ONE TARGET TO ALIGN THEM ALL: LIDAR, RGB AND EVENT CAMERAS EXTRINSIC CALIBRATION FOR AUTONOMOUS DRIVING

Andrea Bertogalli*
DEIB
Politecnico di Milano
Milan, IT
andrea.bertogalli@mail.polimi.it

Giacomo Boracchi
DEIB
Politecnico di Milano
Milan, IT
giacomo.boracchi@polimi.it

Luca Magri
DEIB
Politecnico di Milano
Milan, IT
luca.magri@polimi.it

ABSTRACT

We present a novel *multi-modal extrinsic calibration* framework designed to simultaneously estimate the relative poses between event cameras, LiDARs, and RGB cameras, with particular focus on the challenging event camera calibration. Core of our approach is a novel 3D calibration target, specifically designed and constructed to be concurrently perceived by all three sensing modalities. The target encodes features in planes, ChArUco, and active LED patterns – each tailored to the unique characteristics of LiDARs, RGB cameras, and event cameras respectively. This unique design enables a one-shot, joint extrinsic calibration process, in contrast to existing approaches that typically rely on separate, pairwise calibrations. Our calibration pipeline is designed to accurately calibrate complex vision systems in the context of autonomous driving, where precise multi-sensor alignment is critical. We validate our approach through an extensive experimental evaluation on a custom built dataset, recorded with an advanced autonomous driving sensor setup, confirming the accuracy and robustness of our method. Further implementation details are available online at the supplementary material page.

1 Introduction

In autonomous driving, a diverse set of sensing modalities is required to gain a complete understanding of the environment. RGB cameras provide semantic information, LiDARs offer precise 3D geometric structure, and event cameras are particularly effective in challenging conditions such as high-speed motion, low light or high dynamic range scenes. While these sensors are complementary, their integration relies critically on accurate *extrinsic calibration*, *i.e.*, the estimation of the relative poses between each pair of sensors in the system. Among these sensor modalities, event cameras present the most significant calibration challenges due to their fundamentally different sensing mechanism and sparse, asynchronous output. While LiDAR-RGB calibration is well-established in the literature, robust event camera calibration, particularly in multi-sensor contexts, remains an open problem that our work specifically addresses.

Specifically, as illustrated in Figure 1 (b), extrinsic calibration defines the spatial transformation that aligns the coordinate systems of different sensors, enabling consistent sensors fusion and coherent scene reconstruction. The goal is to compute the rotation matrices R_{ij} and translation vectors $\mathbf{t}_{ij}$ that relate the coordinate systems of each sensors' pair (i, j). Note that, unless otherwise specified, the term calibration refers exclusively to the estimation of extrinsic parameters, thus the relative pose of the sensors, as we assume intrinsic parameters already estimated by classical approaches for each sensor.

We focus on the estimation of the relative pose of a suite of *heterogeneous sensors* including an event camera, LIDARs and RGB cameras. This is essential in multi-sensor frameworks such as those used in autonomous driving for object detection Sgaravatti et al. [2025], Bai et al. [2022], Yin et al. [2023], but also applies to other robotics and perception

*Corresponding author: andrea.bertogalli@mail.polimi.it

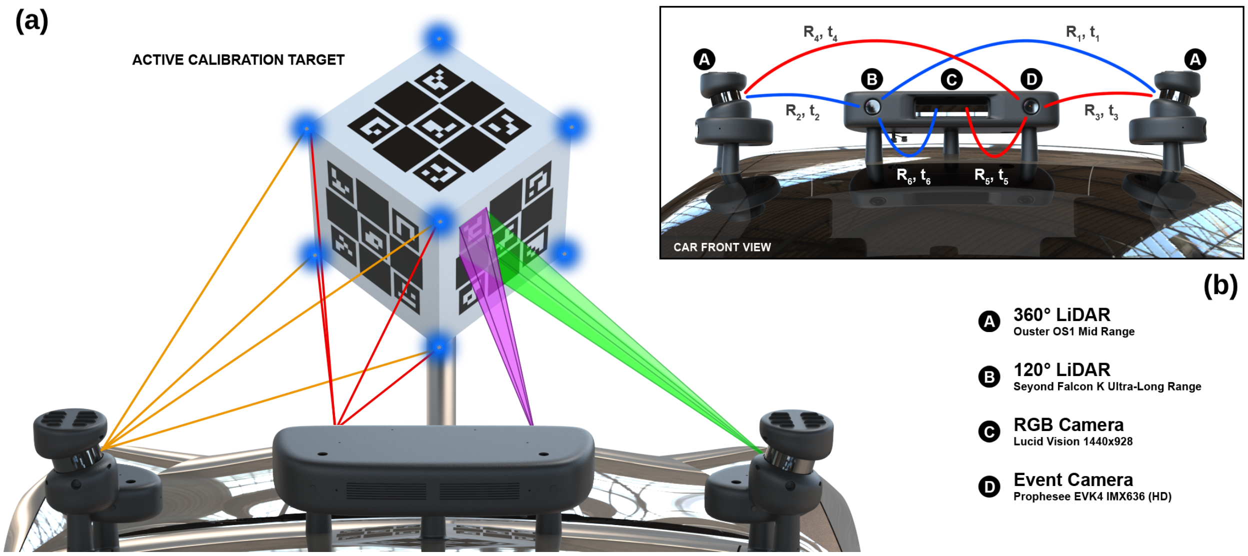


Figure 1: An overview of our calibration target. (a) Our target allows simultaneous features detection across all reference frames. Specifically, the target design addresses event camera detection challenges through frequency-coded LEDs placed at the corners of the cube. (b) The sensor suite along with the estimated relative poses (R_{ij}, t_{ij}) between each sensor pair (i, j). Note that not all RGB cameras are visualized for clarity.

tasks as odometry estimation Censi and Scaramuzza [2014], Zuo et al. [2019], drone tracking Magrini et al. [2024], obstacle detection Wu et al. [2023] and a broad range of additional use cases.

Unfortunately, this task is inherently challenging, as the diverse sensor modalities make it non-trivial to establish unambiguous and reliable correspondences, since there is no direct one-to-one mapping between features of each modality. In particular, establishing correspondences between the sparse events returned by an event camera and the dense LiDAR pointcloud remains a significant challenge. The calibration is further complicated by the high precision required to ensure an accurate spatial alignment of the sensors. These characteristics make traditional calibration techniques less effective. Moreover, classical calibration methods are not suited for event cameras, as they address completely different data. Most of the few methods that address the calibration of an event camera with a LiDAR rely on double-projection schemes Zhu et al. [2018], Gehrig et al. [2021]. These approaches, instead of calibrating the event camera directly with the LiDAR, follow a two-step approach: the event camera is calibrated with an auxiliary sensor (tipically an RGB camera), which is in turn aligned with the LiDAR. This approach suffers from reduced robustness and precision, as the final accuracy depends on the compounded error of multiple transformations, and often requires multiple calibration rigs. In contrast, our method overcomes these limitations by providing a *single*, *direct* estimate of all relative poses *simultaneously* during a single data acquisition.

To achieve this one-shot multi-modal calibration, we introduce a custom 3D calibration target comprising a three-faced ArUco cube with blinking LEDs on each corner as shown in Figure 1 (a). Our target is specifically designed to provide salient features across all three sensor modalities. In addition, our calibration scheme includes a novel feature detection approach for both LiDAR and events in order to provide an accurate and simultaneous feature matching between the two. Our method is well suited for autonomous driving scenarios where the intrinsic parameters of each sensor are typically constant over time and the calibration scene is static. Furthermore, temporal synchronization issues, that are common in other methods due to the different sampling rates of the involved sensors, are inherently avoided by design. As demonstrated in our experiments, our method enables precise and unified extrinsic calibration across multiple modalities, showing improved accuracy and robustness compared to state-of-the-art alternatives Song et al. [2018], Xing et al. [2023], Yan et al. [2023].

To our knowledge, no existing work has proposed a pipeline capable of directly calibrating a LiDAR with both event and RGB cameras simultaneously. Thus in the following section we focus on works that address the calibration of pairs of sensors.

2 Related works

Given the relatively recent adoption of event cameras, the body of literature on multi-sensor calibration involving them is still rather limited. In the following, we briefly review the main existing approaches that address calibrations between pairs of different sensor modalities, considering both target-based and targetless approaches and also automatic and manual approaches. In practice, target-based calibration methods are often preferred due to their higher accuracy, repeatability, and ease of use, especially in controlled environments, unlike target-free approaches, which typically rely on motion or scene assumptions that may not hold consistently.

LiDAR - event camera extrinsic calibration The literature on extrinsic calibration between an event camera and a LiDAR is scarce. Most datasets Gehrig et al. [2021], Zhu et al. [2018] use the double projection scheme to project points between the event camera's reference frame and the LiDAR's one, passing through the reference frame of an RGB camera. This approach is simpler than direct calibration but more imprecise, as it compounds the errors of two projections.

There are only few direct calibration methods between LiDARs and event cameras. In Ta *et al.* Ta et al. [2023] the authors exploit the fact that most recent event cameras can capture the reflection of the laser beams projected by a LiDAR sensor, and this is used to perform an automatic target-free calibration by minimizing an entropy loss. Jiao *et al.* Jiao et al. [2023] proposes a calibration based on deep-frame reconstruction with the well-known E2VID model Rebecq et al. [2019]. Another automatic target-free approach is presented by Xing *et al.* Xing et al. [2023] which exploits motion to generate events and calibrate the sensors. The only method directly comparable to ours is that of Song *et al.* Song et al. [2018], which introduces an automatic target-based event camera–LiDAR calibration using a custom calibration board with four illuminated circular holes. While these circular patterns can be captured by both the LiDAR and the event camera, a different target is required to calibrate the RGB camera with the LiDAR, and the resulting performance is inferior to that of our multi-modal approach.

LiDAR - RGB camera calibration Extrinsic calibration between LiDARs and RGB cameras is a well explored task Li et al. [2023], An et al. [2024], with existing approaches generally falling into four main categories. A first group includes *manual target-based methods*, where a user manually identifies correspondences on a known calibration target before performing the calibration. These methods commonly use checkerboard patterns Geiger et al. [2012], Zhang and Pless [2004], Zhou and Deng [2012], ArUco markers Dhall et al. [2017], Yoo et al. [2018], or other types of fiducial points Guindel et al. [2017], Velás et al. [2014], Hassanein et al. [2016], Pusztai and Hajder [2017].

A second group consists of *automatic target-based methods*, where no manual annotation is required, making them particularly well-suited for real-world applications in controlled environments. For example, Tóth et al. [2020], Beltrán et al. [2022], Yan et al. [2023], Grammatikopoulos et al. [2022], Zhou et al. [2018] propose fully automatic calibration pipelines based on designed targets. In particular, Yan *et al.* Yan et al. [2023] introduce a custom target composed of circular holes and a checkerboard, enabling reliable feature detection in both LiDAR and RGB modalities—a design also adopted by Song et al. [2018] for event-LiDAR calibration. Grammatikopoulos *et al.* Grammatikopoulos et al. [2022] propose a target which exploits two crossing retroreflective stripes and an AprilTag to calibrate a LiDAR and a camera, but it requires a dynamic scene and a consequent temporal calibration between the two sensors. Similarly Zhou *et al.* Zhou et al. [2018] employ an acquisition of a classic checkerboard, exploiting lines and planes correspondences to calibrate the sensors.

A third class involves *manual targetless methods*, where correspondences are manually selected directly between RGB images and LiDAR pointclouds Peršić et al. [2021]. While flexible, these approaches are typically labor-intensive and less repeatable. Lastly, *automatic targetless methods* aim to infer extrinsic parameters directly from environmental features such as planes or edges, relying on assumptions about the scene's structure or motion Li et al. [2017], Liu et al. [2018], Yuan et al. [2021]. Although promising for online or in-field calibration, these methods often fall short in terms of precision and robustness compared to target-based alternatives.

RGB camera - event camera extrinsic calibration The most common approach for calibrating an event camera with an RGB camera is to apply standard stereo calibration techniques. The is achieved by reconstructing a temporally synchronized frame-like representation of the event stream either with deep learning Rebecq et al. [2019], Muglikar et al. [2021] or with other techniques Hu et al. [2024], Dubeau et al. [2020], Creß et al. [2024]. This reconstruction provides two synchronized data streams, from which features are detected and matched to perform stereo calibration.

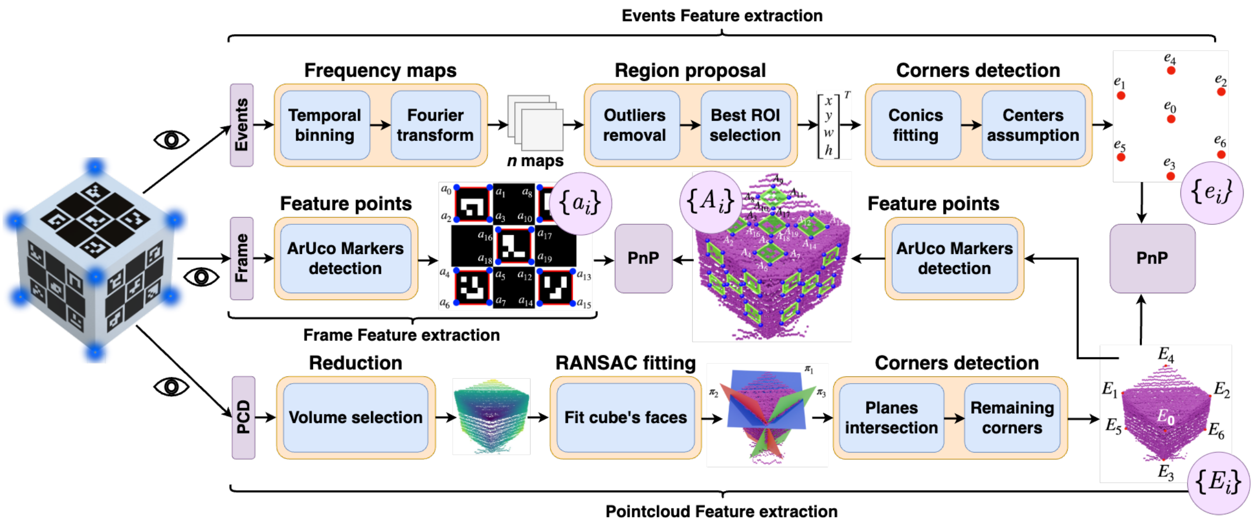


Figure 2: Our calibration pipeline consists in three feature detection branches on the same calibration target. Our novel event feature detection branch detects the 7 cube's corners e_i, the RGB feature detection detects the ArUcos' corners a_i and the pointcloud feature detection branch detects the 3D points corresponding to cube's corner E_i and ArUcos' corners A_i. Finally we apply the PnP algorithm on the correspondences $A_i \leftrightarrow a_i$ and $E_i \leftrightarrow e_i$.

3 Method

The core idea of our method is to leverage a custom-designed calibration target that interacts with all sensing modalities and enables the *simultaneous* detection of the same physical keypoints across the different reference frames involved. Our target consists in a cube with blinking LEDs at its corners, as illustrated in Figure 1 and described in Section 3.1. Our calibration framework is structured into three main keypoints extraction modules, one for each modality involved, as illustrated in Figure 2. The core innovation lies in detecting blinking LEDs placed on the cube's corners $\{E_i\}$ to extract keypoints $\{e_i\}$ from the event stream through frequency analysis (Section 3.2). The three orthogonal faces of the cube are exploited to extract the corresponding corners $\{E_i\}$ in the LiDAR pointcloud (Section 3.3), while ArUco markers $\{A_i\}$ provide anchor points $\{a_i\}$ for RGB cameras (Section 3.4). Crucially, the physical layout of the target ensures that the extracted features are not only detectable, but also uniquely matched across modalities. This allows us to compute correspondences $\{e_i \leftrightarrow E_i\}$ and $\{a_i \leftrightarrow A_i\}$ without ambiguity. Once these cross-sensor correspondences are established, we estimate the extrinsic transformations using a standard Perspective-n-Point (PnP) formulation (Section 3.5).

3.1 3D multi-modal target

Our custom target design is specifically optimized for event camera feature detection challenges. We adapt the blinking LED pattern to easily detect unique keypoints and establish reliable correspondences between sparse event data and dense LiDARs pointcloud.

We only consider the three adjacent faces of the cube facing the sensor suite, that are made by assembling together 3 PVC planes. On each face we place a different 3×3 ChArUco board to uniquely identify the markers' features $\{a_0, a_1, ..., a_{59}\}$ from the RGB camera, while the 3D geometry of the object allows to identify the corners $\{E_0, E_1, ..., E_6\}$ and Aruco's 3D positions $\{A_0, A_1, ..., A_{59}\}$ in the same reference system of the pointcloud. On the seven visible corners $\{E_i\}$ of the 3 faces we installed 7 ellipsoidal LED diodes. LEDs are controlled by an ESP32 microcontroller which makes them blink at a specific frequency (reported in Figure 3). Frequencies have been chosen in order to mitigate the risk of ambiguity in uniquely identifying the LED in the subsequent analysis. The blinking behavior is essential to trigger events while the blinking frequency is used to uniquely identify each LED. Each LED represents a feature in the event stream $\{e_0, e_1, ..., e_6\}$. Note that the measures of the cube are known, allowing a precise modeling of the target.

3.2 Identifying keypoints in the event stream

The procedure to identify the position of the keypoints in the event stream is composed by three main stages.

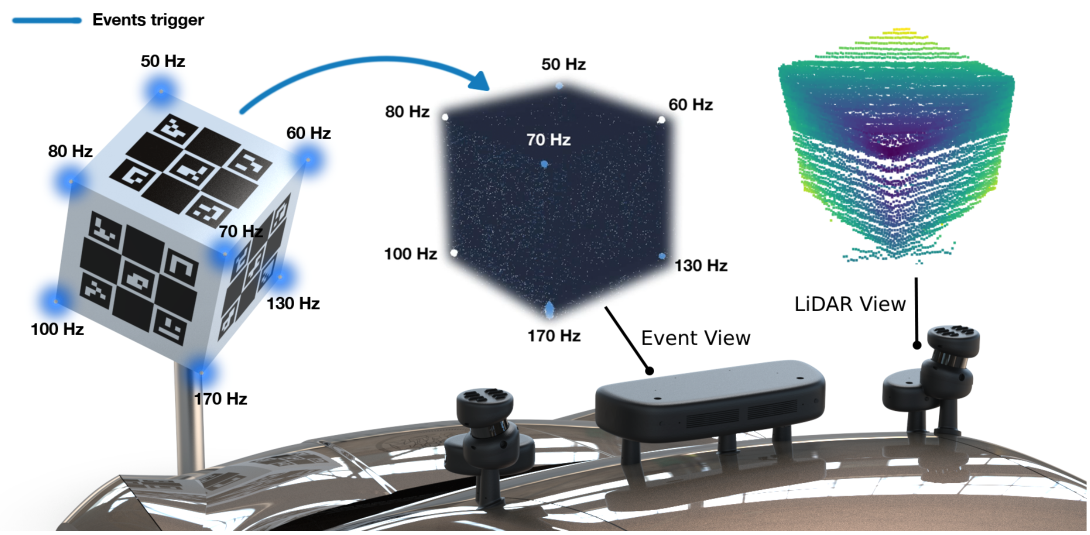


Figure 3: The operating principle of our calibration target. Each LED on the cube blinks at a specific frequency, resulting in a continuous event generation on the event camera yielding a square wave signal for each pixel. In the event view is shown a cumulated frame-like representation of the event stream over a *33.333 ms* time period. The LiDAR view emphasizes the cube inside the pointcloud, where darker colors represent points that are closer to the sensor.

i) Frequency analysis: Pixels of an event camera respond to changes in logarithmic brightness intensity $L = \log(I)$. An event is a tuple $\eta_k = (u_k, t_k, p_k)$, where $u_k = (x_k, y_k)^T$ is the pixel location, t_k the timestamp (typically with a temporal resolution of around 10 μs), and $p_k \in \{-1, +1\}$ the polarity. In particular, an event is triggered when the change in intensity L exceeds a threshold C since the last event at u_k, with polarity:

$$p_k = \begin{cases} +1, & L(u_k, t_k) - L(u_k, t_k - \Delta t) \geq C \\ -1, & L(u_k, t_k) - L(u_k, t_k - \Delta t) \leq -C, \end{cases} \tag{1}$$

where $\Delta t = t_k - t_{k-1}$ is the elapsed time between two consecutive events. Note that pixels with small intensity variation do not give rise to any event. When the LEDs on the target turn on and off repeatedly each pixel generates a stream of events that can be interpreted as a periodic square wave signal. Given an event stream we can compute the blinking frequency of each blinking LED by means of a *frequency map*, defined as a matrix $\mathbf{M} \in \mathbb{R}^{w \times h \times 1}$ having the same dimension $(w,\ h)$ of the event camera frame. A frequency map stores for each pixel its blinking frequency value in Hz as depicted in Figure 4 (a), where frequencies are color-coded. Computing frequencies from an event stream is an already explored task Zhao et al. [2024], Pfrommer [2022], Hoseini et al. [2017], Censi et al. [2013], in our implementation we adopt a method similar to Zhao et al. [2024]. The frequency detection is limited on the predefined range of frequencies (10-200hz), acting like a band-pass filter to filter out environmental noise.

In order to construct a frequency map $\mathbf{M}$ we compute m temporal bins of length Δt:

$$B_j = [t_0 + (j-1)\Delta t, t_0 + j\Delta t) \quad \text{for} \quad j = 1, 2, \ldots, m. \tag{2}$$

For each bin, we derive a value of the signal by considering the polarity of the events in the bin. Once we have reconstructed the square wave for each pixel, we apply the Fast Fourier Transform to obtain the frequency map and consequently the blinking frequency of the LEDs.

ii) Best frequency map estimation: During the acquisition, it is very likely that interferences arise in the event stream and spoil the resulting frequency map. Thus, rather than considering a single frequency map for the entire stream, we compute n frequency maps $\mathbf{M}_i$, each one for a portion of the stream. This allows us to strengthen the frequency estimation process. On each $\mathbf{M}_i$ we compute the smallest bounding box R_i that encloses the non-null frequency region, depicted in white in Figure 4 (a). Then we discard all the bounding boxes that are deemed outliers according to the 3-σ rule. Eventually, we compute a reference bounding box R having as vertices the means of the corresponding vertices of all the remaining R_i. We then select the frequency map $\mathbf{M}_{\bar{i}}$ whose bounding box $R_{\bar{i}}$ minimizes the Manhattan distance (L_1-norm) with R:

$$\bar{i} = \arg\min_j \|R_j - R\|_1. \tag{3}$$

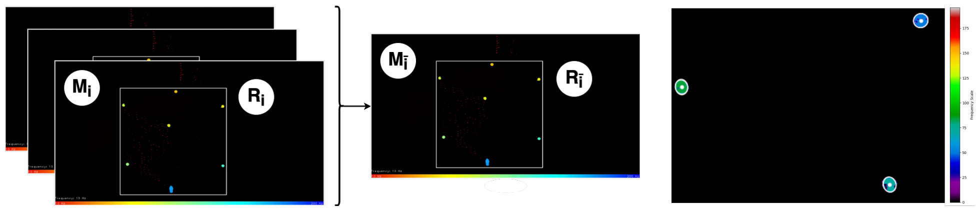


Figure 4: On the left: n frequency maps $\mathbf{M}_i$ with their bounding boxes R_i are analyzed to yield the best frequency map $\mathbf{M}_{\bar{i}}$ along with the bounding box $R_{\bar{i}}$. On the right, the figure shows some of the detected points (the centers of the fitted ellipses) e_i, which are identified using the associated frequency. Note that only three ellipses are shown for simplicity.

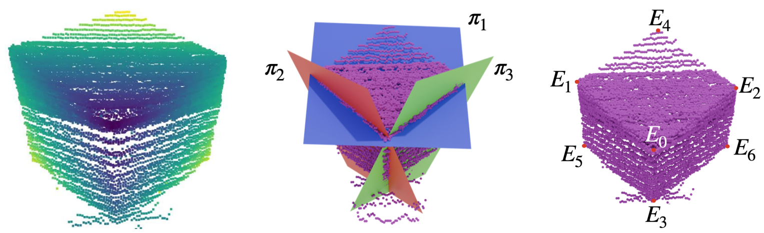


Figure 5: The output of the three steps to detect the features (the seven cube's corner E_i) from the LiDAR pointcloud. (a) the reduced pointcloud we are considering. (b) the fitted planes which represent the cube's faces. (c) the seven detected points E_i, which are known exploiting the geometry of the cube.

iii) Ellipses fitting and location estimation: As last step, inside the selected bounding box $R_{\bar{i}}$ we fit in $\mathbf{M}_{\bar{i}}$ seven ellipses, one for each LED, using a constrained least squares Fitzgibbon et al. [1999] on the edges Canny [1986] detected in the frequency map. At this point, we assume that each ellipse's center $\{e_0, e_1, ..., e_6\}$ corresponds to a cube's corner as illustrated in Figure 4 (b).

3.3 Identifying 3D corners on the pointcloud

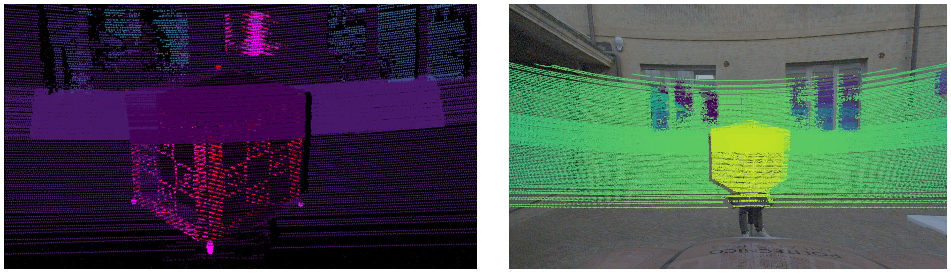
To identify the 3D positions $\{E_i\}$ of the cube's corners we proceed as follows. First, starting from a cropped version of the LiDAR pointcloud (Figure 5 (a)), we fit 3 planes π_1, π_2, π_3 using Sequential Ransac (Figure 5 (b)). Then E_0 is computed as the intersection of all the three planes π_1, π_2, and π_3. Since the dimensions of the cube are known by design, the remaining six points $\{E_1, E_2, ..., E_6\}$ are derived through standard geometric computations, as illustrated in Figure 5 (c).

3.4 ArUco markers detection in RGB and pointcloud

Starting from the cube's corners $\{E_i\}$ identified in the pointcloud, the ArUco markers $\{A_i\}$ are identified exploiting the known measures of each cube's face. In this way we can identify 4 points (the corners) for each ArUco marker in the pointcloud, for a total of 60 points $\{A_0, A_1, ..., A_{59}\}$. The corresponding markers $\{a_0, a_1, ..., a_{59}\}$ are retrieved from the RGB camera frames using the marker detection procedure from Garrido-Juardo *et al.* Garrido-Jurado et al. [2014].

3.5 Pose estimation

For both calibrations, the one between LiDAR and event camera and the one between LiDAR and RGB camera, the calibration procedure is made by employing the well-known Perspective-n-points algorithm Zheng et al. [2013] applied on the correspondences found. In particular, for the calibration between the LiDAR and the event camera we use the 7 cube's corners identified in both the event camera frame and in the pointcloud, building a set of correspondences $\{E_i \leftrightarrow e_i\}$ with $i = 0, \ldots, 6$. For the calibration between the LiDAR and the RGB camera we use the corners of the ArUco markers identified in both the camera image plane and in the pointcloud, building the set of correspondences $\{A_i \leftrightarrow a_i\}$ with $i = 0, \ldots, 59$.



(a) (b)

Figure 6: Qualitative analysis results. The alignment between the pointcloud and the event stream in (a) and the alignement between the pointcloud and the RGB image (b) are satisfactory. From (a) can also be noticed that the cube's faces are illuminated by the LiDAR and the reflection triggers events.

4 Experimental validation

All the experiments are conducted on a custom recorded dataset composed of 17 sequences featuring our calibration target in different light conditions (morning, noon and evening) and positions. The dataset has been recorded on a Maserati GranCabrio Folgore equipped with three LiDARs, three RGB cameras and one event camera, all directly calibrated with our method (Figure 1).

Figure of merits. We quantitatively evaluate calibration performances in terms of *average reprojection error* $E_{\text{mean}} = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2}$, where the reprojection error measures the image distance in pixels between a projected point (x,y) and a measured one $(\hat{x},\hat{y})$. We consider the average error between detected points in the target reference frame j and the projection of the same points projected from the source reference frame i to the target j by means of the projection matrix $P_{i,j} = [R_{i,j}|t_{i,j}]$. The errors are also averaged by calibration type, yielding one value for the event camera–LiDAR calibration and another for the RGB–LiDAR calibration. This is necessary since multiple LiDARs are calibrated with the event camera and multiple LiDARs are calibrated with different RGB cameras. To qualitatively evaluate the calibrations involving LiDARs, the pointcloud is projected onto the image plane (Figure 6) and the alignment between the two is evaluated.

Alternative methods. In terms of performances, we can compare our method against those that use the reprojection error as performance measure. In fact, since we used a custom calibration target, we can not evaluate our method on standard datasets provided with ground truth. As regards the calibration between the event camera and the LiDARs we consider Song *et al.* Song et al. [2018] that uses a custom calibration target and Xing *et al.* Xing et al. [2023] that instead relies on a targetless calibration approach. As regards the calibration between RGB cameras and LiDAR, we consider Yan *et al.* Yan et al. [2023], Grammatikopoulos *et al.* Grammatikopoulos et al. [2022] and Zhou *et al.* Zhou et al. [2018] which all exploit different calibration targets.

Results. Results are collated in Tab 1. For what concerns the more challenging calibration between the event camera and the LiDAR, the benefits of our multi-modal calibration target are more evident, as our performances are considerably better than the ones obtained by similar methods. As regards the calibration between the RGB camera and the LiDAR the results are comparable, the difference with the best method is only 0.03 pixels, which is indeed a very small deviation. The advantage of our approach compared to the others is that we can calibrate the LiDAR with both the event and the RGB camera with a single calibration target and with a unique pipeline, in contrast to previously published methods which take into account only one type of calibration.

The impact of a geometric correction. As explained in Section 3.2 we assume that the centers $\{e_i\}$ of the ellipses coincide with the projections of the corners of the cube $\{E_i\}$. A more sophisticated approximation is to treat each LED diode as a 3D ellipsoid, and to consider its projection as an ellipse in the image plane. Given at least two ellipses-ellipsoids correspondences (we have seven) we can estimate the event camera pose by exploiting the algorithm proposed by Gaudilliere *et al.* Gaudillière et al. [2020] that leverages the geometric properties of quadric projection to accurately localize a perspective camera. The performances of this refinement applied to a synthetic scene (where the quadric definition of each ellipsoid is correct) are good (0.3 of average Jaccard distance between detected ellipses and projected ellipsoids) but even if this method is theoretically correct, the accuracy drops in real world scenes where we have to estimate the ellipsoidal model of a LED, therefore we decided to simply approximate the projections of the corners with the ellipses' centers.

Method	Event camera - LiDAR	RGB Camera - LiDAR
Song *et al.* Song et al. [2018]	4.691	-
Xing *et al.* Xing et al. [2023]	3.161	-
Yan *et al.* Yan et al. [2023]	-	1.104
Grammatikopoulos *et al.* Grammatikopoulos et al. [2022]	-	11.4826 *
Zhou *et al.* Zhou et al. [2018]	-	21.69332 *
Our	**2.732**	**1.134**

Table 1: The quantitative results (in pixels) compared with other similar works on the same metric. Our method is the only one that takes into account both the calibrations directly. An asterisk * indicates values computed from the vertical reprojection error and the horizontal reprojection error reported by the authors.

5 Practical implementations

The work proposed has been implemented to calibrate a real autonomous driving framework of a fully autonomous Maserati GrancCabrio Folgore. Further details are available at https://andberto.github.io/One-target-to-align-them-all/. The sensors suite is adapted from Sgaravatti et al. [2024], Pieroni et al. [2024] with the addition of the event cameras. As introduced in Figure 1 the sensors employed are:

- **A** Two Ouster LiDARs 360° Mid Range, which capture a dense 3D representation of the surroundings (128 scan-lines) and a distance range up to 200m.
- **B** One Seyond Falcon K Ultra-Long Range which, with a 120° horizontal FoV, samples distances up to 500m on 152 scan-lines.
- **C** Up to three Lucid Vision IP68 RGB Cameras.
- **D** One EVK4 Prophesee provided with the IMX636 (HD) sensor.

The data is exchanged with the elaboration system through the well-known ROS2 Macenski *et al.* Macenski et al. [2022].

6 Conclusions

In this work, we successfully addressed the challenge of event camera calibration in multi-sensor autonomous driving systems. Our novel calibration target and feature extraction pipeline specifically tackle the difficulties inherent in event-based sensing, while simultaneously enabling standard RGB and LiDAR calibration within a unified framework.

Our approach lends itself well to extensions and adaptations, in fact the calibration pipeline could be extended to also include a stereo calibration between the two cameras following an approach similar to the one presented by Garcia *et al.* Garcia-D'Urso et al. [2025]. The calibration marker could be extended to include more feature points by adding more LEDs in order to increase the accuracy of the PnP algorithm. In general, our approach lays the foundation for a wide range of applications that require a robust, accurate, and flexible extrinsic calibration as a starting point for complex sensor systems, as for example multi-modal autonomous driving object detection frameworks.

Acknowledgments

The authors gratefully acknowledge Riccardo Pieroni, Andrea Corno, and Matteo Savaresi, along with the AIDA team, for their invaluable support and for generously providing access to their autonomous driving suite, vehicle, and calibration framework.

References

Carlo Sgaravatti, Roberto Basla, Riccardo Pieroni, Matteo Corno, Sergio M Savaresi, Luca Magri, and Giacomo Boracchi. A multimodal hybrid late-cascade fusion network for enhanced 3d object detection. *arXiv preprint arXiv:2504.18419*, 2025.

Xuyang Bai, Zeyu Hu, Xinge Zhu, Qingqiu Huang, Yilun Chen, Hongbo Fu, and Chiew-Lan Tai. Transfusion: Robust lidar-camera fusion for 3d object detection with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1090–1099, 2022.

Zixuan Yin, Han Sun, Ningzhong Liu, Huiyu Zhou, and Jiaquan Shen. Fgfusion: Fine-grained lidar-camera fusion for 3d object detection. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pages 505–517. Springer, 2023.

Andrea Censi and Davide Scaramuzza. Low-latency event-based visual odometry. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 703–710. IEEE, 2014.

Xingxing Zuo, Patrick Geneva, Woosik Lee, Yong Liu, and Guoquan Huang. Lic-fusion: Lidar-inertial-camera odometry. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5848–5854. IEEE, 2019.

Gabriele Magrini, Federico Becattini, Pietro Pala, Alberto Del Bimbo, and Antonio Porta. Neuromorphic drone detection: an event-rgb multimodal approach. *arXiv preprint arXiv:2409.16099*, 2024.

Yue Wu, Jingao Xu, Danyang Li, Yadong Xie, Hao Cao, Fan Li, and Zheng Yang. Flytracker: Motion tracking and obstacle detection for drones using event cameras. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*, pages 1–10. IEEE, 2023.

Alex Zihao Zhu, Dinesh Thakur, Tolga Özaslan, Bernd Pfrommer, Vijay Kumar, and Kostas Daniilidis. The multivehicle stereo event camera dataset: An event camera dataset for 3d perception. *IEEE Robotics and Automation Letters*, 3(3): 2032–2039, 2018.

Mathias Gehrig, Willem Aarents, Daniel Gehrig, and Davide Scaramuzza. Dsec: A stereo event camera dataset for driving scenarios. *IEEE Robotics and Automation Letters*, 6(3):4947–4954, 2021.

Rihui Song, Zhihua Jiang, Yanghao Li, Yunxiao Shan, and Kai Huang. Calibration of event-based camera and 3d lidar. In *2018 WRC Symposium on Advanced Robotics and Automation (WRC SARA)*, pages 289–295. IEEE, 2018.

Wanli Xing, Shijie Lin, Lei Yang, and Jia Pan. Target-free extrinsic calibration of event-lidar dyad using edge correspondences. *IEEE Robotics and Automation Letters*, 8(7):4020–4027, 2023.

Guohang Yan, Feiyu He, Chunlei Shi, Pengjin Wei, Xinyu Cai, and Yikang Li. Joint camera intrinsic and lidar-camera extrinsic calibration. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11446–11452. IEEE, 2023.

Kevin Ta, David Bruggemann, Tim Brödermann, Christos Sakaridis, and Luc Van Gool. L2e: Lasers to events for 6-dof extrinsic calibration of lidars and event cameras. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11425–11431. IEEE, 2023.

Jianhao Jiao, Feiyi Chen, Hexiang Wei, Jin Wu, and Ming Liu. Lce-calib: automatic lidar-frame/event camera extrinsic calibration with a globally optimal solution. *IEEE/ASME Transactions on Mechatronics*, 28(5):2988–2999, 2023.

Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE transactions on pattern analysis and machine intelligence*, 43(6):1964–1980, 2019.

Xingchen Li, Yuxuan Xiao, Beibei Wang, Haojie Ren, Yanyong Zhang, and Jianmin Ji. Automatic targetless lidar–camera calibration: a survey. *Artificial Intelligence Review*, 56(9):9949–9987, 2023.

Pei An, Junfeng Ding, Siwen Quan, Jiaqi Yang, You Yang, Qiong Liu, and Jie Ma. Survey of extrinsic calibration on lidar-camera system for intelligent vehicle: Challenges, approaches, and trends. *IEEE Transactions on Intelligent Transportation Systems*, 25(11):15342–15366, 2024. doi:10.1109/TITS.2024.3419758.

Andreas Geiger, Frank Moosmann, Ömer Car, and Bernhard Schuster. Automatic camera and range sensor calibration using a single shot. In *2012 IEEE international conference on robotics and automation*, pages 3936–3943. IEEE, 2012.

Qilong Zhang and Robert Pless. Extrinsic calibration of a camera and laser range finder (improves camera calibration). In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 3, pages 2301–2306. IEEE, 2004.

Lipu Zhou and Zhidong Deng. A new algorithm for computing the projection matrix between a lidar and a camera based on line correspondences. In *2012 IV International Congress on Ultra Modern Telecommunications and Control Systems*, pages 436–441. IEEE, 2012.

Ankit Dhall, Kunal Chelani, Vishnu Radhakrishnan, and K Madhava Krishna. Lidar-camera calibration using 3d-3d point correspondences. *arXiv preprint arXiv:1705.09785*, 2017.

Joong-Sun Yoo, Do-Hyeong Kim, and Gon-Woo Kim. Improved lidar-camera calibration using marker detection based on 3d plane extraction. *Journal of Electrical Engineering and Technology*, 13(6):2530–2544, 2018.

Carlos Guindel, Jorge Beltrán, David Martín, and Fernando García. Automatic extrinsic calibration for lidar-stereo vehicle sensor setups. In *2017 IEEE 20th international conference on intelligent transportation systems (ITSC)*, pages 1–6. IEEE, 2017.

Martin Velás, Michal Španěl, Zdeněk Materna, and Adam Herout. Calibration of rgb camera with velodyne lidar. In *WSCG 2014 Conference on Computer Graphics, Visualization and Computer Vision*, Plzeň, Czech Republic, 2014. Václav Skala - UNION Agency.

M Hassanein, A Moussa, and N El-Sheimy. A new automatic system calibration of multi-cameras and lidar sensors. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41:589–594, 2016.

Zoltan Pusztai and Levente Hajder. Accurate calibration of lidar-camera systems using ordinary boxes. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 394–402, 2017.

Tekla Tóth, Zoltán Pusztai, and Levente Hajder. Automatic lidar-camera calibration of extrinsic parameters using a spherical target. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8580–8586. IEEE, 2020.

Jorge Beltrán, Carlos Guindel, Arturo De La Escalera, and Fernando García. Automatic extrinsic calibration method for lidar and camera sensor setups. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):17677–17689, 2022.

Lazaros Grammatikopoulos, Anastasios Papanagnou, Antonios Venianakis, Ilias Kalisperakis, and Christos Stentoumis. An effective camera-to-lidar spatiotemporal calibration based on a simple calibration target. *Sensors*, 22(15):5576, 2022.

Lipu Zhou, Zimo Li, and Michael Kaess. Automatic extrinsic calibration of a camera and a 3d lidar using line and plane correspondences. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5562–5569. IEEE, 2018.

Juraj Peršić, Luka Petrović, Ivan Marković, and Ivan Petrović. Online multi-sensor calibration based on moving object tracking. *Advanced Robotics*, 35(3-4):130–140, 2021.

Tao Li, Jianwu Fang, Yang Zhong, Di Wang, and Jianru Xue. Online high-accurate calibration of rgb+ 3d-lidar for autonomous driving. In *Image and Graphics: 9th International Conference, ICIG 2017, Shanghai, China, September 13-15, 2017, Revised Selected Papers, Part III 9*, pages 254–263. Springer, 2017.

Hao Liu, Yingjian Liu, Xiaoyan Gu, Yingying Wu, Fangchao Qu, and Lei Huang. A deep-learning based multi-modality sensor calibration method for usv. In *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, pages 1–5. IEEE, 2018.

Chongjian Yuan, Xiyuan Liu, Xiaoping Hong, and Fu Zhang. Pixel-level extrinsic self calibration of high resolution lidar and camera in targetless environments. *IEEE Robotics and Automation Letters*, 6(4):7517–7524, 2021.

Manasi Muglikar, Mathias Gehrig, Daniel Gehrig, and Davide Scaramuzza. How to calibrate your event camera. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1403–1409, 2021.

Rui Hu, Jürgen Kogler, Margrit Gelautz, Min Lin, and Yuanqing Xia. A dynamic calibration framework for the event-frame stereo camera system. *IEEE Robotics and Automation Letters*, 2024.

Etienne Dubeau, Mathieu Garon, Benoit Debaque, Raoul de Charette, and Jean-François Lalonde. Rgb-de: Event camera calibration for fast 6-dof object tracking. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 127–135. IEEE, 2020.

Christian Creß, Walter Zimmer, Nils Purschke, Bach Ngoc Doan, Sven Kirchner, Venkatnarayanan Lakshminarasimhan, Leah Strand, and Alois C Knoll. Tumtraf event: Calibration and fusion resulting in a dataset for roadside event-based and rgb cameras. *IEEE Transactions on Intelligent Vehicles*, 2024.

M. Zhao, X. Shen, L. Su, and Z. Dong. Power equipment vibration visualization using intelligent sensing method based on event-sensing principle. *Global Energy Interconnection*, 7(2):228–240, 2024.

Bernd Pfrommer. Frequency cam: Imaging periodic signals in real-time. *arXiv preprint arXiv:2211.00198*, 2022.

Sahar Hoseini, Garrick Orchard, Amirreza Yousefzadeh, Balakrishna Deverakonda, Teresa Serrano-Gotarredona, and Bernabé Linares-Barranco. Passive localization and detection of quadcopter uavs by using dynamic vision sensor. In *2017 5th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, pages 81–85. IEEE, 2017.

Andrea Censi, Jonas Strubel, Christian Brandli, Tobi Delbruck, and Davide Scaramuzza. Low-latency localization by active led markers tracking using a dynamic vision sensor. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 891–898. IEEE, 2013.

Andrew Fitzgibbon, Maurizio Pilu, and Robert B Fisher. Direct least square fitting of ellipses. *IEEE Transactions on pattern analysis and machine intelligence*, 21(5):476–480, 1999.

John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986. doi:10.1109/TPAMI.1986.4767851.

Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6): 2280–2292, 2014.

Yinqiang Zheng, Yubin Kuang, Shigeki Sugimoto, Kalle Astrom, and Masatoshi Okutomi. Revisiting the pnp problem: A fast, general and optimal solution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2344–2351, 2013.

Vincent Gaudillière, Gilles Simon, and Marie-Odile Berger. Perspective-2-ellipsoid: Bridging the gap between object detections and 6-dof camera pose. *IEEE Robotics and Automation Letters*, 5(4):5189–5196, 2020.

Carlo Sgaravatti, Roberto Basla, Riccardo Pieroni, Matteo Corno, Sergio M Savaresi, Luca Magri, and Giacomo Boracchi. A multimodal hybrid late-cascade fusion network for enhanced 3d object detection. In *European Conference on Computer Vision*, pages 339–356. Springer, 2024.

Riccardo Pieroni, Simone Specchia, Matteo Corno, and Sergio Matteo Savaresi. Multi-object tracking with camera-lidar fusion for autonomous driving. In *2024 European Control Conference (ECC)*, pages 2774–2779. IEEE, 2024.

Steven Macenski, Tully Foote, Brian Gerkey, Chris Lalancette, and William Woodall. Robot operating system 2: Design, architecture, and uses in the wild. *Science robotics*, 7(66):eabm6074, 2022.

Nahuel Garcia-D'Urso, Bernabe Sanchez-Sos, Jorge Azorin-Lopez, Andres Fuster-Guillo, Antonio Macia-Lillo, and Higinio Mora-Mora. Marker-based extrinsic calibration method for accurate multi-camera 3d reconstruction. *arXiv preprint arXiv:2505.02539*, 2025.