2020 IEEE International Conference on Robotics and Automation (ICRA)
31 May - 31 August, 2020. Paris, France

A Novel Calibration Method between a Camera and a 3D LiDAR with Infrared Images

Shoubin Chen, Jingbin Liu, Xinlian Liang, Shuming Zhang, Juha Hyyppä, and Ruizhi Chen

***Abstract*—Fusions of LiDARs (light detection and ranging) and cameras have been effectively and widely employed in the communities of autonomous vehicles, virtual reality and mobile mapping systems (MMS) for different purposes, such as localization, high definition map or simultaneous location and mapping. However, the extrinsic calibration between a camera and a 3D LiDAR is a fundamental prerequisite to guarantee its performance. Some previous methods are inaccurate, have calibration error that is several times the beam divergence, and often require special calibration objects, thereby limiting their ubiquitous use for calibration. To overcome these shortcomings, we propose a novel and high-accuracy method for the extrinsic calibration between a camera and a 3D LiDAR. Our approach relies on the infrared images from a camera with an infrared filter, and the 2D-3D corresponding points in a scene with the corners of a wall can be extracted to calculate the six extrinsic parameters. Experiments using the Velodyne VLP-16 sensor show that the method can achieve an extrinsic accuracy at the level of the beam divergence, which is fully analyzed and validated from two different aspects. Therefore, the calibration method in this paper is highly accurate, effective and does not require special complicated calibration objects; thus, it meets the requirements of practical applications.**

I. Introduction

Environmental perception is a key technology for autonomous vehicles, virtual reality and mobile mapping systems (MMS). In general, there are two categories of perception sensors that are mounted on platforms: (i) range sensors (e.g., LiDARs (light detection and ranging), radars, and sonars) and (ii) cameras (e.g., perspective, stereo, omnidirectional). Range sensors with active illumination provide accurate 3D position information even in night light. However, these 3D points are discrete and sparse with poor radiation or color information. The cameras are an inexpensive and well-studied solution for machine vision and provide high resolution color images. However, cameras can be used only in appropriate lighting conditions, and problems may occur due to occlusion, shadows or night light. As such, these methods compensate for the shortcomings of the other. Thus, combinations of LiDARs and cameras have been effectively employed in many application scenarios, including the Waymo Self-Driving Car, the Apollo Self-Driving Platform, NavVis

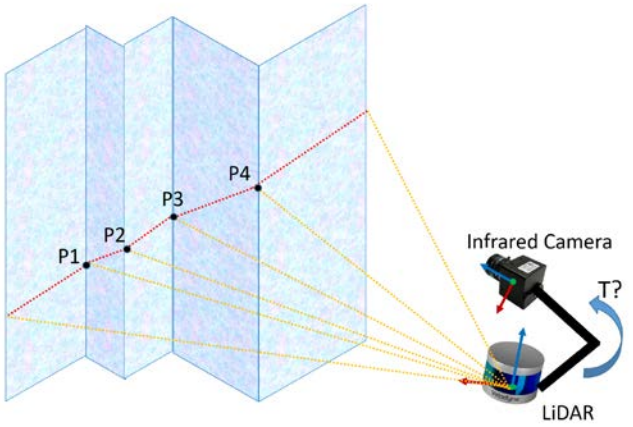

Figure 1. Observation of the corners of a wall by using a camera and a 3D LiDAR. The yellow lines present the field of view of the LiDAR. The red and V-shaped break lines draw and represent the laser's trajectories. For the break point P (black) on the break line, the 2D pixel and 3D coordinate can be respectively obtained from manually pinpricking the infrared image and processing the point cloud of the LiDAR data. After establishing a PnP problem with many 2D-3D/pixel-laser corresponding points, the transformation parameters T can be solved by using a numerical nonlinear optimization method.

3D Mapping Trolleys [1] and Google Cartographer backpacks [2], for different tasks, including localization, high definition maps (HD Maps), and simultaneous location and mapping (SLAM) [3-5].

The extrinsic calibration between a camera and a 3D LiDAR is a fundamental prerequisite for its applications, and it plays essential and vital roles in accomplishing high-accuracy localization, mapping, and SLAM. Extrinsic parameters represent the translation and rotation between the sensor frame and the base reference frame [6, 7] and are required to integrate and fuse all of the measurements from different sensors into a global reference frame. Moreover, the accuracies of mapping and SLAM are strongly influenced by the extrinsic errors and are very sensitive to the extrinsic calibration [8]. Particularly in the case of a long measurement range, small rotation errors can cause significant distortions in the mapping.

Calibrating a camera-LiDAR system has been studied for many years. These studies can be classified into two main

*This work was supported in part by the Natural Science Fund of China (Project No. 41874031), the Technology Innovation Program of Hubei Province (Project No. 2018AAA070), the Natural Science Fund of Hubei Province (Project No. 2018CFA007), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Project No.18YJCZH242) and the Humanities and Social Sciences Fund of Hubei Province (Project No. 18Q059).

Shoubin Chen, Jingbin Liu, Shuming Zhang and Ruizhi Chen is with State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: [shoubin.chen, jingbin.liu, ruizhi.chen]@whu.edu.cn).

Xinlian Liang, and Juha Hyyppä was with Department of Remote Sensing and Photogrammetry and the Center of Excellence in Laser Scanning Research, Finnish Geospatial Research Institute, Masala 02430, Finland (e-mail: xinlian.liang, juha.hyyppa@nls.fi).

categories. The first category needs some special geometrical calibration objects to obtain 2D-3D corresponding points (CPs) or geometric constraints between a single camera and a LiDAR. Rodriguez et al. [9] and Alismail et al. [10] used a black circle-based planar board to estimate the 3D coordinates of the center of a circle and the normal vector of the plane and built an usual absolute orientation problem. Mirzaei et al. [11] proposed a novel algorithm for jointly estimating 3D LIDAR–camera intrinsic and extrinsic calibration with a plane. Park et al. [12] used a white, homogeneous, planar polygonal board for calibration. The method with an arbitrary trihedron was published by Gong et al. [13], and a planar object containing four circular holes in front of a white background was utilized in Velas et al. [14]. In 2018, a new LiDAR-camera calibration was introduced that used ordinary cardboard boxes [15]. Some researches [16, 17] on autonomous vehicles (AVs) established the transformation between camera and LiDAR with geometric constraints in road scene. These 2D-3D geometric constraints included some errors because of sparse point clouds, thus influencing the calibration accuracy.

The second category is mainly based on stereo reconstruction and point cloud matching. The KITTI Calibration Toolbox that was published by Geiger et al. [18] is a classical and representative method. It requires several chessboard boards to reconstruct a stereo pair of two camera images from different positions, and then it matches the point clouds from the two cameras' stereo reconstruction and the one from the LiDAR. Hassanein et al. [19] developed a new automatic calibration method with two cameras and a well-textured object. The camera system needs to be calibrated a priori; thus, a sparse point cloud can be reconstructed from speeded up robust features (SURF). Then, this point cloud is registered to that of the LiDAR sensor by using the iterative closed point (ICP). The point cloud from the cameras' stereo reconstruction is unstable and inaccurate, thus influencing the accuracy of the matching and calibration.

However, there are some drawbacks in the abovementioned methods. First, the accuracies of these calibrations are too low because of the inaccuracy of the 2D-3D geometric constraints or the inaccuracy of the point cloud from the cameras' stereo reconstruction, thus resulting in obvious errors in the environmental perception. The results in [15] show that the translation and angular errors of the proposed techniques in [12], [15], [18] and [19] exceed 0.1 m and 0.5°, respectively, in the case of a virtual Velodyne HDL-64 with a fixed 0.03 Gaussian standard deviation and zero mean. Comparing that the beam divergence of each laser is 0.18° (horizontal) x 0.07° (vertical), the calibration error is several times the beam divergence. In addition, these previous methods usually special complicated calibration objects, such as chessboard boards, geometric boxes and so on, which limit the ubiquitous use of the calibration.

In this work, we try to overcome these shortcomings by proposing a novel and high-accuracy method for the extrinsic calibration between a camera and a 3D LiDAR. Our approach relies on infrared images from the camera. Infrared images record the laser footprints in a common view, which is a novelty and the most important idea in this paper. While a simple V-shaped object, such as the corner of a wall, is simultaneously imaged by the infrared camera and scanned by the LiDAR, the 2D-3D corresponding points can be respectively extracted from the infrared images and the point clouds to calculate the six extrinsic parameters. The experimental results with the Velodyne VLP-16 sensor show that the method can achieve an extrinsic accuracy at the level of the beam divergence (0.18°x 0.07°), which is fully analyzed and validated from two different aspects. There are two main advantages of the proposed method. The first is that it achieves high accuracy at the level of the beam divergence. The second is that the corners of walls are everywhere and easily available; thus, the method does not require special complicated calibration objects.

II. METHODS

The basic principle of the extrinsic calibration between a camera and a 3D LiDAR is to establish the point correspondences or geometric feature constraints between their observations and then to estimate the transformation matrix T between sensors in a global reference frame. As shown in Figure 1, assuming that the camera body frame is labeled as the global reference frame in our calibration system, our target is to obtain the transformations T between the reference frame and the LiDAR coordinate frame, including the rotation R and the translation t. The camera assembled with an infrared filter can capture the laser footprints as infrared images, although the laser trajectories are not visible to the naked eye. When the LiDAR scans a corner of the wall with a V-shaped structure, the V-shaped break lines (red lines) can draw and represent the laser trajectories. The 2D pixel of the break point P on the break line can be easily pinpricked out from the infrared image, and the corresponding 3D point can also be extracted with the point cloud processing of the LiDAR data. While changing the pose of the sensor system relative to the V-shaped scene, we can obtain a series of 2D-3D/pixel-laser corresponding points to establish the Perspective-n-Point problem—PnP in short—in computer vision, which can be solved by using a numerical nonlinear optimization method based on a least-square bundle adjustment (BA) with an initial estimation of the efficient PnP (EPnP) [20].

A. Processing Procedure

The extrinsic calibration process between a camera and a 3D LiDAR using infrared images is shown in Figure 2. There are two main processes: constructing 2D-3D corresponding points, and calibration. The input data includes the infrared images from the camera and the point cloud of the multiple-channel LiDAR in a scene with V-shaped objects. The result is the transformation matrix between these sensor body frames.

In the step of constructing 2D-3D corresponding points, the processing is different for the data from different sensor types. For the multiple-channel LiDAR, we need to preprocess the point cloud with an appropriate filter condition, segment each V-shape, and then extract the break points. The infrared images require the undistorted calibration and manual pinpricking of the break's pixel coordinates. Then, multiple pairs of CPs can be obtained from lots of observations.

Finally, the basic calibration model is built based on the geometric constraint of the multiple pairs of CPs. After this, the calibration equation is solved by using a least-squares

bundle adjustment solution, resulting in the calibration parameters. There are many ways to calculate the derivative. In this paper, the Lie algebraic perturbation model is used to calculate the derivative of the calibration equation.

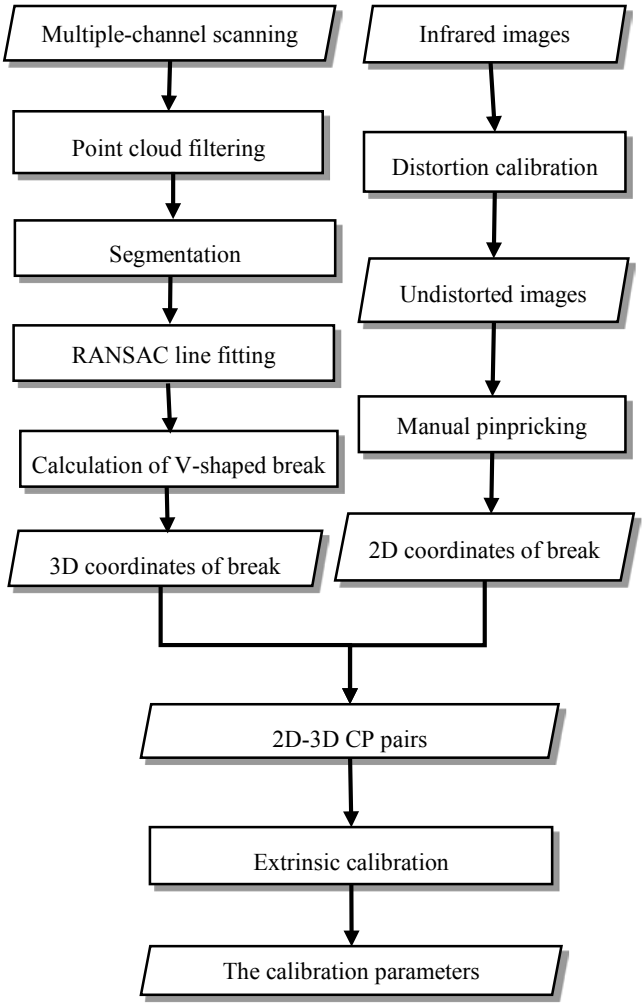

Figure 2. Extrinsic calibration process of a LiDAR-camera system using infrared images. Step 1: constructing 2D-3D corresponding points, where the processing is different for the data from different sensor types. Step 2: establishing and solving the extrinsic calibration model.

B. Constructing 2D-3D Corresponding Points

1) Filtering

Each scan from a LiDAR sensor is composed of hundreds or even thousands of 3D points. According to the size of the corner of the wall and the relative positions between the sensors and the scene, we can establish an appropriate coordinate interval as a filter condition to crop the area of interest w.r.t. the corner of the wall.

2) Segmentation

The multiple V-shaped trajectories are recorded in the point cloud data when the multichannel LiDAR such as the Velodyne VLP-16 scans the V-shaped object. By calculating the altitude angle of the LiDAR points, we can segment the point cloud from different channels. The points at the same altitude are labeled with the same channel number.

3) 3D Break Extract on the V-shaped Line

The random sample consensus (RANSAC) is an iterative and nondeterministic algorithm that is employed to estimate the parameters of a mathematical model from a set of observed data that contains certain outliers [21]. Assuming that there is a V-shaped trajectory within the observations from the LiDAR, we can detect the line model with the RANSAC algorithm.

The model parameters of the two lines will be obtained after RANSAC line fitting, and then the break of the V-shape that is composed by these two lines needs to be calculated. In three-dimensional Euclidean geometry, the intersection of the two lines can be an empty set, a point, or a line [22]. In the general case, two lines will not precisely intersect at a single point, and there is a distance between the two 3D lines. A least-squares solution can be derived that minimizes the sum of the perpendicular distances from the unique solution point to both lines. If the distance of the two lines is less than a threshold ε_d , the solution point is selected as the 3D break point. The threshold ε_d was set to 1 mm in this test.

4) 2D-3D Corresponding Points

When the LiDAR scans the corner of the wall, the images from the infrared camera record the laser trajectories. Unfortunately, low-cost pinhole cameras inherently come with significant distortions. Therefore, camera distortion calibration is a prerequisite and necessary step in order to extract accurate metric information from 2D undistorted images. A flexible and classical method is "Zhang's Camera Calibration" [23], which is provided in a popular MATLAB toolbox [24]. For original distorted infrared images, we use this toolbox to conduct the camera distortion calibration and then obtain undistorted images. The 2D pixel coordinate $\boldsymbol{u}(u,v)$ corresponding to the V-shape's 3D break point P(X,Y,Z) can be manually pinpricked from the undistorted infrared image and become a pair of CPs for the transformation matrix between the camera and the LIDAR. The correspondences can be quickly associated by hand because of the limited rings. While the camera doesn't observe the entire reflection of the LiDAR off the target, we can obtain the top/bottom ring number by blocking the LiDAR sensor partly.

For the multichannel LiDAR, we can obtain multiple pairs of CPs from one observation with one image and its corresponding point cloud. To avoid a poor or abnormal state of the calibration model, the collected CPs should be distributed as uniformly as possible in the common field of view of the camera and LiDAR. In the practical experiment, the size and number of corners of a wall are limited, and one scene cannot usually meet the uniform distribution condition of CPs. Therefore, more scenes using different sensor system poses relative to the wall should be collected, which will increase the uniformity of the distribution and the numbers of CPs.

C. Calibration with Bundle Adjustment

1) Calibration Model

The above processing has resulted in a set of CPs whose 3D coordinates and 2D image projections are known in the LiDAR

coordinate system and in the camera coordinate system, respectively. Then, it is necessary to determine the rotation and translation of a camera and LiDAR coordinate system. It also called the Perspective-n-Point problem—PnP for short—in computer vision. There are many ways to solve typical PnP problems, such as the P3P with just three points [25], the direct linear transformation (DLT) [26], the efficient PnP (EPnP) [20], and the UPnP [27]. In this paper, we solve the PnP problem by using a numerical nonlinear optimization method based on a least-squares bundle adjustment with an initial estimation from the EPnP.

If a set of pairs of CPs completely exist, according to the pinhole camera model, the geometric constraint can be written as (1) for CP pair i:

$$\forall i, u_i = \frac{1}{z_i'} K \bullet P'_i = \frac{1}{z_i'} K \bullet T \bullet P_i. \quad (1)$$

where K is the intrinsic matrix of the camera, and $P'_i(x'_i, y'_i, z'_i) = T \bullet P_i$, P'_i is the correspondence after transforming the point P_i with the transformation matrix T.

Then, the projection error is as follows (2):

$$e_i = u_i - \frac{1}{z_i'} K \bullet T \bullet P_i. \quad (2)$$

Equation (2) is the basic calibration model in this paper. The objective is to solve the rigid transformation matrix T that minimizes the projection error. The projection error minimization metric (3) is the objective optimization function of the least-squares bundle adjustment:

$$\min_{R,t} F = \frac{1}{2} \sum_{i=1}^{n} \left\| u_i - \frac{1}{z_i'} K \bullet T \bullet P_i \right\|_2^2. \quad (3)$$

2) Solution of Equations

For the convenience of taking the derivative of (2), the Lie algebra form ξ of the transformation matrix $T = \exp(\xi^\wedge)$ can be introduced as follows[8, 28, 29]:

$$T = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$$

$$\theta = \arccos \frac{tr(R) - 1}{2}, Ra = a$$

$$B = \frac{\sin\theta}{\theta} I + (1 - \frac{\sin\theta}{\theta}) aa^T + \frac{1 - \cos\theta}{\theta} a^\wedge \quad (4)$$

$$\rho = B^T t, \phi = \theta a$$

$$\xi^\wedge = \begin{bmatrix} \phi^\wedge & \rho \\ 0^T & 0 \end{bmatrix}, \xi = \begin{bmatrix} \rho \\ \phi \end{bmatrix}$$

where ξ is a six-dimensional vector corresponding to the six degrees of freedom of the transformation matrix T; the first dimension ρ is related to the translation but does not equal t; the last dimension ϕ is the Lie algebra or rotation vector responding to rotation matrix R; θ and a are respectively the rotation angle and rotation axis of the rotation vector ϕ in Rodrigues' formula; and B is an intermediate variable according to the Lie algebra and Lie group theory[30].

Then, using ξ to represent the pose, taking partial derivative and linearization for Equation (3), the error equation can be rewritten as follows:

$$e_i = u_i - \frac{1}{z_i'} K \bullet \exp(\xi^\wedge) \bullet P_i = \frac{\partial e_i}{\partial \xi} d\xi - l_i. \quad (5)$$

where $-l_i$ is the constant vector in a Taylor expansion, and the gradient is the first-order term.

Based on the Lie algebraic perturbation model, taking the derivative of (5) with respect to ξ can be expressed as follows:

$$\frac{\partial e_i}{\partial \delta\xi} = \frac{\partial e_i}{\partial P_i'} \bullet \frac{\partial P_i'}{\partial \delta\xi} = \frac{\partial e_i}{\partial P_i'} \bullet \frac{\partial (\exp(\xi^\wedge) \bullet P_i)}{\partial \delta\xi}$$

$$= - \begin{bmatrix} \frac{f_x}{z_i'} & 0 & -\frac{f_x x_i'}{z_i'^2} \\ 0 & \frac{f_y}{z_i'} & -\frac{f_x y_i'}{z_i'^2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & z_i' & -y_i' \\ 0 & 1 & 0 & -z_i' & 0 & x_i' \\ 0 & 0 & 1 & y_i' & -x_i' & 0 \end{bmatrix}$$

$$= - \begin{bmatrix} \frac{f_x}{z_i'} & 0 & -\frac{f_x x_i'}{z_i'^2} & -\frac{f_x x_i' y_i'}{z_i'^2} & f_x + \frac{f_x x_i'^2}{z_i'^2} & -\frac{f_y y_i'}{z_i'} \\ 0 & \frac{f_y}{z_i'} & -\frac{f_x y_i'}{z_i'^2} & -f_y - \frac{f_y y_i'^2}{z_i'^2} & \frac{f_y x_i' y_i'}{z_i'^2} & \frac{f_y x_i'}{z_i'} \end{bmatrix} = J_i \quad (6)$$

where $\delta\xi = X = [\delta\rho, \delta\phi]^T$ is the correction for the calibration parameters, and $P_i' = \exp(\xi^\wedge) \bullet P = (x_i', y_i', z_i')^T$.

Considering different CP pairs, the error equation can be written as follows:

$$\begin{cases} e_1 = \frac{\partial e_1}{\partial \xi} d\xi - l_1 \\ \cdots \\ e_n = \frac{\partial e_n}{\partial \xi} d\xi - l_n \end{cases} \quad (7)$$

For convenience, (7) can be simplified as (8):

$$e = J \bullet d\xi - L \quad (8)$$

Equation (8) is the basic error equation, which can be solved through nonlinear iterations of the Levenberg-Marquardt method by minimizing e towards zero.

$$\xi \leftarrow \xi - \left(J^T J + \lambda \mathrm{diag}(J^T J) \right)^{-1} \left(J^T L \right) \quad (9)$$

where λ is a factor that is determined by the Levenberg-Marquardt method.

III. Experiments and Analysis

A. Data Introduction

The VLP-16 LiDAR from the Silicon Valley-based manufacture Velodyne is an active sensor that emits near-

infrared light at 903 nm. The Velodyne VLP-16 performance parameters are shown in TABLE I.

TABLE I
THE VELODYNE VLP-16 PERFORMANCE PARAMETER TABLE

Item	Velodyne VLP-16
Announced	November 2014
Measurement Principle	Time of Flight
Channels	16
Light Source	903 nm
Range	100 m
Accuracy	±3 cm (Typical)
Beam Size @ Screen	12.7 mm (Horizontal) x 9.5 mm (Vertical)
Beam Divergence	0.18° (Horizontal) x 0.07° (Vertical)
Field of View (Horizontal)	360°
Field of View (Vertical)	+15.0° to -15.0° (30°)
Angular Resolution (Horizontal)	0.1° – 0.4°
Angular Resolution (Vertical)	2.0°
Rotation Rate	5 Hz – 20 Hz

The camera in this paper is assembled with a near infrared absorption filter, which only allow light above 850 mm to pass. Therefore, the footprint of the laser with the 903 nm light source can be recorded on the infrared images. It is noted that the filter has on effect on the geometric performance of the camera. Therefore, the geometric calibration parameters are invariable before and after the filter assembly. Sometimes we need to turn off the light in the room to reduce the interference of visible light. The angular resolution of the camera is approximately 0.002° from the calculations, and the field of view is approximately 80° in the horizontal and approximately 60° in the vertical directions.

To sufficiently validate the proposed method in the experiment, several scenes between a camera and a 3D LiDAR observing the corners of the wall were collected by moving the pose of the sensor system. Figure 3 shows some detailed information of the experimental scenes covering the corners of the wall.

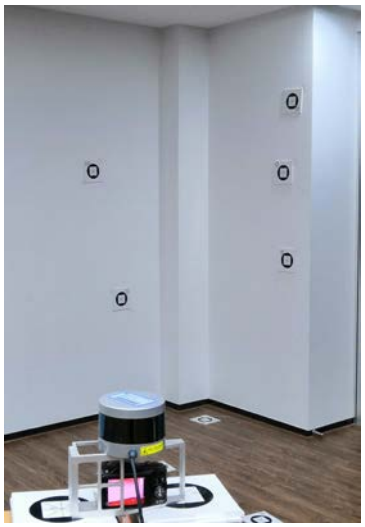

Figure 3. A LiDAR-camera system observing the corners of the wall. Several scenes were collected by changing the pose of the sensor system.

B. Calibration and Accuracy Analysis

The external parameters calculated according to the calibration model are translated into an equal and simpler interpretation style consisting of the Euler extrinsic rotation angles of yaw-pitch-roll (Y-P-R) around a fixed Z-Y-X axis and the translation. The values of the calibration parameters are shown in TABLE II, and the frame configurations are shown in Figure 4.

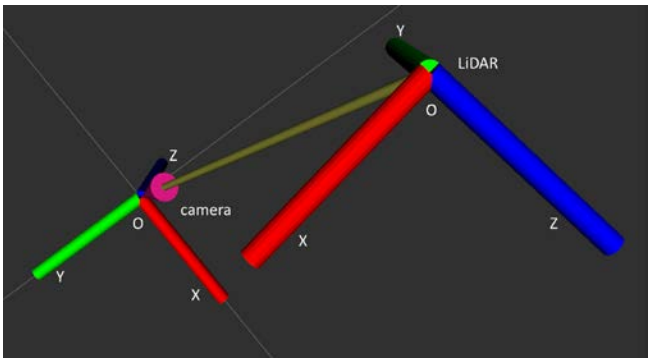

Figure 4. The laser frame configurations corresponding to the calculated transformation parameters, which is defined by a six dimensional vector, including the Euler extrinsic rotation angles of the YPR around a fixed Z-Y-X axis and the translation.

TABLE II
RESULTING CALIBRATION PARAMETERS

Calibration Parameters	Rotation angles (°)			Translation (cm)		
	Y	P	R	X	Y	Z
Values	79.47	0.08	90.72	-0.250	-8.235	-10.589
Standard deviation	0.053	0.058	0.049	0.438	0.378	0.093

At this point, we have estimated the rigid transformation between the camera and the laser range finder frame. Although the ground truth for the relative pose of the sensors can be used to compare the estimated transformation with the exact transformation among the sensors, requiring the availability of the ground truth complicates an absolute evaluation of the method. Considering the difficulty that is involved in precisely measuring the correct pose between a pair of sensors, we cannot guarantee high precision even if the obtained calibration is correct. This problem is particularly difficult with rotations (i.e., it is possible to have a fairly reasonable evaluation of the translation with simple measurements using a tape measure), which is critical since small errors in a rotating system may result in large errors within the final registered data[8]. In this paper, the accuracy of the calibration results can be estimated from two aspects: 1) the standard deviation of the estimated parameters, 2) the reprojection error of the control points.

1) Standard Deviation of Estimated Parameters

First, analyzing the standard deviation of the estimated parameters with the covariance of the solution is one way to assess the quality of the solution that is returned by a nonlinear least squares solver. Under the assumption that measurement errors are normally distributed, the standard deviation of the i-th estimated parameter is defined as follows [31]:

$$\sigma_0 = \sqrt{\frac{1}{2n-m} \bullet \sum_{i=1}^{2n} e_i^2}, \quad Cov = \left(J^T \bullet J\right)^{-1}, \quad \sigma_i = \sigma_0 \bullet \sqrt{Cov(i,i)} \tag{10}$$

where σ_0 represents the root mean squared error that is defined by n, which is the number of control points; m is the number of estimated parameters (3 rotations and 3 translations); and e is the residual of the nonlinear objective function. σ_0^2 is called the variance factor and is the unit weight variance or the prior variance factor. J represents the Jacobian matrix of the last LM-algorithm iteration, and Cov is the covariance matrix of the estimated parameters.

In our experiment, the standard deviations of the Euler extrinsic rotation angles of the yaw-pitch-roll (YPR) around a fixed Z-Y-X axis and the translation are d the translation are listed in TABLE II. It shows that the accuracy of the rotation and translation are approximately 0.05° and are better than 0.01 m, respectively. The method is obviously more accurate than the proposed techniques in [12], [15], [18] and [19], which translation and angular errors exceed 0.1 m and 0.5°, for a virtual Velodyne HDL-64 with a fixed 0.03 Gaussian standard deviation and zero mean.

2) Reprojection Error of Control Points

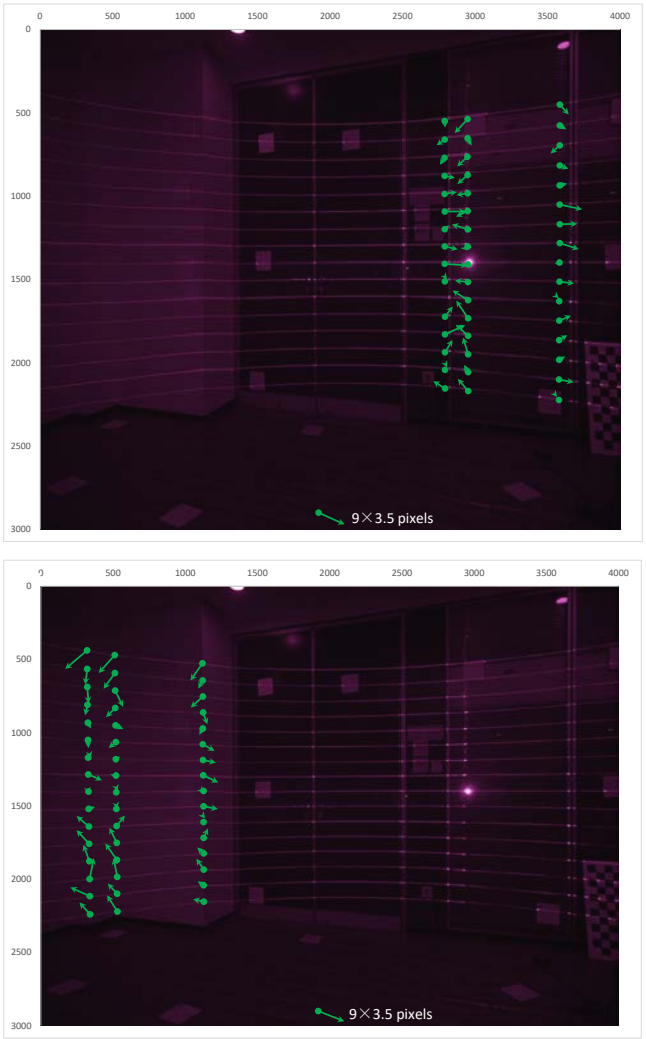

Figure 5. The distribution and residual errors of the control points in scenes 01 (top) and 02 (bottom). The red points represent the 2D pixel coordinates from infrared images, and the red arrows represent the magnitude and direction of the residual errors.

TABLE III
THE REPROJECTION ERROR OF CONTROL POINTS

Scene	CP Pairs	region	RMSE (Pixel)			Mean (Pixel)		
			x	y	$\sqrt{x^2+y^2}$	x	y	$\sqrt{x^2+y^2}$
01	47	Right	3.98	2.18	1.92	0.60	-0.22	4.07
02	48	Left	3.22	3.76	2.24	-0.78	0.13	4.37
all	95	total	3.60	3.07	2.08	-0.10	-0.04	4.22

Second, the reprojection error of the control points is one indication of the quality of the calibration process. Once the calibration is computed, the 3D control point is reprojected on the infrared images where it appears. A geometric error corresponding to the image distance between a projected point and a measured one is the reprojection error, which is defined in (2). This error depends on the quality of the calibration (position and orientation), as well as on the quality of the marked point on the images (the position and zoom level at which the point is marked).

The angular resolution of the camera is 0.02°, and the beam divergence of each laser is 0.18° (horizontal) x 0.07° (vertical). Therefore, a beam will take up 9 x 3.5 pixels on the infrared images. In TABLE III, the mean values of the residuals along each coordinate axis are small and close to zero, and the RMSE values of the residuals are approximately 4 x 3.5 pixels at the level of beam divergence. Figure 5 shows the residual distribution of the control points in scenes 01 and 02, which indicates that there is no systematic error after calibration. Therefore, the extrinsic orientation accuracy after calibration is highly accurate at the level of beam divergence and is effective for further application.

IV. CONCLUSION

This paper presents a novel methodology for calibrating the extrinsic parameters between a camera and a 3D LiDAR using infrared images. The proposed method provides high-accuracy solutions. Moreover, the method does not require complicated calibration objects and simply uses the corners of walls, which are ubiquitous and widely available. Experiments are conducted using a sensor system consisting of a Velodyne VLP-16 LiDAR and a camera that is assembled with an infrared filter, and the system is pointed towards to the corners of the wall. The appropriate distribution and high accuracy of the 2D-3D corresponding points bring the fine performance of the calibration. Two different accuracy assessment methods consistently and sufficiently reflect that the calibration is highly accurate at the level of the beam divergence.

However, there are some weaknesses in this method. Infrared photography is easily affected by visible light, so the method can not work outdoors. Second, the IR cameras are not typically used on robots and AVs, and cameras on robots and AVs are not easy to install an IR filter. In the future, we plan to conduct some research to overcome these limitations and increase its availability.

REFERENCES

[1] Navvis. (20 April 2018). *Digitizing indoors—NavVis*. Available: http://www.navvis.com

[2] Google. (20 April 2018). *Google Cartographer backpack*. Available: https://opensource.google.com/projects/cartographer

[3] J. Zhang and S. Singh, "LOAM : Lidar Odometry and Mapping in real-time," in *Robotics: Science and Systems Conference*, 2014.
[4] V. V. Lehtola, J. P. Virtanen, A. Kukko, H. Kaartinen, and H. Hyyppä, "Localization of mobile laser scanner using classical mechanics," *Isprs Journal of Photogrammetry & Remote Sensing,* vol. 99, pp. 25-29, 2015.
[5] V. V. Lehtola, H. Kaartinen, A. Nüchter, R. Kaijaluoto, A. Kukko, P. Litkey*, et al.*, "Comparison of the Selected State-Of-The-Art 3D Indoor Scanning and Point Cloud Generation Methods," *Remote Sensing,* vol. 9, p. 796, 2017.
[6] J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel*, et al.*, "Towards fully autonomous driving: Systems and algorithms," in *Intelligent Vehicles Symposium*, 2011, pp. 163-168.
[7] D. Yin, J. Liu, T. Wu, K. Liu, J. Hyyppä, and R. Chen, "Extrinsic Calibration of 2D Laser Rangefinders Using an Existing Cuboid-Shaped Corridor as the Reference," *Sensors,* vol. 18, p. 4371, 2018.
[8] S. Chen, J. Liu, T. Wu, W. Huang, K. Liu, D. Yin*, et al.*, "Extrinsic Calibration of 2D Laser Rangefinders Based on a Mobile Sphere," *Remote Sensing,* vol. 10, p. 1176, 2018.
[9] A. R. F. S., V. Fremont, and P. Bonnifait, "Extrinsic calibration between a multi-layer lidar and a camera," in *IEEE International Conference on Multisensor Fusion & Integration for Intelligent Systems*, 2008.
[10] H. S. Alismail, D. L. Baker, and B. Browning, "Automatic Calibration of a Range Sensor and Camera System," in *Second International Conference on 3d Imaging*, 2012.
[11] F. M. Mirzaei, D. G. Kottas, and S. I. Roumeliotis, "3D LIDAR–camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization," *The International Journal of Robotics Research,* vol. 31, pp. 452-467, 2012.
[12] P. Yoonsu, Y. Seokmin, W. Chee Sun, C. Kyungeun, U. Kyhyun, and S. Sungdae, "Calibration between color camera and 3D LIDAR instruments with a polygonal planar board," *Sensors,* vol. 14, pp. 5333-5353, 2014.
[13] G. Xiaojin, L. Ying, and L. Jilin, "3D LIDAR-camera extrinsic calibration using an arbitrary trihedron," *Sensors,* vol. 13, pp. 1902-1918, 2013.
[14] M. Veľas, M. Španěl, Z. Materna, and A. Herout, "Calibration of RGB camera with velodyne LiDAR," 2014.
[15] Z. Pusztai, I. Eichhardt, and L. Hajder, "Accurate Calibration of Multi-LiDAR-Multi-Camera Systems," *Sensors,* vol. 18, pp. 2139-, 2018.
[16] J. Jeong, Y. Cho, and A. Kim, "The Road is Enough! Extrinsic Calibration of Non-overlapping Stereo Camera and LiDAR using Road Information," *IEEE Robotics and Automation Letters,* vol. 4, pp. 2831-2838, 2019.
[17] J. Jiang, P. Xue, S. Chen, Z. Liu, X. Zhang, and N. Zheng, "Line feature based extrinsic calibration of LiDAR and camera," in *2018 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, 2018, pp. 1-6.
[18] A. Geiger, F. Moosmann, O. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *IEEE International Conference on Robotics & Automation*, 2012, pp. 3936-3943.
[19] M. Hassanein, A. Moussa, and N. El-Sheimy, *A NEW AUTOMATIC SYSTEM CALIBRATION OF MULTI-CAMERAS AND LIDAR SENSORS* vol. XLI-B1, 2016.
[20] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision,* vol. 81, p. 155, 2009.
[21] M. A. Fischler and R. C. Bolles, *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*: ACM, 1981.
[22] F. Antonio, "Faster line segment intersection," in *Graphics Gems III (IBM Version)*, ed: Elsevier, 1992, pp. 199-202.
[23] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 22, 2000.
[24] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A toolbox for easily calibrating omnidirectional cameras," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 5695-5701.
[25] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE transactions on pattern analysis and machine intelligence,* vol. 25, pp. 930-943, 2003.
[26] Y. Abdel-Aziz, H. Karara, and M. Hauck, "Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry," *Photogrammetric Engineering & Remote Sensing,* vol. 81, pp. 103-107, 2015.
[27] A. Penate-Sanchez, J. Andrade-Cetto, and F. Moreno-Noguer, "Exhaustive linearization for robust camera pose and focal length estimation," *IEEE transactions on pattern analysis and machine intelligence,* vol. 35, pp. 2387-2400, 2013.
[28] S. Hassani, *Representation of Lie Groups and Lie Algebras*: Springer International Publishing, 2013.
[29] T. D. Barfoot, "State estimation for robotics: A matrix lie group approach," 2017.
[30] V. S. Varadarajan, "Lie groups, Lie algebras, and their representations," vol. 659, pp. xiv,351, 1974.
[31] V. Fremont and P. Bonnifait, "Extrinsic calibration between a multi-layer lidar and a camera," in *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2008, pp. 214-219.