An Overlap-free Calibration Method for LiDAR-Camera Platforms based on Environmental Perception

Dedong Zhang, *Student Member, IEEE*, Lingfei Ma, *Member, IEEE*, Zheng Gong, Weikai Tan, John Zelek, and Jonathan Li, *IEEE Fellow*

***Abstract*— Indoor environments are challenging for multi-sensor calibrations. Traditional calibration methods use the target structure for camera and LiDAR calibration. Those approaches not only require pre-processed data and offline calculations, but also face challenges in low-light and object-occluded indoor environments. We proposed an automatic calibration method using trajectory constraints on the LiDAR-Camera. The proposed method first obtains the accurate LiDAR trajectory by the LiDAR-SLAM algorithm. At the same time, the problem of visual SLAM trajectory drift in the indoor environment is improved by graphical optimization using the rigid relative position invariance between sensors during displacement. Thus, extrinsic calibration is achieved by using the relative relationship between sensor trajectories. This method has higher robustness than the target-based calibration methods. The experimental results show that our algorithm has higher accuracy than the target-based calibration in the underground environment. The rotation root-mean-square error (RMSE) improves from 6.637° to 0.564°, and the translation RMSE improves from 0.197 m to 0.082 m.**

***Index Terms*—Mobile Laser Scanning, Multi-sensor calibration, Low-cost platform, Point Cloud, Simultaneous localization and mapping.**

I. Introduction

Compared to outdoor environments, indoor environments can be challenging for 2D cameras and 3D LiDAR sensors because they often lack GNSS (Global Navigation Satellite System) signal reception, which is used to accurately determine the position of the sensor [1]. Additionally, indoor environments often have complex structural layouts, such as walls, columns, and furniture, which can occlude the view of the sensors and make it difficult to accurately map the environment. Furthermore, poor lighting and lack of texture can reduce the visibility of the sensor and decrease accuracy when processing the data. As a result, using a single LiDAR sensor to collect data and create 3D reconstructions in indoor environments can be extremely challenging [2]. To overcome these challenges, researchers and engineers have developed various techniques to help improve the accuracy and reliability of indoor mobile mapping systems. Cao *et al*. [3] used a low-channel LiDAR, a camera, and a brushless motor with a crank-link system to replace the traditional LiDAR for 3D object detection. Yin *et al*. [4] used 3D LiDAR with a monocular camera and an inertial measurement unit (IMU) for a SLAM algorithm. Ye *et al.* [5] and Xie *et al.* [6] proposed a target-based calibration method using LiDARs and cameras together. Also, multi-sensor indoor mapping platforms have limitations on the device size and weight, so there is a move towards low cost and fast data collection [7].

In contrast to single-sensor methods, multi-sensor algorithms require the external calibration of the sensors. There are two main categories of existing approaches for this: target-based calibration and target-free calibration. The target-based calibration uses the same objects or structures in the fields of view (FoVs) of different sensors to extract the external parameters between sensors [8]. The target-free calibration obtains the coordinate conversion relationship by satisfying certain constraint relationships with the change in pose after sensor motion [9][10]. However, the calibration quality can be affected by the illumination of indoor environments, occlusions of indoor objects, and even the data accuracy of cameras and LiDAR sensors during the actual calibration process. Many current calibration algorithms are designed based on 360° LiDAR and specific friendly environments (e.g., no occlusion, sufficient light, and clear 3D texture). However, the high cost of 360° LiDAR limits its widespread use in indoor settings. Solid-state LiDARs, which are cheaper and produce denser point clouds within their limited FoVs, are a more practical choice for data collection in indoor environments. Some of the challenges currently encountered in the external calibration of

This research was partially funded by the Natural Science and Engineering Research Council of Canada under Grant No. RGPIN-2022-03741, the National Natural Science Foundation of China under Grant No. 42101451, and also partially funded by the Emerging Interdisciplinary Project of Central University of Finance and Economics from China. (Corresponding authors: Lingfei Ma and Jonathan Li).

Dedong Zhang and John Zelek are with the Department of System Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: dedong.zhang@uwaterloo.ca; jzelek@uwaterloo.ca).

Lingfei Ma is with the School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 102206, China (e-mail: l53ma@cufe.edu.cn)

Zheng Gong is with the School of Computer Engineering, Jimei University, Xiamen 361021, China (e-mail: 202261000067@jmu.edu.cn).

Weikai Tan and Jonathan Li are with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada, and Jonathan Li is also with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: weikai.tan@uwaterloo.ca; junli@uwaterloo.ca).

narrow FoV LiDAR sensors and cameras include:

A. Object Occlusion and Limited View Overlap Between Sensors

Current methods utilize target detection and indoor-specific environmental structures for calibration. Most of these methods are migrated from the methods for automatic calibration of outdoor environments and do not address issues such as high occlusion and lack of 3D structures for sensor FoV overlap in indoor environments. For example, Fu et al. [10] proposed a method to obtain good calibration when both the camera and LiDAR can get clear edge information but faces challenges when the edges are obscured or unclear due to the effects of sensor drift. Shu et al. [11] used an image-focused scheme for indoor localization and stitching of point clouds. The drift problem arises when this scheme is used in occlusions and illumination variations. Therefore, it creates the dilemma that indoor automatic calibration algorithms can only be performed in particular friendly environments.

B. Pre-Processing and Offline Processing Requirements

Existing target-based automatic calibration methods require searching against dense point clouds and images synthesized from multiple data frames. Therefore, as the number of sensors increases, the computation time for auto-calibration keeps increasing accordingly [7][8]. Offline processing also makes it impossible to verify the calibration results in real-time. Consequently, a large amount of repetitive work content is added.

In this study, a trajectory-based indoor auto-calibration algorithm is used for the first time for the narrow FoV LiDAR sensors and cameras. LiDAR motion trajectories are obtained with the simultaneous localization and mapping (SLAM) algorithm [12], and a second independent camera trajectory is

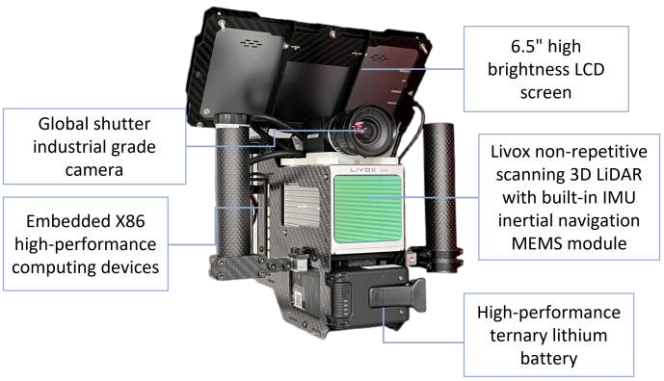


Fig. 1 The multi-sensor mobile platform system used in this study.

obtained simultaneously without considering the view overlap. The objective of the method is to solve the problem of possible errors introduced in the camera trajectory using the graph optimization. Afterward, an extrinsic between the sensors is obtained through the trajectory relationship. This method can be applied to most indoor environments. Moreover, the acquisition of external parameters is made using the real-time computation of the SLAM algorithm, eliminating the need for extensive pre-processing and offline adjustment of the data at an early stage.

II. Data Acquisition and Coordinate System

A. Data Acquisition

This experiment is conducted on a handheld multi-sensor platform. Fig 1 shows this system platform designed for collecting indoor data. The platform consists of a non-repetitive scanning LiDAR (Livox Horizon) with a built-in IMU module and a camera (HIKROBOT MV-CE060-10UC). The non-repetitive scanning and 32-line equivalent LiDAR has a horizontal FoV of 81.7° and a vertical FoV of 25.1°. A wide-angle camera lens (MVL-KF0818M-12MP) with a focal length of 8mm and an aperture of F1.8 is used to ensure that the FoV of the photo and the LiDAR point cloud are consistent. The specified image size collected by the camera is 1520 (width) × 568 (height) pixels.

As for data synchronization, the device uses the network-based high-precision time synchronization protocol PTP (Precision Time Protocol, 1588 V2) [13]. The camera acquires a data frame using the PPS pulse and obtains the time index t. Because the Livox Horizon consists of 24,000 points in one frame, the global time index t is used to find the nearest timestamp in the point clouds. Using the nearest timestamp as a reference, we can get 24,000 points below the moment where the timestamp is located in the point cloud data of that frame. To mitigate the impact of exposure on timestamp synchronization, we took the following measures: selecting the appropriate device and using the proper trigger method. First, we used a camera with alternating exposure hardware, which allows the camera to begin exposing the next frame before the current frame has finished being exposed and read out. Also, this operation can reduce the impact of exposure time on the output time compared to cameras with non-alternating exposure. Second, we used the camera's hardware trigger mode, which involves an MCU generating a pulse signal to trigger the camera and LiDAR. This mode is in contrast to soft triggering, which involves calling an API operation, and is less prone to delayed exposure because it directly accesses the sensor's internal registers for read and write operations.

B. LiDAR-Camera Coordinate System

The point $P\ (x, y, z)$ acquired by the LiDAR is in a local Cartesian coordinate system where the LiDAR sensor is located at $(0, 0, 0)$. Therefore, the point coordinate P for LiDAR can be calculated by

$$P = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} R * \cos(\omega) * \sin(\alpha) \\ R * \cos(\omega) * \cos(\alpha) \\ R * \sin(\alpha) \end{bmatrix} \quad (1)$$

Where ω is azimuth angle, α is polar angle, R is the length of the line segment from the point to the origin. The coordinates of the image data captured by the camera are represented by (U, V), and the 3D point cloud captured by the LiDAR is represented by (X, Y, Z). The goal is to build a transformation matrix M that maps the 3D points (x, y, z) to the 2D points (u, v):

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_u & 0 & u_0 \\ 0 & f_u & v_0 \\ 0 & 0 & 1 \end{pmatrix} (R \quad t) \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$
$$= \begin{pmatrix} f_u & 0 & u_0 \\ 0 & f_u & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{x1} & r_{y1} & r_{z1} & t_x \\ r_{x2} & r_{y2} & r_{z2} & t_y \\ r_{x3} & r_{y3} & r_{z3} & t_z \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = M \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$
$$= \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (2)$$

where the matrices (f_u, f_v, u_0, v_0) are the camera parameters. f_u and f_v are the scale factors, which indicate the effective focal lengths in the horizontal and vertical directions in the XY-axis direction, respectively. u_0 and v_0 are the center points of the image plane, also known as the principal point coordinates. In this equation, R is the 3×3 rotation matrix, and t is the translation vector.

III. Target-Based Manual Calibration

The target-based calibration is used as a comparison for the trajectory-based calibration in this experiment. This target-based method uses calibration boards to collect data in two common indoor scenarios. The steps of data acquisition and processing are as follows.

A. Target Data Acquisition and Pre-Processing

Synchronized frames of point cloud and image data are collected at a range of 5 to 20 m from the mobile mapping platform, ensuring that the target is in the FoV. The images are first resolved for distortion. Accordingly, the calibration board's feature points (i.e., planes, edges, and corner points) in both the image and the point cloud are manually labeled.

B. External Reference Calibration

The LiDAR and the camera are considered rigidly connected structures, and the intrinsic results of the camera have been obtained. Then, the extrinsic calibration solution can be regarded as a Perspective-n-Point (PnP) problem [14]. The point cloud is coarsely registered to the image by the obtained rotation matrix with translation vectors. In this case, the transformation relationship between the camera and the image pixel points is:

$$[u \quad v \quad 1]^T = \frac{1}{Z_c} M [X_c \quad Y_c \quad Z_c]^T \quad (3)$$

Where T is the transpose operation, M is the camera intrinsic. $(X_c \quad Y_c \quad Z_c)$ are the coordinates in the camera coordinate system. The conversion relationship between the coordinates of the camera coordinate system and the LiDAR coordinate system is:

$$[X_c \quad Y_c \quad Z_c \quad 1]^T = \begin{bmatrix} R_t & t_t \\ 0 \quad 0 \quad 0 & 1 \end{bmatrix} [X_l \quad Y_l \quad Z_l \quad 1]^T \quad (4)$$

where $(X_l \quad Y_l \quad Z_l)$ are the position coordinates in the LiDAR coordinate system. R_t denotes the 3×3 rotation matrix of the point cloud in the LiDAR coordinate system to the image pixels in the camera coordinate system. t_t denotes the 3D translation vector.

C. Fine Registration

According to the calibration board plane at different attitudes, a series of linear equations can be obtained to solve the calibration parameters. Such equations can be obtained by Eq. 2 as follows:

$$\begin{cases} u = \dfrac{m_{11}x + m_{12}y + m_{13}z + m_{14}}{m_{31}x + m_{32}y + m_{33}z + m_{34}} \\ v = \dfrac{m_{21}x + m_{22}y + m_{23}z + m_{24}}{m_{31}x + m_{32}y + m_{33}z + m_{34}} \end{cases} \quad (5)$$

The extrinsic calibration is completed by registering the corresponding point cloud to the image. Afterward, the errors are manually calculated.

IV. Trajectory-Based Auto-Calibration

This method utilizes trajectories as a reference for the inter-data optimization to avoid the unnecessary pre-processing step of the target-based calibration. Specifically, the camera-to-LiDAR registration relationship is obtained by exploiting the features that the trajectories of the LiDAR and the camera are similar under a rigid connection structure. Moreover, it is necessary to have relatively stable two trajectories of both the LiDAR and camera. One of the main contributions of this paper also lie in solving the problem of optimizing the two trajectories in the case of large errors.

With the optimization, the trajectory of the camera motion as well as the LiDAR motion trajectory, can be obtained. The camera motion $\mathrm{R}^{\mathrm{i}}_{\mathrm{Camera}}$, $\mathrm{t}^{\mathrm{i}}_{\mathrm{Camera}}$ and LiDAR motion $\mathrm{R}^{\mathrm{i}}_{\mathrm{Lidar}}$, $\mathrm{t}^{\mathrm{i}}_{\mathrm{Lidar}}$ at any given moment satisfy:

$$\mathrm{R}^{\mathrm{i}}_{\mathrm{Camera}}\mathrm{R} = \mathrm{R}\mathrm{R}^{\mathrm{i}}_{\mathrm{Lidar}} \quad (6)$$
$$\mathrm{R}^{\mathrm{i}}_{\mathrm{Camera}}\mathrm{t} = \mathrm{R}\mathrm{t}^{\mathrm{i}}_{\mathrm{Lidar}} + \mathrm{t} \quad (7)$$

The minimization nonlinear optimization cost function is introduced, and the mathematical model can be expressed as follows:

$$\mathrm{R} = \arg\min_{\mathrm{R}} \sum_{\mathrm{i}} \left| \mathrm{R}^{\mathrm{i}}_{\mathrm{Camera}}\mathrm{R} - \mathrm{R}\mathrm{R}^{\mathrm{i}}_{\mathrm{Lidar}} \right| \quad (8)$$

The optimized R and t can be found by the least-squares method. The above serves as a general solution to the calibration problem, and the following section explains specifically if the optimization of trajectories is achieved.

A. Integration of LiDAR Trajectory and Camera Pose

Firstly, the LiDAR-IMU tightly coupled trajectory estimation and map construction based on a LiDAR-SLAM (LIO-SAM) framework is implemented using the built-in IMU and point cloud data [14]. LiDAR obtains a more accurate sub-map M with the assistance of IMU [15]. The LiDAR trajectory obtained by this algorithm is not equivalent to the ground truth. Still, it can be approximated as an accurate trajectory due to the small cumulative error over a short period, which is negligible. Then, the camera image is used separately for the visual SLAM calculation to obtain the camera trajectory information [16]. Hence, the probabilistic roadmap in the visual SLAM is set up

using the feature that the LiDAR trajectory approximates the ground truth. The probabilistic roadmap is not a precise position, but it ensures that the visual SLAM trajectory drift is within a controllable range. It can effectively cope with the long-time drift of visual SLAM, which cannot achieve loop closure detection for a period of time. The pose diagram after adding the probabilistic roadmap of the LiDAR trajectory is shown in Fig. 2.

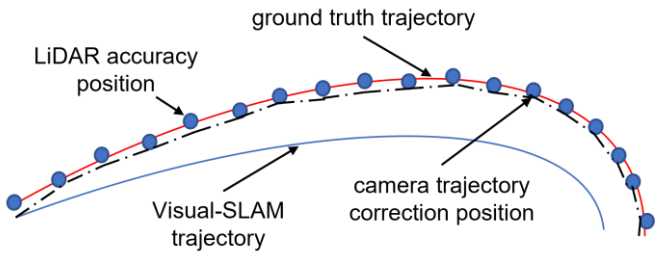


Fig. 2. Visual-SLAM trajectory correction and optimization in accordance with LiDAR trajectory.

This method uses a 6D vector $\xi \in \mathbb{R}^{1\times 6}$ to express the camera poses, and the transformation between any two poses (e.g., positional edges) in the set of poses $\varepsilon = \{\xi_1, \xi_2, \ldots \xi_n\}$ can be expressed by:

$$\Delta\xi_{ij} = \xi_i^{-1} \cdot \xi_j^{-1} = \ln\left[\exp\left(-\widehat{\xi_i}\right)\exp\left(-\widehat{\xi_j}\right)\right]^{\vee} \quad (9)$$

where ^ is the vector antisymmetric matrixing operation and ∨ is the anti-symmetric matrix-to-vector operation. $\Delta\xi_{ij}$ denotes the motion between neighboring poses motion, which the transformation matrix can be also described by:

$$T_{ij} = T_i^{-1}T_j \quad (10)$$

where T_i and T_j denote the corresponding positional transformation matrix of nodes i and j, respectively. The existence of errors e_{ij} makes Eq. (9) or (10) not hold exactly in practice. To solve this problem, the idea of positional optimization is used, which is to construct the least squares error to optimize the variables and minimize the sum of squared errors.

$$\begin{aligned} e_{ij} &= \ln\left(T_{ij}^{-1}T_i^{-1}T_j\right)^{\vee} \\ &= \ln\left\{\exp\left(-\Delta\widehat{\xi_{ij}}\right)\exp\left(-\widehat{\xi_i}\right)\exp\left(-\widehat{\xi_j}\right)\right\}^{\vee} \end{aligned} \quad (11)$$

After the derivation of the error function with respect to the optimization variables, the positional optimization algorithm can be expressed as follows:

$$\xi_{new} = \min_{\xi}\frac{1}{2}\sum_{i,j\in\varepsilon} e_{ij}^{T}\Sigma_{ij}^{-1}e_{ij} \quad (12)$$

Where ξ_{new} is the optimized pose, Σ_{ij}^{-1} is the information matrix. It is the inverse of covariance matrix for the positional estimation variables. The main work of this part is to substitute the trajectory pose data ξ_i of LiDAR at the same time to obtain the optimal pose estimation for visual SLAM.

B. Integration Of Lidar Trajectory and Visual SLAM Poses

After the camera's initial time and position results are aligned with the LiDAR time, the camera position point ξi and its corresponding LiDAR trajectory data observation value g are obtained after N poses. The alignment error is estimated first. Then, the LiDAR local coordinate system position data and the corresponding poses are integrated into this estimate to control the overall error within a specific range. It is assumed that the estimation accuracy of the image alignment position is linearly related to the camera movement speed and conforms to the Gaussian distribution model. The estimation accuracy is inversely proportional to the system movement speed. To be more specific, when the system movement is too violent, the difference between the distance or rotation of the two images becomes more significant over time, resulting in the decrease of the image pose estimation accuracy.

Assume that the system moving speed is $v = S/t$, where t is the image alignment sampling period constant. Here it can also be taken as being the frequency of the camera in the system.S is the position estimation of two points of the total distance. Let Σ_{r0} be the camera position estimation covariance matrix when the velocity is 0 and define $\Sigma_{rv} = v\Sigma_{r0}/k$ as the image position estimation covariance matrix in the case of velocity v, where k is the image position estimation error gain constant. In practice, only observations with error $S' = S + \delta S$ can be obtained, and assuming that δS increases linearly with S, the point position estimation covariance matrix for the case of velocity v is defined as follows:

$$\Sigma_{rv} = \frac{(v+\lambda)\Sigma_{r0}}{k} \quad (13)$$

where λ is the position covariance error estimated at velocity v. k is still the image position estimation error gain constant. According to Eq. (13), the point covariance matrix Σ_{ir} (i =1, 2, ..., n) of the position pose at any moment can be approximated (i is the system position sequence). The accuracy estimation matrix for the integration of arbitrary camera pose and LiDAR trajectory position data can be expressed as follows:

$$A = \sum_{i=1}^{N} {\Sigma_{ir}}^{-1} + {\Sigma_{gj}}^{-1} \quad (14)$$

where Σ_{ir} is the covariance matrix of each pose node, and Σ_{gj} is the covariance matrix of the current LiDAR trajectory position data. A is the data integration accuracy estimation matrix. Then, the current pose node position parameters of the image and the lidar trajectory position after the integration of the pose position parameters are calculated by:

$$\xi_{iT} = A^{-1}{\Sigma_{ir}}^{-1}T_r + A^{-1}{\Sigma_{gj}}^{-1}g_j \quad (15)$$

Where ξ_{iT} is the pose at current position. T_r is the current system position data. This part of LiDAR trajectory and camera pose data integration uses the Kalman estimation for the image trajectory optimization model construction to achieve the purpose of LiDAR trajectory to camera pose optimization.

V. Experimental Results

A. Target-based Experiment

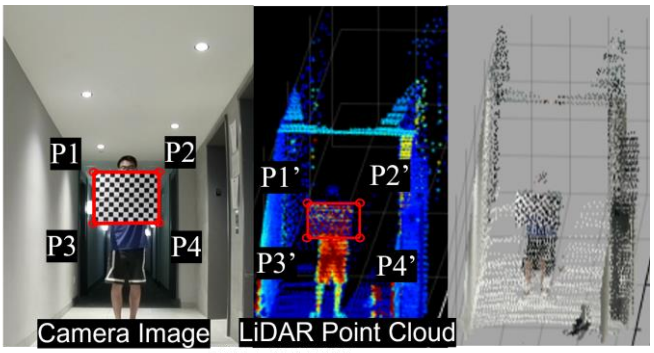
The target-based calibration approach uses a calibration board of size 650 × 500 mm, with each frame of the board measuring 50 × 50 mm. The board was placed in a room at a distance of 3 m from the data collection device. To minimize the impact of manual annotation results on LiDAR point cloud density and image clarity, 10s of manual calibration data was

TABLE I
CALIBRATION ERRORS OF TWO METHODS

Error	Target-based in Scenario A	Ours in Scenario A	Target-based in Scenario B	Ours in Scenario B
x (m)	0.076	0.025	0.248	0.015
y (m)	0.203	0.045	0.083	0.046
z (m)	0.055	0.027	0.116	0.085
Roll (deg)	6.672	0.615	2.323	0.539
Yall (deg)	4.587	0.238	1.665	0.410
Pitch *(deg)*	2.230	0.475	9.183	0.115
RMS_T *(m)*	0.116	0.055	0.197	0.082
RMS_R (deg)	*4.319*	*0.435*	*6.637*	*0.564*

considered a set. Accordingly, a total of 10 sets of experiments were conducted in two scenes (i.e., the office walkway and the underground parking garage.) to ensure the consistency of the results of the comparison experiments. The camera's internal parameters were calculated from the same calibration plate at a distance of 1 m from 20 sets of photos with different angles and positions. The same internal parameters were also used in the experiments of the trajectory-based calibration algorithm. The experimental results were compared uniformly using images and point clouds with calibration plates in both sets of scenes.

In two data sets, the vertices of the target are $P1{\sim}P4$ and $P1'{\sim}P4'$, respectively. The error of the calibration matrix was analyzed according to the obtained calibration results. According to the calibration matrix, the mean distance and the total minimum variance between the corresponding points can be obtained. The data and projection results are shown in Fig.3, and the error analysis results are listed in Table I.



(a) (b) (c)

Fig. 3. The experimental scene of target-based calibration under (a) image, (b) under LiDAR point clouds, and (c) image information registered to the point cloud using the external parameter matrix results.

B. Trajectory-based Experiment

Two data sets were acquired in two scenes using a multi-sensor mobile mapping platform. Each data set contains 530 frames of the point cloud, IMU, and image data. The LIO-SAM algorithm was used to compute the LiDAR data stitching while obtaining the moving trajectory and pose data. Considering that the speed of data collection may affect the accuracy of the data. Before starting the official experiment, we measured three sets of data within the normal walking range of a pedestrian, approximately 1.0m/s, 1.2m/s and 1.5m/s (the average speed was obtained from the measured distance and time spent). The obtained results show that the walking speed is negligible to the results at the current LIDAR and camera frequencies.

Furthermore, according to the ORB-SLAM method, the camera also gets the running trajectory. Since the sensors are rigidly connected, the scaling of the camera trajectory results can be obtained by the length of the two trajectories. The camera trajectory was iteratively optimized using the graph optimization algorithm presented above. Fig 4(a) shows the results obtained in the underground parking lot scenario. Optimizing the camera trajectory keeps the camera and LiDAR trajectories in a stable relative position. The position relationship between the camera and the LiDAR was found using the trajectory. Hence, external rotation and translation matrices were obtained. As seen by the results in Figs 4(b) and 4(c), the optimized two trajectories achieve considerable improvements and the relative position relationships remain stable over the time that the system platform is moved. Through experimentation, it was discovered that in a well-lit office walkway environment, which has rich textures, it is possible to obtain high-quality trajectory data using the SLAM results. The camera trajectories and LiDAR trajectories match almost perfectly after scaling, and the proposed method is not effective for such minor improvements. This is because the scene has enough textures, allowing for accurate tracking and mapping. In contrast, if the environment lacks sufficient textures, it may be more difficult to obtain high-quality trajectory data using the Visual-SLAM. Moreover, it is in this environment that the proposed algorithms gain the greatest advantage.

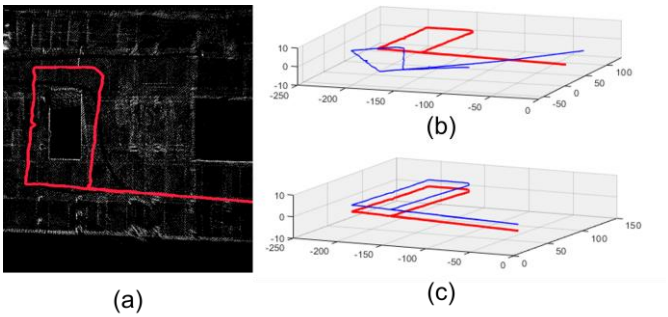


Fig. 4. (a) The LIO-SAM obtains LiDAR trajectory. (b) Optimized scaled camera (blue) trajectory and LiDAR (red) trajectory. (c) The camera (blue) trajectory and the LiDAR (red) trajectory obtained by the graph optimization method.

Fig. 5 illustrates the projection of the results obtained from calibrating a set of data collected from an underground parking lot using the proposed method. The projected data are accurately registered, particularly in regards to the pillars, vehicles, and the 3D texture transformation areas on the ceiling. It is worth mentioning that a part of the wall as a background is blank because the point cloud in Figure 5(b) is obscured by the column. This missing data can be observed clearly in Fig. 5(d). This issue can be resolved by performing multiple scans of the data from various angles. Table I shows that our method has a clear advantage over the target-based calibration methods in indoor environments with low brightness and high occlusion.

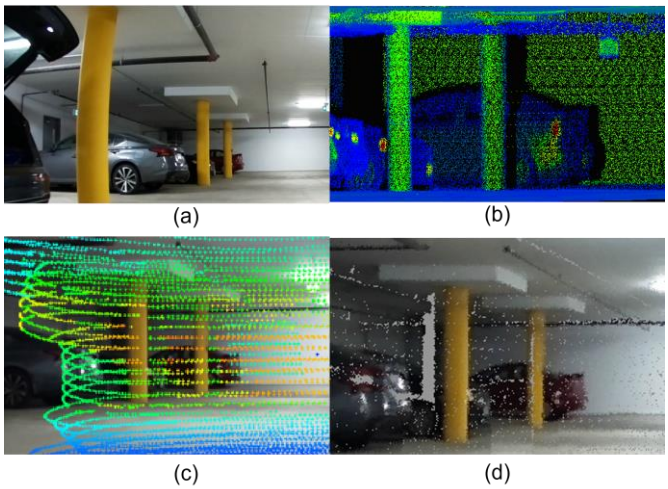


Fig. 5. Validation results after obtaining the calibration matrix by the proposed method. (a) Images in the underground scene. (b) 10 s LiDAR point cloud at the same location, (c) Projection of a certain number of points from the point cloud onto the photograph. (d) Coloring of the point cloud using the image, with no point cloud present in the blank part because of angular occlusion.

VI. Conclusion

In this paper, we have developed a novel automatic calibration method to achieve high accuracy in indoor environments. Since the trajectory generation in the SLAM algorithm does not depend on a single target as the reference, the method of using trajectories as a calibration basis achieves stable acquisition of external parameters in indoor environments with insufficient light and object occlusions. Meanwhile, the graph optimization solves the problem that the visual SLAM trajectory offset cannot determine stable extrinsic. During the data acquisition process, our algorithm obtains the computed trajectory information in real-time and completes the external calibration of the sensor based on the optimized trajectory information. According to the experimental results, the RMSE can be improved from a maximum of 6.637° to 0.564°, and the RMSE of translation can be improved from a maximum of 0.197 m to 0.082 m. These results demonstrate that the calibration accuracy is significantly improved compared to other target-based methods.

References

[1] Z. Kang, J. Yang, Z. Yang, and S. Cheng, "A review of techniques for 3D reconstruction of indoor environments," *ISPRS Int. J. of Geo-Info.*, vol. 9, no. 5, 330, 2020. https://doi.org/10.3390/ijgi9050330.

[2] V. Sequeira, J. G. M. Goncalves, and M. I. Ribeiro, "3D reconstruction of indoor environments," in *Proc. 3rd IEEE Int.Conf. Image Proce.*, 1996, vol. 2, pp. 405–408 vol.2. doi: 10.1109/ICIP.1996.560851.

[3] M. Cao, P. Su, H. Chen, S. Tang, and Y. Liu, "3-D dense rangefinder sensor with a low-cost scanning mechanism," *IEEE Trans. Instrum. Measu.*, 2021, doi: 10.1109/TIM.2020.3016415.

[4] J. Yin, D. Luo, F. Yan, and Y. Zhuang, "A novel lidar-assisted monocular visual SLAM framework for mobile robots in outdoor environments," *IEEE Trans. Instru. Measu.*, 2022, doi: 10.1109/TIM.2022.3190031.

[5] C. Ye, H. Pan, and H. Gao, "Keypoint-based LiDAR-camera online calibration with robust geometric network," *IEEE Trans. Instru. Measu.*, 2022, doi: 10.1109/TIM.2021.3129882.

[6] S. Xie, D. Yang, K. Jiang, and Y. Zhong, "Pixels and 3-D points alignment method for the fusion of camera and LiDAR data," *IEEE Trans. Instru. Measu.*, vol. 68, no. 10, pp. 3661–3676, 2019, doi: 10.1109/TIM.2018.2879705.

[8] X. Liu, C. Yuan, and F. Zhang, "Targetless extrinsic calibration of multiple small FoV LiDARs and cameras using adaptive voxelization," *IEEE Trans. Instru. Measu.*, 2022, doi: 10.1109/TIM.2022.3176889.

[9] X. Liu, C. Yuan, and F. Zhang, "Targetless extrinsic calibration of multiple small FoV LiDARs and cameras using adaptive voxelization," *IEEE Trans. Instrum. Measu.*, 2022, doi: 10.1109/TIM.2022.3176889.

[10] G. Koo, J. Kang, B. Jang, and N. Doh, "Precise Camera-LiDAR Extrinsic Calibration based on a Weighting Strategy using Analytic Plane Covariances," *IEEE Trans. Instrum. Measu.*,, pp. 1–1, 2022, doi: 10.1109/TIM.2022.3202549.

[11] T. Fu, H. Yu, W. Yang, Y. Hu, and S. Scherer, "Targetless extrinsic calibration of stereo, thermal, and laser sensors in structured environments," *IEEE Trans. Instru. Measu.*, 2022, doi: 10.1109/TIM.2022.3204338.

[12] M. Shu, G. Chen, and Z. Zhang, "Efficient image-based indoor localization with MEMS aid on the mobile device," *ISPRS J. Photogram. Remote Sens.*, vol. 185, pp. 85–110, Mar. 2022, doi: 10.1016/j.isprsjprs.2022.01.010.

[13] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "LIO-SAM: Tightly-coupled Lidar inertial odometry via smoothing and mapping." arXiv, 2020. doi: 10.48550/arXiv.2007.00258.

[14] IEEE precision clock synchronization protocol, IEEE Standard 1588, 2008.

[15] E. Kim and S.-Y. Park, "Extrinsic calibration between camera and LiDAR sensors by matching multiple 3D planes," *Sensors*, vol. 20, no. 1, Jan. 2020, doi: 10.3390/s20010052.

[16] J. Zhang and S. Singh, "LOAM: Lidar odometry and mapping in real-time," *Proc. Robotics: Sci. Syst. Conf.*, pp. 109–111, Jan. 2014.

[17] 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, doi: 10.1109/TRO.2015.2463671.

Dedong Zhang received a B.Sc.and M.Sc. degrees in Electrical and Computer Engineering from Baylor University, USA in 2016 and 2018, respectively. He is working toward a Ph.D. degree at the Department of Systems Design Engineering, University of Waterloo, Canada.

He is specialized in optical, IMU, and LiDAR remote sensing, machine learning, artificial intelligence, and simultaneous localization and mapping, His current research interest includes computer vision, localization and mapping, and multi-sensor fusion.

Lingfei Ma (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in geomatics engineering from the University of Waterloo, Waterloo, ON, Canada, in 2015, 2017, and 2020, respectively. He is currently an Assistant Professor of Geospatial Data Science with the Central University of Finance and Economics, Beijing, China.

He has published more than 30 papers in refereed journals and conferences, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, ISPRS Journal of Photogrammetry and Remote Sensing, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and IEEE-CVPRW. His research interests include autonomous driving, mobile laser scanning, intelligent processing of point clouds, 3D scene modeling, and machine learning. He was a recipient of the 2020 National Best Ph.D. Thesis Award granted by the Canadian Remote Sensing Society. He serves as the Guest Editor for International Journal of Applied Earth Observation and Geoinformation.

Zheng Gong received his Ph.D. degree in information and communications engineering with the Fujian Key Laboratory of Sensing and Computing for Smart Cities at the School of Information Science and Engineering, Xiamen University, China, in 2019. From August 2018 to August 2019, he is a visiting PhD student with the Department of Systems Design Engineering, University of Waterloo, Canada. He is currently an Associate Professor in Jimei University.He published a paper in the IEEE Transactions on Instrumentation and Measurement and has co-authored several research papers published in refereed journals and flagship conferences such as TITS, ISPRS, IGARSS and MMT. His current research interests are backpack LiDAR systems, multi-sensor calibration, multi-source data registration, point cloud processing, indoor mapping, 3D computer vision.

Weikai Tan received a B.Sc. degree from Nanjing University, China, in 2014 and B.Sc. and M.Sc. degrees in Geomatics from the University of Waterloo, Canada, in 2014 and 2017, respectively. He is working toward a Ph.D. degree in Geography at the Department of Geography and Environmental Management, University of Waterloo.

He is specialized in optical, SAR, and LiDAR remote sensing, machine learning, artificial intelligence, and environmental monitoring. His current research interest includes urban 3D high-definition mapping and multi-sensor fusion.

John Zelek received the Ph.D. degree in philosophy of electrical engineering from the Centre for Intelligent Machines (CIM), McGill University, Montreal, QC, Canada, in 1996.

He is currently a Professor of the Systems Design Engineering Department, University of Waterloo, Waterloo, ON, Canada and co-Director of the VIP (Vision Image Processing) lab and a member of the Waterloo AI Institute. His research focus is on producing 3D maps from cameras and depth sensors. These 3D maps can be used by robots (vehicles and UAVs) to navigate and localize or by people for various augmented applications. Other research interests include automating infrastructure monitoring as well as automating quality assurance in automated manufacturing. He has published over 300 refereed papers and has been a co-founder of 5 different startup companies from the University of Waterloo and has been an advisor for various other companies.

Jonathan Li (Fellow, IEEE) received the Ph.D. degree in geomatics engineering from the University of Cape Town, South Africa, in 2000. He is currently a professor of geomatics and systems design engineering with the University of Waterloo, Canada. His main research interests include AI-based information extraction from earth observation images, LiDAR point clouds, 3D vision, and GeoAI. He has coauthored more than 530 publications, more than 330 of which were published in refereed journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, ISPRS-JPRS, RSE, and JAG. He has also published papers in flagship conferences in computer vision and AI, including CVPR, AAAI, and IJCAI. He has supervised more than 120 master's and Ph.D. students as well as post-doctoral fellows to completion. He is the Editor-in-Chief of the International Journal of Applied Earth Observation and Geoinformation (JAG) and an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, and Canadian Journal of Remote Sensing. He is a fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada.