LiDAR-Camera Calibration Under Arbitrary Configurations: Observability and Methods

Bo Fu, Yue Wang, *Member, IEEE*, Xiaqing Ding, Yanmei Jiao, Li Tang, and Rong Xiong, *Member, IEEE*

***Abstract*— LiDAR-camera calibration is a precondition for many multi-sensor systems that fuse data from LiDAR and camera. However, the constraint from common field of view and the requirement for time synchronization make the calibration a challenging problem. In this paper, we propose a novel LiDAR-camera calibration method aiming to eliminate these two constraints. Specifically, we capture a scan of 3-D LiDAR when both the environment and the sensors are stationary, then move the camera to reconstruct the 3-D environment using the sequentially obtained images. Finally, we align 3-D visual points to the laser scan based on a tightly couple graph optimization method to calculate the extrinsic parameter between LiDAR and camera. Under this design, the configuration of these two sensors is free from the common field-of-view constraint due to the extended view from the moving camera. In addition, we also eliminate the requirement for time synchronization as we only use the single scan of laser data when the sensors are stationary. We theoretically derive the conditions of minimal observability for our method and prove that the accuracy of calibration is improved by collecting more observations from multiple scattered calibration targets. In addition, the proposed method is beneficial to not only plane measurement error-based calibration targets, such as chessboards, but also other point measurement error-based calibration targets, such as boxes and polygonal boards. We validate our method on both simulation and real-world data sets. Experiments show that our method achieves higher accuracy than other comparable methods, which is in accordance with our theoretical analysis.**

***Index Terms*— Arbitrary configuration, eliminating time variable, LiDAR and camera calibration, observability.**

NOMENCLATURE

$\{\mathcal{C}\}$	The reference camera coordinate system.
$\{\mathcal{L}\}$	The reference LiDAR coordinate system.
${}^{\mathcal{C}}p_c$	A 3-D visual feature point (triangulated from image pixels) on the chessboard in $\{\mathcal{C}\}$, ${}^{\mathcal{C}}p_c \in \mathbb{R}^3$.
${}^{\mathcal{L}}p_c$	A 3-D visual feature point on the chessboard in $\{\mathcal{L}\}$, ${}^{\mathcal{L}}p_c \in \mathbb{R}^3$.
${}^{\mathcal{L}}p_f$	A 3-D visual feature point in $\{\mathcal{L}\}$, ${}^{\mathcal{L}}p_f \in \mathbb{R}^3$.
${}^{\mathcal{C}}p_f$	A 3-D visual feature point in $\{\mathcal{C}\}$, ${}^{\mathcal{C}}p_f \in \mathbb{R}^3$.
p_ℓ	A 3-D laser point in $\{\mathcal{L}\}$, $p_\ell \in \mathbb{R}^3$.
n_ℓ	The normal vector of p_ℓ, $n_\ell \in \mathbb{R}^3$.
${}^{\mathcal{L}}_{\mathcal{C}_k}x$	The pose of the camera at time t_k in $\{\mathcal{L}\}$, ${}^{\mathcal{L}}_{\mathcal{C}_k}x \in SE(3)$.
${}^{\mathcal{L}}_{\mathcal{C}}x$	The extrinsic parameter of $\{\mathcal{C}\}$ with respect to $\{\mathcal{L}\}$, ${}^{\mathcal{L}}_{\mathcal{C}}x \in SE(3)$.
${}^{\mathcal{C}}_{\mathcal{C}_k}x$	The pose of the camera at time t_k in $\{\mathcal{C}\}$, ${}^{\mathcal{C}}_{\mathcal{C}_k}x \in SE(3)$.

Manuscript received March 8, 2019; revised July 7, 2019; accepted July 18, 2019. Date of publication July 29, 2019; date of current version May 12, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1300400, and in part by the National Nature Science Foundation of China under Grant U1609210. The Associate Editor coordinating the review process was Pekka Keränen. *(Corresponding authors: Yue Wang; Rong Xiong.)*

The authors are with the State Key Laboratory of Industrial Control and Technology, Zhejiang University, Hangzhou 310058, China (e-mail: wangyue@iipc.zju.edu.cn; rxiong@zju.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2019.2931526

I. INTRODUCTION

A PERCEPTION system that employs only one sensor will not be robust. For example, LiDAR-based odometry [1] will fail when working in a long corridor, and the camera-based algorithm [2]–[4] cannot be applied to a textureless scene [5]. Fusing the visual and laser data can eliminate the outliers from the algorithm and solve various limitations for the algorithms imposed by the single sensor. For example, the fusion of the range sensor and the camera can improve the accuracy of object detection [6]. What is more, heterogeneous localization methods, such as visual localization on a laser map [7], can enable low-cost and long-term localization. The precondition of all the above algorithms is the calibration of different sensors, and to that end, we focus on the extrinsic calibration of the LiDAR and camera in this paper.

Numerous efforts have been carried out to perform LiDAR–camera extrinsic calibration [8]–[10]. The current calibration approaches can be classified into two groups [11]: one is appearance-based and the other is motion-based. The appearance-based methods can calculate the extrinsic parameter by directly matching 2-D images with 3-D points on the laser point cloud. In the motion-based methods, the motion of the camera is estimated from images, while the motion of the LiDAR is estimated from the laser points, and then calibration is performed by aligning the two trajectories.

First, we will consider *the appearance-based methods*. Methods such as [12] and [13] use targets that can be detected on both 2-D images and 3-D laser point clouds. Geiger *et al*. [14] presented a method to automatically calibrate the extrinsic parameter with one shot of multiple chessboards, which recovered the 3-D structure from the detected image corners. After that, the approach used the constraint that the chessboard planes should coincide with the detected LiDAR planes to perform calibration. The method was applied in the KITTI data set [15] to calibrate the extrinsic parameter between the cameras and the LiDAR sensor. Unlike the approaches above, Wang *et al*. [16] utilized the reflectance

intensity to estimate the corners of the chessboard from the 3-D laser point cloud. If the corners of the 3-D laser point cloud are identified, the extrinsic calibration is converted to a 3-D–2-D matching problem. However, these algorithms always require the sensors sharing a common field of view, which some application scenarios cannot satisfy. Even in the application scenario where the condition is met, the requirement of the common field of view constrains the scale of the scene and limits the number of targets that can be detected, thus affecting the accuracy of the calibration, which prevents the utilization of pinhole cameras from the LiDAR–camera system. In some methods, panoramic or wide-angle cameras are used to solve this problem [17]. Some methods lead to the tedious focus process in order to expand the field of view such as [10].

On the other hand, *the motion-based methods* [18], [19] perform calibration by aligning the estimated motion trajectories. Early motion-based calibration methods were based on hand-eye calibration [20]. In [11], the initial extrinsic parameter is calculated from scale-free camera motion and LiDAR motion. Next, the camera motion is recalculated using the initial extrinsic parameter and the point cloud from the LiDAR, and then, the extrinsic parameter is calculated again using the motion, and this is repeated until the estimate converges. However, the motion-based method is a loosely coupled calibration method that cannot lead to high calibration accuracy. In addition, the motion-based calibration method needs to complete time synchronization before performing calibration, which is not easy in some cases. In scenarios where time synchronization is not completed, an additional variable (i.e., time offset) should be introduced. In [18], they propose a method to obtain the motion of a sensor in 2-D–3-D calibration and estimate the extrinsic parameter and the time offset between the sensors. Obviously, introducing new variables will reduce the calibration accuracy.

In this paper, we propose a hybrid calibration method, which combines the advantages of appearance-based calibration and motion-based calibration. The demonstration of the proposed method is shown in Fig. 1. In our method, a number of chessboards in various poses are placed around the sensors, and one frame laser scan of the chessboards is obtained under stationary. Then, the sensors are moved around to obtain images of each chessboard to reconstruct the 3-D visual point cloud. Note that this differs from previous approaches [8], [9], which require multiple images and the LiDAR data of a single chessboard presented at different poses as inputs; the hidden limitation of these methods is that a common field of view between sensors is needed.

Our method expands the camera's field of view by moving the sensor, so even though there is no common view at the starting position, the LiDAR and the "expanded camera" can also have overlap in their measurement ranges, which removes the configuration limitation for a common field of view. Moreover, the extended field of view obtained can remove the constraints of the observed scale of the scene and increase the number of chessboards that can be detected, which can lead to an increase in accuracy. Additionally, we only use the first frame of laser as a map. In this way, we eliminate the time variable (i.e., time offset) from the spatial extrinsic parameter estimating, which means that we do not need to solve time variable (i.e., time offset) and spatial variable (i.e., ${}^{\mathcal{L}}_{\mathcal{C}}x$) together. So our method is applicable to the cases lacking time synchronization and will not introduce additional variables. As part of our contribution, we also examine the observability properties of our system and present the minimal necessary conditions for estimating the LiDAR–camera extrinsic parameter. Furthermore, we derive the influence of the angle and distance between calibration targets on the calibration accuracy, which proves that sharing a larger field of view between sensors is beneficial for better calibration accuracy. The relevant theory provides a guideline for designing high-accuracy calibration procedures.

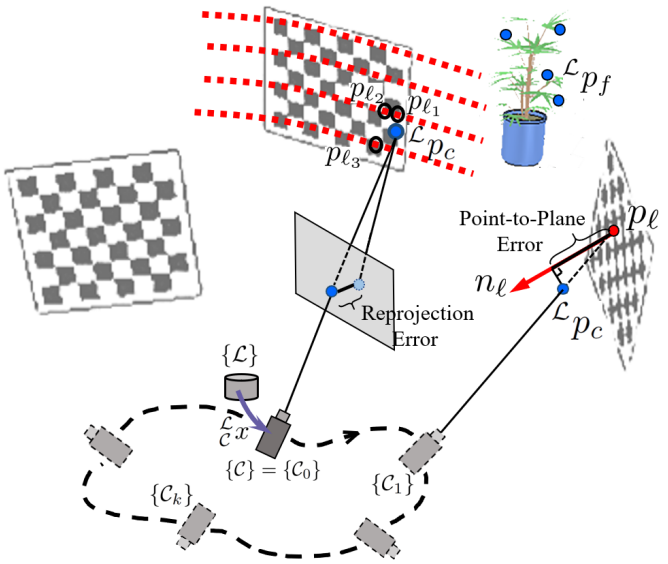


Fig. 1. Mathematical model of the calibration system.

This paper is structured as follows. Section II gives a detailed description of the proposed method. Then, we prove the theory in Sections III and IV and evaluate it in Section V. We present our conclusions in Section VI.

II. Calibration Method

The extrinsic parameter between the camera and the LiDAR is the relative pose ${}^{\mathcal{L}}_{\mathcal{C}}x$ of the camera coordinates $\{\mathcal{C}\}$ with respect to the LiDAR coordinates $\{\mathcal{L}\}$. Thus, a visual 3-D point ${}^{\mathcal{C}}p_c$ represented in $\{\mathcal{C}\}$ can be transformed into $\{\mathcal{L}\}$ via ${}^{\mathcal{L}}p_c = {}^{\mathcal{L}}_{\mathcal{C}}x \otimes {}^{\mathcal{C}}p_c$, where $\otimes$ represents the multiplication between $SE(3)$ and the homogeneous coordinates of $\mathbb{R}^3$, and the result is represented in $\mathbb{R}^3$. Assumed that ${}^{\mathcal{C}}p_c$ is on a plane, then we have the model of the LiDAR vision system as

$$n_\ell^{\mathrm{T}}({}^{\mathcal{L}}_{\mathcal{C}}x \otimes {}^{\mathcal{C}}p_c - p_\ell) = 0 \tag{1}$$

where p_ℓ and n_ℓ is a laser point and its normal on the same plane. During calibration, ${}^{\mathcal{L}}_{\mathcal{C}}x$ is the unknown. The basic idea for estimation is to utilize the model (1) for constraints' formulation. Following this idea, there are three problems to solve: detection of the plane from vision data, detection of

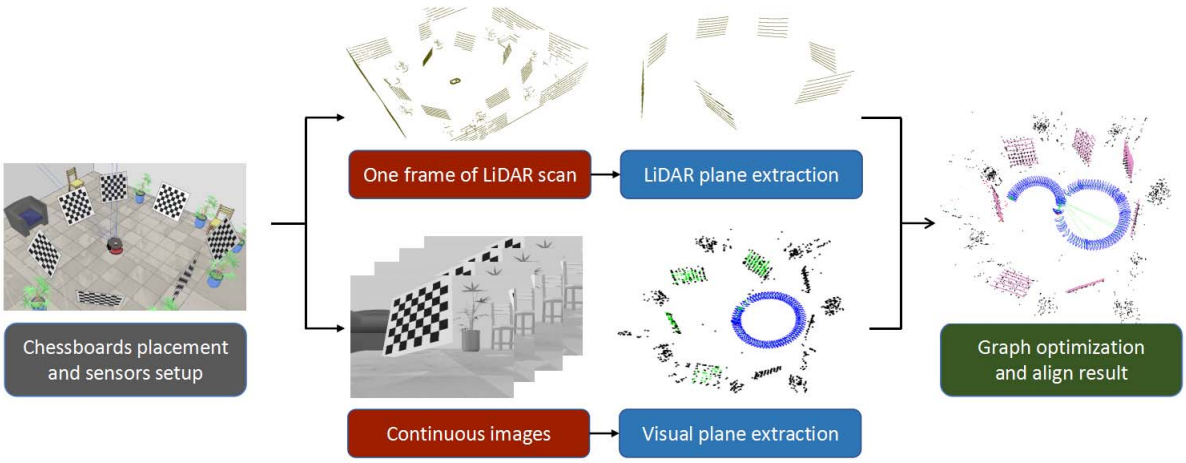


Fig. 2. Overview (using a simulation experiment as an example). First, we extract the chessboard corner points from the camera images and then reconstruct the 3-D visual point clouds. Second, we filter out the chessboard plane in the obtained laser data. Third, we optimize the point-to-plane error to estimate the extrinsic parameter.

the corresponding plane from LiDAR data, and design of the estimator for ${}^{\mathcal{L}}_{\mathcal{C}}x$.

We propose a method consisting of three modules. The overview is shown in Fig. 2. To simplify the detection of plane in both models of data, we place several chessboards in the scene [21], [22]. The first module utilizes the image stream acquired by the moving camera to map the surrounding environment using simultaneous localization and mapping (SLAM) [23] and extracts the boards using the detector generally employed in camera intrinsic parameters' calibration. The second module is to extract these boards in the first LiDAR scan using plane model-based random sample consensus (RANSAC) [24] with the aid of mechanic extrinsic parameter. After the two steps, we have the laser points and visual points on the planes required in model (1). The last module is a regularized optimizer to estimate ${}^{\mathcal{L}}_{\mathcal{C}}x$ by including the constraints of natural features, yielding a more accurate result.

Note that for visual detection, a sequence of image is used, while for LiDAR detection, only the first scan is used. The main reason for this processing step is to eliminate the time offset between the camera and LiDAR. Specifically, as shown in Fig. 1, in the sensor system, the extrinsic parameter ${}^{\mathcal{L}}_{\mathcal{C}}x$ is actually the first camera pose in $\{\mathcal{L}\}$ (i.e., ${}^{\mathcal{L}}_{\mathcal{C}_0}x$). Denote the pose of LiDAR in $\{\mathcal{L}\}$ at time t_0 as ${}^{\mathcal{L}}_{\mathcal{L}_0}x$, and ${}^{\mathcal{L}}_{\mathcal{L}_0}x = I_4$ (since $\{\mathcal{L}\} = \{\mathcal{L}_0\}$), where I_4 is a 4×4 identity matrix. In the ideal case, we have

$$ {}^{\mathcal{L}}_{\mathcal{L}_0}x = {}^{\mathcal{L}}_{\mathcal{C}_0}x\,{}^{\mathcal{L}}_{\mathcal{C}}x^{-1}. \tag{2} $$

When there is a time synchronization error, the equation becomes

$$ {}^{\mathcal{L}}_{\mathcal{L}_\delta}x = {}^{\mathcal{L}}_{\mathcal{C}_0}x\,{}^{\mathcal{L}}_{\mathcal{C}}x^{-1} \tag{3} $$

where ${}^{\mathcal{L}}_{\mathcal{L}_\delta}x$ is the LiDAR pose in $\{\mathcal{L}\}$ at time $t_0 + \delta t$, and δt is a small time offset. When motion-based method is applied, the displacement from t_0 to $t_0 + \delta t$ is ignored, causing the calibration error. However, as we only use the first LiDAR scan and the two sensors can be stationary when the first LiDAR scan is acquired, we have

$$ {}^{\mathcal{L}}_{\mathcal{L}_0}x = {}^{\mathcal{L}}_{\mathcal{L}_\delta}x = {}^{\mathcal{L}}_{\mathcal{C}_0}x\,{}^{\mathcal{L}}_{\mathcal{C}}x^{-1}. \tag{4} $$

As a result, the proposed calibration method is applicable for cases without time synchronization and no additional variables for estimation are introduced. In sequel, we present the three modules in detail.

A. Visual Plane Extraction

The aim of this module is to build the 3-D points on the chessboard from continuous camera images. We run ORB-SLAM [25] first to estimate the trajectory of the camera $\{{}^{\mathcal{C}}_{\mathcal{C}_k}x\}$, as well as the 3-D visual feature points $\{{}^{\mathcal{C}}\bar{p}_f\}$. In the case of a monocular camera, we use the scale of the chessboard to compute the metric, while for the stereo camera, the scale is known with the aid of baseline. Due to the existence of observation error, the quality of the reconstructed 3-D points varies, which is expressed by the depth uncertainty in ORB-SLAM [25]. To reduce error, we only pick the high-quality 3-D visual points for further processing. The quality of a point is measured by

$$ \text{Score}_c = N_a - \gamma\, N_b \tag{5} $$

where N_a is the number of frames that observe the point (higher is better) and N_b is the depth uncertainty calculated in ORB-SLAM (lower is better). γ is a tradeoff parameter, empirically choosing 0.1 in our experiments.

We remove points with scores less than a certain threshold (empirically choosing 0.8 in our experiments), i.e., poor reconstruction quality. After this step, we obtain the filtered visual points $\{{}^{\mathcal{C}}p_f\} \subseteq \{{}^{\mathcal{C}}\bar{p}_f\}$. Then, to distinguish the 3-D points of the chessboard in $\{{}^{\mathcal{C}}p_f\}$, we extract the corners of the chessboard [26] in each image. If any of the reconstructed points $\{{}^{\mathcal{C}}p_f\}$ has image projections in accordance to the chessboard corners, we assign the points as chessboard points $\{{}^{\mathcal{C}}p_c\} \subseteq \{{}^{\mathcal{C}}p_f\}$. Finally, we perform global bundle

adjustment [27] to further improve the accuracy of the points $\{^{\mathcal{C}}p_f\}$ and poses $\{^{\mathcal{C}}_{\mathcal{C}_k}x\}$.

This step provide the points on planes $\{^{\mathcal{C}}p_c\}$ required in (1). Even though building 3-D points of chessboard corners using only one static image is possible, the crucial advantage of utilizing a moving camera is the expansion of the field of view. Therefore, multiple chessboards can be observed and represented in $\{\mathcal{C}\}$. Besides, multiple chessboards can be placed scattered around the sensor system, which is able to provide better constraints, leading to superior calibration performance, as shown in theoretic analysis and experiments' parts.

B. LiDAR Plane Extraction

The second module set is to extract the planes in the first LiDAR scan and find their correspondence to visual points on chessboards $\{^{\mathcal{C}}p_c\}$. We begin the processing by computing the normals for each laser point. Then, region growing [28] is applied to cluster the laser points with similar properties. Specifically, the criteria of the region growing are whether the neighboring point has similar normal. After this step, we have several hypotheses, i.e., clusters, where each one potentially contains points on the chessboard. To further filter the hypotheses, we apply plane-based RANSAC [24] to pick hypothesis with sufficient inliers (empirically choosing 800 in our experiments). As a result, we have a set of points $\{p_\ell\}$ and corresponding normals $\{n_\ell\}$ acquired from the chessboards.

Before the calculation of the extrinsic parameter, we have to find the data association between $\{p_\ell\}$ and $\{^{\mathcal{C}}p_c\}$, or the model (1) may become incorrect constraints, leading to failure in extrinsic parameter estimation. To obtain the data association, the mechanical extrinsic parameter is used as a rough value of $^{\mathcal{L}}_{\mathcal{C}}x$ to transform $^{\mathcal{C}}p_c$ to $^{\mathcal{L}}p_c$. Then, we build a K-dimensional tree (KD-tree) structure [29] for laser points, so that for each visual point, the nearest three laser points can be searched efficiently, denoted as a 4-tuple $(^{\mathcal{C}}p_c, p_{\ell_1}, p_{\ell_2}, p_{\ell_3})$. Generally, such method may cause incorrect point-to-point data association. However, both $\{p_\ell\}$ and $\{^{\mathcal{C}}p_c\}$ are only the points on the chessboards, thus very sparse. Therefore, it is almost impossible that a point on one chessboard is associated with one on another chessboard. After the initial data association, we evaluate the model fitness of a 4-tuple

$$\text{Score}_\ell = \sum_{i=1}^{3} |n_{\ell_i}^{\mathrm{T}}(^{\mathcal{L}}p_c - p_{\ell_i})| \tag{6}$$

where n_{ℓ_1}, n_{ℓ_2}, and n_{ℓ_3} are the normal vectors corresponding to p_{ℓ_1}, p_{ℓ_2}, and p_{ℓ_3}. We remove the 4-tuple with scores larger than a certain threshold (empirically choosing 0.3 in our experiments), i.e., poor model fitness. The remaining tuples specify all the parameters in model (1), formulating the constraints for extrinsic parameter calibration.

Note that the parameters used in plane extraction from visual and LiDAR data can be sensitive when the chessboards are far from the sensor system. Thanks to the moving camera, we can place the chessboards near the system without concerning limited space for multiple chessboards as shown in the experiments.

C. Optimization for Calibration

Given 4-tuples $(^{\mathcal{C}}p_c, p_{\ell_1}, p_{\ell_2}, p_{\ell_3})$ and the normals corresponding to the laser points, we can formulate the equation system based on the model (1). However, this formulation intrinsically treats the result of the ORB-SLAM as rigid observation, regardless of the unequal uncertainty in each map point. To consider this property, we add the global bundle adjustment [27] as a constraint in the cost function in addition to (1). Therefore, the visual features which are not from the chessboard are also employed, i.e., $\{^{\mathcal{C}}p_f\}\backslash\{^{\mathcal{C}}p_c\}$. As the equation system is over determined, we formulate the problem as optimization with the cost function

$$E = \sum_{k,i} E_{\text{proj}}\left(^{\mathcal{L}}_{\mathcal{C}_k}x, {}^{\mathcal{L}}p_{f_i}\right) + \sum_{i} E_{pl}\left(^{\mathcal{L}}p_{c_i}\right) \tag{7}$$

where the first part indicates the cost of bundle adjustment, while the second part stands for the model error. Specifically, $E_{proj}(^{\mathcal{L}}_{\mathcal{C}_k}x, {}^{\mathcal{L}}p_{f_i})$ represents the reprojection error term for the kth camera pose and the ith feature point

$$E_{\text{proj}}(^{\mathcal{L}}_{\mathcal{C}_k}x, {}^{\mathcal{L}}p_{f_i}) = \rho\left(h_{ik}^{\mathrm{T}}\Omega_{ik}h_{ik}\right) \tag{8}$$

with

$$h_{ik} \triangleq \pi\left(^{\mathcal{L}}p_{f_i}, ^{\mathcal{L}}_{\mathcal{C}_k}x\right) - u_{ik} \tag{9}$$

where $\rho(\cdot)$ is the Huber robust cost function [30], $\pi(\cdot,\cdot)$ is the image projection function that projects the first entry to the image with pose specified by the second entry, and u_{ik} denotes the corresponding image feature point. Ω_{ik} is the information matrix encoding the uncertainty of each measurement yielded by ORB detection [25]. Note that the visual feature points and the camera trajectory are now represented in the coordinates of $\{\mathcal{L}\}$.

$E_{pl}(^{\mathcal{L}}p_{c_i})$ is the model error. Given the ith 4-tuple denoted as $(^{\mathcal{C}}p_{c_i}, p_{\ell_{i,1}}, p_{\ell_{i,2}}, p_{\ell_{i,3}})$, we have

$$E_{pl}(^{\mathcal{L}}p_{c_i}) = \sum_{j=1}^{3} \rho\left(y_{ij}^{\mathrm{T}}\Omega_L y_{ij}\right) \tag{10}$$

with

$$y_{ij} \triangleq n_{\ell_{i,j}}^{\mathrm{T}}\left(^{\mathcal{L}}p_{c_i} - p_{\ell_{i,j}}\right) \tag{11}$$

where y_{ij} is the point-to-plane error and Ω_L is the information matrix determined by covariance of the noise in LiDAR measurement, which can be checked in the datasheet of the LiDAR. Note that $^{\mathcal{L}}p_{c_i}$ are not new variables, and they are part of the visual features $^{\mathcal{L}}p_{f_i}$, which are generated from the chessboards.

We solve this optimization problem with the Gauss–Newton algorithm [31] implemented in g2o [32]. The graph of the whole optimization problem is shown in Fig. 3. The structure of the problem is equivalent to the localization of a moving camera in the map built by the first scan of LiDAR. Therefore, the first pose $^{\mathcal{L}}_{\mathcal{C}_0}\hat{x}$ of the resultant estimated camera trajectory $\{^{\mathcal{L}}_{\mathcal{C}_k}\hat{x}\}$ is the extrinsic parameter between the camera and the LiDAR as mentioned before.

The main reason that we state all the variables in $\{\mathcal{L}\}$ is the reduction of unknowns. If we represent the camera trajectory

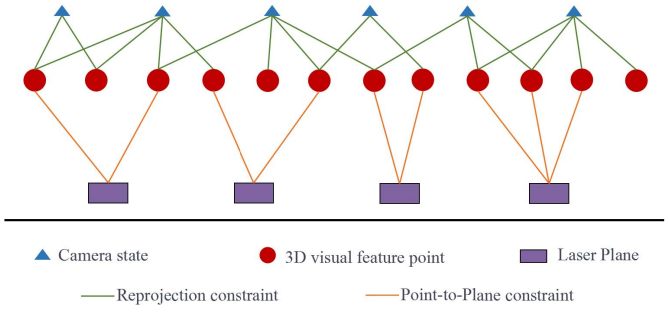


Fig. 3. Optimization of the camera state $\{{}^{\mathcal{L}}_{\mathcal{C}_k}x\}$, 3-D visual feature points $\{{}^{\mathcal{L}}p_f\}$ with the reprojection constraint, the points belonging to the chessboard are represented as $\{{}^{\mathcal{L}}p_c\}$ with the point-to-plane constraint. The LiDAR points on the laser plane (i.e., $\{p_\ell\}$) are observations which will not change with time.

and visual feature points in coordinates $\{\mathcal{C}\}$, an extra variable is needed as extrinsic parameter. In addition, the proposed method is for the points on the plane, which is the chessboard in our experiments. The method can also be extended to the point-based calibration targets when the form of the model error is different, i.e., from point-to-plane error to point-to-point error, which is shown in Appendix B.

III. Observability Analysis

The observability analysis is able to identify the degenerated case, which provides practice guideline when applying the proposed method for calibration, thus very important. We first model the proposed calibration process as a dynamic system, of which the observability is then determined by looking at the rank deficient of the observability matrix. Finally, the number of the chessboard is connected to the observability analysis to reflect the practical results. In sequel, we present the observability analysis in this order.

A. Dynamic System Modeling

The state of the system is the unknowns in (7), i.e., ${}^{\mathcal{L}}_{\mathcal{C}_k}x$ and $\{{}^{\mathcal{L}}p_{f_i}\}$ and $\{{}^{\mathcal{L}}p_{c_i}\}$. To simplify the analysis, we assume that all features are on chessboards, $\{{}^{\mathcal{L}}p_{f_i}\} = \{{}^{\mathcal{L}}p_{c_i}\}$, and thus, we only keep $\{{}^{\mathcal{L}}p_{c_i}\}$ in the state. We further represent ${}^{\mathcal{L}}_{\mathcal{C}_k}x$ by a translation vector ${}^{\mathcal{L}}_{\mathcal{C}_k}p$ and a rotation matrix ${}^{\mathcal{C}_k}_{\mathcal{L}}R$. Thus, the final definition of the state at time t_k consists of ${}^{\mathcal{C}_k}_{\mathcal{L}}R$, ${}^{\mathcal{L}}_{\mathcal{C}_k}p$, and $\{{}^{\mathcal{L}}p_{c_i}\}$.

1) State Propagation: The error state of ${}^{\mathcal{C}_k}_{\mathcal{L}}R$ and ${}^{\mathcal{L}}_{\mathcal{C}_k}p$ is defined as

$$\xi_k = [\,{}^{\mathcal{L}}_{\mathcal{C}_k}\tilde{\theta}^{\mathrm{T}} \quad {}^{\mathcal{L}}_{\mathcal{C}_k}\tilde{p}^{\mathrm{T}}\,]^T \tag{12}$$

with ${}^{\mathcal{L}}_{\mathcal{C}_k}\tilde{\theta}$ satisfying

$${}^{\mathcal{C}_k}_{\mathcal{L}}R = {}^{\mathcal{C}_k}_{\mathcal{L}}\hat{R}\big(I_3 - \lfloor {}^{\mathcal{L}}_{\mathcal{C}_k}\tilde{\theta} \times \rfloor\big) \tag{13}$$

where ${}^{\mathcal{C}_k}_{\mathcal{L}}\hat{R}$ is the estimation of ${}^{\mathcal{C}_k}_{\mathcal{L}}R$, $\lfloor\cdot\times\rfloor$ is the skew matrix expansion of a vector $\cdot \in \mathbb{R}^3$ [33]. And we also have: ${}^{\mathcal{L}}_{\mathcal{C}_k}\tilde{p}^{\mathrm{T}} = {}^{\mathcal{L}}_{\mathcal{C}_k}p^{\mathrm{T}} - {}^{\mathcal{L}}_{\mathcal{C}_k}\hat{p}^{\mathrm{T}}$, where ${}^{\mathcal{L}}_{\mathcal{C}_k}\hat{p}$ is the estimation of ${}^{\mathcal{L}}_{\mathcal{C}_k}p$. Then, the propagation of the rotation error state is

$${}^{\mathcal{L}}_{\mathcal{C}_{k+1}}\tilde{\theta} \simeq {}^{\mathcal{L}}_{\mathcal{C}_k}\tilde{\theta} + {}^{\mathcal{C}_k}_{\mathcal{L}}\hat{R}^{\mathrm{T}}\,{}^{\mathcal{C}_k}_{\mathcal{C}_{k+1}}\tilde{\theta} \tag{14}$$

where ${}^{\mathcal{C}_k}_{\mathcal{C}_{k+1}}\tilde{\theta} \in \mathbb{R}^3$ describing the relative camera rotation error between $\{\mathcal{C}_{k+1}\}$ and $\{\mathcal{C}_k\}$ in $\{\mathcal{C}_k\}$. For propagation of translation error state, we have

$${}^{\mathcal{L}}_{\mathcal{C}_{k+1}}\tilde{p} \simeq -\big\lfloor \big({}^{\mathcal{C}_k}_{\mathcal{L}}\hat{R}^{\mathrm{T}}\,{}^{\mathcal{C}_k}_{\mathcal{C}_{k+1}}\hat{p}\big) \times \big\rfloor\, {}^{\mathcal{L}}_{\mathcal{C}_k}\tilde{\theta} + {}^{\mathcal{C}_k}_{\mathcal{L}}\hat{R}^{\mathrm{T}}\,{}^{\mathcal{C}_k}_{\mathcal{C}_{k+1}}\tilde{p} + {}^{\mathcal{L}}_{\mathcal{C}_k}\tilde{p} \tag{15}$$

where ${}^{\mathcal{C}_k}_{\mathcal{C}_{k+1}}\tilde{p}$ is the relative camera translation error between $\{\mathcal{C}_{k+1}\}$ and $\{\mathcal{C}_k\}$ in $\{\mathcal{C}_k\}$. As the derivation of error state is not the focus of this paper, we refer to [33] for readers who has interests. For clearance in the following derivation, we denote the shorthand notation of ${}^{\mathcal{C}_k}_{\mathcal{L}}\hat{R}$ as $\hat{R}_k$. Then, we write the error state propagation function in a matrix form as

$$\xi_{k+1} = \phi_{\xi_k}\xi_k + \alpha_k \tag{16}$$

where

$$\phi_{\xi_k} \triangleq \begin{bmatrix} I_3 & 0_{3\times3} \\ -\big\lfloor \big(\hat{R}_k^{\mathrm{T}}\,{}^{\mathcal{C}_k}_{\mathcal{C}_{k+1}}\hat{p}\big) \times \big\rfloor & I_3 \end{bmatrix} \quad \alpha_k \triangleq \begin{bmatrix} \hat{R}_k^{\mathrm{T}}\,{}^{\mathcal{C}_k}_{\mathcal{C}_{k+1}}\tilde{\theta} \\ \hat{R}_k^{\mathrm{T}}\,{}^{\mathcal{C}_k}_{\mathcal{C}_{k+1}}\tilde{p} \end{bmatrix} \tag{17}$$

with $0_{3\times3}$ indicating a 3×3 zero matrix. When augmented with m static visual features $\{{}^{\mathcal{L}}p_{c_i}\}$, we have the full state propagation function in the following:

$$\begin{bmatrix} \xi_{k+1} \\ {}^{\mathcal{L}}p_{c_{1\ldots M}} \end{bmatrix} = \phi_k \begin{bmatrix} \xi_k \\ {}^{\mathcal{L}}p_{c_{1\ldots M}} \end{bmatrix} + \begin{bmatrix} \alpha_k \\ 0_{3m\times1} \end{bmatrix} \tag{18}$$

where ${}^{\mathcal{L}}p_{c_{1\ldots M}}$ is $\{{}^{\mathcal{L}}p_{c_i}\}$ stacked in column, and

$$\phi_k \triangleq \begin{bmatrix} \phi_{\xi_k} & 0_{6\times3m} \\ 0_{3m\times6} & I_{3m} \end{bmatrix}. \tag{19}$$

2) Linearized Measurements: According to the cost function (7), we have two kinds of measurement models: the reprojection error (9) and the point-to-plane error (11). The Jacobian of the reprojection error measurement H_{ik} is

$$H_{ik} = \begin{bmatrix} H_{\mathcal{C}_{ik}} & 0_{2\times3} & \ldots & H_{f_{ik}} & \ldots & 0_{2\times3} \end{bmatrix} \tag{20}$$

where the entries are defined as

$$H_{f_{ik}} \triangleq \frac{\partial h_{ik}}{\partial {}^{\mathcal{L}}p_{c_i}} = J_{ik}\cdot\hat{R}_k \tag{21}$$

$$H_{\mathcal{C}_{ik}} \triangleq \frac{\partial h_{ik}}{\partial \xi_k} = H_{f_{ik}}\left[\big\lfloor \big({}^{\mathcal{L}}\hat{p}_{c_i} - {}^{\mathcal{L}}_{\mathcal{C}_k}\hat{p}\big) \times \big\rfloor \; -I_3\right] \tag{22}$$

where J_{ik} is the Jacobian of reprojection measurement with respect to ${}^{\mathcal{C}_k}p_{c_i}$, the ith visual feature in camera coordinates $\{\mathcal{C}_k\}$

$$J_{ik} \triangleq \frac{1}{{}^{\mathcal{C}_k}\hat{p}_{c_{i_z}}}\begin{bmatrix} 1 & 0 & \dfrac{-{}^{\mathcal{C}_k}\hat{p}_{c_{i_x}}}{{}^{\mathcal{C}_k}\hat{p}_{c_{i_z}}} \\ 0 & 1 & \dfrac{-{}^{\mathcal{C}_k}\hat{p}_{c_{i_y}}}{{}^{\mathcal{C}_k}\hat{p}_{c_{i_z}}} \end{bmatrix} \tag{23}$$

$${}^{\mathcal{C}_k}\hat{p}_{c_i} \triangleq \begin{bmatrix} {}^{\mathcal{C}_k}\hat{p}_{c_{i_x}} & {}^{\mathcal{C}_k}\hat{p}_{c_{i_y}} & {}^{\mathcal{C}_k}\hat{p}_{c_{i_z}} \end{bmatrix}^{\mathrm{T}} \tag{24}$$

where ${}^{\mathcal{C}_k}\hat{p}_{c_i}$ is the estimate of ${}^{\mathcal{C}_k}p_{c_i}$.

For the point-to-plane error measurement (11), we have the Jacobian as

$$Y_{ij} = \begin{bmatrix} 0_{1\times 6} & 0_{1\times 3} & \dots & n_\ell^{\mathrm{T}} & \dots & 0_{1\times 3} \end{bmatrix} \tag{25}$$

where Y_{ij} is the Jacobian of the point-to-plane error measurement (11) with respect to ξ_k and ith visual feature ${}^{\mathcal{L}}p_{c_i}$. As the camera pose is not involved in this class of measurements, the Jacobian is much simpler than (20). As a whole, the Jacobian of the measurement with respect to ξ_k and ith visual feature ${}^{\mathcal{L}}p_{c_i}$ can be written as

$$Q_{ik} \triangleq \begin{bmatrix} H_{ik} \\ Y_{ij} \end{bmatrix}. \tag{26}$$

B. Rank Deficient of Observability Matrix

The observability matrix for the time interval between time t_s and t_{s+w} is defined as

$$M \triangleq \begin{bmatrix} Q_s \\ Q_{s+1}\phi_s \\ \vdots \\ Q_{s+w}\phi_{s+w-1}\cdots\phi_s \end{bmatrix} \tag{27}$$

where Q_s is constructed by stacking Q_{is} for all features in column. Directly analyze the null-space of M is not easy. Instead, our idea to investigate the rank deficiency of M consists of two steps. In the first step, we build an observability matrix for the sub-system with only one class of measurement (9). In the second step, we substitute the null-space derived from the sub-system into the observability matrix for the full system to see the change of rank deficiency when different numbers of chessboards are given, leading to the final practical results.

1) Null-Space of Sub-System: For each block in (27), we define the shorthand notation

$$M_k = Q_k\phi_{k-1}\dots\phi_s. \tag{28}$$

Given the sub-system with only one type of measurement, we have

$$\breve{M}_k = H_k\phi_{k-1}\dots\phi_s \tag{29}$$

where H_k is constructed by stacking H_{ik} for all features in column. We further define that

$$\breve{M}_{ik} \triangleq H_{ik}\phi_{k-1}\dots\phi_s = J_{ik}\hat{R}_k \begin{bmatrix} \Gamma_{ik} & -I_3 & 0_{3\times 3} & \dots & I_3 & \dots & 0_{3\times 3} \end{bmatrix} \tag{30}$$

where we have

$$\Gamma_{ik} \triangleq \left\lfloor \left({}^{\mathcal{L}}\hat{p}_{c_i} - {}^{\mathcal{L}}_{\mathcal{C}_s}\hat{p}\right)\times \right\rfloor. \tag{31}$$

Directly, we have $\breve{M}_k$ represented by

$$\breve{M}_k \triangleq \begin{bmatrix} \breve{M}_{1k}^T & \dots & \breve{M}_{ik}^T & \dots & \breve{M}_{mk}^T \end{bmatrix}^{\mathrm{T}}. \tag{32}$$

At this point, we derive the null-space of $\breve{M} = \begin{bmatrix} \breve{M}_s^T & \dots & \breve{M}_k^T & \dots & \breve{M}_{s+w}^T \end{bmatrix}^{\mathrm{T}}$ as

$$N \triangleq \begin{bmatrix} 0_{3\times 3} & I_3 \\ I_3 & -\left\lfloor {}^{\mathcal{L}}_{\mathcal{C}_s}\hat{p}\times \right\rfloor \\ I_3 & -\left\lfloor {}^{\mathcal{L}}\hat{p}_{c_1}\times \right\rfloor \\ \vdots & \vdots \\ I_3 & -\left\lfloor {}^{\mathcal{L}}\hat{p}_{c_m}\times \right\rfloor \end{bmatrix}. \tag{33}$$

It is easy to verify that $\breve{M}_{ik}N = 0_{2\times 6}$. Since this holds for any i and any k (i.e., for all blocks of the observability matrix), we conclude that $\breve{M}N = 0_{2m(w+1)\times 6}$. As a result, the rank deficient of the observability matrix for the sub-system is 6.

Note that the sub-system can be regarded as a scenario that none of the chessboards is detected, so that there is no data association between the LiDAR scan and visual features. It is thus intuitive to understand the rank deficiency that the extrinsic parameter cannot be determined in such a scenario. In other words, the original problem is equivalent to localization of moving camera in LiDAR scan as mentioned above, which degenerates to general SLAM problem when no data association between LiDAR scan and visual features is found. For general SLAM system, the rank deficiency is 6 in accordance to the result derived in [33]. The unobservable directions are corresponding to the first camera pose, ${}^{\mathcal{L}}_{\mathcal{C}_0}x$, which is just the extrinsic parameter in our scenario.

2) Rank Deficient of Calibration System: When the full system is considered, we have each block of (28)

$$M_{ik} \triangleq \begin{bmatrix} & & & \breve{M}_{ik} & & \\ 0_{1\times 3} & 0_{1\times 3} & 0_{1\times 3} & \dots & n_\ell^{\mathrm{T}} & \dots & 0_{1\times 3} \end{bmatrix}. \tag{34}$$

To look into the rank deficiency of the observability matrix M, we can multiply M with the null-space N in (33), and if any of the direction becomes non-zero, the rank deficiency decreases correspondingly.

a) Observation of one plane: Assuming there are two points on the observed chessboard plane, namely the ith and the jth visual feature points, we have

$$M_{(i,j)k} \triangleq \begin{bmatrix} & & & \breve{M}_{ik} & & & \\ & & & \breve{M}_{jk} & & & \\ 0_{1\times 3} & 0_{1\times 3} & 0_{1\times 3} & \dots & n_\ell^{\mathrm{T}} & \dots & 0_{1\times 3} \\ 0_{1\times 3} & 0_{1\times 3} & 0_{1\times 3} & \dots & \dots & n_\ell^{\mathrm{T}} & 0_{1\times 3} \end{bmatrix}. \tag{35}$$

Note that for both visual feature points, the normal of the chessboard is the same, denoted as n_ℓ. Then, we have

$$M_{(i,j)k}N_1 = 0_{6\times 3} \tag{36}$$

where N_1 is the null-space, which is a linear combination of N. The details of derivation are described in Appendix A. Since this holds for any i, j, and k, we conclude that $MN_1 = 0_{3m(w+1)\times 3}$, which means the rank deficiency is 3. The result is explained as when observing only one chessboard, any translation parallel to the plane's normal and any rotation around the plane's normal are unobservable.

b) Observation of two planes: Assuming the ith feature and the jth feature lie on two chessboards, of which the normals are denoted by n_{ℓ_a} and n_{ℓ_b}, respectively, we have

$$M_{(i,j)k} \triangleq \begin{bmatrix} & \breve{M}_{ik} & & & & & \\ & \breve{M}_{jk} & & & & & \\ 0_{1\times3} & 0_{1\times3} & 0_{1\times3} & \dots & n_{\ell_a}^{\mathrm{T}} & \dots & 0_{1\times3} \\ 0_{1\times3} & 0_{1\times3} & 0_{1\times3} & \dots & \dots & n_{\ell_b}^{\mathrm{T}} & 0_{1\times3} \end{bmatrix}. \quad (37)$$

Note that the third and fourth rows in (37) are different from that in (35). For this block of observability matrix, we have

$$M_{(i,j)k} N_2 = 0_{6\times1} \quad (38)$$

where N_2 is the null-space which is described in Appendix A. Since this holds for any i, j, and k, we conclude that $MN_2 = 0_{3m(w+1)\times1}$. Therefore, when observing two chessboards, translation along the direction perpendicular to the normals of the two planes is unobservable.

c) Observation of three planes: Similar to the previous derivation process, we conclude that when three planes with non-collinear normals are observed, the rank deficiency is 0. That is to say, the calibration system is fully observable at least three non-parallel chessboards are observed. Intrinsically, this result suggests that a larger common field of view between the camera and the LiDAR is needed to guarantee the observability and reliable detection of the calibration targets, which is difficult for the static method as shown in experiments. The details of derivation are described in Appendix A.

d) Observability of other calibration targets: As mentioned in Section II, switching the model to the point-to-point error makes the method applicable to other types of calibration targets, i.e., corners of polygonal planar boards [34] and boxes [35]. Accordingly, the result of the observability analysis for the point-to-point error is derived by switching the last rows in blocks of observability matrix, that is (35) and (37). The results for point-based calibration targets are when observing only one point, any rotation is unobservable, while observing two points, one degree of freedom of the rotation is unobservable. When three non-collinear points are observed, the extrinsic parameter is determined. Please refer to Appendix B for derivation details.

IV. Placement of Calibration Targets

The previous analysis concluded that in order to calibrate the six-DoF extrinsic parameter, three chessboards are needed at least. How to place these three chessboards in space to get the better calibration accuracy is what we will discuss next. The following theory can provide a guideline for designing high-accuracy calibration procedures.

To simplify the problem we are analyzing, we derive the problem in 2-D and consider two chessboards that can still provide insights into real-world applications. Analyzing the calibration accuracy refers to analyzing the uncertainty of the extrinsic parameter, which is measured by the determination of the inverse covariance matrix. The larger the determinant of the inverse covariance matrix, the smaller the uncertainty of estimated extrinsic parameter, thus more accurate. Specifically, we are going to investigate the influence of the angle and distance between two chessboards.

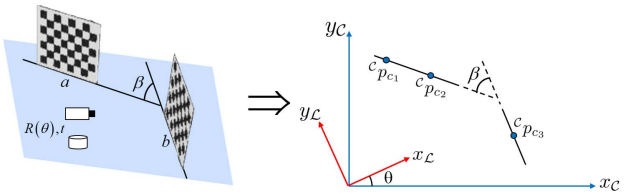


Fig. 4. Specific explanation of the parameters and representation of 2-D.

A. Angle Between Calibration Targets

For clearance, we keep the notation but with a slight abuse. In this section, all the notations indicate for the entities in 2-D to support the analysis. As shown in Fig. 4, the angle of chessboard b relative to chessboard a is β, the angle of $\{\mathcal{C}\}$ relative to $\{\mathcal{L}\}$ is θ, and the extrinsic parameter ${}^{\mathcal{L}}_{\mathcal{C}}x$ is reduced to three DoFs represented by

$$R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad t = \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \quad (39)$$

Then, the point-to-plane error (11) is modified as

$$y = n_\ell^{\mathrm{T}}(R(\theta)\,{}^{\mathcal{C}}p_c + t - p_\ell) \quad (40)$$

where ${}^{\mathcal{C}}p_c$ is the 2-D visual point in the camera coordinate system, $R(\theta)$ and t are the 2-D extrinsic parameters, p_ℓ is the 2-D laser point in the laser coordinate system, and n_ℓ is the 2-D normal vector of p_ℓ.

To describe the uncertainty of the extrinsic parameter, we apply a linearized propagation based on the Jacobian matrix

$$J \triangleq \begin{bmatrix} n_\ell^{\mathrm{T}}(R(\theta)'\,{}^{\mathcal{C}}p_c) & n_\ell^{\mathrm{T}} \end{bmatrix} \quad (41)$$

where $R(\theta)'$ denotes the derivatives of $R(\theta)$ with respect to θ. Assuming an isometric covariance matrix in the measurement, we can derive the following inverse covariance matrix of extrinsic parameter as:

$$H \triangleq J^{\mathrm{T}}J. \quad (42)$$

In order to explore the placement of the chessboards, we have further simplified the situation. Assume that the normal vector of the chessboard a is

$$n_{\ell_a} \triangleq \begin{bmatrix} 1 & 0 \end{bmatrix}^{\mathrm{T}}. \quad (43)$$

Then, the normal vector of the chessboard b is

$$n_{\ell_b} \triangleq \begin{bmatrix} \cos\beta & \sin\beta \end{bmatrix}^{\mathrm{T}}. \quad (44)$$

Thereafter, we take two points ${}^{\mathcal{C}}p_{c_1}$ and ${}^{\mathcal{C}}p_{c_2}$ from the chessboard a and one point ${}^{\mathcal{C}}p_{c_3}$ from the chessboard b. Following (42), we have:

$$\begin{aligned} H &\triangleq J_1^{\mathrm{T}}J_1 + J_2^{\mathrm{T}}J_2 + J_3^{\mathrm{T}}J_3 \\ J_1 &\triangleq \begin{bmatrix} n_{\ell_a}^{\mathrm{T}}(R(\theta)'\,{}^{\mathcal{C}}p_{c_1}) & n_{\ell_a}^{\mathrm{T}} \end{bmatrix} \\ J_2 &\triangleq \begin{bmatrix} n_{\ell_a}^{\mathrm{T}}(R(\theta)'\,{}^{\mathcal{C}}p_{c_2}) & n_{\ell_a}^{\mathrm{T}} \end{bmatrix} \\ J_3 &\triangleq \begin{bmatrix} n_{\ell_b}^{\mathrm{T}}(R(\theta)'\,{}^{\mathcal{C}}p_{c_3}) & n_{\ell_b}^{\mathrm{T}} \end{bmatrix} \end{aligned} \quad (45)$$

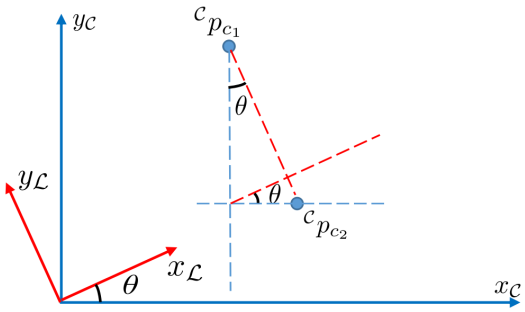


Fig. 5. Schematic of (47).

of which the determinant is derived analytically

$$|H| = (\kappa_1 + \kappa_2)^2 \sin^2(\beta) \tag{46}$$

where ${}^{\mathcal{C}}p_{c_1} \triangleq \left[p_{c_{1_x}}\ p_{c_{1_y}}\right]^{\mathrm{T}}$, ${}^{\mathcal{C}}p_{c_2} \triangleq \left[p_{c_{2_x}}\ p_{c_{2_y}}\right]^{\mathrm{T}}$, $\kappa_1 \triangleq \sin\theta(p_{c_{1_x}} - p_{c_{2_x}})$, and $\kappa_2 \triangleq \cos\theta(p_{c_{1_y}} - p_{c_{2_x}})$.

It can be seen that when $\beta = \pi/2$, $|H|$ takes the maximum value, that is, when the angle between the two chessboards is 90°, the uncertainty of the estimated extrinsic parameter is the smallest.

B. Distance Between Calibration Targets

The conclusion of Section IV-A is that the calibration error uncertainty is smallest when the two chessboards are placed orthogonal to each other. In this section, when the angle between the two chessboards is fixed, we discuss the effect of the distance between two chessboards.

Following (46), when the angle is fixed, we have:

$$|H| = (\kappa_1 + \kappa_2)^2. \tag{47}$$

The illustration of (47) can be seen in Fig. 5.

We can find that when the relative angle between the chessboards is fixed in 90°, the uncertainty of the extrinsic parameter can be reduced by increasing the relative distance between the two visual observations on one chessboard. More intuitively, this result requires a very large chessboard so that visual observations on the chessboard can be larger. However, the requirement of large chessboard and the orthogonality between the two chessboards are hard to be satisfied at the same time in reality due to the limited view of camera. These results suggest that it is possible to improve the calibration accuracy by moving the camera since more scattered observations can be made.

Comments:

1) Combining the analysis of observability and the minimal necessary conditions for calibration, we conclude that at least three chessboards are required and more chessboards can lead to better calibration accuracy.
2) With the same number of calibration targets, a scattered placement is better than a centralized one, which is expected to be true in 3-D.
3) The extended camera field of view obtained by our method meets the requirement of observing multiple calibration targets, which is difficult in those methods that keep the sensors stationary.

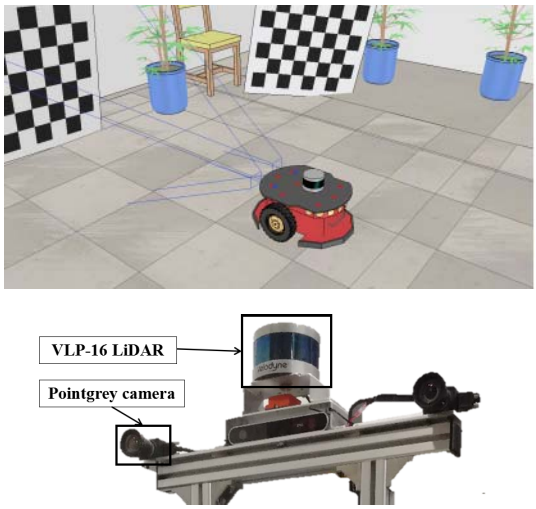


Fig. 6. Sensor configuration. Top: VLP-16 LiDAR and vision sensor in the simulation environment. Bottom: VLP-16 LiDAR and Pointgrey camera in the real world.

4) Observing multiple calibration targets arranged in various poses by the extended field of view also gives our method an advantage compared to the appearance-based method.

V. Experimental Results

In order to evaluate our method, we performed simulation verification and real-world experiments separately. In the simulation experiments, we showed that the placement theory derived in Section IV is reasonable by performing our calibration method with calibration targets placed in scattered and centralized arrangements, respectively. Then, we made seven experiments in accuracy comparison section in order to compare with other methods. We built a simulation environment in V-REP [36], using a stereo vision sensor and a Velodyne VLP-16 LiDAR to obtain data, as shown in Fig. 6 (top). In the real-world experiments, the comparison with other methods is also performed demonstrating the practicality of the proposed method. As shown in Fig. 6 (bottom), we fixed two Pointgrey cameras with a Velodyne VLP-16 LiDAR on the robot to perform the real-world experiments.

A. V-REP Simulation

In the simulation environment, the final calibration ${}^{\mathcal{L}}_{\mathcal{C}}x$ is expressed as rotation R and translation t. R and t are compared against ground truths R_g and t_g, which are obtained from V-REP. Following [14], for the translation error, we computed $\|t - t_g\|$ in meters. For the rotation error, we first computed the relative rotation $\delta R = R^{-1}R_g$ and represented it in degrees by axis-angle representation.

1) Theoretical Verification: We first verify the theoretically derived conclusions. Due to the existence of observation errors, the angle and distance between the calibration targets are highly coupled, and it is impossible to perform control variables to verify the influence of angle and distance separately. We only verify the final conclusions derived from the theory. As shown in Fig. 7(a) and (b), we placed four chessboards

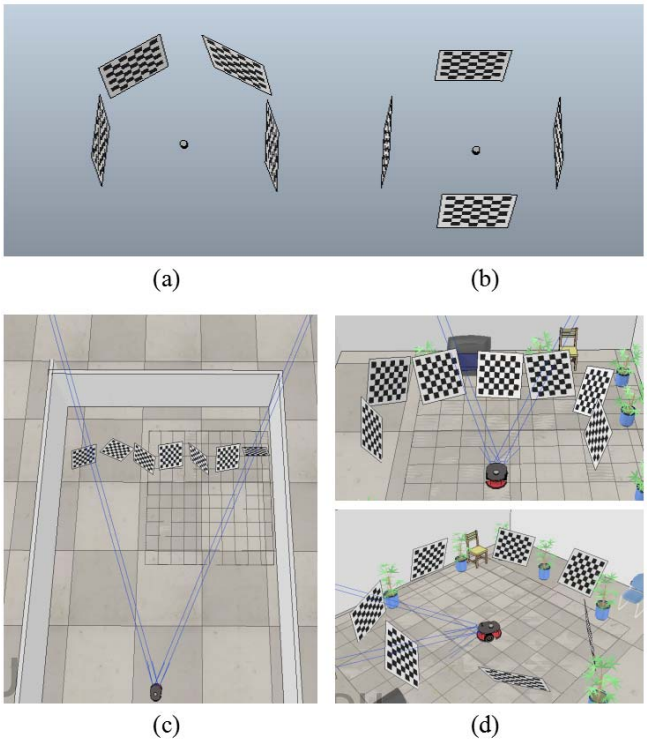


Fig. 7. Scenes for obtaining data with chessboards. (a) and (b) Obtaining data in the scene in which four chessboards are placed nearly vertical to the ground. (c) Placement of the chessboards and the sensors to collect the data required by the KITTI single-shot method. (d) Obtaining data in the scene in which seven chessboards are placed in various poses.

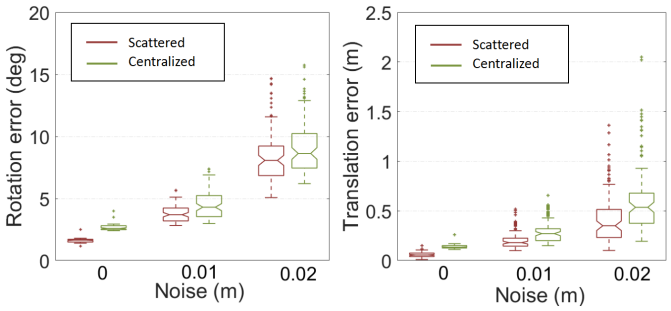


Fig. 8. Theoretical verification results: errors from the ground truth of the calibration result by our method with chessboards placed centralized and scattered.

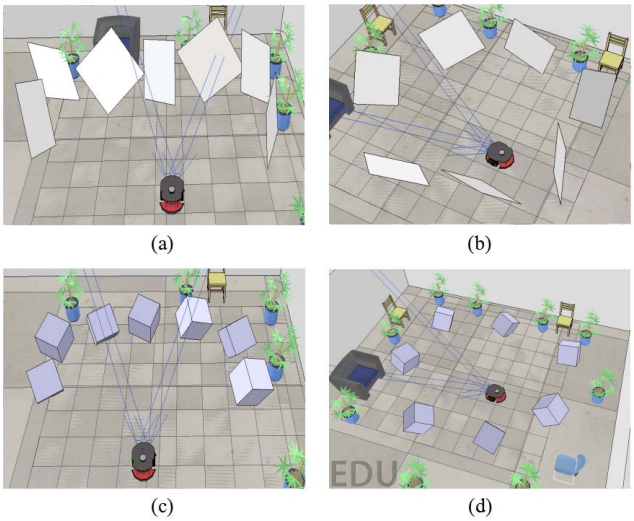


Fig. 9. Scenes for obtaining data. (a) and (b) Obtatining data with polygonal planar boards placed centralized and scattered. (c) and (d) Obtatining data with boxes placed centralized and scattered.

around the sensor and used our method to obtain data for calibration. In order to simulate the situation in 2-D, four chessboards are placed nearly vertical to the ground, and one set is centralized in front of the field of view, while the other is scattered around the sensor. For providing sufficient constraints for calibration, we made the experiment in which the chessboards are placed at a 5° angle to the direction of gravity.

In this experiment, we added Gaussian noise $\mathcal{N}(0, \sigma^2)$ to the laser data for varying values of σ and carried out calibration. We calculated the errors between the calibration result and ground truth, which is used for drawing box plot to evaluate the error mean and dispersion of the error. This process is repeated 100 times and the dispersion is used to evaluate the uncertainty of the estimated extrinsic parameter. The results, shown in Fig. 8, indicate that the calibration accuracy is better when the angle between the two chessboards is 90°, as shown in Fig. 7(b), which is consistent with the conclusion of Section IV. As shown in Fig. 7(a) and (b), in order to obtain a larger common field of view between the two sensors to observe four chessboards, the camera's extended field of view obtained by our method is needed.

2) Accuracy Comparison: Next, we compared our method with the *KITTI single-shot method* [14] on calibration accuracy. The KITTI single shot calibration method can automatically give the extrinsic calibration results in one acquisition, which is convenient to use. As shown in Fig. 7(c), the method requires placing multiple chessboards in front of the field of view and obtaining the LiDAR and camera data in one shot, respectively. However, if one wants to obtain as much data as possible from the chessboards, the sensors should be placed much farther away from the chessboards, which can be easily seen from Fig. 7(c). Once the sensor is too far from the chessboards, it is often difficult to extract the corner points from the obtained camera image, and the laser lines hitting the chessboards are also reduced. Another main limiting assumption of the single-shot method is the common field of view between camera and LiDAR.

As shown in Fig. 7(d) (top and bottom), we then obtained two sets of data for seven chessboards centralized and scattered in a certain field of view and evaluated proposed method (labeled as *chessboard method*). Unlike the KITTI single-shot method, when we captured the camera images, we moved the robot around to obtain images of each chessboard, and then reconstructed the 3-D visual points in the space.

Our method does not limit the calibration target, so we used polygonal planar boards [34] [labeled as *polygonal method*, as shown in Fig. 9(a) and (b)] and boxes [35] [labeled as *box method*, as shown in Fig. 9(c) and (d)] as calibration targets for a complement to this paper. As the detection of the corner is not the point in this paper, for the LiDAR point clouds,

TABLE I

ACCURACY COMPARISON RESULTS: ERRORS OF THE CALIBRATION RESULT BY THE SINGLE SHOT METHOD, THE CHESSBOARD METHOD, THE POLYGONAL METHOD, AND THE BOX METHOD

Error	Centralized Placement		Scattered Placement	
	Translation (m)	Rotation (deg)	Translation (m)	Rotation (deg)
Single shot	0.1910	1.3030	-	-
Chessboard	0.0270	0.8250	0.0083	0.6470
Polygonal	0.0130	0.4130	0.0078	0.3570
Box	0.0091	0.3680	0.0075	0.2730

we manually extracted the corner points of the polygonal planar boards or the calibration boxes from V-REP and added noise; for the image data, we manually picked the planar boards and boxes' feature points from the image feature points. The experiments are set to examine whether the theory is feasible and effective in the scenario when point measurement error-based calibration target is considered. In order to verify the conclusions of the theoretical derivation, we obtained two sets of data for seven calibration targets centralized and scattered in a certain field of view for the polygonal method and the box method.

The results, shown in Table I, indicate that the proposed method achieves better calibration results than the KITTI single-shot method, because the camera's field of view limits the number of laser lines hitting the chessboards and the number of observed chessboards in the single-shot method. It can also be seen in Table I that polygons and boxes achieve higher accuracy than chessboards since the point-to-point error model provides more constraints than the point-to-plane error. The important finding is that, for each calibration target, the calibration results of the scattered placement are better than the centralized placement, which is consistent with the theoretic results in Section IV, suggesting that our method is a general back end to various kinds of detector.

B. Real-World Experiment

We conducted real-world experiments to compare the proposed method with three calibration methods: *the KITTI single shot calibration method* [14]; *MO methods*: the method needs multiple pairs of images and LiDAR data of a single calibration target presented in different poses when the sensor system keeps statically, similar to [8] and [9]; and *motion-based* calibration method: the method estimates the extrinsic parameter by aligning the trajectories of camera and LiDAR.

We obtained one set of data for the KITTI single shot, two sets of data in which the chessboards were placed scattered and centralized for our method, and two sets of data for the motion-based calibration. And we obtained one set of data for the MO calibration method, including 67 corresponding laser scans and camera images under different poses. The 67 pairs of laser scans and camera images were divided into two parts. The first 30 data pairs were used for the MO method to calibrate the extrinsic parameter and the last 37 remaining data pairs were used to blind test the accuracy of all methods. We evaluated the calibration results estimated by different methods with the point-to-plane error, which is computed for each visual point according to (11). The average point-to-plane

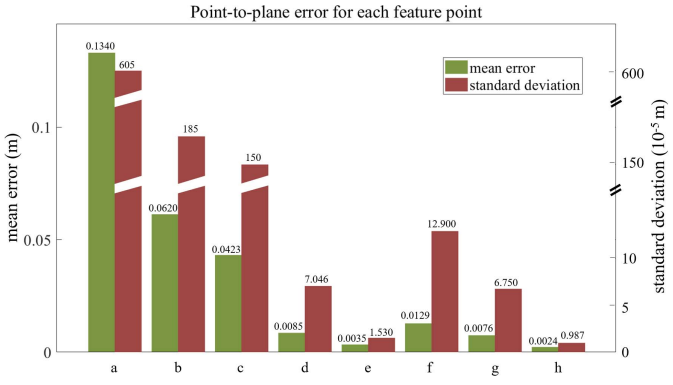


Fig. 10. Real-world experiment results: calibration errors for (a) motion-based non-sync, (b) motion-based with sync, (c) MO-10, (d) MO-20, (e) MO-30, (f) KITTI single shot, (g) centralized, and (h) scattered.

error for each feature point of each method is given by

$$\varepsilon_{pl} = \frac{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{3}\left(n_{\ell_{i,j}}^{\mathrm{T}}\left({}^{\mathcal{L}}p_{c_i} - p_{\ell_{i,j}}\right)\right)^2}}{3m} \tag{48}$$

where m is the number of visual feature points in the data, $p_{\ell_{i,j}}$ is an associated laser point for ${}^{\mathcal{L}}p_{c_i}$ from data association above, and $n_{\ell_{i,j}}$ is the normal vector of $p_{\ell_{i,j}}$. In order to show the uncertainty in real-world experiment, we used the bootstrapping method [37] that relies on random sampling with replacement in test data to approximate the variance. We conducted 100 samples to get 100 sets for calculating point-to-plane error, which is used to get mean error and standard deviation. We use the standard deviation to evaluate the uncertainty of the blind testing error. The calibration error is shown in Fig. 10.

The KITTI single shot calibration method does not give an accurate result in the case of the Velodyne VLP-16 LiDAR due to the insufficient lines, as shown in Fig. 10. The calibration results of our *scattered* placement calibration method are better than the calibration results of the *centralized* placement, which is consistent with the conclusion of Section IV. This result proves once again that when the calibration target is static, a larger common field of view between the two sensors obtained by our method can lead to a better calibration accuracy.

We made two experiments for the motion-based method, one is called the *motion-based non-sync method*, for which we did not complete the hardware synchronization between the two sensors. The other is called the *motion-based with sync method*, for which we completed a rough hardware synchronization by finding the nearest neighbor on the timestamps of the two sensors. For the motion-based calibration method, time synchronization is not easy; even if a rough time synchronization is completed, the calibration result is not as good as the proposed method, which eliminates the time variable, as shown in Fig. 10.

For the MO calibration method, we did another set of experiments by varying the number of laser and image data pairs to explore the relation between the accuracy and the data number. We randomly selected 10 (MO-10),

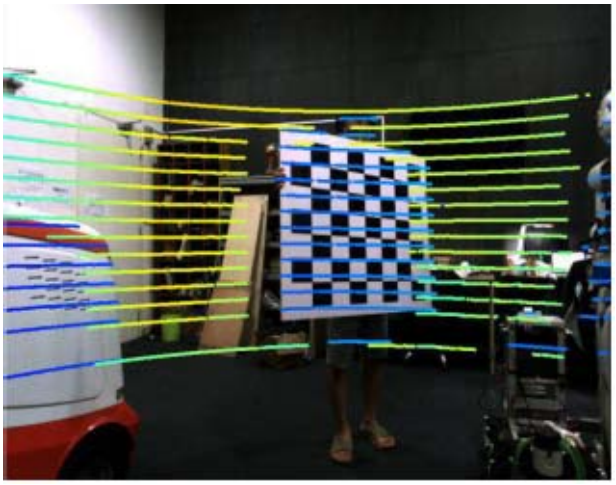


Fig. 11. Results of our calibration of the laser data re-projected into the image. The yellow points on the chessboard are the points observed by LiDAR but not observed by the camera due to the occlusion.

20 (MO-20), and 30 (MO-30) pairs of laser scans and camera images from the aforementioned 30 data pairs to make a comparison experiment. It can be seen in Fig. 10 that with more data used for calibration, the accuracy of the MO calibration method also increases and MO-30 achieves results that are only worse than our calibration method. However, the MO method is time-consuming to use and needs a common field of view between the sensors, which limits its use to some extent.

Compared with other methods, we have summary about our method. First, the extended field of view obtained by our method can improve the chance to satisfy the proposed theoretic results, thus improving the calibration accuracy, and our method can be applied to a case with an arbitrary configuration, which is beneficial and necessary in practical use. Second, our method eliminates the time variable from the spatial extrinsic parameter estimating, so it is applicable to the cases lacking time synchronization and will not introduce additional variables. That is to say, our error term does not include laser motion estimation error and time offset error. Third, the proposed method may be applied for calibrating multiple cameras and LiDAR devices. When the proposed method is used for calibration between two LiDAR devices under arbitrary configurations, one of the LiDAR devices can be considered as a camera to run laser SLAM [38]. Finally, we show an example of the laser data reprojected into the image, which allows us to see the accuracy of the calibration qualitatively, as shown in Fig. 11.

VI. Conclusion

We proposed a LiDAR-camera extrinsic calibration method eliminating the time variable regardless of the limitation of sharing a common field of view. Furthermore, we analyzed the observability of the calibration system and derived how the calibration targets can be placed better to improve the accuracy of the extrinsic calibration. Then, we made a full comparison with other methods through both simulation and real-world experiments, which showed that our method has more chance to satisfy the theoretic findings, thus achieving a better calibration result. In order to simplify the calibration, our next work is to study a calibration method that does not require calibration target and still maintains calibration accuracy.

Appendix

A. Observability With Point-to-Plane Error

What follows is the analysis about the minimal necessary conditions of the chessboards setting to solve the accurate six-DoF extrinsic calibration problem.

1) Observation of One Plane: Assuming there are two points on the observed chessboard plane, namely the ith and the jth visual feature points, we have

$$M_{(i,j)k} \triangleq \begin{bmatrix} & & & & \breve{M}_{ik} & & \\ & & & & \breve{M}_{jk} & & \\ 0_{1\times3} & 0_{1\times3} & 0_{1\times3} & \dots & n_\ell^{\mathrm{T}} & \dots & 0_{1\times3} \\ 0_{1\times3} & 0_{1\times3} & 0_{1\times3} & \dots & \dots & n_\ell^{\mathrm{T}} & 0_{1\times3} \end{bmatrix}. \quad (49)$$

Note that for both visual feature points, the normal of the chessboard is the same, denoted as n_ℓ. Then, we can have

$$M_{(i,j)k}N = \begin{bmatrix} 0_{4\times1} & 0_{4\times1} & 0_{4\times1} & 0_{4\times1} & 0_{4\times1} \\ n_1 & n_2 & n_3 & n_3p_{i2} - n_2p_{i3} & -n_3p_{i1} + n_1p_{i3} \\ n_1 & n_2 & n_3 & n_3p_{j2} - n_2p_{j3} & -n_3p_{j1} + n_1p_{j3} \end{bmatrix} \quad (50)$$

where $n_\ell^{\mathrm{T}} \triangleq \begin{bmatrix} n_1 & n_2 & n_3 \end{bmatrix}$, ${}^{\mathcal{L}}p_{c_i} \triangleq \begin{bmatrix} p_{i1} & p_{i2} & p_{i3} \end{bmatrix}^{\mathrm{T}}$, and ${}^{\mathcal{L}}p_{c_j} \triangleq \begin{bmatrix} p_{j1} & p_{j2} & p_{j3} \end{bmatrix}^{\mathrm{T}}$. Next, we perform an elementary linear transformation on the above equation. That is, both sides of (50) are multiplied by matrix A_1

$$A_1 \triangleq \begin{bmatrix} 1 & -\frac{n_2}{n_1} & -\frac{n_3}{n_1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & n_1 \\ 0 & 0 & 0 & 0 & 1 & n_2 \\ 0 & 0 & 0 & 0 & 0 & n_3 \end{bmatrix} \quad (51)$$

$$M_{(i,j)k}NA_1 = \begin{bmatrix} 0_{4\times1} & 0_{4\times1} & 0_{4\times1} & 0_{4\times1} & 0_{4\times1} \\ n_1 & 0 & 0 & n_3p_{i2} - n_2p_{i3} & -n_3p_{i1} + n_1p_{i3} \\ n_1 & 0 & 0 & n_3p_{j2} - n_2p_{j3} & -n_3p_{j1} + n_1p_{j3} \end{bmatrix}. \quad (52)$$

As shown, there are three columns of $M_{(i,j)k}NA_1$ that become all zeros. So, the second, third, and sixth columns of NA_1 are the null-space of $M_{(i,j)k}$, and we denote it as N_1 and can get

$$M_{(i,j)k}N_1 = 0_{6\times3}. \quad (53)$$

Since this holds for any i, j, and k, we conclude that $MN_1 = 0_{3m(w+1)\times3}$. Note that the first three columns of NA_1 correspond to global translations of the state vector, while

the last three columns to global rotations. Therefore, when observing only one plane, any translation parallel to the plane's normal and any rotation around the plane's normal vector are unobservable.

2) Observation of Two Planes: Assuming that the ith feature and the jth feature lie on two chessboards, of which the normals are denoted by n_{ℓ_a} and n_{ℓ_b}, respectively, we have

$$M_{(i,j)k} \triangleq \begin{bmatrix} & & & \breve{M}_{ik} & & \\ & & & \breve{M}_{jk} & & \\ 0_{1\times3} & 0_{1\times3} & 0_{1\times3} & \cdots & n_{\ell_a}^{\mathrm{T}} & \cdots & 0_{1\times3} \\ 0_{1\times3} & 0_{1\times3} & 0_{1\times3} & \cdots & \cdots & n_{\ell_b}^{\mathrm{T}} & 0_{1\times3} \end{bmatrix}. \quad (54)$$

Then, we can obtain

$$M_{(i,j)k}N = \begin{bmatrix} 0_{4\times1} & 0_{4\times1} & 0_{4\times1} & 0_{4\times1} & 0_{4\times1} \\ n_{a_1} & n_{a_2} & n_{a_3} & n_{a_3}p_{i2} - n_{a_2}p_{i3} & -n_{a_3}p_{i1} + n_{a_1}p_{i3} \\ n_{b_1} & n_{b_2} & n_{b_3} & n_{b_3}p_{j2} - n_{b_2}p_{j3} & -n_{b_3}p_{j1} + n_{b_1}p_{j3} \end{bmatrix} \begin{matrix} \\ n_{a_2}p_{i1} - n_{a_1}p_{i2} \\ n_{b_2}p_{j1} - n_{b_1}p_{j2} \end{matrix} \quad (55)$$

where $n_{\ell_a}^{\mathrm{T}} \triangleq \begin{bmatrix} n_{a_1} & n_{a_2} & n_{a_3} \end{bmatrix}$ and $n_{\ell_b}^{\mathrm{T}} \triangleq \begin{bmatrix} n_{b_1} & n_{b_2} & n_{b_3} \end{bmatrix}$.

Next, we perform an elementary linear transformation on the above equation. That is, both sides of (55) are multiplied by matrix A_2

$$A_2 \triangleq \begin{bmatrix} 1 & -\frac{n_{a_2}}{n_{a_1}} & \frac{n_{a2}}{n_{a1}}\Lambda - \frac{n_{a3}}{n_{a1}} & 0 & 0 & 0 \\ 0 & 1 & -\Lambda & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (56)$$

$$\Lambda \triangleq \frac{n_{b3}n_{a1} - n_{b1}n_{a3}}{n_{b2}n_{a1} - n_{b1}n_{a2}} \quad (57)$$

$$M_{(i,j)k}NA_2 = \begin{bmatrix} 0_{4\times1} & 0_{4\times1} & 0_{4\times1} & 0_{4\times1} & 0_{4\times1} \\ n_{a_1} & 0 & 0 & n_{a_3}p_{i2} - n_{a_2}p_{i3} & -n_{a_3}p_{i1} + n_{a_1}p_{i3} \\ n_{b_1} & n_{b2} - \frac{n_{b1}n_{a2}}{n_{a1}} & 0 & n_{b_3}p_{j2} - n_{b_2}p_{j3} & -n_{b_3}p_{j1} + n_{b_1}p_{j3} \end{bmatrix} \begin{matrix} \\ n_{a_2}p_{i1} - n_{a_1}p_{i2} \\ n_{b_2}p_{j1} - n_{b_1}p_{j2} \end{matrix}. \quad (58)$$

As shown, the third column of NA_2 is the null-space of $M_{(i,j)k}$, and we denote it as N_2 and can get

$$M_{(i,j)k}N_2 = 0_{6\times1}. \quad (59)$$

Since this holds for any i, j, and k, we conclude that $MN_2 = 0_{3m(w+1)\times1}$. Therefore, when observing two chessboards, translation along the direction perpendicular to the normals of the two planes is unobservable.

3) Observation of Three Planes: Similar to the previous derivation process, we conclude that when three planes with non-collinear normals are observed, the rank deficiency is 0. That is to say, the calibration system is fully observable, at least three non-parallel chessboards are observed.

B. Observability With Point-to-Point Error

Our calibration method can use different calibration targets, for example, polygonal planar boards [34] or boxes [35]. For the data obtained by above calibration targets, the error measurement is the point-to-point error measurement y_i for feature i

$$y_i = {}^{\mathcal{L}}p_{c_i} - p_{\ell_i}. \quad (60)$$

Thus, the measurement Jacobian matrix Q_{ik} at time t_k for feature i is given by

$$Q_{ik} \triangleq \begin{bmatrix} H_{ik} \\ Y_i \end{bmatrix} = \begin{bmatrix} H_{\mathcal{C}_{ik}} & 0_{2\times3} & \cdots & H_{f_{ik}} & \cdots & 0_{2\times3} \\ 0_{3\times3} & 0_{3\times3} & \cdots & I_3 & \cdots & 0_{3\times3} \end{bmatrix} \quad (61)$$

where Y_i refers to the Jacobian matrix of y_i with respect to ξ_k and ith visual feature ${}^{\mathcal{L}}p_{c_i}$. What follows is the analysis about the minimal necessary conditions of the point pairs setting to solve the accurate six-DoF extrinsic calibration problem.

1) Observation of One Point: Assuming there is a point on the observed chessboard plane, namely the ith visual feature points, we have

$$M_{ik} \triangleq \begin{bmatrix} & & & \breve{M}_{ik} & & \\ 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & \cdots & I_3 & \cdots & 0_{3\times3} \end{bmatrix}. \quad (62)$$

Then, we can have

$$M_{ik}N = \begin{bmatrix} 0_{2\times1} & 0_{2\times1} & 0_{2\times1} & 0_{2\times1} & 0_{2\times1} & 0_{2\times1} \\ 1 & 0 & 0 & 0 & p_{i3} & -p_{i2} \\ 0 & 1 & 0 & -p_{i3} & 0 & p_{i1} \\ 0 & 0 & 1 & p_{i2} & -p_{i1} & 0 \end{bmatrix} \quad (63)$$

where ${}^{\mathcal{L}}p_{c_i} \triangleq \begin{bmatrix} p_{i1} & p_{i2} & p_{i3} \end{bmatrix}^{\mathrm{T}}$. Next, we perform an elementary linear transformation on the above equation. That is, both sides of (63) are multiplied by matrix A_3

$$A_3 \triangleq \begin{bmatrix} 1 & 0 & 0 & 0 & -p_{i3} & p_{i2} \\ 0 & 1 & 0 & p_{i3} & 0 & -p_{i1} \\ 0 & 0 & 1 & -p_{i2} & p_{i1} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (64)$$

$$M_{ik}NA_3 = \begin{bmatrix} 0_{2\times1} & 0_{2\times1} & 0_{2\times1} & 0_{2\times1} & 0_{2\times1} & 0_{2\times1} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \quad (65)$$

As shown, there are three columns of $M_{ik}NA_3$ that become all zeros. So the fourth, fifth, and sixth columns of NA_3 are the null-space of M_k, and we denote it as N_3 and can get

$$M_{ik}N_3 = 0_{5\times3}. \quad (66)$$

Since this holds for any i and k, we conclude that $MN_3 = 0_{5m(w+1)\times3}$, which means the rank deficiency is 3. Therefore, when observing only one point, any rotation is unobservable.

2) Observation of Two Points: The observability matrix of features i and j at time t_k is as follows:

$$M_{(i,j)k} \triangleq \begin{bmatrix} & & & \breve{M}_{ik} & & & \\ & & & \breve{M}_{jk} & & & \\ 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & \cdots & I_3 & \cdots & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & 0_{3\times3} & \cdots & \cdots & I_3 & 0_{3\times3} \end{bmatrix}. \quad (67)$$

Then, we can have

$$M_{(i,j)k}N = \begin{bmatrix} 0_{4\times1} & 0_{4\times1} & 0_{4\times1} & 0_{4\times1} & 0_{4\times1} & 0_{4\times1} \\ 1 & 0 & 0 & 0 & p_{i3} & -p_{i2} \\ 0 & 1 & 0 & -p_{i3} & 0 & p_{i1} \\ 0 & 0 & 1 & p_{i2} & -p_{i1} & 0 \\ 1 & 0 & 0 & 0 & p_{j3} & -p_{j2} \\ 0 & 1 & 0 & -p_{j3} & 0 & p_{j1} \\ 0 & 0 & 1 & p_{j2} & -p_{j1} & 0 \end{bmatrix} \tag{68}$$

where ${}^{\mathcal{L}}p_{c_j} \triangleq \begin{bmatrix} p_{j1} & p_{j2} & p_{j3} \end{bmatrix}^{\mathrm{T}}$. Next, we perform an elementary linear transformation on the above equation. That is, both sides of (68) are multiplied by matrix A_4

$$A_4 \triangleq \begin{bmatrix} 1 & 0 & 0 & \dfrac{-p_{i2}p_{j3}+p_{i3}p_{j2}}{p_{i1}-p_{j1}} & -p_{j3} & p_{j2} \\ 0 & 1 & 0 & p_{j3}-\dfrac{p_{j1}(p_{i2}-p_{j2})}{p_{i1}-p_{j1}} & 0 & -p_{j1} \\ 0 & 0 & 1 & -p_{j2}+\dfrac{p_{j1}(p_{i3}-p_{j3})}{p_{i1}-p_{j1}} & p_{j1} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \dfrac{p_{i2}-p_{j2}}{p_{i1}-p_{j1}} & 1 & 0 \\ 0 & 0 & 0 & \dfrac{p_{i3}-p_{j3}}{p_{i1}-p_{j1}} & 0 & 1 \end{bmatrix} \tag{69}$$

$$M_{(i,j)k}NA_4 = \begin{bmatrix} 0_{4\times1} & 0_{4\times1} & 0_{4\times1} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0_{4\times1} & 0_{4\times1} & 0_{4\times1} \\ 0 & p_{i3}-p_{j3} & -p_{i2}+p_{j2} \\ 0 & 0 & p_{i1}-p_{j1} \\ 0 & -p_{i1}+p_{j1} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{70}$$

As shown, the fourth column of NA_4 is the null-space of $M_{(i,j)k}$, and we denote it as N_4 and can get

$$M_{(i,j)k}N_4 = 0_{10\times1}. \tag{71}$$

Since this holds for any i, j, and k, we conclude that $MN_4 = 0_{5m(w+1)\times1}$. Therefore, when observing two points, one degree of freedom of the rotation is unobservable.

3) Observation of Three Points: Similar to the previous derivation process, we conclude that when three non-collinear points are observed, we can determine all the unknowns.

REFERENCES

[1] Y. Zhuang, N. Jiang, H. Hu, and F. Yan, "3-D-laser-based scene measurement and place recognition for mobile robots in dynamic indoor environments," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 2, pp. 438–450, Feb. 2013.

[2] S. Hussmann and T. Liepert, "Three-dimensional TOF robot vision system," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 1, pp. 141–146, Jan. 2009.

[3] S. Zhu and G. Yang, "Noncontact 3-D coordinate measurement of cross-cutting feature points on the surface of a large-scale workpiece based on the machine vision method," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 7, pp. 1874–1887, Jul. 2010.

[4] Y. Li, Y. F. Li, Q. L. Wang, D. Xu, and M. Tan, "Measurement and defect detection of the weld bead based on online vision inspection," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 7, pp. 1841–1849, Jul. 2010.

[5] L. Tang, Y. Wang, X. Ding, H. Yin, R. Xiong, and S. Huang, "Topological local-metric framework for mobile robots navigation: A long term perspective," *Auto. Robots*, vol. 43, no. 1, pp. 197–211, Jan. 2019.

[6] Q. Luo, H. Ma, Y. Wang, L. Tang, and R. Xiong, "3D-SSD: Learning hierarchical features from RGB-D images for amodal 3D object detection," 2017, *arXiv:1711.00238*. [Online]. Available: https://arxiv.org/abs/1711.00238

[7] X. Ding, Y. Wang, D. Li, L. Tang, H. Yin, and R. Xiong, "Laser map aided visual inertial localization in changing environment," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 4794–4801.

[8] B. Zheng, X. Huang, R. Ishikawa, T. Oishi, and K. Ikeuchi, "A new flying range sensor: Aerial scan in omni-directions," in *Proc. Int. Conf. 3D Vis. (3DV)*, Oct. 2015, pp. 623–631.

[9] J. Zhang, M. Kaess, and S. Singh, "A real-time method for depth enhanced visual odometry," *Auton. Robots*, vol. 41, no. 1, pp. 31–43, Jan. 2017.

[10] J. L. L. Galilea, J.-M. Lavest, C. A. L. Vazquez, A. G. Vicente, and B. I. Munoz, "Calibration of a high-accuracy 3-D coordinate measurement sensor based on laser beam and CMOS camera," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 9, pp. 3341–3346, Sep. 2009.

[11] R. Ishikawa, T. Oishi, and K. Ikeuchi, "LiDAR and camera calibration using motions estimated by sensor fusion odometry," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 7342–7349.

[12] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," in *Proc. IROS*, vol. 3, Sep./Oct. 2004, pp. 2301–2306.

[13] S. A. R. F, V. Fremont, and P. Bonnifait, "Extrinsic calibration between a multi-layer lidar and a camera," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst.*, Aug. 2008, pp. 214–219.

[14] A. Geiger, F. Moosmann, Ö. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2012, pp. 3936–3943.

[15] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.

[16] W. Wang, K. Sakurada, and N. Kawaguchi, "Reflectance intensity assisted automatic and accurate extrinsic calibration of 3D LiDAR and panoramic camera using a printed chessboard," *Remote Sens.*, vol. 9, no. 8, p. 851, Aug. 2017.

[17] F. M. Mirzaei, D. G. Kottas, and S. I. Roumeliotis, "3D LIDAR–camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization," *Int. J. Robot. Res.*, vol. 31, no. 4, pp. 452–467, Apr. 2012.

[18] Z. Taylor and J. Nieto, "Motion-based calibration of multimodal sensor extrinsics and timing offset estimation," *IEEE Trans. Robot.*, vol. 32, no. 5, pp. 1215–1229, Oct. 2016.

[19] Y. C. Shiu and S. Ahmad, "Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX=XB," *IEEE Trans. Robot. Autom.*, vol. 5, no. 1, pp. 16–29, Feb. 1989.

[20] J. D. Hol, T. B. Schön, and F. Gustafsson, "Modeling and calibration of inertial and vision sensors," *Int. J. Robot. Res.*, vol. 29, nos. 2–3, pp. 231–244, Feb. 2010.

[21] S. Zhan and R. Chung, "Use of LCD panel for calibrating structured-light-based range sensing system," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 11, pp. 2623–2630, Nov. 2008.

[22] J. Liu, Y. Li, and S. Chen, "Robust camera calibration by optimal localization of spatial control points," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 12, pp. 3076–3087, Dec. 2014.

[23] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robot. Autom. Mag.*, vol. 13, no. 2, pp. 99–110, Jun. 2006.

[24] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[25] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.

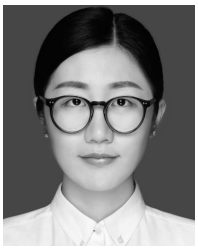
[26] G. Bradski, "The OpenCV library," *Dr Dobb's J. Softw. Tools*, vol. 25, pp. 120–125, Nov. 2000.
[27] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—A modern synthesis," in *Proc. Int. Workshop Vis. Algorithms*. Berlin, Germany: Springer, 1999, pp. 298–372.
[28] T. Rabbani, F. Van Den Heuvel, and G. Vosselmann, "Segmentation of point clouds using smoothness constraint," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 36, no. 5, pp. 248–253, 2006.
[29] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, 1975.
[30] P. J. Huber, "Robust estimation of a location parameter," in *Breakthroughs in Statistics*. New York, NY, USA: Springer, 1992, pp. 492–518.
[31] A. Bjõrck, *Numerical Methods for Least Squares Problems*, vol. 51. Philadelphia, PA, USA: SIAM, 1996.
[32] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G^2o: A general framework for graph optimization," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2011, pp. 3607–3613.
[33] M. Li and A. I. Mourikis, "High-precision, consistent EKF-based visual-inertial odometry," *Int. J. Robot. Res.*, vol. 32, no. 6, pp. 690–711, 2013.
[34] Y. Park, S. Yun, C. S. Won, K. Cho, K. Um, and S. Sim, "Calibration between color camera and 3D LIDAR instruments with a polygonal planar board," *Sensors*, vol. 14, no. 3, pp. 5333–5353, 2014.
[35] Z. Pusztai and L. Hajder, "Accurate calibration of LiDAR-camera systems using ordinary boxes," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Oct. 2017, pp. 394–402.
[36] E. Rohmer, S. P. N. Singh, and M. Freese, "V-REP: A versatile and scalable robot simulation framework," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 1321–1326.
[37] B. Efron, "Second thoughts on the bootstrap," *Stat. Sci.*, vol. 18, no. 2, pp. 135–140, 2003.
[38] J. Zhang and S. Singh, "Low-drift and real-time lidar odometry and mapping," *Auton. Robots*, vol. 41, no. 2, pp. 401–416, Feb. 2017.

Bo Fu received the B.S. degree from the Department of Control Science and Engineering, Shandong University, Jinan, China, in 2017. He is currently pursuing the Ph.D. degree with the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China.

His current research interests include multisensor calibration and sensor fusion.

Yue Wang (M'16) received the Ph.D. degree from the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2016.

He is currently a Lecturer with the Department of Control Science and Engineering, Zhejiang University. His current research interests include mobile robotics and robot perception.

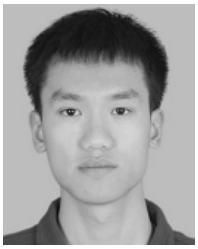


Xiaqing Ding received the B.S. degree from the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2016, where she is currently pursuing the Ph.D. degree.

Her current research interests include simultaneous localization and mapping and vision-based localization.

Yanmei Jiao received the B.S. degree from the Department of Computer Science and Control Engineering, Nankai University, Tianjin, China, in 2017. She is currently pursuing the Ph.D. degree with the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China.

Her current research interests include computer vision and vision-based localization.



Li Tang received the B.S. degree from the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2015, where he is currently pursuing the Ph.D. degree.

His current research interests include vision-based localization and autonomous navigation.

Rong Xiong (M'07) received the Ph.D. degree from the Department of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2009.

She is currently a Professor with the Department of Control Science and Engineering, Zhejiang University. Her current research interests include motion planning and simultaneous localization and mapping.