A CAMERA-LIDAR CALIBRATION METHOD ASSISTED BY INDOOR SPATIAL STRUCTURE

Chenming Ye [123], Zhizhong Kang [123]*, Xiaoyu Guo [123]

[1] School of Land Science and Technology, China University of Geosciences, Xueyuan Road, Beijing, 100083 CN - yechenming@email.cugb.edu.cn, zzkang@cugb.edu.cn, wuxiaoxie0601@163.com.
[2] Research Center of Lunar and Planetary Remote Sensing Exploration, China University of Geosciences (Beijing) , No. 29 Xueyuan Road, Haidian District, Beijing - yechenming@email.cugb.edu.cn, zzkang@cugb.edu.cn, wuxiaoxie0601@163.com.
[3] Subcenter of International Cooperation and Research on Lunar and Planetary Exploration, Center of Space Exploration, Ministry of Education of The People's Republic of China, No. 29 Xueyuan Road, Haidian District, Beijing 100083 China - yechenming@email.cugb.edu.cn, zzkang@cugb.edu.cn, wuxiaoxie0601@163.com.

KEY WORDS: multi-sensor fusion, BaySAC, calibration, vanishing point detection, indoor 3D.

ABSTRACT:

The calibration of the camera and LiDAR is one of the basis for the construction of the multi-sensor fusion mapping system. Planar features of walls and grounds in indoor environments provides effective constraints for multi-sensor calibration. In this paper, we proposed a new camera-LiDAR calibration method with the constraint of indoor spatial structure. Using the image and point cloud data collected by sensors, visual odometry and LiDAR odometry can be constructed to calculate the transformation between sensors. Based on visual odometry and LiDAR odometry, structural parameters in indoor environment are extracted from images and point cloud to constrain rotation estimation between sensors. In the method proposed in this paper, lines are extracted from the images acquired by the camera and used to estimate and track vanishing points. The direction estimated with vanishing points is used as a global constraint to optimize the rotation parameter estimation of the camera. The fitted planes from the point cloud acquired by the LiDAR are used to compute a set of orthogonal normal vectors corresponding to the ground and wall surfaces, which are used as global constraints to optimize the rotation parameter estimation of the LiDAR. The calibration method proposed in this paper is targetless and only constrained by the indoor spatial structure. The result shows that the proposed indoor spatial structure constraint calibration method can calibrate LiDAR and camera without generating cumulative errors during the rotation estimation process.

1. INTRODUCTION

In the indoor activities of robots, indoor 3D reconstruction (Gu, Zhou et al. 2020), mapping and navigation (Wang, Dai et al. 2020) have always been a research hotspot in the field of robotics. In order to establish a robust indoor surveying and mapping system, it is necessary to accurately calibrate various sensors of the robot and determine their relative spatial positions to obtain more accurate positioning results. Lidar and camera are important indoor data acquisition sensors, and calibration between these two sensors is crucial. On the basis of continuous frame matching, the calibration in this paper adds constraints on the walls and floors of the indoor space, which have orthogonal characteristics in the indoor space. At the same time, these planes can be recognized and parameterized in both images and point clouds. Therefore, this paper introduce the method of identifying and calculating these plane parameters from images and point clouds, and corresponding them to the rotation parameters of the camera and LiDAR, in order to calibrate the rotation relationship between the camera and LiDAR.

Among existing calibration methods, most camera-LiDAR calibration methods require the assistance of a calibration board(Kim and Park 2019), which identifies the calibration chessboard plane from the camera image and LiDAR point cloud, and calibrates the sensor by aligning the plane normal vector. The calibration board can assist visual sensors in precise calibration, but it also limits the calibration environment. Therefore, some studies tried to attempt to calibrate sensors using scene data collected by sensors instead of just using calibration chessboards. Liu, Zhang et al. (2022) proposed a targetless calibration method for LiDAR, IMU, and camera, which uses the IMU coordinate system as a benchmark and calibrates the transformation between sensors by aligning the data collected by the visual odometry,

LiDAR odometry, and IMU. Koide, Oishi et al. (2023) used the Super Glue image matching pipeline to find the 2D-3D correspondence between LiDAR and camera data, and estimated the initial transformation of LiDAR and camera via RANSAC. After giving an initial estimate, the transformation estimation is refined through LiDAR camera calibration based on normalized information distance.

In the images acquired from indoor environment, edge features are the key elements to describe the boundary of walls and floors. By extracting line features from these edge features, the spatial structure of the indoor environment can be briefly represented in the image plane. Hough transform can be used to find lines from a large number of edge features in the image (Bao, Zhang et al. 2005). A more efficient method is to use gradients and gradient directions around pixels, combined with region growth, to obtain lines in the image(Grompone von Gioi, Jakubowicz et al. 2010). Due to the constraints of perspective geometry, there exists a definitive relationship between lines in the image plane, and these lines can be classified by vanishing points algorithms. Vanishing points, which are the intersection points of parallel lines in the image plane, can be used to classify lines based on their pointing directions. Therefore, in indoor three-dimensional space, the position and direction of each vanishing point in the camera can be determined by parallel line segments in different directions, and then the rotation parameters of the camera relative to the indoor walls and grounds can be determined. In the vanishing point detection algorithm, Lu, Yaoy et al. (2017) proposed a 2-line RANSAC algorithm that can quickly select two lines to determine a set of orthogonal vanishing point normal vectors, and determine the set of vanishing points with the highest probability of accuracy based on the hypothesis test results. Li, Kim et al. (2020) combined random sampling and Branch and bound method to obtain the vanishing point estimation while ensuring

---

* Corresponding author

the global optimality of the vanishing point. After correctly estimating the vanishing point, spatial structure information is obtained from the image, and it is also necessary to obtain spatial structure information from the point cloud to correspond with the image.

As an important information describing spatial structure in point clouds, planes play an important role in both localization and scene understanding. Among them, the LiDAR odometry utilizes the LOAM algorithm to calculate the smoothness of each point based on the coordinates between each point and adjacent points, and determines whether the point belongs to a plane through a threshold. Finally, the plane features are described by using the normal vectors of all planar points in the point cloud. However, in real indoor environment, the front and back planes of walls, doors and windows have similar plane parameters. and these planes may be combined incorrectly. To solve this problem, Zhou, Koppel et al. (2021) pointed the normal vector of the plane to the origin of the sensor coordinate system when calculating the plane parameters, and the front and back of the plane would have opposite normal vectors due to different scanning directions. Furthermore, the direction of the plane normal vector can be directly used to distinguish the front and back sides of these double-sided objects. In the indoor environment, the plane fitting efficiency of point clouds is also a key factor of plane extraction and registration. Magri and Fusiello (2018) proposed a variant of the j-linkage algorithm based on min-hashed to quickly identify and obtain plane parameters in point clouds, which has a high efficiency in processing large point clouds. Wang, Peng et al. (2018) used a single line LiDAR to search for indoor planes and inflection points. During the point cloud registering process, the midpoint and corner points of the line features are used as feature points in order to improve the efficiency of the plane registering process.

we proposed a new method of camera-LiDAR calibration with only constrained by indoor spatial structure in this paper. This is a targetless and global environmental feature based calibration method, so it is basically not affected by the accumulated error of the odometry during the calibration process. It is a globally stable and convenient calibration method, and the characteristics of spatial structure constraints can be directly applied to SLAM mapping systems as globally stable features to improve positioning accuracy.

2. METHODOLOGY

We proposed a new camera-LiDAR calibration method constrained by indoor spatial structure in this paper. By utilizing images and point cloud data collected by sensors, visual odometry and LiDAR odometry can be constructed to calculate spatial transformation parameters between sensors. On the basis of LiDAR odometry and visual odometry(Liu, Zhang et al. 2022), the structural parameters of indoor spatial environment are extracted from image and point cloud data respectively to constrain the rotation estimation and time synchronization errors between sensors. The global structural parameters of image are constructed in the direction pointed by the vanishing points, while the global structural parameters of point cloud are constructed in the direction pointed by the normal vectors of the walls and ground.
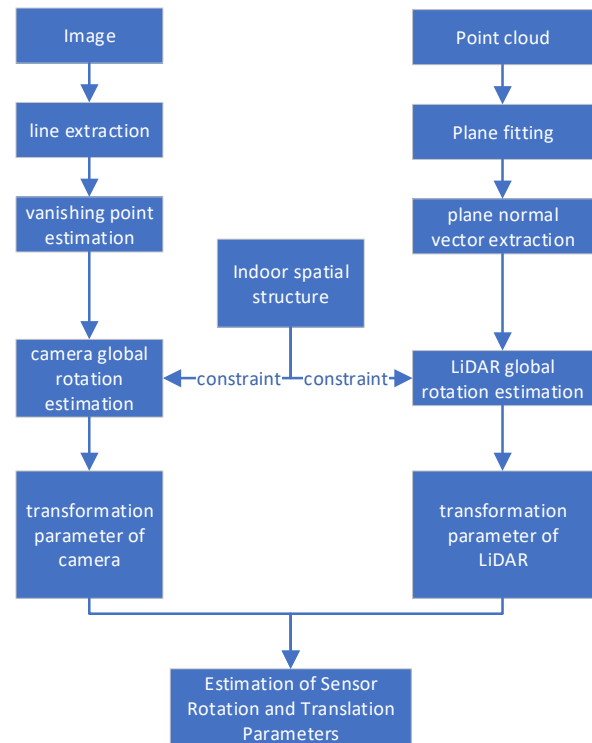

Figure 1. Flowchart of the proposed calibration method.

2.1 Global rotation extracted from images constrained by indoor spatial structure

In the method of estimating camera rotation using indoor acquired image data, feature point matching of consecutive frames is the main approach. These methods estimate the rotation between two frames of images by comparing the relative positions of feature points between consecutive frames. However, this continuous estimation process can result in cumulative errors, which affect the overall accuracy of the image rotation estimation. The vanishing point is the imaging intersection point of a set of parallel lines projected on the camera plane in three-dimensional space. The vanishing point can be estimated by the intersection points generated by multiple sets of lines extracted from the image. Since the estimation of vanishing point is only limited to the direction of the line in the space, the direction of vanishing point estimated from the image can directly represent the relative rotation of the camera and the indoor space. Compared with the feature point rotation estimation method, the vanishing point estimation rotation method can be estimated separately in each frame without cumulative error.

In this paper, LSD algorithm (Grompone von Gioi, Jakubowicz et al. 2010) is used to extract lines from images taken by cameras, and BaySAC algorithm (Kang, Zhang et al. 2014) for estimating model parameters is used as the algorithm to estimate and track vanishing point. When using BaySAC algorithm to estimate the vanishing point, it is necessary to obtain the Prior probability of all lines. In this paper, due to the small displacement of the vanishing point between consecutive images, the Prior probability is given by the position of the vanishing point in the previous frame, and the formula for calculating the Prior probability is

$$P_i = \begin{cases} 1 - \dfrac{D_i}{m} & (D_i < m) \\ 0 & (D_i < m) \end{cases} \tag{1}$$

Wherein, is the Prior probability of line i, is the distance from line i to the assumed vanishing point, and m is the predefined threshold. The adaptive threshold m can be defined by the mean square error of the global point. To accurately estimate the rotation, it is essential to precisely estimate the vanishing points in all three orthogonal directions. Therefore, the prior probability of three vanishing points are estimated for each line. The Prior probability of each line in three directions corresponding to the vanishing point will be compared. One corresponding to the highest probability will be selected as the inliers of the vanishing point hypothesis.

After obtaining the prior probability of all lines, select a group of lines with the highest prior probability to construct the minimum sampling set, and calculate the position of the intersection point of the group of lines from the minimum sampling set, as the vanishing point hypothesis of the cycle. Afterwards, the probabilities of all lines will be updated, and the simplified Bayesian probability update formula is

$$P_t(i \in I) = \begin{cases} \frac{k}{D} P_{t-1}(i \in I) & i \in H_t \\ P_{t-1}(i \in I) & i \notin H_t \end{cases} \quad (2)$$

Among them, I is the set of local points, $P_t(i \in I)$ and $P_{t-1}(i \in I)$ are the probabilities corresponding to the lines in the t and t-1 iterations, respectively. k is the number of lines in this iteration with a distance less than the threshold from the vanishing point, w and D is the total number of lines detected by the LSD algorithm in the current image. After iterating to probability convergence or reaching the number of cycles, calculate the optimal vanishing point parameters from the set of lines with the highest number of local points.

After obtaining the position of the vanishing point, calculate the normal vector of the direction pointed by the vanishing point relative to the camera coordinate system based on the camera focal length and the coordinates of the vanishing point in the image plane. After obtaining the normal vector, it is necessary to calculate the relative rotation between the camera and the direction of the vanishing point in space. In order to correspond to the order of each vanishing point between the previous and subsequent frames, we use the product of the quantity between the vanishing point vectors of the previous frame and the vanishing point vectors of the following frame to calculate the relative relationship between these vanishing point vectors. The calculation formula is

$$\theta_{ij} = VD_{last_i} \cdot VD_{now_j} \quad (3)$$

Where θ_{ij} is the dot product of two vanishing directions, VD_{last_i} is one of the three vanishing directions of the previous frame, p_{now_j} is one of the three vanishing directions of the current frame. A preset threshold θ is used to judge the consistency of the two vector. As shown in Fig.2, for each vanishing point p_i in the previous frame, the calculated maximum θ indicates the association between the vanishing direction of vanishing point p_j in the subsequent frame and the vanishing subsequent of vanishing point p_i in the previous frame. In addition, a third-order Identity matrix is used as the coordinate axis alignment of the vanishing points in the first frame. From the vanishing direction VD_i of the current frame and the current camera coordinate system C_{frame_i}, the rotation of the current camera coordinate $R_{C_{frame\_i}}$ relative to the world coordinate system W of the indoor environment can be obtained, which can be used as the rotation parameter of the camera coordinate in the current frame. According to the vector corresponding to the Vanishing point of

the previous frame and that of the current frame, the Rotation matrix between R_i two frames can be calculated.
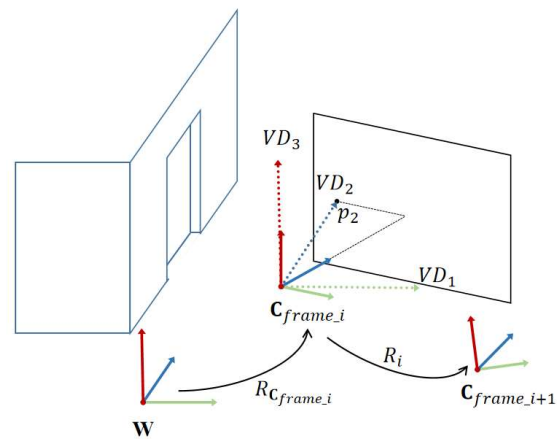

Figure 2. Vanishing point detection and rotation estimation.

In indoor environment, a set of orthogonal Vanishing point estimation results will be used as global constraints to optimize camera rotation parameters estimation. The process involves tracking the vanishing point across consecutive frames, aligning each group of orthogonal directions, and solving for the rotation matrix as the continuous estimation result of the camera's rotation under the global constraints.

2.2 Global rotation extracted from point clouds constrained by indoor spatial structure

In the method of estimating LiDAR rotation using indoor point cloud data, the main focus is on extracting line and plane features from point clouds and matching features between consecutive frames. In this paper, planes are identified and fitted in order to independently extract the relative rotation between the LiDAR and the spatial plane from each frame of the point cloud. Based on the distribution of normal vectors of each plane in the point cloud, the LiDAR sensor coordinate system is counted as a set of rotations orthogonal to the walls and ground planes in the point cloud.

When collecting plane parameters from LiDAR point clouds, this paper uses the region growing method (Poppinga, Vaskevicius et al. 2008) combined with BaySAC to estimate plane parameters. Select adjacent points from random seed points and use BaySAC to fit the plane parameters. In the process of point cloud plane fitting, the parameter estimation and probability update formulas of BaySAC are the same as Eq. (1) and Eq. (2). The parameters of the plane i in a frame of point cloud can be represented as

$$A_i x + B_i y + C_i z + D_i = 0 \quad (4)$$

Based on the fitted plane parameters, continue to search for adjacent points using the region growth method until all adjacent points belonging to the plane are correctly identified. After all points in the point cloud are searched, record the normal vector $N_i = (A_i, B_i, C_i)$ of the plane and the number of points in the plane.

Global rotation can be estimated from the plane normal vectors in the calculated point cloud. Take the normal vector of each plane as the observation value, and use the rotation of the LiDAR relative to walls as the parameters to be estimated. In order to quickly and effectively obtain global rotation, this paper adopts a basic assumption that the plane containing the most points in the recorded plane is the indoor ground or walls, and the normal vector of the plane is recorded as the first directional constraint.

Randomly select a plane parameters from all planes except this plane, and calculate the inner product of the normal vector and the vector constrained by the first direction. When the inner product is less than the threshold, these two directions are considered to be basically orthogonal. To satisfy the orthogonal constraint of rotation, the projection of this direction on the first plane is taken and normalized as the second direction constraint. The third directional constraint is obtained by multiplying the first two vectors.

Based on the existing three directional constraints, verify the orthogonality of all plane normal vectors with these directional constraints, and record the number of points in the plane corresponding to all plane normal vectors orthogonal to this set of directional constraints.

$$\arg\max_{i,j} \sum_{i=1}^{M} \sum_{J=1}^{3} \delta\left(N_i, v_j\right) \tag{5}$$

And

$$\delta\left(N_i, v_j\right) = \begin{cases} num_{p_i} & if \ N_i \cdot v_j > \theta \\ 0 & otherwise \end{cases} \tag{6}$$

Where, N_i is the normal vector of plane i in all M planes. v_j is one of the estimated three directional vectors j. num_{p_i} is the number of points contained in plane i. When the product of the normal vector N_i and the direction vector v_j is higher than the threshold θ, the value of the function $\delta\left(N_i, v_j\right)$ will be recorded as the number of points in the plane. When the number of plane points corresponding to the direction vector is the highest, it is considered that the correct direction estimation has been obtained. With the method in this section, the global rotation of LiDAR relative to indoor space can be accurately estimated. Based on the spatial structural constraints identified in the image and point cloud, the rotation of the camera and LiDAR relative to the indoor space wall was obtained. Therefore, the rotation between the camera and the LiDAR will be calculated based on the relative relationship between these rotations.

2.3 Solving camera-LiDAR rotation parameters based on spatial structure constraints

The global rotation of the indoor spatial structure, determined by both the camera and the LiDAR, is constrained by the walls and ground of the indoor environment. Through their relative rotation, the rotation matrix between the camera and the lidar can be calculated. In this section, construct a coordinate system orthogonal to the walls and ground of the indoor space as the world coordinate system. As shown in Figure 3, the vanishing directions VD_1、VD_2 and VD_3 extracted from the image correspond to the dashed arrows of camera, which are parallel to the axis of the world coordinate system W. Calculate the rotation between the dashed arrow coordinate system and the camera coordinate system to obtain the global rotation of the camera R_{camera} . Similarly, the direction vector $\vec{v_1}$, $\vec{v_2}$ and $\vec{v_3}$ calculated based on the plane normal vector in the point cloud corresponds to the dashed arrow of LiDAR, and these directions are also parallel to the world coordinate system axis, which can also obtain the global rotation of the LiDAR. Based on the two sets of global rotations calculated from multi frame point clouds and images, the rotation parameters between the camera and the LiDAR can be calculated and optimized.
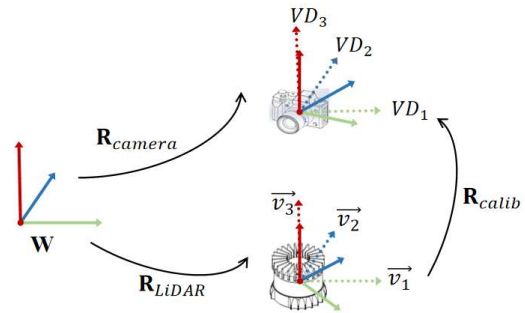

Figure 3. Calculate the rotation of the camera and LiDAR constrained by the indoor spatial structure separately, and use these rotation parameters to estimate the transformation parameters between sensors.

3. EXPERIMENTAL RESULTS

Corresponding to the above methods, the experiment in this paper is divided into three parts. The first part is the vanishing point detection and line classification results of the image. The second part is the point cloud plane fitting results. Finally, the rotation result were aligned in the camera coordinate system to obtain the final calibration result.

A group of point clouds and image data in the indoor environment are collected using the indoor mobile acquisition platform of multi-sensor, and used for the experiment of registration algorithm. According to the algorithm proposed in this paper, experiments were conducted on image and point cloud preprocessing and global rotation extraction, respectively.

Fig.4 shows the results of image vanishing point extraction. The left side is a frame of original image from the collected continuous image data, and the right side is the classification result of the recognized lines and vanishing point detection lines in the image. It can be seen that the lines in Fig.4 are correctly classified and correspond to three vanishing points. According to this vanishing point detection result, we continuously recognize vanishing points from the video, and calculate the camera Rotation matrix according to the camera global rotation estimation method mentioned in Section 2.

Figure 4. The results of vanishing point detection and line classification. left: Raw image data. right: Results of vanishing point detection, with different colored lines corresponding to different vanishing points

For the original point cloud obtained by LiDAR, all the planes were extracted using the plane fitting algorithm proposed in this paper, and the extraction results are shown in Figure 5. Among them, (a) displays the point cloud corresponding to the system time and image, (b) displays the results of plane fitting, and (c) displays the point cloud classification results determined by the direction of the plane normal vector.
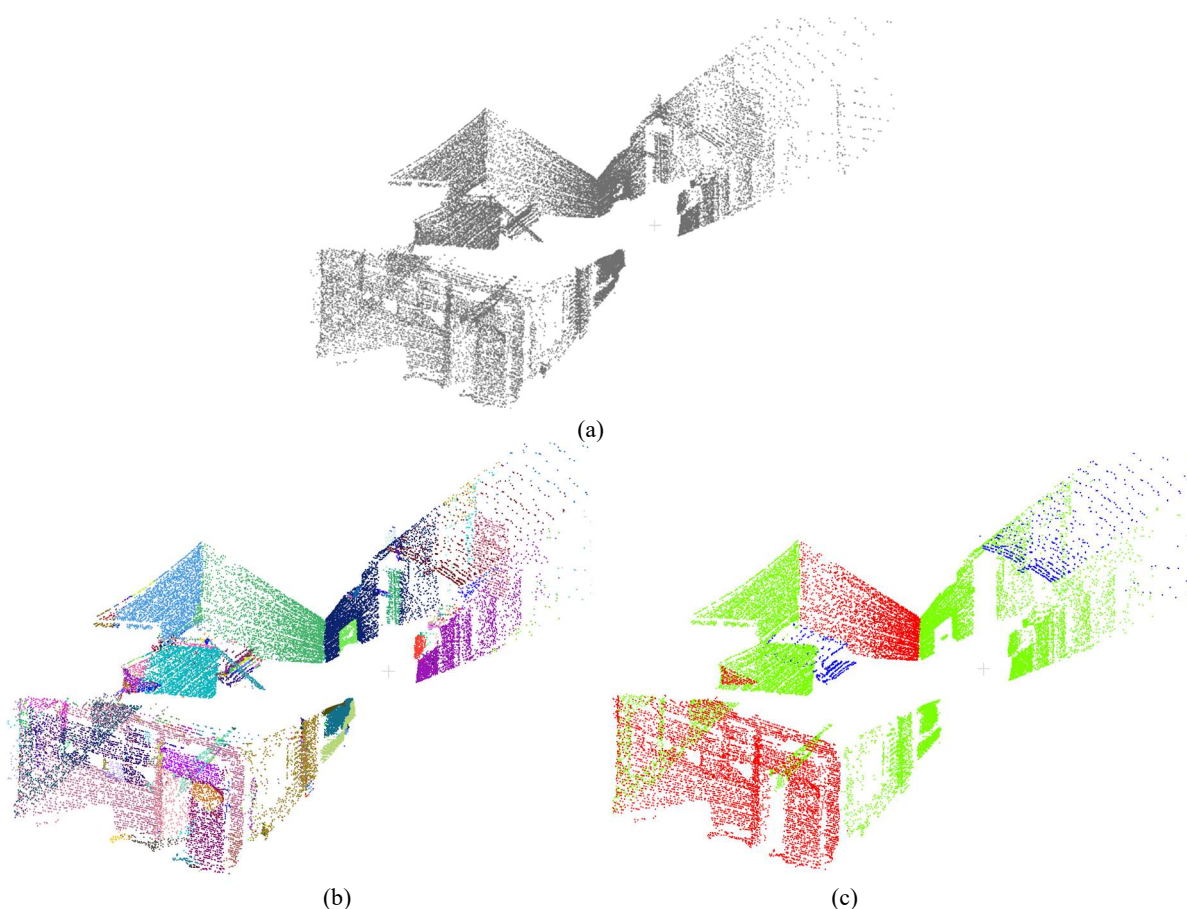

(a)

(b)   (c)

Figure 5. Results of plane fitting and classification.(a) Raw data of indoor point cloud. (b) The results of plane fitting using the method proposed in this paper. (c) Result of plane classification based on normal vector direction.

In Fig.5(c), it can be seen that the walls, floors, and ceilings in the point cloud are all identified correctly. Correspondingly to the global rotation estimation of the aforementioned image, the global rotation of the point cloud is also obtained using the method described in section 2. After calculating the Rotation matrix of the lidar, the yaw angle of each frame is continuously recognized and displayed in Figure 5.

The global rotation calculated by the camera and LiDAR is determined relative to the indoor space environment, and there is a problem of corresponding rotation axes. Therefore, using the rotation parameters of continuous frames of the camera and LiDAR can solve the problem of corresponding axes. Based on the rotation of the camera, calculate the rotation parameters of the LiDAR relative to the camera and align them to the corresponding rotation of the camera, as shown in Fig.6. The optimized continuous rotation of the camera LiDAR can be obtained.

Fig.6 shows the rotation of the camera and LiDAR in each direction in global space. Taking the first chart as an example, the horizontal axis represents the number of frames of the camera and the vertical axis represents the degree of rotation. the blue line corresponds to the rotation of the LiDAR, and the red line corresponds to the rotation of the camera. Among them, point cloud plane recognition has higher robustness and corresponding rotations are smoother and more stable. On the other hand, the result of vanishing point detection may have a small number of incorrectly identified and classified lines, causing the shift of vanishing point, resulting in noise during rotation.

Fig.6 shows the rotation of the camera and LiDAR in each direction in global space. Taking the first chart as an example, the horizontal axis represents the number of frames of the camera and the vertical axis represents the degree of rotation. the blue line corresponds to the rotation of the LiDAR, and the red line corresponds to the rotation of the camera. Among them, point cloud plane recognition has higher robustness and corresponding rotations are smoother and more stable. On the other hand, the result of vanishing point detection may have a small number of incorrectly identified and classified lines, causing the shift of vanishing point, resulting in noise during rotation.
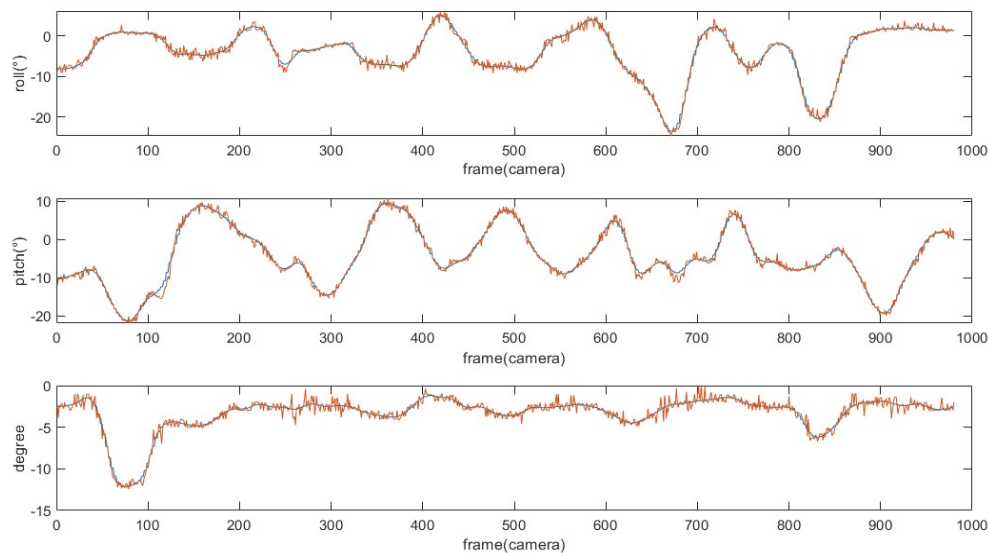
Figure 6. Global rotation results of camera and LiDAR.

4. CONCLUSION

We proposed a camera-LiDAR calibration method for indoor spatial structure constraints in this paper. The global rotation of the camera relative to the world coordinate system is determined using the detected vanishing points in the image, and the global rotation of the LiDAR relative to the world coordinate system is determined using the direction of the detected plane and plane normal vector in the point cloud. Both sets of rotations are based on the world coordinate system constrained by indoor space, so after aligning the various directions of the coordinate axis, the rotation transformation between sensors can be quickly determined. The experimental results show that the vanishing point detection and plane fitting algorithms used in this paper have high robustness and can smoothly estimate the global rotation of two sets of sensors. The final calibration results show that the indoor spatial structure can assist and effectively achieve targetless calibration of sensors.

REFERENCES

Bao, P., et al. (2005). "Canny edge detection enhancement by scale multiplication." Ieee Transactions on Pattern Analysis and Machine Intelligence **27**(9): 1485-U1481.DOI: 10.1109/tpami.2005.173

Grompone von Gioi, R., et al. (2010). "LSD: a fast line segment detector with a false detection control." IEEE Trans Pattern Anal Mach Intell **32**(4): 722-732.DOI: 10.1109/TPAMI.2008.300

Gu, P., et al. (2020). "A 3D reconstruction method using multisensor fusion in large-scale indoor scenes." Complexity **2020**: 1-14

Kang, Z., et al. (2014). "An Optimized BaySAC Algorithm for Efficient Fitting of Primitives in Point Clouds." Ieee Geoscience and Remote Sensing Letters **11**(6): 1096-1100.DOI: 10.1109/lgrs.2013.2286856

Kim, E.-S. and S.-Y. Park (2019). Extrinsic calibration of a camera-LIDAR multi sensor system using a planar chessboard. 2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN), IEEE

Koide, K., et al. (2023). "General, single-shot, target-less, and automatic LiDAR-camera extrinsic calibration toolbox." arXiv preprint arXiv:2302.05094

Li, H., et al. (2020). Globally optimal and efficient vanishing point estimation in atlanta world. Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16, Springer.DOI: 10.1007/978-3-030-58542-6_10

Liu, H., et al. (2022). Spatiotemporal LiDAR-IMU-Camera Calibration: A Targetless and IMU-Centric Approach Based on Continuous-time Batch Optimization. 2022 34th Chinese Control and Decision Conference (CCDC), IEEE

Lu, X., et al. (2017). 2-Line Exhaustive Searching for Real-Time Vanishing Point Estimation in Manhattan World. 2017 IEEE Winter Conference on Applications of Computer Vision (WACV): 345-353.DOI: 10.1109/wacv.2017.45

Magri, L. and A. Fusiello (2018). Reconstruction of interior walls from point cloud data with min-hashed J-Linkage. 2018 International Conference on 3D Vision (3DV), IEEE

Poppinga, J., et al. (2008). Fast plane detection and polygonalization in noisy 3D range images. 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE

Wang, C., et al. (2020). "Isprs Benchmark on Multisensory Indoor Mapping and Positioning." ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences **V-5-2020**: 117-123.DOI: 10.5194/isprs-annals-V-5-2020-117-2020

Wang, Y.-T., et al. (2018). "A single LiDAR-based feature fusion indoor localization algorithm." Sensors **18**(4): 1294

Zhou, L., et al. (2021). "LiDAR SLAM with plane adjustment for indoor environment." Ieee Robotics and Automation Letters **6**(4): 7073-7080