Received: 30 May 2022 | Revised: 12 April 2023 | Accepted: 13 May 2023

IET Image Processing

DOI: 10.1049/ipr2.12831

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

The optimal triangulation method is not really optimal

Seyed-Mahdi Nasiri[1] | **Reshad Hosseini**[1,2] | **Hadi Moradi**[1,3]

[1]School of ECE, College of Engineering, University of Tehran, Tehran, Iran

[2]School of Computer Science, Institute of Research in Fundamental Sciences (IPM), Tehran, Iran

[3]Intelligent Systems Research Institute, SKKU, South Korea

Correspondence
Reshad Hosseini, School of ECE, College of Engineering, University of Tehran, Tehran, Iran.
Email: reshad.hosseini@ut.ac.ir

Abstract

Triangulation refers to the problem of finding a 3D point from its 2D projections on multiple images taken at different camera poses. For solving this problem, it is a common practice to use the so-called optimal triangulation method. But, the method can be optimal only if we assume no uncertainty in the camera parameters. In this paper, an extensive comparison between various existing methods for the calibrated cameras is performed, in which different poses are considered. Furthermore, uncertainty sensitivity analysis is conducted for extrinsic parameters of cameras. The results show that the optimal triangulation method is actually not the best choice when there is uncertainty in extrinsic parameters. Interestingly, it can be observed that the simple midpoint method works equally well or outperforms the optimal triangulation and many other methods. Apart from its high performance, the midpoint method has a simple closed form solution for multiple views while the optimal triangulation method is hard to be used for more than two views. Therefore, in contrast to the common practice, we argue that the simple midpoint method can be a good choice in the *structure-from-motion* process where there is uncertainty in extrinsic camera parameters.

1 | INTRODUCTION

One of the fundamental tasks in 3D vision is reconstructing a point in 3D from its projections on camera images in two or multiple views. This task, which is called triangulation, is used extensively in machine vision and robotics applications such as stereo vision [1, 2], motion estimation and tracking [3, 4], mapping [5] and *structure-from-motion* [6–8]. It is obvious that a 3D point can be simply found by intersecting the lines of sight of each projection when no noise is present. But in practice, due to several sources of noise such as uncertainties in relative camera poses, errors in cameras intrinsic parameters, and subpixel inaccuracies in the position of matched points, all the lines do not necessarily intersect at one point or at all. There have been many attempts to solve the triangulation problem in the presence of uncertainties [9–11].

The common method for solving triangulation in two views is casting it as a nonlinear optimization problem, which is called the optimal method. In such a problem, a new point is found as close as possible to the measured point in each view so that the lines of sights for the new points intersect. In [9], the authors compared several triangulation methods on several simulated datasets. They observed that the optimal triangulation method, which we call it the L_2 method (see Figure 1a) in this paper, outperforms other methods. Apart from good performance, the authors argued that the L_2 triangulation method has a nice property to be projective invariant. In the performed simulations, the authors considered the uncertainty in the position of corresponding points and not camera parameters.

In the majority of triangulation problems, there are uncertainties in both the parameters of cameras and measured corresponding points. Thus, a question arises here, "Is the L_2 triangulation method still an optimal method?" To the best of our knowledge, this question has not been investigated previously. Thus in this paper, we evaluate the performance of the triangulation methods in the calibrated *structure-from-motion* setting. We chose the calibrated *structure-from-motion* because after the work of [12] that has become the natural choice for *structure-from-motion* applications. This is due to the fact that knowing intrinsic calibration improves the accuracy and robustness of the structure-from-motion methods [12, 13]. In other words, the problem in the calibrated setting is to estimate extrinsic parameters as well as 3D points corresponding to the observed points in different views.

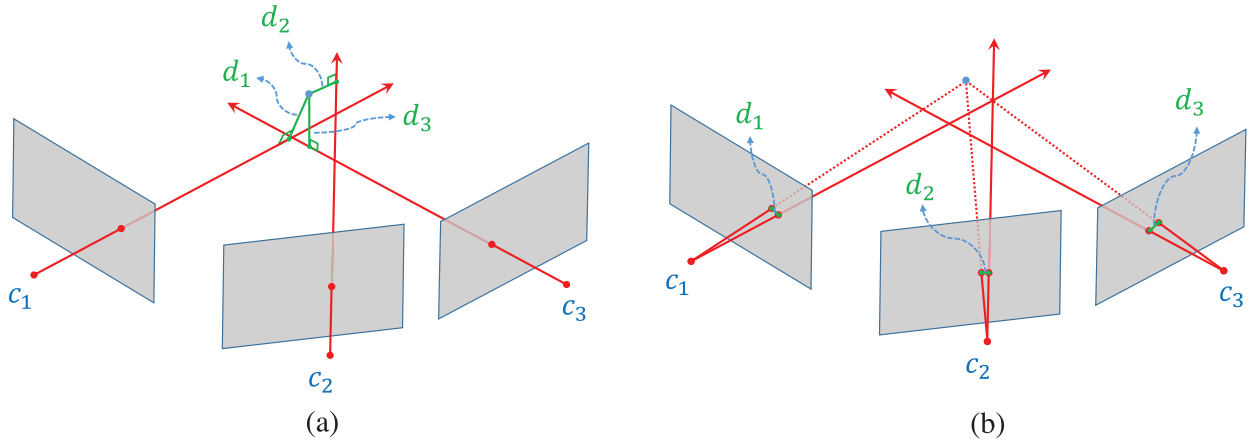

FIGURE 1 Triangulation methods: (a) Mid-point method finds a 3D point which has the minimum sum of squared distances from the lines of sights. (b) L_2 triangulation method finds a 3D point which has the minimum sum of squared 2D distances between its projections and the corresponding points on images.

In this paper, we will show that when uncertainty exists in camera extrinsic parameters, the L_2 triangulation method is no longer a state-of-the-art method. Interestingly, a simple mid-point method, that is, the mid-point of lines of sights in different views (see Figure 1a), works much better than the L2 triangulation method in practice. The mid-point method not only gives better results but it can also be generalized to any number of views with no difficulties. In contrast, the L_2 triangulation method is normally used for two views where the roots of a polynomial of degree 6 needs to be computed [9]. For more than two views, the method becomes computationally expensive and hard. For example, in the case of three views, the optimal solution is one of the real solutions among the set of 47 general roots of a polynomial equation [14]. Finally, although the mid-point method is not projective or affine invariant [9], but this lack of invariance is not important in calibrated camera situations.

Through extensive simulations, both on synthetic and real datasets, we have validated the lower triangulation error of the mid-point method compared to the L2 triangulation method.

We have assessed the error when uncertainties in relative cameras poses exist. We have also assessed the error when uncertainties are caused by a commonly used *structure-from-motion* approach, where first the essential matrix[1] is estimated from point correspondences and then relative poses are estimated [12] and finally the 3D structure is obtained. In all of these experiments, we observed that the mid-point method outperforms other different approaches. Thus, we suggest that unlike the common practice, the mid-point method should be used in *structure-from-motion* applications.

2 | RELATED WORKS

The L_2 triangulation approach, which is known as an optimal method, finds a 3D point that minimizes the L_2 reprojection errors in the image domain [15]. This leads to find the optimal solution, using maximum-likelihood, under the assumption of Gaussian noise in positions of projections. In [9], the authors have shown that minimizing the L_2 reprojection errors, for the case of two images, can be reduced to finding the stationary points of a 6^{th} degree polynomial and selecting the best points by evaluating the objective function. In [16], the authors proposed a Gröbner basis based algorithm for minimizing the L_2 reprojection errors in the case of three image observations. They showed that the optimal solution is one of the real solutions among a set of 47 general roots of a certain polynomial equation. Since their approach has a significant computational cost, an alternative method of [14] can be used. In this method, a technique was introduced that improves the numerical stability of Gröbner basis solvers and significantly reduces the computational costs. In [17], the authors proposed the MPLM method which uses the mid-point method as the initializer of the Levenberg-Marquardt (LM) numerical optimization method for minimizing the L_2 cost function. In this paper, we used the same approach for more than two views for the L_2 triangulation approach.

Because of the non-convexity and complexity of solving the L_2 norm [18], other cost functions have been considered in the literature. For instance, a choice which is robust to outliers is to minimize the L_1 reprojection errors. In the proposed approach of [9], the L_1 optimal solution is found in closed form by solving a polynomial of degree 8. They also stated that the L_1 optimization has slightly less 3D error than the L_2 optimization in real experiments.

Another popular approach is to find L_∞ answer which is optimal under the assumption of uniform noise [19–23]. However, it is well known that L_∞-norm optimization is extremely vulnerable to outliers [21]. Minimizing the median L_∞ error employed in [24] increased the robustness of the triangulation process against outliers.

Angular errors were studied in [10] and closed-form optimal solutions were derived for L_1, L_2, and L_∞ angular errors.

Since the midpoint method is labelled inaccurate in the literature, the authors in [25] proposed a modified loss function

[1] The essential matrix corresponding to a pair of cameras with relative orientation, R, and translation, t, is defined as $E = [t]_\times R$.

and the iteratively reweighted midpoint method to solve it. An incremental triangulation method was proposed in [26] that tries to solve the same cost function of [25] more efficiently. This method is performed in two steps. First, the initial 3D reconstruction is achieved with a few first views. Second, new views are added in a sequence and for each new view the 3D reconstruction is refined.

The closest work to our paper is [27], that proposed a new alternative midpoint method for two-view triangulation. It compared several methods, including the L_2 and the traditional midpoint method on some synthetic two-view data. They perturbed camera poses in addition to the projected points in their experiments, and observed that the midpoint method results in a less 3D triangulation error.

Our study goes beyond [27] in several aspects:

1. Our experiments are not limited to the two-view triangulation.
2. We investigate the sensitivity of different methods to extrinsic parameters of cameras from different aspects.
3. We evaluate the performance of different methods on a real dataset in addition to synthetic data.
4. Most importantly, we evaluate the performance of different triangulation approaches in a common calibrated SfM process [8, 28].

We show that the midpoint method has less sensitivity to uncertainties in cameras extrinsic parameters. This means that as the error in the camera poses increases, the L_2 triangulation method is no longer the optimal method and the midpoint method outperforms it. It is also shown that the midpoint has more accuracy in 3D reconstruction in a full *structure-from-motion* process where positions and orientations of the cameras are estimated from observations.

3 | PRELIMINARIES

In this paper, we evaluate the performance of triangulation methods for calibrated cameras. Therefore, the term camera parameters is used for camera extrinsic parameter throughout the paper. In this paper, "N" represents the number of points whose positions should be estimated from their projections on "N_c" cameras.

3.1 | Camera model and parameters

Let $\mathbf{u}$ be the projection of a point $\mathbf{p}$ on the image plane of a camera. The projection is obtained by $\mathbf{u} = P[\mathbf{p}; 1]$, where P is the camera matrix. The camera matrix P is given by $P = K[R \mid -R\mathbf{c}]$, where K is the camera calibration matrix, and R and $\mathbf{c}$ are the orientation and position of the camera with respect to a world coordinate system. The line of sight of the camera image is the line that passes through the camera point $\mathbf{c}$ along direction $R^{-1}K^{-1}\mathbf{u}$.

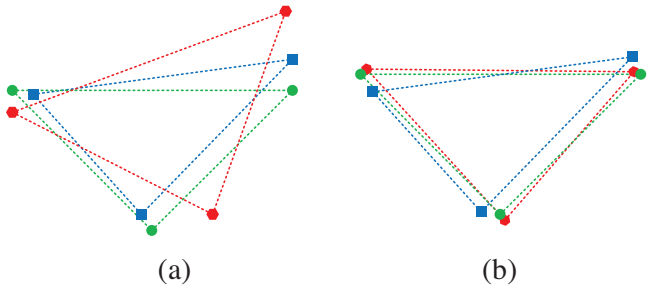

FIGURE 2 The ground truth point cloud (green circles) and the estimated point clouds (blue squares and red hexagons). As it is obvious in (a), the blue points have a smaller sum squared distance to the ground truth versus the red points. But after finding the best rotation, translation, and scale for both red and blue points, done for (b), it reveals that the red points are the better estimation of the ground truth green points.

3.2 | Optimal triangulation

In the optimal triangulation method a 3D point $\mathbf{p}$ is found such that its projected points on the image planes of cameras, $\hat{\mathbf{u}}_i, i \in \{1, \dots, N_c\}$, have the minimum sum Euclidean distance from measurements $\mathbf{u}_i$s. Hence, it minimizes the following cost function:

$$f(\mathbf{p}) = \sum_{i=1}^{N_c} d(\hat{\mathbf{u}}_i, \mathbf{u}_i)^2, \quad \hat{\mathbf{u}}_i = P[\mathbf{p}; 1],\ i \in \{1, \dots, N_c\}, \tag{1}$$

in which $d(\hat{\mathbf{u}}_i, \mathbf{u}_i)$ is the Euclidean distance between the projected point and its measurement in the i^{th} image. Assuming independent Gaussian noise in the image domain and known cameras' poses, this method provides the maximum-likelihood estimation of 3D points.

3.3 | Midpoint triangulation

Another simple triangulation method is to find a $3D$ point $\mathbf{p}$ that minimizes $3D$ distances from the lines of sights. The goal of this method is to minimize the following cost function:

$$f(\mathbf{p}) = \sum_{i=1}^{N_c} d(\mathbf{p}, \mathbf{l}_i)^2, \tag{2}$$

in which $\mathbf{l}_i$s are the lines of sights and $d(\mathbf{p}, \mathbf{l}_i)$ is the distance between $\mathbf{p}$ and $\mathbf{l}_i$. For any number of cameras, minimizing (2) is a linear least squares problem and can be solved in a closed form.

3.4 | Accuracy of the reconstruction

A point cloud reconstructed by a *structure-from-motion* procedure is obtained up to a scaled Euclidean transformation (a more general projective ambiguity exists in the uncalibrated approach). Suppose that $\hat{\mathbf{p}}_i$s, $i \in \{1, \dots, N\}$ are estimated points, and $\mathbf{p}_i$s are the ground truth (with known correspondences).

As shown in Figure 2, the accuracy of the estimation is obtained by finding a scaled Euclidean transformation such that

the estimated point cloud are aligned to the ground truth as much as possible. Mathematically speaking

$$\min_{R,\mathbf{t},s} \sum_{i=1}^{N} d\left(sR\hat{\mathbf{p}}_i + \mathbf{t}, \mathbf{p}_i\right)^2, \tag{3}$$

where R, $\mathbf{t}$, and s are the rotation matrix, translation, and scale parameters of the scaled Euclidean transformation.

3.5 | Reconstruction in a *structure-from-motion* process

A typical *structure-from-motion* framework [8, 29, 30] comprises the following steps:

1. Pairwise images registration
   - Feature extraction and matching [31–33]
   - Finding relative rotations and translations between all pairs of images via the matched features [12, 13, 34]
2. Camera pose estimation
   - Solving the viewing graph created by the pairwise image registrations to find camera positions and orientations [29, 35–40]
3. Triangulation
   - Reconstructing 3D points by triangulating corresponding points.

   The obtained camera poses and 3D reconstructed points are usually refined by a *bundle adjustment* step. In this paper, we only consider up to the triangulation adjustment and exclude the bundle adjustment step.

4 | EXPERIMENTS

The experiments are two-fold. First, the sensitivity of the different triangulation methods to the error in camera poses are compared. Then, the accuracy of different triangulation methods are evaluated in a full reconstruction procedure on synthetic and real datasets. In the experiments, the midpoint method is compared to the L_2 method [9] , iteratively reweighted midpoint [25], minimizing L_1 reprojection errors [9], and minimizing L_1 and L_2 angular errors [10].The Q-Sweep method of [24] is also added to the experiments on the real dataset.

4.1 | Sensitivities

The comparison criteria for the sensitivity analysis on synthetic datasets are:

- Position error sensitivity: The sensitivity of the error in the position of a single triangulated point.
- Distance error sensitivity: The sensitivity of the error in the distance between two triangulated points.
- Angle error sensitivity: The sensitivity of the error in the angles of a triangle composed of three triangulated points.

To evaluate the mentioned sensitivities for different triangulation methods, three configurations are considered for two-views triangulation:

1. Conf. 1) $\mathbf{c}_1 = [-5, -1, 0]^T, \mathbf{c}_2 = [-5, +1, 0]^T$ and the both cameras point at origin.
2. Conf. 2) $\mathbf{c}_1 = [-12, 0, 0]^T, \mathbf{c}_2 = [-2, 0, 0]^T$ and the both cameras point at origin.
3. Conf. 3) $\mathbf{c}_1 = [-10, 2, -1]^T, \mathbf{c}_2 = [-5, -2, 1]^T$ and the both cameras baselines are aligned with the global coordinate x direction.

The first two configurations are the same as the two configurations in [9]. In fact, the first configuration simulates a camera moving forward and looking straight ahead and the second configuration simulates an aerial imaging procedure. Since these two configurations are special cases, a more general configuration is added to them.

4.1.1 | Position error sensitivity

In this part the sensitivity of $3D$ error of different triangulation methods to errors in positions and orientations of the cameras are evaluated. For each configuration, a point $\mathbf{p}$ is placed in a sphere centered at the origin with diameter of 0.5. Then, the projected points on the cameras are obtained. The positions of the cameras are perturbed by random Gaussian noise vectors. The cameras' orientations are perturbed by random rotations consist of random rotation axes and Gaussian random angles. The reconstructed point $\hat{\mathbf{p}}$, is obtained by different triangulation methods. Euclidean distance between points is used to compute $3D$ error:

$$e = d(\hat{\mathbf{p}}, \mathbf{p}). \tag{4}$$

This procedure repeated 100 times for each configuration and for each noise level. The standard deviation of the position Gaussian noise is 0.01, and the standard deviation of the angle Gaussian noise is 0.1 degree for the noise level 1. The standard deviations are multiplied by the noise levels. Figure 3 shows the mean error of different triangulation methods in configurations 1, 2, and 3.

4.1.2 | Distance error sensitivity

To evaluate the distance error sensitivity of different methods, two points $\mathbf{p}_1$ and $\mathbf{p}_2$ are randomly placed in the sphere described in Section 4.1.1. The projections are computed, cameras positions and orientations are perturbed by the described noises in Section 4.1.1, and different triangulation methods are applied to find two estimated $3D$ points $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_2$. The error

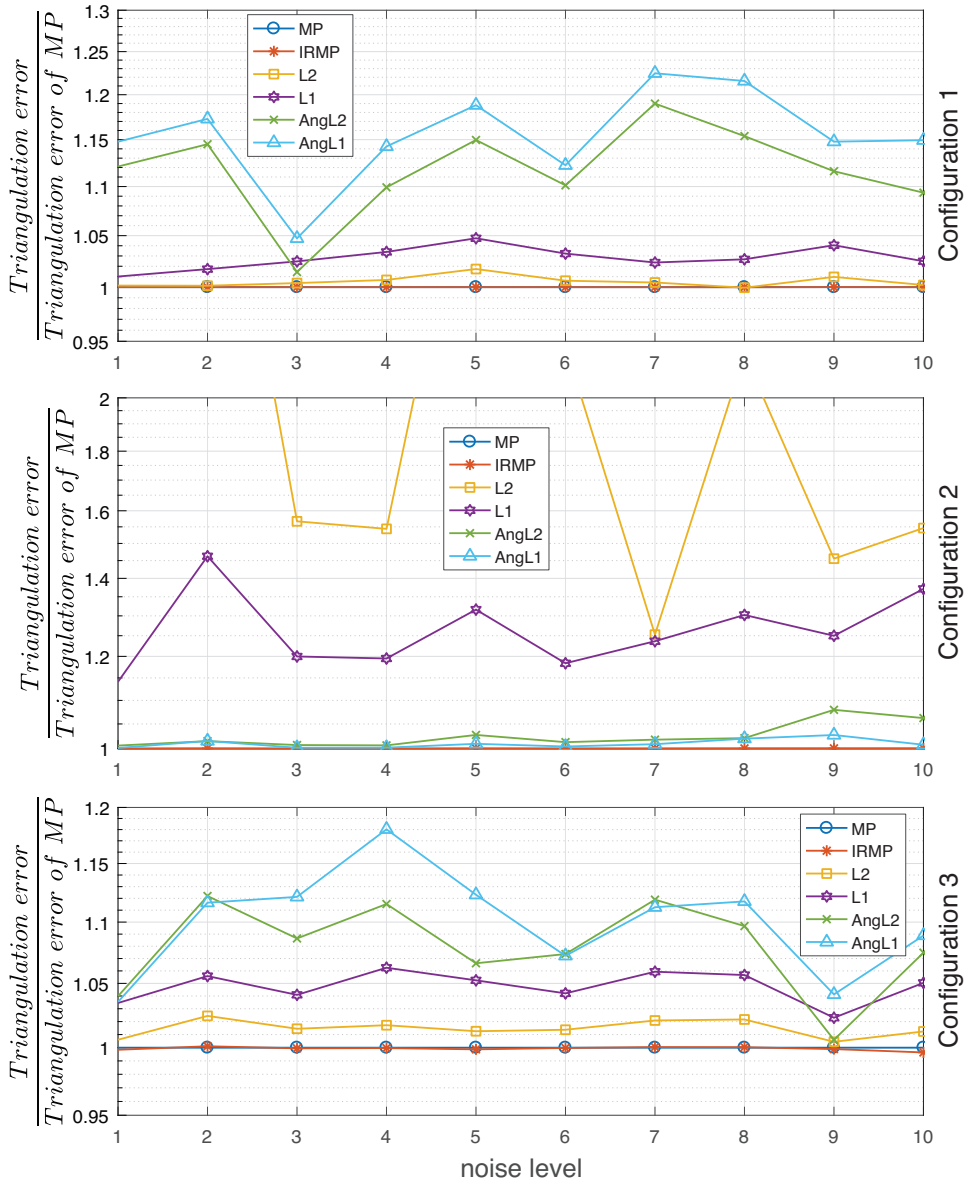

FIGURE 3 The position error (4) sensitivity of different methods in—from top to bottom—configurations 1, 2, 3.

is defined as the absolute value of difference of the distance between $\mathbf{p}_1$ and $\mathbf{p}_2$, and the distance between $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_2$:

$$e = \left| d(\mathbf{p}_1, \mathbf{p}_2) - d(\hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2) \right|. \tag{5}$$

This procedure is repeated 100 times for each configuration and for each noise level. The mean error of different triangulation methods are shown in Figure 4.

4.1.3 | Angle error sensitivity

To compute the angle error sensitivity of different methods, three points $\mathbf{p}_1, \mathbf{p}_2$, and $\mathbf{p}_3$ are randomly placed in the aforementioned sphere and the projections are computed. The positions and orientations of the cameras are perturbed by the noise described in Subsection 4.1.1 and three $3D$ points $\hat{\mathbf{p}}_1$, $\hat{\mathbf{p}}_2$, $\hat{\mathbf{p}}_3$ are estimated by different triangulation methods. The error is defined as the absolute value of the difference between the angle between two vectors $\mathbf{p}_2 - \mathbf{p}_1$ and $\mathbf{p}_3 - \mathbf{p}_1$, and the angle between two vectors $\hat{\mathbf{p}}_2 - \hat{\mathbf{p}}_1$ and $\hat{\mathbf{p}}_3 - \hat{\mathbf{p}}_1$:

$$e = \left| \measuredangle(\hat{\mathbf{p}}_2 - \hat{\mathbf{p}}_1, \hat{\mathbf{p}}_3 - \hat{\mathbf{p}}_1) - \measuredangle(\mathbf{p}_2 - \mathbf{p}_1, \mathbf{p}_3 - \mathbf{p}_1) \right|. \tag{6}$$

This procedure is repeated 100 times for each configuration and for each noise level. The mean error of different triangulation methods are shown in Figure 5.

From these experiments, it can be concluded that the midpoint method and its variant are the best performing methods when there is uncertainty in the cameras parameters. In real-

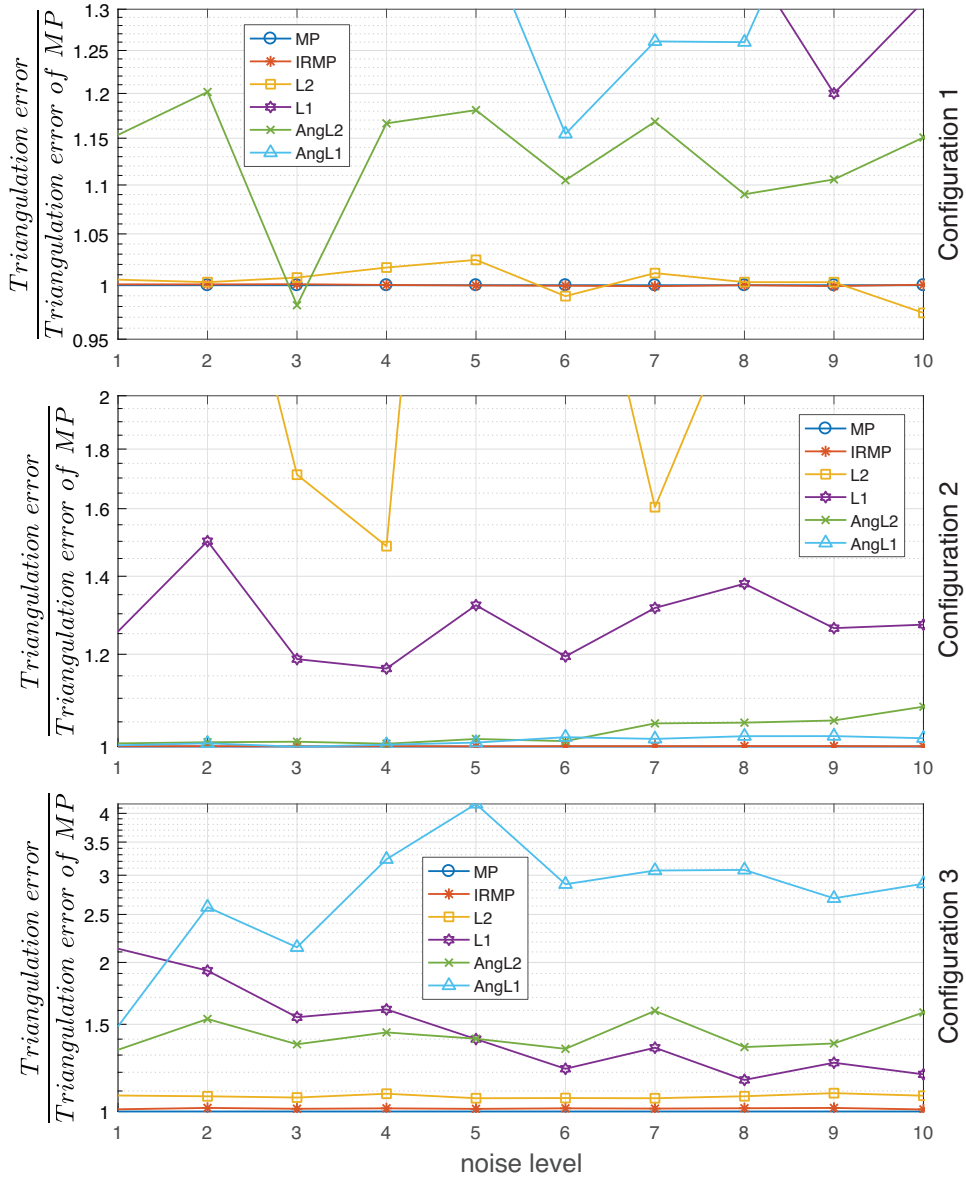

FIGURE 4 The distance error (5) sensitivity of different methods in—from top to bottom—Configurations 1, 2, 3.

ity, there are uncertainty in both cameras parameters and image points. This is addressed in the following subsection.

4.2 | Full reconstruction procedure on synthetic datasets

In the following experiments, the performance of the above methods in a full structure-from-motion reconstruction procedure is assessed. The datasets have uncertainties in image points and consequently there is uncertainty in cameras extrinsic parameters. In this section, first the performance in the case of two views is evaluated, where the poses of the cameras are computed through their essential matrix. Then, the performance for the case of more than two cameras are evaluated, in which an additional viewing graph optimization is needed to be solved for computing the camera's poses. In the case of two cameras, the closed form solution of [9] is used for the L_2 optimization problem. For the case of more than two cameras, the LM numerical optimization method is used with the initial point obtained by the midpoint method as suggested in [17].

4.2.1 | Two cameras

To evaluate the performance of different triangulation methods for the case of two cameras, the following steps are done on the simulation setup of Figure 6:

1. $N = 20$ points are randomly selected in a box.

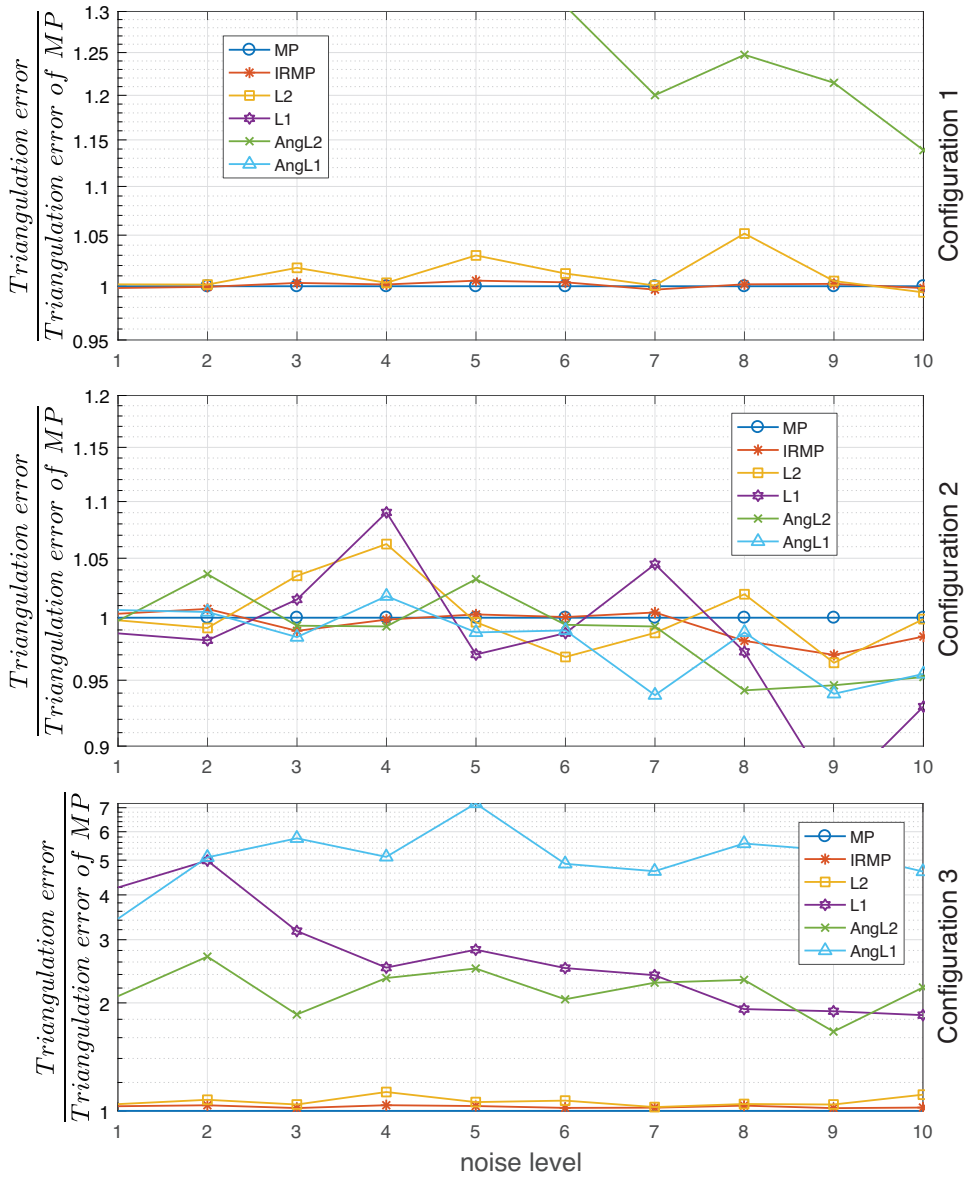

FIGURE 5 The angle error (6) sensitivity of different methods in—from top to bottom—Configurations 1, 2, 3.

2. The points are projected on the cameras.
3. The projections are displaced byunbiased Gaussian random noises with standard deviations of one pixel.
4. The essential matrix between the two cameras is estimated by the method of [13].
5. The relative rotation and translation are estimated.
6. The poses of the cameras are calculated from the relative observations.
7. The corresponding $3D$ points are reconstructed by triangulation.
8. The best rotation, translation, and scale that make the $3D$ triangulated points match the selected points in the box are obtained and the errors are computed.

The procedure is repeated 100 times.

The cameras have the same calibration matrix

$$K = \begin{bmatrix} 300 & 0 & 320 \\ 0 & 300 & 240 \\ 0 & 0 & 1 \end{bmatrix}, \tag{7}$$

and the images have 640×480 pixels.

Figure 7 shows the mean and standard deviation of triangulation errors of all 20 points for different methods in the first 10 experiments. Table 1 shows the mean, median, standard deviation, minimum, and maximum of the mean error of all 100 experiments. High standard deviations have been originated by the difference in the points used for computing cameras' parameters, and it is not related to any uncertainties in the

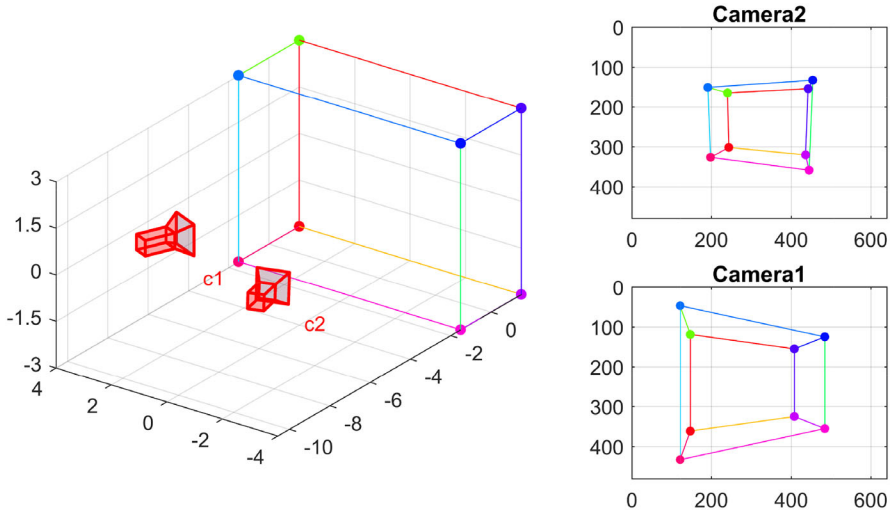

FIGURE 6 The configuration of the cameras and the box, and the images of the box on the cameras. The cameras are placed at $\mathbf{c}_1 = [-7, 3, 0]^T, \mathbf{c}_2 = [-10, -3, 1]^T$. Both cameras looking at the origin. The dimensions of the box are $3 \times 8 \times 6$ and, it is centered at the origin.

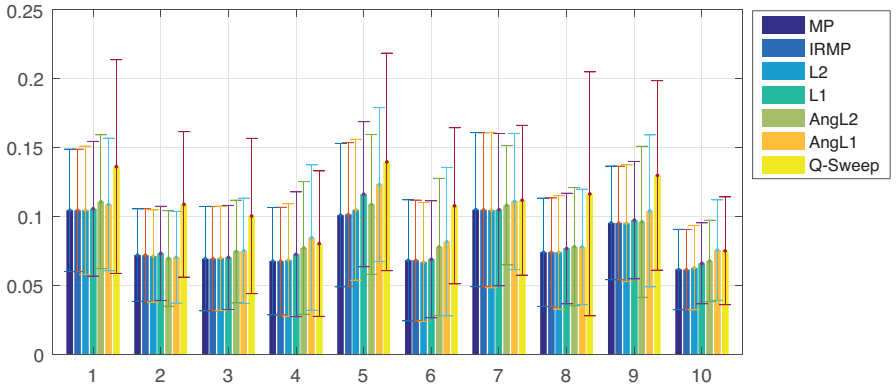

FIGURE 7 The mean and standard deviation of 3D triangulation errors for different methods in the first ten random runs on the synthetic dataset. The errors are computed for random 20 points selected in a region. The cameras and the region containing the points are shown in Figure 6.

TABLE 1 The mean, standard deviation, median, minimum, and maximum of 100 mean 3D triangulation error for different methods. The errors are computed for random 20 points selected in a region. The cameras and the region containing the points are shown in Figure 6.

Method	Mean	Median	Std	Min	Max
MP	**0.0951**	0.0814	0.0580	0.0486	0.3648
IRMP	**0.0951**	**0.0813**	0.0580	0.0485	0.3647
L2	0.0958	0.0816	0.0602	0.0484	0.3732
L1	0.0978	0.0847	0.0605	0.0518	0.3708
AngL2	0.1027	0.0825	0.0648	0.0522	0.3957
AngL1	0.1043	0.0861	0.0617	0.0547	0.3646
Q-Sweep	0.1193	0.1094	0.0524	0.0663	0.3541

experiments or in the methods. In other words if weput the midpoint method as the baseline and plot the difference between the error of a method and this baseline method, we get a very small standard variation.

4.2.2 | More than two cameras

If N_c cameras ($N_c > 2$) are involved in the reconstruction process, the essential matrices and consequently the relative positions and orientations are computed for every 2-combinations of N_c cameras. The $\binom{N_c}{2}$ relative observations of orientations and directions create a viewing graph which should be solved to estimate the cameras' poses. In the 4^{th} step of the reconstruction procedure, the essential matrices are estimated for any 2-combinations of N_c cameras, and in the 5^{th} step, a viewing graph is created from relative measurements. Obtained viewing graph is solved in step 6 to estimate the positions and orientations of the cameras.

To evaluate the performance of different methods on multiview triangulation, another camera is added to the mentioned two-camera setup as shown in Figure 8. The mean and standard deviation of triangulation error of all 20 points for different methods in the first 10 experiments are shown in Figure 9. Again, the mean, median, standard deviation, minimum,

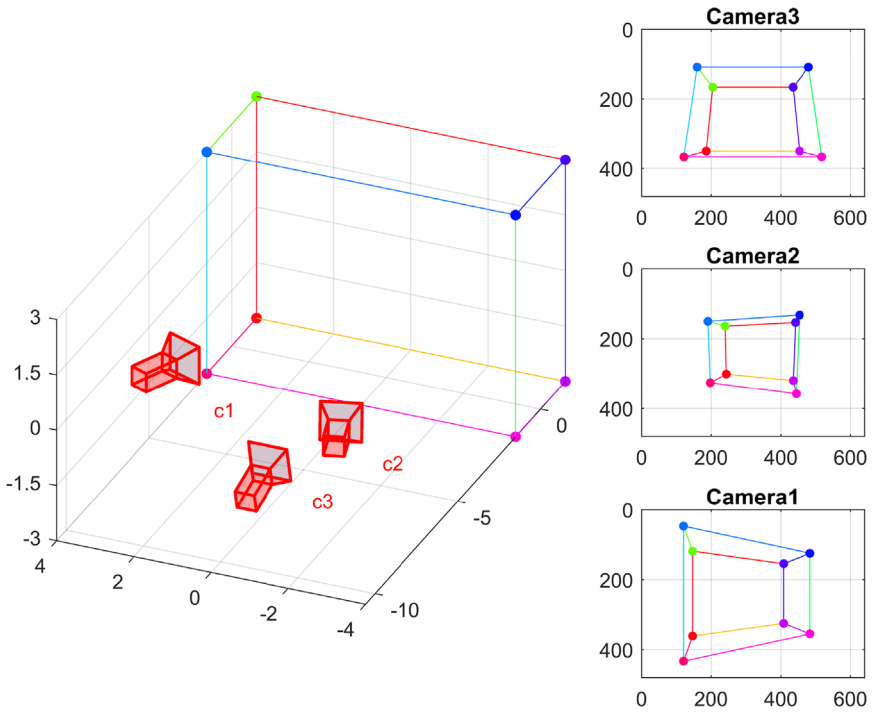

FIGURE 8 The configuration of the box and cameras $\mathbf{c}_1$ and $\mathbf{c}_2$ are the same as Figure 6, and the third camera is placed at $\mathbf{c}_3 = [-8, 0, -2]^T$.

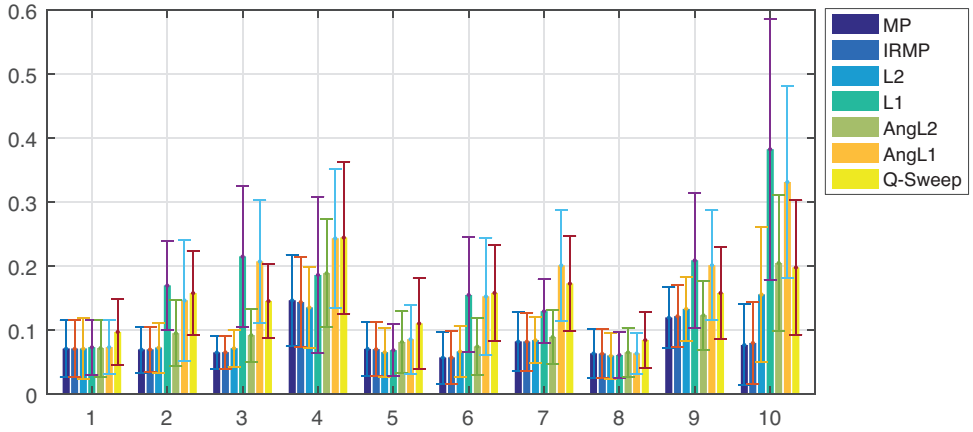

FIGURE 9 The mean and standard deviation of 3D triangulation errors for different methods in the first ten random runs on the synthetic dataset. The errors are computed for random 20 points selected in a region. The cameras and the region containing the points are shown in Figure 8.

and maximum of the mean error of all 100 experiments are computed and are shown in Table 2.

4.3 | Full reconstruction procedure on a real dataset

In this part, "Fountain-P11" dataset is used to evaluate the triangulation methods. The SURF feature correspondences [31] are used to find the essential matrices between all 2-combinations of cameras[2]. The test process is the same as the process in the

TABLE 2 The mean, standard deviation, median, minimum, and maximum of 100 mean 3D triangulation error for different methods. The errors are computed for random 20 points selected in a region. The cameras and the region containing the points are shown in Figure 8.

Method	Mean	Median	Std	Min	Max
MP	**0.0962**	**0.0829**	0.0462	0.0443	0.3076
IRMP	0.0975	**0.0829**	0.0463	0.0437	0.3097
L2	0.1063	0.0861	0.0582	0.0438	0.3579
L1	0.1880	0.1557	0.1405	0.0538	0.8181
AngL2	0.1304	0.1063	0.0760	0.0541	0.5187
AngL1	0.1949	0.1614	0.1158	0.0564	0.6609
Q-Sweep	0.1835	0.1728	0.0734	0.0846	0.4525

[2] Fountain-P11 dataset consists of 11 images from different perspectives. We remove the first and last images, which have a few number of feature correspondences and find the essential matrices for all 2-combinations of all other 9 images.

TABLE 3 The mean, standard deviation, median, minimum, and maximum of mean 3D triangulation error of all 2-view experiments on Fountain-P11 dataset.

Method	Mean $\times 10^3$	Median $\times 10^3$	Std $\times 10^3$	Min $\times 10^3$	Max $\times 10^3$
MP	**4.34**	**3.76**	2.22	1.10	9.26
IRMP	**4.34**	**3.76**	2.22	1.09	9.26
L2	4.36	3.79	2.22	1.07	9.22
L1	4.44	3.94	2.23	1.18	9.23
AngL2	4.36	3.78	2.22	1.08	9.16
AngL1	4.41	3.88	2.23	1.25	9.11
Q-Sweep	4.35	3.79	2.23	1.09	9.24

TABLE 4 The mean, standard deviation, median, minimum, and maximum of mean 3D triangulation error of all 2-view experiments on Fountain-P11 dataset.

Method	Mean $\times 10^3$	Median $\times 10^3$	Std $\times 10^3$	Min $\times 10^3$	Max $\times 10^3$
MP	**3.14**	**2.47**	2.10	1.16	14.7
IRMP	3.19	2.59	2.19	1.17	15.8
L2	3.70	3.11	3.06	1.15	23.6
L1	4.20	3.57	3.16	1.54	24.9
AngL2	6.57	4.61	8.96	1.67	63.7
AngL1	5.22	4.40	3.95	1.91	30.2
Q-Sweep	4.49	3.85	2.10	1.79	9.40

synthetic datasets except that the projection in step 2 is replaced by the feature matching. Furthermore, no noise is added to the points. For each selected pair of cameras, the process is repeated 10 times for different random corresponding points. Table 3 shows the results of different triangulation methods in the mentioned process.

The experiment is repeated for triangulation by three cameras. Again the process is the same as the process in the synthetic datasets of subsection 4.2.2 with the feature matching used in step 2. The process is repeated 10 times for any 3-combinations of the cameras. The results are shown in Table 4. As it can be seen in the results, the midpoint method and its iterative reweighted version outperform the other methods and have less mean 3D reconstruction error in all of experiments.

5 | CONCLUSION

In this paper, different triangulation methods were evaluated in terms of 3D reconstruction accuracy in a calibrated *structure-from-motion* setting. It was shown that the midpoint triangulation method, which has a closed-form solution for any number of cameras, is less sensitive to error in cameras' extrinsic parameters in comparison to several well-known methods. The performance of different methods in a *structure-from-motion* process were evaluated in synthetic and real datasets through extensive experiments. It was shown that the midpoint triangulation method outperforms the commonly used L_2 triangulation method [9] in typical practical applications, where cameras' extrinsic parameters are computed based on image registration and consequently have uncertainties. It is well-known that the midpoint method is the simplest and fastest existing non-incremental method for triangulation. Therefore, one can conclude that the midpoint method is one of the best choices for real-time SfM applications, where both accuracy and speed are important factors.

AUTHOR CONTRIBUTIONS

Seyed-Mahdi Nasiri: Conceptualization, formal analysis, methodology, software, writing - original draft. Reshad Hosseini: Conceptualization, methodology, supervision, writing - review and editing. Hadi Moradi: Conceptualization, writing - review and editing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

DATA AVAILABILITY STATEMENT

The real data associated with this paper is a public benchmark data which is freely available. The codes of all implementations used in the experiments are also available via http://visionlab.ut.ac.ir/resources/triangulation.zip.

CREDIT CONTRIBUTION STATEMENT

S-M.N. is the main author, and helped with conceptualization, doing formal analysis, developing methodology and software, and writing the original draft of the paper. R.H. is the supervisor of the team, and helped with conceptualization, developing methodology and reviewing and editing the paper. H.M. helped with conceptualization, and reviewing and editing the paper.

ORCID

Reshad Hosseini https://orcid.org/0000-0002-3669-760X

REFERENCES

1. Ferrer, J., Garcia, R.: Bias reduction for stereo triangulation. Electron. Lett. 46(25), 1665–1666 (2010)
2. Tippetts, B., Lee, D.J., Lillywhite, K., Archibald, J.: Review of stereo vision algorithms and their suitability for resource-limited systems. J. Real-Time Image Process. 11(1), 5–25 (2016)
3. Wu, W., Xu, M., Liang, Q., Mei, L., Peng, Y.: Multi-camera 3d ball tracking framework for sports video. IET Image Proc. 14(15), 3751–3761 (2020)
4. Zou, C., He, B., Zhang, L., Zhang, J.: Scene flow for 3d laser scanner and camera system. IET Image Proc. 12(4), 612–618 (2018)
5. Castle, R.O., Klein, G., Murray, D.W.: Wide-area augmented reality using camera tracking and mapping in multiple regions. Comput. Vis. Image Understand. 115(6), 854–867 (2011)

6. Bartoli, A., Sturm, P.: Structure-from-motion using lines: Representation, triangulation, and bundle adjustment. Comput. Vis. Image Understand. 100(3), 416–441 (2005)
7. Chatkaewmanee, P., Dailey, M.N.: Object virtual viewing using adaptive tri-view morphing. IET Image Proc. 7(6), 586–595 (2013)
8. Ramalingam, S., Lodha, S.K., Sturm, P.: A generic structure-from-motion framework. Comput. Vis. Image Understand. 103(3), 218–228 (2006)
9. Hartley, R.I., Sturm, P.: Triangulation. Comput. Vis. Image Understand. 68(2), 146–157 (1997)
10. Lee, S.H., Civera, J.: Closed-form optimal two-view triangulation based on angular errors. In: IEEE International Conference on Computer Vision, pp. 2681–2689. IEEE, Piscataway (2019)
11. Lindstrom, P.: Triangulation made easy. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1554–1561. IEEE, Piscataway (2010)
12. Nister, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. Pattern Anal. Mach. Intell. 26(6), 756–770 (2004)
13. Kukelova, Z., Bujnak, M., Pajdla, T.: Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In: British Machine Vision Conference, vol. 2, pp. 56.1–56.10 (2008)
14. Byröd, M., Josephson, K., Åström, K.: Improving numerical accuracy of Gröbner basis polynomial equation solvers. In: IEEE International Conference on Computer Vision, pp. 449–456. IEEE, Piscataway (2007)
15. Kanatani, K.: Statistical Optimization for Geometric Computation: Theory and Practice. Elsevier, New York (1996)
16. Stewenius, H., Schaffalitzky, F., Nister, D.: How hard is 3-view triangulation really? In: IEEE International Conference on Computer Vision, vol. 1, pp. 686–693. IEEE, Piscataway (2005)
17. Chen, J., Wu, D., Song, P., Deng, F., He, Y., Pang, S.: Multi-view triangulation: Systematic comparison and an improved method. IEEE Access 8, 21017–21027 (2020)
18. Hartley, R., Kahl, F., Olsson, C., Seo, Y.: Verifying global minima for L_2 minimization problems in multiple view geometry. Int. J. Comput. Vision 101(2), 288–304 (2013)
19. Hartley, R., Kahl, F.: Optimal algorithms in multiview geometry. In: Asian Conference on Computer Vision, pp. 13–34. Springer, Cham (2007)
20. Hartley, R., Schaffalitzky, F.: L_∞ minimization in geometric reconstruction problems. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 504–509. IEEE, Piscataway (2004)
21. Li, H.: A practical algorithm for L_∞ triangulation with outliers. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE, Piscataway (2007)
22. Olsson, C., Eriksson, A., Hartley, R.: Outlier removal using duality. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1450–1457. IEEE, Piscataway (2010)
23. Sim, K., Hartley, R.: Removing outliers using the L_∞ norm. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 485–494. IEEE, Piscataway (2006)
24. Zhang, Q., Chin, T.J., Suter, D.: Quasiconvex plane sweep for triangulation with outliers. In: IEEE International Conference on Computer Vision, pp. 920–928. IEEE, Piscataway (2017)
25. Yang, K., Fang, W., Zhao, Y., Deng, N.: Iteratively reweighted midpoint method for fast multiple view triangulation. IEEE Rob. Autom. Lett. 4(2), 708–715 (2019)
26. Fang, W., Yang, K., Li, H.: Propagation-based incremental triangulation for multiple views 3d reconstruction. Chin. Opt. Lett. 19(2), 021101 (2021)
27. Lee, S.H., Civera, J.: Triangulation: Why optimize? In: Sidorov, K., Hicks, Y. (eds.) Proceedings of the British Machine Vision Conference, pp. 61.1–61.12. BMVA Press, London (2019)
28. Toldo, R., Gherardi, R., Farenzena, M., Fusiello, A.: Hierarchical structure-and-motion recovery from uncalibrated images. Comput. Vis. Image Understand. 140, 127–143 (2015)
29. Sweeney, C., Sattler, T., Hollerer, T., Turk, M., Pollefeys, M.: Optimizing the viewing graph for structure-from-motion. In: IEEE International Conference on Computer Vision, pp. 801–809. IEEE, Piscataway (2015)
30. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4104–4113. IEEE, Piscataway (2016)
31. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). Comput. Vis. Image Understand. 110(3), 346–359 (2008)
32. Lo, T.W.R., Siebert, J.P.: Local feature extraction and matching on range images: 2.5D SIFT. Comput. Vis. Image Understand. 113(12), 1235–1250 (2009)
33. Lowe, D.G.: Object recognition from local scale-invariant features. In: IEEE International Conference on Computer Vision, vol. 2, pp. 1150–1157. IEEE, Piscataway (1999)
34. Torr, P.H., Zisserman, A.: Mlesac: A new robust estimator with application to estimating image geometry. Comput. Vis. Image Understand. 78(1), 138–156 (2000)
35. Arrigoni, F., Rossi, B., Fragneto, P., Fusiello, A.: Robust synchronization in SO(3) and SE(3) via low-rank and sparse matrix decomposition. Comput. Vis. Image Understand. 174, 95–113 (2018)
36. Chatterjee, A., Govindu, V.M.: Robust relative rotation averaging. IEEE Trans. Pattern Anal. Mach. Intell. 40(4), 958–972 (2017)
37. Hartley, R., Trumpf, J., Dai, Y., Li, H.: Rotation averaging. Int. J. Comput. Vision 103(3), 267–305 (2013)
38. Jiang, N., Cui, Z., Tan, P.: A global linear method for camera pose registration. In: IEEE International Conference on Computer Vision, pp. 481–488. IEEE, Piscataway (2013)
39. Ozyesil, O., Singer, A.: Robust camera location estimation by convex programming. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2674–2683. IEEE, Piscataway (2015)
40. Zhu, S., Zhang, R., Zhou, L., Shen, T., Fang, T., Tan, P., Quan, L.: Very large-scale global sfm by distributed motion averaging. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4568–4577. IEEE, Piscataway (2018)

How to cite this article: Nasiri, S.-M., Hosseini, R., Moradi, H.: The optimal triangulation method is not really optimal. IET Image Process. 1–11 (2023). https://doi.org/10.1049/ipr2.12831