MULi-Ev: Maintaining Unperturbed LiDAR-Event Calibration

Mathieu Cocheteux, Julien Moreau, Franck Davoine

MULi-Ev: Maintaining Unperturbed LiDAR-Event Calibration

Mathieu Cocheteux[1], Julien Moreau[1], Franck Davoine[2]
[1]Université de technologie de Compiègne, CNRS, Heudiasyc, France
[2]CNRS, INSA Lyon, UCBL, LIRIS, UMR5205, France
{mathieu.cocheteux, julien.moreau}@hds.utc.fr, franck.davoine@cnrs.fr

Abstract

Despite the increasing interest in enhancing perception systems for autonomous vehicles, the online calibration between event cameras and LiDAR—two sensors pivotal in capturing comprehensive environmental information—remains unexplored. We introduce MULi-Ev, the first online, deep learning-based framework tailored for the extrinsic calibration of event cameras with LiDAR. This advancement is instrumental for the seamless integration of LiDAR and event cameras, enabling dynamic, real-time calibration adjustments that are essential for maintaining optimal sensor alignment amidst varying operational conditions. Rigorously evaluated against the real-world scenarios presented in the DSEC dataset, MULi-Ev not only achieves substantial improvements in calibration accuracy but also sets a new standard for integrating LiDAR with event cameras in mobile platforms. Our findings reveal the potential of MULi-Ev to bolster the safety, reliability, and overall performance of perception systems in autonomous driving, marking a significant step forward in their real-world deployment and effectiveness.

1. Introduction

Autonomous driving technologies are on the brink of revolutionizing transportation, announcing a new era of enhanced safety, efficiency, and accessibility. At the heart of this transformation is the development of advanced perception systems that accurately interpret and navigate the complexities of the real world, such as the sharing of the road with other transport modalities (*e.g.* bikes, pedestrians, buses, etc.). A critical element in crafting such systems is sensor calibration. In this work we focus on extrinsic calibration between LiDAR and event cameras, a subject that still remains too little explored today.

Event cameras, which capture dynamic scenes with high temporal resolution and excel in various lighting conditions, can significantly reduce or help leverage motion blur [4, 10]. On the other hand, LiDAR sensors offer detailed depth information vital for precise object detection and environmental mapping. The integration of these complementary technologies promises to substantially elevate vehicle perception capabilities. However, no method has yet been proposed to provide accurate, real-time calibration between these sensors.

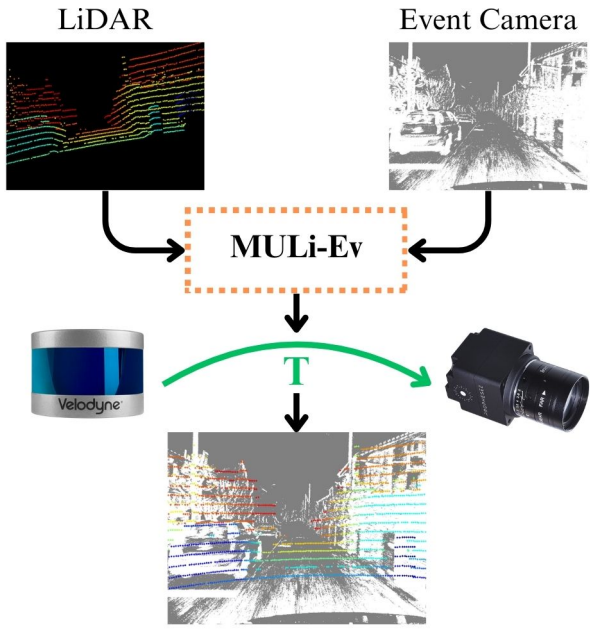


Figure 1. Overview of the MULi-Ev calibration workflow. This process integrates LiDAR point clouds and event camera data into the MULi-Ev network to compute accurate extrinsic calibration parameters (the rigid transformation in $SO(3)$ between the two sensors' reference frames, here represented by T). These parameters enable real-time, precise sensor alignment, facilitating enhanced perception for autonomous vehicles in dynamic scenarios.

Traditional calibration methods [8, 15, 16, 18] perform well under controlled conditions but are unusable in the dynamic, real-world environments autonomous vehicles encounter. These methods often necessitate cumbersome manual adjustments or specific calibration targets, unsuitable for the on-the-fly recalibration needs of operational ve-

hicles. Furthermore, the sparse and asynchronous nature of event camera data introduces additional challenges for the calibration process.

To address these challenges, we propose a novel deep-learning framework trained specifically for the online calibration of event cameras and LiDAR sensors (Figure 1). This approach not only simplifies the calibration process but also allows onboard online calibration on the vehicle, ensuring consistent sensor alignment. By enabling the joint use of these sensors, our method helps leveraging the complementary strengths of event cameras and LiDAR in other tasks, significantly enhancing the vehicle's perception system, enabling more accurate object detection and scene interpretation across a diverse range of driving scenarios.

Our contributions include:

1. The introduction of a deep-learning framework for online calibration between event cameras and LiDAR, enabling real-time, accurate sensor alignment—a first for this sensor combination.
2. The validation of our method against the DSEC dataset, showing marked improvements in calibration precision compared to existing methods.
3. The capability for on-the-fly recalibration introduced by our framework directly addresses the challenge of maintaining sensor alignment in dynamic scenarios, a crucial step toward enhancing the robustness of autonomous driving systems in real-world conditions.

The sections that follow will explore related works to contextualize our contributions within the broader research landscape, describe our methodology in detail, present an exhaustive evaluation of our framework against existing state-of-the-art methods, and conclude with a discussion on the broader implications of our findings and potential avenues for future research.

2. Related Works

2.1. Event Camera and LiDAR Calibration

The calibration of extrinsic parameters between event cameras and LiDAR is a necessity to leverage their combined capabilities for enhanced perception in autonomous systems. Unlike traditional cameras, event cameras capture pixel-level changes in light intensity asynchronously, presenting unique challenges for calibration with LiDAR, which provides sparse spatial depth information. A few offline calibration methods [8, 15, 16, 18] have been proposed.

Song *et al*. [15] made an early contribution with a 3D marker designed for this purpose. Although pioneering, their method necessitates specific, often impractical setup conditions. To address these limitations, Xing *et al*. [18] proposed a target-free calibration approach, utilizing natural edge correspondences in the data from both sensors. This innovative method simplifies the calibration process, but is still performed offline. Jiao *et al*. [8] introduced LCE-Calib, an automatic method that streamlines the calibration process, enhancing robustness and adaptability across various conditions. Building on these advancements, Ta *et al*. [16] introduced L2E, a novel automatic pipeline for direct and temporally-decoupled 6-DoF calibration between event cameras and LiDARs, which better leverages the specificities of event data to improve results.

This progression of techniques underscores a shift towards methods that are not only more versatile but also suited for real-world deployment. However, no method has been proposed until now for the online calibration of this sensor combination.

2.2. Deep Learning in Extrinsic Calibration

While deep learning has revolutionized many aspects of autonomous driving technology, its application to extrinsic calibration between event cameras and LiDAR remains unexplored. Our work introduces the first deep learning-based method for this specific task. However, the groundwork laid by methodologies for RGB cameras and LiDAR calibration [2, 3, 7, 9, 11, 14, 17] provides a valuable reference point. For instance, RegNet [14] by Schneider *et al*. leverages convolutional neural networks (CNNs) for sensor registration, predicting the 6-DOF parameters between RGB cameras and LiDAR without manual intervention, marking an early milestone in learning-based calibration. Following this, CalibNet [7] by Iyer *et al*. further refines the approach with a geometrically supervised network, enhancing the automation and accuracy of the calibration process. LCCNet [11], introduced by Lv *et al*., represents a significant advancement by utilizing a cost volume network to articulate the correlation between RGB images and depth images derived from LiDAR data, achieving substantial improvements in calibration precision. These methods underscore the potential of integrating deep learning into the calibration workflow, offering insights into feature correlation and end-to-end model training that are instrumental for our approach.

The existing body of work on RGB and LiDAR calibration delineates a path towards automated, real-time calibration solutions. By adapting and extending these methodologies, our research pioneers the application of deep learning for calibrating event cameras with LiDAR, aiming to harness the unique advantages of event cameras for enhanced autonomous vehicle perception and navigation.

3. Methodology

Our methodology introduces a deep-learning framework designed for the online calibration of event cameras and LiDAR sensors, aimed at autonomous driving applications. This section describes the overall architecture of our model,

Event Representation	Dimensions	Polarity	Temporality
Event Frame	$H \times W$	✗	✗
Voxel Grid	$B \times H \times W$	✗	✓
Time Surface	$H \times W$	(✓)	(✓)

Table 1. Comparison of some event representations considered for our method, and their properties. H and W represent respectively height and width.

the representation of event data, the calibration process, and details of our training procedure.

3.1. Architecture

Our calibration framework integrates event camera and LiDAR data through a unified deep learning architecture, similarly to UniCal [2], as illustrated in Figure 2. Leveraging a single MobileViTv2 [12] backbone for feature extraction and a custom-designed regression head, the framework achieves precise calibration parameter estimation.

Feature Extraction Backbone: Central to our approach is the MobileViTv2 [12] backbone, chosen for its fast inference speed and its ability to efficiently process multi-modal data. This facilitates handling event and LiDAR pseudo-images within a single backbone. By feeding both modalities into separate input channels, our model concurrently processes event camera and LiDAR data, learning intricate correlations between these two modalities. This unified processing not only streamlines the architecture but also bolsters the model's feature extraction capabilities, crucial for accurate extrinsic calibration.

Custom Regression Head: Focused on extrinsic calibration parameters, the regression head begins with a common layer that identifies features applicable to both translation and rotation, benefiting from shared data characteristics. Subsequently, the architecture divides into translation and rotation pathways, each comprising two layers designed specifically for their respective parameter sets. This specialization accounts for the unique aspects of translation (x, y, z) and rotation (roll, pitch, yaw) parameters, such as scale and unit differences, thereby enhancing the model's calibration precision.

3.2. Event Representation

In developing our calibration framework, a critical consideration was the optimal representation of event data captured by event cameras. Event cameras generate data in a fundamentally different manner from traditional cameras, recording changes in intensity for each pixel asynchronously. The data is structured as a flow of events, necessitating a thoughtful binning approach to transform it into a new representation for effective processing and integration with LiDAR data.

Our investigation encompassed various formats for representing event data, including:

- The event frame [13] representation, which accumulates events into a 2D image, where the intensity of a pixel corresponds to the number of events that occurred at that location within the specified accumulation time.
- The voxel grid [19] representation, which extends this concept into three dimensions, adding a temporal depth to the accumulation.
- The time surface [1] representation, which encodes the most recent timestamp of an event at each pixel, capturing the temporal dynamics more explicitly.

Each binning strategy offers distinct advantages in terms of capturing the spatial and temporal dynamics of the scene, which are recapitulated in Table 1. However, our primary objective was to identify a representation that not only simplifies the calibration process but also enhances performance by preserving essential geometric information such as edges, without unnecessarily complicating the model with temporal details that are less critical for our specific calibration task.

Ultimately, we found that event frame representation was the most effective approach. This decision was driven by several key factors:

- **Simplicity**: The event frame representation aligns closely with conventional data types used in deep learning, allowing for a more straightforward integration into our calibration framework.
- **Performance**: Through empirical testing (detailed in Section 4.5), we observed that the event frame provided superior performance in terms of calibration accuracy. This improvement is attributed to the format's effectiveness in preserving the geometric integrity of the scene.

In summary, the event frame representation emerged as the superior choice for our online calibration method, balancing simplicity, performance, and geometric fidelity. This finding underscores the importance of matching the data representation format with the specific requirements of the task, especially in the context of sensor fusion and calibration.

3.3. Training Procedure

3.3.1 Model Training

We introduce artificial decalibrations into the dataset, akin to the strategy employed by RegNet [14]. This involves systematically applying random offsets to the calibration parameters between the event cameras and LiDAR. The network is then tasked with predicting these offsets, effectively learning to correct the artificially induced decalibrations.

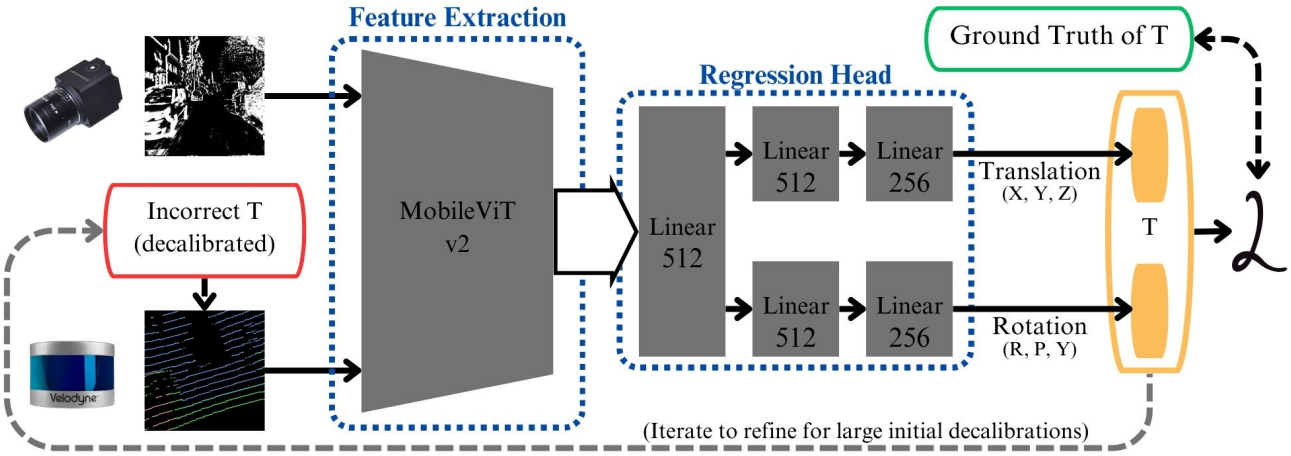


Figure 2. Overall architecture of MULi-Ev. The initial decalibrated extrinsic parameters T (three for rotation, and three for translation) are used to project the LiDAR point cloud into the event camera frame. Both input are then fed to a MobileViTv2 [12] backbone for feature extraction. The features are passed to a regression head, which regresses separately translation and rotation parameters. Together, they compose the output T, which the loss $\mathcal{L}$ compares to the known ground truth.

The smallest range used during our training focuses on recalibrating the most common yet most challenging and subtle decalibrations within $\pm 1°$ and $\pm 10cm$. However, our model is capable, using an approach similar to [3, 7, 11, 14], to correct larger decalibrations, by iterating through a cascade of networks trained on larger decalibrations. For our experiments, we use a cascade of two networks. A first network with a larger training range of up to $\pm 10°$ and $\pm 100cm$, giving us a rough estimate of the parameters (with an average error of $0.47°$ and $3.03cm$), well within the training range of the second network, trained on the $\pm 1°$ and $\pm 10cm$ range.

3.3.2 Optimization and Evaluation

Throughout the training, we employ Mean Square Error (MSE) regression losses (wildly used for regression tasks, and specifically on calibration tasks [14]) to minimize the difference between the predicted calibration parameters and the ground truth, derived from the original, unaltered DSEC [5] data. The model is trained with the Adam optimizer and a learning rate of 0.0001. Continuous evaluation on a validation set, separate from the training data, allows us to monitor the model's performance and adjust the training parameters accordingly to avoid overfitting and ensure optimal generalization.

4. Experiments

The effectiveness of our proposed deep-learning framework for online calibration of event cameras and LiDAR sensors is demonstrated through a series of experiments using the DSEC [5] dataset. This section outlines our experimental setup, evaluation metrics, comparisons with existing works, and the results achieved.

Split	Area	Time	Environment	Sequences
Training	Interlaken	Day	Rural	5
	Thun	Day	Suburban	1
	Zurich City	Day/Night	Urban	35
Test	Interlaken	Day	Rural	3
	Thun	Day	Suburban	2
	Zurich City	Day/Night	Urban	7

Table 2. Subsets of the DSEC [5] dataset by location of capture, and their characteristics.

4.1. Dataset

For our experiments, we leverage the DSEC dataset [5], a pioneering resource offering high-resolution stereo event camera data for driving scenarios and LiDAR. More specifically it relies on a Velodyne VLP-16 LiDAR (a 16 channels LiDAR), and Prophesee Gen3.1 monochrome event cameras with a 640×480 resolution. This dataset is particularly notable for its inclusion of challenging illumination conditions, ranging from night driving to direct sunlight scenarios, as well as urban, suburban, and rural environments, making it an ideal benchmark for our calibration framework. Its composition is detailed in Table 2.

4.2. Preprocessing

Preprocessing temporally aligns LiDAR and event camera data for our calibration framework, before entering the network. The steps include:

- **Projection of LiDAR Data:** Initial (erroneous) calibra-

tion parameters are used to project LiDAR point clouds into the event camera frame.

- **Temporal Synchronization:** LiDAR timestamps are used to synchronize the LiDAR data with asynchronous events from the event camera, ensuring accurate event accumulation over LiDAR scans.
- **Event Accumulation:** A 50ms window (evaluated in Section 4.5) is used for event accumulation, balancing scene representation detail with data volume.
- **Transformation to Event Frame:** Accumulated events are converted into an event frame (also evaluated in Section 4.5), preparing the data for neural network processing.

Data normalization is applied as a standard step, bringing LiDAR and event camera data to a same scale for optimal feature extraction.

4.3. Evaluation Metrics

We employ the Mean Absolute Error (MAE) to gauge the accuracy of our calibration technique, both for translational and rotational parameters. The MAE for translation components is defined as the average of the absolute discrepancies between the predicted and actual translation vectors, with each component's error given by:

$$\text{MAE}_{\text{trans}} = \frac{1}{N}\sum_{i=1}^{N} \left\|\mathbf{t}_{\text{pred},i} - \mathbf{t}_{\text{gt},i}\right\|_2 , \tag{1}$$

where $\mathbf{t}_{\text{pred},i}$ and $\mathbf{t}_{\text{gt},i}$ represent the predicted and ground truth translation vectors for the i-th sample, respectively, and N denotes the number of test samples.

For rotation, our network outputs Euler angles, which are converted into rotation matrices to facilitate a robust error computation. The angles are then converted back into Euler form to report errors in a more interpretable fashion. Consequently, the MAE for rotational components—Roll, Pitch, and Yaw—is calculated as follows:

$$\text{MAE}_{\text{rot}} = \frac{1}{N}\sum_{i=1}^{N} \left\|\text{Euler}(\mathbf{R}_{\text{rel},i})\right\| , \tag{2}$$

where $\mathbf{R}_{\text{rel},i}$ represents the relative rotation matrix for the i-th sample, obtained by the operation $\mathbf{R}_{\text{pred},i} \times \mathbf{R}_{\text{gt},i}^{-1}$. The function $\text{Euler}(\cdot)$ converts this matrix to Euler angles, expressing the rotational discrepancy in terms of Roll, Pitch, and Yaw. The norm $\|\cdot\|$ then quantifies the magnitude of these angles, yielding the rotational error in degrees. This methodology allows for a precise measurement of rotational calibration performance across the dataset.

4.4. Experimental Results

Existing methods [8, 15, 16, 18] being offline approaches, their authors chose to evaluate them on a few scenes that they captured themselves. However, considering our online, deep learning-based approach, we evaluated our method in a more systematic way, on the publicly available DSEC [5] dataset presented in Section 4.1. Moreover, most existing works measure the quality of their results through non-absolute, sensor-dependent metrics, such as reprojection error, which is more suitable when using targets, and can be affected by sensor resolution and lens distortion. One of the most recent works, LCE [8], is the most suitable for comparison with our method, as it not only offers state-of-the-art results, but also uses similar sensors (notably the same LiDAR, Velodyne VLP-16). It also communicates results in the same absolute metric as our work, measuring the Mean Absolute Error on rotation and translation.

Method	Translation Error (cm)	Rotation Error (deg)	Online	Execution Time (s)
L2E [16]	*N/A*	*N/A*	No	134
LCE-Calib [8]	1.5	0.3	No	*N/A*
MULi-Ev (Ours)	**0.81**	**0.10**	**Yes**	$<$ **0.1**

Table 3. Comparison of MULi-Ev to the state of the art.

Location	Translation Error (cm)	Rotation Error (deg)
Interlaken	1.07	0.12
Thun	0.59	0.08
Zurich City	0.40	0.08

Table 4. Evaluation of the mean absolute error of MULi-Ev on the location subsets of DSEC [5].

General Results Analysis: As demonstrated by the results in Table 3, MULi-Ev achieves superior calibration accuracy, reducing the translation error to an average of $0.81cm$ and rotation error to $0.1°$. These results are illustrated qualitatively in Figure 4 and detailed in box plots in Figure 3. Distinctively, MULi-Ev achieves these results while being, to our knowledge, the first online, targetless calibration method for this sensor setup. It bridges a significant gap in real-time operational needs while surpassing existing offline, target-dependent methods, such as [8]. Finally, MULi-Ev being deep learning-based, it manages to reach this accuracy in an execution time inferior to $0.1s$ on a GPU, while an offline method like [16] takes about $134s$ with its fastest optimizer.

Box Plots Analysis: Interestingly, it can be noticed in Figure 3 that results on translation axis Z and rotation axis Pitch tend to be less regular. This was also found in works focused on RGB-LiDAR calibration such as [11], and was thus expected. It can be explained by the physical nature of these axes that align with the vertical dimension, in which the LiDAR points density is much lower (the vertical res-

olution of the VLP-16 LiDAR is only $2°$, while its horizontal resolution is between $0.1°$ and $0.4°$). This low vertical resolution of the VLP-16 is mostly due to it having only 16 LiDAR rings, compared to 64 for the Velodyne HDL-64E used in the KITTI [6] dataset, which was most commonly used to evaluate RGB-LiDAR calibration methods [2, 3, 7, 9, 11, 14, 17].

Influence of the Environment: To further analyze the behavior of MULi-Ev on different types of scenes, we measured the average errors per location. The results are available in Table 4, while the characteristics and number of sequences in these locations were reported in Table 2. We observe in Table 4 that the best results are obtained in Zurich City, while the least accurate results were for scenes captured in Interlaken. This was expected, and we can infer from it two possible explanations: first, 35 sequences from Zurich were included in the training set, compared to 5 for Interlaken; second, scenes in Interlaken happen to be mostly rural, and thus to have generally less available features, especially long vertical edges like the ones offered by buildings. However, MULi-Ev performed better on the Thun scenes than Interlaken scenes, despite having even less sequences (only 1 for training). This tends to confirm our second hypothesis, as Thun offers more of a suburban environment, with enough human-built structures to offer more linear features. Another interesting fact is that while Interlaken and Thun scenes were all recorded by day, Zurich City sequences include night scenes, and still obtains the best accuracy, suggesting that MULi-Ev might adapt quite well to varying lighting conditions. Overall, results are at least on par with the state of the art or better for all three location subsets. Qualitative results in Figure 4 show successful recalibrations in different environment: in a tunnel, in a suburban zone, and on a rural road.

4.5. Ablation Study

To assess the impact of event data representation and accumulation time on the calibration accuracy, we conducted experiments on the DSEC [5] dataset.

Event Representations: The three different event representations considered and detailed in Section 3.2 (event frame, voxel grid, and time surface) were evaluated to determine the best performing one.

Accumulation Times: For the event frame representation, accumulation times of $30ms$, $50ms$, and $80ms$ were tested. These intervals were selected to explore the trade-off between temporal resolution and the richness of accumulated event information, potentially affecting the calibration's accuracy and robustness.

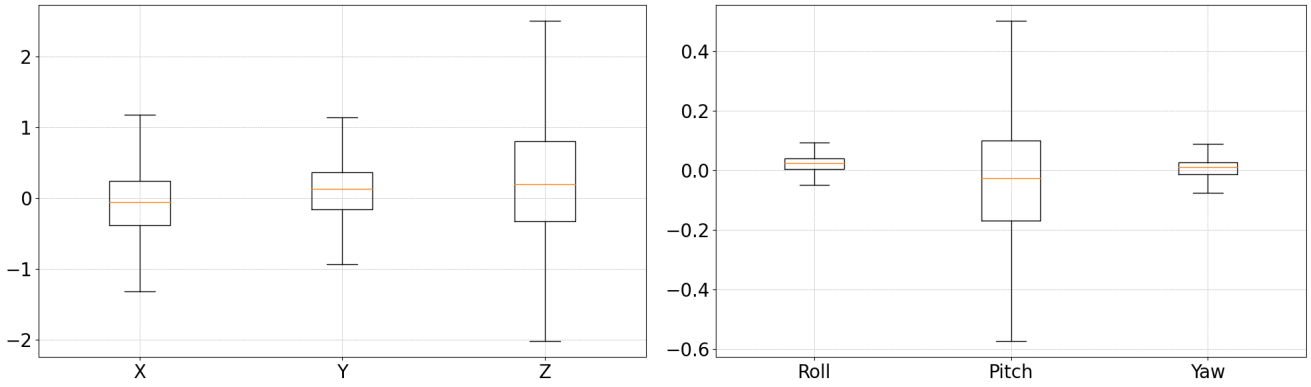
Configuration	Translation Error (cm)	Rotation Error (deg)
Event Frame (30 ms)	1.01	0.12
Event Frame (50 ms)	**0.81**	**0.10**
Event Frame (80 ms)	0.85	0.11
Voxel Grid (50 ms)	0.88	0.11
Time Surface (50 ms)	1.17	0.23

Table 5. Results of ablation experiments on DSEC [5] to determine the influence of event representation on the final calibration result (average error).

Ablation Results: Results reported in Table 5 indicate that the event frame representation, with an accumulation time of $50ms$, achieved the highest calibration accuracy. This suggests that a longer accumulation time might lead to higher noise levels, degrading the result. Conversely, shorter accumulation times, while offering fresher data, may not accumulate enough events to adequately represent the scene for effective calibration. The voxel grid and time surface representations, despite their more complex encoding of event data, did not yield improvements in calibration accuracy over the optimized event frame representation. These observations underscore the importance of the choice of event representation and accumulation period to optimize the results of our method.

4.6. Discussion

Results of our experiments in Table 3 demonstrate that MULi-Ev can provide better accuracy than existing offline works, and this in diverse environments (as demonstrated in Table 4) while being the first online method proposed for this sensor combination. As a comparison, deep learning-based methods for RGB-LiDAR sensor setups have initially reached MAE of $0.28°$ and $6cm$ [14], while more recent approaches reached $0.03°$ and $0.36cm$[11], showing there is probably still potential for improving the accuracy offered by MULi-Ev (currently $0.1°$ and $0.81cm$). MULi-Ev not only enhances operational convenience by eliminating the need for impractical calibration targets but also excels in dynamic environments where rapid recalibration is essential, thanks to its execution time of less than $0.1s$ offering on-the-fly recalibration capability. By ensuring immediate recalibration to maintain performance and safety, MULi-Ev can contribute to the robustness of autonomous navigation systems in real-world applications. Finally, results from Table 5 suggest that our choice of the simple event frame for event representation delivers the best results while simplifying the implementation of our method.



(a) Translation errors (cm)

(b) Rotation errors (degrees)

Figure 3. Box plots of translation and rotation errors on the test set of DSEC [5].

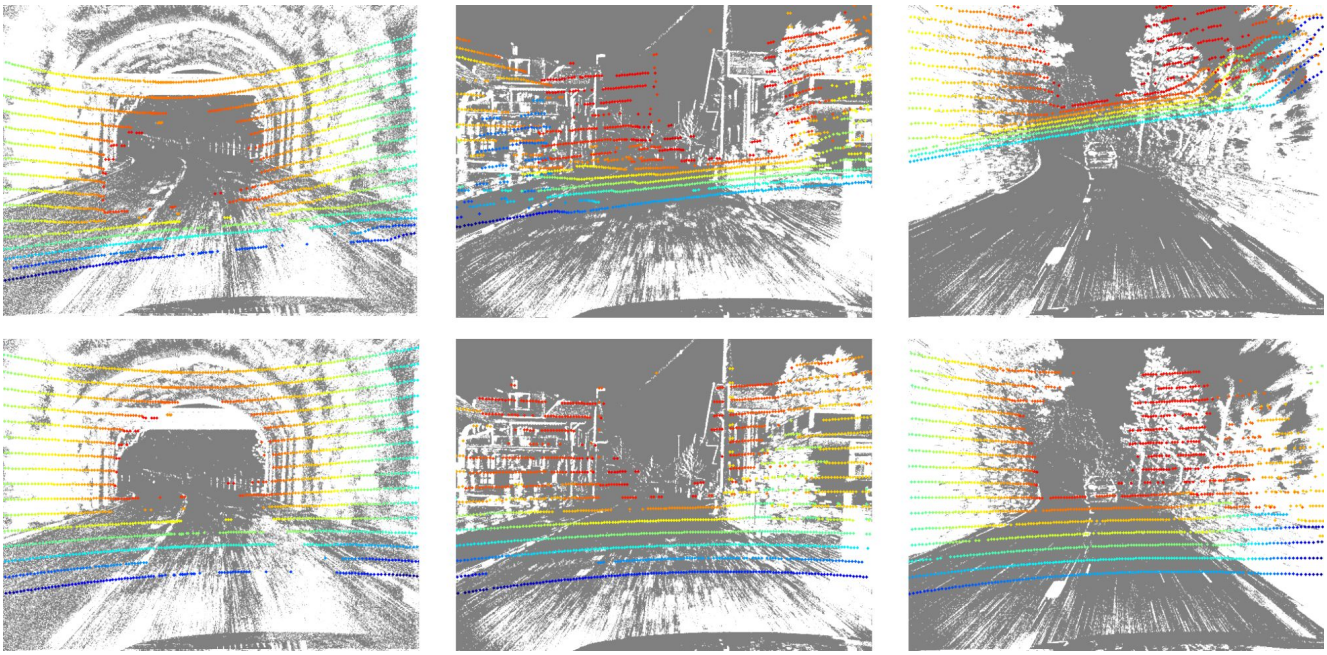


Figure 4. Qualitative results on DSEC [5], showing three examples of recalibration in diverse environments. Images show the LiDAR pointclouds projected on the event frame. The top lane presents random decalibrations applied to the setup, while the bottom lane presents the correction proposed by MULi-Ev.

5. Conclusion

In this work, we introduced MULi-Ev, a pioneering framework that establishes the feasibility of online, targetless calibration between event cameras and LiDAR. This innovation marks a significant departure from traditional, offline calibration methods, offering enhanced calibration accuracy and operational flexibility. The real-time capabilities of MULi-Ev not only pave the way for immediate sensor recalibration—a critical requirement for the dynamic environments encountered in autonomous driving—but also open up new avenues for adaptive sensor fusion in operational vehicles.

Looking ahead, we aim to further refine MULi-Ev's robustness and precision, with a particular focus on monitoring and adapting to the temporal evolution of calibration parameters. Such enhancements will ensure that MULi-Ev continues to deliver accurate sensor alignment even as conditions change over time. Additionally, we are interested in expanding the applicability of our framework to incorporate a wider array of sensor types and configurations. This expansion will enable more comprehensive and nuanced perception capabilities, ultimately facilitating the development

of more sophisticated autonomous systems.

As we move forward, our focus on refining MULi-Ev is aligned with the evolving demands of autonomous vehicle technology. By addressing the real-world challenges of sensor calibration and integration, MULi-Ev contributes to improving the safety, reliability, and performance of these systems. Our efforts to enhance sensor fusion and adaptability reflect a practical step towards achieving more robust and reliable autonomous driving capabilities.

Acknowledgements

This work was granted access to the HPC resources on the supercomputer Jean Zay of IDRIS under the allocation 2023-AD011014065 made by GENCI.
This work has been carried out within SIVALab, joint laboratory between Renault and Heudiasyc (CNRS / Université de technologie de Compiègne).

References

[1] Ryad Benosman, Charles Clercq, Xavier Lagorce, Sio-Hoi Ieng, and Chiara Bartolozzi. Event-based visual flow. *IEEE transactions on neural networks and learning systems*, 25 (2):407–417, 2013. 3

[2] Mathieu Cocheteux, Aaron Low, and Marius Bruehlmeier. UniCal: a single-branch transformer-based model for camera-to-lidar calibration and validation. *arXiv preprint arXiv:2304.09715*, 2023. 2, 3, 6

[3] Mathieu Cocheteux, Julien Moreau, and Franck Davoine. PseudoCal: Towards initialisation-free deep learning-based camera-lidar self-calibration. In *34th British Machine Vision Conference 2023, BMVC 2023, Aberdeen, UK, November 20-24, 2023*. BMVA, 2023. 2, 4, 6

[4] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):154–180, 2020. 1

[5] 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. 4, 5, 6, 7

[6] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013. 6

[7] Ganesh Iyer, R Karnik Ram, J Krishna Murthy, and K Madhava Krishna. CalibNet: Geometrically supervised extrinsic calibration using 3d spatial transformer networks. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1110–1117. IEEE, 2018. 2, 4, 6

[8] 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*, 2023. 1, 2, 5

[9] Xin Jing, Xiaqing Ding, Rong Xiong, Huanjun Deng, and Yue Wang. Dxq-net: differentiable lidar-camera extrinsic calibration using quality-aware flow. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6235–6241. IEEE, 2022. 2, 6

[10] You Li, Julien Moreau, and Javier Ibanez-Guzman. Emergent visual sensors for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 24(5):4716–4737, 2023. 1

[11] Xudong Lv, Boya Wang, Ziwen Dou, Dong Ye, and Shuo Wang. LCCNet: Lidar and camera self-calibration using cost volume network. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2888–2895, 2021. 2, 4, 5, 6

[12] Sachin Mehta and Mohammad Rastegari. Separable self-attention for mobile vision transformers. *Transactions on Machine Learning Research*, 2022. 3, 4

[13] Henri Rebecq, Timo Horstschaefer, and Davide Scaramuzza. Real-time visual-inertial odometry for event cameras using keyframe-based nonlinear optimization. 2017. 3

[14] Nick Schneider, Florian Piewak, Christoph Stiller, and Uwe Franke. RegNet: Multimodal sensor registration using deep neural networks. In *2017 IEEE intelligent vehicles symposium (IV)*, pages 1803–1810. IEEE, 2017. 2, 3, 4, 6

[15] 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. 1, 2, 5

[16] 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. 1, 2, 5

[17] Shan Wu, Amnir Hadachi, Damien Vivet, and Yadu Prabhakar. This is the way: Sensors auto-calibration approach based on deep learning for self-driving cars. *IEEE Sensors Journal*, 21(24):27779–27788, 2021. 2, 6

[18] 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*, 2023. 1, 2, 5

[19] Alex Zihao Zhu, Liangzhe Yuan, Kenneth Chaney, and Kostas Daniilidis. Unsupervised event-based learning of optical flow, depth, and egomotion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 989–997, 2019. 3