Event-Based Vision for Robust SLAM: An Evaluation Using Hyper E2VID Event Reconstruction Algorithm

Hamza Anwar*[1,2], Anayat Ullah[1,2]
[1]Department of Electronic Engineering, Balochistan University of IT, Engineering and Management Sciences, Quetta, Pakistan.
[2]Control, Automotive and Robotics Lab, National Centre of Robotics and Automation, Rawalpindi, Pakistan.
***Correspondence:** hamzaanwar5893@gmail.com

Citation | Anwar. H, Ullah. A, "Event-Based Vision for Robust SLAM: An Evaluation Using Hyper E2VID Event Reconstruction Algorithm", IJIST, Special Issue. pp 131-145, Oct 2024
Received | Oct 08, 2024 **Revised |** Oct 14, 2024 **Accepted |** Oct 18, 2024 **Published |** Oct 22, 2024.

This paper investigates the limitations of traditional visual sensors in challenging environments by integrating event-based cameras with visual SLAM (Simultaneous Localization and Mapping). The work presents a novel comparison between a visual-only SLAM implementation using the state-of-the-art HyperE2VID reconstruction method and conventional frame-based SLAM. Traditional cameras struggle in low dynamic range and motion blur scenarios, limitations that are addressed by event-based cameras, which offer high temporal resolution and robustness in such conditions. The study employs the HyperE2VID algorithm to reconstruct event frames from event data, which are then processed through the SLAM pipeline and compared with conventional frames. Performance metrics, including Absolute Pose Error (APE) and feature tracking performance, were evaluated by contrasting visual SLAM implementations on reconstructed images against those from traditional cameras across three event camera dataset sequences: Dynamic-6DoF, Poster-6DoF, and Slider depth sequence. Experimental results demonstrate that event-based cameras yield higher-quality reconstructions, significantly outperforming conventional cameras, especially in scenarios marked by motion blur and low dynamic range. Among the tested sequences, the Poster-6DoF sequence exhibited the best performance due to its information-rich scenes, while the Slider depth sequence faced challenges related to drag and scaling, as it lacked rotational motion. Although the APE values for the Slider depth sequence were the lowest, it did experience trajectory drift. In contrast, the Poster-6DoF sequence displayed superior overall performance, with reconstructions closely aligning with those produced by conventional camera-based SLAM. The Dynamic-6DoF sequence showed the poorest performance, marked by high absolute pose error and trajectory drift. Overall, these findings highlight the substantial improvements that event-based cameras can bring to SLAM systems operating in challenging environments characterized by motion blur and low dynamic ranges.

Keywords: SLAM, Event Camera, Neuromorphic, Feature Detection, Computer Vision.

Introduction:

In recent decades, the field of robotics has experienced significant advancements in sensors and algorithms, particularly in simultaneous localization and mapping (SLAM). Visual sensors are commonly favored in SLAM due to their compact size, lower power consumption, greater data richness, and affordability compared to alternatives like LiDAR. However, challenges such as motion blur (illustrated in Figure 1a and 1b) and low dynamic range (depicted in Figure 1c) can lead to feature loss and blurriness in images, thereby hindering both robot localization and efficient environmental mapping.

(a) Motion blur caused by moving sensor beside the mirror in the background scene.

(b) Motion blur caused by motion in the scene by the falling bottles [1]

(c) Low (left) vs. high (right) dynamic range [2]. High dynamic range can enlighten the image, making it capable of efficient feature detection/tracking

Figure 1. Most occurring issues in visual sensors - Motion Blur and Low Dynamic Range

To address the challenges faced by traditional visual sensors, neuromorphic retinas—commonly referred to as neuromorphic cameras or event-based cameras—were introduced in the mid-2000s. The first commercially viable event camera, the Dynamic Vision Sensor (DVS), was developed by the Institute of Neuroinformatics in Zurich [3]. Unlike conventional cameras, which capture images at fixed intervals, event cameras asynchronously detect pixel-wise brightness changes, generating a continuous stream of events, as illustrated in Figure 2 and represented by Equation (1):

$$e = x, y, p, t \quad (1)$$

In this context, \(x\) and \(y\) denote the pixel coordinates, \(p\) represents pixel polarity, and \(t\) indicates the time of the event occurrence. A value of \(p = -1\) signifies a transition from bright to dark, while \(p = 1\) indicates a transition from dark to light in edge intensity. The asynchronous, pixel-wise output of event cameras provides high temporal resolution, low power consumption, and enhanced robustness against motion blur and dynamic range challenges.

Literature Review:

Since the introduction of Mono-SLAM [5] in the early 2000s, SLAM algorithms have seen significant advancements in speed and efficiency, with notable examples including ORB-SLAM, LSD-SLAM, and VINS-Mono. While these algorithms perform well under moderate motion and proper lighting conditions, their efficiency is compromised by motion blur and low dynamic range data. This limitation has created a demand for alternative visual sensors better suited for high-speed robotics and SLAM applications. References [6] and [7] both present event-based visual-inertial odometry techniques that utilize event frames alongside an extended Kalman filter (EKF) backend. While [6] demonstrated minimal drift, its reliance on an event-only sparse frontend limited efficiency due to the availability of features. In contrast, [7] addressed this issue by incorporating an Inertial Measurement Unit (IMU) to acquire odometry, resulting in reduced pose errors. However, this algorithm still struggled in texture-less and motion-less environments, where event generation and feature availability were absent.

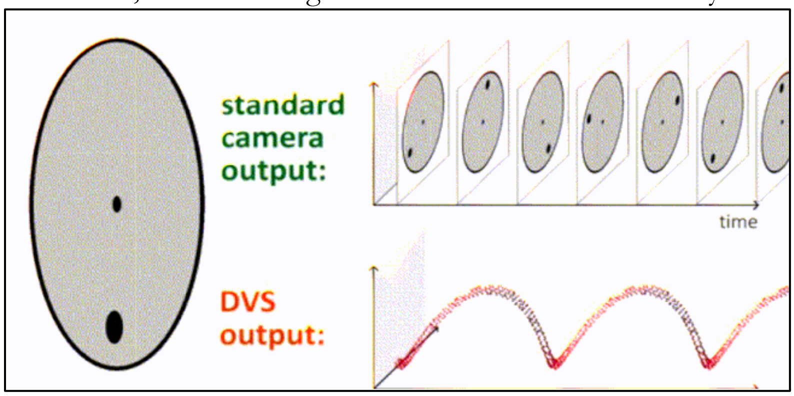


Figure 2. Output's comparison of standard camera with Event camera. (Adapted from [4]).

A. R. Vidal et al. [8] proposed a visual-inertial odometry pipeline that integrates event frames, standard frames, and IMU data. This approach uniquely leverages the complementary strengths of both standard and event cameras, enhancing accuracy in both normal and extreme motion scenarios. However, its optimization-based backend results in slower processing compared to other filter-based methods. Mahlknecht et al. [9] introduced a KLT-VIO [10]-based visual-inertial pipeline designed for drone-based planetary exploration. This pipeline incorporates event cameras alongside a laser range finder to improve scale estimation, utilizing an EKF-based backend that offers faster performance than optimization-based approaches. Nevertheless, its reliance on frames for feature extraction limits exploration in cave environments to areas illuminated by residual light from the entrance [9]. [11] presents an innovative event-based visual-inertial odometry pipeline that utilizes reconstructions of event frames through the time surface method, facilitating separate feature tracking and loop closure. However, the computational complexity inherent in the graph-based optimization results in slower performance. [12] introduces a monocular visual-inertial odometry (VIO) system that leverages both point and line features from event cameras. By integrating high-frequency event data with IMU measurements, this system improves pose estimation accuracy and robustness, particularly in dynamic and low-light conditions, thereby surpassing conventional VIO systems. [13] proposes an algorithm that fuses events, visual, and inertial measurements for precise state estimation and 3D dense mapping. Through tightly coupled fusion of the event camera and IMU, the system integrates feature matching with direct alignment for event-based 2D-2D alignment and reprojection, achieving robust real-time performance and making significant contributions to visual-inertial SLAM. Conventional computer vision algorithms typically require frames to process data, which has led to the development of several reconstruction algorithms, such as E2VID [14][15] and Hyper E2VID [16]. This paper specifically focuses on evaluating SLAM performance using the HyperE2VID algorithm, a state-of-the-art reconstruction method for event-based data.

To facilitate this evaluation, EVREAL [17] was employed to reconstruct frames from the event camera dataset (ECD). Both the reconstructed and real frames underwent feature extraction and tracking using the Shi-Tomasi feature detector and ORB descriptor.

Subsequently, the performance of SLAM was analyzed using PYSLAM across multiple sequences.

Objectives of the Study:

The objectives of this research are as follows:

- To compare the accuracy, efficiency, and robustness of SLAM using real camera frame images versus HyperE2VID-reconstructed images. This comparison will involve various SLAM metrics, including absolute pose errors and feature stability.
- To assess the applicability of the state-of-the-art HyperE2VID algorithm for SLAM tasks, evaluating whether the reconstructed images from event cameras deliver similar or improved performance compared to real images in terms of feature tracking and localization accuracy.

Novelty Statement:

This research is novel in two key aspects: first, it presents a unique comparison of visual-only SLAM implementations using the state-of-the-art HyperE2VID reconstruction method alongside conventional frame-based SLAM. No prior work has utilized a visual-only SLAM pipeline in this manner. Second, it is the first study to evaluate SLAM performance using HyperE2VID, which has not been applied in this context before.

Methodology:

This study aims to analyse the performance of various event-based reconstruction algorithms for SLAM, using conventional camera frames as a baseline for comparison. The experimentation deliberately excludes additional sensors, such as Inertial Measurement Units (IMUs), to prevent any undue advantage to the least performing reconstruction algorithm, ensuring a clearer and more direct comparison.

Data Collection:

Three sequences from the Event Camera Dataset (ECD) [18] were selected: Dynamic-6DoF, Poster-6DoF, and Slider Depth. The data were collected using the DAVIS-240, a dynamic and active-pixel vision sensor that combines both event and conventional camera capabilities, featuring a resolution of 240x180 pixels at 24 FPS. These datasets are utilized to analyze scenarios involving fast motion and varying dynamic ranges.

Image Reconstruction:

Conventional SLAM algorithms necessitate frame-based inputs, which is why the state-of-the-art event-reconstruction algorithm HyperE2VID is employed to transform event streams into frames using the EVREAL framework. HyperE2VID utilizes hyper-networks for E2VID, enhancing efficiency by generating per-pixel adaptive filters that improve both memory efficiency and the accuracy of the target network. Additionally, the model's context-aware fusion and curriculum learning techniques further enhance output quality while reducing training time and complexity.

Figure 3 presents a comparison between HyperE2VID (left) and the ground truth images (right), showing a lower Mean Squared Error (MSE) alongside moderate Structural Similarity Index (SSIM) scores. While HyperE2VID exhibits a lower SSIM, it generates human-interpretable images, as indicated by the higher LPIPS scores, resulting in better-lit and clearer scene representations.

For the comparison analysis, the ground truth information from the ECD dataset was set aside due to its higher resolution compared to the HyperE2VID reconstructions. Down-sampling the ground truth would lead to significant information loss, making it an ineffective basis for comparison. Therefore, conventional camera images from the same dataset were utilized as a baseline. Both the reconstructed images and the conventional camera images were processed through the same SLAM algorithms to facilitate a comprehensive comparison. Figure 4 presents a qualitative analysis of three sequences, with real camera frames on the left and the

HyperE2VID reconstructions on the right. The analysis indicates that the reconstructions are well-lit and contain more detailed information than the conventional frames.

Simultaneous Localization and Mapping Algorithm:

PYSLAM was chosen for the lightweight implementation of the SLAM algorithm. It integrates various feature detection and tracking techniques, enabling experimentation with algorithms such as ORB [19], ORB2 [20], SIFT [21], SURF [22], and Shi-Tomasi corners [23].

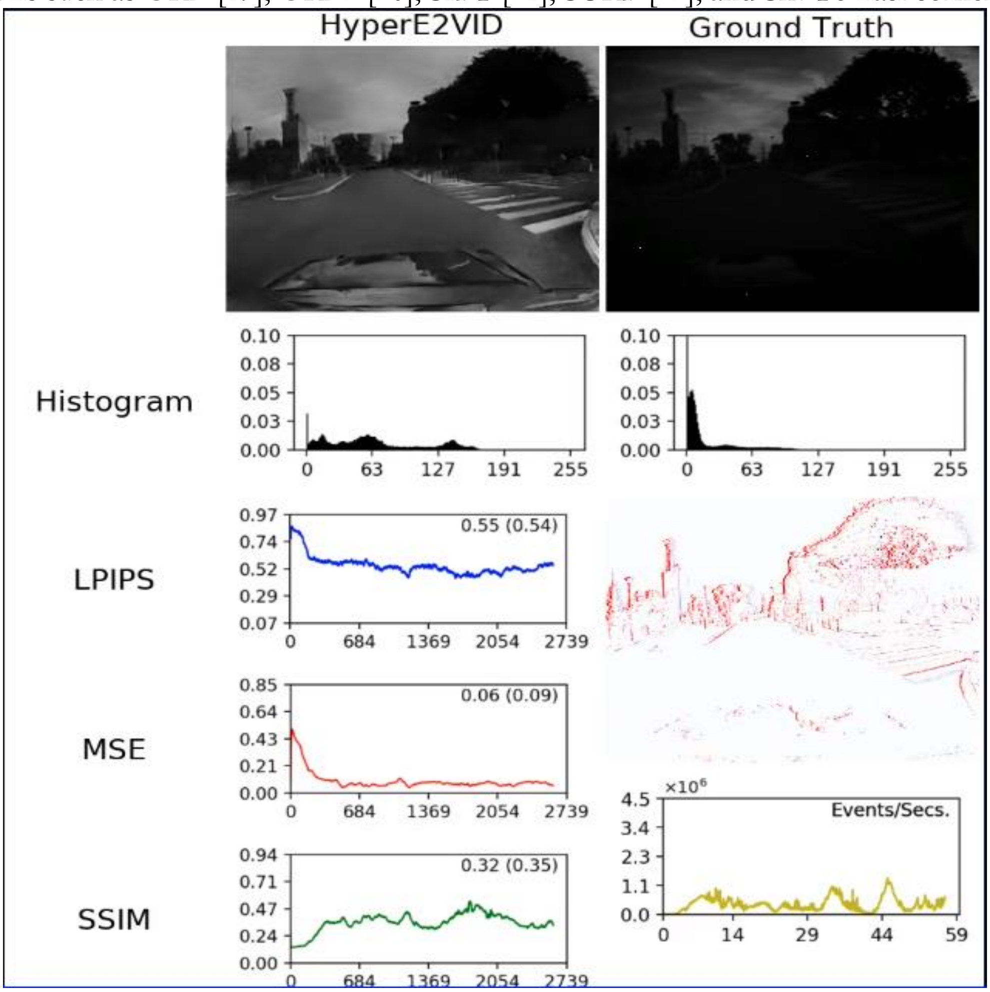


Figure 3. Results from EVREAL platform comparing the selected technique on LPIPS, SSIM and MSE metrics.

Shi-Tomasi corners are particularly robust against image noise and rotation, making them well-suited for low-resolution environments [23]. In contrast, ORB provides a fast and efficient feature tracking mechanism, which is essential for real-time SLAM applications [19]. Experimental results confirm that the most accurate outcomes were achieved by combining Shi-Tomasi corner detection with ORB feature tracking.

Figure 5 illustrates the experimentation process, which includes the following steps:

- **Processing the Event Camera Dataset:** The dataset is converted into a suitable format for input into EVREAL.
- **Image Reconstruction:** The preprocessed data is sent to the EVREAL tool to reconstruct images from the event data.
- **Grouping Reconstructed Images:** The reconstructed images are organized into two categories: real images and HyperE2VID images.
- **SLAM Pipeline Presentation:** Each group is presented to the SLAM pipeline using a combination of the Shi-Tomasi feature detector and the ORB feature descriptor.

Performance Analysis:

PYSLAM is applied to both event-reconstructed images and real images, allowing for performance analysis across various datasets.

(a)

(b)

(c)

Figure 4. Comparison of conventional camera frames (left) with HyperE2VID frames (right) in three sequences. (a) Dynamic-6DoF, (b) Poster-6DoF, (c) Slider Depth.

Result and Discussion:

To analyze the trajectories generated by the SLAM implementations, we employed Evolution of Odometry (EVO), an open-source Python library, to compare the resulting trajectories and conduct a quantitative analysis. This analysis utilized metrics such as Absolute Pose Error (APE), along with mean and standard deviation. Additionally, SLAM performance was evaluated in relation to feature matching accuracy and lighting conditions. This section presents the findings from the analysis of all three sequences.

Feature Tracking Analysis:

PYSLAM, utilizing the Shi-Tomasi feature detector and ORB feature tracker, was applied to both HyperE2VID reconstructed images and real frame-based images across all three sequences of the ECD dataset. The performance of feature detection and tracking is depicted in Figure 6. In this analysis, the black line represents the descriptor distance threshold \ (\phi \), which serves as the cutoff for feature matching. When feature distances fall below this threshold, the tracker loses the feature and necessitates a new keyframe. In the dynamic-6DoF sequence, feature matching for HyperE2VID reconstructed frames (Figure 6 (b)) frequently dipped below the descriptor distance threshold, leading to inadequate feature retention and indicating a need for frequent keyframe reinitializations. In contrast, the real frame-based tracking (Figure 6 (a)) demonstrated more consistent performance.

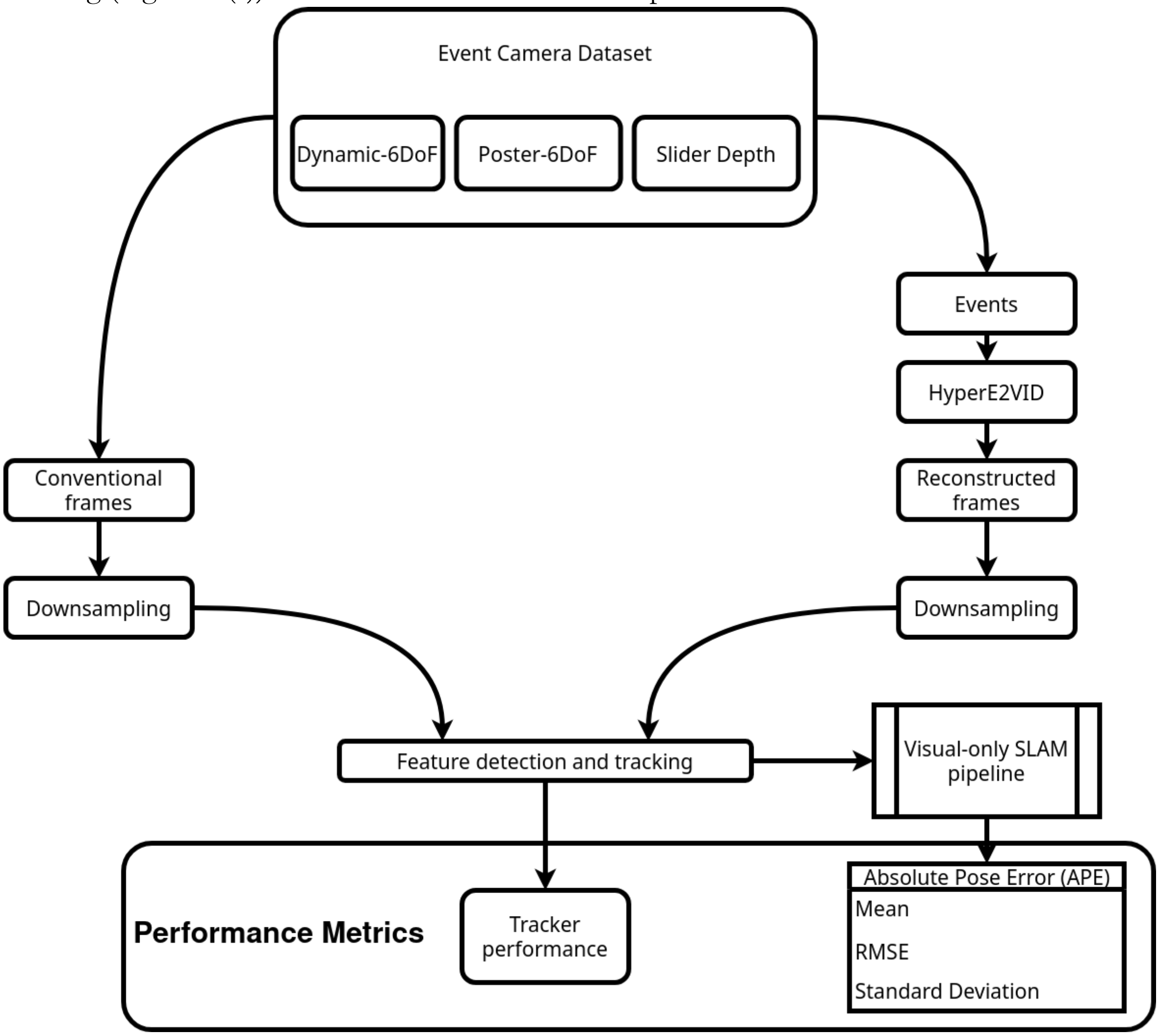


Figure 5. Study Flow Diagram of proposed Comparison between Event-based and conventional image-based V-SLAM pipeline.

The parameters for the conducted experimentation are detailed in Table 1.

The poster-6DoF sequence exhibits a close alignment between the real frames (Figure 6 (c)) and the reconstructed frames (Figure 6 (d)), thanks to the sequence's highly textured scene, which produced rich event data and led to improved reconstructions. Figures 6 (e) and 6 (f) display the feature performance graphs for conventional frames and HyperE2VID reconstructions, respectively. The reconstructions consistently maintained a higher number of features, staying well above the descriptor distance threshold and even outperforming the tracker based on real frames.

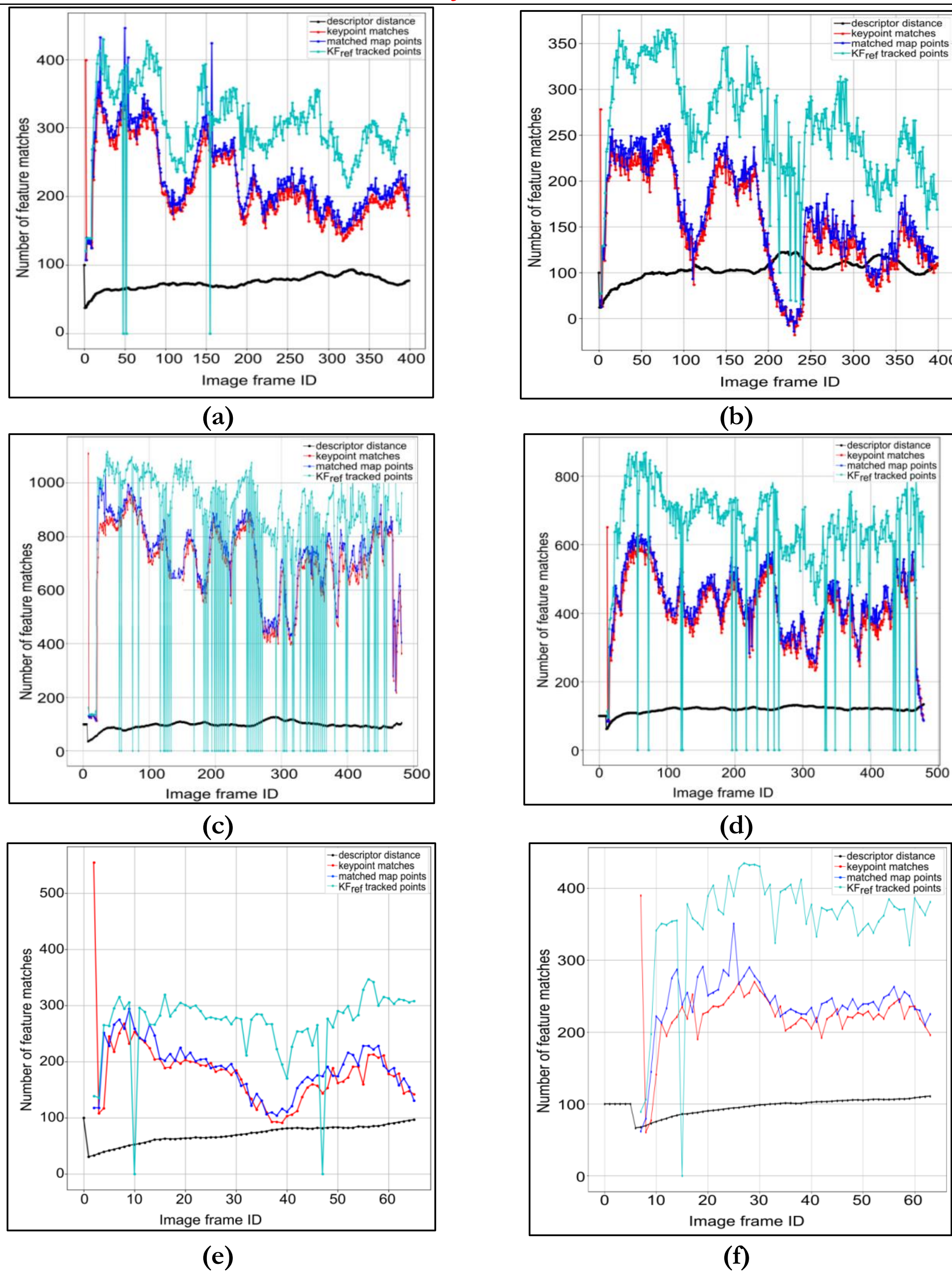


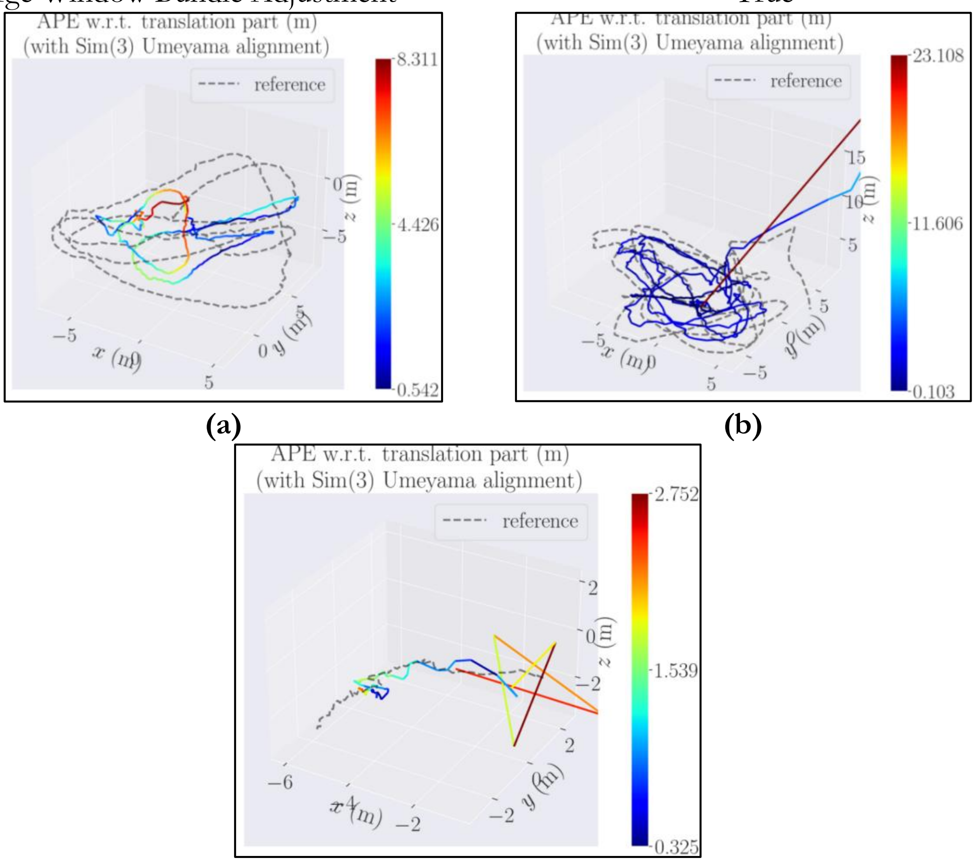
Figure 6. Feature detector and tracker performance evaluation for Dynamic-6DoF, poster-6 DoF and slider depth sequences.

Trajectory and Drift Analysis:

Figure 7 (a), (b), and (c) present a comparison of 3D trajectories for the dynamic-6DoF, poster-6DoF, and slider depth sequences, respectively. The dotted line represents the trajectory of conventional frames, while the HyperE2VID reconstructions are depicted in a rainbow spectrum, with dark red indicating the farthest trajectory drag and dark blue representing the least. In Figure 7 (a), trajectory drag accumulates due to scaling issues, leading to a higher Absolute Pose Error (APE). Some sections of the trajectory exhibit minimal drag, while others demonstrate significant scaling problems. Figure 7 (b) illustrates the poster-6DoF sequence, where the event-based trajectory (predominantly blue) closely aligns with the conventional camera's dotted path, resulting in a smaller APE and minimal scaling issues. In Figure 7 (c), the slider depth sequence shows significant drift and scaling challenges in the reconstruction, attributed to the absence of rotational motion.

Table 1. Parameters for the conducted experimentation.

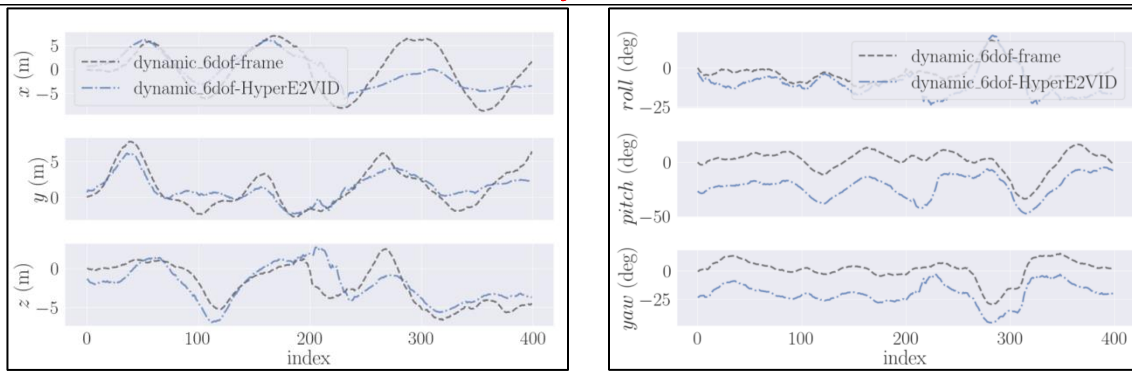
Parameter	Value
Detector type	Shi-Tomasi (Good Features to Track)
Descriptor type	ORB
Number of levels	8
Number of desired features in a single frame	2000
Scale factor	1.9
Minimum number of features	500
Large Window Bundle Adjustment	True



(c)

Figure 7. Trajectory and drift analysis of the Dynamic-6DoF, Poster-6DoF and Slider depth sequences. Frame based (black dotted line) and Hyper E2VID (rainbow).

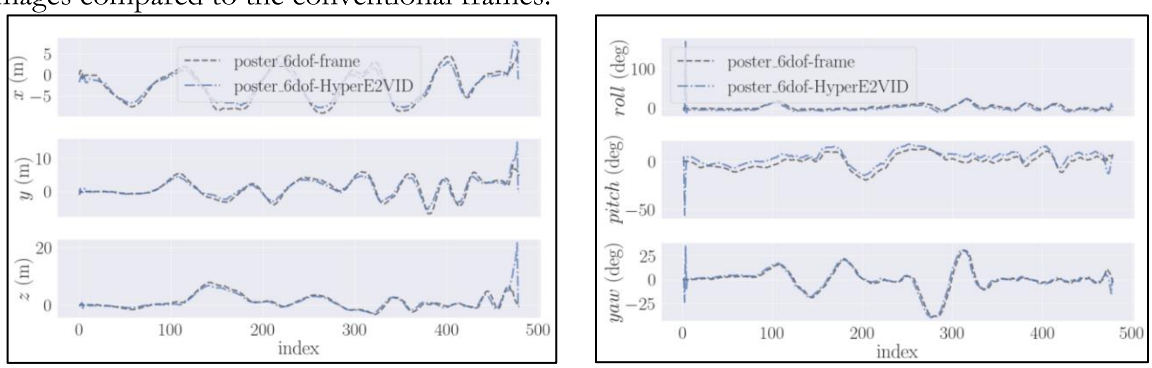
Figure 8 illustrates the trajectory error in two dimensions for both translational and rotational components. As depicted, the reconstructed trajectory experiences slight drift and scaling issues, with the translational error remaining below five meters, and the rotational drift under 25 degrees, as shown in Figures 8 (a) and (b) respectively.



(a) **(b)**

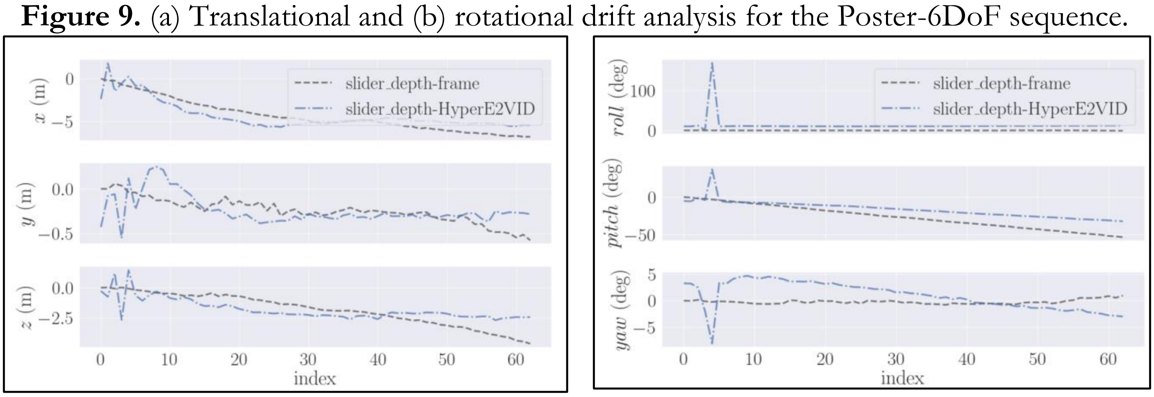
Figure 8. (a) Translational and (b) rotational drift analysis for the Dynamic-6DoF sequence.

Similarly, the Poster-6DoF sequence exhibits a comparatively smaller absolute pose error of 0.1 meters across the entire trajectory. The results provide insights into the trajectory in the translational (x, y, z) directions, as shown in Figure 9 (a), and analyze the rotation (roll, pitch, and yaw) in Figure 9 (b). The trajectory of the event camera closely follows the conventional frame-based reference trajectory, indicated by the dotted line, in both translational and rotational components. It can also be deduced that, in this specific Poster-6DoF sequence, the event camera outperformed the conventional camera due to its ability to capture better edges and a higher number of events. Consequently, the event-based reconstructions produced much clearer images compared to the conventional frames.



(a) **(b)**

Figure 9. (a) Translational and (b) rotational drift analysis for the Poster-6DoF sequence.

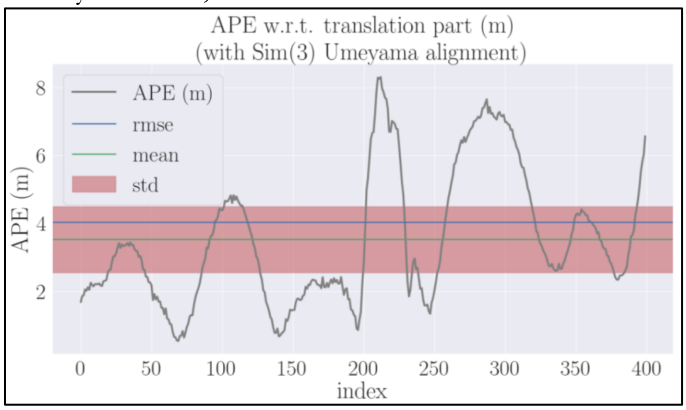


(a) **(b)**

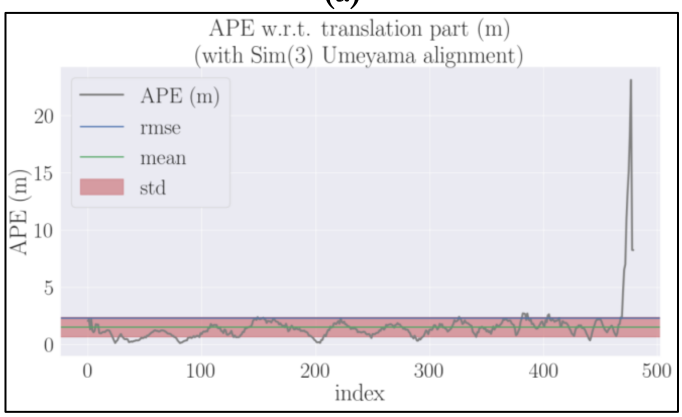
Figure 10. (a) Translational and (b) rotational drift analysis for the Slider depth sequence.

The third sequence, referred to as slider depth, was analyzed as illustrated in Figure 9. This sequence lacks any rotational components and consists solely of lateral translations, resulting in less effective reconstruction and significant scaling issues. This is evident in Figures 10 (a) and 10 (b), which depict the translation and rotational elements, respectively. The trajectory based on event frames shows noticeable drift from the traditional camera-based trajectory, primarily due to the absence of rotational motion in the scene. In contrast, during the rotational roll, the drift remains minor and does not accumulate significantly, as shown in Figure 10 (b).

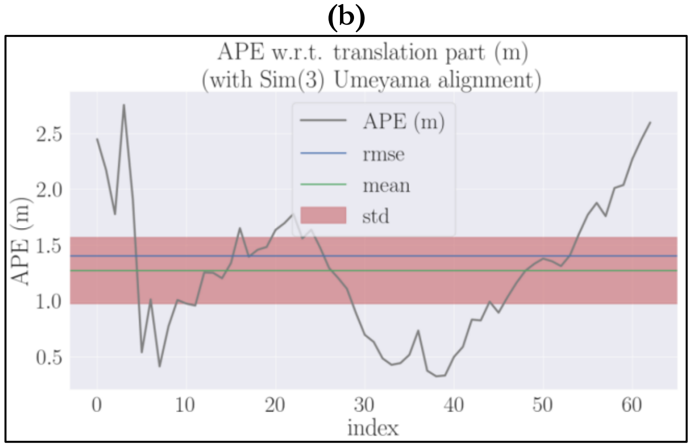
From this analysis, various metrics—including Absolute Pose Error (APE), Root Mean Square Error (RMSE), mean, and standard deviation (std)—were calculated and are presented in Figure 11. The results indicate that the APE for the trajectory of the Dynamic-6DoF sequence (Figure 11 (a)) fluctuates between 1 and 8 meters, with a mean value of 3.5 meters, a standard deviation of approximately 2 meters, and an RMSE of about 4 meters.



(a)



(b)



(c)

Figure 11. Absolute Pose Error for Dynamic-6DoF, Poster-6DoF and Slider depth sequences.

Figure 11 (b) illustrates that the Poster-6DoF sequence exhibits an exceptionally low APE, attributed to the availability of superior event information and more effective reconstructions, yielding a root mean square error of 2.5 meters. The trajectory reflects a mean APE value of 2 meters, with a standard deviation of only 1.5 meters.

In addition, Figure 11 (c) presents the APE values for the slider depth sequence, which fluctuate between 0.5 and 2.5 meters. The analysis reveals a mean absolute pose error (APE) of approximately 1.3 meters, an RMSE of 1.4 meters, and a standard deviation of about 0.5 meters. Although the APE for the slider depth sequence shows relatively favourable values, the trajectory experiences significant drift, diverging considerably from the reference trajectory.

Discussions:

The analysis of the visual SLAM algorithm across the three sequences reveals important insights into its performance relative to the characteristics of the data used. Notably, the Poster-6DoF sequence yielded the most favourable results, primarily due to its event-rich and edge-dense environment. The abundance of edge information facilitates robust feature detection and tracking, resulting in a low absolute pose error (APE) of 2.0 meters, a root means square error (RMSE) of 2.5 meters, and a standard deviation of 1.5 meters. This underscores the significance of comprehensive visual features in enhancing the accuracy of the SLAM algorithm.

In contrast, the Slider Depth sequence exhibited poorer qualitative performance due to its intrinsic limitations, despite achieving the lowest mean APE value of 1.3 meters. Trajectory analysis revealed significant scaling and drift issues, primarily stemming from the sequence's lack of rotational motion. The RMSE was recorded at 1.04 meters, with a standard deviation of only 0.5 meters, while the APE for the Slider Depth sequence varied from 0 to 2.5 meters. These metrics indicate that, while the absolute error appears favourable, the trajectory's consistency and alignment with the reference trajectory were severely compromised, as demonstrated by the trajectory analysis.

With a mean APE of 3.5 meters, an RMSE of 4.0 meters, and a standard deviation of 2.0 meters, the Dynamic-6DoF sequence exhibited intermediate performance. The dynamic conditions during data collection contributed to greater challenges in precise feature tracking and odometry estimation, leading to comparatively higher APE values. In summary, the performance of the SLAM algorithm is significantly influenced by the nature of the sequences employed. A summary of the results is presented in Table 2, highlighting the APE metrics across the three sequences.

Table 2. Absolute Pose Error (APE) for the three sequences showing its mean, RMSE and standard deviation.

Absolute Pose Error (APE) (Meters)	Dynamic-6DoF Sequence	Poster-6DoF Sequence	Slider Depth Sequence
Mean	3.5	2.0	1.3
RMSE	4.0	2.5	1.4
Standard Deviation	2.0	1.5	0.5

Conclusion:

This work investigates the advancements and challenges associated with vision-based Simultaneous Localization and Mapping (SLAM) algorithms, with a particular emphasis on the potential of event-based cameras to overcome the limitations of traditional visual sensors, such as motion blur and dynamic range issues. The study includes a comprehensive review of state-of-the-art visual and visual-inertial SLAM algorithms, highlighting their performance under various conditions while discussing the limitations inherent in their reliance on conventional cameras.

To address these challenges, this study introduces neuromorphic or event cameras, exploring their operation and notable advantages, including high temporal resolution and resilience to motion blur. To evaluate the effectiveness of event-based SLAM, the EVREAL framework was utilized, and reconstructions were performed using the cutting-edge algorithm HyperE2VID to generate images from event data. A visual SLAM pipeline was then implemented using the Shi-Tomasi feature detector and ORB feature tracker, allowing for a comparison of results against traditional camera images using metrics such as Absolute Pose Error (APE) and feature tracking performance.

The findings indicate that event-based cameras provide significant improvements, yielding more efficient reconstructed frames in scenarios characterized by high-speed motion and challenging lighting conditions. The images reconstructed from event data demonstrated performance that is comparable, if not superior, in SLAM tasks, thereby validating the feasibility and advantages of integrating event-based cameras into SLAM systems. Furthermore, it is anticipated that these results could be enhanced by incorporating an IMU or other sensors through sensor fusion with event-based cameras, enabling real-time, high-fidelity mapping and localization in dynamic environments. Additionally, employing an event-based pipeline, such as spiking neural networks for event processing, could offer a computationally efficient and robust solution, fully leveraging the capabilities of event cameras.

Acknowledgement: This research is conducted at Control Automotive and Robotics Lab (CARL-BUITEMS), funded by National Center of Robotics and Automation (NCRA) with the collaboration of Higher Education Commission (HEC) of Pakistan. We hereby declare that this manuscript has not been previously published, nor is it under consideration for publication elsewhere.

Author's Contribution: The experimentation for this study was conducted by Hamza Anwar, under the supervision of his MS supervisor, Anayat Ullah. All authors have contributed significantly to this work and are in full agreement with the content of the manuscript. The project idea originated from Anayat Ullah, while Hamza Anwar was responsible for the initial draft of the manuscript. Anayat Ullah reviewed and revised the paper to its final version.

Conflict of Interest: We report no conflict of interest.

Project Details: Control, Automotive and Robotics Lab (CARL) was developed in 2018 under the HEC call for the Establishment of National Center of Robotics and Automation (NCRA). CARL is one of the 11 labs in 13 Universities established under this call. CARL's mandate is to promote AI and Robotics in Baluchistan and work on Disaster Management and Human Machine Interaction for the wellbeing and economic improvement of Pakistan in general and Baluchistan in particular. The total funding of the project is 7.983 million PKR. Project duration was 2018 through 2021 which was extended up to 2025.

References:

1. Poladchegivara.persiangig.com. "Motion Blur." [Online; accessed Aug 3, 2022]. Aug. 2013. URL: http://poladchegivara.persiangig.com/image/picbank/Motion20Blur/.
2. Alex. "Low Dynamic Range." [Online; accessed Aug 3, 2022]. Nov. 2015. URL: https://www.dronetrest.com/t/fpv-cameras-for-your-drone-what-you-need-to-know-before-you-buy-one/1441.
3. Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. "A 128-128 120 dB 15 us Latency Asynchronous Temporal Contrast Vision Sensor." IEEE Journal of Solid- State Circuits, 2008, 43(2), pp. 566–576. ISSN: 0018-9200. DOI: 10.1109/JSSC.2007.914337. URL: http://ieeexplore.ieee.org/document/4444573/ (visited on 06/04/2024).
4. Elias Mueggler, Basil Huber, and Davide Scaramuzza. "Event-based, 6-DOF pose tracking for high-speed maneuvers." 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014, pp. 2761–2768. DOI: 10.1109/IROS.2014.6942940.
5. Andrew J Davison et al. "MonoSLAM: Real-time single camera SLAM." IEEE transactions on pattern analysis and machine intelligence, 2007, 29(6), pp. 1052–1067.
6. Alex Zihao Zhu, Nikolay Atanasov, and Kostas Daniilidis. "Event-based visual inertial odometry". In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017, pp. 5391–5399.

7. Henri Rebecq, Timo Horstschaefer, and Davide Scaramuzza. "Real-time Visual- Inertial Odometry for Event Cameras using Keyframe-based Nonlinear Optimization." BMVC, 2017.
8. Antoni Rosinol Vidal et al. "Ultimate SLAM? Combining events, images, and IMU for robust visual SLAM in HDR and high-speed scenarios". In: IEEE Robotics and Automation Letters 3.2 (2018), pp. 994–1001.
9. Florian Mahlknecht et al. "Exploring Event Camera-based Odometry for Planetary Robots." arXiv preprint arXiv:2204.05880, 2022.
10. Jeff Delaune, David S Bayard, and Roland Brockers. "Range-visual-inertial odometry: Scale observability without excitation." IEEE Robotics and Automation Letters, 2021, 6(2), pp. 2421–2428.
11. Weipeng Guan and Peng Lu. "Monocular Event Visual Inertial Odometry based on Event-corner using Sliding Windows Graph-based Optimization." 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2022, pp. 2438–2445. DOI: 10.1109/IROS47612.2022.9981970.
12. Weipeng Guan et al. "PL-EVIO: Robust Monocular Event-based Visual Inertial Odometry with Point and Line Features." IEEE Transactions on Automation Science and Engineering, 2023.
13. Weipeng Guan et al. "EVI-SAM: Robust, Real-time, Tightly-coupled Event-Visual-Inertial State Estimation and 3D Dense Mapping." Advanced Intelligent Systems, 2024.
14. Henri Rebecq et al. "High Speed and High Dynamic Range Video with an Event Camera." IEEE Trans. Pattern Anal. Mach. Intell., 2021, 43(6), pp. 1964–1980. ISSN: 0162-8828, 2160-9292, 1939-3539. DOI: 10.1109/TPAMI.2019.2963386. URL: https://ieeexplore.ieee.org/document/8946715/ (visited on 05/25/2024).
15. Henri Rebecq et al. "Events-To-Video: Bringing Modern Computer Vision to Event Cameras." 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, June 2019, pp. 3852–3861. ISBN: 978-1-72813-293-8. DOI: 10.1109/CVPR.2019.00398. URL: https://ieeexplore.ieee.org/document/8953722/ (visited on 05/25/2024).
16. Burak Ercan et al. "HyperE2VID: Improving Event-Based Video Reconstruction via Hypernetworks." IEEE Trans. on Image Process., 2024, 33, pp. 1826–1837. ISSN: 1057-7149, 1941-0042. DOI: 10.1109/TIP.2024.3372460. URL: https://ieeexplore.ieee.org/document/10462903/ (visited on 05/25/2024).
17. Burak Ercan et al. "EVREAL: Towards a Comprehensive Benchmark and Analysis Suite for Event-based Video Reconstruction." 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2023, pp. 3943–3952. URL: https://api.semanticscholar.org/CorpusID:258427067.
18. Elias Mueggler et al. "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM." The International Journal of Robotics Research, 2017, 36(2), pp. 142–149. ISSN: 0278-3649, 1741-3176. DOI: 10.1177/0278364917691115. URL: http://journals.sagepub.com/doi/10.1177/0278364917691115 (visited on 06/05/2024).
19. Ethan Rublee et al. "ORB: an efficient alternative to SIFT or SURF." 2011, pp. 2564–2571. DOI: 10.1109/ICCV.2011.6126544.
20. Raúl Mur-Artal and Juan D. Tardós. "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras." IEEE Transactions on Robotics, 2017, 33(5), pp. 1255–1262. DOI: 10.1109/TRO.2017.2705103.
21. David G. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints." Int. J. Comput. Vision, 2004, 60(2), pp. 91–110. ISSN: 0920-5691. DOI:

10.1023/B.0000029664.99615.94. URL: http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94.

22. Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. "SURF: Speeded up robust features." Springer, 2006, vol. 3951, pp. 404–417. ISBN: 978-3-540-33832-1. DOI: 10.1007/11744023_32.
23. Jianbo Shi and Carlo Tomasi. "Good features to track." 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1994, pp. 593–600. URL: https://api.semanticscholar.org/CorpusID:778478.