ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

CVL Computer Vision Lab

Calibrating AV Sensors

MEMS Lidar, Frame Camera, Event Camera, and Radar

Semester Thesis

Kevin Ta
Master of Science in Robotics, Systems and Control

Advisors: David Brüggemann, Tim Brödermann, Christos Sakaridis
Supervisor: Prof. Dr. Luc van Gool

April 28, 2022

Calibrating AV Sensors: MEMS Lidar, Frame Camera, Event Camera, and Radar

Kevin Ta
Computer Vision Lab
ETH Zürich
kevita@ethz.ch

***Abstract*—This report explores the use of mutual information, stereo calibration, and Fourier-Mellin Transforms to calibrate an autonomous driving sensor suite which includes a monocular frame-based camera, a monocular event-based camera, a MEMS lidar, and a rotating radar. By leveraging the sensing modality similarities, we formulate bi-sensor extrinsic calibration and joint multi-sensor calibration. Nominal calibrations for radar-lidar are completed using frequency space methods and event camera-camera calibration results are performed traditional stereo-based methods. Our experiments show how intrinsic and extrinsic calibrations are highly sensitive to the quality of the calibration target and the properties of the unstructured environment. We also present the synergistic relationship between event cameras and lidars as the event-driven sensors exhibit high dynamic range that can capture the lidar laser returns outside the wavelengths of visible light. Multiple optimizers were evaluated in an information-based paradigm for their associated robustness to noise and computational performance. We show that event-driven sensors and lidars achieve a high degree of repeatability regardless of optimizer and present a novel method for extrinsic calibration between these two sensors.**

The source code for camera-lidar calibration is available at https://github.com/kev-in-ta/l2c-info-calib. The forked code for the radar-lidar calibration is available at https://github.com/kev-in-ta/radar_to_lidar_calib.

I. Introduction

In recent years, the development of autonomous vehicles has accelerated due to both academic research and industry efforts. In these approaches, many sensing modalities have become part of the standard autonomous driving suite, including cameras, lidars, and radars [2], [3], [12], [20]. These sensors have enabled significant advances in perception, localization, and odometry. A key aspect of enabling these sensors is through careful calibration of their intrinsic and extrinsic parameters which details how each sensor is positioned relative to each other and how the sensor signals are properly mapped into a global frame. Figure 1 shows how calibration is important for geometric alignment between an RGB frame-based camera and a lidar sensor. Without calibration, the points are projected incorrectly into the image space, likely leading to negative consequences in downstream perception.

One type of sensor which has had limited exploration in the realm of autonomous vehicle perception is the event camera, an asynchronous sensor which extracts pixel-wise visual differences as opposed to traditional cameras which capture entire frames at regular intervals. Event cameras have the capacity to augment the existing standard suite of sensors by providing a dense scene representation, capturing fast moving objects at higher frequencies, reducing data bandwidth in static scenes, and exhibiting high dynamic range in low-light and adverse conditions [6].

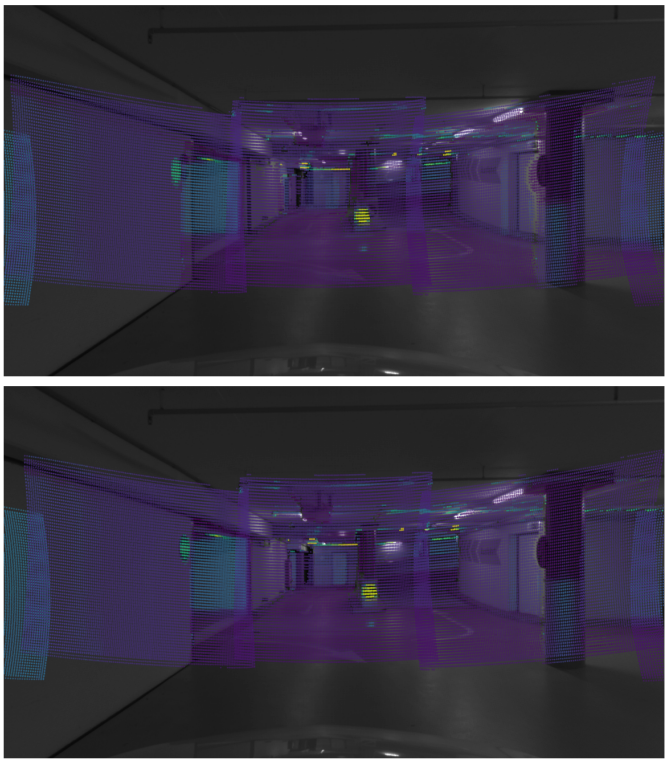

Fig. 1. Scene 34 (Top) Uncalibrated scene with lidar points projected using seed values. (Bottom) Calibrated scene showing proper lidar point projection alignment.

Many sensors are effectively intrinsically calibrated by the vendor or manufacturer, reducing the total amount of calibration required. However, some sensors still require intrinsic calibration processes due to their customizability. Cameras can have many different configurations affecting image quality, including which lens is equipped and the adjustment to the focus. This variability requires that the sensors are calibrated intrinsically first before extrinsic geometric alignment. Standard methods for intrinsic calibration, such as those based on the seminal work by Zhang [21], exist and are widely implemented.

Methods for sensor fusion aim to leverage the advantages of different sensors for more robust perception, especially in challenging environments and adverse conditions [20]. To fully enable the effective use of multi-sensor set-ups, accurate calibration is pivotal in ensuring that detected features and objects are geometrically aligned. Due to variability in fabrication of sensors and installation onto the robot platform, design schematics and CAD models may not be accurate, especially in the case of small extrinsic rotations.

Many sensors share sensing modalities in how they interface with their surroundings or in how they represent their surroundings. Typical cameras are sensitive to visible light with wavelengths between 380-750 nm. Lidars typically emit near infrared light with wavelengths between 800-1100 nm. Event cameras exhibit high dynamic range around the typical wavelengths of visible light and may be partially sensitive to wavelengths outside the typical range of visible light. In terms of representation, cameras project the 3D environment to a dense 2D image that lacks depth information, while Lidars capture the full 3D environment as a sparse 3D point cloud with direct depth information. The radar used in this experiment is a rotating sensor which captures a 2D slice of the 3D environment with direct depth information.

Calibration methods can be categorized as structured calibration or unstructured calibration. Structured calibration uses defined physical markers that can be accurately located by multiple sensors. For example, the use of chessboards for cameras, planar targets for Lidars, and dihedral targets for radars. Solving for extrinsics between sensors is accomplished with targets that can be identified by multiple sensors, allowing for relative calibration and alignment between sensors. As an example of a multi-modal target, lidars can detect structured edges from chessboard targets while cameras can detect the chessboard pattern.

Unstructured calibration, sometimes referred to as automatic calibration, foregoes easily identifiable targets for simpler hardware set-ups and greater scene flexibility. Unstructured calibration aims to identify shared features in scenes such as edges [10] or planes. These methods look to extract features identifiable in multiple sensors that are present in everyday scenes or environments. Some of these methods aim to jointly calibrate for geometric and temporal differences between sensors, while other methods focus solely on geometric alignment.

Balancing the flexibility of unstructured calibration methods and the geometric motivation of structured methods is an important consideration when calibrating a suite of sensors. In this report, we present a hybrid structured-unstructured method for geometrically aligning a novel suite of sensors, which include a monocular frame camera, a monocular event camera, a MEMS lidar, and a spinning radar. Following intrinsic calibration of the frame-based and event-based cameras, we leverage an information-based unstructured calibration framework for camera-lidar extrinsic calibration. We also present target-based stereo calibration for event camera-camera alignment and a novel method to perform a fully-consistent event camera-camera-lidar calibration in a semi-structured fashion. Finally, we perform unstructured radar-lidar calibration using frequency domain methods to complete the full extrinsic calibration of the sensor suite.

II. Related Work

A significant number of autonomous driving datasets have developed over the last decade. In 2017, the Oxford RobotCar team released a multimodal dataset including lidar, cameras, and GNSS positional sensors under a variety of conditions [12]. This dataset was expanded in 2020 to include a spinning radar sensor [2]. DSEC, a stereo-based event camera dataset was released in 2021 with lidar, stereo event cameras, and stereo frame cameras [8]. In 2022, the Boreas dataset, a dataset that captures many adverse conditions, was released with a platform that included cameras, lidar, and radar [3]. In all of these datasets, intrinsic and extrinsic calibration parameters are necessary to handle the multi-modal data. In comparison to the aforementioned datasets, the EFCL vehicle platform will includes a MEMS lidar, a spinning radar, a monocular frame-based camera, and a monocular event-based camera.

Intrinsic calibration of cameras is a well-established field with widely available tools such as OpenCV [7] or Kalibr [15]. These standard methods have ROS implementations and are implemented in multiple programming languages allowing for their ease of use in robotics and autonomous vehicle development. Extension of these methods have recently been made for event cameras where state-of-the-art event-to-video reconstruction is used to reconstruct frames of chessboard patterns for the purpose of intrinsic calibration [13]. Standard methods can then be used to complete the intrinsic calibration of the camera.

Many methods have been proposed for extrinsic calibration between different sensors. Structured stereo calibration methods are common to calibrate in multi-camera set-ups and are implemented in common libraries like OpenCV and Kalibr. These tools identify planar targets with different potential patterns including chessboard, ArUco, ChArUco, and AprilTag patterns and performs PnP (pose) optimization to extrinsically calibrate each sensor.

Lidar-camera methods have also been heavily explored in research. These include unstructured mutual information maximization schemes [16], [19], edge alignment methods, or plane fitting methods. The first approach leverages the similar sensitivity to wavelengths of light between cameras and lidars to maximize their joint intensity mutual information. Kang and Doh propose a probabilistic edge detection method to maximize the detected edge overlap between lidar scans and images [10]. An et al. fused 3D-2D and 3D-3D point correspondences with structured planar targets and unstructured environmental objects for robust calibration [1]. In this work, we explore the use of information frameworks for camera-lidar calibration.

Radar calibration research often explores structured methods for calibration where specialized targets that maximize radar returns are developed for accurate localization. Different methods of calibration targets have been explored for camera-radar

calibration [14], lidar-radar calibration [11], [17], and camera-lidar-radar calibration [5]. Additional, there are efforts to explore unstructured automatic lidar-radar calibration through feature point registration between the two sensors [9].

Radar calibration, due to the coarse nature of radar signals, are sometimes performed in SE(2) as opposed to SE(3). This reduces the complexity down to a single rotation and planar translation, and allows for 2D point registration in the birds-eye-view (BEV) representation. This allows for certain methods such as the Fourier-Mellin Transform (FMT) [18] for point registration. Such methods were employed for 2D radar odometry by Cen and Newman [4]. For radar-lidar calibration, we adapt the calibration method used by the Boreas dataset using FMT which incorporates feature extraction methods developed by Cen and Newman.

Domhof et al. [5] also propose a framework for calibration based on the connectivity of extrinsic calibrations. Extrinsic calibration can be performed in a minimally-connected fashion where there are no loops in the extrinsic relationship graphs between sensors. They proposes a full-connected approach to calibration where multiple sensors optimize for their joint extrinsics. This approach ensures fully consistent calibration by ensuring loop closure constraints.

This works differentiates itself from past work by exploring extrinsic calibration of two emerging sensors for autonomous systems, an event-based vision sensor and a MEMS lidar. Additionally, we propose a fully-connected calibration scheme between camera-camera-lidars system where stereo calibration and mutual information optimization are used to build self-consistent extrinsics.

III. Methodology

For camera intrinsic calibrations, we use the standard camera pinhole and distortion model. the Pinhole model, K, is described by the focal length (F_x, F_y) and principal point (C_x, C_y) parameters as described in Eq. (1). The distortion model is described by the 5 distortion parameters (k_1, k_2, p_1, p_2, k_3) where k_n are the radial correction terms for barrel/pincushion distortion and p_n are the tangential correction terms for image skew.

$$\mathrm{K} = \begin{bmatrix} F_x & 0 & C_x \\ 0 & F_y & C_y \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

For representation of our extrinsics calibration as homogeneous transformations, we collapse the transformations down to a 6 parameter vector given by the translation parameters (x, y, z) and the rotational vector (axis-angle) representation (v_1, v_2, v_3) where norm$||\vec{v}||$ is the angle of rotation in radians. Thus the full representation of the extrinsic transformations is (x, y, z, v_1, v_2, v_3). The translation parameters will be referred to as the translation vector t. The expanded rotation into its matrix form will be referred to as R.

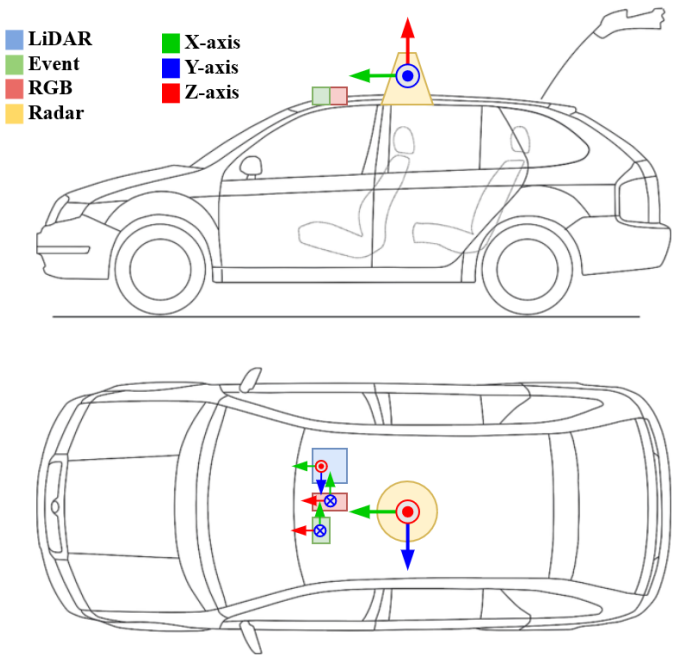

Fig. 2. Sensor set-up on the EFCL data collection vehicle and the associated default co-ordinate systems for each sensor.

A. Sensor Specifications

The EFCL vehicle platform is equipped with the Lucid Vision Lab's Triton 5.4 MP Model frame-based camera, Prophesee's Gen4.1 event-based vision sensor, the RoboSense RS-LiDAR-M1 MEMS lidar, and Navtech's CIR spinning radar sensor. Table I gives an overview of the sensor information.

Table I. Overview of sensors on EFCL data collection vehicle.

Sensor	Type	Resolution	HFoV	VFoV	Freq.
TRI054	frame	1920×1080	77°	48°	20.8 fps
Gen4.1	event	1280×720	64°	39°	50 Gev/s
RS-LiDAR-M1	lidar	600×175	120°	25°	15 Hz
CIRDEV	radar	400 / sweep	360°	1.8°	4 Hz

The frame camera, event camera, and the MEMS lidar exhibit a similar forward field of view, while the radar captures the full 360° view around the vehicle. Figure 2 shows the sensor set-up as mounted on the vehicle and the default co-ordinate systems for each sensor.

B. Intrinsic Calibration

To calibrate the intrinsic parameters for the frame camera, a number of chessboard images are captured. We ensure a variety of positions and orientations are represented to improve the calibration through varied pose constraints. These images are then run through a calibration pipeline developed with OpenCV to extract corner points which are then used to calibrate the pinhole camera model parameters and the standard distortion model.

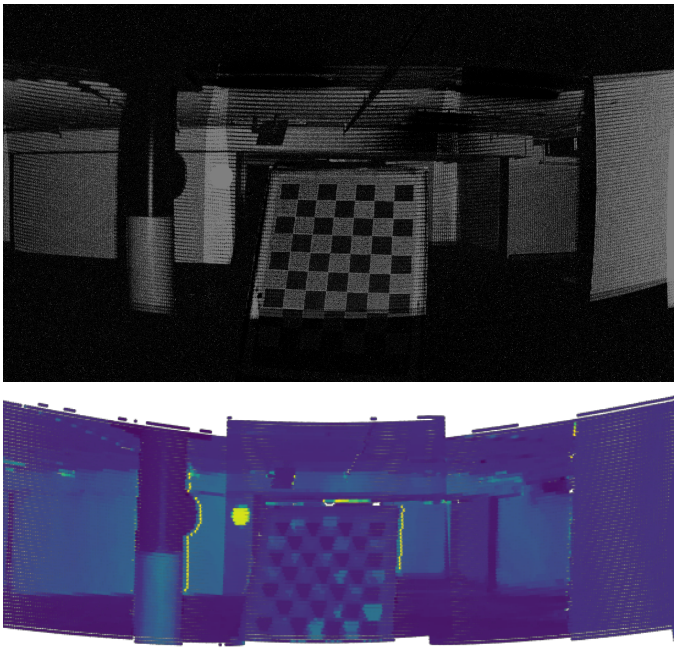

Fig. 3. Scene 48 (Top) Accumulated events in the synthesized event camera image. (Bottom) Cropped projected lidar scans into synthetic image space.

To calibrate the intrinsic parameters for the event camera, the process described in [13] is used where events are reconstructed into video frames where the chessboard patterns can be extracted. The reconstructed frames can then be calibrated with the pipeline described above for the standard camera intrinsic parameters.

The RoboSense lidar and the Navtech radar are calibrated by the manufacturers and left unmodified aside from mounting to the vehicle platform and are simply taken at face value.

C. Event Camera-Lidar Optical Wavelength Detectability

Event-based cameras in static scenes will typically just return noise due to a lack of motion. However, the Prophesee Gen4.1 event camera exhibits higher sensitivity to wavelengths of light outside the visible range, picking up the 905 nm beams emitted by the RoboSense lidar. Consequently, the event camera captures an "illuminated" scene from the emitted lidar beams. A synthetic image can be generated from the accumulated events for the event camera which is shown in Figure 3.

This synergistic property between the event camera and lidar enables the generation of synthetic images for calibration without the need of target or scene motion and without the inclusion of temporal calibration. The decoupling of geometric calibration from temporal calibration is of significant benefit in calibration processes as it reduces the optimization space. An additional advantage is the ability to simplify calibration set-up design where external active elements or platform motion may be required for full spatial-temporal calibration.

D. Camera-Lidar Extrinsic Calibration

As cameras and lidars operate at similar wavelengths, their sensing modalities are also quite comparable. In particular, the intensity of lidar returns are highly correlated to the grayscale values found in an image. This correlation is exploited to formulate a mutual information maximization problem to calibrate the camera to the lidar.

1) Mutual Information Formulation: Mutual information (MI) is a measure of statistical dependence between random variables, indicating how much information one variable contains regarding the other. MI can be described multiple ways, but we take the same entropy based representation used in [16]. Here the MI is defined in terms of the entropy of the random variables X and Y, and their respective joint entropy $\mathrm{H}(X, Y)$.

$$\mathrm{MI}(X, Y) = \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X, Y) \tag{2}$$

The entropy denotes a measure of uncertainty within one variable, while the joint entropy represents the uncertainty present in the event of a co-observation of X and Y. Note that mutual information cannot be negative as the maximum uncertainty the joint entropy can have is at most equivalent to the uncertainty of variables X and Y in the situation where X and Y are perfectly uncorrelated. The entropies of random variable X, Y, and their joint entropy are described in Eq. (3, 4, 5).

$$\mathrm{H}(X) = -\sum_{x \in X} \mathrm{p}_X(x) \log\, \mathrm{p}_X(x) \tag{3}$$

$$\mathrm{H}(Y) = -\sum_{y \in Y} \mathrm{p}_Y(y) \log\, \mathrm{p}_Y(y) \tag{4}$$

$$\mathrm{H}(X, Y) = -\sum_{x \in X} \sum_{y \in Y} \mathrm{p}_{XY}(x, y) \log\, \mathrm{p}_{XY}(x, y) \tag{5}$$

2) Probability Distribution Formulation: We use the same formulation as Pandey et al. [16] for the approximation of the probability distribution through the intensity and grayscale histograms of the lidar scan and image respectively. Let $\{\mathbf{P}_i; i = 1, 2, ..., n\}$ be the set of homogeneous 3D points in the 3D scan. Then $\{\mathrm{X}_i; i = 1, 2, ..., n\}$ would be the set of intensity returns for each point in the set $\mathbf{P}$. The lidar points can be projected into the image space through the extrinsics (R, $\mathbf{t}$) and the intrinsics K as shown in Eq. (6).

$$\mathbf{p}_i = \mathrm{K}[\mathrm{R}|\mathrm{t}]\mathbf{P}_i \tag{6}$$

The location of the projected point are then used to acquire the associate grayscale value in the image:

$$\mathrm{Y}_k = \mathrm{I}(\mathbf{p}_i) \tag{7}$$

From the accumulated points, a histogram is generated from the discretized intensity values and total points, n, that lie within the valid image region as described in Eq. (8).

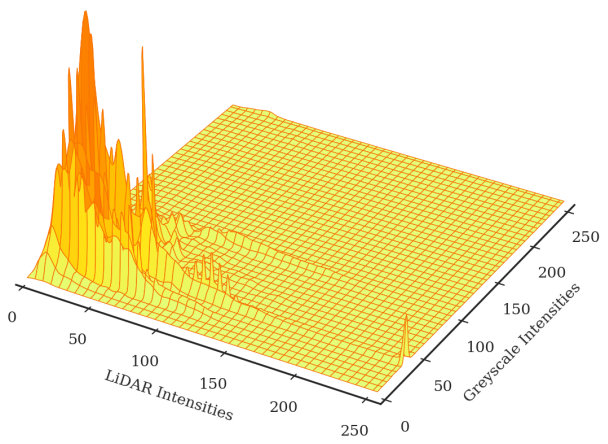

Fig. 4. Smoothed joint histogram as the approximate joint probability distribution between the image and lidar intensities.

$$\hat{p}(X = k) = \frac{x_k}{n}, \qquad k \in [0, 255] \tag{8}$$

The raw histograms, however, are quite noisy and the optimization requires a smooth function to find suitable solutions. To address this issue, we smooth our histograms using Gaussian blurring convolutions, a method that is nearly equivalent to kernel density estimation (KDE) in the case where the histograms bins are equally spaced and the points already discretized. Gaussian blurring runs orders of magnitude faster than true KDE algorithms. In the 1D case these are nearly equivalent aside from the marginal difference in finite and infinite support between the two methods. In the 2D case these methods are also nearly equivalent, except for the marginal covariance terms and the same support issues present in the 1D case. We use the same kernel smoothing parameters through Silverman's rule-of-thumb for the 1D case to ensure the probability distributions are consistent. Eq. (9) shows the kernel density approximation for the 2D joint probability.

$$\begin{aligned} \hat{p}(X = k) &= \frac{1}{n} \sum_{i=1}^{n} K_\Omega \left(\begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} X_i \\ Y_i \end{bmatrix} \right) \\ \Omega &= f \times n^{-1/5} \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix} \\ f &= \left(\frac{(d+2)}{4} \right)^{\frac{-1}{d+4}} \end{aligned} \tag{9}$$

Where Ω is the KDE smoothing parameter (standard deviation of the Gaussian blur), f is Silverman's factor, n is the total number of points, and $d = 1$ for the one-dimensional case of Silverman's rule-of-thumb. Figure 4 shows the joint probability distribution of the lidar intensities and the image grayscale values.

3) Optimization: With the MI criteria defined and our approximations of the probability distributions, we formulate an optimization problem as follows:

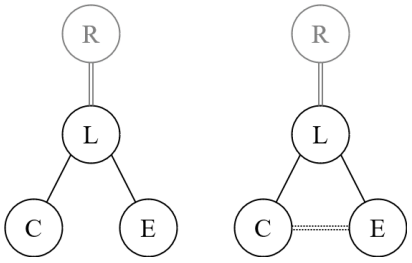

Fig. 5. R - radar; L - lidar; C - frame camera; E - event camera. Single solid line - MI-based optimization; Double solid line - FMT point registration; Double dotted line - stereo calibration. (Left) Minimally-connected pose estimation. (Right) Fully-connected pose estimation between L/C/E.

$$\hat{\Theta} = \arg\max_{\Theta} \text{MI}(X, Y; \Theta) \tag{10}$$

where $\Theta = (x, y, z, v_1, v_2, v_3)$, when parameterized by the axis-angle form, and is maximized at the correct extrinsic parameters. Effective optimization of this function is heavily assisted by smooth convex functions. The KDE helps smooth out the optimization process, but convexity is not ensured, in part due to the choice of angle parameterization.

The optimization process is accomplished with the standard optimizers available in SciPy. These optimizers include robust non-gradient methods such as "Nelder-Mead" or the "Powell" shooting method, as well as gradient based methods like the "Broyden-Fletcher-Goldfarb-Shanno" (BFGS) algorithm, conjugate gradient methods, or "Sequential Least SQuares Programming" (SLSQP).

E. Event Camera-Camera Extrinsic Calibration

To calibrate the monocular frame camera to the event camera, we use established stereo calibration methods, such as those implemented in OpenCV. Here, we can use co-occurrence of chessboards in the field-of-view (FoV) to calibrate the relative positions of the two cameras by optimizing the reprojection error.

F. Camera-Lidar Calibration with Stereo Constraints

The above methods optimize the calibration between two sensors without regard for the consistency of extrinsic calibration between all three sensors. Calibration can be performed in a minimally-connected or fully-connected manner [5]. Figure 5 shows the two possible configurations. In our procedure, camera-camera calibration is performed using stereo calibration, camera-lidar calibration is performed using intensity mutual information maximization, and radar-lidar is performed using frequency domain FMT for point registration.

Extending the previous methods, we utilize the stereo calibration and optimize for both the mutual information in the RGB camera-lidar and event camera-lidar calibration whereby the optimization ensures the stereo constraint is satisfied. The optimization objective function is shown in Eq. (11).

$$\text{F}(X, Y_1, Y_2) = \text{MI}(X, Y_1) + \alpha \text{MI}(X, Y_2) \tag{11}$$

Where Y_1 is the RGB camera image, Y_2 is the event camera image, and α is a weighting factor that controls which image is prioritized in the optimization. In our experiments, we use equal weighting with $\alpha = 1$.

G. Radar-Lidar Extrinsic Calibration

Both lidar and radar sensors directly acquire range information. The sensing modalities between radar and imaging in unstructured scenes is fundamentally quite different, and is thus not unified with the methods described above. This is noted in the connectivity graph shown in Figure 5 where the radar is shown to be solely connected to the lidar.

Consequently, we only exploit the inherent structure similarity, direct depth measurement, of the lidar and radar returns to perform point registration. Specifically, we perform a SE(2) calibration in the frequency domain using FMT. This method was employed in [3] for the Boreas dataset and we employ those methods with minor changes to suit the data format and configuration available to the EFCL vehicle platform.

A pivotal aspect of this method is key-point extraction of the noisy radar signal to get a sparse 2D point map for registration. The feature extraction follows from [4] which detects landmark objects from the power-range spectrum and was originally used for the purpose of radar odometry.

This method is primarily useful in calibrating the relative yaw rotations between the radar and lidar. Due to the coarse nature of the radar returns (azimuth resolution of 0.9° and range resolution of 4.4 cm), the positional calibration from this method is an order of magnitude less precise than hand measurements of the installed sensors on the vehicle. As a result, the measured positional seed values for the as-built radar configuration will be used as the extrinsics parameters for calibration.

IV. Experimental Results

To complete full extrinsic calibration, 93 scenes were captured in an underground parking lot at the ETH Zürich Zentrum campus. These scenes are divided roughly into two main subgroups, the unstructured garage scenes and the stationary "structured" scenes where a chessboard in different positions is visible in the FoV. Radar returns were only captured in the last two sub-datasets (Chessboard #3 Garage #2). A breakdown of these calibration scenes is shown in Table II.

Table II. Calibration datasets.

Calibration Set	Scenes	RADAR
Chessboard #1	1 - 10	✗
Garage #1	11 - 35	✗
Chessboard #2	36 - 40	✗
Whiteboard #1	41 - 45	✗
Chessboard #3	46 - 65	✓
Garage #2	66 - 93	✓

As all sensors are calibrated to the lidar, we present seed calibration values with respect to the lidar coordinate frame. Table III shows the initial seed calibration parameterized with ZYX Euler angles for interpretability. In the lidar coordinate frame, the forward direction is the x-axis, to the left is the y-axis, and the upward direction is the z-axis.

Table III. Extrinsic seed calibrations relative to lidar.

Sensor	x (m)	y (m)	z (m)	$\alpha(^\circ)$	$\beta(^\circ)$	$\gamma(^\circ)$
TRI054	-0.050	-0.106	0.00	90	0	90
Gen4.1	-0.050	-0.188	0.00	90	0	90
CIRDEV	-0.400	-0.150	0.20	0	0	0

Multiple optimizer choices were tested to evaluate their performance in terms of robustness, accuracy, and computational speed. The principal choices were between non-gradient based methods like "Nelder-Mead" or the "Powell" shooting method compared to numeric gradient based methods like "L-BFGS-B" and "SLSQP".

When using a sufficient number of scenes the optimizers converge to an optimal solution in the parameter space. Figure 6 shows how the 1D variations across each axes of translation rotation shows convergence to a local optima for the event camera-lidar calibration.

A. Intrinsic Calibration Results

Prior to extrinsic calibration, intrinsic calibration is required as we need to determine the correct mapping from 3D space into the 2D image. This mapping is critical for the point projection step in extrinsic calibration and in future sensor fusion. Additionally, the dataset will also provide rectilinearized images, or undistorted images based on the intrinsic calibration, to simplify the image loading and training of deep learning perception models. For the Triton 5.4 MP camera, 326 valid images were captured for calibration with an 9x7 checkerboard grid (8x6 valid points). For the Prophesee Gen4.1, image reconstruction was performed using E2VID, a tool that converts events into image frames. These frames can them be used through any standard camera intrinsic calibration process. 558 valid images were used for the event camera calibration.

The two sensors were calibrated using the OpenCV implementation of camera calibration. Table IV shows the calculated camera model properties while Table V shows the distortion parameters after calibration.

Table IV. Camera intrinsic calibration pinhole results.

Camera	RP Error	F_x	F_y	C_x	C_y
TRI054	**0.536**	1200.72	1201.14	939.19	538.03
Gen4.1	**0.904**	1027.79	1029.03	615.96	342.07

The reprojection errors for the intrinsic calibrations are quite high for calibration, with typical reprojection error targets lying between 0.1-0.3 pixels. There are many possible sources as for why the reprojection error is high. For example, many images exhibit high degrees of motion blur or are generally quite dark in the event-image reconstruction. This effect could

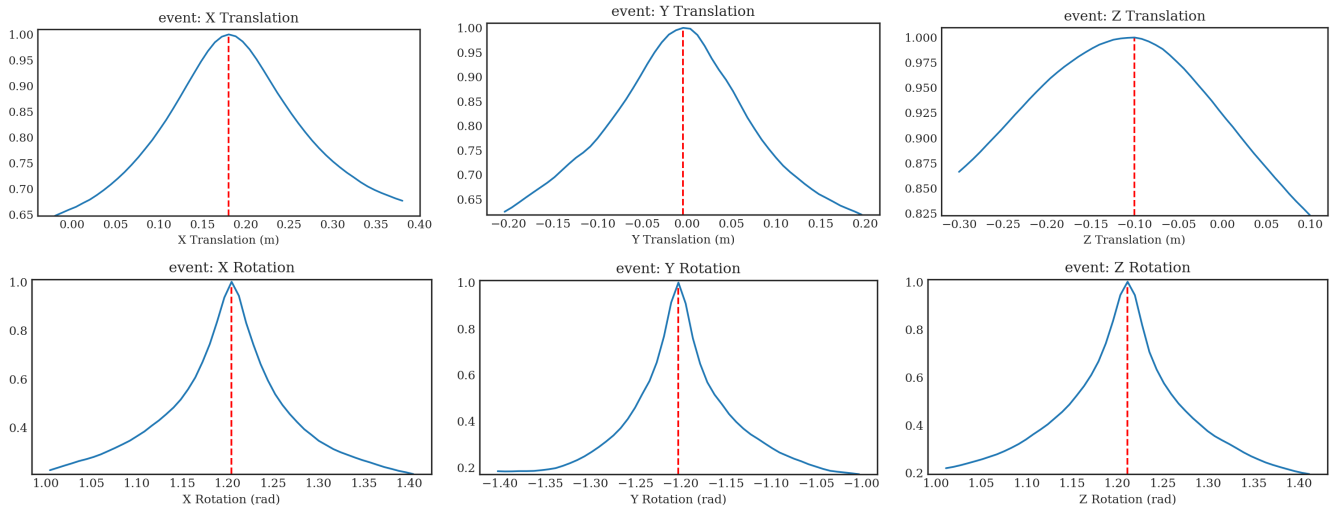

Fig. 6. Normalized cost function optima when exploring the cost landscape around the found optimal calibration parameters.

Table V. Camera intrinsic calibration distortion results.

Camera	K_1	K_2	P_1	P_2	K_3
TRI054	-0.3553	0.1638	0.0002	0.0001	-0.0414
Gen4.1	-0.4408	0.2570	0.0012	0.0003	-0.0903

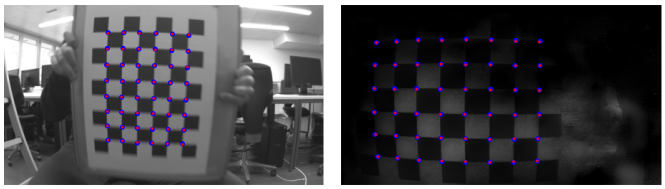

Fig. 7. Images showing the poor quality of certain calibration images for RGB (left) and event (right) cameras.

be a result of poorly illuminated environments that increase the exposure time or gain compensation, exacerbating the motion blurring issues. Figure 7 shows examples of poor quality images that exhibit high reprojection error that were used in the calibration process.

Other potential sources of error include the use of non-flat or warped patterns on the target. These could result from the pattern being printed on a non-flat surface, blurring issues due to lighting or lens focus, or the printing and application process of the pattern causing poor orthogonality or uniformity in the final pattern. As the target used in this initial calibration process was a printed A4-size paper target taped onto the side of an old computer case, the integrity of the pattern is potentially suspect.

To address the previous point regarding poor target quality, we generated new calibration results using a commercial metrology-grade chessboard target from https://calib.io/. The chessboard is an 800 mm x 600 mm target with a 12x9 (11x8 valid points) chessboard pattern. These targets are constructed from aluminum composite materials that exhibit a high degree of environmental stability. The tolerances on these targets are specified to be within 0.1mm + 0.3mm/m (at 20°C). While the chessboard target was replaced, the lighting conditions were similarly poor. We see significant improvements to the reprojection error for both the RGB camera and event camera with this target. The RGB camera calibration used 80 images while the event camera calibration used 71 images. Table VI shows the re-calibrated camera model properties while Table VII shows the re-calibrated distortion parameters.

Table VI. Re-calibrated camera intrinsic calibration pinhole results.

Camera	RP Error	F_x	F_y	C_x	C_y
TRI054	**0.263**	1215.16	1215.58	931.47	541.55
Gen4.1	**0.667**	1043.98	1044.39	620.35	343.76

Table VII. Re-calibrated camera intrinsic calibration distortion results.

Camera	K_1	K_2	P_1	P_2	K_3
TRI054	-0.3657	0.1945	-0.0004	0.0001	-0.0638
Gen4.1	-0.4558	0.2994	0.0001	0.0001	-0.1391

B. Camera-Lidar Mutual Information Formulation

With accurate intrinsic calibrations, we can complete extrinsic calibration with both cameras using the MI optimization framework. For well-posed optimization, the objective function is ideally smooth and convex to allow for numeric gradient or shooting methods to converge towards a global optimum. As described in the methodology section, we used smoothed intensity distributions as proxies for our probability distributions. When we explore the cost landscape around the optimal calibration results, we find the solution lies within a local optima as expected for the MI optimization process.

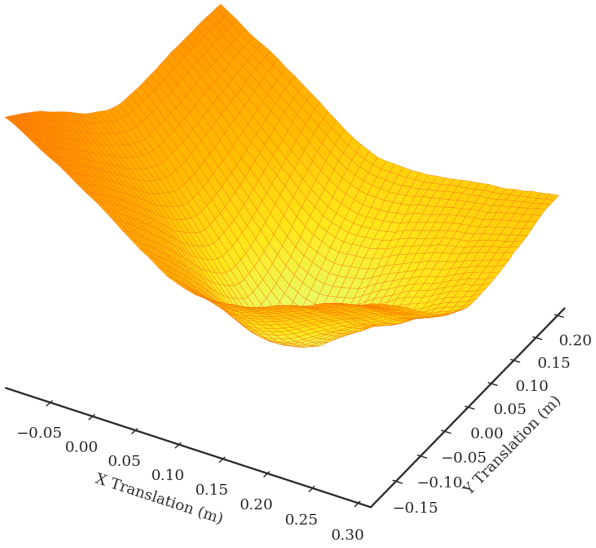

Fig. 8. Cost surface with 28 scans in the Garage #2 subset displaying a clear optima when varying the x and y position from the optimally found parameters.

Figure 8 shows the clear local optima when we vary two of the positional axes.

C. Camera-Lidar Calibration

As mentioned in the methodology, we use the axis-angle parameterization to avoid singularity configurations in Euler parameterization where two parameters may not be fully independent. This parameter independence leads to a more convex optimization space and improves the repeatability of the calibration results.

1) Noise-free Full Calibration: We present the extrinsic parameters when calibrating across all 93 scenes with the seed values determined from the CAD model. We find that the unbounded methods, "CG" and "BFGS," deviate wildly from the seed calibrations for the RGB camera-lidar calibration and their results have been omitted. For the RGB frame camera, we find that the positional calibration results are within 5 to 10 mm, when we exclude the "Nelder-Mead" results, which have poor positional calibration performance. The rotation components are within 0.002 radians or 0.1°. The event camera-lidar calibration achieves more similar results with positional parameters varying by 2 to 5 mm (again excluding "Nelder-Mead") and rotational parameters within 0.001 radians or 0.05°. A full summary of these calibration results can be found in Table VIII.

2) Optimizer and Scene Robustness: The axis-angle representations appears to work reasonably well with most of the optimizers. To ensure robustness, however, we perform two main experiments using the axis-angle representation. We test each optimizer with 40 repeated runs of 40 randomly sub-selected scenes for these two experiments. The first experiment uses the CAD model seed values for initial parameters for calibration whereas the second experiment induces an additional

Table VIII. Extrinsic calibration results for transforming lidar points into the respective camera frames with different optimizers. NM - "Nelder-Mead", Po = "Powell", BF - "BFGS", LB - "L-BFGS-B", CG - conjugate-gradient, SL - "SLSQP".

Camera	x (m)	y (m)	z (m)	v_1	v_2	v_3
TRI054 NM	0.1063	0.0000	-0.0512	1.223	-1.210	1.187
TRI054 Po	0.1009	0.0019	-0.0163	1.224	-1.208	1.188
TRI054 LB	0.1042	0.0069	-0.0216	1.225	-1.210	1.187
TRI054 SL	0.1053	0.0106	-0.0210	1.225	-1.210	1.187
Gen4.1 NM	0.1851	0.0001	-0.0508	1.204	-1.207	1.214
Gen4.1 Po	0.1835	-0.0028	-0.0335	1.203	-1.207	1.214
Gen4.1 BF	0.1849	-0.0041	-0.0317	1.203	-1.207	1.214
Gen4.1 LB	0.1852	-0.0034	-0.0317	1.203	-1.207	1.214
Gen4.1 CG	0.1829	-0.0044	-0.0326	1.204	-1.207	1.214
Gen4.1 SL	0.1878	-0.0023	-0.0340	1.203	-1.208	1.214

stochastic noise in the range of ± 5 cm or ± 0.05 rad to each parameter.

In the RGB camera-lidar calibration, we find the conjugate-gradient and default "BFGS" algorithms, two unbounded methods, often perform poorly in all parameters in the noise-less optimization and converge to parameters far from the true solution. The "Nelder-Mead" simplex refinement method works poorly for optimizing translation parameters, often staying close to the seed parameters. The "Powell" shooting method performs well in optimizing to repeatable values regardless of initialization, but tends to have the highest run-time for convergence.The "L-BFGS-B" algorithm and the "SLSQP" method both exhibit more repeatable results in the noiseless and noise-induced experiments. The "SLSQP" results in more stable parameterization in the noiseless experiment, but can diverge from the repeated results in the noise-induced experiment.

Conversely, we find that the event camera-lidar calibration exhibits a high degree of stability and repeatability with a wide range of optimizers. The exception in this case is the "Nelder-Mead" method which exhibits the same issues regarding the convergence of translation parameters where it tends to remain close to the initial seed values. Figure 9 shows the parameter optimization for the event camera for both the noiseless and noise-induced experiments.

In terms of calibration repeatability, we analyzed the results of the noise-free experiments with the different optimizers. Specifically, we calculate the mean result and the associated standard deviation of the 40 experiments where 40 scenes are sub-sampled from the 93 total scenes to look at the process robustness. We have excluded the "CG" and "BFGS" methods, two unbounded methods, for the RGB camera due to their wild divergence from the seed calibration skewing the results heavily.

The RGB results show good repeatability in translation with standard deviations between 0.3-0.9 cm, excluding the "Nelder-Mead" results which we have previously shown to have poor positional optimization. The rotational standard deviations tend to be below 0.002 radians or within 0.1°. We

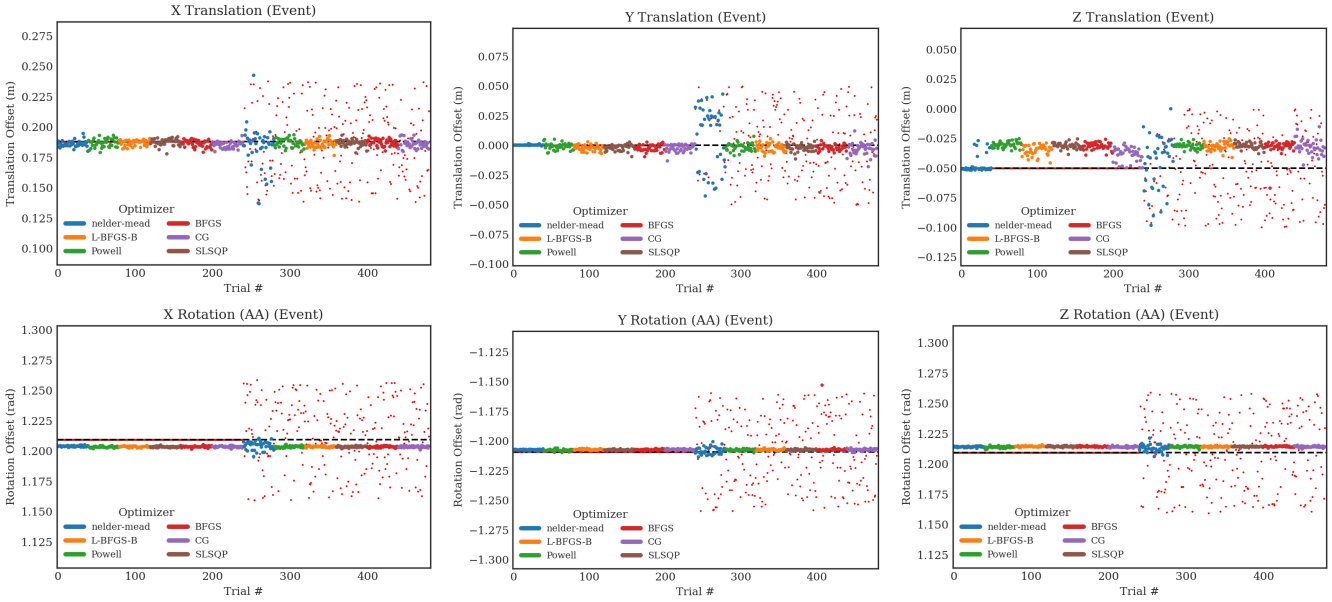

Fig. 9. Optimized extrinsic parameters for the event camera-lidar calibration under the axis-angle representation.

find that the event camera-lidar calibration exhibits excellent repeatability with translational standard deviations typically below 0.3 cm and rotational standard deviations around 0.0005 radians or around 0.03°. Table IX summarizes the full repeatability results.

D. MI Optimization Computational Analysis

A key consideration for calibration at-scale is the computational load of calibration methods. We explore the difference in computational speed depending on the optimizer choice and the number of scenes used in calibration.

1) Optimizer Choice: Between the six evaluated optimization methods, we explore their computational performance under similar conditions. We ran each optimizer 40 times with 40 randomly selected scenes used in the optimization process. "SLSQP" runs the fastest of the six, followed by "Nelder-Mead", "L-BFGS-B", "BFGS", "CG", and "Powell" respectively. Many of the optimizers show high degrees of variance based on the randomized initial seeds for these experiments. The "Powell" shooting method, in particular, shows significant run-time variation for its optimization. Figure 10 shows the comparison between different optimizers.

These runs were completed on a computational cluster equipped with Intel® Xeon® Processor E5 CPUs with clock-speeds ranging from 2.1 to 2.7 GHz. With a 40 scene calibrations taking on average ~100 seconds with "SLSQP" optimization and ~300 seconds with "L-BFGS-B" optimization, these computational times are within reasonable limits for acquiring accurate calibration.

2) Time Complexity: The optimization process was also partially completed locally on a laptop (Dell XPS 13 9350) with an Intel® Core™ i7-6560U CPU@2.20GHz x4 and 8 GB RAM. Figure 11 graphs the optimization computational time for the "Nelder-Mead" method using simplex refinement as a function of the number of processed scenes.

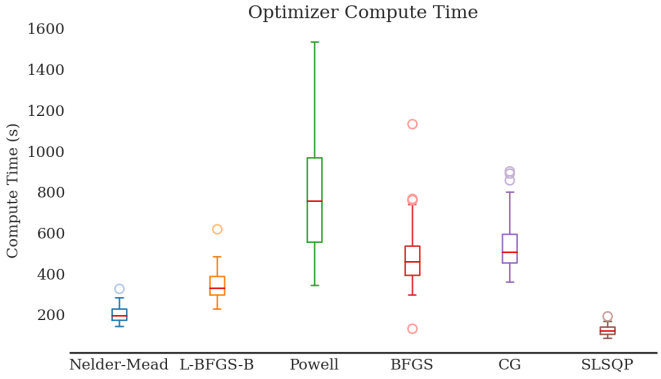

Fig. 10. Comparison of the average compute time between different optimizers.

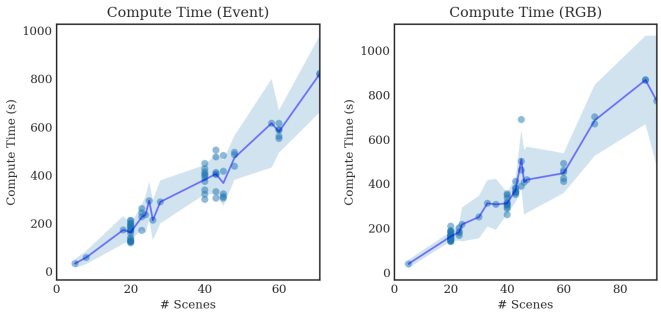

Fig. 11. Compute time showing an approximately linear O(n) scaling with of scenes.

The function to acquire the mutual information scales

Table IX. Extrinsic calibration mean parameters in the noise-free calibration experiments with the standard deviations reported in brackets. Angles reported in radians. NM - "Nelder-Mead", Po = "Powell", BF - "BFGS", LB - "L-BFGS-B", CG - conjugate-gradient, SL - "SLSQP".

Camera	x (m)	y (m)	z (m)	$v_1(rad)$	$v_2(rad)$	$v_3(rad)$
TRI054 NM	0.10831 (0.00083)	0.00002 (0.00005)	-0.05057 (0.00058)	1.22333 (0.00084)	-1.20969 (0.00103)	1.18790 (0.00120)
TRI054 Po	0.10284 (0.00271)	0.00682 (0.00477)	-0.01617 (0.00393)	1.22470 (0.00089)	-1.20932 (0.00165)	1.18711 (0.00175)
TRI054 LB	0.10164 (0.00599)	0.00842 (0.00511)	-0.02761 (0.00861)	1.22457 (0.00110)	-1.21005 (0.00143)	1.18587 (0.00213)
TRI054 SL	0.10211 (0.00328)	0.00748 (0.00360)	-0.01686 (0.00602)	1.22439 (0.00065)	-1.20963 (0.00136)	1.18619 (0.00170)
Gen4.1 NM	0.18592 (0.00259)	0.00021 (0.00033)	-0.04757 (0.00718)	1.20391 (0.00052)	-1.20760 (0.00051)	1.21410 (0.00045)
Gen4.1 Po	0.18678 (0.00334)	-0.00092 (0.00273)	-0.03040 (0.00242)	1.20321 (0.00056)	-1.20790 (0.00077)	1.21385 (0.00072)
Gen4.1 BF	0.18607 (0.00279)	-0.00236 (0.00267)	-0.03171 (0.00298)	1.20351 (0.00064)	-1.20739 (0.00045)	1.21430 (0.00044)
Gen4.1 LB	0.18678 (0.00228)	-0.00228 (0.00232)	-0.03509 (0.00426)	1.20352 (0.00054)	-1.20739 (0.00050)	1.21447 (0.00044)
Gen4.1 CG	0.18512 (0.00263)	-0.00238 (0.00298)	-0.03862 (0.00579)	1.20346 (0.00065)	-1.20736 (0.00057)	1.21403 (0.00056)
Gen4.1 SL	0.18744 (0.00280)	-0.00206 (0.00243)	-0.03130 (0.00240)	1.20336 (0.00054)	-1.20763 (0.00054)	1.21435 (0.00050)

linearly with the number of projections of the lidar scan into an image and the subsequent calculation of the intensity histogram counts for each projected point, achieving an O(n) scaling with respect to the number of scenes. Knowing the linear scaling of the optimization process based on the number of scenes, informed trade-offs can be made depending on the speed-accuracy requirements of the calibration process.

E. Stereo Calibration

Several scenes in the "Chessboard #1" and "Chessboard #3" contain co-visible chessboard images between the RGB and event camera. The 3D pose can be extracted from these chessboards and their extrinsic calibrations determined from the co-visible 3D points. This generates an independent calibration that can be used to compared to the camera-lidar approach in the previous section.

In the event camera image reconstruction, the illumination effect is driven by the lidar laser returns. As the MEMS lidar has a vertical FoV of 25°, only the middle band of the event camera is illuminated, as the recti-linearized event camera frame has a FoV of 39°. Consequently, we cannot place the chessboard pattern for calibration in an arbitrary position, but in positions that are within the full FoV of the RGB camera and the applicable FoV of the event camera.

Stereo calibration results are compared to the MI calibration, generated from using all the available scenes and optimized with the "L-BFGS-B" optimizer, in Table X. We see that the calibration results differ by approximately 5 mm in the horizontal and depth direction. The calibration result differs by approximately 15 mm in the y direction, potentially due to the limited vertical range in the camera resolution and in the positioning of the target. The absolute angular difference between the two configurations, as calculated from Eq. (12), is 0.18°.

$$\theta = \arccos \frac{\operatorname{tr} R - 1}{2}, \qquad \text{where } R = R_1 R_2^T \tag{12}$$

The calibration process reports a reprojection error of 0.6007[1]. This reprojection error is quite high compared to typical target values for reprojection error. As described in the intrinsic calibration process, the same issues are potential causes for the high reprojection error including poor pattern integrity and poor lens focus leading to blurring. We also compare the reprojection error calculated from taking the pose estimate from only the RGB camera frame and reprojecting with the stereo and the MI-based extrinsic parameters into the event camera space. We find that this process retrieves an error of 1.03 pixels for the stereo calibration and 1.47 pixels for the MI-based calibration.

Table X. Extrinsic calibration results between the RGB and event camera and the associated reprojection error from the RGB image to the event image.

Method	RGB→ E (px)	x (m)	y (m)	z (m)	|v| (rad)
Stereo	**1.0294**	0.0745	0.0033	-0.0177	0.0257
MI	**1.4746**	0.0807	-0.0104	-0.0120	0.0286

F. Stereo-constrained Camera-Lidar Calibration

As shown in the previous section, the stereo calibration and the individual camera-lidar are similar, but not fully consistent. To address this consistency issue, we perform a joint optimization of both camera-lidar calibrations while maintaining the stereo calibration as a constraint. The results in the calibrations for the RGB camera are found in Table XI.

G. Radar-Lidar Calibration

Radar calibration is the final step for the full set of extrinsic calibrations. The spinning radar can capture direct depth information and is known to be more robust to inclement weather conditions.

The radar-lidar calibration process uses the "Garage #2" sub-dataset which contained varied unstructured scenes. The sub-dataset contains 28 scene pairs captured in an indoor environment. The indoor environment poses a number of

[1]This value is calculated from the optimized pose that minimizes the joint reprojection error from both cameras as opposed to a pose estimate generated from just one camera.

Table XI. Extrinsic calibration results for joint optimization of both cameras. Angles reported in radians. NM - "Nelder-Mead", Po = "Powell", BF - "BFGS", LB - "L-BFGS-B", CG - conjugate-gradient, SL - "SLSQP".

Camera	x (m)	y (m)	z (m)	v_1	v_2	v_3
TRI054 NM	0.1083	0.0001	-0.0494	1.219	-1.214	1.196
TRI054 Po	0.1201	-0.0010	-0.0058	1.218	-1.216	1.197
TRI054 BF	0.1187	-0.0014	-0.0045	1.218	-1.216	1.197
TRI054 LB	0.1223	-0.0008	-0.0086	1.217	-1.216	1.197
TRI054 CG	0.1208	-0.0005	-0.0010	1.218	-1.216	1.197
TRI054 SL	0.1219	-0.0003	-0.0038	1.218	-1.216	1.197

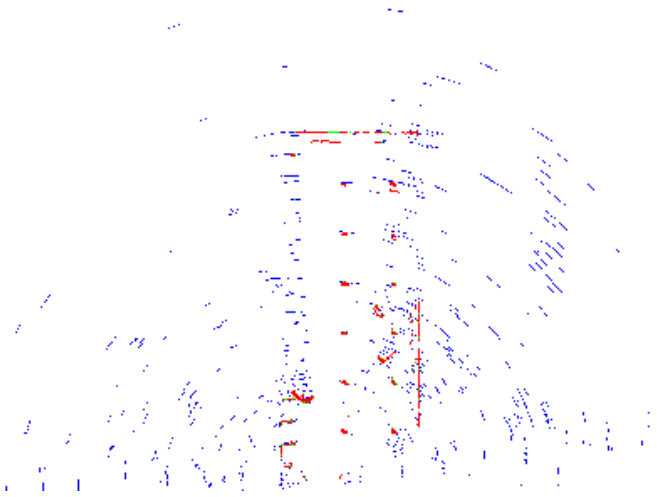

Fig. 12. Bird's-eye-view (BEV) of the sliced lidar returns and the filtered radar returns. Lidar returns are in red, radar returns in blue, and co-observed points in green.

challenges. Indoor scenes tend to exhibit high degrees of noise in the radar returns due to the returns that rebound off of the walls and ceilings of the enclosed space. To account for the high degree of noise in the scene, we set the feature detector thresholds higher than for outdoor scenes to try and isolate true landmarks or key-points from the noise. Figure 12 shows the overhead view of the result of the radar-lidar calibration.

The calibration process determines the yaw rotation of the lidar to the radar frame to be 0.868±0.293°. We also conducted repeatability studies where we perform 300 trials where we randomly sub-select calibration from 5 scenes, 10 scenes, and 15 scenes. We find that the calibrations achieve 0.876±0.116°, 0.864±0.073°, and 0.870±0.054° respectively in these repeated trials. With 10 scenes, we see repeatability of results within 0.1°, which is in line with what [3] found in their calibrations with 10 scenes.

For the positional calibration values, we measured these values in both the CAD model and in the as-built configuration. The values of both measurements are listed in Table XII.

V. Discussion

Calibrating an autonomous vehicle sensor suite requires many considerations due to the different sensing modalities unique to each sensor. Due to the sensor resolution, there may

Table XII. Positions of the radar relative to the lidar in the CAD model and as measured.

Camera	x (m)	y (m)	z (m)
CAD	-0.40	0.15	0.20
As-built Measurement	-0.38	0.17	0.24

be practical limits to the accuracy of achievable calibration. Regardless, accurate calibration remains a pivotal aspect for multi-modal perception in robotics and in other autonomous systems. The effect of poor calibration can be seen in the pre-calibrated images seen in Figure 13.

For camera-lidar calibration, mutual information provides a well-motivated approach for calibration in loosely structured scenes. This method is applicable to scenes without targets, but may also exploit standard targets like chessboards to extend calibration with stereo methods. When multiple calibration methods are employed, we can leverage these different calibration methods to build confidence in calibration results within the multi-sensor framework.

A. Event Camera-Lidar Geometric Calibration

From the different calibration methods, we found that event camera-lidar calibration to express the greatest stability and accuracy. MI approaches are particularly applicable in the calibration of event cameras and lidars where the event camera's high dynamic range can directly capture the lidar laser returns, resulting in a direct correlation between the illumination source and the returned signal for both sensors. Additionally, this approach allows for the system to decouple the geometric and temporal calibration between these two sensors, permitting the calibration of an event camera without complicated motion triggering of the event camera to detect edges or other features for alignment.

B. Stereo Constraint Extensions

The use of stereo calibration as a hard constraint to jointly optimize multi-camera-lidar calibration is a simple extension of the optimization without introducing additional optimizable parameters. However, a joint objective function which incorporates both the MI score and the stereo reprojection error would allow for the calibration process to perform a more complete optimization of all parameters. This joint function could potentially be formulated from the a weighted sum of the RGB-lidar MI score, the event-lidar MI score, and the reprojection error for the stereo calibration. Another alternative would be to use the stereo constraint as a soft penalty which penalizes the calibration of the camera-lidar systems if their result deviates too far from the stereo calibration.

C. Impact of Intrinsic Calibration

All calibration of the camera systems inherently rely on the accuracy of the intrinsic calibration results. The intrinsic calibration has a significant impact on both stereo calibration, where typically the intrinsic values are fixed to reduce the

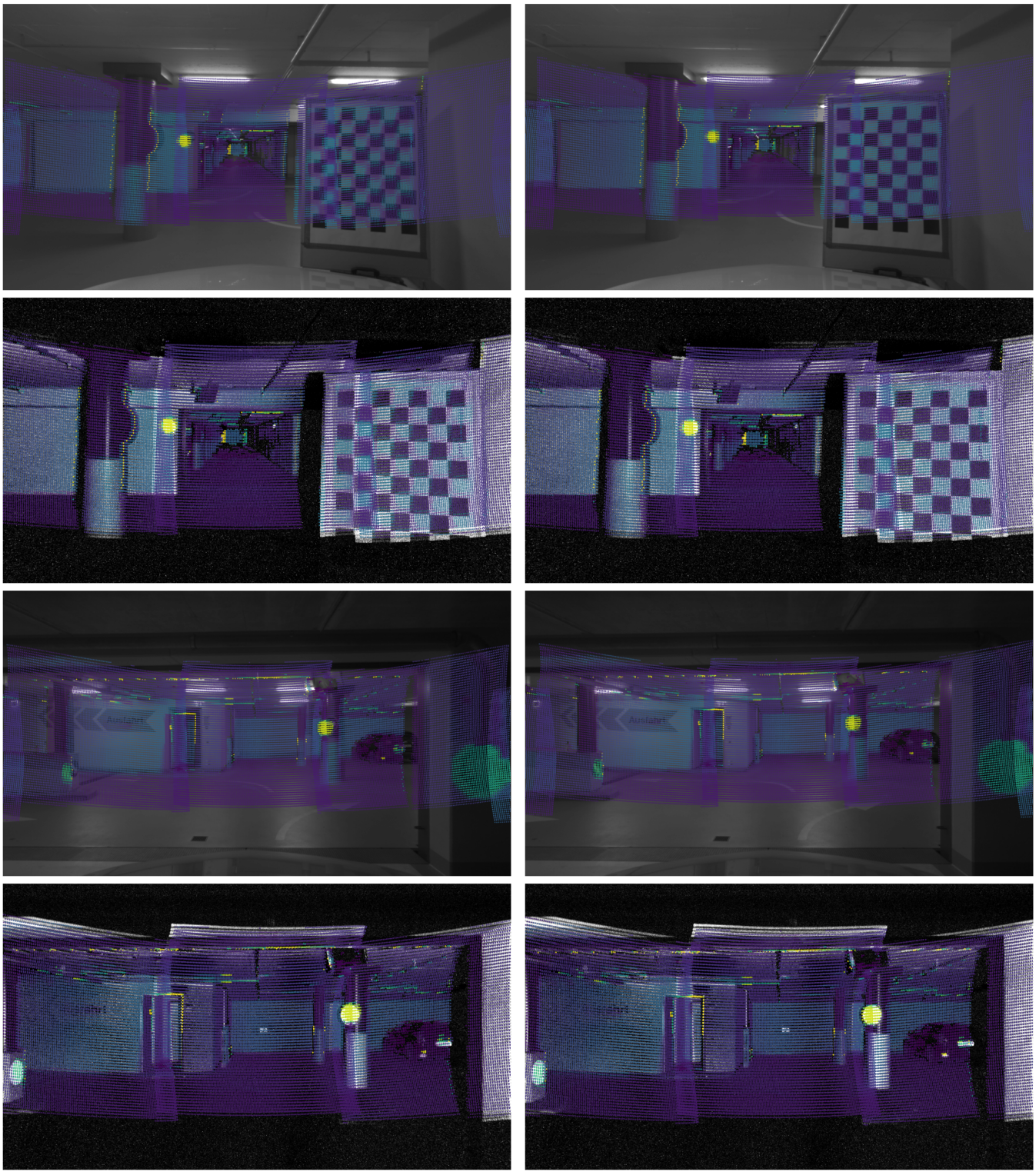

Fig. 13. Scene 47 (Top 4) and scene 77 (Bottom 4) showing RGB frame camera-lidar and event camera-lidar alignment before and after calibration. (Left) Pre-calibrated lidar point projections. (Right) Calibrated lidar point projections.

optimization space, and the camera-lidar calibrations where the intrinsics affect how points are projected into the 2D image. If the intrinsic calibration results are inaccurate, then the extrinsics will also converge to inaccurate results.

Consequently, the retrieval of accurate intrinsics is fundamental to accurate extrinsic calibration. To ensure accurate intrinsics, the quality of the hardware set-up is of significant importance. We note the significant improvements to the reprojection error when we transition from the initial target to the commercial target. As such, intrinsic calibration should carefully considered when calibrating sensors for the purpose of autonomous driving and in generalized robotics.

D. Optimizer Choice

In our experiments, we found that the "L-BFGS-B" algorithm typically had the best stability across optimizers in terms of result performance. This optimizer tended to be slower, however, than the "Nelder-Mead" simplex algorithm or the "SLSQP" algorithm. "SLSQP" offered a much faster optimization and led to stable results when seed calibrations were close to the true calibrations. "SLSQP," tended to be less robust to random seed noise in the calibration initialization.

Additionally, these optimizers were not fine-tuned to the process of calibration and their performance could be improved in both stability and computation speed. Extensive testing of termination criteria, bounding, step size, and other parameters for each optimizer was not performed. For example, the optimization used 3-step numeric differences for gradient calculations, which tends to be slower than 2-step numeric differences but is potentially more stable. Optimizer tuning could improve both the optimization time and result repeatability across different scene selections.

E. Radar-Lidar Calibration

The radar-lidar optimization was performed in an indoor environment with extensive noise due to the radar returns from the ceiling and ground elements. Repeating these results in an outdoor setting would limit the noise in the radar scans and and allow for greater confidence in the radar-lidar calibration. Additionally, the lidar is different from most other lidars used in perception datasets as it does not exhibit a full 360° horizontal FoV.

VI. Conclusion

In this report, we explore the complete calibration of four sensors with unique sensing modalities in the use of an autonomous driving dataset. We showcase an MI optimization-based calibration process for calibrating camera-lidar systems and a method extension to the stereo-lidar calibration case. Additionally, we show that lidars and event cameras have synergistic compatibility for calibration, enabling image reconstruction from events triggered by the lidar lasers. The calibration process incorporates both information from the unstructured scene and potentially from structured targets, allowing for calibration flexibility. We also showcase a radar-lidar calibration performed using frequency space information for yaw rotation calibration.

References

[1] P. An, T. Ma, K. Yu, B. Fang, J. Zhang, W. Fu, and J. Ma. Geometric calibration for LiDAR-camera system fusing 3D-2D and 3D-3D point correspondences. *Optics Express*, 28(2):2122, 2020.

[2] D. Barnes, M. Gadd, P. Murcutt, P. Newman, and I. Posner. The Oxford Radar RobotCar Dataset: A Radar Extension to the Oxford RobotCar Dataset. *Proceedings - IEEE International Conference on Robotics and Automation*, pages 6433–6438, 2020.

[3] K. Burnett, D. J. Yoon, Y. Wu, A. Z. Li, H. Zhang, S. Lu, J. Qian, W.-K. Tseng, A. Lambert, K. Y. K. Leung, A. P. Schoellig, and T. D. Barfoot. Boreas: A Multi-Season Autonomous Driving Dataset. *arXiv*, 2022.

[4] S. H. Cen and P. Newman. Precise Ego-Motion Estimation with Millimeter-Wave Radar under Diverse and Challenging Conditions. *Proceedings - IEEE International Conference on Robotics and Automation*, pages 6045–6052, 2018.

[5] J. Domhof, J. F. Kooij, and D. M. Gavrila. A Joint Extrinsic Calibration Tool for Radar, Camera and Lidar. *IEEE Transactions on Intelligent Vehicles*, 6(3):571–582, 2021.

[6] G. Gallego, T. Delbruck, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza. Event-Based Vision: A Survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):154–180, 2022.

[7] B. Gary. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 25(2236121):120–123, 2008.

[8] M. Gehrig, W. Aarents, D. Gehrig, and D. Scaramuzza. DSEC: A Stereo Event Camera Dataset for Driving Scenarios. *IEEE Robotics and Automation Letters*, 6(3):4947–4954, 2021.

[9] L. Heng. Automatic targetless extrinsic calibration of multiple 3D LiDARs and radars. *IEEE International Conference on Intelligent Robots and Systems*, pages 10669–10675, 2020.

[10] J. Kang and N. L. Doh. Automatic targetless camera–LIDAR calibration by aligning edge with Gaussian mixture model. *Journal of Field Robotics*, 37(1):158–179, 2020.

[11] C. L. Lee, Y. H. Hsueh, C. C. Wang, and W. C. Lin. Extrinsic and Temporal Calibration of Automotive Radar and 3D LiDAR. *IEEE International Conference on Intelligent Robots and Systems*, pages 9976–9983, 2020.

[12] W. Maddern, G. Pascoe, C. Linegar, and P. Newman. 1 year, 1000 km: The Oxford RobotCar dataset. *International Journal of Robotics Research*, 36(1):3–15, 1 2017.

[13] M. Muglikar, M. Gehrig, D. Gehrig, and D. Scaramuzza. How to calibrate your event camera. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1403–1409, 2021.

[14] J. Oh, K. S. Kim, M. Park, and S. Kim. A Comparative Study on Camera-Radar Calibration Methods. In *2018 15th International Conference on Control, Automation, Robotics and Vision, ICARCV 2018*, pages 1057–1062. IEEE, 2018.

[15] L. Oth, P. Furgale, L. Kneip, and R. Siegwart. Rolling shutter camera calibration. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1360–1367, 2013.

[16] G. Pandey, J. R. McBride, S. Savarese, and R. M. Eustice. Automatic Extrinsic Calibration of Vision and Lidar by Maximizing Mutual Information. *Journal of Field Robotics*, 32(5):696–722, 8 2015.

[17] J. Persic, I. Markovic, and I. Petrovic. Extrinsic 6DoF calibration of 3D LiDAR and radar. *2017 European Conference on Mobile Robots, ECMR 2017*, 2017.

[18] B. Srinivasa Reddy and B. N. Chatterji. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE Transactions on Image Processing*, 5(8):1266–1271, 8 1996.

[19] Z. Taylor and J. Nieto. A mutual information approach to automatic calibration of camera and lidar in natural environments. *Australasian Conference on Robotics and Automation, ACRA*, (December), 2012.

[20] D. J. Yeong, G. Velasco-hernandez, J. Barry, and J. Walsh. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, 21(6):1–37, 2021.

[21] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

APPENDIX A
EULER ANGLE PARAMETERIZATION

Another choice for interpretable rotation parameterization are ZYX Euler angles (α, β, γ). In this case, the full representation of the extrinsic transformations is $(x, y, z, \alpha, \beta, \gamma)$.

A. ZYX Euler Parameterization

Early experiments focused on using the ZYX Euler angle parameterization. However, the general outputs for optimization with ZYX Euler angles had mixed results. The "Nelder-Mead" method exhibited the greatest optimization stability, whereas other optimizers often terminated at the provided bounds. Figure 14 shows relative stability in the calibration results for the y-axis (pitch) rotation of the calibration.

In contrast, the x-axis and z-axis rotations displayed significantly greater amounts of variance in the optimization parameters. This was especially apparent in the RGB frame camera which typically exhibited higher variance than the event camera calibration. Figure 15 shows the greater variations in the optimized parameters in the case of the RGB frame camera.

This variance in the rotational angles arises due to the ill-posed nature of the optimization problem when parameterized with Euler angles in singular configurations. When we evaluate the distance between the rotation parameters by finding the rotational difference in their matrix forms, we see that each of the optimized parameters are much closer than when inspecting parameters separately. Figure 16 compares the angular difference magnitude to a median result in the set of optimized result, showing how although the variance in parameters is high, the resulting solutions are similar.

While this indicates that the solutions are similar, it may lead to potentially poor optimization stability due to the redundancy of two parameters creating a ridge structure in the optimization space. Figure 17 shows the ridge structure in the optimization space when we look at the 2D slice of the cost surface for Z and X rotations.

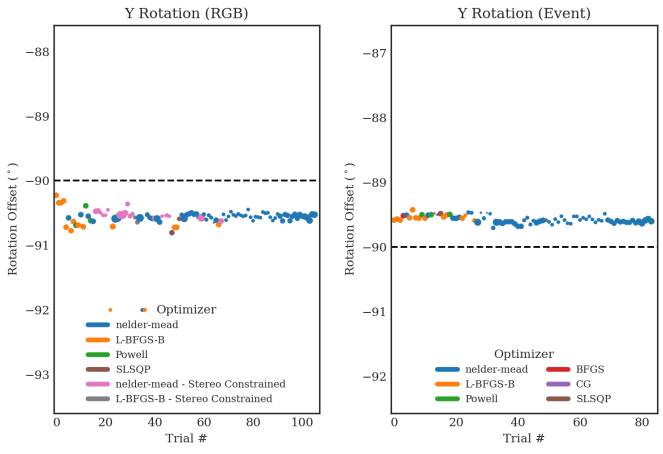

Fig. 14. Graphs showing y-axis (pitch) rotation optimization for both the RGB and event cameras. Dashed line shows the seed value and the size of the marker is proportional to the number of scenes used in the optimization.

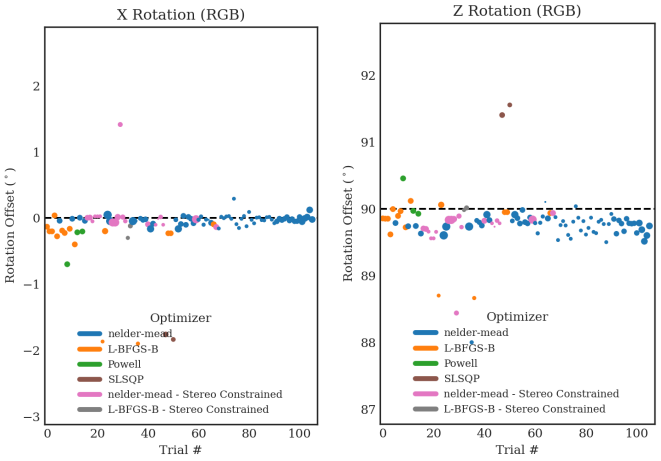

Fig. 15. Graphs showing x-axis (roll) and z-axis (yaw) rotation optimization. Dashed line shows the seed value and the size of the marker is proportional to the number of scenes used in the optimization.

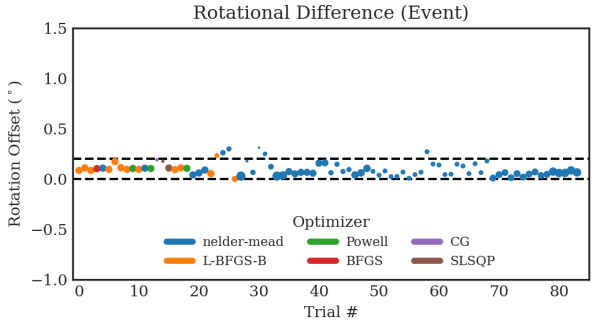

Fig. 16. Rotation difference to the median rotation result showcasing the angular similarity between different optimized parameters.

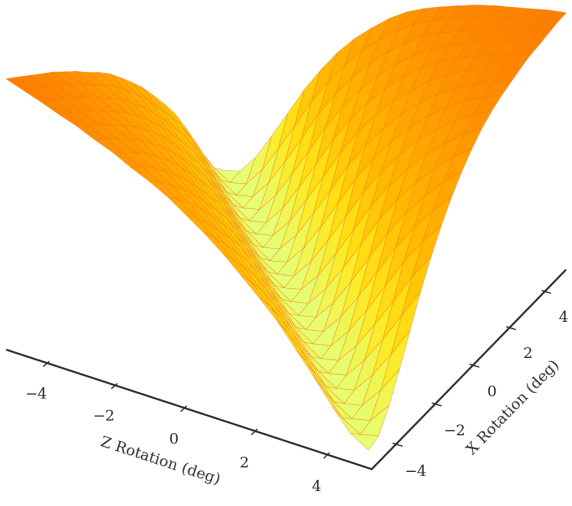

Fig. 17. Ridge structure (valley in minimization) in the value function between the Z and X rotation spaces.