An Overview of Event-Camera Based Approaches for SLAM

Mehdi Raza Khorasani
Robotics Engineering
University of Genoa
6164555@studenti.unige.it
Genoa, Italy

Tanvir Rahman Sajal
Robotics Engineering
University of Genoa
6007201@studenti.unige.it
Genoa, Italy

Ozan Pali
Robotics Engineering
University of Genoa
5831146@studenti.unige.it
Genoa, Italy

Girum Molla Desalegn
Robotics Engineering
University of Genoa
6020433@studenti.unige.it
Genoa, Italy

***Abstract*—This paper presents an exploration of the current landscape of event cameras, a novel type of vision sensor, and their application in Simultaneous Localization and Mapping (SLAM). Event cameras, unlike traditional cameras, capture intensity changes at each pixel independently and asynchronously, resulting in high temporal resolution, low latency, low power consumption, and better High Dynamic Range (HDR). However, their asynchronous and sparse output presents challenges for traditional frame-based vision algorithms. SLAM, a technique widely used for autonomous navigation and map generation, has traditionally been implemented using vision-based or Lidar-based methods. These methods suffer from high computational costs, poor low light performance, motion blur, and low HDR. The use of event cameras in SLAM aims to overcome these limitations. Various methods to process event data for SLAM have been developed. This paper provides a detailed review of these methods, their strengths and challenges.**

Keywords:- SLAM, Localization, Mapping, Tracking, Visual Odometry, Event-based camera, HDR, SNN, Motion Tracking, Bio-Inspired Vision. Sensor Fusion, Stereo Vision

I. Introduction

Event cameras, a novel type of vision sensor, have emerged as a promising alternative to traditional cameras due to their unique data-driven sensing mechanism. Unlike traditional cameras that capture frames at a fixed rate, event cameras capture intensity changes at each pixel independently and asynchronously. This results in high temporal resolution, low latency, low power consumption, and better High Dynamic Range (HDR). However, the asynchronous and sparse output of event cameras presents challenges for traditional frame-based vision algorithms.

Simultaneous Localization and Mapping (SLAM), a technique widely used for autonomous navigation and map generation, has been traditionally implemented using vision-based or Lidar-based methods. These methods, however, suffer from high computational costs, poor low light performance, motion blur, and low HDR. The use of event cameras in SLAM aims to overcome these limitations.

The unique output of event cameras has led to the development of various methods to process event data for SLAM. These methods can be broadly categorized into feature-based methods, direct methods, motion compensated methods, and deep learning-based methods. Each of these methods has its own strengths and challenges, and the choice of method depends on the specific application and task at hand.

This paper provides a comprehensive overview of event cameras and their application in SLAM. Section II and Section III build a high level context of event cameras and the approaches used in literature to achieve SLAM. Section IV provides a detailed review of state-of-the-art approaches for SLAM using event cameras and the challenges. Section V concludes the discussion.

II. Event cameras

Event cameras, as compared to traditional cameras differ in the way they capture information. A traditional camera works by capturing frames of images at a specific rate (by use of a clock). An event-based camera, in contrast, captures the intensity changes of each pixel independently. The output does not have a fixed rate and is asynchronous [6]. Each pixel remembers the last sent intensity level and continuously monitors for changes in that intensity level. If the intensity change exceeds a particular threshold, the pixel emits an event with the coordinates of the pixel, a timestamp, and sometimes the polarity of the change [6]. This mechanism makes event cameras as data-driven sensors i.e. their output depends on the amount of brightness change in the environment (which can be caused by motion for example) [6]. In the context of capturing motion, the number of events from the camera is directly proportional to the perceived motion.

Some advantages of event cameras are as follows:

1) **High temporal resolution**: The monitoring resolution of intensity is 1MHz [6], therefore event cameras can capture very fast motions without suffering from motion blur.
2) **Low Latency**: Low latency in capturing changes is achieved as each pixel operates independently [6]. Compared to a traditional camera which captures whole frames, there is no conception of global exposure time in event cameras.
3) **Low power consumption**: Since event cameras only capture changes in brightness intensity, redundant information is majorly removed when compared to a traditional camera where whole frames are captured and transmitted. [6]

4) **Better HDR**: Since each individual pixel captures the intensity change independently, we can capture very dark and very bright images. [6]

Since there is no free lunch in engineering, the following are some of the challenges of using event cameras:

1) **Async and Sparse output**: Since the output of the event camera is not a stream of frames, frame-based vision algorithms cannot be directly applied [6]
2) **Noise**: Like all vision sensors, these cameras also suffer from shot noise in photons and transistor noise [6].

Mathematically, the change in intensity (brightness) can be modeled using the photo-current I as follows:

$$L = log(I)$$

An event $e_k(x_k, y_k, t_k)$ is triggered at pixel (x_k, y_k) when

$$|\Delta L| = C$$

where C is a constant. The polarity can be represented as $p_k = sgn(\Delta L)$

Combining both equations, we can write $\Delta L = p_k C$

III. SLAM

For a robot to carry out a given task autonomously, it should be capable of navigating to a given goal, while updating its position in real-time and simultaneously generating a map of the environment. To this goal, SLAM is one of the most frequently employed techniques. In general, SLAM algorithms can be categorized into the following two types in the literature:

1) Vision-based SLAM (VSLAM)
2) Lidar Based SLAM

Traditional vision-based SLAMs are popular in the literature as they employ feature-matching techniques on frames of images to build a map of the surroundings. The VSLAM approaches generally suffer from high computational costs, bad low light performance, motion blur, and low HDR. Additionally, dynamic environments further degrade the performance as feature matching becomes increasingly difficult in such scenarios. [14].

The above characteristics are almost always present in real-life scenarios, hence the use of event cameras to achieve SLAM aims to eliminate these limitations of traditional vision-based sensors, as event cameras inherently do not suffer from these problems [14].

Event based SLAM

The approach to achieve SLAM using event cameras has been an emerging research question mainly due to the reason that traditional SLAM algorithms cannot be applied directly due to the different nature of output data.

Each event is caused by the apparent motion of intensity edges, hence the maps obtained by the majority of developed SLAM algorithms using event cameras consist of scene-edges [6]. The challenge here is to estimate the intensity gradient as the event camera only outputs intensity changes.

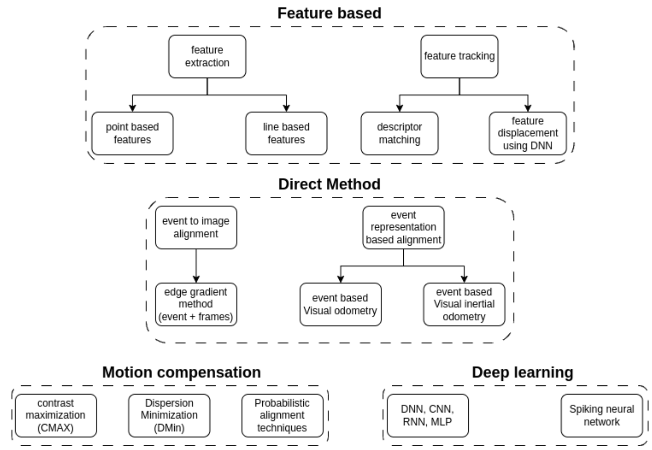

Fig. 1. summary of different SLAM approaches in literature

There are different methods available in the literature to process event data. These methods and algorithms can be selected based on the application and task at hand. [14] summarizes the available methods in literature in the following four categories (see Fig. 1)

1) Feature-based methods
2) Direct method
3) Motion compensated method
4) Deep Learning-based methods

Feature based methods

Feature-based algorithms primarily consist of two components:

1) Extraction and tracking of features
2) Tracking and mapping of the camera

Feature extraction

During feature extraction, features resistant to motion, noise, and illumination changes are first identified. The identified features are then tracked to fuse identical points across events. Using these features, the relative poses of the camera and the 3D landmarks of the features are estimated.

Feature extraction methods identify shapes from within the event stream. They are classified into two types:

1) Point-based features
2) Line-based features

Point-based features: Point-based features, such as the intersection of event edges, have been extracted using traditional techniques like frame-based corner detectors [15].

However, these methods face challenges like computational complexity and sensitivity to motion changes. To overcome these, learning-based approaches have been proposed, which recurrent neural networks (RNNs) to improve corner detection stability [4].

Line-based features: Line-based features are clusters of events along straight lines, that have been extracted using methods such as Line Segment Detector (LSD) [8]. An example of such an approach is [9], which applies LSD to a motion-compensated event stream.

Feature tracking

Feature-tracking algorithms are used to link events to the relevant features, updating their models of features like motion trajectory, locations, and 2D rigid body transformations. Methods like descriptor matching, and deep neural networks (DNN) that use DNN to predict feature displacement in subsequent events. Some trackers align local patches of the brightness incremental image from event data with feature patterns and estimate brightness changes. Feature tracking often involves modeling feature motions on the image plane, with methods that use some optimization steps for convergence [14].

Direct method

Direct methods for Visual Simultaneous Localization and Mapping (VSLAM) do not require explicit data association and directly align event data. These methods are divided into the following two:

1) Event-to-image alignment
2) Event representation-based alignment

Event-to-image alignment: Event-image alignment techniques, such as the Edge Gradient Method (EGM), leverage the link between brightness variations from events and frames. This technique fuses visual images and event data, correlating event data with corresponding frame pixels to estimate camera positions and depths. [10] presents an EDS with a direct monocular visual odometry using events and frames.

Event representation-based alignment: Event representation-based alignment techniques transform event data into 2D image-like representations. The Event-based visual odometry (EVO) method presented in [11] introduces a geometric strategy based on edge patterns for matching data from an event stream. Several techniques are proposed for estimating camera posture and velocity from event data. Some methods use non-linear optimization to process groupings of events concurrently to reduce the computational cost associated with updating camera positions on an event-by-event basis [14]

Motion compensated method

Motion-compensation techniques, based on event alignment, use the event frame as the fundamental event representation. These algorithms optimize event alignment in the motion-compensated event frame to predict camera motion parameters, providing clear images and reducing motion blur over a longer temporal window. However, these techniques may lead to event collapse, where a series of events accumulates into a line or a point within the event frame. The approaches are categorized into Contrast Maximization (CMax), Dispersion Minimization (DMin), and Probabilistic Alignment techniques. [14]

Deep learning based methods

Deep learning techniques have shown significant potential in Visual Simultaneous Localization and Mapping (VSLAM) algorithms [21], [16], [17], [19], [20], [18]. However, traditional Deep Neural Networks (DNNs) like Multi-Layer Perceptron networks (MLPs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) struggle with the sparse and asynchronous nature of event data from event cameras. Current DNNs often require conversion to voxel grids or event-frame-based representations to process event data. In contrast, Spiking Neural Networks (SNNs) can process individual event data directly without pre-processing. Event-based deep learning can be categorized into supervised and unsupervised learning techniques [14]

IV. LITERATURE REVIEW

[6] explains that event cameras are sensors inspired by biology. Unlike traditional cameras that take pictures at set times, these cameras detect changes in brightness for each pixel and record these changes as events. However, new methods are necessary to process the data from these cameras. This paper reviews the principles of how event cameras work, the sensors available, their applications, and the algorithms used to process their data. It also discusses the challenges and future opportunities in this field.

Event processing in event-based cameras involves three key aspects: event representations, biologically inspired visual processing, and methods for event handling. Events are represented as frames, time surfaces, or voxel grids to capture the scene's spatial and temporal dynamics, making the data easier to analyze. Biologically inspired visual processing uses models like spiking neural networks (SNNs) to mimic how human neurons process visual information, aiming for fast and efficient processing. Event handling methods include real-time filters for immediate updates and batch methods for deeper analysis using machine learning. These strategies help event-based cameras manage fast movements, work in various lighting conditions and reduce data redundancy, outperforming traditional cameras [6].

Event-based system capability is particularly advantageous for SLAM in dynamic and challenging environments. The asynchronous nature and high dynamic range of event cameras enable robust performance in scenarios with rapid motion and varying lighting conditions [6].

Continued research and development in this field are essential to fully unlock the capabilities of event-based vision and address existing challenges in SLAM applications. We have reviewed the recent papers.

A. Event-Based Line SLAM

[2] addresses a significant gap in the field of event-based vision systems, particularly focusing on the challenge of efficiently and accurately performing (SLAM) using event-based cameras. This paper bridges this gap by introducing a real-time system that leverages the advantage of event cameras to handle dynamic environments and challenging lighting conditions.

In the experiments, the authors implemented the event-based line SLAM algorithm using the DAVIS 240 C and 346 camera models to test its performance in various human-made environments with different depths, camera speeds, and lighting conditions. They conducted tests in sequences such as Officelarge, Office Lshape, Officefar, and the synthetic Trihedron dataset, as well as the Dynamic6DOF sequence for dynamic conditions. The algorithm was initialized using bootstrapping methods like Events + Bundle Adjustment (E+BA) or marker-based approaches to establish the initial camera pose and map. The tracking module processed events in small temporal windows, estimating the camera pose at a high rate by minimizing the event-line reprojection error with a Lie-formulated error-state Kalman filter. Simultaneously, the mapping module employed the space sweep algorithm to create a semi-dense 3D grid and extract 3D lines, which were optimized through bundle adjustment to refine both camera poses and line parameters.

[2]The tables below provide a comparison among state-of-the-art approaches such as EVO: A geometric approach to event-based 6-DOF parallel tracking and mapping in real time [12], ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM [1] and SVO: Fast semi-direct monocular visual odometry [5].

TABLE I
"RMSE UNDER LIGHTING CONDITIONS AND RAPID MOTION"

Dataset		Theirs		EVO	
		HDR	Fast	HDR	Fast
Office_large challenging	[m]	0.08	0.12	0.19	0.25
	[deg]	7.17	9.74	8.19	13.52
Trihedron challenging	[m]	0.03	0.05	0.05	0.13
	[deg]	2.55	3.16	3.36	4.18

TABLE II
RMSE FOR REGULAR MOTION AND REAL-TIME FACTORS

Dataset		Ours	EVO	ORB3	SVO
Office_large	[m]	0.07	0.08	0.11	0.37
	[deg]	6.49	7.23	9.47	8.75
Trihedron	[m]	0.03	0.05	0.05	0.12
	[deg]	1.74	2.40	4.46	4.23
Office L_shape	[m]	0.03	0.04	0.03	0.05
	[deg]	1.97	2.48	0.84	2.77
Office_far	[m]	0.04	0.05	0.04	0.08
	[deg]	2.21	2.83	3.36	2.79
Dynamic_6dof	[m]	0.05	0.06	0.06	0.10
	[deg]	2.6	2.9	2.79	3.51
RTF	Tracking	3.57	2.56	-	-
	Mapping	0.31	0.37	-	-
	Tracking and Mapping	-	-	1.31	1.45
Rate [1e6 ev/s]		0.9 - 1.95	0.95 - 1.7	-	-

The table I, shows that the proposed method achieves significantly lower RMSE values in both meters and degrees across the Office large and Trihedron datasets, indicating superior performance in challenging environments. The table II, highlights the proposed method's consistently lower RMSE values and higher real-time factors across various sequences, demonstrating its robustness and efficiency.

The approach achieves a camera pose estimation rate of approximately 3.3 kHz and operates effectively in dynamic scenarios with velocities exceeding 2.5 m/s and challenging lighting conditions [2].

B. Event-Based Visual Odometry

The Event-Based Visual Odometry method solves the SLAM problem with the 6-DOF motion of an event camera in natural scenes in real-time on the CPU. The method is geometric and does not require estimation of the image density, which ends up getting rid of the propagation of errors due to the estimations. The paper proposes a new approach for event-based pose tracking based on image-to-model alignment which is also used in frame-based using edge maps. On the other hand, the paper uses the Event-based Multi-View Stereo (EMVS) method for mapping. [12]

1) Pose Tracking: In contrast to the frame-based approach, geometric alignment error is used instead of photometric error for the image alignment strategy. For the registration process; the event image obtained by combining a limited number of events into an edge map and projected semi-dense 3D map of the scene according to a known pose of the event camera are used. Images are registered using an inverse compositional method called Lucas-Kanade(LK) by iteratively computing the incremental pose ΔT that minimizes the equation

$$\sum_u (M(W(u;\Delta T)) - I(W(u;T)))^2 \tag{1}$$

where M is a semi-dense projected map and I is the event image, W is warp. The equation measures the registration error between the measured image using events and the predicted image from the projection of the 3D edge map.

Updating warp W leads to the update of T rigid-body transformation from the frame of M to the frame of I.

$$T \leftarrow T \cdot (\Delta T)^{-1} \tag{2}$$

In the inverse method (1), the projected map M is adjusted until it aligns with the warped event image produced by the current estimate of the registration transformation T. The 3D rigid-body warp is derived in the following way

$$W(u;T) := \pi\left(T \cdot \pi^{-1}(u, d_u)\right) \tag{3}$$

where u represents a point in the image plane of M, T is a rigid-body transformation, π and π^{-1} represent the camera projection and inverse projection respectively, and d_u is the known depth of the 3D point that projects onto pixel u

The LK method is efficient computationally. M remains constant during the iteration leading derivatives to be pre-computed. Also, these computations can be re-used for aligning several event images I concerning the same M.

2) Mapping: The paper utilizes (EMVS) method for mapping. It discretizes the space using a voxel grid called Disparity Space Image-DSI. The point cloud of locations is obtained with the local maxima of DSI in two steps. Firstly, the DSI is transformed into both a depth map and an accompanying 2D confidence map. Secondly, the confidence map undergoes adaptive thresholding to retain the most confident local maxima from the DSI, resulting in a semi-dense depth map, which is subsequently converted into a point cloud.

C. Event-Based Stereo Visual Odometry

Event-Based stereo visual odometry method handles the perception problem using a stereo stream of events instead of using additional sensors that limit the speed and dynamic range of the system. [22] Their method also produces ego-motion of the stereo rig and 3D scene map. The paper proposes;

1) A new mapping technique that optimizes an objective function aimed at evaluating spatio-temporal consistency across stereo event streams.
2) A fusion approach leveraging the probabilistic properties of the estimated inverse depth to enhance the density and accuracy of the reconstructed 3D structure
3) An innovative camera tracking technique based on 3D-2D registration that utilizes the inherent distance field properties of a compact and efficient event representation
4) A comprehensive experimental evaluation on both publicly available and self-collected datasets shows that the system is computationally efficient, running in real-time on a standard CPU. The software, stereo rig design, and datasets used have been open-sourced.

D. Event-Based Stereo Visual Inertial Odometry

The Event-Based Stereo Visual Inertial Odometry method utilizes the inertial measurement that exceeds the performance of the other image-based and event-based methods. [3] The paper contributes with the followings:

1) Proposes purely Event-Based Stereo visual odometry pipeline with sliding windows graph-based optimization that manages the robust state estimation under harsh scenarios such as aggressive motion and low light situations. Also empower the method by integrating stereo event streams, stereo image frames and IMU.
2) Designs geometry-based spatial and temporal data associations within consecutive stereo event streams to handle the event-based stereo feature tracking and mapping problem.
3) Attain state-of-the-art performance on publicly available datasets. And they run their ESVIO estimator on a quadrotor working onboard.

The paper mentions that previous works have the problem of losing the visual tracks in textureless areas.

E. Multi DVS Stereo Depth Estimation

Depth estimation is done in [7] using two synchronized and calibrated event cameras that are rigidly attached. The input consists of a stereo event stream and the poses of the cameras, aiming to produce a semi-dense depth map or a 3D point cloud of the scene. The method uses space sweeping to create Disparity Space Images (DSIs) from each camera's data, which are then fused into a single DSI. This fused DSI is used to extract the 3D structure of the scene. The process involves fusion across cameras for better depth accuracy, with an optional shuffling block to further refine the results. It contributes to a new approach to event-based stereo semi-dense depth map construction focusing on correspondence-free methods for efficiency and extensibility.

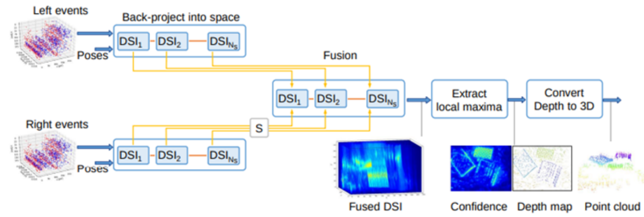

Fig. 2. Map extraction from fused cameras

The proposed algorithm for stereo event-based multi-view stereo (EMVS) in [7] integrates data from two event cameras to generate accurate depth maps. The process begins with establishing a common Disparity Space Image (DSI) location, ensuring both cameras reference a common point along the left camera's trajectory to prevent resampling errors. Each camera's events are then back-projected into their respective DSIs, preserving the geometric alignment of event origins. Next, the DSIs from both cameras are fused into a unified DSI. Various fusion metrics such as arithmetic mean, geometric mean, harmonic mean, and quadratic mean are considered, with the harmonic mean preferred for its emphasis on intersecting parts of viewing rays, which creates the 3D structure. Subsequently, depth maps are extracted by identifying maxima along each viewing ray through the reference view, leveraging a confidence or "contrast" map derived from the fused DSI. An Adaptive Gaussian Thresholding (AGT) step follows to select pixels with the highest local values, yielding a semi-dense depth map. A median filter is applied to eliminate isolated points, ensuring smoother depth estimates. For sequences spanning multiple time intervals, the algorithm in [7] employs temporal fusion by dividing the interval into sub-intervals. DSIs are constructed for each sub-interval and camera, enabling the fusion of all DSIs into a single comprehensive representation (shown in Figure 2).

[7] evaluates the performance of its event-based stereo depth estimation methods against several state-of-the-art techniques, including GTS, SGM, and ESVO, as well as the monocular EMVS method. The findings, demonstrated through both qualitative and quantitative analyses, show that the proposed methods consistently outperform existing approaches across multiple metrics on the MVSEC indoor sequences. The methods excel in capturing finer details and maintaining accuracy, with notable improvements in error rates and outlier reduction. The robustness of the methods is also confirmed on various datasets (UZH, DSEC, TUM-VIE, and EVIMO2), highlighting advantages such as higher accuracy, outlier rejection, and faster convergence due to additional parallax from stereo setups. Furthermore, [7] addresses the impact of spatial resolution, contrast thresholds, and computational efficiency, emphasizing the methods' adaptability and effectiveness even with shuffled event sequences. Overall, the results demonstrate the superiority of the proposed stereo methods in producing precise and detailed depth maps, significantly enhancing performance compared to current state-of-

the-art techniques.

F. SNN SLAM Algorithm for Fused Event-Based Camera and Radar

A pioneering bio-inspired SLAM system that fuses data from an event camera and radar using Spiking Neural Networks (SNNs), which continuously learn and adapt through Spike-Timing-Dependent Plasticity (STDP), mimicking brain functions is introduced in [13]. Unlike most fusion-based SLAMs and learning-based SLAM systems that require pre-training a Deep Neural Network (DNN) on a pre-captured dataset of the environment, this DVS-Radar fusion SNN learns in real-time as the drone navigates. The SNN outputs are integrated into a bio-inspired RatSLAM back-end for loop closure detection and map correction(shown in Figure 3). It contributes to enhance navigation robustness by leveraging both visual and radar data in a cohesive manner.

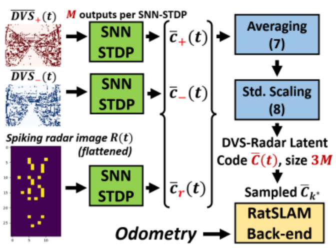

Fig. 3. SNN-STDP architecture

The proposed algorithm in [13] integrates a DVS-Radar fusion SLAM system using Spiking Neural Networks (SNNs) with Spike-Timing-Dependent Plasticity (STDP) for real-time learning and inference. Data from two primary sources, the Dynamic Vision Sensor (DVS) and Radar, undergo specific preprocessing steps tailored to their unique characteristics. The DVS captures events that are flattened into vectors, with each event's polarity (+ or -) channeled into corresponding SNN-STDP networks. Meanwhile, Radar provides detections in the form of co-ordinates and velocity for each detection, transformed into a binary spiking image. These radar detections are quantized and integrated into the SNN-STDP framework as a third module. Within the SNN-STDP ensembles, each spiking radar image (R) is temporally oversampled to match the DVS time resolution. Outputs of positive polarity DVS events, negative DVS events, and radar detections from the three ensembles are concatenated and averaged over a 0.1s time window to produce a fused latent code. This code is subsequently scaled to accommodate covariate shifts during drone navigation. Integration with the RatSLAM back-end involves feeding the fused latent code alongside raw radar-gyroscope odometry data. The latter includes radar-derived drone heading velocity and gyroscope-measured yaw rotation velocity. Odometry updates in the algorithm are computed iteratively by first updating the yaw rotation based on the yaw rotation velocity and the gyroscope sampling period, which is one millisecond. Next, the drone's X-coordinate is updated by adding the product of the heading velocity, the cosine of the updated yaw rotation, and the radar frame period, which is forty milliseconds. Similarly, the drone's Y-coordinate is updated by adding the product of the heading velocity, the sine of the updated yaw rotation, and the radar frame period. Additionally, radar-based obstacle aggregation refines drone poses and obstacle coordinates based on drone's yaw angle and the drone's coordinates at that time step. Radar detections are adjusted for the drone's yaw angle and position to reconstruct obstacles and walls. This process aids in loop closure detection and map correction within the RatSLAM framework, refining past poses and obstacle coordinates based on updated information [13].

Testing was done across challenging scenarios in [13]. Under normal lighting conditions, the proposed method achieves significantly lower mean absolute error in both localization (MAEL) and mapping (MAEM) compared to state-of-the-art RGB-based solutions, as well as ORB feature-based methods. By fusing DVS and radar data, the system outperforms individual modalities (DVS-only and Radar-only), leveraging the complementary strengths of each sensor type for robust navigation and obstacle modeling in complex environments with high visual ambiguities. Importantly, the DVS-Radar setup maintains its superior performance even under challenging low-light conditions, demonstrating robustness where RGB-based systems typically falter. This approach stands out due to its real-time, online learning capability via unsupervised STDP adaptation, contrasting with traditional DNN-based SLAM systems that require extensive offline training. These findings highlight the potential of integrating low-power SNN processors for deploying environmentally robust SLAM solutions.

Conclusion

The reviewed literature highlights significant advancements and diverse applications of event-based vision systems in the field of SLAM (Simultaneous Localization and Mapping). Event cameras, inspired by biological systems, offer distinct advantages over traditional frame-based cameras, including high dynamic range and asynchronous operation, which are particularly beneficial in dynamic and challenging environments. The studies discussed various methodologies and algorithms tailored for event cameras, such as event processing techniques, stereo visual odometry, and integration with other sensors like radar using innovative approaches such as Spiking Neural Networks (SNNs) with Spike-Timing-Dependent Plasticity (STDP). These advancements enable robust real-time navigation, accurate mapping, and obstacle detection even under low-light conditions, showcasing the potential for enhancing autonomous systems' performance and reliability in complex scenarios. Future research directions focus on refining processing algorithms, addressing challenges related to noise management, algorithm initialization and exploring further applications across domains to fully exploit the capabilities of event-based vision systems.

References

[1] Carlos Campos, Richard Elvira, Juan J Gómez Rodríguez, José MM Montiel, and Juan D Tardós. Orb-slam3: An accurate open-source library

for visual, visual–inertial, and multimap slam. *IEEE Transactions on Robotics*, 37(6):1874–1890, 2021.

[2] William Chamorro, Joan Sola, and Juan Andrade-Cetto. Event-based line slam in real-time. *IEEE Robotics and Automation Letters*, 7(3):8146–8153, 2022.

[3] Peiyu Chen, Weipeng Guan, and Peng Lu. Esvio: Event-based stereo visual inertial odometry. *IEEE Robotics and Automation Letters*, 2023.

[4] P. Chiberre et al. Detecting stable keypoints from events through image gradient prediction. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2021.

[5] Christian Forster, Matia Pizzoli, and Davide Scaramuzza. Svo: Fast semi-direct monocular visual odometry. pages 15–22, 2014.

[6] Guillermo Gallego, Tobi Delbruck, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jorg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-based vision: A survey. *IEEE*, 2022.

[7] Suman Ghosh and Guillermo Gallego. Multi-event-camera depth estimation and outlier rejection by refocused events fusion. *Advanced Intelligent Systems*, 4(12), 2022.

[8] R.G.v. Gioi et al. Lsd: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4):722–732, 2010.

[9] W. Guan et al. Pl-evio: Robust monocular event-based visual inertial odometry with point and line features. *IEEE Transactions on Automation Science and Engineering*, pages 1–17, 2023.

[10] J. Hidalgo-Carrió, G. Gallego, and D. Scaramuzza. Event-aided direct sparse odometry. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2022.

[11] H. Rebecq et al. Evo: A geometric approach to event-based 6-dof parallel tracking and mapping in real time. *IEEE Robotics and Automation Letters*, 2(2):593–600, 2017.

[12] Henri Rebecq, Timo Horstschäfer, Guillermo Gallego, and Davide Scaramuzza. Evo: A geometric approach to event-based 6-dof parallel tracking and mapping in real time. *IEEE Robotics and Automation Letters*, 2(2):593–600, 2016.

[13] Ali Safa, Tim Verbelen, Ilja Ocket, André Bourdoux, Hichem Sahli, Francky Catthoor, and Georges Gielen. Fusing event-based camera and radar for slam using spiking neural networks with continual stdp learning. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2782–2788, 2023.

[14] Sangay Tenzin, Alexander Rassau, and Douglas Chai. Application of event cameras and neuromorphic computing to vslam: A survey. 2024.

[15] V. Vasco, A. Glover, and C. Bartolozzi. Fast event-based harris corner detection exploiting the advantages of event-driven cameras. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016.

[16] S. Wang et al. Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017.

[17] N. Yang et al. D3vo: Deep depth, deep pose and deep uncertainty for monocular visual odometry. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2020.

[18] S. Zhang, J. Zhang, and D. Tao. Towards scale-aware, robust, and generalizable unsupervised monocular depth estimation by integrating imu motion dynamics. In *Computer Vision – ECCV 2022*, Cham, 2022. Springer Nature Switzerland.

[19] S. Zhang, J. Zhang, and D. Tao. Towards scale consistent monocular visual odometry by learning from the virtual world. In *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022.

[20] H. Zhao et al. Jperceiver: Joint perception network for depth, pose and layout estimation in driving scenes. In *Computer Vision – ECCV 2022*, Cham, 2022. Springer Nature Switzerland.

[21] T. Zhou et al. Unsupervised learning of depth and ego-motion from video. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.

[22] Yi Zhou, Guillermo Gallego, and Shaojie Shen. Event-based stereo visual odometry. *IEEE Transactions on Robotics*, 37(5):1433–1450, 2021.