

# L2E: Lasers to Events for 6-DoF Extrinsic Calibration of Lidars and Event Cameras

**Conference Paper****Author(s):**

Ta, Kevin; Brüggemann, David; Brödermann, Tim; Sakaridis, Christos; Van Gool, Luc

**Publication date:**

2023

**Permanent link:**

<https://doi.org/10.3929/ethz-b-000626204>

**Rights / license:**

[Creative Commons Attribution 4.0 International](#)

**Originally published in:**

<https://doi.org/10.1109/icra48891.2023.10161220>

# L2E: Lasers to Events for 6-DoF Extrinsic Calibration of Lidars and Event Cameras

Kevin Ta<sup>1</sup> David Bruggemann<sup>1</sup> Tim Brödermann<sup>1</sup> Christos Sakaridis<sup>1</sup> Luc Van Gool<sup>1,2</sup>

**Abstract**—As neuromorphic technology is maturing, its application to robotics and autonomous vehicle systems has become an area of active research. In particular, event cameras have emerged as a compelling alternative to frame-based cameras in low-power and latency-demanding applications. To enable event cameras to operate alongside staple sensors like lidar in perception tasks, we propose a direct, temporally-decoupled extrinsic calibration method between event cameras and lidars. The high dynamic range, high temporal resolution, and low-latency operation of event cameras are exploited to directly register lidar laser returns, allowing information-based correlation methods to optimize for the 6-DoF extrinsic calibration between the two sensors. This paper presents the first direct calibration method between event cameras and lidars, removing dependencies on frame-based camera intermediaries and/or highly-accurate hand measurements. Code: <https://github.com/kev-in-ta/l2e>

## I. INTRODUCTION

In the last decade, the development of many robotics applications—*e.g.* autonomous vehicles—has accelerated due to both academic and industrial research. The choice of sensors is pivotal for optimizing the perception of a robot’s surroundings. Besides frame-based camera and radar, lidar has become a core perception technology for autonomous robots, drones, and vehicles. By detecting the reflection of emitted laser pulses, lidars generate 3D point cloud representations of a scene.

More recently, event cameras have been explored as a potential alternative to traditional frame-based cameras, offering advantages in low-power and latency-demanding applications. Event-based vision derives from neuromorphic engineering, aimed at replicating fundamental, biological neural functions [1]. An event camera asynchronously extracts individual pixel-wise events that correspond to luminosity changes. This contrasts with traditional cameras that capture entire frames at regular intervals, even in fully static scenes.

To maximize performance and robustness, fundamental robot tasks such as perception, localization, and odometry often rely on multiple sensors, whose outputs are fused [2]. Importantly, multiple sensors offer complementary information, *e.g.* adverse weather conditions corrupt event camera and lidar signals to different degrees. This said, a key aspect of enabling these sensors is to accurately calibrate their intrinsic and extrinsic parameters. These parameters specify how each

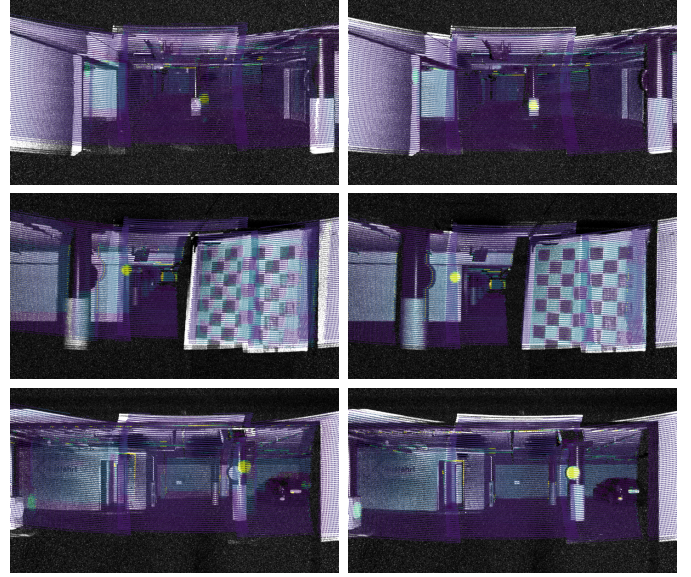


Fig. 1. (Left) Uncalibrated scene showing improperly projected lidar points. (Right) Calibrated scene using L2E showing improved lidar projection alignment. Accumulated event maps are shown in grayscale and lidar points are overlaid in colour.

modality represents the environment and how each sensor is positioned relative to the others [3]. Without calibration, sensor fusion methods incorrectly associate spatial features, as seen in Fig. 1, which negatively impacts downstream perception.

As event cameras represent a still immature technology, methods have yet to be fully explored for the direct extrinsic calibration between them and other perception sensors. Current event camera calibration methods involve confounding factors such as time synchronization for temporal correction, dynamic target handling introducing elastic deformations [4] and motion artifacts, dynamic scenes requiring ego-motion compensation, or calibration intermediaries introducing propagation errors. Instead, we argue that reliable and repeatable alignment of physical sensors benefits from *strong environmental controls*—such as static indoor scenes and temporal decoupling. Accordingly, we explore extrinsic lidar-event camera calibration in a fully static environment, thereby eliminating potential confounding factors.

Despite the increasing interest in event cameras, few large-scale multi-sensor datasets include event cameras to date. To the best of our knowledge, there exist only two datasets—both automotive—which contain both lidar and event cameras: MVSEC [5] and DSEC [6]. However, neither performs direct

This work was supported by the ETH Future Computing Laboratory (EFCL), financed by a donation from Huawei Technologies.

<sup>1</sup>Kevin Ta, David Bruggemann, Tim Brödermann, Christos Sakaridis, and Luc Van Gool are with the Computer Vision Lab, ETH Zurich, 8092 Zurich, Switzerland. [contact@kevinta.dev](mailto:contact@kevinta.dev)

<sup>2</sup>Luc Van Gool is with the Department of Electrical Engineering, KU Leuven, 3000 Leuven, Belgium.

calibration between event cameras and lidars.

In this paper we present L2E, a novel automatic pipeline using information-based optimization for calibrating a lidar with respect to an event-based camera. It leverages event-based structured light for static correspondence matching. In particular, we show how a dense lidar scan collected by a RoboSense RS-LiDAR-M1 can be easily and reliably correlated to a series of events registered by a Prophesee GEN4.1 event camera, and how we can achieve robust calibration using accumulated event maps. Our key contributions can be summarized as follows:

- 1) We present the first direct and temporally-decoupled 6-DoF calibration between an event camera and a lidar using accumulated event maps and lidar point clouds.
- 2) We propose a novel approach for synthesizing a static scene representation through accumulated event maps from the lidar-correlated event activity.
- 3) We investigate how scene selection and scene count affect calibration accuracy, repeatability, and computational performance.

## II. RELATED WORK

Neuromorphic sensors have rarely been integrated in the perception stack of autonomous robots. They suffer from a lack of maturity in neuromorphic technology. Yet, event cameras offer distinct advantages beyond the current suite of standard sensors. With their high temporal resolution and dynamic range, these sensors can operate in high-speed and over-saturated environments that may pose problems for frame-based cameras [1]. The event-based paradigm has the potential to reduce data bandwidth in static scenes without sacrificing the advantageous dense representation of imaging technology.

Recently, new datasets have emerged that incorporate event camera systems. There are two datasets containing event cameras and lidar sensors—both automotive. MVSEC [5] uses two DAVIS346 event cameras and a Velodyne VLP-16 lidar. The DAVIS346 camera captures a relatively low spatial resolution ( $346 \times 260$ ). DSEC [6] uses two Prophesee GEN3.1 event cameras and a Velodyne VLP-16 lidar. The Prophesee GEN3.1 operates at a higher spatial resolution ( $640 \times 480$ ) compared to the DAVIS346 camera, but at a lower spatial resolution than the current GEN4.1 sensors ( $1280 \times 720$ ).

For calibration, MVSEC attempted to use the grayscale image produced by the DAVIS346’s Active Pixel Sensor (APS) with the Camera and Range Calibration Toolbox [7]. The APS design is able to simultaneously act as a monochrome frame-based and event camera. However, the calibration results were found to be inaccurate, and the authors resorted to CAD measurements and manual adjustment for extrinsic parameters. DSEC uses the Prophesee event cameras which do not have the APS. Instead, rotation-only calibration is achieved using stereo methods and indirect calibration. Both datasets take fixed translation parameters from CAD models and solely calibrate for the 3-DoF rotation.

Table I. Specifications of the MEMS lidar and event camera.

Sensor	Resolution	HFoV	VFoV	Freq.
RS-LiDAR-M1	75k points	120°	25°	10 Hz
GEN4.1	1280 × 720	63°	38°	1.06 Gev/s

Intrinsic and stereo calibration of cameras is a well-established field with widely available tools such as OpenCV [8] or Kalibr [9]. Extensions of these methods have recently been made for event cameras where state-of-the-art event-to-video reconstruction is used to interpolate frames of moving checkerboard patterns for the purpose of intrinsic and stereo calibration [10]. DSEC uses this image reconstruction method to employ stereo calibration between the event and frame-based cameras using the Kalibr toolbox. The authors then perform extrinsic calibration between a frame-based camera stereo pair and a lidar sensor by performing modified point-to-plane ICP [11] from the stereo point cloud generated by SGM [12] and the lidar point cloud.

Calibration methods have been explored quite intensively between traditional cameras and lidars. These include automatic mutual information maximization schemes [13], [14], edge alignment methods [15], [16], hand-selected point correspondences [17], [18], and plane alignment methods [7], [15]. These methods vary from being completely unstructured, semi-structured [19], to using manufactured targets, such as standard planar checkerboards [20] or custom multi-modal targets [21]. These methods are designed with traditional camera images in mind and are not directly applicable to an event-based data stream.

Our basic observation is that lidar laser returns can trigger events, generating highly correlated signals in both sensors. Previous work has explored the use of event-based structured light in the visible spectrum using line or raster-pattern scans for 3D reconstruction [22], [23], [24] and stereo calibration [25]. We can naturally formulate the one-to-one correspondence matching under a mutual information framework, such as the one proposed by Pandey *et al.* [13]. Our work thus extends the use of mutual information frameworks from traditional cameras to event cameras.

## III. SENSORS OVERVIEW

### A. Characteristics

Our experimental sensor setup consists of a RoboSense RS-LiDAR-M1 MEMS lidar [26] and a Prophesee GEN4.1 event camera [27]. These sensors are mounted on top of a car, facing forward. Their nominal parameters are listed in Table I and the sensor installation is shown in Fig. 2.

The Prophesee GEN4.1 event-based vision system utilizes a state-of-the-art event vision sensor [28] and is capable of higher spatial resolution and smaller pixel sizes ( $4.86 \times 4.86 \mu\text{m}$ ) than the previous generation ( $15 \times 15 \mu\text{m}$ ) [29]. This modern event-driven sensor can operate between 5 and 100,000 lux, with a full dynamic range in excess of 110 dB.

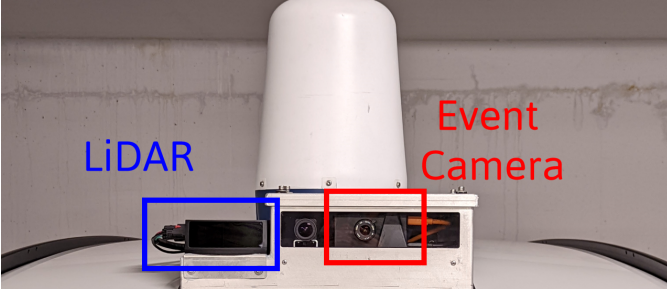


Fig. 2. Annotated image of the as-built sensor platform.

Such silicon-based image sensors are naturally sensitive to wavelengths up to 1000 nm [30]. Modern event cameras also utilize IR-correcting lenses as opposed to IR cut filters that would otherwise inhibit the sensor quantum efficiency (QE) and dynamic range. Additional improvement such as back-illumination also improve NIR sensitivity [31], [32]. As the most common industrial lidar wavelength is 905 nm due to its robustness in adverse conditions [33], most commercial lidars are capable of direct registration by modern event cameras.

The RS-LiDAR-M1 is a 905 nm MEMS lidar that captures a large frontal field-of-view (120° horizontal) while maintaining a very small profile. With an angular resolution of 0.2°, the generated point cloud densely overlaps the event camera view. The Prophesee GEN4.1 event camera can thus directly register the laser signals generated by the RS-LiDAR-M1.

#### B. Intrinsic Calibration

The RoboSense lidar is calibrated by the manufacturer and assumed to be accurate. For the event camera, we employ the intrinsic calibration process described in [10]. Accordingly, a moving target is captured as a series of events and then reconstructed as video frames with E2VID [34], [35]. The frames enable standard camera calibration using OpenCV [8]. The employed target is a commercially-available metrology-grade checkerboard, with a reported accuracy within 0.1 mm + 0.3 mm/m (at 20°C). We calibrate using the standard pinhole camera model and the Brown–Conrady distortion model for calibration. The pinhole model,  $\mathbf{K}$ , is described by the focal length ( $F_x, F_y$ ) and principal point ( $C_x, C_y$ ) parameters.

### IV. EVENT CAMERA-LIDAR EXTRINSIC CALIBRATION

In this section, our proposed event camera-lidar extrinsic calibration method—L2E—is described in detail. To represent extrinsic calibrations 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 the norm  $\|v\|$  is the angle of rotation in radians. Thus the full representation of the extrinsic transformation,  $\Theta$ , is given by  $(x, y, z, v_1, v_2, v_3)$ . The translation vector will be referred to as  $t$ . The expanded rotation matrix will be referred to as  $\mathbf{R}$ .

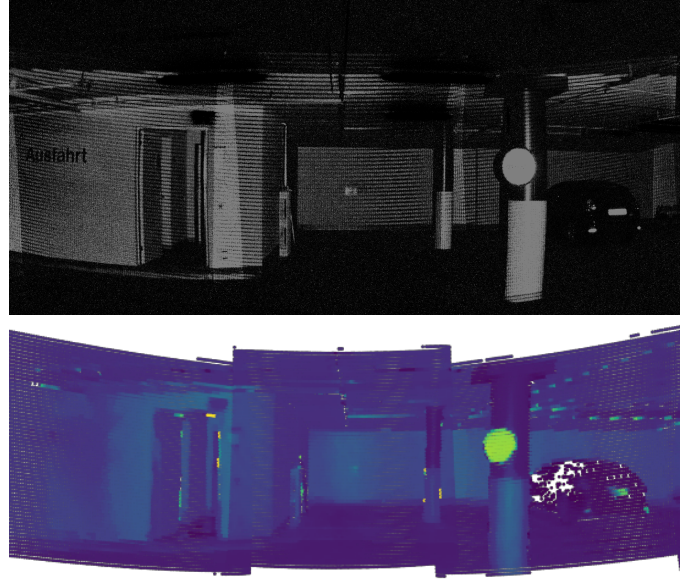


Fig. 3. (Top) Accumulated event map gathered over three seconds. (Bottom) Cropped projected lidar scans. The accumulated events in the static scene are correlated with the laser returns from lidar.

#### A. Accumulated Event Map

The events registered by the event camera are provided in the following format:  $(t_e, x, y, \pm)$ . Events are recorded with the time of the event, the location of the event, and the polarity of the event—*i.e.* whether the intensity has increased or decreased. Passing over an edge triggers a single positive (dark-to-bright) or negative (bright-to-dark) polarity event. In contrast, a laser pulse triggers both a positive and a negative event. As a lidar continuously scans a scene, we can simply increment the location for every triggered event, regardless of polarity, to get a corresponding accumulated event map,  $\mathcal{E}$ .

After accumulating events for 3 seconds, we clip the event map such that  $\mathcal{E}(x, y) \in [0, 127]$  to prevent scaling issues with abnormally high event activity at a pixel location. Additionally, we apply Gaussian smoothing to the accumulated event map for smoother optimization. An accumulated event map and its corresponding lidar projection are shown in Fig. 3.

#### B. Mutual Information

As events are directly triggered by the lidar returns, the correlation between the lidar point cloud and accumulated event map can be exploited with mutual information optimization.

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 in multiple ways, but we take the same entropy-based representation used in [13]. MI is defined in terms of the entropy of the random variables  $L$  and  $E$ , and their respective joint entropy  $H(L, E)$ :

$$MI(L, E) = H(L) + H(E) - H(L, E) \quad (1)$$

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  $L$  and  $E$ . We take the random variables  $L$  and  $E$  to be the lidar return intensities and the event activity in the accumulated event map, respectively. The entropies of random variables  $L$  and  $E$  and their joint entropy are described in Eq. (2, 3, 4) from their respective probabilities  $p_L$ ,  $p_E$ , and  $p_{LE}$ .

$$H(L) = - \sum_{l \in L} p_L(l) \log p_L(l) \quad (2)$$

$$H(E) = - \sum_{e \in E} p_E(e) \log p_E(e) \quad (3)$$

$$H(L, E) = - \sum_{l \in L} \sum_{e \in E} p_{LE}(l, e) \log p_{LE}(l, e) \quad (4)$$

### C. Probability Distribution Approximation

We approximate the probability distribution from the intensity and activity histograms of the lidar scan and accumulated event map respectively, as was done in [13]. Let  $\mathcal{P} = \{\mathbf{P}_i : i = 1, 2, \dots, n\}$  be the set of 3D points in the 3D scan and  $\{L_i : i = 1, 2, \dots, n\}$  be the set of intensity returns for each point in the set  $\mathcal{P}$ . The lidar points can be projected into the image space through the searched-for extrinsics  $(\mathbf{R}, \mathbf{t})$  and the known intrinsics  $\mathbf{K}$ .

$$\mathbf{p}_i = \mathbf{K}[\mathbf{R}|\mathbf{t}]\mathbf{P}_i \quad (5)$$

The location of the projected point,  $\mathbf{p}_i$ , is then used to retrieve the associated event activity in the accumulated event map image  $\mathcal{E}$  as shown in Eq. (6).

$$E_i = \mathcal{E}(\mathbf{p}_i) \quad (6)$$

Through accumulating 2D projected points that lie within the camera view, histograms  $\text{hist}_L$  and  $\text{hist}_E$  are generated from the discretized lidar intensities of each point and the event activity at the projected location. A joint 2D histogram is also generated from each intensity-activity pair. This process is detailed in Algorithm 1.

---

#### Algorithm 1 Raw Histogram Generation

---

**Input** 3D point cloud  $\mathcal{P}$  and 2D accumulated event map  $\mathcal{E}$

**Output** 1D lidar hist, 1D event hist, 2D joint hist

Initialize:  $\text{hist}_L$ ,  $\text{hist}_E$ , and  $\text{hist}_{LE}$

```

for  $(\mathbf{P}_i, L_i) \in \mathcal{P}$  do                                ▷ 3D point & intensity
   $\mathbf{p}_i \leftarrow \mathbf{K}[\mathbf{R}|\mathbf{t}]\mathbf{P}_i$                                 ▷ 2D projection
  if  $\mathbf{p}_i$  in image bounds then
     $E_i \leftarrow \mathcal{E}(\mathbf{p}_i)$ 
     $\text{hist}_L(L_i) \leftarrow \text{hist}_L(L_i) + 1$ 
     $\text{hist}_E(E_i) \leftarrow \text{hist}_E(E_i) + 1$ 
     $\text{hist}_{LE}(L_i, E_i) \leftarrow \text{hist}_{LE}(L_i, E_i) + 1$ 
  end if
end for

```

---

Subsequently, these raw histograms are normalized to acquire noisy approximations of the probability distribution functions  $\hat{p}_L$ ,  $\hat{p}_E$ , and  $\hat{p}_{LE}$ . The histograms are normalized by  $n$ , the total number of points that project into the valid

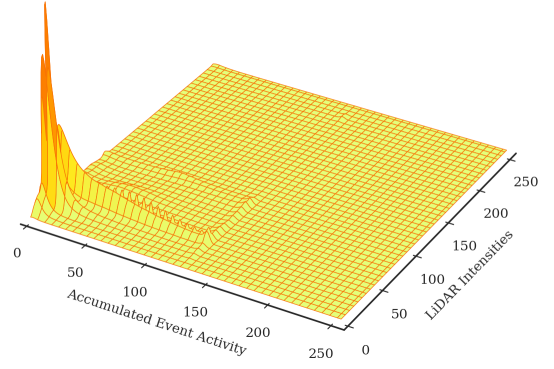


Fig. 4. Smoothed joint histogram as the approximate joint probability distribution between the accumulated event map and lidar intensities.

image region. To suppress noise, we perform a kernel density estimation (KDE) using Silverman’s rule-of-thumb [36]. In practice, we apply Gaussian blurring convolutions, a method that is nearly equivalent to kernel density estimation (KDE) in the case where histogram bins are spaced equivalently to the discretized point values. Fig. 4 shows the joint probability distribution of the lidar intensities and the accumulated event map values. The smoothed normalized histograms are finally used to get an MI estimate via Eq. (1)-(4).

### D. Optimization Formulation

With the formulated MI objective function and our approximations of the probability distributions, we formulate an optimization problem as follows:

$$\hat{\Theta} = \arg \max_{\Theta} \text{MI}(L, E; \Theta) \quad (7)$$

where  $\Theta = (x, y, z, v_1, v_2, v_3)$  and MI is maximized at the correct extrinsic parameters given an arbitrary set of scenes. Effective optimization of this function is aided by greater smoothness and convexity in the objective function.

The optimization process can be theoretically accomplished by any numeric minimization algorithm. In this work, we show that multiple optimizers are effective. In particular we find that bounded numeric gradient-based methods are the most effective. Bounding is particularly critical to prevent the ill-posed optimization space where few or no points are projected into the accumulated event map. We implement these processes using the SciPy library [37].

## V. EXPERIMENTS

### A. Extrinsic Calibration Dataset

Our calibration dataset consists of static scenes recorded over 3 seconds divided into two classes: checkerboard scenes and checkerless scenes. All scenes were collected in an underground garage. The checkerboard scenes were taken in a set position with a checkerboard target placed in close proximity to both the event camera and lidar. The checkerless scenes are captured in multiple locations without a checkerboard. Our dataset consists of 35 checkerboard scenes and 58 checkerless scenes for a total of 93 scenes.

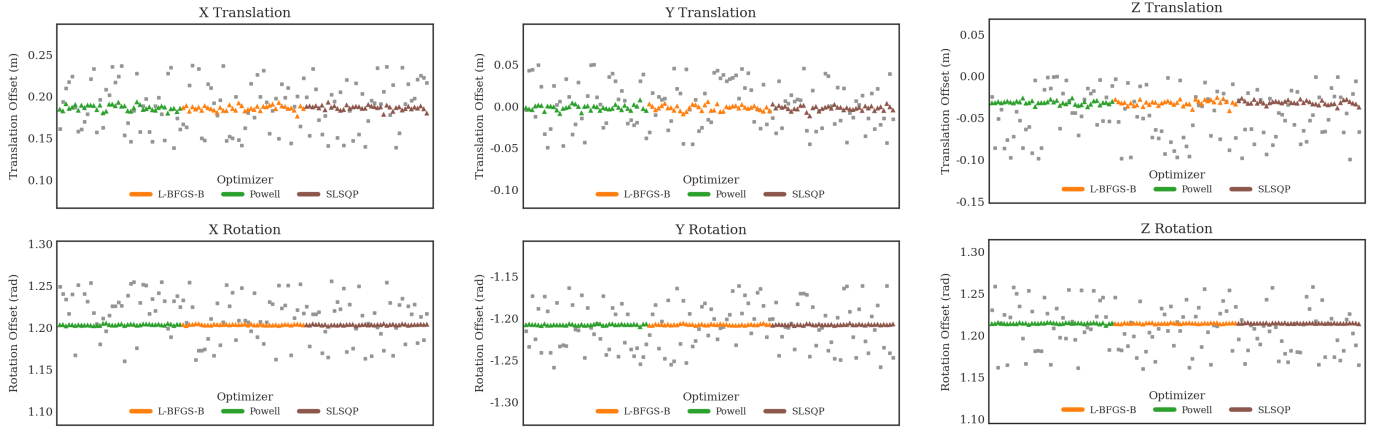


Fig. 5. Optimized extrinsic calibration parameters for the translation and rotation where grey squares are the initial seed values. Best viewed on a screen.

### B. Convergence Assessment

We perform noise-induced experiments to evaluate robustness and run-time of the calibration for different optimizers. Specifically, we calculate the mean and standard deviation results of 40 calibrations, where the measured parameters are induced with uniform noise in the range of 0.1 m and 0.1 radians. Unless otherwise specified, 40 scenes are respectively sub-sampled from the total of 93 scenes.

**Robustness Analysis.** Some tested optimizers failed to robustly converge. Specifically, the Nelder-Mead simplex refinement method performed poorly in converging to the consensus translation parameters. Unbounded methods also occasionally converged in the ill-posed portion of the optimization space. We find that bounded Broyden-Fletcher-Goldfarb-Shanno (L-BFGS-B), bounded Sequential Quadratic Programming (SLSQP), and the gradient-free Powell’s Method perform well. Fig. 5 shows the convergence robustness to seed errors for these optimizers. The convergence of calibration results in the presence of significant noise indicates a convergence basin robust against initialization errors. The rotation calibration is highly consistent, converging to results within a standard deviation of 0.0007 rad (0.04°) against uniformly induced noise of 0.1 rad (5.7°) for each axis. The translation calibration is also consistent, but exhibits a measurable standard deviation of 3 mm against the 100 mm of uniform noise induced in each translation axes.

**Computational Time Analysis.** We find that of the three optimizers, SLSQP performs significantly faster than L-BFGS-B and Powell’s Method. Powell’s Method takes up to an order of magnitude longer than SLSQP and multiple times longer than L-BFGS-B. We thus choose SLSQP as the default optimizer for L2E. A full summary is shown in Table II. We can also analyze the computational time in terms of the number of optimized scenes,  $m$ . More scenes lead to increased robustness, but longer computation. We can theorize that the optimization process has a complexity of  $O(m)$ , as the expected number of optimization steps should not depend on the scene count. This assumes that the underlying cost

Table II. Average and standard deviation of run-times across 40 runs.

Optimizer	Powell’s Method	L-BFGS-B	SLSQP
Time (s)	934.4 ± 251.2	363.0 ± 94.2	134.2 ± 25.2

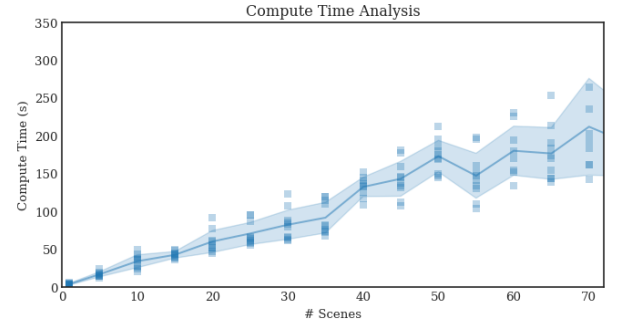


Fig. 6. Computation time as a function of the number of scenes used in SLSQP optimization.

landscape structure is independent of the number of scenes used as scene sampling is a random process. Consequently, the computation depends on the mutual information calculation, which must calculate the projection and mutual information for each individual scene. This calculation should scale linearly with the number of scenes used, which is observed in our experiments shown in Fig. 6.

**Impact of Number of Scenes.** One of the key components for optimization is the collection of sufficient scenes for optimization. We investigate the number of scenes required for consistent calibration in Fig. 7. As expected, increasing the number of scenes used in optimization decreases repeatability variance. The variance in calibration has substantially decreased when using 20 or more scenes in the optimization.

### C. Performance Comparison

**Checkerboard vs. Checkerless Scenes.** We perform repeated experiments using 20 sub-sampled scenes within the

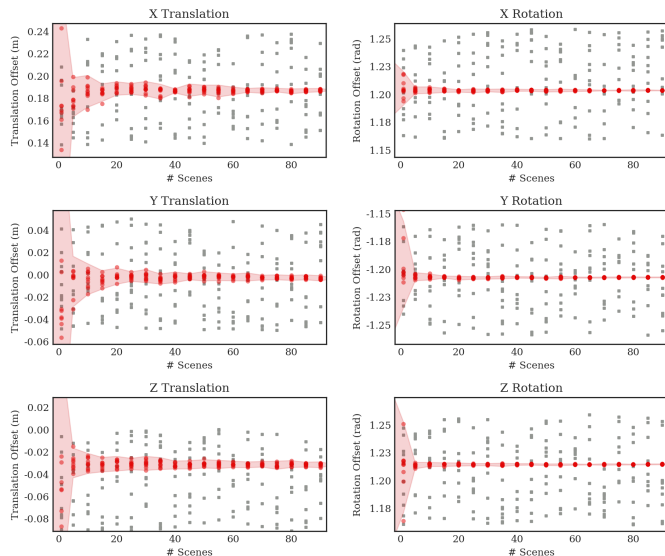


Fig. 7. Calibration results in red from SLSQP optimization for increasing number of scenes from left to right. Grey squares represent the initial seeds.

Table III. Averaged MI scores comparison between our calibration procedure (MI optimization), the two-stage baseline, and the uncalibrated results.

Calibration	Full Eval.	Checker Eval.	Checkerless Eval.
Ours (Mixed Set)	<b>0.44119</b>	0.53282	0.36729
Ours (Checker-only)	0.44082	<b>0.53421</b>	0.36550
Ours (Checkerless)	0.44064	0.53092	<b>0.36782</b>
2-Stage Baseline	0.41794	0.51664	0.33834
Uncalibrated	0.28767	0.33834	0.23873

full, checkerboard-only, and checkerless subsets. We evaluate the per-scene MI for held-out scenes. Results from these calibrations are shown in Table III, rows 1-3. We report marginally higher average MI on the full evaluation when using the checkerboard-only subset vs. checkerless-only subset.

**Two-stage Calibration Comparison.** We compare our method to a two-stage method that first stereo-calibrates the event camera to a standard camera intermediary requiring time-synchronized and motion-dependent event-to-video reconstruction [10] and then calibrates to the lidar using the Lidar-Camera Toolbox [15]. Our method achieves better visual alignment, as evidenced by Fig. 8, and better information agreement, as seen in Table III, row 4. The two stage-process propagates errors for each calibration process (lidar to camera, camera to event camera), leading to greater errors in the indirectly connected sensors.

## VI. DISCUSSION

**Overcoming time-synchronization and motion dependence.** A key benefit of our approach is overcoming challenges that make traditional methods of calibration impossible. Event-driven sensors are incapable of providing information in static scenes, thus demanding temporal synchronization and scene motion for extrinsic calibration. In contrast, our approach

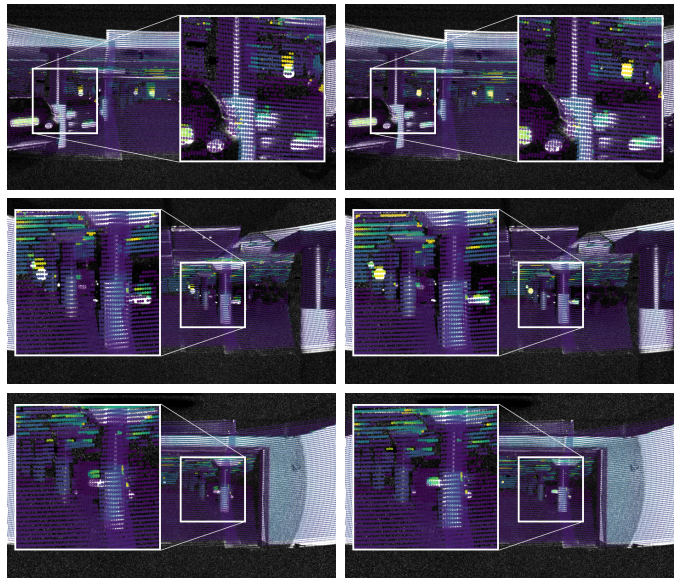


Fig. 8. (Left) Baseline. (Right) L2E calibration. Best viewed on a screen.

decouples the effects of time synchronization from geometric calibration.

**Considerations in Scene Selection.** L2E can be used in an unstructured environment with any combination of static scenes. However, we find that L2E tends to perform slightly better when observing high-texture objects—*i.e.* our checkerboards. This phenomenon is likely a result of the proximity of the checkerboard targets to both sensors, providing stronger rotation and translation constraints to the optimization. Poorer optimization in outdoor scenes was observed in [13], where errors were partially attributed to having fewer near-field 3D points in outdoor scenes. Consequently, the use of close-proximity objects with a reasonable amount of texture should have a beneficial effect on the calibration results. We note that precise geometric controls—*e.g.* checkerboard target tolerances—are not required, as opposed to traditional camera calibration methods.

**Managing NIR Sensitivity.** Our method relies on the registration of lidar laser signals by the event camera. However, NIR sensitivity may be undesirable during standard operation. To address this concern, the Prophesee GEN4.1 contains bias tuning parameters that can eliminate high frequency flicker effects, which we have found to be able to eliminate registration of lidar pulses.

## VII. CONCLUSION

This work presents L2E, the first direct extrinsic calibration method between event-based cameras and lidars using event-based structured light. L2E offers a flexible automatic approach to extrinsic calibration that leverages direct correlation between the lidar active signals and the corresponding registered events without the need for precise time synchronization or dynamic scenes. As event-driven vision technology continues to mature, we hope that our direct calibration method will enable further research on event cameras and sensor fusion.



## REFERENCES

- [1] 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*, 2022. 1, 2
- [2] S. Jusoh and S. Almajali, "A Systematic Review on Fusion Techniques and Approaches Used in Applications," *IEEE Access*, 2020. 1
- [3] D. J. Yeong, G. Velasco-hernandez, J. Barry, and J. Walsh, "Sensor and sensor fusion technology in autonomous vehicles: A review," *Sensors*, 2021. 1
- [4] A. Hagemann, M. Knorr, and C. Stiller, "Modeling dynamic target deformation in camera calibration," in *Winter Conference on Applications of Computer Vision (WACV)*. IEEE/CVF, 2022. 1
- [5] A. Z. Zhu, D. Thakur, T. Özasan, B. Pfrommer, V. Kumar, and K. Daniilidis, "The Multivehicle Stereo Event Camera Dataset: An Event Camera Dataset for 3D Perception," *IEEE Robotics and Automation Letters (RA-L)*, 2018. 1, 2
- [6] M. Gehrig, W. Aarents, D. Gehrig, and D. Scaramuzza, "DSEC: A Stereo Event Camera Dataset for Driving Scenarios," *IEEE Robotics and Automation Letters (RA-L)*, 2021. 1, 2
- [7] A. Geiger, F. Moosmann, O. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *International Conference on Robotics and Automation (ICRA)*. IEEE, 2012. 2
- [8] B. Gary, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2008. 2, 3
- [9] L. Oth, P. Furgale, L. Kneip, and R. Siegwart, "Rolling shutter camera calibration," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE/CVF, 2013. 2
- [10] M. Muglikar, M. Gehrig, D. Gehrig, and D. Scaramuzza, "How to calibrate your event camera," in *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE/CVF, 2021. 2, 3, 6
- [11] Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," in *International Conference on Robotics and Automation (ICRA)*. IEEE, 1991. 2
- [12] H. Hirschmüller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008. 2
- [13] 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*, 2015. 2, 3, 4, 6
- [14] Z. Taylor and J. Nieto, "A mutual information approach to automatic calibration of camera and lidar in natural environments," in *Australasian Conference on Robotics and Automation (ACRA)*. ARAA, 2012. 2
- [15] L. Zhou, Z. Li, and M. Kaess, "Automatic Extrinsic Calibration of a Camera and a 3D LiDAR Using Line and Plane Correspondences," in *International Conference on Intelligent Robots and Systems (IROS)*. IEEE/RSJ, 2018. 2, 6
- [16] J. Kang and N. L. Doh, "Automatic targetless camera-LIDAR calibration by aligning edge with Gaussian mixture model," *Journal of Field Robotics*, 2020. 2
- [17] D. Scaramuzza, A. Harati, and R. Siegwart, "Extrinsic self calibration of a camera and a 3D laser range finder from natural scenes," in *International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2007. 2
- [18] R. Unnikrishnan and M. Hebert, "Fast Extrinsic Calibration of a Laser Rangefinder to a Camera," Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-05-09, 7 2005. 2
- [19] 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*, 2020. 2
- [20] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000. 2
- [21] J. Domhof, J. F. Kooij, and D. M. Gavrila, "A Joint Extrinsic Calibration Tool for Radar, Camera and Lidar," *IEEE Transactions on Intelligent Vehicles*, 2021. 2
- [22] C. Brandli, T. A. Mantel, M. Hutter, M. A. Höpfinger, R. Berner, R. Siegwart, and T. Delbruck, "Adaptive pulsed laser line extraction for terrain reconstruction using a dynamic vision sensor," *Frontiers in Neuroscience*, 2014. 2
- [23] M. Muglikar, G. Gallego, and D. Scaramuzza, "ESL: Event-based Structured Light," in *International Conference on 3D Vision (3DV)*. IEEE, 12 2021. 2
- [24] N. Matsuda, O. Cossairt, and M. Gupta, "MC3D: Motion Contrast 3D Scanning," *International Conference on Computational Photography (ICCP)*, 2015. 2
- [25] J. N. Martel, J. Müller, J. Conradt, and Y. Sandamirskaya, "An Active Approach to Solving the Stereo Matching Problem using Event-Based Sensors," in *International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2018. 2
- [26] RoboSense, "RS-LiDAR-M1 Brochure EN," 2021. 2
- [27] Prophesee, "Prophesee Evaluation Kit 2 - HD Brochure," 2021. 2
- [28] T. Finateu, A. Niwa, D. Matolin, K. Tsuchimoto, A. Mascheroni, E. Reynaud, P. Mostafalu, F. Brady, L. Chotard, F. LeGoff, H. Takahashi, H. Wakabayashi, Y. Oike, and C. Posch, "1280x720 Back-Illuminated Stacked Temporal Contrast Event-Based Vision Sensor with 4.86µm Pixels, 1.066GEPS Readout, Programmable Event-Rate Controller and Compressive Data-Formatting Pipeline," in *International Solid-State Circuits Conference (ISSCC)*. IEEE, 2020. 2
- [29] Prophesee, "Prophesee Metavision Packaged Sensor Product Brief," 2019. 2
- [30] L. C. P. Gouveia and B. Choubey, "Advances on CMOS image sensors," *Sensor Review*, 2016. 3
- [31] P. Fereyre, M. Guilon, V. Prevost, F. Ramus, and T. Ligozat, "Back Illuminated System-on-Chip for Night Vision," in *International Symposium on Optoelectronics in Defence and Security (OPTRO)*. 3AF, 2012. 3
- [32] G. Taverni, D. Paul Moeys, C. Li, C. Cavaco, V. Motsnyi, D. San Segundo Bello, and T. Delbruck, "Front and Back Illuminated Dynamic and Active Pixel Vision Sensors Comparison," *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2018. 3
- [33] J. Wojtanowski, M. Zygmunt, M. Kaszczuk, Z. Mierczyk, and M. Muzal, "Comparison of 905 nm and 1550 nm semiconductor laser rangefinders' performance deterioration due to adverse environmental conditions," *Opto-Electronics Review*, 2014. 3
- [34] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "Events-to-video: Bringing modern computer vision to event cameras," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE/CVF, 2019. 3
- [35] —, "High Speed and High Dynamic Range Video with an Event Camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 3
- [36] B. W. Silverman, *Density Estimation for Statistics and Data Analysis Estimation Density*. London: Chapman & Hall/CRC, 1986. 4
- [37] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. P. Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G. L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, and Y. Vázquez-Baeza, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nature Methods*, 2020. 4