Extrinsic Calibration of Camera to LIDAR using a Differentiable Checkerboard Model

Lanke Frank Tarimo Fu Nived Chebrolu Maurice Fallon

***Abstract*— Multi-modal sensing often involves determining correspondences between each domain's signals, which in turn depends on the accurate extrinsic calibration of the sensors. Challengingly, the camera-LIDAR sensor modalities are quite dissimilar and the narrow field of view of most commercial LIDARs means that they observe only a partial view of the camera frustum. We present a framework for extrinsic calibration of a camera and a LIDAR using only a simple off-the-shelf checkerboard. It is designed to operate even when the LIDAR observes a significantly truncated portion of the checkerboard. Current state-of-the-art methods often require bespoke manufactured markers or full observation of the entire checkerboard in both camera and LIDAR data which is prohibitive. By contrast, our novel algorithm directly aligns the LIDAR intensity pattern to the camera-detected checkerboard pattern using our differentiable formulation. The key step for achieving accurate extrinsics estimation is the use of the spatial derivatives provided by the differentiable checkerboard pattern, and jointly optimizing over all views. In our experiments, we achieve calibration accuracy in the order of 2-4 mm and demonstrate a 30% error reduction compared to state-of-the-art approaches. We are able to achieve this improvement while using only partial LIDAR views of the checkerboard that allows for a simpler data capture process. We also demonstrate the generalizability of our approach to different combinations of LIDARs and cameras with varying sparsity patterns and noise levels.**

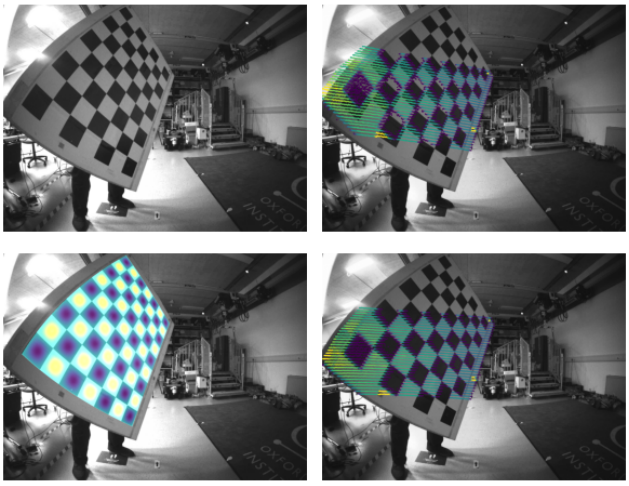


Fig. 1: An example setup where the LIDAR has a significantly narrower field of view than the camera. Significant portions of the checkerboard (and crucially, their edges) are not scanned by the LIDAR. Top-right shows the extrinsics estimate from point-to-plane matching (an existing approach). Bottom-left shows the warped differentiable checkerboard in the image space, generated using the homography of the detected checkerboard corners in the image. Bottom-right shows the results of our method after alignment.

I. Introduction

Extrinsic sensor calibration is the process of determining the relative spatial transforms between different sensors. For robots that employ multiple sensors, accurate extrinsic calibration is crucial when inferring correspondences between input signals. Many existing state-of-the-art approaches for camera-LIDAR extrinsic calibration rely on specially machined calibration targets to match the LIDAR-detected geometric features to the camera-detected textural features[1], [15]. Other approaches that use off-the-shelf targets rely on matching simpler LIDAR-detected target geometries — e.g. planes [5], [13] and edges [17] to their corresponding camera image counterparts. The efficient formulation of Zhou et al. [17] allowed for obtaining state-of-the-art calibration accuracy requiring fewer LIDAR calibration target detections compared to previous approaches [5], [13].

In this work, we present a target-based framework for camera-LIDAR extrinsic calibration that jointly utilizes the intensity pattern observed by the LIDAR over all views, in addition to the planar geometry of the calibration target. Using a differentiable approximation to the checkerboard intensity pattern, our approach corrects the visual misalignment between the projected LIDAR intensity pattern and the camera-detected checker pattern to fine-tune the extrinsic parameters.

Existing approaches using plane or edge features typically require that all edges of the calibration target are visible in the LIDAR scan for a data sample to be useful for estimating extrinsic parameters. In practice, this requirement can be limiting as the LIDAR's field of view is usually much narrower than the camera's. This restricts the valid data samples to be collected from further away, where the camera's observations are less accurate. In our approach, we benefit from the continuous nature of the differentiable formulation and are not constrained by the need to observe the entire checkerboard in each LIDAR scan. Our differen-

Support for this work has been provided by the Horizon Europe project DigiForest (101070405) and a Royal Society University Research Fellowship (M. Fallon). This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Oxford Robotics Institute, Dept. of Engineering Science, University of Oxford, UK. {fu, nived, mfallon}@robots.ox.ac.uk

tiable formulation enables us to incorporate information from even partial views of the checkerboard in the LIDAR scan. This approach is especially useful for sensor setups involving wide field of view cameras where the camera frustum is considerably larger than that of the LIDAR.

In summary, our main contributions are:

- A novel optimization formulation based on a differentiable checkerboard intensity pattern model that can be used to fine-tune the in-plane SE(2) pose of a checkerboard in a LIDAR scan (a parameter unobservable in plane-to-plane matching)
- Camera-LIDAR extrinsic calibration pipeline that achieves state-of-the-art accuracy using partial LIDAR views of the calibration target allowing for a convenient data collection procedure.

We evaluate our performance by estimating the stereo-baseline width of a calibrated, machined stereo camera pair. Our approach achieves a calibration accuracy within 4 mm of the ground truth while only using partial views of the checkerboard in the LIDAR. Additionally, we demonstrate the versatility of our method by evaluating the calibration results on multiple camera-LIDAR setups including on a sparse 16-channel and a dense 128-channel scanner with a wide-angle fisheye and mild-distortion camera respectively.

II. Related Work

In broad terms, camera-LIDAR extrinsic calibration can be classified into three branches: target-based, target-less and deep-learned methods. Our method falls under the target-based category where the calibration is carried out assuming that each sensor perceives a predefined target, and is typically performed in a controlled environment.

Target-based Calibration Early approaches to camera-LIDAR calibration required manually-labelled correspondences between the camera and LIDAR signals [11], [16], [13]. This is time-consuming and couples the calibration accuracy with the expertise of the human operator. Subsequent work automated the target detection aspect of calibration by using checkerboards detectable in both camera and LIDAR signals which minimizes the impact of operator error.

Several existing methods to automatic target detection require precisely machined geometric patterns in alignment with the visual checker pattern on the calibration target [1], [15]. Whereas, other methods work with simple off-the-shelf printed checkerboard targets [5], [17], [14]. Zhou *et al.* [17] proposed an efficient algorithm using edge geometry to estimate more pose constraints per checkerboard detection than plane-to-plane matching [5]. Wang et al. [14] additionally used intensity information, on top of the edge information to estimate extrinsics. Limitingly, the feature detection step in both theses approaches require the full checkerboard to be visible in each LIDAR scan which limits the range of operation. Additionally, if the printed checker pattern doesn't span the entire physical checkerboard, the offsets between the checker pattern and the edges of the board need to be precisely measured – incurring more work.

Similar to [5], [17], [14], our approach proposes an automated procedure using a standard checkerboard pattern. Unlike plane-to-plane matching by Geiger et al. [5], our differentiable checkerboard formulation enables our approach to fine-tune the in-plane misalignment using direct alignment with the LIDAR intensity pattern. At the same time, our direct alignment formulation relaxes the need to observe the entire calibration target in each LIDAR scan, which is required in [17], [14]. The ability to perform calibration using partial LIDAR views of the target significantly simplifies the data collection procedure, and improves the efficiency of the calibration process.

Classical Target-less Calibration For completeness, we review more general calibration methods which do not require targets and can be used *in the field*. These methods target the automatic fine-tuning of camera-LIDAR extrinsic parameters after deployment but can benefit from an improved pre-deployment target-based calibration method such as ours.

The method of Levinson and Thrun [8] samples different poses around an initial estimate to maximize the correlation between geometric discontinuities in the LIDAR signal and image intensity discontinuities in the camera to perform pose refinement. This assumption that image intensity discontinuities correspond to depth discontinuities in the LIDAR scan doesn't hold for the case of a simple, printed checkerboard, where the checker pattern falls on a single plane. Inspired by Pandey et al. [10] and Scott et al. [12] our method also utilizes the intensity pattern of the LIDAR scan. While [10], [12] formulate the calibration problem as one of matching intensity histograms (an optimization which requires computing numerical gradients), our differentiable formulation of the checkboard intensity pattern provides spatial derivatives by construction, allowing us to efficiently compute the camera-LIDAR pose alignment.

Deep-learning Based Extrinsic Calibration Another arc of research is the development of deep-learning based methods such as [6], [9], [7]. These approaches use images and LIDAR scans of natural scenes and do not require any calibration targets. Still, their training is supervised by ground truth data generated using target-based calibration procedures. The above learned methods [6], [9], [7] were all trained and evaluated on The KITTI dataset[4] which was calibrated using the target-based method of Geiger *et al.* [5]. In short, many deep-learned models still rely on target-based calibration methods either for training data or for ground truth verification, and can hence, benefit from an approach such as ours.

III. Our Approach

A. Problem Formulation

Given a set of 3D points $\mathbf{P}_L \in \mathbb{R}^3$ represented in the LIDAR reference frame (denoted by the subscript L) and a set of camera images captured synchronously showing views of a checkerboard at different poses, the task is to determine the extrinsic parameters of the transform between the camera and the LIDAR sensor $\mathbf{R}_L^C \in \mathbf{SO}(3), \mathbf{t}_L^C \in \mathbb{R}^3$.

We follow the convention that the coordinates of the 3D LIDAR points, in the camera reference frame, are given by:

$$\mathbf{P}_C = \mathbf{R}_L^C \mathbf{P}_L + \mathbf{t}_L^C \tag{1}$$

for convenience, we also use $\mathbf{T}_L^C \in \mathbf{SE}(3)$ to denote the rigid body transform formed by $\mathbf{R}_L^C$ and $\mathbf{t}_L^C$.

B. Calibration Target Design and Detection

In common with [17], our framework requires only a simple checkerboard calibration target to generate data to align the camera and the LIDAR poses. In our experiments we used the 100 cm x 90 cm checkerboard shown in Fig. 1.

We detect the target in the camera frame using a standard checkerboard detection pipeline from OpenCV [2]. From the LIDAR point cloud, we isolate points that lie on the target by segmenting the plane nearest to the estimated checkerboard pose. The checkerboard pose estimate (in the LIDAR frame) is obtained by transforming the target detection from the camera to the LIDAR reference frame using the initial estimate of the extrinsic parameters. We retain these points for subsequent steps.

C. Stage 1: Initial Solve using Point-to-Plane Alignment

As a first step, we solve for the extrinsic parameters using point-to-plane fitting of the checkerboard detections discussed in Sec. III-B. We minimize the point-to-plane residual $r(\mathbf{R}_L^C, \mathbf{t}_L^C)$ as follows:

$$\hat{\mathbf{R}}_L^C, \hat{\mathbf{t}}_L^C = \min_{\mathbf{R}_L^C, \mathbf{t}_L^C} \sum_i \sum_j^{|\mathbf{P}_{Li}|} r_{ij}^2(\mathbf{R}_L^C, \mathbf{t}_L^C) \tag{2}$$

$$r_{ij}(\mathbf{R}_L^C, \mathbf{t}_L^C) = \mathbf{n}_{C_i} \cdot (\mathbf{R}_L^C \mathbf{P}_{Li,j} + \mathbf{t}_L^C) - d_{C_i} \tag{3}$$

where i iterates through the set of checkerboard detections in the camera images, $|\mathbf{P}_{Li}|$ is the cardinality of LIDAR points segmented for the i-th checkerboard detection in the camera.

Note that $\mathbf{n}_{C_i} \in \left\{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\| = 1\right\}$ along with $d_{C_i} \in \mathbb{R}$ define the plane equation of the i-th checkerboard in the camera reference frame, in the form: $\mathbf{n}_{C_i} \cdot \mathbf{P}_C - d_{C_i} = 0$ for the set of points $\mathbf{P}_C \in \mathbb{R}^3$ that are on the checkerboard plane in the camera reference frame. Equation 2 is optimized using the Levenberg-Marquardt algorithm, where we parameterize the pose using the Lie-algebra formulation: $\xi^T = [\mathbf{t}, \omega]^T$ with $\xi \in \mathfrak{se}(3), \mathbf{t} \in \mathbb{R}^3, \omega \in \mathfrak{so}(3)$. The Jacobian at each observed LIDAR point $\mathbf{P}_{Li,j}$ is given by:

$$\frac{\partial r_{i,j}(\mathbf{R}_L^C, \mathbf{t}_L^C)}{\partial \xi} = \mathbf{n}_{\mathbf{C}i} \cdot \left[\mathbf{I} | \lfloor \hat{\mathbf{P}}_{C_{i,j}} \rfloor_\times\right] \tag{4}$$

where,

$$\hat{\mathbf{P}}_{C_{i,j}} = \mathbf{exp}(\xi) \mathbf{P}_{L_{i,j}} \tag{5}$$

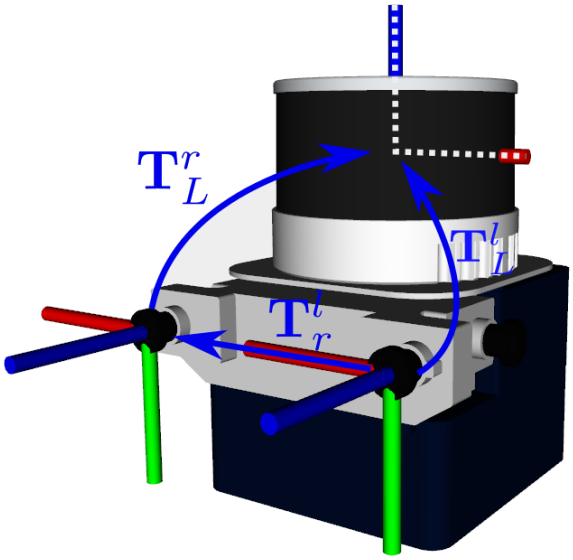


Fig. 2: Our method estimates the transforms from the LIDAR to the left and right cameras $\mathbf{T}_L^r$ and $\mathbf{T}_L^l$ respectively. To evaluate the accuracy of the calibration process, we compare their combination, i.e. $\mathbf{T}_L^l.\mathbf{T}_r^L$, against the transform between the two cameras $\mathbf{T}_r^l$ estimated by traditional stereo camera calibration in Sec . IV.

D. Stage 2: Refinement using the Differentiable Checkboard Model

Depending on the spatial distribution of the data samples, using the point-to-plane alignment alone can result in sub-optimal calibration parameters as depicted in Fig. 1. This is because the plane-to-plane formulation cannot correct for in-plane pose errors. We propose to remedy these in-plane errors using direct alignment with the LIDAR intensity pattern to the camera-detected checkerboard.

In this work, we assume that the camera intrinsic parameters are pre-computed. We can then derive from the checker corner coordinates in the i-th image the homography $\mathbf{H}_i$ such that:

$$s_{ik} \begin{bmatrix} \mathbf{x}_{ik} \\ 1 \end{bmatrix} = \mathbf{H}_i \begin{bmatrix} \mathbf{p}_{ik} \\ 1 \end{bmatrix}. \tag{6}$$

where $\mathbf{H}_i \in \mathbb{R}^{3\times3}$ warps each normalized, camera-detected checkerboard corner $\mathbf{p}_{ik} \in \mathbb{R}^2$ to its corresponding position $\mathbf{x}_{ik} \in \mathbb{R}^2$ on the canonical checkerboard plane. $s_{ik} \in \mathbb{R}$ is a scaling factor. Note that these normalized coordinates $\mathbf{p}_{ik}$ are derived using the distortion and pinhole projection parameters from the given pre-computed camera intrinsic parameters.

With Eq. 6, we can also use the latest estimates of the extrinsic parameters to map each j-th LIDAR point to its corresponding point on the i-th checkerboard plane using:

$$\mathbf{x}_{ij} = \mathbf{H}_i(\mathbf{R}_L^C \mathbf{P}_{Lij} + \mathbf{t}_L^C) \tag{7}$$

where for brevity, we have omitted the scaling term that follows the homography transform. Since $\mathbf{H}_i$ is derived from the camera's detection of the checkerboard, by comparing the LIDAR intensity of $\mathbf{P}_{Lij}$ with its corresponding checkerboard intensity function value at the point $\mathbf{x}_{ij}$ from Eq. 7,

we can get a measure of the agreement between the two domains' signals. Next we discuss our design of a suitable checkerboard intensity function.

Differentiable Checkerboard Pattern: An actual checkerboard pattern is a square wave that is piecewise constant and discontinuous at the transition points. In our framework, however, we require an intensity checker pattern with well defined spatial partial derivatives, so as to transform the intensity misalignments into pose update information. As such, we adopt a differentiable approximation to the checkerboard pattern using sinusoids as seen in Fig. 3.

The board in Fig. 3 is a 2-D function $I(\mathbf{x}) : \mathbf{x} \rightarrow [-1, 1]$ for $\mathbf{x} \in \mathbb{R}^2$ defined as:

$$I(\mathbf{x}) = \begin{cases} \sin(\frac{\pi \cdot x_x}{w} + \frac{\pi}{2}) \cos(\frac{\pi \cdot x_y}{w}), & \text{if } \mathbf{x} \in \mathcal{B} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

where $[x_x, x_y]$ are the coordinates of the point $\mathbf{x}$ in the 2D plane, and $w \in \mathbb{R}^+$ is the spatial scale of each checker square. Note that points that fall beyond the extent of the checker pattern $\mathcal{B}$ return zero.

The corresponding partial derivatives of Eq. 8 are given by the chain-rule as:

$$\frac{\partial I(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} I_y \cdot \frac{\pi}{w} \cdot \cos(\frac{\pi \cdot x_x}{w} + \frac{\pi}{2}) \\ -I_x \cdot \frac{\pi}{w} \cdot \sin(\frac{\pi \cdot x_y}{w}) \end{bmatrix}^T \tag{9}$$

Where $I_x = \sin(\frac{\pi \cdot x_x}{w} + \frac{\pi}{2})$ and $I_y = \cos(\frac{\pi \cdot x_y}{w})$ are the x and y-coordinate contributing factors to $I(\mathbf{x})$ respectively. Similarly all $\mathbf{x} \notin \mathcal{B}$ have all partial derivative components set to zero so that they do not influence the alignment optimization.

We obtain $I_{ref_{ij}}$, i.e., the expected intensity of the j-th LIDAR point in the i-th checkerboard pose by chaining together the coordinate transform from the LIDAR to the camera reference frame, the homography thereafter to the camera-detected checkerboard, and lastly, our differentiable

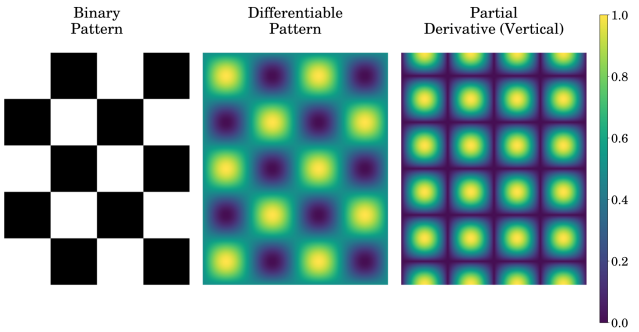


Fig. 3: A side-by-side comparison of an actual checker pattern with the differentiable approximation that is used for direct alignment. Shown on the right is the absolute value of the partial derivative $\nabla_y I$ (with respect to the vertical direction on this page). Notice that the partial derivative achieves its peak values at the vertical transitions between the checkers. The exact values have been scaled for better visualisation.

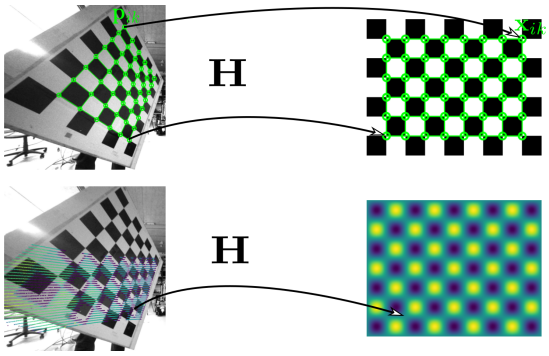


Fig. 4: Two views depicting that the homography detected in the camera image (top) that is used to transform the LIDAR points from their estimated position in the camera frame to the corresponding location on the differentiable checker pattern (bottom). The incorrect initial pose estimate (bottom left) leads to discrepancies between the LIDAR-measured intensities and the intensity values at their projected locations (bottom right). This discrepancy, along with the Jacobian shown in Fig.3 forms the gradient for the optimization.

checker intensity pattern. $I_{ref_{ij}}$ is computed as:

$$I_{ref_{ij}}(\mathbf{R}_L^C, \mathbf{t}_L^C) = I(\mathbf{H}_i(\mathbf{R}_L^C \mathbf{P}_{Lij} + \mathbf{t}_L^C)). \tag{10}$$

We then optimize for the camera-LIDAR extrinsic parameters using a direct alignment approach by minimizing the square of the error $r_{I_{ij}}$ jointly for all the LIDAR points. the error $r_{I_{ij}}$ is defined as:

$$r_{I_{ij}}(\mathbf{R}_L^C, \mathbf{t}_L^C) = I_{ref_{ij}}(\mathbf{R}_L^C, \mathbf{t}_L^C) - I_{L_{ij}}. \tag{11}$$

jointly for all the LIDAR points. Here $I(\cdot)$ is as given in Eq. 8. Note that the expected LIDAR intensity (of a point at the current extrinsic parameter estimates) is compared against the normalized observed intensity I_L reported by the LIDAR. This normalization is needed because, the range of the raw intensity signal $I_{L_{raw_{ij}}}$ reported by LIDARs can vary depending on the manufacturer, software driver used and the distance of the reflecting surface from the LIDAR etc. Therefore, for direct alignment, we normalize the intensity channel of all LIDAR points that fall on the same checkerboard observation independently i.e.

$$I_{L_{ij}} = \frac{I_{L_{raw_{ij}}} - \mu(I_{L_{raw_i}})}{\sigma(I_{L_{raw_i}})} \tag{12}$$

where $\mu(I_{L_i})$ and $\sigma(I_{L_i})$ are the mean and standard deviation of all raw intensity channel measurements of the LIDAR points on the i-th checkerboard detection, respectively.

To minimize the direct alignment residual presented in Eq. 11, we utilize the Jacobian from pose parameters to intensity error define in close form as:

$$\frac{\partial r_{I_{ij}}(\mathbf{R}_L^C, \mathbf{t}_L^C)}{\partial \xi} = \frac{\partial I(\mathbf{x})}{\partial \mathbf{x}}|_{\hat{\mathbf{x}}_{ij}} \frac{\partial \hat{\mathbf{x}}_{ij}}{\partial \hat{\mathbf{P}}_{C_{i,j}}} \left[\mathbf{I} | \lfloor \hat{\mathbf{P}}_{C_{i,j}} \rfloor_\times\right] \tag{13}$$

$$\frac{\partial \hat{\mathbf{x}}_{ij}}{\partial \hat{\mathbf{P}}_{C_{i,j}}} = \frac{\hat{s}_{ij} \mathbf{H}_i - \begin{bmatrix} \hat{\mathbf{x}}_{ij} \\ 1 \end{bmatrix} \mathbf{H}_{(3,:)_i}}{\hat{s}_{ij}^2} \tag{14}$$

where $\hat{\mathbf{P}}_{i,j}$ is as detailed in Eq. 5, $\mathbf{H}_{(3,:)_i}$ is the third row of $\mathbf{H}_i$, and $\hat{\mathbf{x}}_{ij}$ and $\hat{s}_{ij}$ satisfy the equation:

$$\hat{s}_{ij}\begin{bmatrix}\hat{\mathbf{x}}_{ij}\\1\end{bmatrix} = \mathbf{H}_i\hat{\mathbf{P}}_{C_{i,j}} \tag{15}$$

Finally, to refine all in-plane intensity alignment errors jointly, we optimize

$$\hat{\mathbf{R}}_L^C, \hat{\mathbf{t}}_L^C = \min_{\mathbf{R}_L^C, \mathbf{t}_L^C} \sum_i \sum_j^{|P_{L_i}|} r_{I_{ij}}^2(\mathbf{R}_L^C, \mathbf{t}_L^C) \tag{16}$$

using the Levenberg-Marquardt algorithm with the residuals and Jacobians defined in Eq. 11 and Eq. 13, respectively. As the initial guess for the extrinsic parameters, we use the estimates from the point-to-plane matching algorithm which, as seen in Fig. 1, is typically accurate to within one checker square.

IV. Experimental Evaluation

Our method is intended to be an accurate, camera-LIDAR extrinsic calibration framework that is easy to use (requiring only a simple printed checkerboard target) and with a more flexible data collection process.

We present experiments which demonstrate the capabilities of our method and support our key claims, which are: (i) accuracy of the extrinsic calibration which exceeds a state-of-the-art approach using the same data, (ii) adaptability to different LIDAR sensors, and, (iii) robustness to challenging input data — specifically when only a part of the checkerboard is within the LIDAR view.

It is difficult to acquire the ground-truth extrinsic parameters between a camera and a LIDAR; thus, we evaluate the accuracy of our method in a similar fashion as [17]. Namely, for a stereo camera pair, we compute the extrinsic parameters between each camera and the LIDAR independently. Using these two individual camera-LIDAR pose estimates, we compute the relative pose between the two cameras which we then compare against ground truth camera-to-camera poses computed directly using common observations between the cameras. In this case we use the stereo calibration routines in the Kalibr library [3].

Formally, for a perfectly calibrated three sensor system we expect

$$\hat{\mathbf{T}}_L^l(\hat{\mathbf{T}}_L^r)^{-1}\mathbf{T}_l^r = \Delta\hat{\mathbf{T}} = \mathbf{I}_{4\times4} \tag{17}$$

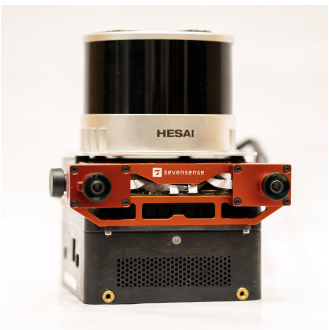


Fig. 5: Our main test device: a 32-beam Hesai XT32 LIDAR and a Sevensense Alphasense fisheye stereo camera.

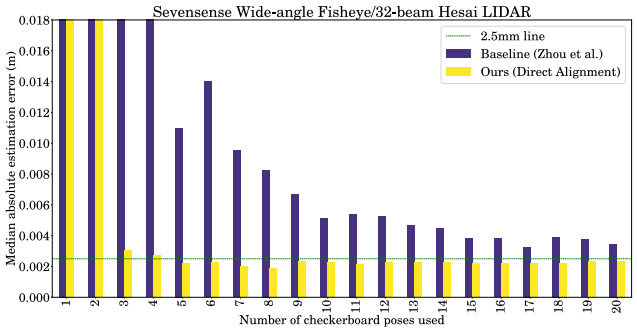


Fig. 6: Hesai XT32 Results: Our method outperforms the baseline approach, achieving roughly 1mm lower estimation error – a 30% error reduction.

where $\hat{\mathbf{T}}_L^{(\cdot)}$ denotes (for our stereo-pair experiments) the estimated pose of the LIDAR (L) with respect to the two cameras (l and r). $\mathbf{T}_l^r$ is the ground truth relative pose between the two cameras computed using the Kalibr library. $\Delta\hat{\mathbf{T}}$ is the estimation error which, when both two individual camera to LIDAR calibrations are perfect, should equal identity. A visual layout of these reference frames is shown in Fig. 2.

A. Comparison with State-of-the-Art

We begin our evaluation by comparing our calibration accuracy to that of the state-of-the-art approach of Zhou and Kaess [17] which is available in MATLAB's LIDAR toolbox. Our main test device is shown in Fig. 5, it consists of a Sevensense Alphasense fisheye stereo pair with ground truth stereo baseline of about 11cm and a 32-beam Hesai XT32 LIDAR. The Alphasense imagers are Sony IMX-273s fitted with wide-angle fisheye lenses with 165 degrees field of view.

We collected 30 synchronized sets of stereo images and LIDAR scans with the calibration target placed at different poses. For a fair comparison, the calibration targets in this experiment were placed such that all the four edges were visible to the LIDAR — a restriction necessary to operate the baseline method [17].

We then chose different, random subsets of the image/LIDAR scan pairs (without replacement) and ran the calibration algorithms 100 times for each subset size used, to compare the trend of accuracy as the number of calibration target detections increases.

As illustrated in Fig. 6, our method achieve better accuracy than state-of-the-art method of [17]. It also converges quicker to our best estimate (in terms of median error). This shows that for the same amount of input test data, our method achieves better accuracy. Still, we add a note of caution when comparing the MATLAB implementation of [17] because the sensor rigs, datasets and parameter tuning were all our own selections.

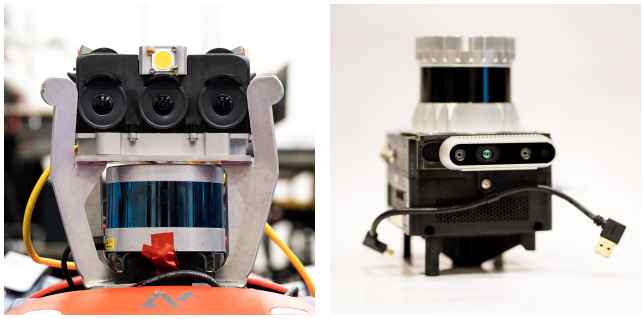
B. Applicability to different LIDAR sensors

Next, we demonstrate that our method can generalize to two different sensor setups: first an Ouster OS0-128 and then

a Velodyne VLP16. We use the same metric described earlier in this section of comparing the stereo baseline estimated using the two individual camera-LIDAR transforms computed by our algorithm, to the ground truth stereo width computed by Kalibr [3]. Here, from a set of 17 image/LIDAR scan pairs, we selected random subsets of various sizes with which we ran our calibration algorithm (100 times for each subset size) to show the accuracy trend with increasing calibration target poses, for the different sensor setups. Here we plot the accuracy of our algorithm against the plane-to-plane matching approach [5].

Ouster OS0-128 LIDAR is a dense but noisy LIDAR with a 90 degree field of view and is shown in Fig. 7b. It was paired with an Intel RealSense D435i (87 degrees field of view) in an older sensor rig (effectively a pinhole stereo camera). Nonetheless, our calibration performance behaves in a similar manner — as shown in Fig. 8. Due to the increased field of view, the checkerboard target is more commonly in the field of view and one can see that our method achieves a slightly lower residual error of 2mm. The same trend of the calibration accuracy improving with an increasing number of calibration targets is observed.

Velodyne VLP16 LIDAR is part of a sensor payload attached to the ANYbotics ANYmal C quadruped in Fig 7a as well as a set of Sevensense Alphasense fisheye cameras. With just 16 beams, this is a particularly sparse sensor with just 30 degrees field of view.



(a) Velodyne VLP16 LIDAR and a Sevensense Alphasense camera

(b) Ouster OS0 128-beam LIDAR and an Intel RealSense D435i

Fig. 7: Two additional sensor rigs used for evaluation in Sec. IV-B

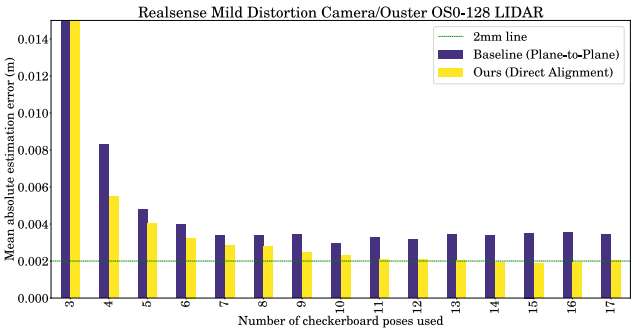


Fig. 8: Ouster OS0-128 Results: Even with this noisier sensor, our method outperforms plane-to-plane matching for the same number of checkerboard targets used.

Figure 9 shows that both plane-to-plane matching and our method improve in calibration accuracy with an increased number of calibration images used. Due to the sparsity of the LIDAR scans, the results for this sensor suite are slightly worse than those of the denser LIDAR setups (4mm error). The same trends and comparative results are observed with the proposed method and the plane-to-plane matching algorithm.

C. Calibration using partial calibration target LIDAR views

The method of Zhou and Kaess contains a pre-filtering phase where non-complete target detections are filtered out and do not contribute to the calibration estimation. If incomplete LIDAR scans of the checkerboard are not filtered, the edge detection algorithm fails as seen in Fig . 10 and leads to an incorrect calibration. Furthermore this filtering often rejects scans where a significant portion of the calibration target is visible in the LIDAR view. Scans which our method, on the other hand, can still use to accurately calibrate.

From a dataset of 20 image/LIDAR scan pairs (collected using our main test device Fig . 5), we identified 7 examples which were rejected by the MATLAB implementation of [17]. We show some illustrative examples of the rejected data samples in Fig. 11b. Next, we demonstrate that our method achieves accurate calibration even only using this set of image/LIDAR scans rejected by the pre-filtering phase of [17]. From an initial estimate computed by point-to-plane matching, we jointly optimized the 7 image/LIDAR scan pairs using our direct alignment formulation to achieve a stereo-baseline translation error of 2 mm. Fig. 11c shows the result of our calibration, demonstrating that we attain accurate calibration even using data that would have been incorrectly detected by [17] (had they not been filtered out).

To summarize, our evaluation shows that our method outperforms the calibration accuracy achieved by the state-of-the-art method [17], in a setting where all camera and LIDAR views observe the full checkerboard. We further demonstrated that our method works well with three of most popular LIDAR sensors available. Finally, we showed that our method accurately calibrates even when the LIDAR sees

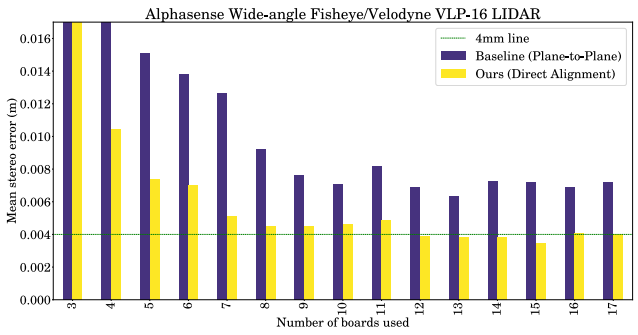


Fig. 9: Velodyne VLP16 Results: Calibration accuracy with this sparser sensor is higher than the other sensor suites. Still, our method can perform the plane-to-plane matching approach, showing lower mean error for the same number of calibration targets.

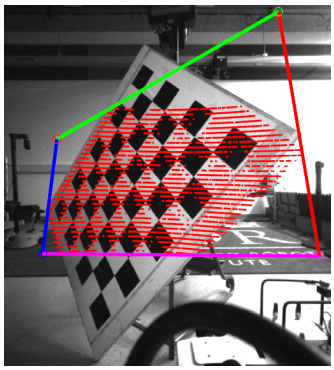


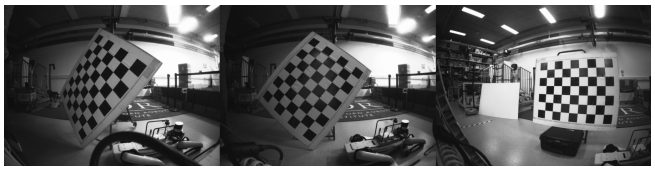
Fig. 10: An example LIDAR scan where the edge-detection algorithm of [17] would fail and result in incorrect calibration. Examples such as these are automatically filtered out in the baseline calibration method as they are unusable. We show in 11c that our method calibrates accurately even using examples such as these.

only partial views of the calibration target — which improves the efficiency of calibration data collection.

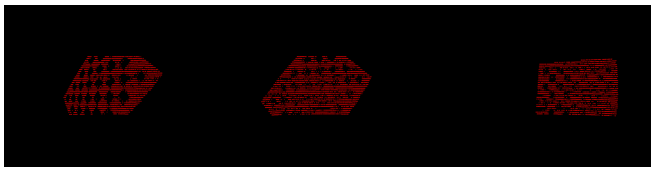
V. Conclusion and Future Work

In this paper, we presented a novel formulation of a differentiable checkerboard intensity pattern with which we perform direct alignment of the camera-detected checkerboard to the intensity pattern as observed by the LIDAR. By exploiting the gradient information analytically defined by our differentiable model, our method retrieves accurate camera-LIDAR calibration transforms — even when using only partial LIDAR views of the calibration target. This is especially beneficial in wide field-of-view camera setups, where calibration data ought to be collected closer to the camera (where they are more accurate). We showed that our method could be applied to sensors from 3 different LIDAR manufacturers achieving a typical calibration accuracy in the order of $2-4$mm and outperformed a state-of-the-art algorithm.

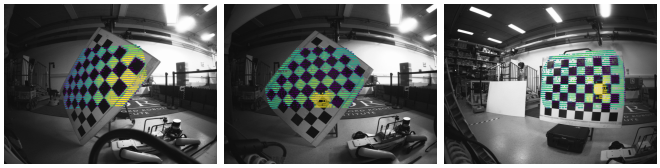
In future work, we aim to extend our differentiable formulation to the pixel space of the camera, so that we can simultaneously calibrate both extrinsic as well as intrinsic parameters, further simplifying the overall calibration process. Additionally, we also intend to explore jointly calibrating of all sensors simultaneously (e.g. 2 cameras and 1 LIDAR), taking advantage of the common calibration target observations which link the observations.



(a) Challenging data samples which are not captured fully in the lidar.



(b) Data samples that were filtered by the baseline method.



(c) These data samples can still be used by our proposed method.

Fig. 11: The baseline method [17] needs to filter out dense LIDAR views of the checkerboard — even when placed at a moderate range while our approach can still accurately process those samples.

References

[1] J. Beltrán, C. Guindel, and F.A. García. Automatic extrinsic calibration method for lidar and camera sensor setups. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–13, March 2021.

[2] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[3] P.T. Furgale, M. Moos, and R. Kümmerle. Kalibr: An open-source toolbox using the kalman filter for rigid body pose estimation from a variety of sensors. In *2013 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4443–4450. IEEE, 2013.

[4] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012.

[5] A. Geiger, F. Moosmann, Ö. Car, and B. Schuster. Automatic camera and range sensor calibration using a single shot. In *IEEE International Conference on Robotics and Automation*, pages 3936–3943, 2012.

[6] G. Iyer, R. KarnikRam, K.M. Jatavallabhula, and K.M. Krishna. Calibnet: Geometrically supervised extrinsic calibration using 3d spatial transformer networks. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1110–1117, 2018.

[7] X. Jing, X. Ding, R. Xiong, H. Deng, and Y. Wang. DXQ-Net: Differentiable lidar-camera extrinsic calibration using quality-aware flow. In *arXiv: 2203.09385*, 2022.

[8] J. Levinson and S. Thrun. Automatic online calibration of cameras and lasers. In *Robotics: Science and Systems*, 2013.

[9] X. Lv, B. Wang, Z. Dou, D. Ye, and S. Wang. LCCNet: LiDAR and camera self-calibration using cost volume network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2888–2895, 2021.

[10] G. Pandey, J. McBride, S. Savarese, and R. Eustice. Automatic targetless extrinsic calibration of a 3d lidar and camera by maximizing mutual information. In *AAAI Conference on Artificial Intelligence*, 2012.

[11] D. Scaramuzza, A. Harati, and R. Siegwart. Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4164–4169, 2007.

[12] T. Scott, A.A. Morye, P. Piniés, L.M. Paz, I. Posner, and P. Newman. Exploiting known unknowns: Scene induced cross-calibration of lidar-stereo systems. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3647–3653, 2015.

[13] R. Unnikrishnan and M. Hebert. Fast extrinsic calibration of a laser range finder to a camera. Technical report, Carnegie Mellon University, 2005.

[14] 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 Sensing*, 9(8), 2017.

[15] G. Yan, F. He, C. Shi, X. Cai, and Y. Li. Joint camera intrinsic and lidar-camera extrinsic calibration. In *arXiv: 2202.13708*, 2022.

[16] Q. Zhang and R. Pless. Extrinsic calibration of camera and laser range finder. In *IEEE/RSJ Conference on Intelligent Robots and Systems*, pages 2301–2306. 3, 2004.

[17] L. Zhou, Z. Li, and M. Kaess. Automatic extrinsic calibration of a camera and a 3d lidar using line and plane correspondences. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5562–5569, 2018.