

Received January 2, 2020, accepted January 16, 2020, date of publication January 23, 2020, date of current version February 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2969082

Multi-View Triangulation: Systematic Comparison and an Improved Method

JIA CHEN¹, DONGLI WU¹, PENG SONG², FUQIN DENG³,
YING HE⁴, AND SHIYAN PANG¹

¹School of Educational Information Technology, Central China Normal University, Wuhan 430079, China

²École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

³Department of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, China

⁴Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China

Corresponding author: Shiyang Pang (psy2046@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61605054 and Grant 41701389, and in part by the Fundamental Research Funds for the Central Universities under Grant CCNU19QD007 and Grant CCNU19TD007.

ABSTRACT Triangulation is an important task in the 3D reconstruction of computer vision. It seems simple to find the position of a point in 3D space when its 2D perspective projections in multi-view images are given and the corresponding camera projection matrices are known. However, in practice, multiple lines in 3D space do not intersect at one point because of noise. Then how to calculate the optimal 3D point of intersection becomes difficult. While there have been multiple methods trying to solve this problem, there is no systematic comparison between them. In this paper, we reviewed various currently existing variants of triangulation method and compared them through extensive experiments. The speed and accuracy of these methods have been compared using both synthetic and real datasets. We presented the results of experiments and summarized the advantages, limitations, and applicability of these methods so that to provide a guide for users when they need to choose an appropriate triangulation method for their given applications. Moreover, based on above analysis we proposed an improved method which shows better performance.

INDEX TERMS Triangulation, perspective projection, multiple views, 3D reconstruction.

I. INTRODUCTION

Triangulation is an important task in the 3D reconstruction of computer vision [1], [2]. It is usually based on non-rigid structure from motion (NRSFM) [3], [4] or multiple camera systems (MCSs) [5], [6] to obtain the matching points and corresponding projection matrices. For a point \mathbf{X} in 3D space, it is imprinted on the image by the camera, satisfying $\mathbf{x} = \mathbf{P}\mathbf{X}$, where \mathbf{x} is the image point and \mathbf{P} is the camera matrix. Obviously, we can obtain \mathbf{x} if \mathbf{P} and \mathbf{X} are known. While for triangulation, \mathbf{x} and \mathbf{P} of this projection equation are known and we need to find the 3D point \mathbf{X} . In theory, two lines intersect at a point, so we only need two of these equations to figure out the coordinates of a 3D point for the ideal case. That is triangulation. However, the actual situation is not the case. In real life, due to the presence of noise [7], the two lines may not have an intersection, or the intersection may deviate from the true value. For multi-views, this problem is more complicated because multiple lines may

have multiple intersections. Then it becomes a complicated problem to determine the optimal estimation of the real 3D spatial point \mathbf{X} .

Triangulation is a crucial step of the 3D reconstruction and it affects the accuracy of the whole reconstruction result. In addition, triangulation also plays an important role in simultaneous localization and mapping (SLAM) [8], [9]. Therefore, many researchers have studied this issue. Hartley *et al.* [7] proposed the Polynomial method which is aimed at the case when polar line constraint is not satisfied due to noise in the two views. The Polynomial method, also known as the optimal triangulation method, can find the global optimal solution without iteration. This method has good performance, and some researchers [10], [11] have applied this method to 3-view triangulation. However, the Polynomial method is difficult to generalize to multi-view since it is complex to implement and the computational cost is considerable [12].

For multiple views, the simplest method is the Linear method or the Midpoint method [7], [13], [14] that can directly solve the problem. However, these methods usually

The associate editor coordinating the review of this manuscript and approving it for publication was Jinjia Zhou¹.

lead to a large error. Bundle Adjustment (BA) is the most commonly used method at present [15], [16], which first determines the cost function, and then to find the optimal value to minimize the cost function. The definition of the cost function is closely related to the measurement of error. Normally, the measurement of error distance is selected as the L_2 norm, which represents the geometric distance of Euclidean space. However, it is complex and prone to local minimum even in the 2-view case where there could be multiple local minima. Hence, some researchers have explored the application of L_∞ norm. It shows that L_∞ optimization comes down to minimizing a cost function with a single minimum (local or global) on a convex domain [17]. Donné *et al.* [18] proposed the Polyhedron collapse method based on the L_∞ norm. It adopts the L_∞ norm to quantify both the single view reprojection error and the aggregated reprojection error, and it is simpler and faster than previous methods [19], [20]. While taking outliers into account, Zhang *et al.* [21] used the least median squares (LMS) to propose the Q-Sweep method, which improves the robustness of the algorithm.

In this paper, the general methods of triangulation are reviewed, and these methods are compared using synthetic datasets and real datasets. Section 2 discusses these triangulation methods, Section 3 presents the evaluation metrics and datasets used in the experiments, Section 4 compares the experiments and proposes an improved method, and the final section reaches the conclusion.

II. METHODS

In this section, we will introduce triangulation methods from the following four aspects: linear method, midpoint method, L_2 triangulation method and L_∞ triangulation method.

A. LINEAR METHOD

First, suppose that $\mathbf{x}_i = \mathbf{P}_i \mathbf{X}$, where \mathbf{x}_i is the corresponding point of the 3D space point \mathbf{X} on the i -th image and \mathbf{P}_i is its camera matrix. In order to simplify the calculation, we adopt the homogeneous coordinates

$$\mathbf{x}_i = w(u_i, v_i, 1)^T \quad (1)$$

$$\mathbf{P}_i = \begin{bmatrix} \mathbf{p}_i^{1T} \\ \mathbf{p}_i^{2T} \\ \mathbf{p}_i^{3T} \end{bmatrix} \quad (\mathbf{P}_i \in \mathbb{R}^{3 \times 4}) \quad (2)$$

where w is the scale factor so that it does not affect the actual coordinates of the image points.

For the equation $\mathbf{x}_i = \mathbf{P}_i \mathbf{X}$, we can write

$$\begin{aligned} wu_i &= \mathbf{p}_i^{1T} \mathbf{X} \\ wv_i &= \mathbf{p}_i^{2T} \mathbf{X} \\ w &= \mathbf{p}_i^{3T} \mathbf{X} \end{aligned} \quad (3)$$

Since w is a scaling factor, we can eliminate it and then equation (3) can be reformed as

$$\begin{aligned} u_i \mathbf{p}_i^{3T} \mathbf{X} - \mathbf{p}_i^{1T} \mathbf{X} &= 0 \\ v_i \mathbf{p}_i^{3T} \mathbf{X} - \mathbf{p}_i^{2T} \mathbf{X} &= 0 \end{aligned} \quad (4)$$

Since \mathbf{X} is a homogeneous representation of 3D coordinates, we only need to know the values of the two sets \mathbf{x}_i and \mathbf{P}_i to solve it. For multiple views, more than four groups of equations are obtained, and these equations are not equivalent due to the influence of noise. That is to say, there is no exact solution. Instead of finding the exact solution, we can only try to find an approximate solution close to it [15].

Therefore, this problem is equivalent to a least squares problem for solving homogeneous linear equations $\mathbf{A} \mathbf{X} = \mathbf{0}$, i.e.

$$\begin{aligned} \min \|\mathbf{A} \mathbf{X}\| \\ \text{s.t. } \|\mathbf{X}\| = 1 \end{aligned} \quad (5)$$

In other words, the problem turns into a unit singular vector for finding the minimum singular value of \mathbf{A} [22]. We call this method the Linear-Eigen method [7].

However, there is a big error in solving the least square solution of the equation directly. The major reason is that minimizing $\|\mathbf{A} \mathbf{X}\|$ does not make any geometric sense and it is not an error function that minimizes the effect of noise. So, some researchers added a factor measuring the contribution of each equation on this basis and optimized the solution using the iterative method [7]. This method is called the Iterative-Eigen method. However, the Iterative-Eigen method often works longer than the maximum number of iterations, and cannot converge to the optimal solution.

B. MIDPOINT METHOD

In the case of two views, the idea of the midpoint method is to find a common vertical line segment perpendicular to the two rays if two rays have no intersection point in 3D space. Then we can take the midpoint of this line segment as the optimal estimate value of the spatial point \mathbf{X} [7]. However, when extending to multiple views, it is not desirable to look for common perpendicular lines for all rays. So the midpoint method for multiple views is to determine the location of the spatial point \mathbf{X} by minimizing the sum of the squares of the distance from that point to all the light rays [23].

The MVMP (Multiple View Midpoint) method is described in detail in a recent work [24]. Suppose the optical center of the camera is \mathbf{O}_i , and the ray (unit vector) formed by the image point \mathbf{x}_i and the optical center \mathbf{O}_i is \mathbf{b}_i , then the distance from the spatial point \mathbf{X} to ray \mathbf{b}_i can be expressed as

$$d_i = (\mathbf{I} - \mathbf{b}_i \mathbf{b}_i^T)(\mathbf{X} - \mathbf{O}_i) \quad (6)$$

Therefore, in the case of multiple views, the objective function that minimizes the sum of the squares of the distance from the spatial point \mathbf{X} to all the rays is

$$\min d(\mathbf{X}) = \min \sum_{i=1}^N \left\| (\mathbf{I} - \mathbf{b}_i \mathbf{b}_i^T)(\mathbf{X} - \mathbf{O}_i) \right\|^2 \quad (7)$$

Since the minimum value of $d(\mathbf{X})$ is 0, we can make $d(\mathbf{X}) = 0$, i.e.

$$\sum_{i=1}^N \left\| (\mathbf{I} - \mathbf{b}_i \mathbf{b}_i^T)(\mathbf{X} - \mathbf{O}_i) \right\|^2 = 0 \quad (8)$$

This equation is equivalent to

$$\left(\sum_{i=1}^N (\mathbf{I} - \mathbf{b}_i \mathbf{b}_i^T)\right) \mathbf{X} = \left(\sum_{i=1}^N (\mathbf{I} - \mathbf{b}_i \mathbf{b}_i^T) \mathbf{O}_i\right) \quad (9)$$

In this equation, \mathbf{O}_i and \mathbf{b}_i are known, and \mathbf{X} is unknown. It is equivalent to solving a system of triadic equations, which can be easily solved by mathematical methods.

The midpoint method is fast, but when the camera is parallel, the error of this method is too large to be considered as a good method. Therefore, some scholars have improved it on this basis. They believe that the error of the midpoint method mainly comes from the distance from the space point to the camera [24]. Therefore, a penalty factor w can be added to the objective function, so that the objective function becomes

$$\min e(\mathbf{X}) = \min \sum_{i=1}^N \|w_i(\mathbf{X})d(\mathbf{X})\|^2 \quad (10)$$

where, $w_i = 1/\|\mathbf{X} - \mathbf{O}_i\|$. This method is called the IRMP (Iteratively Reweighted MidPoint) method [24].

C. L_2 TRIANGULATION METHOD

In addition to directly figuring out the location of spatial points, more scholars adopt the iterative optimization method based on the Linear-Eigen method. Due to the influence of noise, there is a certain error between the position of the image point and the real position. However, the real value of the image point cannot be known. What we know is that the assumed real image point coordinates can only be obtained through the projection matrix after the spatial point coordinates are known. This is the reprojection error in 3D reconstruction, which is usually defined as

$$\sum_i d(\mathbf{x}_i, \hat{\mathbf{x}}_i) \quad (11)$$

where \mathbf{X}_i is the coordinate of the measured image point, $\hat{\mathbf{x}}_i$ is the coordinate of the image point calculated by the quadratic projection of the spatial point and $d(\bullet)$ is the distance between the two points.

Our goal is to find the point \mathbf{X} that minimizes this cost function, i.e.,

$$\min_{\mathbf{X}} \sum_i \|\mathbf{x}_i - \mathbf{P}_i \mathbf{X}\| \quad (12)$$

which can be further expressed as

$$\min_{\mathbf{X}} \sum_i \left\| \mathbf{x}_i - \frac{\mathbf{p}_i^{1:2} \mathbf{X}}{\mathbf{p}_i^3 \mathbf{X}} \right\|_p \quad (13)$$

where $\mathbf{x}_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix}$, $\|\cdot\|_p$ is the p norm of a matrix, usually the value of p is 1, 2 or ∞ .

Normally, the measurement of error distance is selected as the L_2 norm because it represents the geometric distance of Euclidean space. If we are dealing with the L_2 norm, then (13) can be written as

$$\min_{\mathbf{X}} \sum_i \left(\mathbf{x}_i - \frac{\mathbf{p}_i^{1:2} \mathbf{X}}{\mathbf{p}_i^3 \mathbf{X}} \right)^2 \quad (14)$$

Obviously, it is a problem of nonlinear optimization. The commonly used methods include the gradient descent method, the Newton method, the Levenberg-Marquardt (LM) method, among which the Levenberg-Marquardt method is the most commonly used [25].

D. L_∞ TRIANGULATION METHOD

It exists certain difficulties when take the L_2 norm to define the objective function since it usually gets into the local minimum and the global optimal solution cannot be found. In recent years some scholars explore other methods. Research work [17] by Hartley et al. has inspired many scholars to use L_∞ norm to solve the various geometric problems. This work shows the L_∞ norm can make the cost function significantly easier than the L_2 norm.

When the L_∞ norm is used for triangulation, then two aspects are involved: the single view reprojection error and camera aggregation error. If the L_∞ norm is taken in the single-view reprojection error, i.e.

$$\varepsilon_i \triangleq \max(|u_i - \hat{u}_i|, |v_i - \hat{v}_i|) \quad (15)$$

The value of equation (15) can be converted into

$$\varepsilon_i \triangleq \max(u_i - \hat{u}_i, \hat{u}_i - u_i, v_i - \hat{v}_i, \hat{v}_i - v_i) \quad (16)$$

If the camera aggregation error also takes the L_∞ norm, then the problem is converted to

$$\min_{\mathbf{X}} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_\infty = \min_{\mathbf{X}} \max_i \varepsilon_i \quad (17)$$

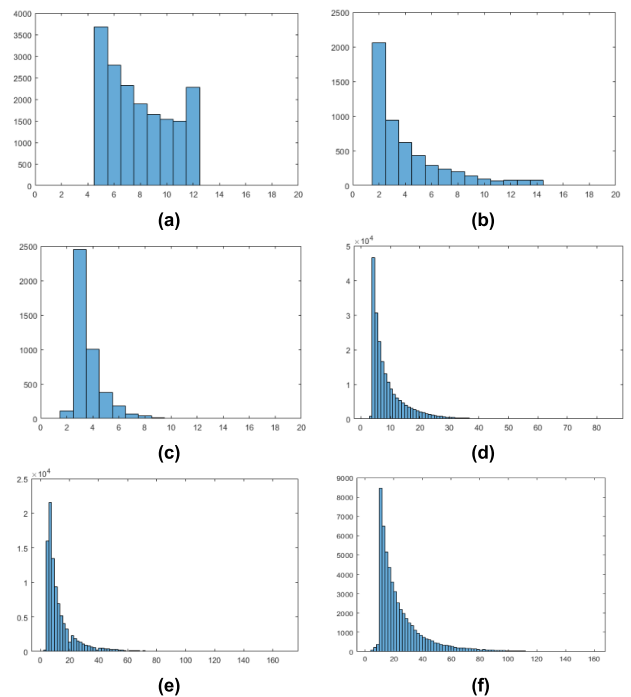


FIGURE 1. The view distribution of the spatial points in each dataset. (a) The Door dataset. (b) The Drinking Fountain dataset. (c) The Gustav Vasa dataset. (d) The Linkoping Cathedral dataset. (e) The UWU dataset. (f) The Notre Dame dataset.

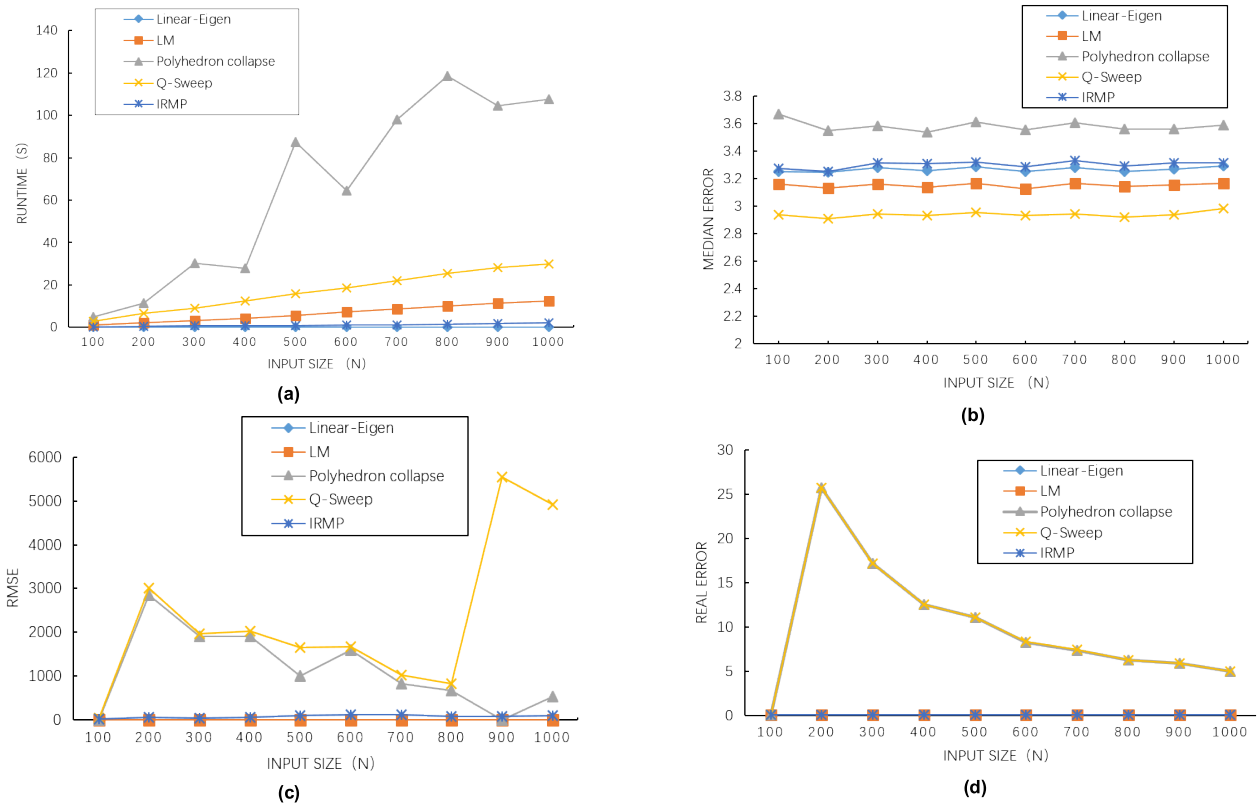


FIGURE 2. Results on synthetic dataset plotted against input size N which is the number of space points. Each point is visible in exactly 36 views. (a) Runtime for discussed methods. (b) Median error for all methods. (c) Root Mean Squared Error for all methods. (d) Real error for all methods.

Some research works [19], [26], [27] used the L_∞ norm on the single view reprojection error while the L_2 norm on the aggregation error. However, more recent works used the L_∞ norm in both aspects. The polyhedron collapse method proposed by Donne *et al.* used the L_∞ norm in both the reprojection error and the aggregation error [18]. Its process only involves unary quadratic equations and some basic algebraic geometry, which is very simple and fast. However, considering the influence of noise, Zhang *et al.* [21] used the method of least median squares (LMS). They proposed the Q-Sweep method which is based on the L_∞ norm and minimizes the median of the reprojection error since the L_∞ norm is easily affected by the outliers. It is

$$\min_{\mathbf{X}} \operatorname{median}_i \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_\infty = \min_{\mathbf{X}} \operatorname{median}_i \varepsilon_i \quad (18)$$

This minimum median method can tolerate 50% of outliers. The Q-Sweep method can effectively deduce the local update step size, and the obtained solution is more accurate.

III. EVALUATION METRICS AND DATASETS

A. EVALUATION METRICS

The classic criterion for evaluating triangulation is to calculate the RMSE (Root Mean Squared Error) of the reprojection error, that is, to calculate the root mean square of the Euclidean distance between the estimated

two-dimensional coordinates of the spatial point reprojection and the two-dimensional coordinates of the image point.

However, the triangulation methods based on the L_∞ norm usually are compared by the convergence error. And due to the difference of the cost functions, it leads to different comparison methods. Here, we uniformly measure the mean value of the median error of the reprojection error, which is consistent with the work [21].

In the synthetic experiment, since the ground truth is known, we take the RMSE of the error compared with ground truth, which is the difference between the results obtained by each method and ground truth. Thereby we estimate whether it can be applied to SLAM or not. The unit for measuring reprojection error in all experiments is pixel.

B. DATASETS

We used six different datasets to compare these methods. The datasets are from [28]–[30] and contain 3 small scenes and 3 large scenes. The small scene datasets include: (1) The Door dataset, which reconstructs 17650 spatial points from 12 images; (2) The Drinking Fountain dataset, which reconstructs 5302 spatial points from 14 images; (3) The Gustav Vasa dataset, which reconstructs 4249 spatial points from 18 images. The large scene datasets include: (1) The Linkoping Cathedral dataset, which reconstructs 202,737 spatial points from 538 images; (2) The UWO dataset, which

reconstructs 97,326 spatial points from 692 images; (3) The Notre Dame dataset, which reconstructs 53,857 spatial points from 761 images.

These datasets were created for 3D reconstruction, so the authors provided relevant parameters for the datasets. In this paper, we use the matching points and camera parameters provided by the datasets. Because each spatial point in the dataset is not captured by all cameras, to better compare the differences in the dataset, we count the view distribution of the spatial points in each dataset. The results are shown in Fig. 1.

IV. EXPERIMENTS

In this paper, we conduct an experimental comparison of the Linear-Eigen method, the LM method, the Polyhedron collapse method, the Q-Sweep method, and the IRMP method by using both synthetic and real datasets. We will compare each method from different aspects: running time (runtime), median error, RMSE and the error with the ground truth. All the experiments were done using MATLAB on a single thread.

A. SYNTHETIC DATASETS EXPERIMENTS

We generate synthetic datasets for triangulation as follows: N space points and M cameras are generated randomly, and all

space points are located in front of the cameras. We obtained the image points through space points and cameras, and then Gaussian noise with a variance of 3 was added to evaluate robustness. For outliers, we add Gaussian noise with a variance of 9 to some image points.

1) NUMBER OF SPACE POINTS

We used 36 cameras and then increased the number of points in space to compare different methods. Fig. 2(a) shows the runtime of the five methods in different numbers of space points. It can be seen that among the five methods, the order of running speed does not change. The Linear-Eigen method is the fastest, followed by the IRMP method, and the Polyhedron collapse method is the slowest. As the number of spatial points increases, the increase of runtime for Q-Sweep method is faster than that for the LM method, and the runtime of Polyhedron collapse method fluctuates greatly. Fig. 2(b) shows the median error of all the methods. The Q-Sweep method committed to minimizing the median error is the best and the Polyhedron collapse method is the worst. All the methods remain basically stable as the number of space points increases. For RMSE of all methods, the LM method works best (see Fig. 2(c)). And Linear-Eigen method and IRMP method work well while the performance of Polyhedron collapse and Q-Sweep method are not stable. When compared

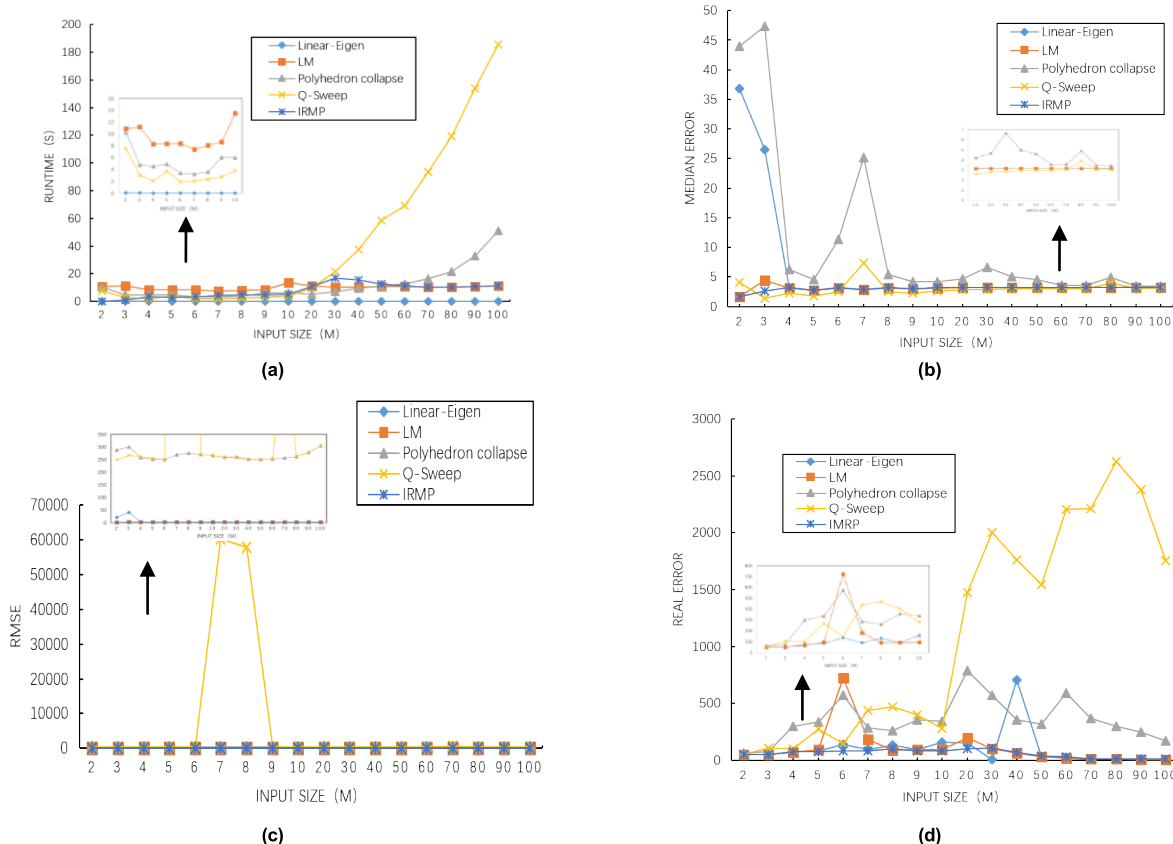


FIGURE 3. Results on synthetic dataset plotted against input size M which is the number of cameras. 1000 space points are visible in all views. Enlarge the detail on the small figures. (a) Runtime for discussed methods. (b) Median error for all methods. (c) Root Mean Squared Error for all methods. (d) Real error for all methods.

with the ground truths, Linear-Eigen, IRMP and LM method have the best effect in Fig. 2(d). And Polyhedron collapse and Q-Sweep method could reduce the gap with the ground truths as the number of points increases.

2) NUMBER OF CAMERAS

We used 1000 space points and then increased the number of cameras one by one to compare different methods. In Fig. 3(a), the Linear-Eigen method is the fastest, and almost not affected by the increasing number of cameras. The LM method has less influence by the number of cameras and maintains a flat growth. The Q-Sweep method is greatly affected by the number of cameras, and as the number of cameras increases, the runtime becomes longer and longer. The Polyhedron collapse method is also affected, but slightly better than the Q-Sweep method. Fig. 3(b) shows the relationship between the median error of the five methods and the number of cameras. As the number of cameras increases, the median error decreases. The most effective method is Q-Sweep. While the Polyhedron collapse method has the worst effect. For RMSE, the Polyhedron collapse and Q-Sweep method are much larger than the others, and the Q-Sweep method has experienced great fluctuations. Compared with the ground truths, the IRMP method works best. The LM method is superior to the Polyhedron collapse method and relatively stable, while the Q-Sweep method fluctuates greatly.

3) OUTLIER RATIO

As the proportion of outliers increases, the runtime remains basically stable except for the IRMP method, while the median error of all methods will increase (see Fig. 4). For the Q-Sweep method, its effect is close to that of LM when the outlier ratio is lower than 50%. However, when the outlier ratio exceeds 50%, its error will increase sharply. In comparison with the ground truths, LM, IRMP and Linear-Eigen method are better, while the fluctuation of Q-Sweep method is very large. And the RMSE of Q-Sweep method is the worst.

B. REAL DATASETS EXPERIMENTS

We used six different datasets described in Section 3.2. Since the ground truth is not always available we compared these methods from the runtime, RMSE, and median error.

Table 1 shows the experimental results. Fig. 5 shows the reconstruction effect of the four methods without the Linear-Eigen method. We do not present the Linear-Eigen method because both RMSE and median error of the Linear-Eigen method are large and the obtained reconstruction effect is bad.

From the perspective of runtime, the fastest way to rebuild a small scene is the Linear-Eigen method, followed by the IRMP method, and the slowest is the LM method. However, when the number of views is large, the Linear-Eigen method is sometimes not the fastest, because the coefficient matrix becomes larger as the number of views increases, and the calculation of singular value decomposition becomes more

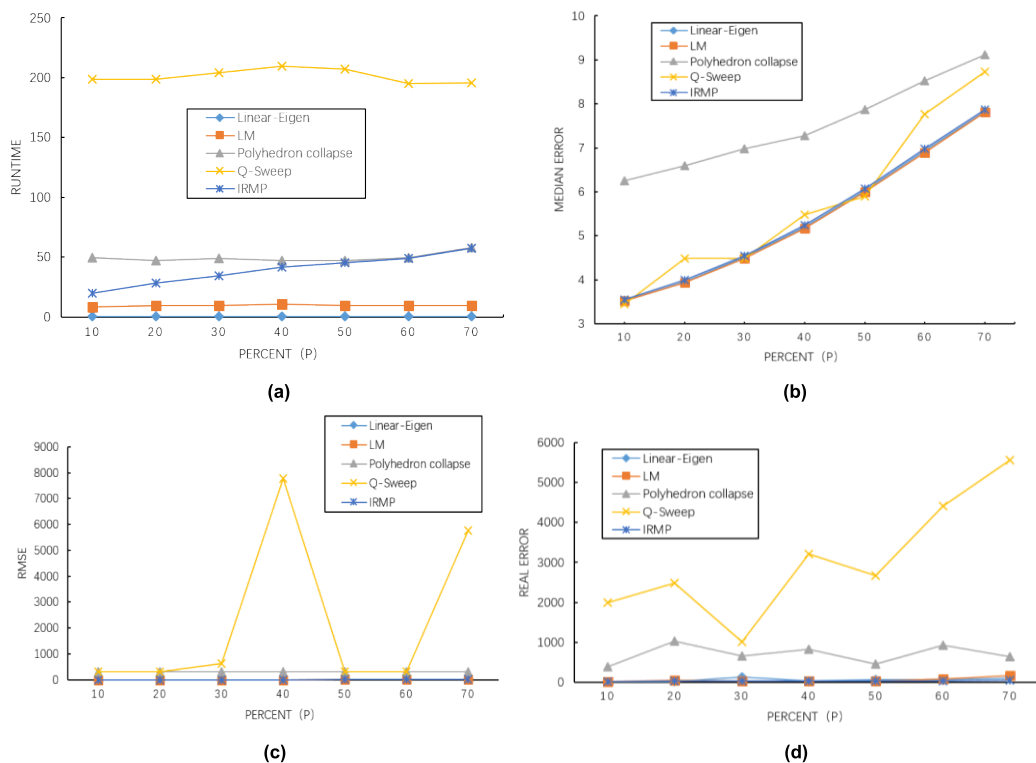


FIGURE 4. Results on synthetic dataset plotted against input size P which is the proportion of outliers. 1000 space points are visible in 100 views. (a) Runtime for discussed methods. (b) Median error for all methods. (c) Root Mean Squared Error for all methods. (d) Real error for all methods.

TABLE 1. Results on the real datasets.

Dataset	Algorithm	Time(s)	RMSE	Median Error
Door (12, 17650)	Linear-Eigen	1.0533	217.6700	228.2720
	LM	242.4006	15.3282	14.8434
	IRMP	8.7919	0.2451	0.3805
	Polyhedron collapse	34.8080	556.7969	0.440
	Q-Sweep	26.101	556.8225	0.187
Drinking Fountain (14, 5302)	Linear-Eigen	0.2105	1465.2	1971.6
	LM	66.1285	58.9165	111.3934
	IRMP	1.4926	1782.9	3565.2
	Polyhedron collapse	8.738	494.1296	0.339
	Q-Sweep	5.419	494.1442	0.182
Gustav Vasa (18, 4249)	Linear-Eigen	0.1688	479.2903	625.6212
	LM	50.6944	0.5089	1.6799
	IRMP	1.2332	4.6972	9.9186
	Polyhedron collapse	8.189	474.5662	0.747
	Q-Sweep	4.181	474.5965	0.348
Linkoping Cathedral (538, 202737)	Linear-Eigen	97.3	1715.0	1543.3
	LM	3652.4	17.5360	167.8164
	IRMP	899.2213	0.5358	1.8961
	Polyhedron collapse	519.5880	14400	0.9440
	Q-Sweep	528.8	18392	0.384
UWO (692, 97326)	Linear-Eigen	1016.7	1160.3	3359.4
	LM	2409.0	470.1729	1256.2
	IRMP	173.0558	0.6693	2.0964
	Polyhedron collapse	328.797	9835	1.252
	Q-Sweep	603.425	10357	0.481
Notre Dame (761, 53857)	Linear-Eigen	643.0999	392.9571	616.8394
	LM	1225.10	0.909	2.1446
	IRMP	113.9279	0.6621	1.8175
	Polyhedron collapse	242.436	828.4576	1.387
	Q-Sweep	669.336	798.5755	0.496

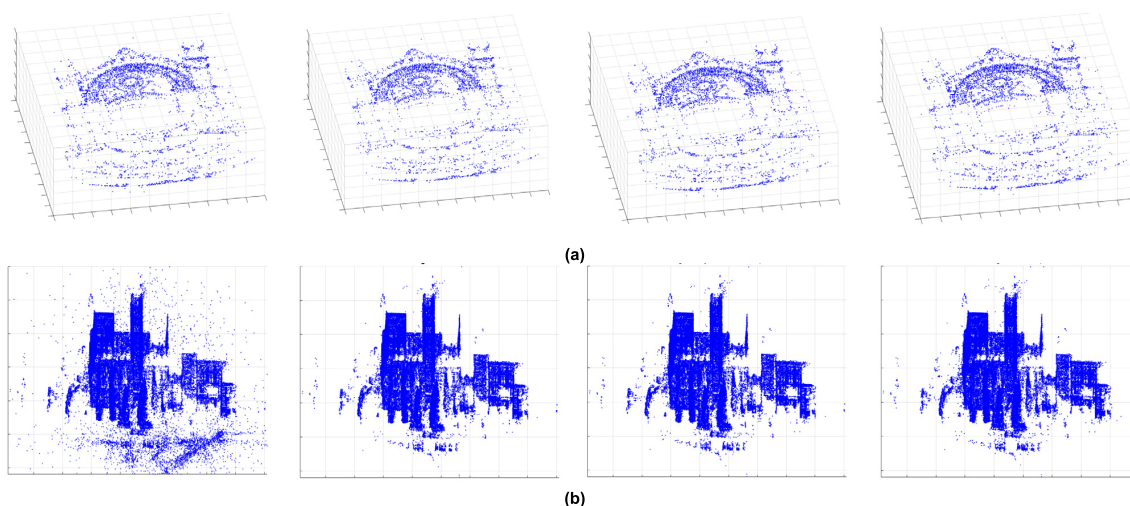


FIGURE 5. From left to right, reconstruction effect by the LM method, IRMP method, Polyhedron collapse method and Q-Sweep method. (a) Using the Drinking Fountain dataset. (b) Using the UWO dataset.

complex. However, the slowest is still the LM method, which takes a longer time as the number of views increases. It should be noted that although these datasets contain many views,

not every point is visible to all cameras. From the view distribution of Fig. 1, the view distribution of the Linkoping Cathedral dataset is mainly concentrated within 10 views,

TABLE 2. Experimental results of MPLP on the real datasets.

Dataset	Algorithm	Time(s)	RMSE	Median Error
Door (12, 17650)	MPLM	201.8039	0.2415	0.3864
	LM	242.4006	15.3282	14.8434
	IRMP	8.7919	0.2451	0.3805
Drinking Fountain (14, 5302)	MPLM	54.7102	2.0297	4.3257
	LM	66.1285	58.9165	111.3934
	IRMP	1.4926	1782.9	3565.2
Gustav Vasa (18, 4249)	MPLM	43.7908	0.5089	1.6798
	LM	50.6944	0.5089	1.6799
	IRMP	1.2332	4.6972	9.9186
Linkoping Cathedral (538, 202737)	MPLM	4621.1	0.5350	1.8839
	LM	3652.4	17.5360	167.8164
	IRMP	899.2213	0.5358	1.8961
UWO (692, 97326)	MPLM	2211.2	0.6683	2.0929
	LM	2409.0	470.1729	1256.2
	IRMP	173.0558	0.6693	2.0964
Notre Dame (761, 53857)	MPLM	529.8683	0.6611	1.8056
	LM	1225.10	0.909	2.1446
	IRMP	113.9279	0.6621	1.8175
Dinosaur (36, 4983)	MPLM	49.3433	0.2869	1.6871
	LM	51.1171	0.2869	1.6898
	IRMP	1.5021	0.287	1.6871
Gustav Vasa (18, 4249)	MPLM	48.9213	0.5089	1.6798
	LM	54.1507	0.5089	1.6799
	IRMP	1.6275	4.6972	9.9186
Skansen Kronan (131, 28371)	MPLM	380.0843	0.5663	1.5116
	LM	404.4972	0.5663	1.5114
	IRMP	28.2940	0.5672	1.5190
Water Tower (173, 53857)	MPLM	159.1302	0.4487	2.1622
	LM	188.7838	17.0662	152.496
	IRMP	13.4299	6.4592	12.4606
Ystad Monestary (290, 139951)	MPLM	1850.6	0.6835	2.1788
	LM	2010.2	253.6541	2577.9
	IRMP	131.6134	1.6042	3.9911
Buddah Statue (322, 156356)	MPLM	1322.7	0.4582	2.7826
	LM	2052.40	0.4585	2.7825
	IRMP	109.8185	4.7300	10.6026

while the view distribution of the other two datasets is concentrated within 10-40 views, so the Linear-Eigen method is still the fastest on the Linkoping Cathedral dataset. Compared with the Q-Sweep method, the Polyhedron collapse method is faster than Q-Sweep when the number of views is small, but the Q-Sweep method takes a longer time when the number of views is large.

From the perspective of RMSE, the best method is the IRMP method. But as seen in Table 1, the experimental results show this method is sometimes not as good as the LM method, and even in the reconstruction of the small scene of the Drinking Fountain dataset, a large error is generated. The RMSE of the LM method is also small, but when it falls into a local minimum, a large error is generated, and LM cannot make the median error as small as the IRMP method. For the

median error, the Q-Sweep method is the best, followed by the Polyhedron collapse method, but the RMSE of the two methods is relatively large. From the reconstruction effect, the Q-Sweep method can reduce the noise outside the target and converge on the reconstructed object compared with the Polyhedron collapse method.

V. DISCUSSION AND IMPROVEMENT

A. DISCUSSION

From the experimental results of the synthetic datasets, we can draw some conclusions. In the experiments, the LM method is the slowest, and the Linear-Eigen method is the fastest. When more than 20 views, the time consumption of the Q-Sweep method increases the fastest. When the number

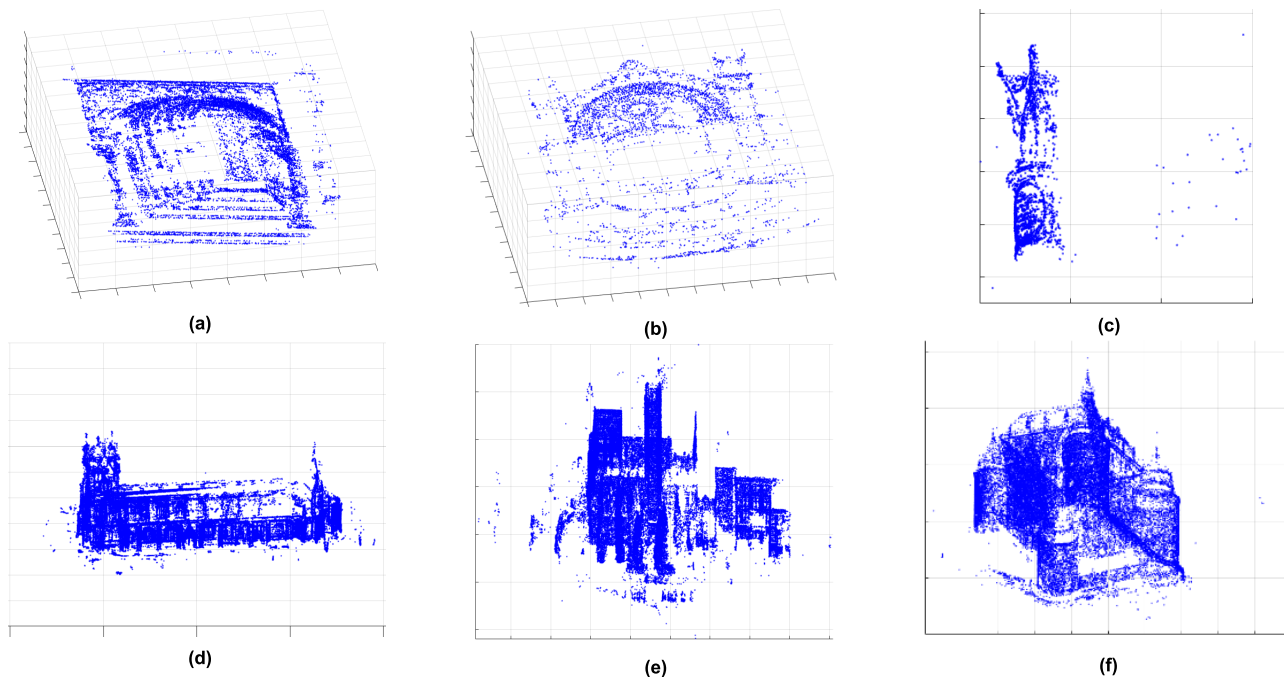


FIGURE 6. The reconstruction effect of the MPLM method on six datasets. (a) The Door dataset. (b) The Drinking Fountain dataset. (c) The Gustav Vasa dataset. (d) The Linköping Cathedral dataset. (e) The UWO dataset. (f) The Notre Dame dataset.

of views exceeds 70, the time consumption of the Polyhedron collapse method increases dramatically. The proportion of outliers has a significant impact on the IRMP method and has little influence on other methods. For median errors, the Q-Sweep method always maintains the best results.

With regard to RMSE and error with ground truths, the errors of the Q-Sweep method and the Polyhedron collapse method are always much larger than other methods. So they are not suitable for tasks that need to take into account the real position of the object, such as SLAM.

From the experimental results of the real scene datasets, the IRMP method performs well for speed, RMSE, and median error, but there are large errors in individual cases. The LM method takes a long time, and the accuracy is lower than IRMP because it is susceptible to the initial value and falls into the local minimum. The median error of the Q-Sweep method is always the smallest, followed by the Polyhedron collapse method, but their RMSEs are still large.

B. IMPROVEMENT

Based on the above experiments, we find that each method has its own advantages and disadvantages, and cannot always maintain good results. Since there are large differences between the results obtained by using the L_∞ norm method and the ground truths, we mainly study other methods. The L_2 norm method is easy to fall into local minimum due to the non-convexity of the function. The result of the LM method in the experiment reflects this characteristic. The distance representation of the IRMP method also takes the Euclidean

distance, so this is also the case. However, we find that the results of LM and IRMP are not consistent. In the experiments, the LM method on the UWO dataset performs poorly, while IRMP performs poorly on the Drinking Fountain dataset. This is due to the different cost function constructions and the different initial values.

The cost function of IRMP is to minimize the deviation angle between the estimated point and the optical center and the image point and the optical center in the space, and the cost function of LM is to minimize the distance between the image point and the 2D point which is the re-projection of the estimated point. The cost function is completely different, and their initial values are also different. The initial value of IRMP is obtained by the MVMP method, which is an improvement on the MVMP method, while the initial value of LM is obtained by the linear method. Both The initial values and the cost functions are different, so they behave differently when they fall into a local extreme value. For the poor performance of IRMP on Drinking Fountain dataset, we analyze the reason that IRMP and MVMP have similar cost functions. Although the MVMP can get a better initial value, the initial value obtained by it will more easily cause the IRMP to fall into a local extreme value and cannot escape. If this initial value is placed in another cost function that is completely different from the MVMP, it may iterate normally.

Inspired by this discovery, we can combine the advantages of the LM method and the IRMP method. Considering that the LM method is susceptible to the initial value, we use the initial value taken by the IRMP method. We use MVMP to select the initial value and then use LM to conduct

follow up optimization. We call this new method MPLM. The parameters used by the LM method are the projection matrix \mathbf{P} and the image point \mathbf{x} , while the parameters used by the MVMP method are the optical center \mathbf{O} and the image point \mathbf{x} . We can obtain the optical center \mathbf{O} from the projection matrix \mathbf{P} . Assuming a projection matrix $\mathbf{P} = [\mathbf{p}_1 \ \mathbf{p}_2 \ \mathbf{p}_3 \ \mathbf{p}_4]$ (\mathbf{p}_i represents the i -th column of the matrix \mathbf{P}), the homogeneous coordinates of the optical center \mathbf{O} are expressed as $\mathbf{O} = (x, y, z, t)^T$, then there is

$$\begin{aligned} x &= \det([\mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4]) \\ y &= -\det([\mathbf{p}_1, \mathbf{p}_3, \mathbf{p}_4]) \\ z &= \det([\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_4]) \\ t &= -\det([\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3]) \end{aligned} \quad (19)$$

In addition, since the MVMP method only involves solving a system of equations containing three linear equations, the complexity is not increased as the number of views increases. So the initial value is also solved very quickly.

The experimental results are shown in Table 2. To better measure the proposed approach, we add some datasets (Dinosaur, Gustav Vasa, Skansen Kronan, Water Tower, Ystad Monastery, Buddah Statue) [28]–[31]. The experimental results show that compared with the LM method, the MPLM method can shorten the calculation time and jump out of the local minimum. Although IRMP takes a shorter time than MPLM, MPLM is more accurate than IRMP. The reconstruction effect is shown in Fig. 6.

VI. CONCLUSION

In this paper, we have compared various currently existing variants of triangulation method using both synthetic datasets and real datasets. The Linear-Eigen method and the LM method are conventional methods. The Polyhedron collapse method, the Q-Sweep method, and the IRMP method are proposed in recent years. We have evaluated these methods with RMSE and median error. It can be seen from the experimental results that the performance of IRMP method is better with regard to RMSE, while the Q-Sweep method is better with regard to median error. From the synthetic experiment, we find that the results of the Q-Sweep method and the Polyhedron collapse method have large errors with the ground truth, although these two methods perform better in the median error. Between the Q-Sweep method and the Polyhedron collapse method, the Q-Sweep method has better robustness and accuracy. However, when the number of views is large, it takes a longer time than the Polyhedron collapse method, especially it takes much longer time than the LM method when most points are captured by more than 40 cameras. However, most datasets contain many views, and most of them are distributed within 20 views, so the Q-Sweep method is worth choosing.

Through experiments, we have some other discoveries. The Linear-Eigen method is simple and easy to use and is often used as an initial estimate for other methods, but when the number of views is large, it is even slower than the

IRMP method. Through experimental exploration, we presented an improved method (MPLM) which uses MVMP to select the initial value and then uses LM to conduct follow up optimization. Experimental results show that the effect of this combination is better. Compared with the previous LM method, runtime and accuracy of the MPLM method are improved. Compared with the IRMP method, although the IRMP method is faster, the MPLM method is more accurate.

REFERENCES

- [1] J. Chen, "Human body shape and motion tracking by hierarchical weighted ICP," presented at the ISVC, Las Vegas, NV, USA, 2011.
- [2] P. Song, X. Wu, and M. Y. Wang, "Volumetric stereo and silhouette fusion for image-based modeling," *Vis. Comput.*, vol. 26, no. 12, pp. 1435–1450, Dec. 2010.
- [3] I. Khan, "Robust sparse and dense nonrigid structure from motion," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 841–850, Apr. 2018.
- [4] Y. Gao and A. L. Yuille, "Symmetric non-rigid structure from motion for category-specific object structure estimation," presented at the ECCV, Amsterdam, The Netherlands, 2016.
- [5] D. Xiao, J. Li, and K. Li, "Robust precise dynamic point reconstruction from multi-view," *IEEE Access*, vol. 7, pp. 22408–22420, 2019.
- [6] S. Urban, S. Wursthorn, J. Leitloff, and S. Hinz, "MultiCol bundle adjustment: A generic method for pose estimation, simultaneous self-calibration and reconstruction for arbitrary multi-camera systems," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 234–252, Jan. 2017.
- [7] R. I. Hartley and P. Sturm, "Triangulation," *Comput. Vis. Image Understand.*, vol. 68, no. 2, pp. 146–157, 1997.
- [8] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, and V. Kumar, "Robust stereo visual inertial odometry for fast autonomous flight," *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 965–972, Apr. 2018.
- [9] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.
- [10] H. Stewenius, F. Schaffalitzky, and D. Nister, "How hard is 3-view triangulation really?" presented at the ICCV, Beijing, China, 2005.
- [11] M. Byrod, K. Josephson, and K. Astrom, "Fast optimal three view triangulation," presented at the ACCV, Tokyo, Japan, 2007.
- [12] P. Lindstrom, "Triangulation made easy," presented at the IEEE CVPR, San Francisco, CA, USA, 2010.
- [13] S. H. and J. Civera, "Triangulation: Why optimize?" presented at the BMVC, Cardiff, Wales, 2019.
- [14] P. A. Beardsley, A. Zisserman, and D. W. Murray, "Navigation using affine structure from motion," presented at the ECCV, Berlin, Germany, 1994.
- [15] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
- [16] B. Triggs, "Bundle adjustment—A modern synthesis," presented at the IWVA, Berlin, Germany, 2000.
- [17] R. Hartley and F. Schaffalitzky, " L_∞ minimization in geometric reconstruction problems," presented at the CVPR, Washington, DC, USA, 2004.
- [18] S. Donné, B. Goossens, and W. Philips, "Point triangulation through polyhedron collapse using the ℓ_∞ norm," presented at the ICCV, Santiago, Chile, 2015.
- [19] S. Agarwal, N. Snavely, and S. M. Seitz, "Fast algorithms for L_∞ problems in multiview geometry," presented at the IEEE CVPR, Anchorage, Alaska, 2008.
- [20] A. Eriksson and M. Isaksson, "Pseudoconvex proximal splitting for L_∞ problems in multiview geometry," presented at the IEEE CVPR, Columbus, OH, USA, 2014.
- [21] Q. Zhang, T.-J. Chin, and D. Suter, "Quasiconvex plane sweep for triangulation with outliers," presented at the ICCV, Venice, Italy, 2017.
- [22] E. G. Kogbetliantz, "Solution of linear equations by diagonalization of coefficients matrix," *Quart. Appl. Math.*, vol. 13, no. 2, pp. 123–132, Jul. 1955.
- [23] P. Sturm, S. Ramalingam, and S. Lodha, "On calibration, structure from motion and multi-view geometry for generic camera models," presented at the 2nd ISPRS Panoramic Photogram. Workshop, Berlin, Germany, 2005.

[24] K. Yang, W. Fang, Y. Zhao, and N. Deng, "Iteratively reweighted midpoint method for fast multiple view triangulation," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 708–715, Apr. 2019.

[25] K. Madsen, H. B. Nielsen, and O. Tingleff, *Methods for Non-Linear Least Squares Problems*, 2nd ed. Lyngby, Denmark: IMM, 2004.

[26] Z. Dai, Y. Wu, F. Zhang, and H. Wang, "A novel fast method for L_∞ problems in multiview geometry," presented at the ECCV, Florence, Italy, 2012.

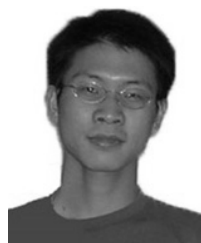
[27] F. Kahl and R. Hartley, "Multiple-view geometry under the L_∞ -norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1603–1617, Sep. 2008.

[28] N. Snavely, S. M. Seitz, and R. Szeliski, "Photo tourism: Exploring photo collections in 3D," *ACM ToG*, vol. 25, no. 3, pp. 835–846, 2006.

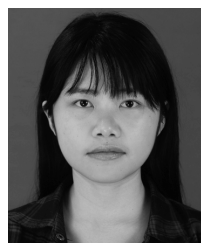
[29] C. Olsson and O. Enqvist, "Stable structure from motion for unordered image collections," presented at the SCIA, Ystad, Sweden, 2011.

[30] O. Enqvist, F. Kahl, and C. Olsson, "Non-sequential structure from motion," presented at the ICCV Workshops, Barcelona, Spain, 2011.

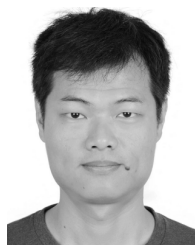
[31] *VGG Multi-view Data*. Accessed: Dec. 30, 2019. [Online]. Available: <http://www.robots.ox.ac.uk/~vgg/data/mview/>



JIA CHEN was born in Hubei, China, in 1982. He received the M.S. degree in electrical engineering and the Ph.D. degree in control science and engineering from the Harbin Institute of Technology (HIT), China, in 2012. Then, he worked at the Samsung Research Institute, Beijing, as a leading Research and Development Engineer. He joined the School of Information Technology, Central China Normal University. From 2017 to 2018, he visited the Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, Guildford, Surrey, U.K., as a Visiting Scholar. He currently teaches at the Department of Digital Media Technology, School of Educational Information Technology, Central China Normal University. His current research interests focus on VR/AR, 3D reconstruction, 3D motion capture, and educational information technology.



DONGLI WU was born in Henan, China, in 1996. She received the B.S. degree from the School of Computer and Information Science, Southwest University, China, in 2018. She is currently pursuing the M.S. degree with the School of Educational Information Technology, Central China Normal University. Her current research interests focus on camera calibration and 3D reconstruction.



PENG SONG received the B.S. and M.S. degrees in control science and engineering from the Harbin Institute of Technology, in 2007 and 2009, respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2013. From 2013 to 2014, he worked as a Research Fellow at Nanyang Technological University (NTU), Singapore. From 2014 to 2017, he was an Associate Professor with the University of Science and Technology of China (USTC). He holds a Postdoctoral Fellowship at EPFL, working with Mark Pauly. He has already published seven SIGGRAPH or SIGGRAPH Asia papers. He will join Singapore University of Technology and Design (SUTD) as an Assistant Professor, in 2020. His research interests include computer graphics, computer vision, and human-computer interaction.



FUQIN DENG received the B.S. degree in applied mathematics from the Harbin Institute of Technology, Harbin, China, in 2005, the M.S. degree in control science and engineering from the Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China, in 2007, and the Ph.D. degree in ECE from The University of Hong Kong, Hong Kong. He is currently a Professor with Wuyi University. His current research interests include image processing, pattern recognition, computer vision, and machine vision applications.



YING HE was born in Beijing, China, in 1984. She received the M.S. and Ph.D. degrees from the Harbin Institute of Technology (HIT), China, in 2018. She currently holds a Postdoctoral Fellowship at the Graduate School at Shenzhen, Tsinghua University. Her current research interests focus on 3D reconstruction, simultaneous localization and mapping (SLAM), and robot visual navigation.



SHIYANG PANG received the B.S., M.S., and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, China, in 2009, 2012, and 2015, respectively. From 2016 to 2019, she held a Postdoctoral Fellowship at Wuhan University. She currently teaches at the Department of Digital Media Technology, School of Information Technology, Central China Normal University. Her current research interests include 3D reconstruction, computer graphics, computer vision, and remote sensing.

...