


RESEARCH ARTICLE

SLAMB&MAI: a comprehensive methodology for SLAM benchmark and map accuracy improvement

Shengshu Liu , Erhui Sun and Xin Dong

Department of Mechanical, Materials and Manufacturing Engineering, University of Nottingham, Nottingham NG7 2RD, UK
Corresponding author: Xin Dong; Email: xin.dong@nottingham.ac.uk

Received: 1 August 2023; **Revised:** 21 December 2023; **Accepted:** 25 December 2023;
First published online: 30 January 2024

Keywords: comprehensive and objective SLAM benchmark; map ground truth; map error correction; map accuracy improvement; ground truth acquisition

Abstract

SLAM Benchmark plays a pivotal role in the field by providing a common ground for performance evaluation. In this paper, a novel methodology of simultaneous localization and mapping benchmark and map accuracy improvement (SLAMB&MAI) is introduced. It can objectively evaluate errors of localization and mapping, and further improve map accuracy by utilizing evaluation results as feedback. The proposed benchmark transforms all elements into a global frame and measures the errors between them. The comprehensiveness consists in the benchmark of both localization and mapping, and the objectivity consists in the consideration of the correlation between localization and mapping by the preservation of the original pose relations between all reference frames. The map accuracy improvement is realized by first obtaining the optimization that minimizes the errors between the estimated trajectory and ground truth trajectory and then applying it to the estimated map. The experimental results showed that the map accuracy can be improved by an average of 15%. The optimization that yields minimal localization errors is obtained by the proposed Centre Point Registration-Iterative Closest Point (CPR-ICP). This proposed Iterative Closest Point (ICP) variant pre-aligns two point clouds by their centroids and least square planes and then uses traditional ICP to minimize the error between them. The experimental results showed that CPR-ICP outperformed traditional ICP, especially in cases involving large-scale environments. To the extent of our knowledge, this is the first work that can not only objectively benchmark both localization and mapping but also revise the estimated map and increase its accuracy, which provides insights into the acquisition of ground truth map and robot navigation.

1. Introduction

Simultaneous localization and mapping (SLAM), also known as concurrent mapping and localization, is one of the fundamental challenges for robotics, dealing with the necessity to build a map of the environment while simultaneously determining the location of the robot within this map. The map is also referred to as a geometry or a 3D surface reconstruction, and these three terms are used in this paper with no distinction. In a general SLAM system, the performances of localization and mapping depend largely on SLAM algorithms, and the quantitative accuracy is determined by SLAM benchmarks. Regarding a general SLAM benchmark procedure, there are two challenges that need to be addressed: (1) How to benchmark SLAM systems comprehensively and objectively regarding the accuracy of localization and mapping. (2) How to revise the estimated map, making it closer to ground truth. The aim of addressing the first challenge is to obtain unbiased evaluations that represent actual SLAM performances, and the aim of addressing the second challenge is to generate high-precision maps that have the potential to be used as ground truth. A comprehensive and objective SLAM benchmark is supposed to evaluate the performances of localization and mapping dependently in a global scope instead of benchmarking them independently. However, most existing SLAM benchmarks only evaluate localization performances but

leave out the assessment on mapping due to the unavailability of the map ground truth; and for those that evaluate both aspects, they evaluate them independently which always leads to biased results.

The current literature on localization benchmark mainly provides open-sourced imagery datasets, pose ground truth, methodologies for acquiring pose ground truth and evaluation metrics, among which TUM RGB-D [1], KITTI [2] and EuRoc [3] are used most widely. The acquisition of pose ground truth mostly makes use of external high-precision apparatus such as high-precision reference maps, RTK-GPS, motion capture systems, laser range finders, sonar systems, etc. In ref. [4], the robot pose ground truth was obtained by using Monte Carlo Localization with the help of a high-precision reference map, and the performances of different localization approaches were benchmarked against this ground truth using Euclidean distance and orientation difference as the evaluation metric. The dataset provided in Rawseeds project [5] was collected by a mobile robot platform equipped with a variety of sensors, which used RTK-GPS apparatus and a fusion of vision-based and laser-based sensors for ground truth acquisition in both outdoor and indoor environments respectively. Meanwhile, the localization accuracy was evaluated against the ground truth with a similar metric used in refs. [4, 6]. Similarly, the utilization of GPS was adopted in ref. [7] for the generation of vision benchmark datasets. Furthermore, the work also proposed a novel Average Orientation Similarity metric for the performance evaluation of object detection and 3D orientation estimation. GPS is mostly used in outdoor environments for the acquisition of pose ground truth, while motion capture systems are more suitable for indoor small-scale environments. In refs. [1] and [8], the motion capture system was used to track reflective visual markers affixed to RGB-D cameras and output camera poses at a certain frame rate. Reference [1] also offered two error metrics to evaluate pose tracking accuracy, which were Relative Pose Error (RPE) and Absolute Trajectory Error (ATE). The RPE was used in ref. [8] for the evaluation of camera trajectory accuracy.

However, the aforementioned methodologies of pose ground truth acquisition are oftentimes laborious and require external high-precision tools that could be unavailable under some circumstances. Therefore, the approach of generating and utilizing synthesized pose ground truth has been adopted by many researches [9–12]. For this approach, the SLAM algorithms were first run in real-world environments, and the estimated camera trajectory was then fitted into the synthesized imagery scenes by rigid body transformation. In synthesized imagery scenes, the transformed poses were used as the ground truth trajectory to control the movements of a virtual camera. Finally, the errors between the estimated trajectory and ground truth trajectory were obtained using the error metric ATE proposed in ref. [1].

The majority of research for mapping benchmarks is actually about the benchmark of 3D reconstruction of small or medium-sized objects with available CAD models as ground truth. A common methodology used in the current literature [13–17] is firstly aligning the estimated map with the ground truth map using a 3D shape registration approach [18–25] then calculating errors based on an error metric [26–29]. This methodology has two major issues:

(1) The methodology is mainly used for the benchmark of 3D reconstruction of small or medium size objects, of which the geometry ground truth is acquired by using high-precision 3D scanners, which are not applicable for room-sized or large-scale scenes that appear in many SLAM scenarios. This issue is a common SLAM problem, which is also formulated as the acquisition of map ground truth for large-scale scenes, especially the scenes that are inaccessible to human beings, such as nuclear facilities and chemical plant pipelines, etc.

(2) An objective mapping benchmark is supposed to preserve the original pose relation between the estimated map and ground truth map, and then calculate the errors between corresponding elements in two maps. Thus, the calculated errors can objectively represent the unbiased performance of the SLAM system. Manual alignments between the estimated map and ground truth map will obfuscate the original errors and result in biased evaluations.

The work presented in ref. [10] addressed the first issue by providing synthetic datasets of a living room scene and an office room scene with both surface and trajectory ground truth. However, because it benchmarked the performances of localization and mapping independently, the second issue was still unsolved. Meister et al. [30] also addressed the first issue by evaluating the 3D reconstruction system KinectFusion against different objects and scenes to investigate its capabilities for generating accurate 3D surface reconstruction that can be used as geometry ground truth. The work concluded

that KinectFusion was reliable in resolving object details with a minimum size of approximately 10 mm and provided a high-quality dataset consisting of 57 scenes that can be used as training data for 3D recognition and depth inpainting. For the second issue, Wasenmüller et al. [8] addressed it by placing all estimates and ground truth in a global coordinate system while preserving the original orientational relation and positional relation between the estimates and ground truth. In this work, a Microsoft Kinect v2 was adopted as the Visual-SLAM sensor. Meanwhile, two other vision systems were used for ground truth acquisition, which were a high-precision 3D scanner for geometry ground truth acquisition and a motion capture system for pose ground truth acquisition. However, the first issue was not addressed in this work. Moreover, because the three vision systems had their respective independent coordinate systems, complex rigid body transformations were needed for the global evaluation, which increased the computational costs. Besides, system errors would accumulate with the increase in the number of sensor systems used, since each sensor system introduces its unique system error into the benchmark process, which increases the likelihood of biased evaluations.

The benchmark methodology SLAMB&MAI proposed in this paper addresses the two aforementioned issues. The objectivity and comprehensiveness of the methodology consist in benchmarking localization and mapping performances dependently in a global coordinate system. Furthermore, the utilization of a high-precision motion capture system for acquiring the ground truth of both poses and the map introduces less computational loads and system errors than the method proposed in ref. [8]. In addition, SLAMB&MAI further reduces map errors and improves the map accuracy by utilizing pose errors, which revises the estimated map and makes it closer to ground truth. Finally, the methodology of revising the map offers insights into the methodologies of acquiring geometry ground truth with SLAM systems.

Unlike the performance investigation of KinectFusion proposed in ref. [30], the proposed SLAMB&MAI method doesn't have a demanding requirement on the accuracy of SLAM systems when acquiring geometry ground truth, because it uses benchmark results as feedbacks to correct geometry errors and generate accurate geometries that are eligible to be used as ground truth, which has the potential to be applied in numerous SLAM systems for acquiring geometry ground truth of different size objects and scenes.

2. Approach for SLAM benchmark and map accuracy improvement

2.1. Application scenarios

In order to obtain the high-precision maps of large-scale scenes that don't have their CAD models as ground truth, such as legacy nuclear facilities, historical heritage sites, chemical plants, etc, a novel methodology is proposed in this section. As shown in Fig. 1, the SLAM camera is firstly moved along a known ground truth trajectory (red solid arrow line), and the estimated trajectory (green dash arrow line) and the estimated map (chromatic point cloud) are generated based on the SLAM algorithm. Then the optimal error correction is calculated by minimizing the error between the ground truth trajectory and the estimated trajectory. This optimal error correction is then applied to optimize the estimated map to generate a high-precision map. Finally, the mapping error is evaluated by comparing the high-precision map with the ground truth map of the environment (here denotes the 3D model of the corridor available in benchmark). The proposed methodology is presented and detailed in Sections 2.2 and 2.3.

2.2. Methodology of a comprehensive and objective SLAM benchmark

Compared with the current literature, the proposed novel benchmark methodology consists in benchmarking localization and mapping in a global coordinate system.

As shown in Fig. 2, a SLAM camera (RGB-D camera in the example) is used to map a 3D object (profile of a house in the example). To realize a comprehensive and objective SLAM benchmark, a high-precision measuring tool (e.g. VICON motion capture system) is needed for acquiring the ground

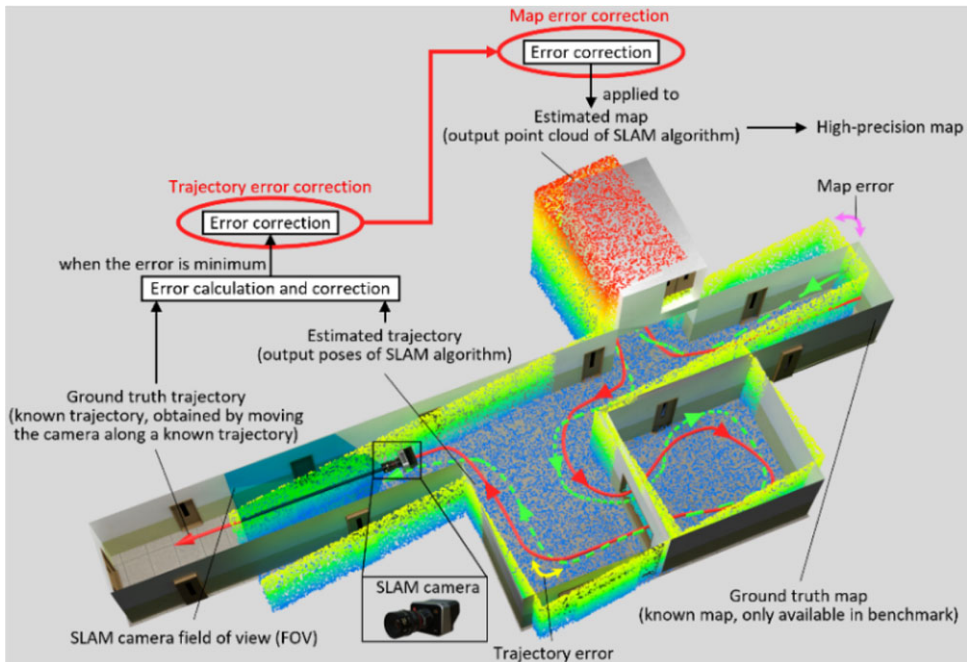


Figure 1. Concept of simultaneous localization and mapping benchmark and map accuracy improvement. The error correction minimizing trajectory errors is then applied to the estimated map to generate the high-precision map. One example of the potential applications is the mapping of a corridor.

truth of the trajectory and map, and a global coordinate system needs to be established so that the ground truth and estimates of the trajectory and map can be transformed into it for further evaluation.

As shown in Fig. 2(a), the trajectory ground truth P^* and map ground truth Q^* are acquired by using the high-accuracy vision system, working with a plurality of markers placed on the SLAM camera and object. The ground truth acquisition system is set up at a high position so that it can see all the markers, and its coordinate system is used as the global coordinate frame G . Several markers are placed at corners of the mapping object so that the ground truth mapping Q^* can be obtained directly without any pose transformation.

As shown in Fig. 2(b), the reference frame of the SLAM camera is denoted as Sc (SLAM camera frame), and the trajectory of its origin is considered as the SLAM camera trajectory. For acquiring the ground truth of the SLAM camera trajectory P^* , a 3D-printed part is attached to the back of the SLAM camera with the attachment of three markers. The reference frame of these three markers is denoted as Sca (SLAM camera attachment frame); its origin is located at the centroid of the triangle encircled by the three markers, and its x and y axes are parallel to the attachment length and width respectively. T_{Sca}^G denotes the pose transformation from Sca to G ; it is being captured by the ground truth acquisition system constantly, and the position of Sc origin within Sca is a known factor; therefore, the ground truth of the SLAM camera trajectory can be obtained.

The markers on the SLAM camera attachment are placed in an asymmetric pattern, and the reason for this is that the symmetry of the rectangle encircled by four markers oftentimes obfuscates the ground truth acquisition system, which causes the pose of Sca to change during the operation. By contrast, the triangle encircled by three markers is asymmetric, which increases the robustness of the pose of Sca and makes it less likely to change during the operation.

The estimated trajectory and map obtained from the SLAM algorithm are originally within Sc , and the pose transformation from Sc to Sca is denoted as T_{Sc}^{Sca} which is also a known factor; therefore, the pose transformation from Sc to G can be expressed as $T_{Sc}^G = T_{Sca}^G T_{Sc}^{Sca}$. In this way, the estimated trajectory

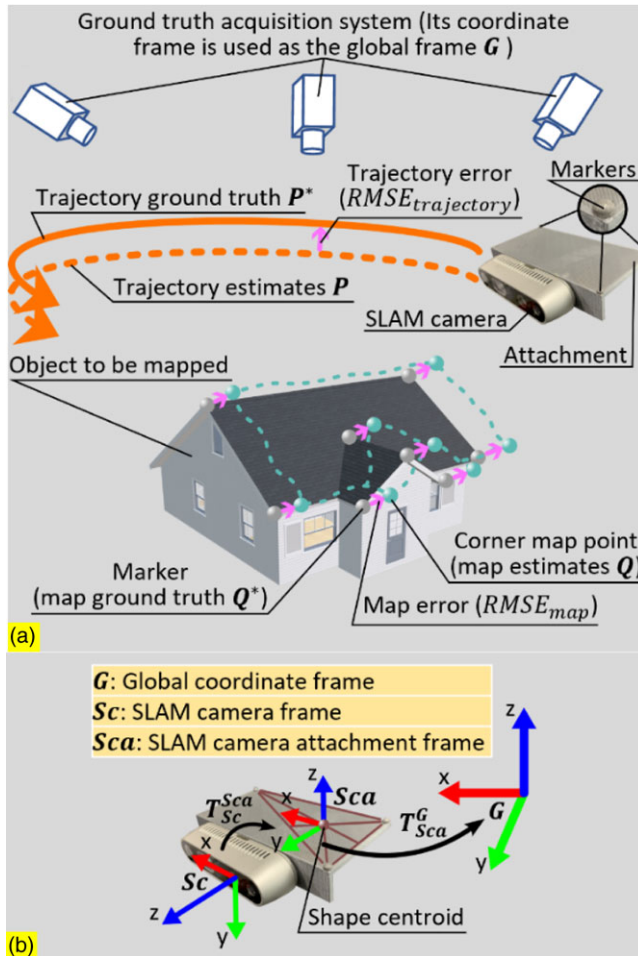


Figure 2. Schematic diagram of the benchmark methodology: (a) General concept of the methodology. To ensure the benchmark is comprehensive and objective, a high-precision measuring tool is used for ground truth acquisition, and its coordinate frame is used as the global frame, which all estimates and ground truth are transformed into. (b) Pose transformations between different frames. Except that the map ground truth Q^* can be obtained directly, the acquisition of the other three variables (map estimates Q , trajectory ground truth P^* , trajectory estimates P) requires Sca to achieve the transformation from their respective reference frames to the global frame G .

P and map Q within the global frame G can be obtained. In the global frame G , for all positions from the estimated camera trajectory $p_1, \dots, p_n \in P$, their respective nearest neighboring positions from the ground truth trajectory are denoted as $p_1^*, \dots, p_n^* \in P^*$; for all positions from the estimated map $q_1, \dots, q_m \in Q$, their respective nearest neighboring points or mesh surfaces are denoted as $q_1^*, \dots, q_m^* \in Q^*$. For the camera trajectory, the nearest neighbor distance is point-to-point distance, and the Root Mean Squared Error (RMSE) overall nearest neighbor distances is expressed in Eq. (1). For the map, the nearest neighbor distance could be either point-to-point distance or point-to-mesh distance [26, 27], and the RMSE of them is collectively expressed in Eq. (2). The RMSE is used as the evaluation metric for error analysis shown in Section 3.

$$RMSE_{trajectory} = \left(\frac{1}{n} \sum_{i=1}^n \|p_i - p_i^*\|^2 \right)^{1/2} \quad (1)$$

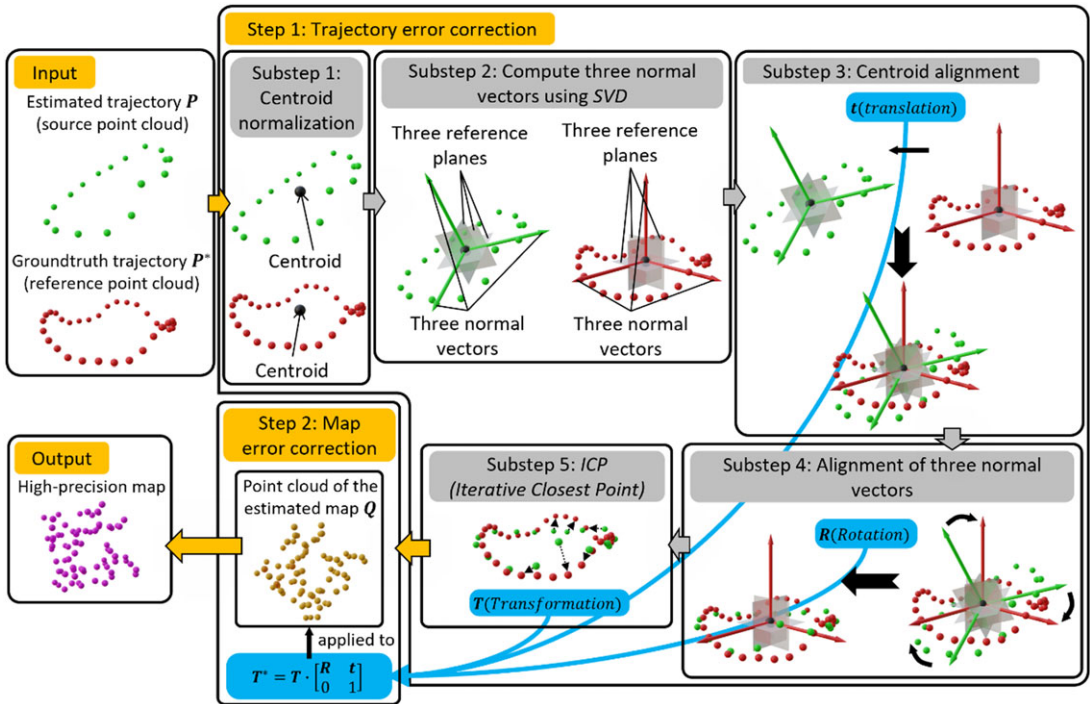


Figure 3. Workflow of the map accuracy improvement system. The approach used in the first step (trajectory error correction) is Centre Point Registration-Iterative Closest Point (CPR-ICP); it is subdivided into five substeps. The first four substeps constitute the CPR operation and are followed by the final ICP step. The output of CPR-ICP is a rigid transformation; it is fed into the step of map error correction and the revised high-precision map is obtained.

$$RMSE_{map} = \left(\frac{1}{m} \sum_{i=1}^m \|q_i - q_i^*\|^2 \right)^{1/2} \tag{2}$$

2.3. Methodology of map accuracy improvement

The motivation of the proposed novel methodology is to improve the accuracy of the estimated map in order to generate a high-precision map (geometry). As shown in Fig. 1, the novel methodology can be formulated as firstly finding the best error optimization between the estimated trajectory and ground truth trajectory and then applying this error optimization on the estimated map to obtain the high-precision map. The rationale behind this is that there exist correlations between the trajectory error and geometry (yellow and pink curve double arrow lines). As shown in Fig. 3, the workflow of the map accuracy improvement system comprises two steps, that is trajectory error correction (Step 1) and map error correction (Step 2). The trajectory error correction is in essence the registration of the point cloud of the estimated trajectory (green point cloud) to the point cloud of the ground truth trajectory (red point cloud). The commonly employed approach is Iterative Closest Point (ICP) [19, 31–33] method. For traditional ICP method, if there is no decent feature matching between two point clouds to start with, it cannot achieve good results. Therefore, a novel ICP, that is Centre Point Registration-ICP (CPR-ICP), is proposed to tackle this problem. The Centre Point Registration (CPR) is at the core of CPR-ICP and it is a pre-alignment process for the two point clouds prior to the ICP operation. The aim of CPR is to bring closer the same features in two point clouds and create a decent feature matching for the subsequent

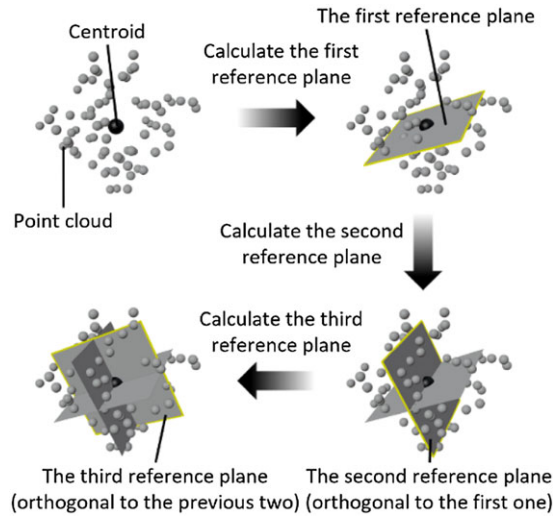


Figure 4. The objective of Centre Point Registration. Being orthogonal to each other, the three reference planes are fitted with the least squares method and they all pass through the centroid of the point cloud.

ICP operation to start with. As shown in Fig. 4, CPR is achieved by finding and aligning three reference planes for each point cloud, and the three reference planes should have the following attributes:

- The first reference plane is achieved by minimizing the sum of the squares of distances between points and the plane.
- The second plane has two attributes: (1) orthogonal to the first reference plane; (2) minimization of the sum of the squares of the distances between the points and the plane.
- The third plane has three attributes: (1) orthogonal to the first reference plane; (2) orthogonal to the second reference plane; (3) minimization of the sum of the squares of the distances between the points and the plane is the minimum.

The problem of finding the three reference planes can be formulated in a mathematical way: given a point cloud $P = \{p_i | p_i = [x_i y_i z_i]^T, i \in \mathbf{N} \cap i \in [1, n]\}$ and a plane $Ax + By + Cz + D = 0$ with a unit vector of $\hat{u} = [ABC]^T$, the sum of the squares of the distances between the points and the plane can be expressed as:

$$f(\hat{u}, D) = \sum_{i=1}^n (\hat{u}^T p_i + D)^2 \tag{3}$$

Find the \hat{u} and D that make (4) hold.

$$\langle \hat{u}, D \rangle = \min_{\hat{u}, D} f(\hat{u}, D) \tag{4}$$

If $f_b(\hat{u}, D) = 0$, $D = -\hat{u}^T \bar{p}$ where $\bar{p} = \frac{\sum_{i=1}^n p_i}{n}$ is the centroid of the point cloud P . This means that regardless of the normal unit vector \hat{u} , the planes making (4) hold definitely pass through the point cloud centroid \bar{p} (as shown in Fig. 4, all three reference planes pass through the centroid); therefore, the problem can be further formulated as: given a point cloud matrix $P = [p_1, p_2, \dots, p_n]$ where $p_i = [x_i y_i z_i]^T$, $\bar{p} = \frac{\sum_{i=1}^n p_i}{n}$ and $\mathbf{1} = [1, 1, \dots, 1]_{n \times 1}^T$, find the unit vector $\hat{u} = [A B C]^T$ that makes (5) hold:

$$\hat{u} = \min_{\hat{u}} \hat{u}^T (P - \bar{p}\mathbf{1}^T) (P - \bar{p}\mathbf{1}^T)^T \hat{u} \tag{5}$$

In (5), $(\mathbf{P} - \bar{\mathbf{p}}\mathbf{1}^T)(\mathbf{P} - \bar{\mathbf{p}}\mathbf{1}^T)^T$ is a symmetric matrix, so it can be factorized into the form of $\mathbf{U}\Sigma\mathbf{U}^T$ using SVD (Singular Value Decomposition) [34]; therefore, (5) can also be expressed as:

$$\hat{\mathbf{u}} = \min_{\hat{\mathbf{u}}} \hat{\mathbf{u}}^T \mathbf{U} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \mathbf{U}^T \hat{\mathbf{u}} \quad (6)$$

In (6), \mathbf{U} is an orthogonal matrix and $\lambda_1, \lambda_2, \lambda_3$ are the singular values and $\lambda_1 \geq \lambda_2 \geq \lambda_3$; therefore, $\mathbf{U}^T \hat{\mathbf{u}}$ is still a unit vector and can be denoted as $\hat{\mathbf{u}}' = [A' B' C']^T$, so (6) can also be expressed as:

$$\hat{\mathbf{u}}' = \min_{\hat{\mathbf{u}}'} \lambda_1 A'^2 + \lambda_2 B'^2 + \lambda_3 C'^2 \quad (7)$$

(7) holds if $\hat{\mathbf{u}}' = [0 0 1]^T$, which means $\hat{\mathbf{u}}$ equals the last column vector of \mathbf{U} ; therefore, the last column vector of \mathbf{U} is the normal vector of the first reference plane. To find the second reference plane, the above calculation is repeated. Since the second reference plane also passes through the centroid and is orthogonal to the first reference plane, its unit vector $\hat{\mathbf{u}}' = [A' B' 0]^T$, so (7) reaches the minimum value λ_2 if $\hat{\mathbf{u}}' = [0 1 0]^T$, which means $\hat{\mathbf{u}}$ equals the second to last column vector of \mathbf{U} ; therefore, the second to last column vector of \mathbf{U} is the normal vector of the second reference plane, and the first column vector of \mathbf{U} is the normal vector of the third reference plane. To sum up, this calculation process firstly normalizes every point in the point cloud by the centroid, and then factorizes the square of the normalized point cloud matrix using SVD; therefore, the CPR operation can be elaborated as the following four steps:

1. Compute the centroid (the mean of all point coordinates $\bar{\mathbf{p}} = \frac{\sum_{i=1}^n \mathbf{p}_i}{n}$) of each point cloud and normalize each point cloud by the centroid, which is corresponding to Substep 1 of Step 1 shown in Fig. 3.
2. Apply SVD to each normalized point cloud and compute the three normal vectors, which is corresponding to Substep 2 of Step 1 shown in Fig. 3.
3. Translate the source point cloud (the estimate) to the reference point cloud (the ground truth) and align their centroids, which is corresponding to Substep 3 of Step 1 shown in Fig. 3.
4. Rotate the translated source point cloud (the estimate) around its centroid and align its three normal vectors to those of the reference point cloud (the ground truth), which is corresponding to Substep 4 of Step 1 shown in Fig. 3.

3. Experiment and analysis

The experimental validation of the proposed methodology is categorized into three different scenes: experiments of small-scale scenes, medium-scale scenes, and large-scale scenes. To clearly demonstrate the advantage of CPR-ICP over the traditional ICP, before it is tested in SLAM environments, it is first tested on the famous ‘‘Stanford Bunny’’ dataset [35–37], as shown in Fig. 5. In this case, the source point cloud (yellow bunny) is deliberately placed in a pose that differs vastly from that of the target point cloud (grey bunny). Under such a circumstance, the pre-alignment (CPR) can generate a decent feature matching between two point clouds; therefore, CPR-ICP demonstrates an ideal alignment as shown in Fig. 5(b). In contrast, due to not having a decent feature matching to start with, the traditional ICP demonstrates a poor alignment between two point clouds as shown in Fig. 5(a).

3.1. Experiments of small-scale scenes

The proposed methodology is then tested on a small-scale SLAM scene. As shown in Fig. 6(a), in the experiment, a Vicon motion capture system is used for acquiring ground truth and providing global frame \mathbf{G} ; a laptop (SLAM laptop) with an Intel Core i9-10980HK CPU (2.40 GHz, 8 Cores, 16 Logical Processors) and 32 gigabytes memory is used to run the SLAM algorithm ORB-SLAM2 [38]; and

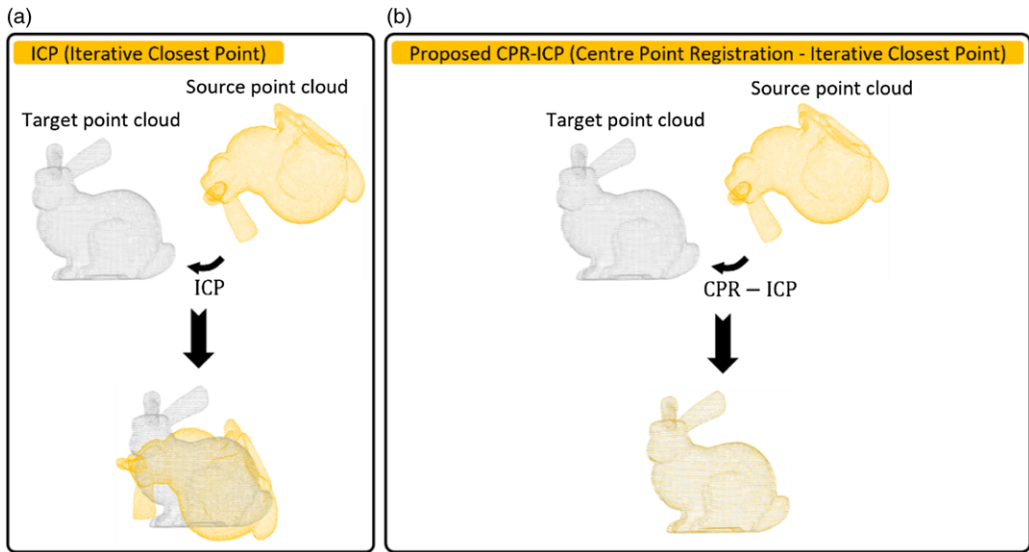


Figure 5. Performance comparison between Iterative Closest Point (ICP) and Centre Point Registration-ICP (CPR-ICP) under the circumstance of a decent initial feature matching being absent. (a) Performance of the traditional ICP: the misalignment between two bunnies demonstrates the poor performance of ICP. It is due to the lack of a decent feature matching to start with. (b) Performance of CPR-ICP: the fine alignment between two bunnies demonstrates the ideal performance of CPR-ICP. CPR-ICP firstly conducts a rough alignment of two point clouds and then conducts the ICP operation, making it robust to achieve an ideal performance under such an extreme circumstance.

an Intel RealSense D435i RGB-D camera is affixed at the top of the SLAM laptop and used as the SLAM vision sensor (SLAM camera). As shown in Fig. 6(b), the map being evaluated consists of three points that are centroids of three rectangular regions on a 75×75 cm table surface. The table surface is partitioned into four equal rectangular regions by tapes with color textures. Among the four rectangular regions, twelve markers are placed at the corners of three rectangular regions; in this way, the centroid coordinates (three blue points) of these three rectangular regions can be captured directly by the motion capture system and used as the ground truth map Q^* , as is shown in Fig. 6(c). The ground truth map P^* , estimated trajectory P and estimated map Q are also shown in the form of point clouds in Fig. 6(c).

During the experimental operation, the SLAM camera scans the table surface constantly to generate its point cloud; in the meantime, the pose of the SLAM camera attachment frame (Sca) relative to the global frame G is being captured constantly by the motion capture system so that P , P^* , Q and Q^* can be obtained within the global frame G using the pose transformations shown in Section 2. Within the generated point cloud of the table surface, for each rectangular region, the four map points at four corners are chosen manually and their coordinates are averaged, and these three mean coordinates are used as the estimated map Q (three yellow points). After P , P^* , Q and Q^* are obtained within the global frame G , for the purpose of benchmarks, the RMSE errors of the trajectory and map are calculated based on Eqs. (1) and (2). Then the error correction function CPR-ICP (obtained in Step 1 shown in Fig. 3) or traditional ICP can be obtained, it is applied on the estimated map Q and the revised map (three purple points) is obtained, as is shown in Fig. 1 and Step 2 in Fig. 3. Then, the RMSE errors between the revised trajectory and map and the ground truth trajectory and map are calculated as well based on Eqs. (1) and (2), respectively.

As shown in Table I, the test results show that the concept of correcting map errors using trajectory error correction works fine; it can reduce the map error by an average of around 15%. Among all test results, for either trajectory or map, ICP and CPR-ICP reduce roughly an equal amount of errors. The

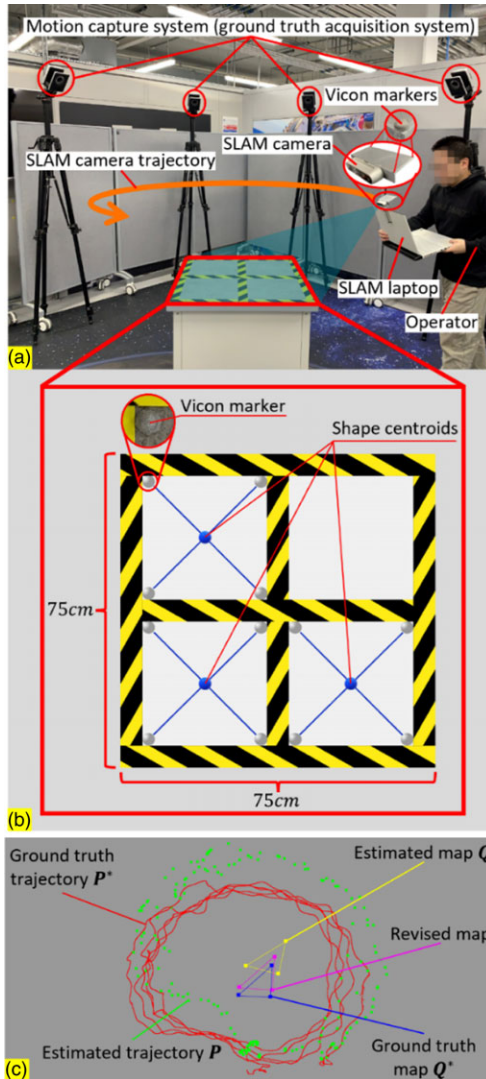


Figure 6. Experimental setup and operation of simultaneous localization and mapping benchmark and map accuracy improvement (SLAMB&MAI). (a) The SLAM system consists of an RGB-D camera (SLAM camera) and a laptop (SLAM laptop); the SLAM camera scans the table surface and feeds the image data into the SLAM laptop to generate the point cloud. During the scanning operation, the SLAM laptop is carried by an operator moving around the table surface, meanwhile, the pose of the SLAM camera is constantly adjusted by the operator to make sure the entire table surface stays in SLAM camera's field of view and the SLAM camera attachment frame (Sca) stays in Vicon's field of view throughout the operation process. After the scanning operation is done, all obtained coordinates are transformed into the global frame G . In the transformed map point cloud, the map points at four corners of the three rectangular regions are chosen manually, and their mean coordinate is used as the map estimate; then the errors of SLAM camera trajectory and map are calculated, and the error correction function of camera trajectory is obtained and applied to the map estimates to obtain the high-precision map. (b) The layout of the map setup on a table surface. (c) The point cloud of the captured data under the global coordinate frame G . The three blue points represent the ground truth map Q^* ; the three yellow points represent the estimated map Q ; the three purple points represent the revised map; the green points represent the estimated trajectory P ; and the dense red points represent the ground truth trajectory P^* .

Table I. Errors of map and trajectory before and after Centre Point Registration-Iterative Closest Point correction for the small-scale scene.

Trial number	Original trajectory errors (mm)	Trajectory errors after CPR-ICP correction (mm)	Amount of trajectory error correction using CPR-ICP (mm/%)	Original map errors (mm)	Map errors after CPR-ICP correction (mm)	Amount of map error correction using CPR-ICP (mm/%)
1	66.59	61.21	5.38 (8.08%)	59.54	51.20	8.34 (14.01%)
2	72.09	63.83	8.26 (11.46%)	67.26	54.22	13.03 (19.38%)
3	105.98	101.24	4.74 (4.48%)	68.73	60.06	8.67 (12.62%)
4	103.99	94.34	9.65 (9.28%)	139.88	112.44	27.45 (19.62%)
5	62.58	57.92	4.66 (7.45%)	65.77	55.21	10.57(16.07%)

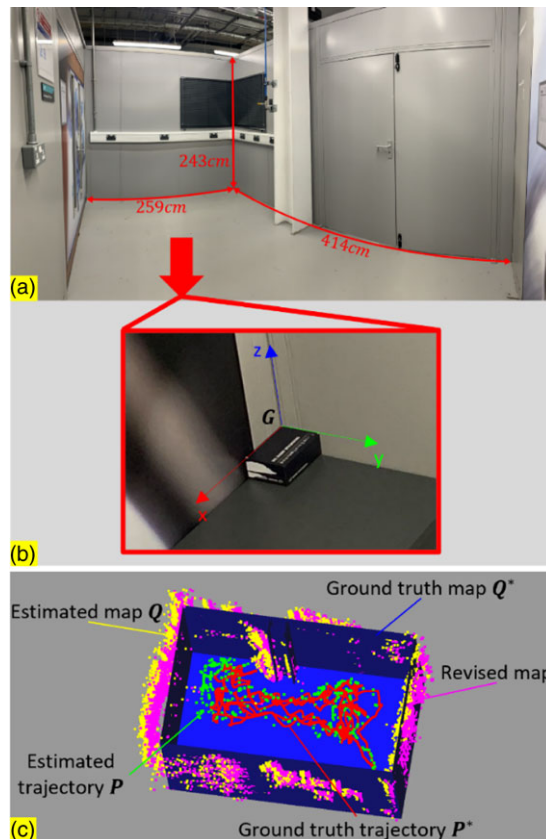


Figure 7. Experiment setup in medium-scale environments and results. (a) A panoramic photo of the medium-scale scene. (b) The global coordinate frame G was established at a corner where two walls meet, and its relative position to the small-scale room is known. (c) The point clouds of the ground truth and estimate of the camera trajectory and map. The blue CAD model represents the ground truth map Q^* , its relative position to G is known; the yellow point cloud represents the estimated map Q ; the purple point cloud represents the revised map; the green point cloud represents the estimated trajectory P ; and the dense red point cloud represents the ground truth trajectory P^* . They were all captured under the global frame G .

Table II. Errors of map and trajectory before and after Centre Point Registration-Iterative Closest Point correction for the medium-scale scene.

Trial number	Original trajectory errors (mm)	Trajectory errors after CPR-ICP correction (mm)	Amount of trajectory error correction using CPR-ICP (mm/%)	Original map errors (mm)	Map errors after CPR-ICP correction (mm)	Amount of map error correction using CPR-ICP (mm/%)
1	116.19	90.27	25.92 (22.30%)	78.98	67.21	11.76 (14.90%)
2	61.15	51.70	9.45 (15.46%)	70.43	62.72	7.71 (10.95%)
3	147.27	100.27	47.00 (31.91%)	79.40	69.95	9.45 (11.91%)
4	124.01	102.53	21.48 (17.32%)	71.13	61.96	9.17 (12.89%)
5	115.67	82.67	33.00 (28.53%)	95.78	82.65	13.13 (13.71%)

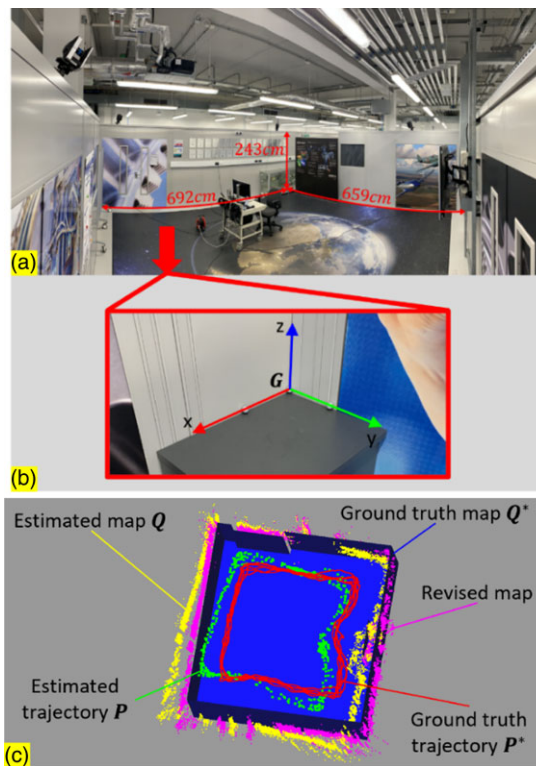


Figure 8. Experiment setup in large-scale environments and results. (a) A panoramic photo of the large-scale scene. (b) The global coordinate frame G is established at a corner where two walls meet, and its relative position to the small-scale room is known. (c) The point clouds of the ground truth and estimate of the camera trajectory and map. The blue CAD model represents the ground truth map Q^* , its relative position to G is known; the yellow point cloud represents the estimated map Q ; the purple point cloud represents the revised map; the green point cloud represents the estimated trajectory P ; and the dense red point cloud represents the ground truth trajectory P^* . They were all captured under the global frame G .

Table III. Errors of map and trajectory before and after Centre Point Registration-Iterative Closest Point correction for the large-scale scene.

Trial number	Original trajectory errors (mm)	Trajectory errors after CPR-ICP correction (mm)	Amount of trajectory error correction using CPR-ICP (mm/%)	Original map errors (mm)	Map errors after CPR-ICP correction (mm)	Amount of map error correction using CPR-ICP (mm/%)
1	444.54	181.47	263.07 (59.18%)	310.52	221.58	88.94 (28.64%)
2	549.83	250.14	299.69 (54.51%)	269.21	226.53	42.68 (15.85%)
3	465.78	215.62	250.16 (53.71%)	285.05	210.35	74.70 (26.21%)
4	314.10	182.03	132.06 (42.04%)	281.08	235.32	45.76 (16.28%)
5	346.53	220.00	126.53 (36.51%)	268.00	200.40	67.60 (25.22%)

reason for this is that in the small-scale scene, the point clouds of estimates and ground truth tend to have a decent initial feature matching between them; therefore, the CPR correction makes a modest contribution. But for a point cloud registration case that doesn't have a decent initial feature matching like the one shown in Fig. 5, the CPR correction makes a significant difference. To further verify the practical applicability of the proposed methodology for scenes with larger scales, experiments in a medium-scale scene and a large-scale scene were conducted and are detailed in 3.2 and 3.3, respectively.

3.2. Experiments of medium-scale scenes

Similar experiments were then carried out in a medium-scale room with the size of 243 cm × 259 cm × 414 cm, as is shown in Fig. 7(a). The captured data shown in Fig. 7(c) are similar to the one shown in Fig. 6(c), except that the ground truth map Q^* is a manually drawn CAD model instead of a point cloud. The global frame G shown in Fig. 7(b) was established at a corner where two walls meet, and the relative position of Q^* to G was pre-obtained by manual measurement. Using the manually drawn CAD model as ground truth map would introduce errors; therefore, the actual performance of the proposed methodology is better than the test results shown in Table II.

As shown in Table II, the concept of correcting map errors using trajectory error correction also works for medium-scale scenes; the map error was reduced by an average of around 12%. As mentioned before, using the manually drawn CAD model as ground truth map would introduce errors, so it's safe to say the proposed methodology is able to reduce the map error by at least 12%. Among all the test results, 40% of them of CPR-ICP are better than that of ICP, which is higher than the 20% for the small-scale scene. This result shows that CPR-ICP tends to outperform ICP in scenes with larger scales, so more experiments were conducted in a large-scale scene and are presented in the next section.

3.3. Experiments of large-scale scenes

The proposed methodology is then tested on a large-scale room with the size of 243cm × 692cm × 659cm as shown in Fig. 8. Same as the test in medium-scale scenes, using the manually drawn CAD model as ground truth map would introduce errors, so the actual performance of the proposed methodology is better than the test results shown in Table III.

As can be seen from Table III, the concept of correcting map errors using trajectory error correction works for large-scale scenes as well; the proposed methodology was able to reduce the map error by around 20%. Among all test results, 80% of them of CPR-ICP are better than that of ICP, and 40% are better by above 15%. Combined with the 0% for small-scale scenes and 40% for medium-scale scenes, the ratios of test results in which CPR-ICP is better than ICP in different scale scenes are shown in Fig. 9.

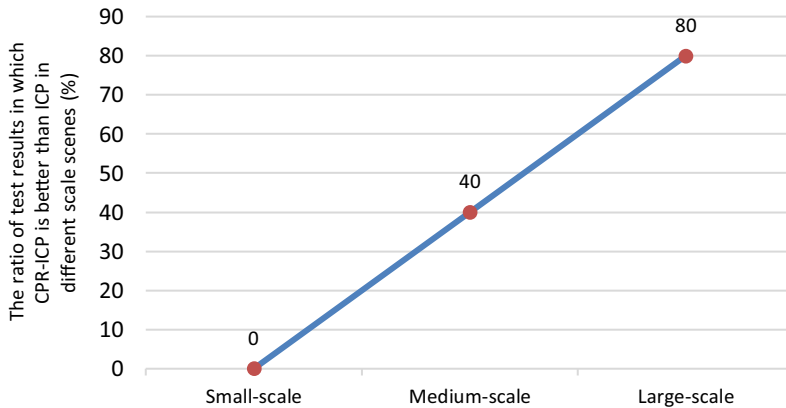


Figure 9. The ratios of test results in which Centre Point Registration-Iterative Closest Point (CPR-ICP) is better than ICP in different scale scenes. The trend shows that the proposed CPR-ICP tends to outperform the traditional ICP in scenes with larger scales.

As can be seen from Fig. 9, in terms of map error correction, the trend shows that the proposed CPR-ICP tends to outperform the traditional ICP in scenes with larger scale. The reason is that the SLAM 3D reconstruction of large-scale scenes tends to have more errors, therefore, a pre-alignment before ICP is necessary.

4. Conclusion

In this paper, the methodology of a comprehensive and objective SLAM benchmark is presented, which can evaluate the SLAM performances unbiasedly. Based on this benchmark, a novel concept of map error correction method by using the correction of trajectory error is presented, and a novel point cloud registration method CPR-ICP is proposed, which firstly aligns two point clouds by their three fitted planes, then applies the ICP operation.

A set of experiments was conducted to validate the feasibility of the proposed point cloud registration method and the map error correction method. The results have shown that the proposed novel registration method outperformed the traditional ICP algorithm, especially in scenes with large scales where a decent initial feature matching is oftentimes absent. The proposed map error correction method has also shown promising performances. After correction, the map error can be reduced by 15%, 12%, and 20% for small-scale, medium-scale, and large-scale scenes, respectively (15.67% on average). This can be considered acceptable for industrial applications, especially for obstacle avoidance in robot navigation, which requires credible obstacle position.

However, for the general applicability of the proposed concept and methodology, they will be further assessed in more SLAM scenarios where different types of sensors and SLAM algorithms are implemented, such as LiDAR-based SLAM systems and other Visual SLAM systems. Moreover, they will also be further assessed against different types of maps or geometries (such as 2D point cloud and 3D mesh) for complex shape objects. The computational cost in these different scenarios will also be considered and benchmarked. In addition, the robustness of the proposed methodology will be further investigated under extreme conditions, such as occlusions, low illumination, etc, where the point cloud is incomplete or has noise. In these cases, the correlations between localization error and mapping error will be further investigated.

Author contributions. Shengshu Liu and Xin Dong conceived and designed the study. Shengshu Liu conducted data gathering. Shengshu Liu performed statistical analyses. Shengshu Liu, Xin Dong, and Erhui Sun wrote the article.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interests. The authors declare no competing interests exist.

Ethical Approval. None.

References

- [1] J. Sturm, N. Engelhard, F. Endres, W. Burgard and D. Cremers, "A Benchmark for the Evaluation of RGB-D SLAM Systems," *In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura-Algarve* (IEEE, 2012) pp. 573–580.
- [2] A. Geiger, P. Lenz, C. Stiller and R. Urtasun, "Vision meets robotics: The kitti dataset," *Int. J. Rob. Res.* **32**(11), 1231–1237 (2013).
- [3] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *Int J Rob Res* **35**(10), 1157–1163 (2016).
- [4] O. Wulf, A. Nuchter, J. Hertzberg and B. Wagner, "Ground Truth Evaluation of Large Urban 6D SLAM," *In: 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA* (IEEE, 2007) pp. 650–657.
- [5] S. Ceriani, G. Fontana, A. Giusti, D. Marzorati, M. Matteucci, D. Migliore, D. Rizzi, D. G. Sorrenti and P. Taddei, "Rawseeds ground truth collection systems for indoor self-localization and mapping," *Auton Robot* **27**(4), 353–371 (2009).
- [6] R. Havangi, "Robust square-root cubature fastSLAM with genetic operators," *Robotica* **39**(4), 665–685 (2021).
- [7] A. Geiger, P. Lenz and R. Urtasun, "Are We Ready for Autonomous Driving? The Kitti Vision Benchmark Suite," *In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI* (IEEE, 2012) pp. 3354–3361.
- [8] O. Wasenmüller, M. Meyer and D. Stricker, "CoRBS: Comprehensive RGB-D Benchmark for SLAM Using Kinect v2," *In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY* (IEEE, 2016) pp. 1–7.
- [9] A. Handa, R. A. Newcombe, A. Angeli and A. J. Davison, "Real-Time Camera Tracking: When is High Frame-Rate Best?," *In: European Conference on Computer Vision*, (Springer, Berlin, Heidelberg, 2012) pp. 225–235.
- [10] J. Funke and T. Pietzsch, "A Framework For Evaluating Visual SLAM," *In: Proceedings of the British Machine Vision Conference (BMVC)* (2009).
- [11] A. Handa, T. Whelan, M. Donald and J. Davison, "A Benchmark for RGB-D Visual Odometry, 3D Reconstruction and SLAM," *In: 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China* (IEEE, 2014) pp. 1524–1531.
- [12] L. Nardi, B. Bodin, M. Zeeshan Zia, J. Mawer, A. Nisbet, P. H. J. Kelly, A. J. Davison, M. Luján, M. F. P. O'Boyle, G. Riley, N. Topham and S. Furber, "Introducing SLAMBench, a Performance and Accuracy Benchmarking Methodology for SLAM," *In: 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA* (IEEE, 2015) pp. 5783–5790.
- [13] R. Klette, A. Koscham, Schlüns K. and V. Rodehorst, "Evaluation of Surface Reconstruction Methods," *In: New Zealand Image and Vision Computing Workshop*, (Citeseer, New Jersey, 1995) pp. 3–12.
- [14] A. Eid and A. Farag, "A Unified Framework for Performance Evaluation of 3-D Reconstruction Techniques," *In: 2004 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC* (IEEE, 2004), p. 33.
- [15] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," *In: 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06), New York, NY* (IEEE, 2006) pp. 519–528.
- [16] A. Bellmann, O. Hellwich, V. Rodehorst and U. Yilmaz, "A Benchmarking Dataset for Performance Evaluation of Automatic Surface Reconstruction Algorithms," *In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN* (IEEE, 2007) pp. 1–8.
- [17] B. Krolla and D. Stricker, "Heterogeneous Dataset Acquisition for a Continuously Expandable Benchmark (CEB)," *In: International Conference on Computer Graphics, Visualization and Computer Vision (WSCG), Pilzen, Czech Republic* (2015).
- [18] Y. Chen and G. Medioni, "Object Modeling by Registration of Multiple Range Images," *In: Proceedings. 1991 IEEE International Conference on Robotics and Automation, Sacramento, CA* (1991) pp. 2724–2729.
- [19] P. J. Besl and N. D. McKay, "Method for Registration of 3-D Shapes," *In: Proceedings SPIE 1611, Sensor Fusion IV: Control Paradigms and Data Structures* (1992).
- [20] X. Li and I. Guskov, "Multi-Scale Features for Approximate Alignment of Point-based Surfaces," *In: Proceedings of the Third Eurographics Symposium on Geometry Processing, Vienna, Austria* (Citeseer, 2005) pp. 217–226.
- [21] D. Aiger, N. J. Mitra and D. Cohen-Or, "4-Points Congruent Sets for Robust Pairwise Surface Registration," *In: ACM SIGGRAPH 2008 Papers, Los Angeles, California* (2008) pp. 1–10.
- [22] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM* **24**(6), 381–395 (1981).
- [23] S. Rusinkiewicz and M. Levoy, "Efficient Variants of the ICP Algorithm," *In: Proceedings Third International Conference on 3-D Digital Imaging and Modeling, Quebec City, QC* (IEEE, 2001) pp. 145–152.
- [24] K.-L. Low. *Linear Least-Squares Optimization for Point-to-Plane Icp Surface Registration*, (University of North Carolina, Chapel Hill, 2004) pp. 1–3.

- [25] A. Censi, "An ICP Variant Using a Point-to-Line Metric," **In: 2008 IEEE International Conference on Robotics and Automation, Pasadena, CA** (IEEE, 2008) pp. 19–25.
- [26] P. Cignoni, C. Rocchini and R. Scopigno, "Metro: Measuring error on simplified surfaces," *Comp Graph For* **17**(2), 167–174 (1998).
- [27] N. Aspert, D. Santa-Cruz and T. Ebrahimi, "Mesh: Measuring Errors Between Surfaces Using the Hausdorff Distance," **In: Proceedings. IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland** (IEEE, 2002) pp. 705–708.
- [28] D. Girardeau-Montaut, "Détection De Changement Sur Des Données géométriques Tridimensionnelles," *Ph.D. Dissertation* (Télécom ParisTech, Paris, France, 2006).
- [29] D. Girardeau-Montaut. *CloudCompare* (EDF R&D Telecom ParisTech, France, 2016) pp. 11.
- [30] S. Meister, P. Kohli, S. Izadi, M. Hämmerle, C. Rother and D. Kondermann, "When Can We Use Kinectfusion for Ground Truth Acquisition," **In: Proceedings Workshop on Color-Depth Camera Fusion in Robotics, Vilamoura, Portugal** (IEEE, 2012).
- [31] K. S. Arun, T. S. Huang and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *IEEE Trans. Pattern Anal* **9**(5), pp. 698–700 (1987).
- [32] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image Vision Comput* **10**(3), 145–155 (1992).
- [33] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int J Comput Vision* **13**(2), 119–152 (1994).
- [34] G. W. Stewart, "On the early history of the singular value decomposition," *SIAM Rev* **35**(4), 551–566 (1993).
- [35] B. Curless and M. Levoy, "A Volumetric Method for Building Complex Models from Range Images," **In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, New York, NY**, (1996) pp. 303–312.
- [36] V. Krishnamurthy and M. Levoy, "Fitting Smooth Surfaces to Dense Polygon Meshes," **In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, New York, NY** (1996) pp. 313–324.
- [37] G. Turk and M. Levoy, "Zippered Polygon Meshes from Range Images," **In: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, New York, NY** (1994) pp. 311–318.
- [38] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.* **33**(5), 1255–1262 (2017).