GRE-SLAM: 6-DoF Pure Event-Based SLAM with Semi-Dense Depth Recovery Assisted Bundle Adjustment

Yang Chen
School of Computer Science and Technology
Tongji University
Shanghai, China
2011439@tongji.edu.cn

Lin Zhang*
School of Computer Science and Technology
Tongji University
Shanghai, China
cslinzhang@tongji.edu.cn

Abstract

Event cameras are innovative bioinspired vision sensors that output pixel-level brightness changes instead of standard intensity frames. Such cameras do not suffer from motion blur and cope well with scenes characterized by high dynamic range, which can benefit classic computer vision tasks such as pose estimation. However, currently developed event-based pose estimation methods either require extra data as inputs (such as IMU data or depths) or lack a global refinement step to alleviate accumulated drifts. To this end, we propose the first 6-DoF pure event-based SLAM system equipped with back-end global optimization, named GRE-SLAM (Globally Refined Event-based SLAM). For robustness and accuracy, first, 6-DoF motion compensation is introduced in the front-end to prepare sharp-edged event frames and a favorable initialization pose, mitigating unstable optimization during event registration brought by sparsity and noise of events. Second, a novel adaptive semi-dense depth recovery algorithm enriches front-end's sparse depths without additional sensors, helping establish long-term edge alignment constraints to support global BA in the back-end. Comprehensive experiments on real-world datasets demonstrate that our method can produce high-accuracy pose estimation results as well as recover a semi-dense depth map for each Image of Warped Events (IWE).

CCS Concepts

• **Computing methodologies** → **Tracking**; *Vision for robotics.*

Keywords

Event camera, Bundle adjustment, Visual odometry, Motion compensation

ACM Reference Format:
Yang Chen and Lin Zhang. 2025. GRE-SLAM: 6-DoF Pure Event-Based SLAM with Semi-Dense Depth Recovery Assisted Bundle Adjustment. In *Proceedings of the 2025 International Conference on Multimedia Retrieval (ICMR '25), June 30-July 3, 2025, Chicago, IL, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3731715.3733352

*Corresponding author

ICMR '25, Chicago, IL, USA

ACM ISBN 979-8-4007-1877-9/2025/06
https://doi.org/10.1145/3731715.3733352

1 Introduction

Unlike conventional cameras that capture full frames at a fixed frame rate, event cameras only report the pixels that have undergone a significant change in brightness [25, 35], thereby reducing redundant data and enabling more efficient processing. The output events of this new camera are visualized in Fig. 1 for an intuitive understanding. Such asynchronous and sparse outputs enable event cameras to exhibit several advantages over traditional cameras, such as high temporal resolution, low power consumption, and high dynamic range. These advantages make event cameras particularly well-suited for applications that require low latency and high temporal precision, such as robot navigation [1], augmented reality [34, 36], and high-speed tracking [32, 40].

In the past decade, an increasing number of researchers have explored the use of event cameras for a key problem in robotics, Simultaneous Localization And Mapping (SLAM) [4, 22, 29]. However, unlocking the advantages of event cameras for SLAM is very challenging due to the fact that the outputs of event cameras are fundamentally different from those of standard cameras. Therefore, traditional vision algorithms cannot be directly applied to event data and innovative SLAM techniques must be investigated.

An ideal event-based SLAM system would not process redundant data to ensure computation efficiency, allowing on-board processing in real-time. Although a few solutions to event-only-based Visual Odometry (VO) or SLAM have been proposed, they suffer from the following two key limitations. First, the majority of them focus on rotation-only motion estimation [13, 20], limiting their usage in common environments where the camera moves with 6 Degrees of Freedom (DoF). Second, all these event-based methods estimate the camera pose for the current set of events within a short-term window, which can only serve as the front-end of a SLAM system. Without a global refinement step as the back-end, the accumulated drifts will obviously damage the system's robustness and trajectory accuracy.

As an attempt to fill in the above research gaps to some extent, we proposed GRE-SLAM (Globally Refined Event-based SLAM), a 6-DoF event-only-based SLAM system for parallel tracking and mapping enhanced by a back-end global optimization. In summary, our contributions are as follows:

- The first **pure event-based 6-DoF** SLAM system with a full **front-end and back-end** structure, GRE-SLAM, is proposed.
- The first **event-only-based 6-DoF global optimization** module is designed, acting as the back-end of GRE-SLAM. In this module, an adaptive semi-dense depth recovery algorithm is designed to enrich the sparse depths corresponding

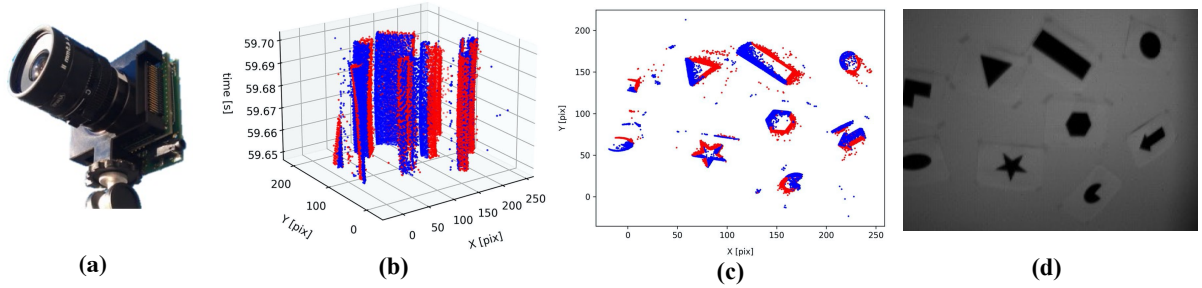


Figure 1: Illustration of the event camera and its output compared to the standard camera image. (a) Event camera (DAVIS240C); (b) Raw event data in the spatio-temporal space; (c) Accumulated events; (d) Intensity image.

to a limited number of short-term events from the front-end through multiple camera observations at different historical times. The output semi-dense depths help establish long-term geometric pose constraints based on more event points with valid 3D positions to support effective global BA.

- The accuracy and efficiency of GRE-SLAM have been corroborated by extensive experiments conducted on both indoor and outdoor environments, with pose accuracy even superior to some multi-sensor SLAM systems (i.e. VINS-Mono and Ultimate-SLAM) in several scenes.

2 Related Work

In order to compare the different characteristics of representative relevant methods more clearly, in Table 1, we summarize them from the following aspects: 1) the DoF of the motion (DoF); 2) whether depth estimation is performed (Depth); 3) the sensor configuration (Sensor); 4) whether equipped with the back-end (BE). It can be seen from Table 1 that our GRE-SLAM is the first pure event-based 6-DoF SLAM system with a complete front-end and back-end structure without additional sensors or priors.

2.1 Hybrid Event-based SLAMs

With the rapid development of multi-modal SLAM systems, researchers have been attempting to introduce event cameras to these systems to cooperate with other sensors such as regular cameras or IMUs. According to the sensor configuration, we can mainly divide current multi-modal event-based SLAM methods into two categories, event-visual SLAM and event-visual-inertial SLAM.

2.1.1 Event-visual SLAM. Censi *et al.* [5] presented the first VO based on an event camera plus a normal CMOS camera to provide the absolute brightness values. Weikersdorfer *et al.* [38] demonstrated the advantage of fusing an event camera with a classic frame-based RGB-D sensor in 3D map reconstruction. In [24], another low-latency visual odometry algorithm was presented for the DAVIS sensor that employs event-based feature tracking. Features are detected in grayscale frames and tracked asynchronously via the event stream. These features are used for both pose optimization and probabilistic mapping, effectively tracking the sensor's 6-DOF motion in natural environments. Gallego *et al.* [11] provided the solution to the problem of accurate, low-latency tracking of an event camera yet with a prior of a photometric depth map, comprising intensity and depth information built via classic dense reconstruction pipelines. EDS [17] is the first method to achieve 6-DOF visual odometry using a direct approach that integrates both events and frames, demonstrating increased robustness and accuracy. Pellerito *et al.* [28] introduced RAMP-VO, the first end-to-end learned image- and event-based VO system, utilizing novel RAMP encoders to fuse asynchronous events with image data, achieving faster inference and more accurate predictions than existing methods.

2.1.2 Event-visual-inertial SLAM. Zhu *et al.* [43] introduced the first algorithm that integrates an event-based tracking system with an IMU to achieve precise 6-DoF camera pose tracking. An Extended Kalman Filter is designed to merge feature tracks and IMU data for an initial pose to eliminate failed tracks, effectively tracking camera motion in challenging scenarios. USLAM [37] emerged as the first comprehensive scheme that fuses events, standard frames, and inertial measurements in a tightly coupled framework. This hybrid pipeline not only improves the tracking accuracy but also unlocks flight scenarios, which were not reachable with traditional visual-inertial odometry, with the assistance of event sensors. Guan *et al.* [15] also introduce a monocular visual-inertial odometry (VIO) for event cameras, involving two distinct event representations based on time surfaces to facilitate event-corner feature tracking (for front-end incremental estimation) and matching (for loop closure detection). PL-EVIO [14] is another robust and real-time event-based VIO that extracts both point-based and line-based event features for pose estimation. ESVIO [6] is an event-based stereo visual-inertial odometry system that facilitates temporal tracking and real-time matching between consecutive stereo event streams, leading to reliable state estimation.

Although the above-mentioned hybrid event-included systems show satisfactory tracking or mapping performance, they require additional sensors/priors and cost more time to process extra data other than events.

2.2 Pure Event-based Solutions

The special nature of event data asks for novel approaches, and full 6-DoF motion estimation with a single event camera remains a challenging problem. Thus quite a few related methods rely on simplifying assumptions. For example, some of them assume that the camera only performs rotational motion [7, 26, 33] while some are based on the hypothesis of plane motion [39, 44]. As a result, these

methods are hindered from being used in real-world applications with complex motions or non-planar scene structures.

Kim *et al.* [21] utilize three decoupled probabilistic filters to respectively estimate 6-DoF camera motion, scene logarithmic intensity gradient, and scene inverse depth relative to a keyframe. Besides, motion compensation [19, 27] is also a powerful tool for raw data processing and pose estimation in event-based VO. Its main motivation comes from estimating motion from events directly in 3D space (e.g. events augmented with depth), without projecting them onto an image plane. EVO [31] is the first to track 6-DoF camera motions while recovering a 3D map of the environment in real time on a standard CPU. Based on the event-based reconstruction work EMVS [30], it consists of two parallel modules, tracking and mapping, following the framework of PTAM [22]. However, its tracking module suffers from unsatisfactory robustness and sometimes fails in practical running.

ESVO [42] maximizes spatio-temporal consistency of stereo event data using a simple representation. The mapping module creates a semi-dense 3D map by probabilistically fusing depth estimates from multiple local viewpoints, while the tracking module recovers the stereo pose through a registration problem linked to the map and event data. DEVO [23] is the first monocular event-only system that excels in many real-world benchmarks, eliminating the need for additional sensors. DEVO sparsely tracks selected event patches over time, utilizing a novel deep patch selection mechanism tailored for event data. However, as a learning-based method, DEVO requires posed images along with depths for training, which takes extra effort to prepare the cumbersome training data.

Most of the above-mentioned VOs lack a global optimization step to keep the consistency of the global trajectory, leaving room for improvement of localization accuracy. Although there exist a few approaches for event-based global pose optimization through global alignment of event packets [16, 20], they are still limited to 3-DoF motion and can only produce a panorama map instead of a scaled 3D geometric map.

Table 1: Comparison on features of typical event-based motion estimation methods

Methods	DoF	Depth	Sensor	BE
Censi [5]	3	×	event+grayscale	×
Weikersdorfer [38]	6	input	event+RGB-D	×
Kueng[24]	6	✓	event+grayscale	×
Gallego [11]	6	input	event+RGB-D	×
EDS [17]	6	✓	event+RGB	✓
RAMP-VO [28]	6	×	event+grayscale	×
Zhu [43]	6	×	event+IMU	×
USLAM [37]	6	✓	event+RGB+IMU	✓
Guan [15]	6	✓	event+grayscale+IMU	✓
PL-EVIO [14]	6	✓	event+grayscale+IMU	✓
ESVIO [6]	6	✓	stereo event+grayscale+IMU	✓
GAE [20]	3	×	event	✓
Cook[7]	3	×	event	×
Liu [26]	3	×	event	×
Weikersdorfer [39]	3	×	event	×
Kim [21]	6	✓	event	×
IncEMin [27]	6	×	event	×
PEME [19]	6	×	event	×
EVO [31]	6	✓	event	×
CMax-SLAM[16]	3	×	event	✓
ESVO [42]	6	✓	stereo event	×
DEVO [23]	6	×	event	×
GRE-SLAM	6	✓	event	✓

3 Methodology

3.1 Framework Overview

Our GRE-SLAM is composed of two parallel threads, the front-end and the back-end, as shown in Fig. 2. On the one hand, the front-end receives input events and recovers the corresponding sparse depth map for each event frame to produce a local pose using event alignment. For more robust alignment, we utilize the effective processing tool for event data, the motion compensation algorithm, to provide an initialized pose and a sharper edge map. On the other hand, in terms of the back-end, we design a two-stage event-based global optimization scheme for better global trajectory consistency. In detail, to prepare for the optimization, an adaptive depth fusion module maintains a set of historical depth maps of Image of Warped Events (IWE) to generate a semi-dense depth map for each IWE, which provides more event points with valid 3D positions to establish long-term pose constraints. After that, the first-stage BA-based optimization is conducted to refine the local structure of the sliding window. Finally, the second-stage global pose graph optimization is employed to further eliminate the accumulated pose error. In the following, we will introduce the front-end VO (Sec. 3.2) and the back-end optimization, including the depth fusion module (Sec. 3.3.1) and event-based global optimization (Sec. 3.3.2) respectively.

3.2 Motion Compensated Direct VO (Front-end)

Since events naturally correspond to scene edges, a typical pure event-based odometry, EVO [31], obtains poses by optimizing edge alignment errors under the direct VO framework [10], [9]. Unfortunately, due to the sparsity and noise of event data, EVO [31] tends to fail when dealing with long-term sequences. Instead, to improve the pose optimization's robustness for long-term motions, we utilize the effective motion compensation algorithm to better prepare the event data for later VO tracking based on the direct paradigm.

Motion compensation. The motion compensation algorithm [12, 19, 27] is expected to remove blur when accumulating individual events in a spatiotemporal neighborhood. Mathematically, each input event, $e_k = (\boldsymbol{x}_k, t_k, p_k)$, is composed of the pixel position $\boldsymbol{x}_k = (x_k, y_k)^T$, the timestamp t_k, and the polarity p_k representing the brightness changes. Let $\mathcal{E} = \{e_k\}_{k=1}^{N_e}$ denote a set of neighboring events within a time interval $\mathcal{T} = \{t_k\}_{k=1}^{N_e}$, where N_e is the event group size. We define the optimizable parameters, angular and linear velocities of the event group, as $\boldsymbol{\omega} \in \mathbb{R}^3$ and $\boldsymbol{\theta} \in \mathbb{R}^3$, respectively. The compensated camera motion from arbitrary time t_k to the reference time t_1 can be formulated by :

$$\Delta \boldsymbol{T}_k^C = \begin{bmatrix} exp_{so(3)}(\hat{\boldsymbol{\omega}}\delta t_k) & \boldsymbol{\theta}\delta t_k \\ \boldsymbol{0}^T & 1 \end{bmatrix}, \tag{1}$$

where δt_k is the time difference, i.e., $\delta t_k = t_k - t_1$, $\hat{\boldsymbol{\omega}} \in \mathbb{R}^{3\times3}$ is the cross product matrix of $\boldsymbol{\omega}$. $exp_{so(3)}(\cdot)$ refers to exponential mapping from $so(3)$ to $SO(3)$.

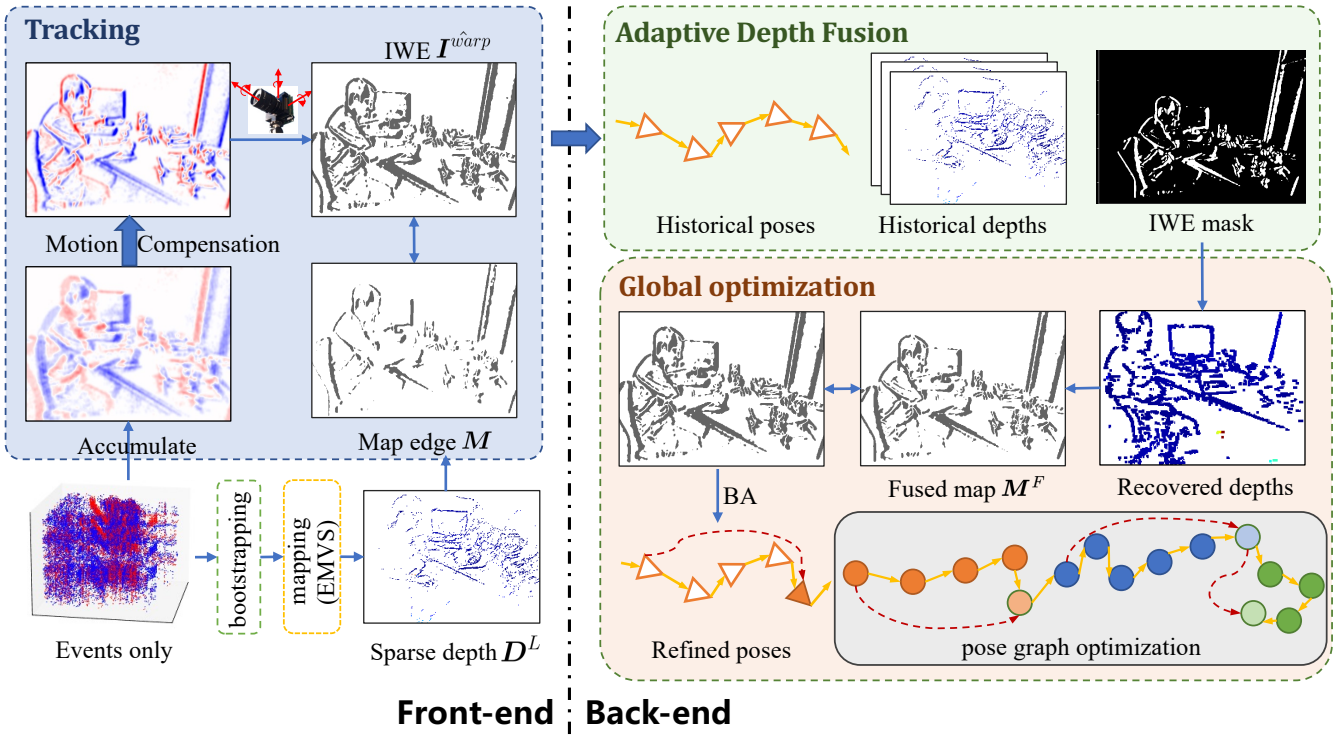


Figure 2: Overview of GRE-SLAM. In GRE-SLAM, only event streams are required as inputs. The front-end estimates short-term 6-DoF camera poses using the motion compensation algorithm to provide the initialization and IWE for edge alignment. The back-end adaptively fuses historical depths based on historical VO poses to establish long-term constraints for global trajectory refinement using pose graph optimization.

Then each event e_k in the group can be warped by the corresponding ΔT_k^C to form the IWE denoted by I^{warp}:

$$x'_k(d_x) = K^{-1} d_x [x_k, y_k, 1]^T, \tag{2}$$

$$W(x_k; \Delta T_k^C, d_x) = \pi(\Delta T_k^C \begin{bmatrix} x'_k(d_x) \\ 1 \end{bmatrix}), \tag{3}$$

$$I^{warp}(x; \Delta T^C, d_x) = \sum_{k=1}^{N_e} p_k \delta_d(x - W(x_k, \Delta T_k^C, d_x)), \tag{4}$$

where $K \in \mathbb{R}^{3\times3}$ is the camera intrinsic matrix, the depth value d_x is set to the median depth of the scene for efficiency as in [37]. $\delta_d(\cdot)$ is the Dirac delta function, $x'_k = [x'_k, y'_k, z'_k]^T$ is the inverse-projected point from x_k to the camera coordinate system and π denotes the camera projection.

Since a sharper IWE indicates better alignment of event data, we maximize the contrast of I^{warp} to optimize the velocities ω and θ:

$$\underset{\omega,\theta}{\text{maximum}} \; \|I^{warp}(\Delta T^C(\omega, \theta), d_x)\|_F^2, \tag{5}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Now, we have obtained the motion compensated frame I^{warp} and the optimized velocities ω and θ. After this optimization, we can obtain the relative camera pose ΔT_k^C at any time $t_k \in [t_1, t_{N_e}]$ for this event group through Eq. 1.

Direct methods. Extending the traditional direct VO pipelines [9, 10] to event-based VO, the geometric alignment error [31] instead of the photometric error is minimized in GRE-SLAM's front-end. Such geometric error is constructed between two edge maps: the accumulated event frame and the edge image of the 3D map projected by the relative transformation ΔT^D denoted by M. Since directly accumulating events will bring motion blur, we utilize the binary map of the compensated IWE I^{warp}, denoted by $I^{\hat{warp}}$, as the event frame. Then we can iteratively compute the incremental pose δT^D following the inverse compositional Lucas-Kanade (LK) method [2] that minimizes:

$$\sum_{x_k} (M(W(x_k; \delta T^D, d'_x)) - I^{\hat{warp}}(W(x_k; \Delta T^D, d'_x))^2, \tag{6}$$

where d'_x is depth estimated by EMVS [30] instead of the median scene depth for better accuracy, and the target ΔT^D will be updated after each iteration:

$$\Delta T^D \leftarrow \Delta T^D \cdot (\delta T)^{-1}. \tag{7}$$

It is worth noticing that ΔT^D is initialized as ΔT^C provided by Eq. 5 instead of the commonly used zero initialization. Thanks to the motion compensation step, our front-end can benefit from the IWE $I^{\hat{warp}}$ without motion blur and a good initialization ΔT^C to enable stable and effective pose optimization, making our GRE-SLAM more robust to complex motion in practice.

Table 2: Absolute Pose Errors on translation ($APE_t/m\downarrow$) and rotation ($APE_r/^\circ\downarrow$) of compared methods on RPG Stereo DAVIS [41]

Method	Input	bin		boxes		desk		monitor		Avg.	
		APE_t	APE_r	APE_t	APE_r	APE_t	APE_r	APE_t	APE_r	APE_t	APE_r
ORB-SLAM3 [4]	F	0.016	**1.11**	0.024	2.48	0.027	**1.53**	0.029	3.49	0.024	**2.15**
VINS-Mono [29]	F+I	0.012	1.20	0.063	**2.29**	0.017	1.89	0.018	3.81	0.028	2.30
EDS [17]	E+F	0.011	1.51	**0.021**	4.86	**0.016**	2.16	0.010	**1.86**	**0.015**	2.60
USLAM [37]	E+F+I	**0.009**	3.51	0.052	2.31	0.019	3.76	**0.009**	4.59	0.022	3.54
IncEmin [27]	E	0.015	4.18	0.021	6.49	0.025	17.46	0.017	4.32	0.020	8.11
PEME-1k [19]	E	0.012	6.56	0.181	7.80	0.042	16.43	0.032	7.78	0.067	9.64
PEME-30k [19]	E	0.010	5.34	0.161	7.75	0.040	13.29	0.028	7.46	0.060	8.46
EVO [31]	E	0.107	30.54	0.120	5.28	0.131	37.63	0.251	28.60	0.152	25.51
GRE-SLAM (no BA)	E	0.022	6.13	0.043	5.46	0.035	5.81	0.021	7.47	0.030	6.22
GRE-SLAM	E	**0.009**	**3.13**	**0.012**	**2.17**	**0.010**	**4.34**	**0.015**	**3.62**	**0.012**	**3.32**

3.3 Semi-Dense Depth Assisted BA (Back-end)

3.3.1 Adaptive Depth Fusion. To ensure the accuracy of the full 6-DoF motion estimation, the quality of depth estimation is of great necessity. However, due to the locality and the sparsity of event points, the depth map for each IWE from the front-end generated by EMVS [30] is also sparse, which contributes to a limited number of valid event points that can be utilized in Eq. 6. Starting from this point, in the back-end, we maintain a depth set storing all the historical depths to produce a global semi-dense depth map for each IWE adaptively.

The algorithm for fusing IWE depth maps is given in Alg. 1. We maintain a set of fused depth maps corresponding to each IWE $\boldsymbol{I}_i^{\hat{warp}}$ denoted by $\mathcal{D}$. Each $\boldsymbol{T}_k^D$ in $\mathcal{P}$ is calculated by accumulating all the past $\Delta\boldsymbol{T}^D$ obtained from Eq. 6 and Eq. 7. For each coming depth map $\boldsymbol{D}_i^L$ from the front-end, we enrich this sparse depth map by projecting past fused depths selected from $\mathcal{D}$ based on the past poses in $\mathcal{P}$ and filter its outliers with the compensated $\boldsymbol{I}_i^{\hat{warp}}$ as the mask. When it comes to the situation where multiple past depths are projected to the same pixel, we determine the final depth value by comparing their adaptive weights. The adaptive weight for each depth point is measured by the timestamp t and the similarity s between its edge and the corresponding IWE. We design this weighting strategy based on two assumptions: 1) earlier depth points are more reliable due to fewer accumulated drifts; 2) a reasonable depth map is supposed to capture the same edges with the IWE.

With this adaptive depth fusing module, the output depth maps are semi-dense, clean and globally consistent since the observations from different camera views with a long time interval can compensate for each other. This dense map assists in global pose optimization by providing long-term geometric constraints from more event points.

3.3.2 Two-Stage Global Optimization. Based on the fused depths $\mathcal{D}$ integrating global geometric information, we are capable of conducting long-term global optimization with a sliding-window strategy. For each window covering several IWEs denoted by $[\boldsymbol{I}_i^{\hat{warp}}, \boldsymbol{I}_j^{\hat{warp}}]$, the designed global optimization scheme consists of two steps. First, the relative pose within a window denoted by $\Delta\boldsymbol{T}_{ij}^W$ is estimated through BA based on the similar optimization objective in Eq. 6, yet with the fused edge map $\boldsymbol{M}_i^F$ extracted from the fused depth $\boldsymbol{D}_i^F$. Specifically, the objective to be minimized is:

$$\sum_{\boldsymbol{x}_k}(\boldsymbol{M}_i^F(\mathbf{W}(\boldsymbol{x}_k;\delta\boldsymbol{T}_{ij}^W)) - \boldsymbol{I}_j^{\hat{warp}}(\mathbf{W}(\boldsymbol{x}_k;\Delta\boldsymbol{T}_{ij}^W))^2, \tag{8}$$

where $\Delta\boldsymbol{T}_{ij}^W$ is initialized with the front-end $\Delta\boldsymbol{T}_i^D$.

The derivatives of the edge map $\boldsymbol{M}$ to the incremental pose $\delta\boldsymbol{T}$ can be obtained by the chain-rule:

$$\frac{\partial\boldsymbol{M}}{\partial\delta\boldsymbol{T}} = \frac{\partial\boldsymbol{M}}{\partial\mathbf{W}}\cdot\frac{\partial\mathbf{W}}{\partial\delta\boldsymbol{T}}, \tag{9}$$

Algorithm 1 Algorithm for Fusing IWE Depths

Input: A local depth map $\boldsymbol{D}_i^L$, an accumulated front-end pose $\boldsymbol{T}_i^D$ corresponding to the i-th IWE $\boldsymbol{I}_i^{\hat{warp}}$ at time t

Output: The updated set of fused depths $\mathcal{D} = \{\boldsymbol{D}_k^F\}_{k=1}^{N_D}$, a set of poses $\mathcal{P} = \{\boldsymbol{T}_k^D\}_{k=1}^{N_D}$, a set of weighted maps $\mathcal{M}^w = \{\boldsymbol{M}_k^w\}_{k=1}^{N_D}$.

initialized $\mathcal{D} = \varnothing, \mathcal{P} = \varnothing, \mathcal{M}^w = \varnothing, N_D = 0, t_{ref} = t_N = 0$, current weight map $\boldsymbol{M}_{curr}^w = \mathbf{0}^{H\times W}$
if $\mathcal{D} == \varnothing$ **then**
  add $\boldsymbol{D}_i^L$ to $\mathcal{D}$, add $\boldsymbol{T}_i^D$ to $\mathcal{P}$, add $\boldsymbol{M}_{curr}^w$ to $\mathcal{M}^w$
else
  for each pixel $\boldsymbol{x}$ in $\boldsymbol{D}_i^L$ **do**
    $s = SSIM(\boldsymbol{D}_i^L, \boldsymbol{I}_i^{\hat{warp}})$
    $\boldsymbol{M}_{curr}^w(\boldsymbol{x}) = exp(-\frac{t-t_{ref}}{t_N})\cdot s$
  for each $\boldsymbol{D}_k^F \in \mathcal{D}$ **do**
    for each pixel $\boldsymbol{x}$ in $\boldsymbol{D}_k^F$ **do**
      $\boldsymbol{x}'' = \mathbf{W}(\boldsymbol{x}, \boldsymbol{T}_k^D)$
      if $\boldsymbol{D}_i^L(\boldsymbol{x}'') == 0$ or $\boldsymbol{M}_k^w(\boldsymbol{x}) > \boldsymbol{M}_{curr}^w(\boldsymbol{x}'')$ **then**
        $\boldsymbol{D}_i^L(\boldsymbol{x}'') = z_{\boldsymbol{x}''}$
        $\boldsymbol{M}_{curr}^w(\boldsymbol{x}'') = \boldsymbol{M}_k^w(\boldsymbol{x})$
    $\boldsymbol{D}_{k+1}^F = \boldsymbol{I}_i^{\hat{warp}}\cdot\boldsymbol{D}_i^L$
    add $\boldsymbol{D}_{k+1}^F$ to $\mathcal{D}$, add $\boldsymbol{T}_i^D$ to $\mathcal{P}$, add $\boldsymbol{M}_{curr}^w$ to $\mathcal{M}^w$
$t_{ref} = t, t_N = t_N + t$
return $\mathcal{D}, \mathcal{P}, \mathcal{M}^w$

Table 3: Absolute Pose Errors on translation ($APE_t/m \downarrow$) and rotation ($APE_r/^\circ \downarrow$) of compared methods on DAVIS 240C [3]

Method	Input	shapes_6dof		boxes_6dof		poster_6dof		hdr_boxes		hdr_poster		dynamic_6dof		Avg.	
		APE_t	APE_r	APE_t	APE_r	APE_t	APE_r	APE_t	APE_r	APE_t	APE_r	APE_t	APE_r	APE_t	APE_r
ORB-SLAM3 [4]	F	0.253	26.79	0.321	13.46	0.305	18.55	0.364	32.58	0.311	25.76	0.102	5.63	0.276	20.46
VINS-Mono [29]	F+I	0.246	23.20	0.412	23.32	0.227	12.85	0.352	22.88	0.313	33.29	0.037	1.73	0.265	19.55
USLAM [37]	E+F+I	0.344	18.09	0.497	11.52	0.288	21.35	0.275	15.71	0.275	22.66	0.313	21.42	0.332	18.46
IncEMin [27]	E	0.293	20.64	0.526	39.89	0.306	15.99	0.326	17.82	0.308	29.59	0.307	20.44	0.344	24.06
PEME-1k [19]	E	0.232	21.34	0.362	27.50	0.295	18.38	0.331	19.45	0.225	23.64	0.328	22.77	0.296	22.18
PEME-30k [19]	E	**0.204**	20.08	0.354	25.51	0.266	15.47	0.298	19.41	**0.221**	21.62	0.314	20.68	0.276	20.46
EVO [31]	E	0.210*	19.64*	**0.328***	52.72*	0.252*	15.77*	0.348*	21.50*	0.239*	**14.09***	0.268*	66.56*	0.274*	31.71*
GRE-SLAM (no BA)	E	0.343	20.32	0.364	28.92	0.305	15.60	0.324	16.89	0.254	26.55	0.304	27.42	0.316	22.62
GRE-SLAM	E	0.230	**18.04**	**0.328**	**25.23**	**0.216**	**13.54**	**0.298**	**16.64**	0.247	24.82	**0.267**	**19.50**	**0.265**	**19.63**

"*" means losing track in the middle of the sequences.

Table 4: Absolute Pose Errors on translation ($APE_t/m \downarrow$) and rotation ($APE_r/^\circ \downarrow$) of compared methods on DAVIS Depth [11]

Method	Input	pipe1		pipe2		bicycle		Avg.	
		APE_t	APE_r	APE_t	APE_r	APE_t	APE_r	APE_t	APE_r
ORB-SLAM [4]	F	0.015	3.35	0.021	5.89	0.050	4.74	0.029	4.66
VINS-Mono [29]	F+I	0.013	4.84	0.023	5.61	0.044	4.82	0.027	5.09
USLAM [37]	E+F+I	0.011	3.62	0.038	22.74	0.045	3.83	0.031	10.06
IncEMin [27]	E	0.054	28.57	0.049	29.53	0.102	16.46	0.068	24.85
PEME-1k [19]	E	0.032	24.53	0.047	20.38	0.087	13.49	0.055	19.47
PEME-30k [19]	E	0.029	20.19	0.036	16.75	0.064	**10.37**	0.043	15.77
EVO [31]	E	0.029*	28.38*	0.031*	27.54*	0.124*	15.97*	0.061*	23.96*
GRE-SLAM (no BA)	E	0.027	14.09	0.048	15.19	0.055	12.58	0.043	13.95
GRE-SLAM	E	**0.026**	**12.83**	**0.035**	**13.77**	**0.032**	12.43	**0.031**	**13.01**

"*" means losing track in the middle of the sequences.

where the first derivative term can be simply given by the gradient $\nabla \boldsymbol{M}$, and the latter term is calculated by :

$$\frac{\partial \boldsymbol{M}}{\partial \delta \boldsymbol{T}} = \begin{bmatrix} \frac{f_x}{z'_k} & 0 & -\frac{f_x x'_k}{z'^2_k} & -\frac{f_x x'_k y'_k}{z'^2_k} & f_x(1+\frac{x'^2_k}{z'^2_k}) & -\frac{f_x y'_k}{z'_k} \\ 0 & \frac{f_y}{z'_k} & -\frac{f_y y'_k}{z'^2_k} & -f_y(1+\frac{y'^2_k}{z'^2_k}) & \frac{f_y x'_k y'_k}{z'^2_k} & \frac{f_y x'_k}{z'_k} \end{bmatrix}, \tag{10}$$

where f_x and f_y are the focal lengths of the event camera. It is worth mentioning that, compared with EVO [31], our adaptive fusion module can output denser depth maps. Thus, more valid event points can be used to construct the error terms of BA for more accurate and robust localization.

After BA, the pose graph optimization is conducted to further refine all past poses $[\boldsymbol{T}_i^D, \boldsymbol{T}_j^D]$ in the sliding window with $\Delta \boldsymbol{T}_{ij}^W$ as a new long-term constraint. The objective for graph optimization within each window is :

$$\begin{aligned} e &= Log((\Delta \boldsymbol{T}_{ij}^D)^{-1}((\boldsymbol{T}_i^D)^{-1}\boldsymbol{T}_j^D))^\vee \\ &= Log((\Delta \boldsymbol{T}_{ij}^W)^{-1}(\Delta \boldsymbol{T}_i^D \cdot \Delta \boldsymbol{T}_{i+1}^D \cdots \Delta \boldsymbol{T}_j^D))^\vee. \end{aligned} \tag{11}$$

4 Experiments

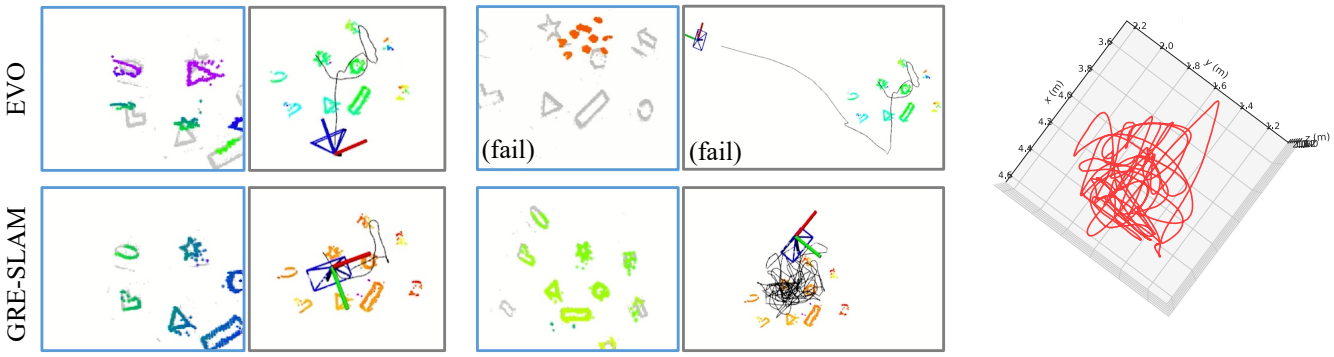
4.1 Experimental Setup

Datasets. We evaluated our approach in terms of both tracking accuracy and the quality of recovered depths on three public real-world datasets where ground truth poses are available by motion caption systems. These three datasets include RPG Stereo DAVIS [41], DAVIS 240C [3], and DAVIS Depth [11]. The former two provide indoor sequences while the latter one covers outdoor scenes. The sequence lengths vary from 10s to 60s, including scenes with high dynamic range, fast motion or dynamics. All the comparative experiments were conducted on the same desktop with a CPU model of AMD Ryzen 9 5900X 12-Core Processor.

Metrics. To measure the accuracy of the estimated poses, we selected two widely used metrics in SLAM, the Absolute Pose Error (APE) on both the translational and the rotational components, represented by APE_t/m and $APE_r/^\circ$, respectively. APE_t is the Euclidean distance between the ground truth translation and the estimated event camera location while APE_r is measured using the angular difference between the estimated rotation and the ground truth rotation.

4.2 Localization Accuracy

Table 2, Table 3, Table 4 report quantitative results of our GRE-SLAM compared with other VOs/SLAMs measured by APE_t and APE_r on three datasets mentioned in Sec. 4.1, respectively. For a more comprehensive comparison, apart from the pure event-based VOs [19, 27, 31], we also provided the results of event-based VO (EDS [17] for the short-duration RPG Stereo DAVIS), event-based VIO (USLAM [37]) and classic frame-based SLAMs (ORB-SLAM [4] and VINS-Mono [29]) for reference. The inputs from different sensors required for each method were also given, denoted by E



(a) short-term (b) long-term (c) GT trajectory

Figure 3: Comparison in both short-term and long-term tracking performance with the classic EVO [31] on the "*shapes_6dof*" scene from the DAVIS 240C [3] dataset. Blue boxes indicate 2D tracking results on the image plane while gray boxes represent the trajectories in the 3D map. It can be seen that EVO loses tracking in the middle of the sequence while our GRE-SLAM maintains accurate pose tracking in the long-term.

(events), F (image frames) and I (inertial measurements) for brevity. The best results are highlighted in bold. It can be seen from these tables that our GRE-SLAM clearly outperforms all event-only-based competitors in almost all scenes. Also, GRE-SLAM even surpasses some schemes with multiple sensors such as VINS-Mono [29] and USLAM [37] in several scenes, demonstrating GRE-SLAM's satisfactory performance even only with an event sensor. It is worth noticing that, in some challenging scenes, such as the HDR scenes in Table 3 (*hdr_boxes* and *hdr_poster*), all the event-based SLAMs outperform the non-event SLAM (ORB-SLAM and VINS-Mono), proving event cameras' advantages in dealing with this intractable environment while frame-based cameras may be disturbed by abrupt illumination changes.

The most relevant work to our GRE-SLAM is EVO [31] since it is the only competitor that also recovers both 6-DoF pose and depth from events only. However, it fails to address the full sequence in most scenes due to its instability of direct pose optimization. In contrast, our motion-compensated direct VO shows clear advantages over it both in robustness and accuracy even without BA. In terms of qualitative results, Fig. 3 represents the visualization results of both short-term and long-term tracking performance of EVO [31] and our GRE-SLAM. EVO exhibits unstable tracking ability in the short term and fails in the long-term performance, demonstrating our GRE-SLAM's accuracy and robustness. Also, Fig. 4 gives the visualization of tracking results of our GRE-SLAM compared with EVO on the "*pipe*1" scene, intuitively demonstrating GRE-SLAM's improvement in global trajectory accuracy.

4.3 Time Efficiency

Table 5 gives the time cost of GRE-SLAM compared two other event-based VOs also relying on motion compensation [19, 27]. The simple yet effective design of both the front-end and the back-end of our GRE-SLAM accounts for our high efficiency. Specifically, our front-end provides the direct pose optimization with a good initialization for fast convergence while the back-end fully utilizes

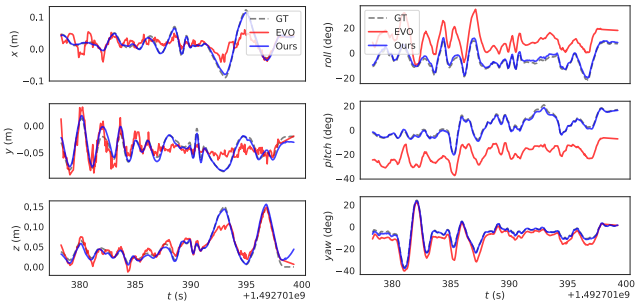


Figure 4: Comparison of estimated trajectories in the XYZ view and the RPY view by EVO and our GRE-SLAM (before EVO fails) on the outdoor "*pipe*1" scene from the DAVIS Depth dataset [11].

the outputs from the front-end without extra processing of events to save time and computational cost.

Table 5: Time cost (μs $\downarrow$) to process each event on tested datasets

Method/Dataset	DAVIS 240C [3]	DAVIS Depth [11]	RPG Stereo DAVIS [41]
EVO[31]	17.6	18.5	16.4
IncEmin [27]	31.4	33.5	33.4
PEME-10k [19]	22.7	23.1	20.5
PEME-30k [19]	26.4	28.1	28.2
GRE-SLAM	**13.7**	**12.1**	**12.7**

4.4 Performance on Semi-dense Depth Recovery

Depth reconstruction is another important task in SLAM since its quality not only influences the quality of 3D scene reconstruction but also affects the tracking accuracy in pose estimation. Fig. 5 shows the qualitative comparison of our GRE-SLAM and EVO [31] on typical sequences from three tested datasets. From Fig. 5, it can

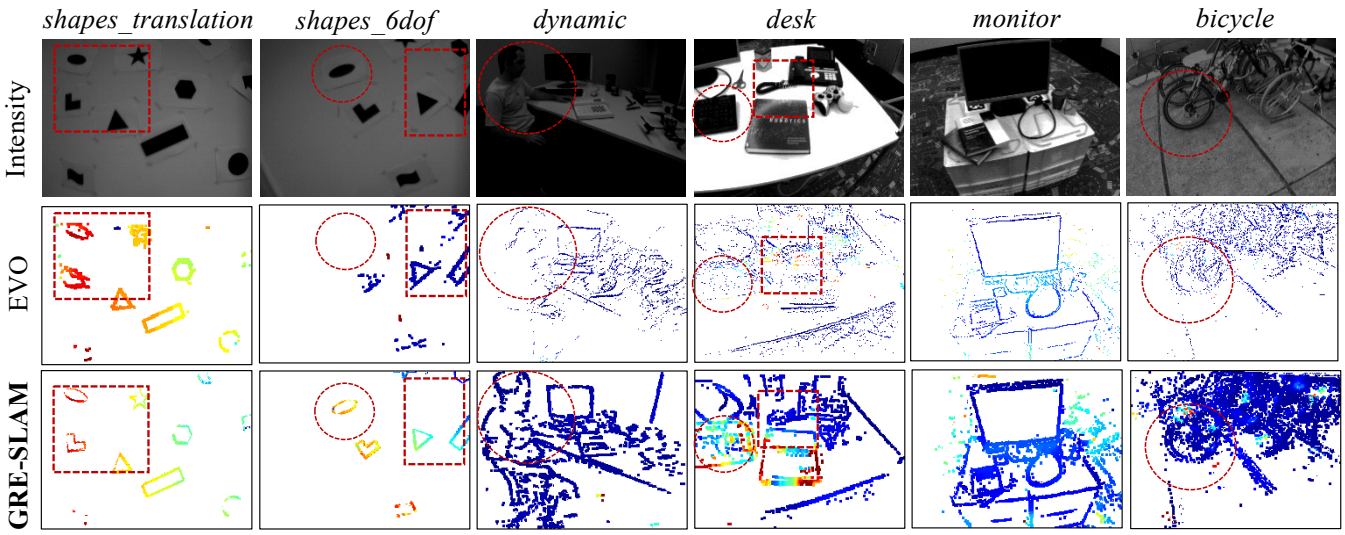


Figure 5: Qualitative comparison of depth reconstruction on three testing sequences from three datasets. Each row depicts the pseudo-colored inverse depth maps generated by corresponding methods. It is worth mentioning that since the timestamps of event frames formed by different methods are not completely aligned, we chose to show results with relatively close perspectives.

be seen the depth estimation of EVO [31] is sparse and usually loses scene details (marked with red dotted circles) or contains obvious outliers (marked with red dotted rectangles), which may cause tracking failure. In comparison, GRE-SLAM produces semi-dense depth maps with satisfactory geometric structures consistent with perception. Although there exist several event-based schemes aimed at recovering dense depths [8, 18], introducing them requires considerable computational time or extra sensors. Instead, GRE-SLAM relies on the adaptive depth fusion strategy to recover more details utilizing multiple observations from different camera views in the past time, while maintaining high time efficiency. Also, the outliers can be effectively filtered out by the motion-compensated mask. What's more, denser and more accurate depths also contribute to the outstanding localization accuracy of GRE-SLAM during global trajectory refinement.

5 Conclusions

In this paper, we presented the first pure event-based SLAM system for the full 6-DoF motion with the front-end and back-end structure, namely GRE-SLAM. Our GRE-SLAM provides a good initialization and a sharp referenced edge map for the front-end optimization using motion compensation, and recovers a semi-dense depth map for each IWE in the back-end to support the global optimization to ease accumulated drifts. A key characteristic of GRE-SLAM is that in its back-end, it utilizes events directly by mining their geometric and spatial information. Thus GRE-SLAM has no dependence on additional sensors or intermediate hand-crafted features. Experiments demonstrated the outperforming localization accuracy of GRE-SLAM. Besides, on challenging HDR scenes, our GRE-SLAM even surpasses the compared multi-modal SLAM methods integrating image frames and IMU. In summary, the goal of our work is to explore innovative SLAM techniques for long-term tracking using the novel event camera sensor, and narrow down the gap in tracking robustness and accuracy between pure event-based SLAM and traditional frame-based SLAM.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62272343.

References

[1] Kristoffer Fogh Andersen, Huy Xuan Pham, Halil Ibrahim Ugurlu, and Erdal Kayacan. 2022. Event-based Navigation for Autonomous Drone Racing with Sparse Gated Recurrent Network. In *2022 European Control Conference (ECC)*. 1342–1348. doi:10.23919/ECC55457.2022.9838538

[2] Simon Baker and Iain Matthews. 2004. Lucas-kanade 20 Years on: A Unifying Framework. *International journal of computer vision* 56 (2004), 221–255.

[3] Christian Brandli, Raphael Berner, Minhao Yang, Shih-Chii Liu, and Tobi Delbruck. 2014. A 240 × 180 130 dB 3 µs Latency Global Shutter Spatiotemporal Vision Sensor. *IEEE Journal of Solid-State Circuits* 49, 10 (2014), 2333–2341. doi:10.1109/JSSC.2014.2342715

[4] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel, and Juan D. Tardós. 2021. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. *IEEE Transactions on Robotics* 37, 6 (2021), 1874–1890. doi:10.1109/TRO.2021.3075644

[5] Andrea Censi and Davide Scaramuzza. 2014. Low-Latency Event-Based Visual Odometry. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*. 703–710. doi:10.1109/ICRA.2014.6906931

[6] Peiyu Chen, Weipeng Guan, and Peng Lu. 2023. ESVIO: Event-Based Stereo Visual Inertial Odometry. *IEEE Robotics and Automation Letters* 8, 6 (2023), 3661–3668. doi:10.1109/LRA.2023.3269950

[7] Matthew Cook, Luca Gugelmann, Florian Jug, Christoph Krautz, and Angelika Steger. 2011. Interacting Maps for Fast Visual Interpretation. In *The 2011 International Joint Conference on Neural Networks*. 770–776. doi:10.1109/IJCNN.2011.6033299

[8] Mingyue Cui, Yuzhang Zhu, Yechang Liu, Yunchao Liu, Gang Chen, and Kai Huang. 2022. Dense Depth-Map Estimation Based on Fusion of Event Camera and Sparse LiDAR. *IEEE Transactions on Instrumentation and Measurement* 71 (2022), 1–11. doi:10.1109/TIM.2022.3144229

[9] Jakob Engel, Thomas Schöps, and Daniel Cremers. 2014. LSD-SLAM: Large-Scale Direct Monocular SLAM. In *Computer Vision – ECCV 2014*, David Fleet,

Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 834–849.

[10] Christian Forster, Matia Pizzoli, and Davide Scaramuzza. 2014. SVO: Fast Semi-Direct Monocular Visual Odometry. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*. 15–22. doi:10.1109/ICRA.2014.6906584

[11] Guillermo Gallego, Jon E.A. Lund, Elias Mueggler, Henri Rebecq, Tobi Delbruck, and Davide Scaramuzza. 2018. Event-Based, 6-DOF Camera Tracking from Photometric Depth Maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 10 (2018), 2402–2412. doi:10.1109/TPAMI.2017.2769655

[12] Guillermo Gallego, Henri Rebecq, and Davide Scaramuzza. 2018. A Unifying Contrast Maximization Framework for Event Cameras, with Applications to Motion, Depth, and Optical Flow Estimation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3867–3876. doi:10.1109/CVPR.2018.00407

[13] Guillermo Gallego and Davide Scaramuzza. 2017. Accurate Angular Velocity Estimation With an Event Camera. *IEEE Robotics and Automation Letters* 2, 2 (2017), 632–639. doi:10.1109/LRA.2016.2647639

[14] Weipeng Guan, Peiyu Chen, Yuhan Xie, and Peng Lu. 2024. PL-EVIO: Robust Monocular Event-Based Visual Inertial Odometry With Point and Line Features. *IEEE Transactions on Automation Science and Engineering* 21, 4 (2024), 6277–6293. doi:10.1109/TASE.2023.3324365

[15] Weipeng Guan and Peng Lu. 2022. Monocular Event Visual Inertial Odometry based on Event-corner using Sliding Windows Graph-based Optimization. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2438–2445. doi:10.1109/IROS47612.2022.9981970

[16] Shuang Guo and Guillermo Gallego. 2024. CMax-SLAM: Event-Based Rotational-Motion Bundle Adjustment and SLAM System Using Contrast Maximization. *IEEE Transactions on Robotics* 40 (2024), 2442–2461. doi:10.1109/TRO.2024.3378443

[17] Javier Hidalgo-Carrio, Guillermo Gallego, and Davide Scaramuzza. 2022. Event-aided Direct Sparse Odometry . In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 5771–5780. doi:10.1109/CVPR52688.2022.00569

[18] Javier Hidalgo-Carrió, Daniel Gehrig, and Davide Scaramuzza. 2020. Learning Monocular Dense Depth from Events. In *2020 International Conference on 3D Vision (3DV)*. 534–542. doi:10.1109/3DV50981.2020.00063

[19] Xueyan Huang, Yueyi Zhang, and Zhiwei Xiong. 2023. Progressive Spatio-temporal Alignment for Efficient Event-based Motion Estimation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1537–1546. doi:10.1109/CVPR52729.2023.00154

[20] Haram Kim and H. Jin Kim. 2021. Real-Time Rotational Motion Estimation With Contrast Maximization Over Globally Aligned Events. *IEEE Robotics and Automation Letters* 6, 3 (2021), 6016–6023. doi:10.1109/LRA.2021.3088793

[21] Hanme Kim, Stefan Leutenegger, and Andrew J Davison. 2016. Real-time 3D Reconstruction and 6-DoF Tracking with An Event Camera. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14*. Springer, 349–364.

[22] Georg Klein and David Murray. 2007. Parallel Tracking and Mapping for Small AR Workspaces. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. 225–234. doi:10.1109/ISMAR.2007.4538852

[23] Simon Klenk, Marvin Motzet, Lukas Koestler, and Daniel Cremers. 2024. Deep Event Visual Odometry. In *2024 International Conference on 3D Vision (3DV)*. 739–749. doi:10.1109/3DV62453.2024.00036

[24] Beat Kueng, Elias Mueggler, Guillermo Gallego, and Davide Scaramuzza. 2016. Low-Latency Visual Odometry using Event-Based Feature Tracks. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 16–23. doi:10.1109/IROS.2016.7758089

[25] Patrick Lichtsteiner, Christoph Posch, and Tobi Delbruck. 2008. A 128× 128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits* 43, 2 (2008), 566–576. doi:10.1109/JSSC.2007.914337

[26] Daqi Liu, Álvaro Parra, and Tat-Jun Chin. 2021. Spatiotemporal Registration for Event-based Visual Odometry. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 4935–4944. doi:10.1109/CVPR46437.2021.00490

[27] Urbano Miguel Nunes and Yiannis Demiris. 2022. Robust Event-Based Vision Model Estimation by Dispersion Minimisation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 12 (2022), 9561–9573. doi:10.1109/TPAMI.2021.3130049

[28] Roberto Pellerito, Marco Cannici, Daniel Gehrig, Joris Belhadj, Olivier Dubois-Matra, Massimo Casasco, and Davide Scaramuzza. 2024. Deep Visual Odometry with Events and Frames. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 8966–8973.

[29] Tong Qin, Peiliang Li, and Shaojie Shen. 2018. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator. *IEEE Transactions on Robotics* 34, 4 (2018), 1004–1020. doi:10.1109/TRO.2018.2853729

[30] Henri Rebecq, Guillermo Gallego, Elias Mueggler, and Davide Scaramuzza. 2018. EMVS: Event-Based Multi-View Stereo—3D Reconstruction with An Event Camera in Real-Time. *International Journal of Computer Vision* 126, 12 (2018), 1394–1414.

[31] Henri Rebecq, Timo Horstschaefer, Guillermo Gallego, and Davide Scaramuzza. 2017. EVO: A Geometric Approach to Event-Based 6-DOF Parallel Tracking and Mapping in Real Time. *IEEE Robotics and Automation Letters* 2, 2 (2017), 593–600. doi:10.1109/LRA.2016.2645143

[32] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. 2021. High Speed and High Dynamic Range Video with an Event Camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 6 (2021), 1964–1980. doi:10.1109/TPAMI.2019.2963386

[33] Christian Reinbacher, Gottfried Munda, and Thomas Pock. 2017. Real-time Panoramic Tracking for Event Cameras. In *2017 IEEE International Conference on Computational Photography (ICCP)*. 1–9. doi:10.1109/ICCPHOT.2017.7951488

[34] Thomas Schöps, Jakob Engel, and Daniel Cremers. 2014. Semi-dense Visual Odometry for AR on A Smartphone. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 145–150. doi:10.1109/ISMAR.2014.6948420

[35] Bongki Son, Yunjae Suh, Sungho Kim, Heejae Jung, Jun-Seok Kim, Changwoo Shin, Keunju Park, Kyoobin Lee, Jinman Park, Jooyeon Woo, Yohan Roh, Hyunku Lee, Yibing Wang, Ilia Ovsiannikov, and Hyunsurk Ryu. 2017. A 640×480 Dynamic Vision Sensor with a 9μm Pixel and 300Meps Address-Event Representation. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)*. 66–67. doi:10.1109/ISSCC.2017.7870263

[36] Yuan Tian and Marc Compere. 2019. A Case Study on Visual-Inertial Odometry using Supervised, Semi-Supervised and Unsupervised Learning Methods. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 203–2034. doi:10.1109/AIVR46125.2019.00043

[37] Antoni Rosinol Vidal, Henri Rebecq, Timo Horstschaefer, and Davide Scaramuzza. 2018. Ultimate SLAM? Combining Events, Images, and IMU for Robust Visual SLAM in HDR and High-Speed Scenarios. *IEEE Robotics and Automation Letters* 3, 2 (2018), 994–1001. doi:10.1109/LRA.2018.2793357

[38] David Weikersdorfer, David B. Adrian, Daniel Cremers, and Jörg Conradt. 2014. Event-based 3D SLAM with A Depth-Augmented Dynamic Vision Sensor. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*. 359–364. doi:10.1109/ICRA.2014.6906882

[39] David Weikersdorfer, Raoul Hoffmann, and Jörg Conradt. 2013. Simultaneous Localization and Mapping for Event-Based Vision Systems. In *Computer Vision Systems: 9th International Conference, ICVS 2013, St. Petersburg, Russia, July 16-18, 2013. Proceedings 9*. Springer, 133–142.

[40] Tianyi Xiong, Jiayi Wu, Botao He, Cornelia Fermuller, Yiannis Aloimonos, Heng Huang, and Christopher Metzler. 2024. Event3DGS: Event-Based 3D Gaussian Splatting for High-Speed Robot Egomotion. In *8th Annual Conference on Robot Learning*. https://openreview.net/forum?id=EyEE7547vy

[41] Yi Zhou, Guillermo Gallego, Henri Rebecq, Laurent Kneip, Hongdong Li, and Davide Scaramuzza. 2018. Semi-dense 3D Reconstruction with A Stereo Event Camera. In *Proceedings of the European conference on computer vision (ECCV)*. 235–251.

[42] Yi Zhou, Guillermo Gallego, and Shaojie Shen. 2021. Event-Based Stereo Visual Odometry. *IEEE Transactions on Robotics* 37, 5 (2021), 1433–1450. doi:10.1109/TRO.2021.3062252

[43] Alex Zihao Zhu, Nikolay Atanasov, and Kostas Daniilidis. 2017. Event-Based Visual Inertial Odometry. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 5816–5824. doi:10.1109/CVPR.2017.616

[44] Dekai Zhu, Zhongcong Xu, Jinhu Dong, Canbo Ye, Yinbai Hu, Hang Su, Zhengfa Liu, and Guang Chen. 2019. Neuromorphic Visual Odometry System For Intelligent Vehicle Application With Bio-inspired Vision Sensor. In *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. 2225–2232. doi:10.1109/ROBIO49542.2019.8961878