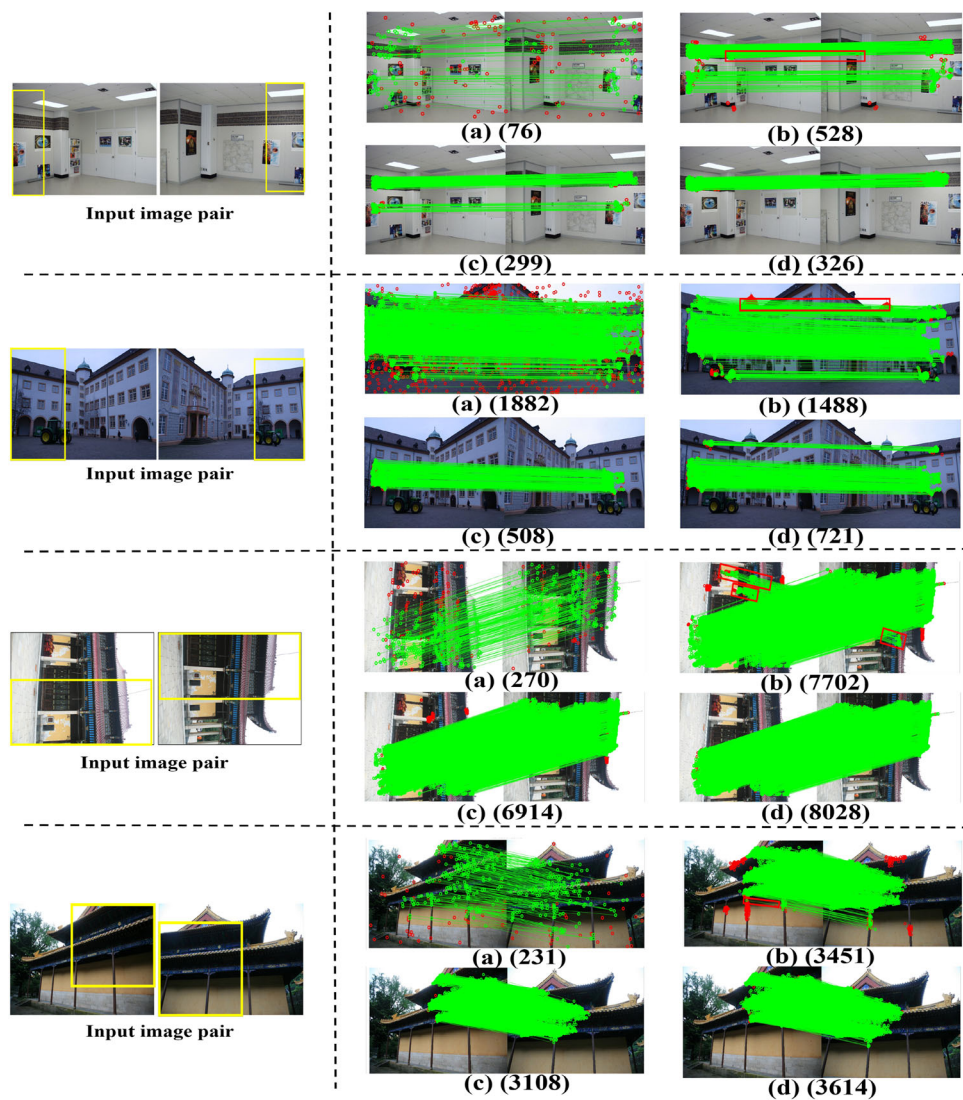
**Table 1** Dataset introduction

Category	Datasets	Image pairs	Image size
1	Castle	56	3072 × 2048
2	Fountain	10	3072 × 2048
3	Herz-Jesu	31	3072 × 2048
4	Shengkelou	102	4368 × 2912
5	Qinghuaxuetang	193	4368 × 2912
6	Qinghuamen	68	4368 × 2912
7	Zhantan	158	4368 × 2912
8	Fayu	290	4368 × 2912
		A total of 908 pairs	

**Fig. 6** A comparison chart of IMS. The number of matches is shown in brackets. **a** SIFT [9, 10]+ratio test; **b** GMS [23, 24]; **c** EMC+GD\_G; **d** EMC+GD\_C (ours)



**Fig. 7** A comparison chart of DMS. The number of matches is shown in brackets. **a** SIFT [9, 10]+ratio test; **b** GMS [23, 24]; **c** EMC+GD\_G; **d** EMC+GD\_C (ours)



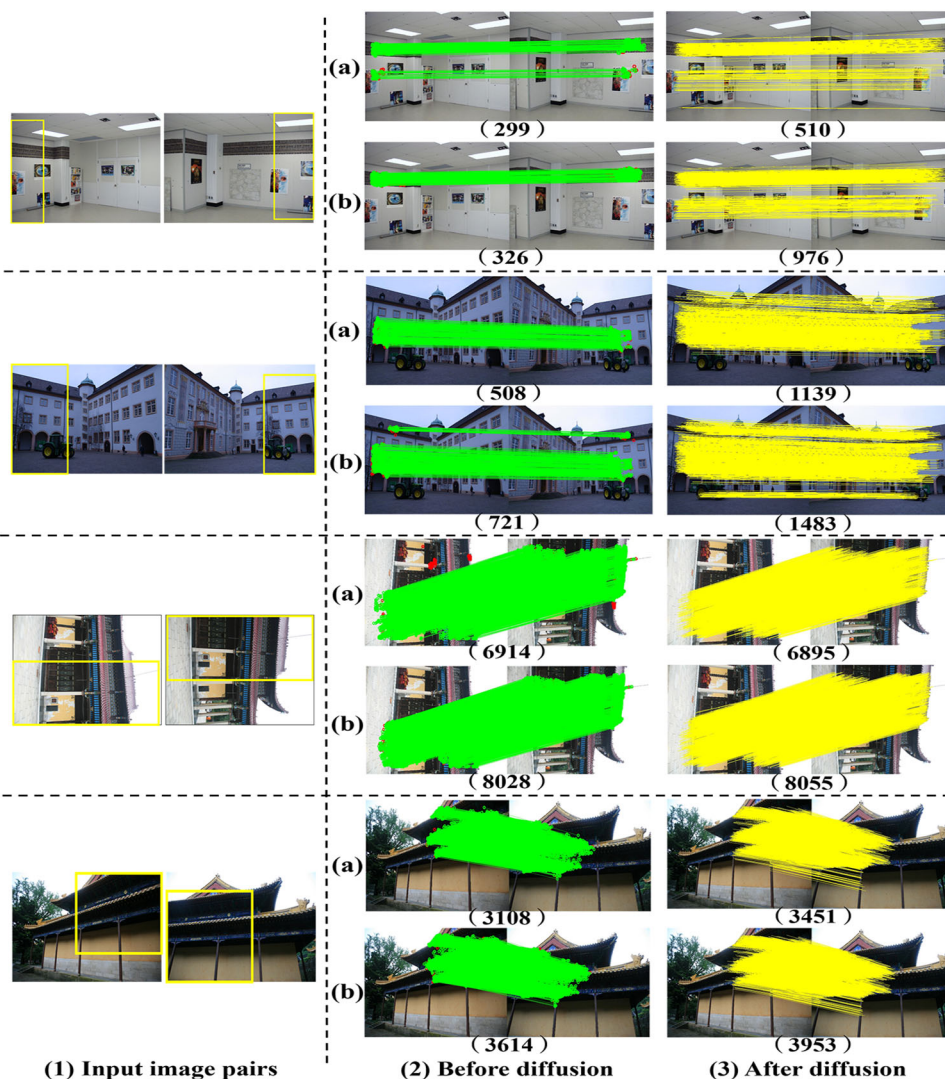
shown in Fig. 6a, and the matching result is seriously unstable. Figure 6b shows the matching results obtained by GMS [23, 24] using the default threshold of 6, and a large number of false matches are obtained in similar or repeated regions. Figure 6c and d are the IMS results obtained by EMC+GD\_G and EMC+GD\_C algorithms using  $\beta=11$ , respectively. As can be seen from the figure, our method reduces the false matches obtained from similar or repeated regions through the idea of EMC, and although the number of matches is relatively reduced, it has higher accuracy than other tested algorithms. In addition, our method obtains more IMS than EMC+GD\_G, which benefits from the circle-based neighborhood partitioning method in this paper. Therefore, our method is more favorable for matching optimization.

## Evaluate DMS

In this section, we evaluate the DMS. The matching results of some image pairs are shown in Fig. 7, where the yellow box represents the overlapping area of the image pair, the red box represents the mismatch caused by the similar or repetitive structure, and the red dot represents the abnormal value filtered out after RANSAC. Figure 7a is the SIFT-based ratio test (SIFT [9, 10]+ratio test) algorithm, and a DMS containing a large number of false matches is obtained. Figure 7b shows that GMS [23, 24] is optimized by RANSAC and has a certain number of mismatches (red box), most of which are caused by similar or repetitive structures. Figure 7c and d achieve relatively high-precision matches, albeit with a drop in the number of matches. Due to the high precision of IMS in this paper, our method also exhibits high precision on DMS, and thanks to the circle-based neighborhood division, our method obtains more matches than EMC+GD\_G.



**Fig. 8** A comparison chart of RMS. The number of matches is shown in brackets. **a** EMC+GD\_G; **b** EMC+GD\_C (ours)



### Evaluate RMS

In this section, we evaluate the RMS. Some image pairs are shown in Fig. 8, where the yellow box represents the overlapping area of the image pair, and the red dots represent the outliers filtered by the RANSAC algorithm. The comparison diagram of EMC+GD\_G algorithm before and after diffusion is shown in Fig. 8a, where  $\gamma=6$ , and the comparison diagram of our EMC+GD\_C algorithm before and after diffusion is shown in Fig. 8b, where  $\gamma=6$ . Matching pairs before diffusion are easily concentrated in a local area of the image, and matching pairs after diffusion are more evenly distributed. In addition, compared with the EMC+GD\_G algorithm, our method has more matching pairs after diffusion, and the distribution is more uniform.

### Feature matching precision ( $P$ ) comparison

The calculation of feature matching  $P$  in this paper uses the calculation formula proposed by Mikołajczyk et al. [38]:

$$1 - P = \frac{F}{T + F} \tag{7}$$

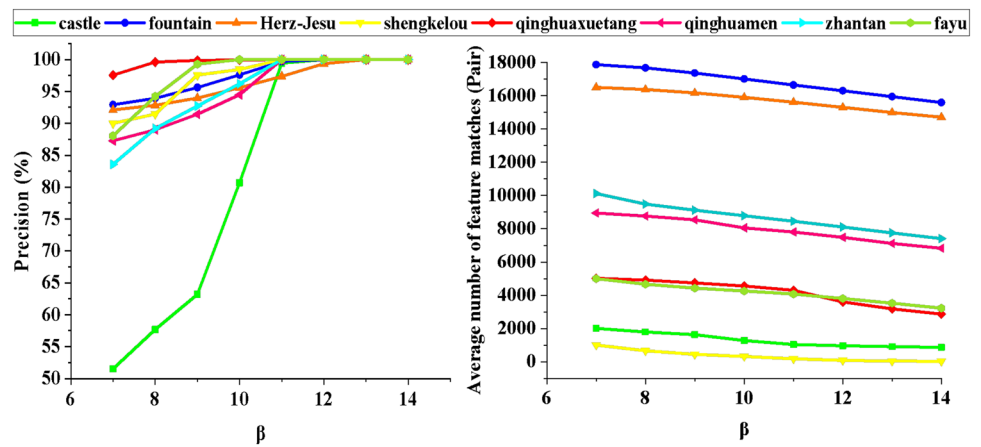
where  $T$  is the number of correct matches;  $F$  is the number of incorrect matches.

### Parametric analysis

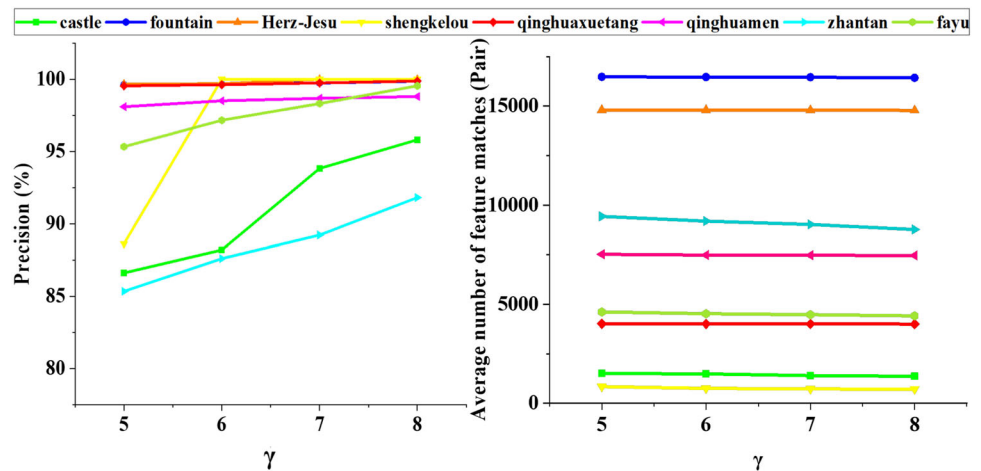
In this section, we analyze the parameters  $\beta$  and  $\gamma$  used in the experiment on eight different data sets [39, 40].

The comparison chart of IMS  $P$  and quantity when  $\beta$  takes different values is shown in Fig. 9. It can be seen from the figure that with the increasing of  $\beta$ , the number of feature matching shows a slow downward trend, and the  $P$  of IMS

**Fig. 9** Comparison of  $P$  and quantity of IMS under different parameters  $\beta$



**Fig. 10** Comparison of  $P$  and quantity under different parameters  $\gamma$



shows a rapid upward trend. Until  $\beta$  increases to 11,  $P$  basically reaches more than 95%, and the subsequent growth is slow. Therefore, considering the  $P$  and the number of feature matching, the final  $\beta$  value is selected as 11.

When performing GD, it is necessary to perform a motion consistency constraint in a small range  $\gamma$  to reduce the mismatch caused by the epipolar constraint. When  $\gamma$  takes different values, the  $P$  and quantity comparison chart of the EMC+GD\_C algorithm in this paper are shown in Fig. 10. As can be seen from the figure, with the increase of  $\gamma$ , the number of feature matching shows a slow decline while  $P$  keeps increasing. When the value of  $\gamma$  is 6–7, the  $P$  begins to show a rapid upward trend, especially for the ‘castle’ dataset. Although the  $P$  still has an upward trend with the increase of  $\gamma$  value, the increase is relatively slow, and the average number of feature matches decreases with the increase of  $\gamma$  value. Therefore, considering both the  $P$  and the number of feature matching, the final  $\gamma$  value range is 6–7.

### The $P$ comparison of different matching methods

In this section, we compare the feature matching  $P$  of different matching methods on 8 sets of 908 pairs of adjacent image data sets [39, 40]. As shown in Fig. 11, our method (using thresholds  $\beta=11$  and  $\gamma=6$ ) and the EMC+GD\_G algorithm have a higher  $P$  than the SIFT-based ratio test algorithm and the GMS algorithm, and our method is the highest. In addition, the average  $P$  of the SIFT-based ratio test, GMS algorithm, EMC+GD\_G algorithm and our method is about 73.75%, 88.64%, 95.88% and 97.82%, respectively. It can be seen that our method is 24.07% higher than the SIFT-based ratio test algorithm, 9.18% higher than the GMS algorithm, and 1.94% higher than the EMC+GD\_G algorithm.

### Feature matching number comparison

In this section, we compare the number of feature matching for different matching methods on the adjacent image data sets [39, 40] in 908. In Fig. 12a, we compare the SIFT-based ratio test (SIFT [9, 10]+ ratio test), GMS [23, 24], EMC+GD\_G, and our method (EMC+GD\_C), respectively.

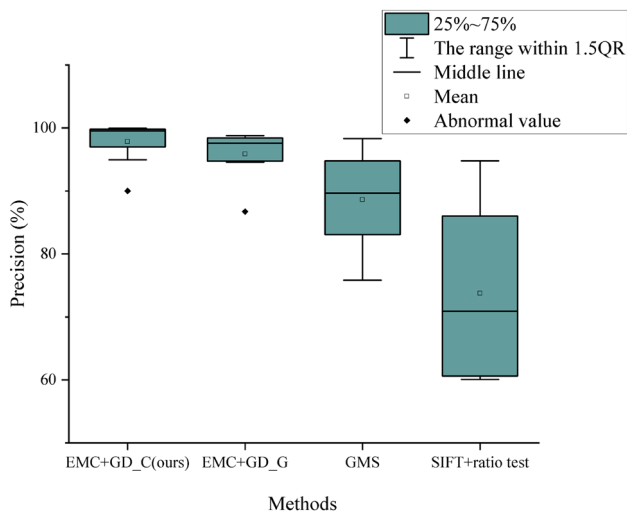


Fig. 11 Comparison chart of  $P$  of different matching methods

As can be seen from the Fig. 12, although the SIFT-based ratio test (SIFT [9, 10]+ ratio test) and GMS [23, 24] algorithm obtained more matching pairs than our method, the matching  $P$  is relatively low (see Fig. 11). In addition, our method and the EMC+GD\_G algorithm have obtained a large number of feature matches while maintaining a high  $P$ , and as shown in Fig. 12b, our method has obtained more matching pairs than the EMC+GD\_G algorithm.

### Computational complexity

In this section, we compare the computational complexity for different matching methods on the adjacent image data sets [39, 40] (a total of 8 groups of 908 pairs). The SIFT-based ratio test (SIFT [9, 10]+ ratio test), GMS [23, 24], EMC+GD\_G, and our method (EMC+GD\_C) are compared in Table 2. As can be seen from the table, the GMS [23, 24] algorithm takes the shortest time, followed by the EMC+GD\_G algorithm, and then the SIFT-based ratio test

(SIFT [9, 10]+ratio test) algorithm, our EMC+GD\_C algorithm takes the longest time.

In summary, our method has the following advantages over the EMC+GD\_G algorithm: (1) higher  $P$ ; (2) more feature matching numbers; (3) more uniform distribution. However, the computational complexity is relatively high, as shown in Table 3.

### Quantitative evaluation

In this section, we verify the data set of Strecha et al. [39], using 91 pairs of sequence images as the test set. The ground truth of this data set includes the projection matrix  $P'$  and the camera internal parameter  $K$ . The rotation matrix  $R_1$  and the translation matrix  $t_1$  can be decomposed by formula (8):

$$P' = K[R_1 \ t_1] \tag{8}$$

The essential matrix  $E$  can be obtained from the fundamental matrix  $F2$  by formula (9):

$$F2 = K^{-T} \times E \times K^{-1}, \quad E = K^T \times F2 \times K \tag{9}$$

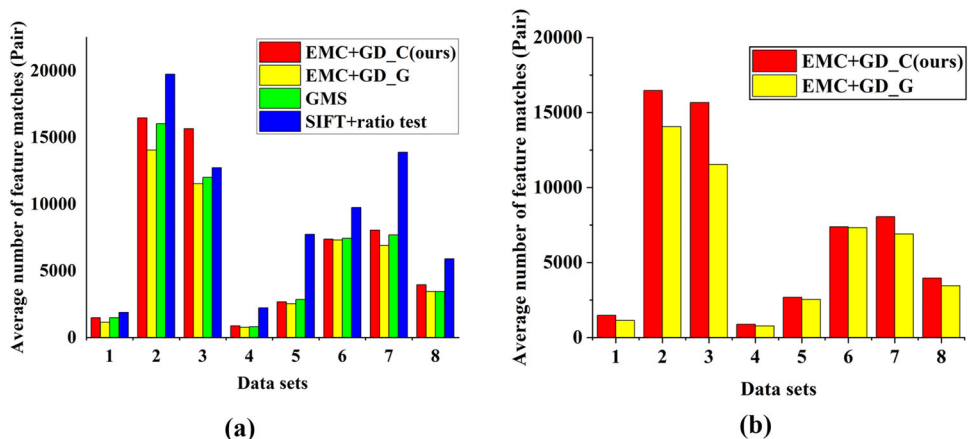
The essential matrix  $E$  can be decomposed into the rotation matrix  $R_2$  and the translation matrix  $t_2$  (see formula (10)):

$$E = [t_2] \times R_2 \tag{10}$$

Since the translation matrix  $t$  is equal when it differs by a scale factor, we do not use an absolute quantity but convert it to the form of a vector, that is, the angular error value  $x^o$  of rotation and translation is used as a variable. Pose estimators often give very incorrect solutions (or crashes) when they fail, so the average error is meaningless. To avoid this situation, we use the success percentage  $SP$  [22] (see formula (11)):

$$SP(x) = \frac{\sum_{i=1}^{N-1} (e_i(R/t) \leq x^o)}{N} \tag{11}$$

Fig. 12 Comparison chart of number of different matching methods



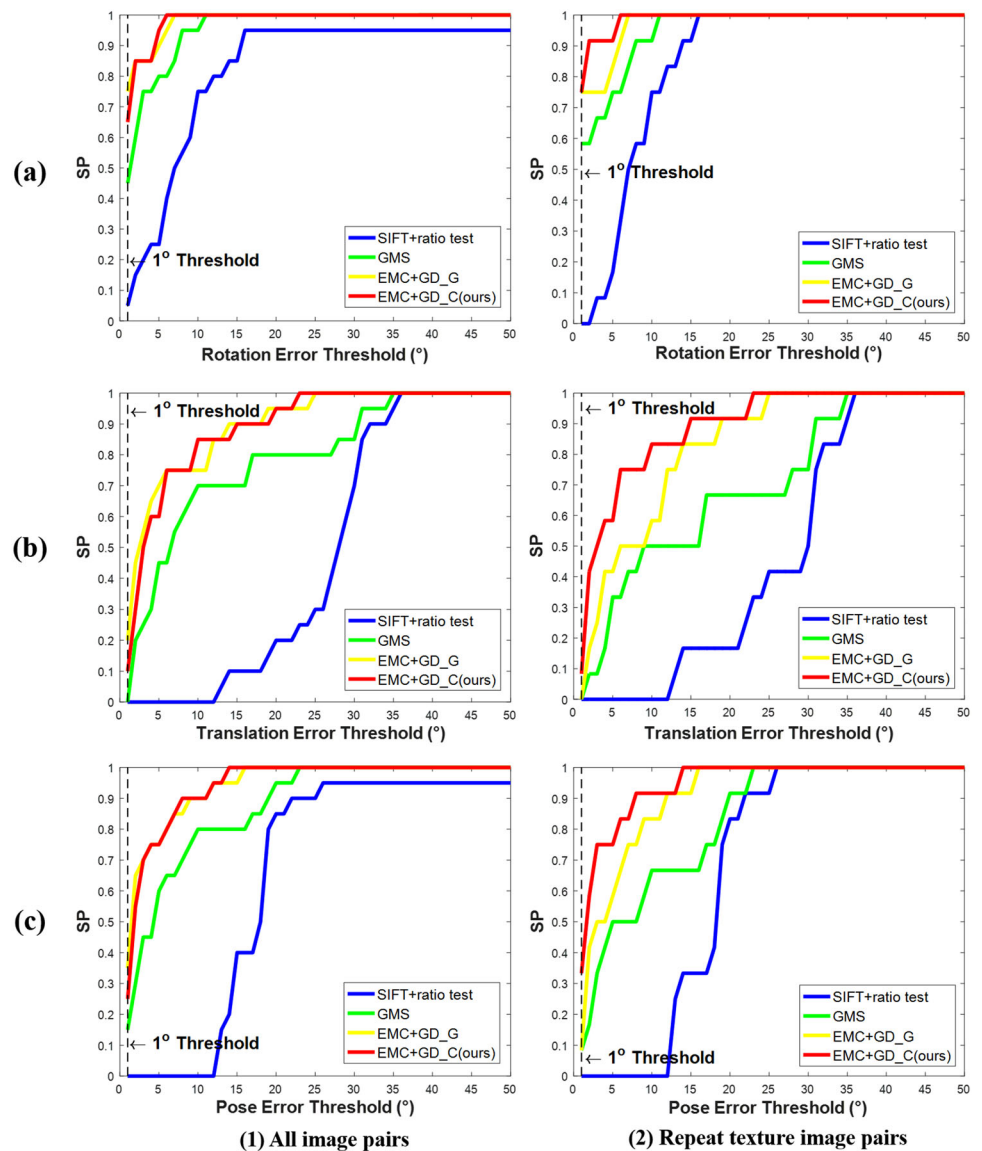
**Table 2** Computational complexity of different matching methods

Method	Average match time/minute							
	1	2	3	4	5	6	7	8
SIFT+ratio test	1.7058	2.1424	2.4042	4.0812	3.6254	3.0667	2.2822	3.5414
GMS	0.3624	0.9713	0.5172	0.7556	0.8232	0.7428	0.6569	0.4817
EMC+GD_G	0.3464	1.4945	1.8092	1.3695	1.5904	1.4295	1.5231	1.0314
EMC+GD_C(ours)	3.9333	16.7721	11.5278	15.1933	18.8500	9.6433	13.9400	6.6167

**Table 3** EMC+GD\_G and EMC+GD\_C algorithm comparison

Method	$P$	Number	Distributed	Time complexity
EMC+GD_C(ours)	High	Much	Relatively more uniform	High
EMC+GD_G	Low	Little	Relatively uniform	Low

**Fig. 13**  $SP$  comparison chart



where  $N$  represent the number of all image pairs,  $e_i(R/t)$  represents rotation error  $e_i(R)$  or translation error  $e_i(t)$  between the  $i$ th image and the  $i$ th plus 1 image, which are respectively expressed as formula (12):

$$e_i(R) = |R_{i1} - R_{i2}|, \quad e_i(t) = |t_{i1} - t_{i2}| \quad (12)$$

where  $|\cdot|$  is the absolute value operation,  $i$  is the  $i$ th image,  $R_1$  and  $t_1$  are obtained in formula (8),  $R_2$  and  $t_2$  are obtained in formula (10).

In Fig. 13a and b, we compare the rotation  $SP$  and translation  $SP$  of the SIFT-based ratio test (SIFT [9, 10]+ratio test), GMS [23, 24], EMC+GD\_G algorithm, and our EMC+GD\_C algorithm, respectively. It can be seen from the Fig. 13a and b, the curve drawn by the threshold  $x^o$  of the rotation or translation error and the  $SP$  is a non-decreasing curve, and the smaller the error value, the higher the precision of the pose estimation. Therefore, the  $1^\circ$  threshold is a value we are more concerned about. Our method has a rela-

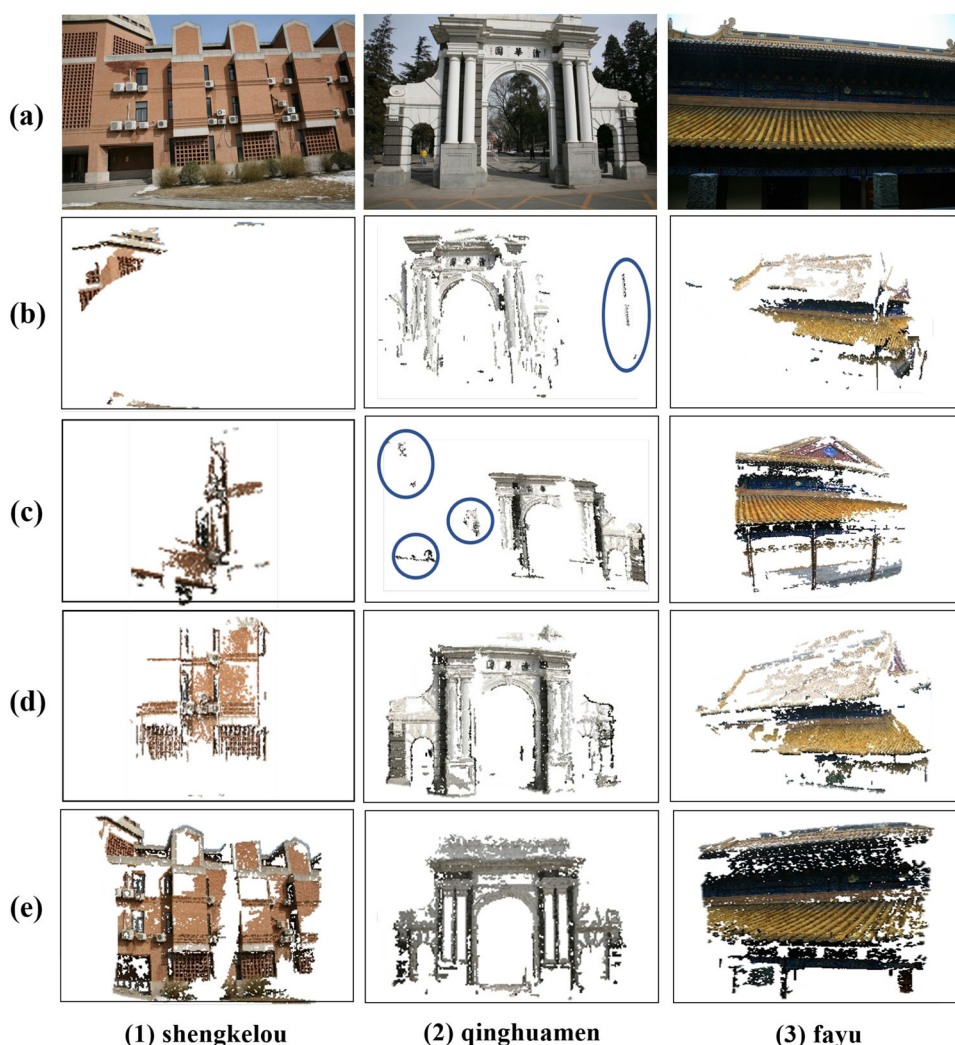
tively higher  $SP$  in both rotation and translation, especially for image pairs with repeated textures.

Since the rotation  $SP$  and translation  $SP$  comparison graphs are both non-decreasing curves, to comprehensively see the superiority of our method, we take the average value of the rotation error and the translation error as the threshold of the pose estimation error (see Fig. 13c). It can be seen from the Fig. 13c, our method has a higher  $SP$  than other test algorithms in pose estimation, especially for image pairs with repeated textures.

### Comparisons of 3D reconstruction results

In this section, we discuss the application of our method in 3D reconstruction and verify it on the adjacent image data sets [39, 40] in 908. The main application reconstruction system is VisualSFM [41]. Some of the dense reconstruction results are shown in Fig. 14a is the input image sequence, Fig. 14b is the point cloud reconstructed by the VisualSFM [41] sys-

**Fig. 14** Dense reconstruction of three scenes. **a** Input image pair; **b** SIFT [9, 10]+ratio test; **c** GMS [23, 24]; **d** EMC+GD\_G; **e** EMC+GD\_C (ours)



tem (SIFT is used by default), which is sparse and easily interrupted. The reasons are as follows: (1) In a 3D scene with large parallax, there are not enough matching points, resulting in inaccurate model fitting, and the same scene is mistaken as a different scene, which will cause the scene to be interrupted; (2) when the number of correct matching points is too small, the point cloud will be sparse. Compared with Fig. 14b, GMS [23, 24] in Fig. 14c obtained more point clouds after dense reconstruction, but the  $P$  of the matching pair is not high, and the model estimation is inaccurate, resulting in noisy point clouds and model distortion. Compared with Fig. 14b and c, the EMC+GD\_G algorithm in Fig. 14d improves the reconstruction quality. In Fig. 14e, our EMC+GD\_C algorithm has obtained better point cloud quantity, reconstruction quality, and reconstruction effect.

## Conclusions and limitations

In this paper, we present a circle-based enhanced motion consistency and guided diffusion feature matching algorithm (EMC+GD\_C) for 3D reconstruction. We divide the neighborhood by drawing circles to increase the number of feature matches. The proposed EMC idea efficiently improves the precision of the IMS. A DMS and a more precise model are obtained based on the RANSAC. By combining guided matching and motion consistency, the idea of GD is proposed, which expands the distribution range of feature point matching. Experiments demonstrate that our method has better performance in the number, precision, and distribution of feature matching, and can achieve better reconstruction effects in the number of point clouds, reconstruction quality, and reconstruction stability.

This paper has gradually improved the quality of feature point matching through three steps of initial matching, reliable matching and robust matching, but it will increase the time complexity of the algorithm. In addition, this paper divides the neighborhood by drawing circles, which increases the number of initial matching sets, but experiments show that it will increase the time complexity of the algorithm. Therefore, in the future, we will consider further optimizing the algorithm by introducing the idea of ‘parallel’ or ‘integration’ to reduce the complexity of the algorithm.

**Acknowledgements** This study was supported by the Natural Science Foundation of Shanxi Province, China (Grant No. 202103021224285) and the National Natural Science Foundation of China (Grant Nos U1931209).

**Data availability** The datasets generated during and/or analysed during the current study are available in the dataset provided by Strecha et al. and the 3D reconstruction datasets provided by the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), [<https://documents.epfl.ch/>

[groups/cv/cvlab-unit/www/data/multiview/index.html](https://groups/cv/cvlab-unit/www/data/multiview/index.html)] and [<http://vision.ia.ac.cn/data>]

## Declarations

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Tian L, Cheng X, Honda M et al (2023) Multi-view 3D human pose reconstruction based on spatial confidence point group for jump analysis in figure skating. *Complex Intell Syst* 9(1):865–879
2. Li Z, Oskarsson M, Heyden A (2022) Detailed 3D human body reconstruction from multi-view images combining voxel super-resolution and learned implicit representation. *Appl Intell* 52:6739–6759
3. Yang B, Wang S, Markham A et al (2020) Robust attentional aggregation of deep feature sets for multi-view 3D reconstruction. *Int J Comput Vision* 128(1):53–73
4. Devi PRS, Baskaran R (2021) SL2E-AFRE: Personalized 3D face reconstruction using autoencoder with simultaneous subspace learning and landmark estimation. *Appl Intell* 51:2253–2268
5. Migliori S, Chiastra C, Bologna M et al (2020) Application of an OCT-based 3D reconstruction framework to the hemodynamic assessment of an ulcerated coronary artery plaque. *Med Eng Phys* 78:74–81
6. Yang JT, Kang ZZ, Zeng LP et al (2021) Semantics-guided reconstruction of indoor navigation elements from 3D colorized points. *ISPRS J Photogramm Remote Sens* 173:238–261
7. Zhu Q, Wang Z, Hu H et al (2020) Leveraging photogrammetric mesh models for aerial-ground feature point matching toward integrated 3D reconstruction. *ISPRS J Photogramm Remote Sens* 166:26–40
8. Bitzidou M, Chrysostomou D, Gasteratos A (2012) Multi-camera 3D object reconstruction for industrial automation. In: *IFIP Int Conference Adv Prod Manag Syst* 526–533
9. Lowe DG (2014) Distinctive image features from scale-invariant keypoints. *Int J Comput Vision* 60(2):91–110
10. Muja M, Lowe DG (2014) Scalable nearest neighbor algorithms for high dimensional data. *IEEE Trans Pattern Anal Mach Intell* 36(11):2227–2240
11. Rublee E, Rabaud V, Konolige K et al (2012) ORB: An efficient alternative to SIFT or SURF. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp 2564–2571
12. Bay H, Tuytelaars T, Gool LV (2006) SURF: speeded up robust features. In: *Proceedings of European Conference on Computer Vision*, pp 404–417

13. Morel JM, Yu GS (2009) ASIFT: a new framework for fully affine invariant image comparison. *SIAM J Imag Sci* 2:438–469
14. Lin WYD, Cheng MM, Lu J et al (2014) Bilateral functions for global motion modeling. In: *Proceedings of European Conference on Computer Vision*, pp 341–356
15. Tan X, Sun C, Sirault X et al (2015) Feature matching in stereo images encouraging uniform spatial distribution. *Pattern Recognit* 48(8):2530–2542
16. Bellavia F, Tegolo D, Valenti C (2014) Keypoint descriptor matching with context-based orientation estimation. *Image Vision Comput* 32(9):559–567
17. Lin WY, Cheng MM, Shuai Z et al (2013) Robust non-parametric data fitting for correspondence modeling. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp 2376–2383
18. Scholefield A, Ghasemi A, Vetterli M (2020) Bound and Conquer: improving triangulation by enforcing consistency. *IEEE Trans Pattern Anal Mach Intell* 42(9):2321–2326
19. Lipman Y, Yagev S, Poranne R et al (2014) Feature matching with bounded distortion. *ACM Trans Graph* 33(3):1–14
20. Maier J, Humenberger M, Murschitz M et al (2016) Guided matching based on statistical optical flow for fast and robust correspondence analysis. In: *Proceedings of European Conference on Computer Vision*, pp 101–117
21. Wang C, Wang L, Liu LQ (2015) Density maximization for improving graph matching with its applications. *IEEE Trans Image Process* 24(7):2110–2123
22. Lin WY, Liu SY, Jiang NJ et al (2016) RepMatch: robust feature matching and pose for reconstructing modern cities. In: *Proceedings of European Conference on Computer Vision*, pp 562–579
23. Bian JW, Lin WY, Matsushita Y et al (2017) GMS: grid-based motion statistics for fast, ultra-robust feature correspondence. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 2828–2837
24. Bian JW, Lin WY, Liu Y et al (2020) GMS: grid-based motion statistics for fast, ultra-robust feature correspondence. *Int J Comput Vision* 128(6):1580–1593
25. Lin WY, Wang F, Cheng MM et al (2018) CODE: coherence based decision boundaries for feature correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34–47
26. Yang L, Huang Q, Li X et al (2022) Dynamic-scale grid structure with weighted-scoring strategy for fast feature matching. *Appl Intell* 52(9):10576–10590
27. Wang LB, Chen BB, Xu P et al (2020) Geometry consistency aware confidence evaluation for feature matching. *Image Vision Comput* 103:103984
28. Nie YY, Hu LH, Zhang JF et al (2020) Feature matching based on grid and multi-density for ancient architectural images. *J Comput Aided Design Comput Graph* 32(3):437–444
29. Ma JY, Zhao J, Jiang JJ et al (2019) Locality preserving matching. *Int J Comput Vision* 127(5):512–531
30. Fischler M, Bolles R (1987) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 24(6):381–395
31. Guo HL, Xia GB, Yan Y (2020) A preference-statistic-based data representation for robust geometric model fitting. *Chinese J Comput* 7(43):1199–1214
32. Xiao Z, Tong H, Qu R et al (2023) CapMatch: semi-supervised contrastive transformer capsule with feature-based knowledge distillation for human activity recognition. *IEEE Trans Neural Networks Learn Syst*. <https://doi.org/10.1109/TNNLS.2023.3344294>
33. Xiao Z, Xing H, Zhao B et al (2023) Deep contrastive representation learning with self-distillation. *IEEE Trans Emerg Topics Comput Intell* 8(1):3–15
34. Lai B, Liu W, Wang C et al (2022) 2D3D-MVPNet: Learning cross-domain feature descriptors for 2D–3D matching based on multi-view projections of point clouds. *Appl Intell* 52(12):14178–14193
35. Hu YF (2011) Research on a three-dimensional reconstruction method based on the feature matching algorithm of a scale-invariant feature transform. *Math Comput Modell* 54(3–4):919–923
36. Stumpf A, Malet JP, Allemand P et al (2013) Robust affine-invariant feature points matching for 3D surface reconstruction of complex landslide scenes. In: *EGU General Assembly*, pp. EGU2013–6203
37. Liu SM, Zhu WQ, Zhang CQ et al (2017) 3D reconstruction of indoor scenes using RGB-D monocular vision. *Microcomput Appl* 1–7
38. Sun K, Tao W, Qian Y (2020) Guide to Match: multi-layer feature matching with a hybrid gaussian mixture model. *IEEE Trans Multimed* 22(9):2246–2261
39. Strecha C, Hansen WV, Gool LV et al (2008) On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 1–8
40. (2018) National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. Datasets for 3D reconstruction [Online], available: <http://vision.ia.ac.cn/data>
41. Wu CC (2011) VisualSfM: A visual structure from motion system. [Online], available: <http://ccwu.me/vsfm/>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.