

# Why Event Cameras Are Still Not Widely Used in Industrial Robotics: A Barrier Analysis for Warehousing Logistics

Lokeshwaran Manohar<sup>1,2</sup>, Karol Piotr Rösner<sup>1,2</sup>, Jérôme Rutinowski<sup>1,2</sup>, Alice Kirchheim<sup>1,2,3</sup>

<sup>1</sup> Chair of Material Handling and Warehousing, TU Dortmund University, 44227 Dortmund, Germany

<sup>2</sup> Lamarr Institute for Machine Learning and Artificial Intelligence, Germany

<sup>3</sup> Fraunhofer Institute for Material Flow and Logistics IML, 44227 Dortmund, Germany

**E**vent cameras report per-pixel intensity changes asynchronously with microsecond latency, a high dynamic range and low power consumption, which makes them a natural fit for the high-speed autonomous mobile robots emerging in warehousing logistics. Despite this fit, they remain largely absent from documented industrial mobile-robot deployments. This paper examines the structural reasons for that gap through a barrier analysis along four dimensions: data availability, tooling and integration maturity, representation heterogeneity, and application fit. Drawing on a structured review of the event-based vision literature and on practical experience, we find that warehouse-specific datasets remain scarce and narrow in speed range even though event data are abundant overall, that drivers and software stacks remain fragmented across vendor and community projects even though calibration is largely solved, that no dominant event representation has emerged, and that the class of scenarios in which event cameras offer a decisive advantage is narrower than commonly assumed. We conclude that event cameras are genuinely useful for a well-defined subset of warehousing applications, above all high-speed navigation and perception under challenging lighting, but that broader adoption depends on progress in data, interfaces and representation together with a selection criterion based on the failure mode of the incumbent sensing stack.

[Keywords: event cameras, autonomous mobile robots, high-speed perception, warehousing logistics, industrial adoption]

## 1 Introduction

Mobile robots in logistics are becoming faster and more dynamic. Research platforms such as LoadRunner demonstrate vehicle speeds of up to approximately 10 m/s, while recent robotics datasets likewise motivate perception for platforms operating in the 5–10 m/s range [1, 2]. Increasing speed tightens the time available for sensing, inference and reaction. It also increases image motion, so long exposure times can produce motion blur precisely when robust perception is most important. These effects do not make conventional cameras unsuitable for warehousing in general, but they create operating conditions in which frame-based perception becomes harder.

Event cameras use a different sensing principle. Instead of acquiring complete images at fixed intervals, each pixel independently reports a change in logarithmic brightness. The resulting asynchronous stream offers microsecond-scale temporal resolution, high dynamic range and reduced exposure-based motion blur [3, 4]. These properties have enabled strong results in high-speed and high-dynamic-range vision, and they appear well matched to fast robotic motion.

Nevertheless, event cameras are still uncommon in publicly documented industrial mobile-robot systems. This gap cannot be explained by sensor specifications alone. A sensor that performs well in a laboratory still has to enter an engineering ecosystem: representative data must exist, drivers and calibration workflows must be maintainable, algorithms must accept the sensor output, interfaces must remain stable over the product lifecycle, and the resulting improvement must justify the additional integration effort. While recent work has begun to investigate event-based object detection in industrial multi-class environments [2, 5], most established detection benchmarks and pretrained models still originate from automotive or other non-warehouse domains.

Prior surveys discuss the general maturity of event-based vision, including hardware, algorithms, datasets and applications [4, 6, 7]. Related gap analyses in unmanned aerial vehicles identify similar issues around datasets, simulation, integration and ground truth [8]. These works are important references for the field, but they do not answer the application-specific question faced by a warehouse integrator: *under which conditions does an event camera provide enough system-level benefit to justify a less mature perception stack?*

This paper addresses that question through a barrier analysis for warehousing logistics. The analysis examines four dimensions:

1. Domain-specific data availability.
2. Tooling and integration maturity.
3. Representation heterogeneity.
4. Application fit.

The analysis distinguishes barriers that can plausibly shrink through technical and ecosystem progress from constraints that follow from the sensing principle itself. The contribution is therefore not another catalogue of event-camera advantages, but an application-oriented assessment of why those advantages have not yet translated into broad warehousing adoption and where they already justify deployment.

## 2 Related Work

An event camera measures brightness changes rather than absolute intensity frames. A commonly used event model represents the  $k$ -th event as

$$e_k = (x_k, y_k, t_k, p_k), \quad (1)$$

where  $(x_k, y_k)$  is the pixel address,  $t_k$  is the timestamp and  $p_k \in \{-1, +1\}$  denotes the polarity of the brightness change. An event is triggered when the change in logarithmic intensity exceeds a contrast threshold [3, 4]. Since pixels operate independently and asynchronously, the output rate follows scene activity rather than a fixed frame rate.

This sensing principle creates the best-known advantages of event cameras, namely very fine temporal resolution, high dynamic range and robustness to exposure-based motion blur [4]. It also creates their central engineering difficulty. The output is a sparse, irregular spatiotemporal stream rather than a dense image tensor. Classical image-processing pipelines therefore do not transfer directly, and modern learning systems usually require an event representation, an event-native network, or a reconstruction step [6, 9].

## 2.1 Perception Requirements in Warehousing Logistics

Warehousing does not define a single perception regime. Conventional automated guided vehicles and many autonomous mobile robots operate in structured environments where frame cameras, RGB-D cameras, lidar and safety scanners already provide mature solutions. Other systems pursue substantially higher vehicle speeds, fast material flow or perception under strong illumination transitions. The relevant variable is therefore not vehicle speed alone but the relative image motion, exposure setting, scene depth, illumination, task latency and required information content.

This distinction is important for event sensing. Event cameras become informative when edges or illumination patterns move across pixels. Conversely, low-motion conditions can yield little task-relevant event activity. Hybrid visual-inertial work has explicitly exploited this complementarity. Conventional images provide rich appearance information under good lighting and limited motion, while events support tracking in high-speed and high-dynamic-range conditions [10]. For warehousing, this suggests that the practical question is often whether an event camera should *complement* an established sensor stack rather than replace it.

## 2.2 Existing Gap Analyses

The limited adoption of event cameras outside research has been discussed at a broader level. Cimarelli et al. review neuromorphic vision hardware, algorithms and applications and identify gaps in hardware maturity, processing, data and real-world transfer [7]. Akanbi and Ayomoh reach a comparable conclusion for UAV applications, emphasizing limited real-world validation, non-standardized simulation, integration effort and ground-truth data [8]. The present paper uses these general findings as a starting point but evaluates them against warehousing requirements and against developments available by 2026.

## 3 Method

This paper uses a structured evidence review and barrier analysis rather than a statistical meta-analysis. Established and recent peer-reviewed surveys of event-based vision were used to identify recurring challenges related to datasets, algorithms, sensor integration, and applications. Relevant primary publications were subsequently examined to assess whether these challenges remain open or have been addressed. For rapidly evolving aspects such as software support, licensing, and standardization, official manufacturer documentation, ROS documentation, and standardization sources were consulted.

Four barrier dimensions were retained because they recur in event-vision surveys and also map directly to an in-

dustrial integration process: data, tooling and integration, representation, and application fit. For each dimension we ask three questions: *What evidence exists in 2026? What practical consequence does the issue have for a warehouse deployment? Has the issue been largely addressed, only partially addressed, or does it remain open?* Application fit is labelled scenario-dependent rather than open because it is not a defect to be solved; it describes the operating conditions under which the sensing principle is useful.

#### 4 Barriers to Industrial Adoption

The event-camera ecosystem has matured considerably in recent years. Large datasets and established learning methods are now available [4, 6], event-camera calibration has progressed from specialized procedures to practical calibration frameworks [11, 12], and ROS integration is available for major sensor ecosystems [13, 14]. The barriers to industrial adoption therefore no longer arise primarily from a complete absence of supporting technology.

Instead, the remaining challenges concern how readily these components can be transferred into warehousing applications. In the following, they are examined in terms of domain-specific data availability, tooling and integration maturity, representation heterogeneity, and application fit. The first three are largely ecosystem-related barriers that may decrease as datasets, interfaces, and software mature. Application fit is different: because event cameras respond to changes in brightness rather than sampling complete images, their benefit inherently depends on the motion, illumination, and perception requirements of the application [4].

##### 4.1 Data Availability and Dataset Limitations

Event data are no longer scarce in absolute terms. Automotive perception has large labeled collections: GEN1 contains more than 39 hours of recordings with car and pedestrian annotations, while the 1MP dataset introduced by Perot et al. contains more than 14 hours and approximately 25 million bounding boxes [15, 16]. DSEC provides stereo event data for driving, and MVSEC and M3ED support a broader set of robotic perception tasks [17, 18, 19]. The barrier is therefore not the absence of event-camera datasets; it is the shortage of datasets that match warehouse objects, motion patterns and evaluation tasks.

Recent datasets demonstrate progress but also illustrate the remaining domain gap. PEDRo targets person detection for robotics with 119 moving-camera recordings and 43,259 manually annotated boxes, yet it is intentionally person-specific [20]. EDAT24 addresses manufacturing assembly actions, but records 400 samples with a static DAVIS240C camera [21]. MTevent is the closest match to warehousing logistics. It provides synchronized stereo-event and RGB recordings across 75 scenes with a mix of

Table 1: Selected event datasets and their relevance to warehousing.

| Dataset             | Domain               | Warehouse limitation / value                                                           |
|---------------------|----------------------|----------------------------------------------------------------------------------------|
| GEN1 / 1MP [15, 16] | Automotive detection | Large-scale bounding-box supervision, but focused on road objects and driving scenes.  |
| PEDRo [20]          | Mobile robotics      | Relevant for human detection, but without warehouse load-carrier classes.              |
| EDAT24 [21]         | Manufacturing        | Industrial setting, but with a static camera and only four primitive actions.          |
| MTevent [2]         | Warehouse            | Closest domain match; 75 scenes and 16 objects, but camera motion is limited to 2 m/s. |

industrial and household objects, 6D-pose-related annotations and moving-object scenarios [2]. Its authors nevertheless identify a key limitation for high-speed mobile robotics: the moving camera reaches only 2 m/s, below the speed capability of the platforms that motivate the dataset [2].

Recent results further indicate that the limitation is not only the amount of available event data, but also its domain relevance. A benchmark of recurrent event-based object detection on MTevent shows that pretraining on the automotive GEN1 dataset improves industrial multi-class detection, whereas initialization from the person-oriented PEDRo dataset performs worse than training from scratch [5]. Large event datasets can therefore not be treated as interchangeable training resources; their value depends on how closely source task and source domain match the target application.

MTevent also reports an RGB FoundationPose baseline with Average Recall 0.22 using ground-truth masks in its dynamic scenes [2]. This result is useful evidence that conventional frame perception can be stressed by the recorded conditions, but it is not a universal crossover threshold between sensing modalities. Such a threshold depends on exposure, target motion, optics, scene geometry and downstream processing.

Synthetic data can reduce the recording burden. ESIM provides physically motivated event simulation with ground-truth motion quantities, while video-to-event conversion allows existing video datasets to be reused for training event-based models [22, 23]. Simulation is therefore an important tool, but it does not remove the need for warehouse recordings: sensor noise, bias settings, illumination flicker and event-rate behaviour remain device- and scene-dependent. Recent work continues to benchmark simulated-event realism and sensor-bias tuning, indicating that reproducibility at the sensing level remains an active topic [24, 25].

For industrial adoption, the consequence is straightforward. A company can acquire a camera today, but it cannot download a mature warehouse benchmark that covers load carriers, humans, occlusion, long range, lighting variation and high vehicle speed at the scale available in automotive vision. The cost of data collection, synchronization, annotation and validation therefore remains part of the adoption cost.

## 4.2 Tooling and Integration Maturity

The software ecosystem has matured considerably, and ROS support itself is no longer a fundamental barrier. Prophesee provides an official ROS wrapper for its event sensors based on OpenEB and targeting ROS Noetic [13]. ROS 2 support is also available through community-maintained projects. For example, the `ros-event-camera` project provides a Metavision-based ROS 2 driver whose documentation explicitly states that it is not written or supported by Prophesee [26]. For iniVation devices, the manufacturer documents a `dv-ros` integration based on DV-Processing [14].

The remaining challenge is therefore not the absence of integration software, but its fragmentation across vendor-specific and community-maintained stacks. For industrial deployment, this increases the effort required to maintain compatibility and software support over the lifetime of a robotic system.

This distinction matters in industrial engineering. Different vendors expose different SDKs, recording formats, bias controls and firmware dependencies. A ROS message can hide part of the transport difference, but acquisition still depends on device-specific libraries and encodings. The same is true for software licensing. Prophesee's Metavision SDK retains open-source OpenEB components, but since version 5.3 commercial or production deployment of applications developed with the SDK requires a separate commercial license agreement [27]. This is not a technical limitation, but it becomes a procurement and lifecycle consideration that is largely absent when using mature commodity frame-camera interfaces.

Calibration has progressed more strongly. Muglikar et al. provided a practical calibration framework that reconstructs intensity-like views from events so conventional target-based calibration can be used [11]. More recent methods include dynamic event-frame stereo calibration [28] and eKalibr, which estimates event-camera intrinsics directly from event geometry [12]. Vendor tooling also provides mono and stereo calibration workflows for current devices [29]. Calibration therefore should no longer be presented as an unsolved fundamental barrier. The remaining difficulty is system integration: heterogeneous rigs may combine event cameras with RGB, depth, lidar, IMU and external synchronization, so calibration, time alignment and revalidation still require specialized engineering.

Overall, the tooling barrier has changed in character. A prototype can be built with available SDKs, ROS packages and calibration tools. The unresolved industrial question is whether the complete stack has the interface stability, vendor support, licensing clarity and maintenance ownership expected over a product lifecycle.

## 4.3 Representation Heterogeneity

Even after events have been acquired reliably, there is no single representation analogous to the RGB image. Common choices include event-count images and histograms, voxel grids that preserve coarse temporal structure, time-surface families, and learned representations optimized jointly with the downstream task [4, 30, 9, 6]. The choice affects memory, latency, sparsity, temporal precision and compatibility with conventional neural-network operators. No representation is uniformly best because those trade-offs depend on the application and model architecture.

This heterogeneity has a direct industrial consequence. A published detector is not defined only by its network weights: it is also tied to the accumulation interval, number of bins, polarity handling, normalization, sensor resolution and preprocessing pipeline used to construct its input. Replacing a camera, changing the event encoding or modifying the latency target can therefore require retraining and revalidation rather than a simple interface substitution. This limits the portability of models and makes headline accuracy values harder to compare across papers.

Fragmentation also extends to data formats and software interfaces. Event-camera ecosystems still rely on different encodings and processing stacks, which limits straightforward interoperability across sensors and software pipelines. Standardization efforts are beginning to address this issue. JPEG XE (ISO/IEC 26112), for example, defines an event coding system intended to support interoperability between event sensing, storage, and processing [31]. However, such standardization is still emerging and has not yet resulted in a broadly adopted cross-vendor format.

Representation heterogeneity is therefore the barrier least likely to disappear through one algorithmic contribution. Progress requires convergence in interfaces, benchmark protocols and interchange formats even if multiple task-specific representations continue to coexist internally.

## 4.4 Application Fit

The first three barriers concern maturity and ecosystem coordination. Application fit is different: it follows from what an event camera measures. Events are generated by temporal contrast. High relative motion, rapid illumination changes and low-latency control naturally create informative event streams, while a nearly static camera observing a nearly static scene can produce little task-relevant activity.

Table 2: Qualitative application-fit guide for warehousing.

| Operating condition       | Event-camera suitability                                                   |
|---------------------------|----------------------------------------------------------------------------|
| High motion / low latency | High – temporal resolution and reduced exposure-based blur are beneficial. |
| Rapid lighting changes    | High – wide dynamic range and asynchronous sensing are advantageous.       |
| Static, well-lit scenes   | Low – frame-based sensing is typically sufficient.                         |
| Colour / texture critical | Low as stand-alone sensing; frame information remains valuable.            |
| Mixed conditions          | Moderate – complementary frame/event sensing may be preferable.            |

Conventional images show the complementary behaviour: they retain absolute appearance and texture regardless of motion, and are often preferable when lighting is benign and temporal resolution is not critical [4, 10].

This means that a single speed limit cannot define the point at which event cameras become worthwhile. Image motion depends on vehicle and object velocity, distance, focal length, camera orientation and vibration. Frame-camera blur additionally depends on exposure time, while the end-to-end reaction delay depends on frame period, image transfer and inference. Lighting, required colour or texture information, and the availability of complementary sensors change the comparison again. The relevant quantity is therefore the failure margin of the *existing perception task*, not vehicle speed by itself.

Peer-reviewed industrial applications support this task-dependent view. Tang et al. use event sensing for surface-defect detection on aluminium substrates in highly dynamic industrial scenes [32]. Ishikawa et al. use event cameras for non-contact motor-vibration sensing and anomaly classification [33]. Fast-object counting has likewise been demonstrated with event data in a production-line-style setting [34]. These applications share a common pattern: the signal of interest is temporal, fast, or difficult to sample efficiently with conventional frames. Their success does not imply that a warehouse navigation stack benefits automatically. Rather, these results show where the sensing principle aligns directly with the task.

For warehousing logistics, the practical implication is not that event cameras require extreme vehicle speed. They are justified when a measurable system limitation persists with the established sensing stack. Such limitations include motion blur, illumination failure, transient-event loss and reaction latency. If conventional RGB, RGB-D or range sensing already satisfies the required performance with adequate margin, the event camera adds integration cost without a demonstrated return.

Table 3: Status of the four barrier dimensions in 2026.

| Barrier and status             | Main finding                                                                                                               |
|--------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| <b>Data</b>                    |                                                                                                                            |
| <i>Partially addressed</i>     | Large event datasets exist, but warehouse-specific data and high-speed recordings remain limited.                          |
| <b>Tooling and integration</b> |                                                                                                                            |
| <i>Partially addressed</i>     | Drivers and calibration are usable, but fragmented software, lifecycle support, and licensing increase integration effort. |
| <b>Representation</b>          |                                                                                                                            |
| <i>Open</i>                    | Multiple event representations coexist, limiting model portability and direct comparison.                                  |
| <b>Application fit</b>         |                                                                                                                            |
| <i>Scenario-dependent</i>      | Benefits are strongest under high relative motion, difficult illumination, or strict latency requirements.                 |

## 5 Discussion

The barrier analysis does not support a binary verdict that event cameras are either “ready” or “not ready” for industry. The four dimensions have matured at different rates, and this unevenness is the actual result. Table 3 summarizes the assessment. Data and tooling are classified as partially addressed. Large-scale event datasets, stable drivers and usable calibration procedures exist, but warehouse-specific and high-speed recordings remain scarce, and the software landscape is fragmented with respect to lifecycle support and licensing. Representation is the only dimension classified as open, because competing event encodings coexist and thereby limit both model portability and the direct comparison of published results. Application fit is not a maturity level at all, but a scenario-dependent property whose benefit is concentrated in regimes with high relative motion, difficult illumination or strict latency requirements. No dimension is therefore fully closed, one remains conceptually unresolved, and the practically decisive criterion lies outside sensor maturity itself. The classification is deliberately conservative. A research solution is not treated as industrially settled merely because a method has been published, while mature software or calibration methods are acknowledged even where the wider ecosystem remains fragmented.

The most important result is that application fit can dominate these maturity issues. If a warehouse perception problem is already solved reliably by a conventional camera, RGB-D sensor or lidar, then better event-camera tooling alone does not create a reason to replace it. Conversely, when the incumbent fails because of fast image motion, illumination dynamics or required reaction time, the additional integration effort can become justified. This explains why

event cameras can be compelling in specialized industrial measurement and inspection tasks while remaining uncommon as a general-purpose warehouse sensor.

This perspective also changes how pilot studies should be designed. A fair comparison should not start from sensor specifications and ask which camera is “better”. It should first define an operating envelope and a system-level failure criterion. Relevant controlled variables include relative motion, exposure time, illumination range, target distance and motion, and the latency budget. Relevant outputs include task accuracy or recall, time-to-detection, reaction distance, compute load and data rate. Engineering effort and life-cycle constraints should then be considered alongside perception performance. The 0.22 Average Recall of the RGB pose baseline on MTeven is evidence of a difficult dynamic regime in one dataset [2]. It should not be interpreted as the point at which all warehouse systems should switch modalities.

A second implication is that complementarity deserves more attention than replacement. Hybrid event-frame systems can use conventional images when appearance information is strong and events when motion or illumination becomes challenging [10]. Warehouses already rely on heterogeneous sensor stacks for safety and navigation, so an event camera may be more realistic as an additional high-temporal-resolution channel than as a replacement for RGB or range sensing.

The analysis further suggests three practical priorities for broader adoption. First, datasets should move from generic event-camera scale toward warehouse coverage: synchronized event/frame recordings, industrial objects and humans, long-range perception, documented exposure and illumination, and motion extending beyond the current 2 m/s MTeven limit. Second, acquisition and storage should become less vendor-dependent. The ROS event-message ecosystem and JPEG XE standardization are encouraging steps, but long-term interoperability will require stable production support. Third, the field needs controlled crossover studies that quantify when an event-assisted perception stack starts to outperform a mature frame-based stack in system-level metrics rather than sensor-level specifications.

Several limitations qualify these conclusions. This paper is a structured, application-oriented evidence review rather than an exhaustive systematic review. Public information on commercial deployments is incomplete, and the absence of a published warehouse deployment does not prove that no proprietary deployment exists. Software and licensing statements rely on official documentation because these facts change rapidly, whereas the technical analysis is grounded primarily in peer-reviewed work. Finally, the application-fit guide in Table 2 is qualitative: its boundaries should be validated experimentally across speed, exposure, illumination,

scene geometry and task type. The present paper identifies the variables that such an evaluation should control rather than claiming a universal numerical threshold.

## 6 Conclusion and Outlook

Event cameras offer properties that are highly relevant to fast and visually difficult robotic operation, but those properties alone do not guarantee a useful warehouse system. This paper analyzed four reasons why adoption remains limited: domain-specific data, tooling and integration, representation heterogeneity, and application fit.

The 2026 evidence shows substantial progress. Calibration is no longer a fundamental unsolved problem, practical ROS and vendor software stacks exist, and event datasets have grown considerably. The remaining barriers are more specific: warehouse datasets still cover a limited operating envelope, software interfaces and data formats remain fragmented, commercial deployment raises questions of long-term software support and licensing, and perception models remain strongly dependent on the chosen event representation. Although standardization efforts are emerging, broad interoperability across event-camera ecosystems has not yet been established.

Most importantly, event cameras should be selected by failure mode rather than novelty. High relative motion, difficult illumination and strict reaction-time requirements are conditions in which asynchronous sensing can add a measurable advantage. Slow, well-lit and appearance-rich tasks usually favour the maturity and simplicity of conventional sensing. For many warehouse systems, the most plausible path to adoption is therefore complementary event sensing in the parts of the operating envelope where the established perception stack loses margin. Quantifying that crossover under controlled warehouse conditions is the key next step from promising sensor technology to defensible industrial deployment.

## REFERENCES

- [1] M. ten Hompel, H. Bayhan, J. Behling, L. Benkenstein, J. Emmerich, G. Follert, M. Grzenia, C. Hammermeister, H. Hasse, D. Hoening *et al.*, “Technical report: Loadrunner®, a new platform approach on collaborative logistics services,” *Logistics Journal: Editorial-Reviewed*, Oct. 2020. [Online]. Available: <https://edrev.logistics-journal.de/article/view/206>
- [2] S. Awasthi, A. Gouda, S. Franke, J. Rutinowski, F. Hoffmann, and M. Roidl, “Mtevent: a multi-task event camera dataset for 6d pose estimation and moving object detection,” in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025, pp. 5063–5071.
- [3] P. Lichtsteiner, C. Posch, and T. Delbruck, “A  $128 \times 128$  120 db 15  $\mu$ s latency asynchronous temporal contrast vision sensor,” *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008.
- [4] G. Gallego, T. Delbrück, 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*, vol. 44, no. 1, pp. 154–180, 2022.
- [5] L. Manohar and M. Roidl, “Benchmarking recurrent event-based object detection for industrial multi-class recognition on mtevent,” *arXiv preprint arXiv:2603.21787*, 2026.
- [6] X. Zheng, Y. Liu, Y. Lu, T. Hua, T. Pan, W. Zhang, D. Tao, and L. Wang, “Deep Learning for Event-based Vision: A Comprehensive Survey and Benchmarks,” Feb. 2023. [Online]. Available: <https://arxiv.org/abs/2302.08890v3>
- [7] C. Cimorelli, J. A. Millan-Romera, H. Voos, and J. L. Sanchez-Lopez, “Hardware, Algorithms, and Applications of the Neuromorphic Vision Sensor: A Review,” *Sensors (Basel)*, vol. 25, no. 19, p. 6208, Oct. 2025. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12526923/>
- [8] I. Akanbi and M. Ayomoh, “Event-Based Vision Application on Autonomous Unmanned Aerial Vehicle: A Systematic Review of Prospects and Challenges,” *Sensors (Basel)*, vol. 26, no. 1, p. 81, Dec. 2025. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12787784/>
- [9] D. Gehrig, A. Loquercio, K. Derpanis, and D. Scaramuzza, “End-to-end learning of representations for asynchronous event-based data,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, 2019, pp. 5632–5642.
- [10] A. R. Vidal, H. Rebecq, T. Horstschaefer, and D. Scaramuzza, “Ultimate slam? combining events, images, and imu for robust visual slam in hdr and high-speed scenarios,” *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 994–1001, 2018.
- [11] M. Muglikar, M. Gehrig, D. Gehrig, and D. Scaramuzza, “How to calibrate your event camera,” in *IEEE Conf. Comput. Vis. Pattern Recog. Workshops (CVPRW)*, June 2021.
- [12] S. Chen, X. Li, L. Yuan, and Z. Liu, “ekalibr: Dynamic intrinsic calibration for event cameras from first principles of events,” *IEEE Robotics and Automation Letters*, vol. 10, no. 7, pp. 7094–7101, 2025.
- [13] Prophesee, “Prophesee ros wrapper: Ros driver for prophesee event-based sensors,” GitHub repository, accessed: 2026-08-07. [Online]. Available: [https://github.com/prophesee-ai/prophesee\\_ros\\_wrapper](https://github.com/prophesee-ai/prophesee_ros_wrapper)
- [14] iniVation AG, “DV-Ros: Ros integration for inivation event cameras,” 2026, accessed: 2026-08-07. [Online]. Available: <https://docs.inivation.com/software/other-software.html>
- [15] P. de Tournemire, D. Nitti, E. Perot, D. Migliore, and A. Sironi, “A Large Scale Event-based Detection Dataset for Automotive,” Jan. 2020. [Online]. Available: <https://arxiv.org/abs/2001.08499v3>
- [16] E. Perot, P. de Tournemire, D. Nitti, J. Masci, and A. Sironi, “Learning to Detect Objects with a 1 Megapixel Event Camera,” in *Advances in Neural Information Processing Systems*, vol. 33. Curran Associates, Inc., 2020, pp. 16 639–16 652. [Online]. Available: <https://papers.nips.cc/paper/2020/hash/c213877427b46fa96cff6c39e837ccee-Abstract.html>
- [17] M. Gehrig, W. Aarents, D. Gehrig, and D. Scaramuzza, “Dsec: A stereo event camera dataset for driving scenarios,” *IEEE Robotics and Automation Letters*, 2021.
- [18] A. Z. Zhu, D. Thakur, T. Özaslan, 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*, vol. 3, no. 3, pp. 2032–2039, 2018.
- [19] K. Chaney, F. Cladera, Z. Wang, A. Bisulco, M. A. Hsieh, C. Korpela, V. Kumar, C. J. Taylor, and K. Daniilidis, “M3ed: Multi-robot, multi-sensor, multi-environment event dataset,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2023, pp. 4015–4022.
- [20] C. Boretti, P. Bich, F. Pareschi, L. Prono, R. Rovatti, and G. Setti, “PEDRo: An event-based dataset for person detection in robotics,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, pp. 4065–4070.
- [21] L. Duarte and P. Neto, “Event-based dataset for the detection and classification of manufacturing assembly tasks,” *Data in Brief*, vol. 54, p. 110340, 2024.

- [22] H. Rebecq, D. Gehrig, and D. Scaramuzza, “ESIM: an open event camera simulator,” Oct. 2018.
- [23] D. Gehrig, M. Gehrig, J. Hidalgo-Carrio, and D. Scaramuzza, “Video to events: Recycling video datasets for event cameras,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 3586–3595.
- [24] A. Ziegler, D. Joseph, T. Gossard, E. Moldovan, and A. Zell, “BiasBench: A reproducible benchmark for tuning the biases of event cameras,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025, pp. 5033–5042.
- [25] K. Chanda, A. Verma, A. Vaghela, Y. Yang, and B. Chakravarthi, “Event quality score (EQS): Assessing the realism of simulated event camera streams via distance in latent space,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025, pp. 5144–5152.
- [26] ros-event-camera, “metavision\_driver: Ros 2 driver for event-based cameras using the metavision sdk,” GitHub repository, accessed: 2026-08-07. [Online]. Available: [https://github.com/ros-event-camera/metavision\\_driver](https://github.com/ros-event-camera/metavision_driver)
- [27] Prophesee, “Metavision sdk 5.3.1 release notes,” 2026, accessed: 2026-08-07. [Online]. Available: [https://docs.prophesee.ai/stable/release\\_notes.html](https://docs.prophesee.ai/stable/release_notes.html)
- [28] R. Hu, J. Kogler, M. Gelautz, M. Lin, and Y. Xia, “A dynamic calibration framework for the event-frame stereo camera system,” *IEEE Robotics and Automation Letters*, vol. 9, no. 12, pp. 11 465–11 472, 2024.
- [29] iniVation AG, “Calibrate an event camera with DV,” 2026, accessed: 2026-08-07. [Online]. Available: <https://docs.inivation.com/software/dv/gui/calibrate-event-camera.html>
- [30] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, “Unsupervised event-based learning of optical flow, depth, and egomotion,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 989–997.
- [31] Joint Photographic Experts Group (JPEG), “JPEG XE: Event coding systems, ITU-T T.JPEG-XE — ISO/IEC 26112,” 2026, standard under development; accessed: 2026-08-07. [Online]. Available: <https://jpeg.org/jpegxe/index.html>
- [32] J. Tang, Z. Gong, and Y. Fan, “High-performance surface defect detection of aluminum substrate based on event camera,” *Journal of Computational Design and Engineering*, vol. 11, no. 6, pp. 232–243, 2024.
- [33] Y. ISHIKAWA, T. SATO, K. IWATA, and Y. SATOH, “Event-camera-based non-contact sensing and anomaly classification of motor vibration,” *Journal of the Japan Society for Precision Engineering*, vol. 92, no. 3, pp. 243–247, 2026.
- [34] K. Bialik, M. Kowalczyk, K. Blachut, and T. Kryjak, “Fast object counting with an event camera,” *Pomiary Automatyka Robotyka*, vol. 27, pp. 79–84, 02 2023.

---

**Lokeshwaran Manohar, M.Sc.**, Research Associate at the Chair of Material Handling and Warehousing (FLW), TU Dortmund University.

**Karol Piotr Rösner, M.Sc.**, Research Associate at the Chair of Material Handling and Warehousing (FLW), TU Dortmund University.

**Dr.-Ing. Jérôme Rutinowski**, Head of Research and Operations at the Chair of Material Handling and Warehousing (FLW), TU Dortmund University.

**Prof. Dr.-Ing. Alice Kirchheim**, Director of Fraunhofer Institute for Material Flow and Logistics IML, Head of the Chair of Material Handling and Warehousing (FLW), TU Dortmund University, and Area Chair Planning & Logistics at the Lamarr Institute for Machine Learning and Artificial Intelligence.

Address: Chair of Material Handling and Warehousing (FLW), LogistikCampus, Joseph-von-Fraunhofer-Str. 2-4, 44227 Dortmund, Germany

E-Mail: [lokeshwaran.manohar@tu-dortmund.de](mailto:lokeshwaran.manohar@tu-dortmund.de)