The following publication G. Gong et al., "TEVIO: Thermal-Aided Event-Based Visual–Inertial Odometry for Robust State Estimation in Challenging Environments," in IEEE Transactions on Instrumentation and Measurement, vol. 74, pp. 1-11, 2025, Art no. 7505211 is available at https://doi.org/10.1109/TIM.2025.3552392.

TEVIO: Thermal-Aided Event-based Visual Inertial Odometry for Robust State Estimation in Challenging Environments

Gu Gong, *Student Member, IEEE*, Fuji Hu, Fangyuan Wang, *Student Member, IEEE*, Muhammed Muddassir, Peng Zhou, *Member, IEEE*, Lu Li, Qiang Wang, Zhen He and David Navarro-Alarcon, *Senior Member, IEEE*

***Abstract*—Event-based visual odometry excels in high dynamic range scenarios but struggles in extremely low-light or low-contrast conditions, motivating the integration of thermal imaging. This paper presents TEVIO, a multi-modal system that fuses thermal imaging, event-based vision, and inertial measurements to address the challenges of visual-inertial odometry in low-light, high-dynamic-range, and low-texture environments. An enhanced time surface map (ETSM) improves feature extraction for high-motion and low-texture scenes. A parallel frequency-varied tracking framework then estimates the pose stably and in high precision. Extensive tests on public event camera datasets and real-world outdoor vehicle experiments show TEVIO's superior tracking accuracy and robustness compared to state-of-the-art monocular methods like EVIO, enabling reliable pose estimation in conditions where conventional approaches fail. A video demonstration is available at https://youtu.be/RfWYU15WwsU.**

***Index Terms*—Dynamic vision sensor, thermal sensor, multi-modal fusion, visual-inertial odometry**

I. Introduction

VISUAL Odometry (VO) has become a fundamental technology in autonomous systems, robotics, and augmented reality by providing accurate localization and mapping without the GPS signal [20], [21]. Integrating an inertial measurement unit (IMU) to the visual odometry to advance visual-inertial odometry (VIO) [7] enhances VO to a more robust framework for estimating the camera's pose . Traditional frame-based cameras suffer from multiple limitations such as frame readout speed [8] and dynamic range [8], as illustrated in Fig. 1. In addition, severe motion, high dynamic range scenes, and low-texture or featureless environments can degrade the performance of standard frame-based approaches. NASA demonstrated the importance of robust perception pipelines by incorporating stereo cameras for Mars exploration rovers [21], and subsequent research has explored multi-modal fusion with LiDAR, range, and inertial sensors to augment depth information for monocular RGB cameras [22]–[26]. While these developments have led to increased position estimation accuracy, traditional frame-based methods still rely heavily on the visible spectrum, making them susceptible to motion blur and poor illumination conditions [27].

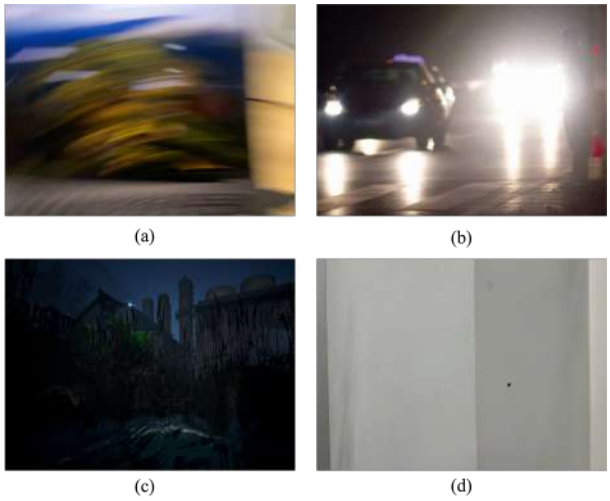


Fig. 1. Challenging scenarios that are unfeasible to a standard frame-based camera: (a) Severe Motion, (b) high dynamic range, (c) low light, (d) featureless

A bio-inspired sensor called an event camera has been introduced to address the inherent shortcomings of frame-based cameras [10]–[16]. Event cameras asynchronously capture changes in brightness with extremely high temporal resolution, significantly improving tracking and mapping performance under challenging conditions. Various event representation approaches have been proposed to process the asynchronous nature of these data, including the commonly used Surface-of-active-event (SAE) [17], [18], which preserves spatio-temporal information. Nonetheless, selecting and tuning the decay parameter remains non-trivial, potentially degrading performance in environments with drastically varying motion or illumination.

Recent advancements in event-based visual odometry (EVO) and event-based visual-inertial odometry (EVIO) have shown promising results, e.g., ESVO [32], which uses a probabilistic framework for depth estimation and camera track-

This work is supported in part by the Research Grants Council (RGC) of Hong Kong under grant 15212721, and in part by the Jiangsu Industrial Technology Research Institute Collaborative Funding Scheme under grant ZG9V. *Corresponding authors: David Navarro-Alarcon and Zhen He.*

G. Gong, F. Hu, F. Wang, M. Muddassir and D. Navarro-Alarcon are with the Department of Mechanical Engineering of The Hong Kong Polytechnic University, Kowloon, Hong Kong. (e-mail: davidgu.gong@connect.polyu.hk, emrys.hu@connect.polyu.hk, fangyuan.wang@connect.polyu.hk, mmudda@polyu.edu.hk, dnavar@polyu.edu.hk)

P. Zhou is with the Department of Computer Science, The University of Hong Kong, Pok Fu Lam, Hong Kong. (e-mail: jeffzhou@hku.hk)

L. Li is with the Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, China. (e-mail: lli@iamt.ac.cn)

Q. Wang and Z. He are with the Department of Control Science and Engineering, Harbin Institute of Technology, China. (e-mail: wangqiang@hit.edu.cn, hezhen@hit.edu.cn)

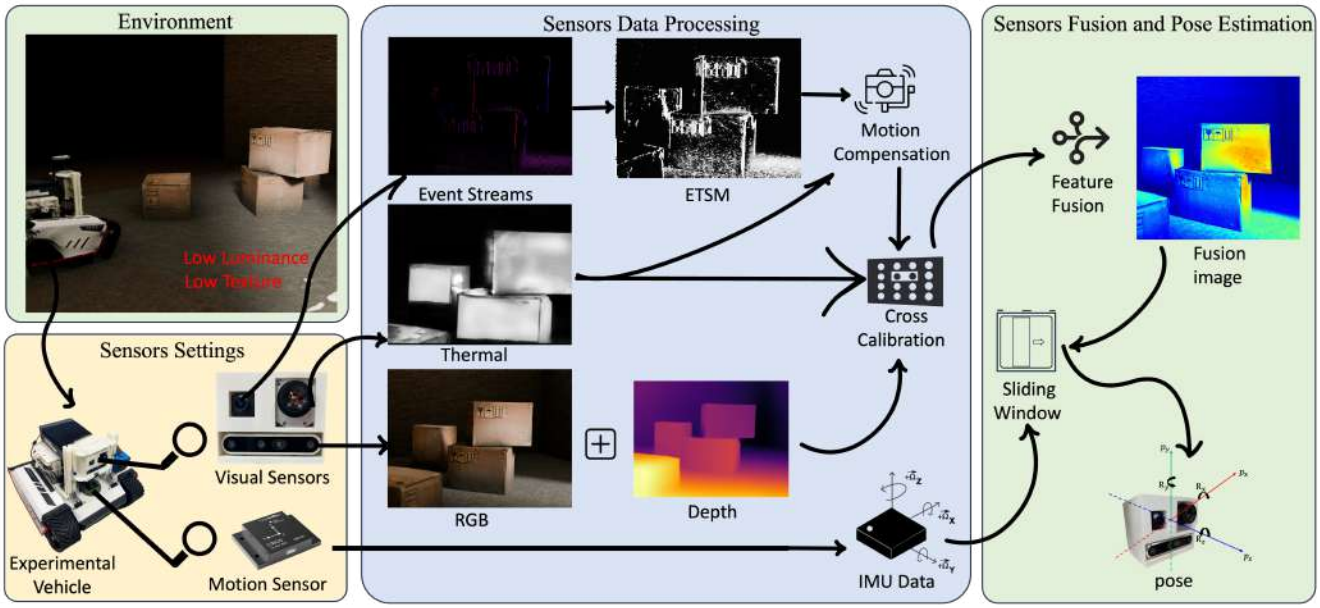


Fig. 2. Overview of proposed **TEVIO** visual inertial odometry pipeline. The system integrates an event camera, a long-wave infrared (LWIR) camera, and an RGB-D camera for initialization and perception. The three visual modalities are fused based on a cross-calibration structure and combined with inertial measurement unit (IMU) data for a tight-coupled fusion VIO using sliding window optimization to estimate the system's 6-degree-of-freedom (6DoF) pose. Please consider switching some icons, such as the sliding window. The RGB and Depth are not at the exact resolution.

ing, and ESVIO [10], a system that tightly fuses IMU with event data. However, these methods face limitations in specific scenarios. In purely dark or low-contrast environments, the absence of texture prevents events from being generated, and in extremely slow-motion or featureless settings, event sparsity can degrade tracking quality. To address such problems, many researchers have turned to thermal sensors (*i.e.,* LWIR cameras), which detect heat signatures rather than visible light and remain robust in darkness or low-texture conditions [2]–[4], [28]. These thermal-based solutions have been shown to reduce scale ambiguity by incorporating ground-plane constraints [3] and improve the robustness of visual-inertial odometry by fusing data from IMU and thermal sensors in darkness [2].

Despite these improvements, thermal imaging alone still inherits the typical problems of frame-based approaches, such as motion blur. Consequently, synergistically fusing thermal and event sensor data becomes a natural avenue for enhancing localization in scenarios suffering from rapid movement, low texture, or drastically changing light conditions. Recently, some efforts have begun to explore event-thermal fusion for nighttime navigation in aerial vehicles [15], [31]. Extended approaches are being investigated to preserve the event camera's dynamic range benefits while leveraging thermal imaging's ability to perceive features invisible to RGB cameras [33], [34], [36]. Furthermore, advanced frameworks for multi-sensor data alignment and motion compensation [29], [30] have confirmed that robust fusion strategies can significantly enhance the overall accuracy and reliability of VO/VIO systems. Recent works have also introduced integrated event-thermal solutions for visual odometry in extreme low-light scenarios [5], [6], highlighting the emerging importance of cross-modality sensor fusion.

To the best of our knowledge, this work presents the first thermal-event-visual method for odometry. The key contributions of this research are as follows:

- A new method for multi-modal visual-inertial odometry to estimate accurate pose in low light conditions.
- An enhanced time surface map (ETSM) with dynamic time-decay to improve the feature extraction quality of the event camera.
- A thermal-guided algorithm that enhances motion estimation in event cameras, reducing blur and improving data alignment by leveraging thermal information's invariance.
- A rigorous test of the proposed pipeline with the publicly available and newly created outdoor datasets.

The rest of this paper is structured as follows: Sec. II presents the preliminaries; Sec. III presents the methodology; Sec. IV reports the conducted experimental validation; Sec. V concludes the paper.

II. Preliminaries

A. Framework Overview

Our framework estimates the pose by implementing two parallel threads: a high-speed front-end thread and a high-precision back-end thread. The front-end thread fuses event, IMU, and thermal data using our proposed Enhanced Time-Surface Map (ETSM), which achieves a maximum generation time of 5 ms. The frame rates of all the sensors in our system are shown in Fig. 3 The back-end thread employs a sliding window optimizer for state estimation using front-end constraints. Our tracking module estimates 6-DOF ego-motion by aligning ETSM-constructed event frames with a semi-dense 3D map built from depth sensor data. The pose tracking process then minimizes a squared edge consistency

measure using a variant of the inverse compositional Lucas-Kanade method. The overview of the proposed system is shown in Fig. 2

B. Event Representations

Event cameras only generate information for the dynamic part of a scene, called an "event," which contains the pixel coordinates of the event, the trigger time, and the polarity. The i-th event in an event sequence is expressed as a set:

$$e_i = \{\mathbf{u}_i, t_i, p_i\} \tag{1}$$

Where e_i denotes the i-th event, $\mathbf{u}_i = [x_i \ y_i]^\top$ the event location on the image plane, t_i the timestamp and p_i the polarity. A Time Surface Map (TSM) [18] is an event representation method that records the motion history of each pixel in a recent time interval. It assigns a value to each pixel based on an exponential decay function, where the value depends on the time elapsed since the last motion event occurred at that pixel location. The exponential decay kernel is given by:

$$\mathcal{T}(\mathbf{u}, t) = \exp\left(-\frac{t - t_{\text{last}}(\mathbf{u})}{\tau}\right), \tag{2}$$

where t denotes the current time, $t_{\text{last}}(\mathbf{u})$ the timestamp of the most recent motion event at pixel $\mathbf{u}$, and τ, the decay rate parameter that needs to be tuned based on the motion dynamics. Pixels with higher values represent more recent motion events, emphasizing areas of recent activity. The TSM values are typically mapped to the range [0, 255] for visualization.

C. IMU Propagation Model

The state of the IMU at the $(i+1)^{\text{th}}$ time step is updated based on its position, velocity, and orientation relative to the world frame. The propagation equations are given by:

$$\begin{aligned}
p^W_{B_{i+1}} &= p^W_{B_i} + v^W_{B_i}\Delta t \\
&\quad + \iint_{t\in[t_i,t_{i+1}]} \left[q^W_{B_t}\cdot(a_t - b_{a_t})\left(q^W_{B_t}\right)^{-1} - g^W\right] dt^2 \\
v^W_{B_{i+1}} &= v^W_{B_i} + \int_{t\in[t_i,t_{i+1}]} \left[q^W_{B_t}\cdot(a_t - b_{a_t})\left(q^W_{B_t}\right)^{-1} - g^W\right] dt \\
q^W_{B_{i+1}} &= q^W_{B_i} \otimes \exp\left(\frac{1}{2}(\omega_t - b_{g_t})\Delta t\right)
\end{aligned} \tag{3}$$

In (3), $p^W_{B_i} \in \mathbb{R}^3$ and $v^W_{B_i} \in \mathbb{R}^3$ represent the position and velocity vectors of the IMU in the world frame at the i^{th} time step, respectively. The quaternion $q^W_{B_t} \in \mathbb{R}^4$ defines the orientation of the IMU frame relative to the world frame at time t. The acceleration $a_t \in \mathbb{R}^3$ and angular velocity $\omega_t \in \mathbb{R}^3$ are measured by the IMU, while $b_{a_t} \in \mathbb{R}^3$ and $b_{g_t} \in \mathbb{R}^3$ denote the accelerometer and gyroscope biases, respectively. The gravity vector $g^W \in \mathbb{R}^3$ is defined in the world frame, and $\Delta t \in \mathbb{R}$ represents the time interval between the i^{th} and $(i+1)^{\text{th}}$ time steps.

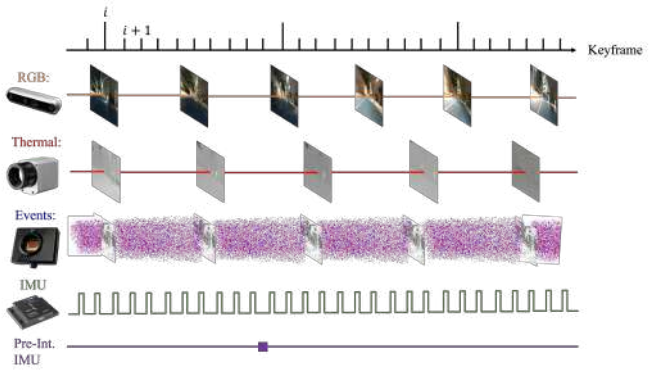


Fig. 3. Sensor data type and their frame rates are used in the multi-modal perception system. The RGB-D camera operates at 30 Hz, while the thermal camera captures frames at 25 Hz. The event camera provides asynchronous events with a temporal resolution of 200 μs. The IMU outputs angular velocities and linear accelerations at 200 Hz.

III. Methodology

The TEVIO pipeline consists of two concurrent processes. The front-end process fuses event, IMU, and thermal information to obtain event images for visual feature point detection and tracking. The back-end process constructs an optimization problem using the constraints from the front-end process to obtain the state estimation.

A. Event Feature Extraction

The event camera suffers from the camera motion. The constant decay [18] rate of the time surface map can cause an image with either too little or too much event data, neither of which is desirable. Here, we propose a novel adaptive time surface that calculates pixel-wise decay rate, $\tau(\mathbf{u})$, based on the surrounding pixels' timestamp, which is calculated by (4):

$$\begin{aligned}
\tau(\mathbf{x}) &= \max\left(\tau_u - \frac{1}{n_{\text{val}}}\sum_{i=0}^{n_q}\left(t - t_{\text{last},i}\right), \tau_l\right), \\
n_q &= \text{count}(\mathbf{u}, n) = |\{\mathbf{u}' \in \mathcal{N}_n(\mathbf{u}) \mid \mathcal{T}_{\text{ref}}(\mathbf{u}') > \delta(\mathbf{u})\}|
\end{aligned} \tag{4}$$

where τ_u and τ_l are the upper and lower bounds of the decay rate. $t_{last,i}$ represents the i th timestamp of the neighborhood n pixels around the reference center pixel at the reference time t_{ref}. The qualified neighborhood n_q pixels are filtered by comparing its TSM value with a threshold function $\delta(\mathbf{u})$ determined by the central pixel's aligned depth and temperature.

By subtracting the average time gap between the current timestamp and the surrounding pixels' timestamps from the upper bound, the ETSM achieves faster decay in high-texture or high-speed scenarios, preventing pixel overlap. Conversely, it allows for slower decay in low-texture or low-speed environments, ensuring sufficient information capture. The resulting decay rates are refined using blur and median blur filters to smooth the output. This adaptive approach significantly improves pixel selection and enhances the distinctness of the time surface, ultimately leading to more robust and accurate event-based vision processing across diverse scenarios.

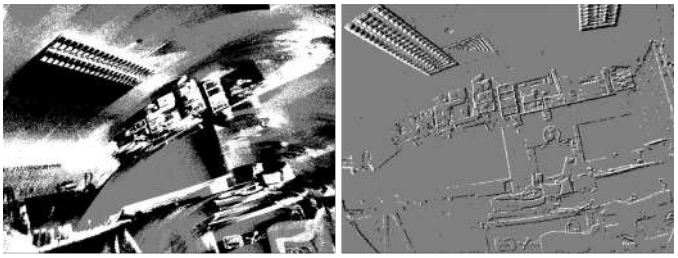


Fig. 4. Raw Event frames (left) and compensated frames (right) for the **ETSM** processing. White pixels indicate positive events, black for negative ones, and gray for where no events occurred at these positions.

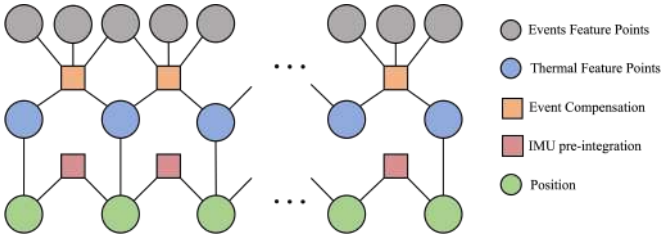


Fig. 5. Back-end optimization. A back-end optimization framework for precise pose estimation. The sliding window optimization incorporates three constraint terms: (1) IMU pre-integration for efficient storage of intermediate inertial data, (2) event data reprojection onto the thermal modality to compute reprojection error, and (3) marginalization to retain information from previously optimized results.

B. Thermal-Guided Event Data Motion Compensation

We apply motion compensation to the events to further ensure that the extraction of features is not affected by the speed of motion. The effectiveness of the algorithm is shown in the Fig. 4

$T(x, y, t)$ represents the temperature at pixel (x, y) at time t. The thermal gradient ∇T is computed as:

$$\nabla T = \left(\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y} \right) \tag{5}$$

We model the motion of events within a sliding window W using a warping function $W(x, y, t; \theta)$, where θ represents the motion parameters. The warping function maps events from their original positions to motion-compensated positions $\hat{e}_i = W(e_i; \theta) = (x'_i, y'_i, t_i, p_i)$, where (x'_i, y'_i) is the motion-compensated position of the event. The proposed algorithm for thermal-based motion compensation of event data involves initializing a sliding window W with N events and corresponding thermal data. For each iteration, thermal gradients ∇T within the window are computed, and initial motion parameters θ_0 are estimated using these gradients. The motion parameters are then optimized by minimizing the objective function:

$$\theta^* = \arg\min_{\theta} \sum_{e_i \in W} \left\| W\left(e_i; \theta\right) - e_i \right\|^2 + \lambda \left\| \nabla T \cdot \frac{\partial W}{\partial \theta} \right\|^2 \tag{6}$$

where λ is a weighting factor balancing the event alignment error and thermal gradient consistency, the event positions are updated using the optimized parameters $\hat{e}_i = W(e_i; \theta^*)$. The sliding window then moves forward, adding new events and thermal data. The motion compensation problem is formulated as a sliding window nonlinear optimization. The objective function $J(\theta)$ combines event alignment error and thermal gradient consistency:

$$J(\theta) = \sum_{e_i \in W} \left\| W\left(e_i; \theta\right) - e_i \right\|^2 + \lambda \left\| \nabla T \cdot \frac{\partial W}{\partial \theta} \right\|^2 \tag{7}$$

The optimization is solved using the Levenberg-Marquardt algorithm $\theta_{k+1} = \theta_k - (H + \mu I)^{-1} \nabla J(\theta_k)$ to solve, where H is the Hessian matrix, μ is the damping factor, and $\nabla J(\theta_k)$ is the gradient of the objective function.

In this algorithm 1, H represents the Hessian matrix, which is used in the optimization process to approximate the second-order derivatives of the objective function. The damping factor μ is a scalar that stabilizes the Levenberg-Marquardt update by preventing large jumps during the optimization. The parameter ϵ is the convergence threshold that determines when the optimization should terminate, specifically when the change in the motion parameters θ between iterations falls below this small predefined value.

To enhance motion estimation, we propose a fusion term $F(\theta)$ that combines event data with thermal information:

$$F(\theta) = \alpha E(\theta) + (1 - \alpha) T(\theta) \tag{8}$$

where $E(\theta)$ is the event-based error term, $T(\theta)$ is the thermal-based error term, and $\alpha \in [0, 1]$ is a fusion parameter.

C. Front-End of the TEVIO

The front-end of TEVIO comprises two primary components: semi-dense 3D map construction and pose tracking. The semi-dense 3D map construction process utilizes a feature detection function F and a depth extraction function D. The feature detection function is formalized as $P = F(I) \rightarrow \{p_i\}_{i=1}^{N}$, where I represents the input RGB image, and (I) identifies feature points p_i using the FAST algorithm for efficient keypoint

Algorithm 1: Thermal-Guided Event Data Motion Compensation

Input: E: Event data, T: Thermal data, N: Window size
Output: $\hat{E}$: Compensated event data

$W \leftarrow$ Initialize sliding window with N events;
while $|E| > 0$ **do**
  $\nabla T \leftarrow (\frac{\partial T}{\partial x}, \frac{\partial T}{\partial y})$;
  $\theta_0 \leftarrow$ Estimate from ∇T;
  for $k = 0$ **to** $max\_iterations$ **do**
    $J(\theta_k) \leftarrow \sum_{e_i \in W} |W(e_i; \theta) - e_i|^2 + \lambda |\nabla T \cdot \frac{\partial W}{\partial \theta}|^2$;
    Compute $H, \nabla J(\theta_k)$;
    $\theta_{k+1} \leftarrow \theta_k - (H + \mu I)^{-1} \nabla J(\theta_k)$;
    if $|\theta_{k+1} - \theta_k| < \epsilon$ **then**
      break;
  $\hat{e}_i \leftarrow W(e_i; \theta^*)$ for all $e_i \in W$;
  $\hat{E} \leftarrow \hat{E} \cup \hat{e}_i$;
  $W \leftarrow$ Update with new events;
return $\hat{E}$

detection. Subsequently, the depth extraction function generates the semi-dense point cloud P_d as follows: $P_d = \{(x_i, y_i, D(x_i, y_i)) \mid (x_i, y_i) \in p_i, \forall i = 1, \ldots, N\}$, where $D(x_i, y_i)$ retrieves the depth value at feature point (x_i, y_i).

The pose tracking approach incrementally refines the camera's 6-DOF transformation by aligning the ETSM $\mathcal{I}$ with a synthesized edge map $\mathcal{M}$ generated from the semi-dense point cloud. This alignment process solves for the rigid motion $\mathbf{T}$ that best superimposes $\mathcal{I}$ onto $\mathcal{M}$. TEVIO employs a variant of the inverse compositional Lucas-Kanade method to minimize a squared edge consistency measure, defined as:

$$\Delta\mathbf{T} \;=\; \underset{\Delta\mathbf{T}}{\operatorname{argmin}} \sum_{\mathbf{q}} \big[\, \mathcal{M}\big(\mathbf{W}(\mathbf{q};\, \Delta\mathbf{T})\big) \;-\; \mathcal{I}\big(\mathbf{W}(\mathbf{q};\, \mathbf{T})\big)\big]^2 \tag{9}$$

where $\mathbf{T}$ is the current estimation of the camera pose, $\Delta\mathbf{T}$ is the incremental update we want to solve for, and $\mathbf{q}$ indexes the relevant pixels in $\mathcal{M}$. The warp function $\mathbf{W}(\mathbf{q};\, \mathbf{T})$ takes into account the depth at pixel $\mathbf{q}$ (denoted $d_{\mathbf{q}}$) and applies a 3D transformation followed by a projection back to the 2D image plane. Once $\Delta\mathbf{T}$ is found, it is composited with the previous estimate in an inverse compositional manner. This refined state estimation from the front-end module is an initial guess for the back-end optimization described in Section E, ensuring a robust and efficient visual-inertial odometry pipeline.

D. Back-End with Sliding Window Nonlinear Optimization

The sliding window optimization problem in TEVIO aims to estimate the optimal state trajectory within a fixed-size window of recent measurements, incorporating thermal data, events, IMU data, and prior information.

To retain the observational data and constraints associated with previous keyframes, a marginalization strategy is implemented to convert them into state-prior constraints within the optimization window. Consequently, the comprehensive cost function of the back-end optimization incorporates inertial measurement unit (IMU) pre-integration constraints, thermal-event reprojection constraints, and marginalized prior constraints. Fig. 5 illustrates the factor graph representation of the back-end optimization components.

IMU pre-integration constraints. The IMU pre-integration provides constraints between consecutive frames regarding position, velocity, and attitude. The pre-integrated measurements are defined as follows:

$$\begin{aligned} p_{B_{i+1}}^{B_i} &= v_{B_i}^{B_i}\Delta t - \frac{1}{2} g^{B_i}\Delta t^2 + \alpha_{B_{i+1}}^{B_i} \\ v_{B_{i+1}}^{B_i} &= v_{B_i}^{B_i} - g^{B_i}\Delta t + \beta_{B_{i+1}}^{B_i} \\ q_{B_{i+1}}^{B_i} &= q_{B_i}^{B_i} \otimes \gamma_{B_{i+1}}^{B_i} \end{aligned} \tag{10}$$

where $\alpha_{B_{i+1}}^{B_i}$, $\beta_{B_{i+1}}^{B_i}$, and $\gamma_{B_{i+1}}^{B_i}$ are the pre-integrated quantities defined as:

$$\begin{aligned} \alpha_{B_{i+1}}^{B_i} &= \iint_{t\in[t_i,t_{i+1}]} R_t^{B_i}(a_t - b_{a_t})dt^2 \\ \beta_{B_{i+1}}^{B_i} &= \int_{t\in[t_i,t_{i+1}]} R_t^{B_i}(a_t - b_{a_t})dt \\ \gamma_{B_{i+1}}^{B_i} &= \int_{t\in[t_i,t_{i+1}]} \frac{1}{2}\Omega(\omega_t - b_{g_t}) q_t^{B_i} dt \end{aligned} \tag{11}$$

The residuals for the IMU pre-integration constraints are constructed as follows:

$$\begin{aligned} \delta\alpha_{B_{i+1}}^{B_i} &= R_W^{B_i}(p_{B_{i+1}}^W - p_{B_i}^W + \frac{1}{2} g^W \Delta t^2 - v_{B_i}^W \Delta t) - \alpha_{B_{i+1}}^{B_i} \\ \delta\beta_{B_{i+1}}^{B_i} &= R_W^{B_i}(v_{B_{i+1}}^W - v_{B_i}^W + g^W \Delta t) - \beta_{B_{i+1}}^{B_i} \\ \delta\theta_{B_{i+1}}^{B_i} &= 2[q_{B_{i+1}}^W \otimes (q_{B_i}^W)^{-1} \otimes \gamma_{B_{i+1}}^{B_i} - 1]_{xyz} \\ \delta b_a &= b_{a_{B_{i+1}}} - b_{a_{B_i}} \\ \delta b_g &= b_{g_{B_{i+1}}} - b_{g_{B_i}} \end{aligned} \tag{12}$$

In the pre-integration functions (12), $b_{a_{B_i}}$ and $b_{g_{B_i}}$ are the accelerometer and gyroscope biases for frame i and $[*]_{xyz}$ represents the vector part of a quaternion. The IMU pre-integration constraint in the sliding window optimization can be expressed as:

$$\begin{aligned} C_{\text{imu}} = \sum_i \Big(& \delta\alpha_{B_{i+1}}^{B_i}{}^T \Sigma_\alpha^{-1} \delta\alpha_{B_{i+1}}^{B_i} + \delta\beta_{B_{i+1}}^{B_i}{}^T \Sigma_\beta^{-1} \delta\beta_{B_{i+1}}^{B_i} \\ & + \delta\theta_{B_{i+1}}^{B_i}{}^T \Sigma_\theta^{-1} \delta\theta_{B_{i+1}}^{B_i} + \delta b_a^T \Sigma_{b_a}^{-1} \delta b_a + \delta b_g^T \Sigma_{b_g}^{-1} \delta b_g \Big) \end{aligned} \tag{13}$$

where Σ_α, Σ_β, Σ_θ, Σ_{b_a}, and Σ_{b_g} are the covariance matrices associated with each residual term.

Thermal-Event Reprojection Constraints Thermal-event reprojection constraints ensure that the observed events in the thermal camera align with the predicted events based on the estimated state. These constraints are formulated as follows:

Given a set of thermal events $\{e_i\}$, each event e_i is characterized by its pixel coordinates (u_i, v_i) and timestamp t_i. The reprojection error for each event is defined as the difference between the observed event position and the projected position based on the current state estimate $\mathbf{x}$:

$$r_i = \pi(\mathbf{T}_{cw} \cdot \mathbf{P}_i) - \mathbf{z}_i \tag{14}$$

In the reprojection error (14), $\pi(\cdot)$ is the projection function, $\mathbf{T}_{cw}$ is the transformation matrix from the world frame to the camera frame, $\mathbf{P}_i$ is the 3D point corresponding to event e_i and $\mathbf{z}_i = (u_i, v_i)$ is the observed pixel position of event e_i.

The total reprojection error for all events in the sliding window is minimized:

$$C_{\text{reproj}} = \sum_i \|r_i\|^2 \tag{15}$$

Marginalized Priori Constraints To manage the optimization dimension while preserving the observational or constraint information from older keyframes, marginalized priors incorporate information from past states that have been marginalized out of the sliding window. The marginalized prior constraint can be expressed as

$$C_{\text{prior}} = \|\mathbf{x}_{\text{marg}} - \mathbf{H}\mathbf{x}_{\text{current}}\|^2 \tag{16}$$

where $\mathbf{x}_{\text{marg}}$ is the marginalized state, $\mathbf{H}$ is the Jacobian matrix relating the marginalized state to the current state, and $\mathbf{x}_{\text{current}}$ is the current state vector.

Sliding Window Optimization The sliding window optimization framework maintains a fixed-size window of recent states and measurements. The overall cost function to be minimized is a combination of the above constraints:

$$C_{\text{total}} = C_{\text{reproj}} + C_{\text{prior}} + C_{\text{imu}} \tag{17}$$

TABLE I
RMSE BETWEEN SOME ALGORITHMS

Sequence	Motion	Ours	ORB3	EVO-L	Ult
hdr_boxes	6 DOF, incr. speed	**0.19**	0.54	0.24	0.28
hdr_posters	6 DOF, incr. speed	**0.20**	0.63	0.38	0.36
shapes_trans	Trans., incr. speed	0.27	0.34	**0.17**	0.26
shapes_6dof	6 DOF, incr. speed	**0.30**	0.73	0.36	0.42
poster_trans	Trans., incr. speed	**0.16**	0.43	0.19	0.29
poster_6dof	6 DOF, incr. speed	**0.15**	0.34	0.19	0.24
dynamic_trans	Rotation, incr.speed	**0.16**	0.33	0.17	0.35
dynamic_6dof	Trans., incr. speed	0.25	0.60	**0.24**	0.40

The optimization problem is solved iteratively, updating the state estimates within the sliding window to minimize C_{total}. This approach ensures the system remains robust to noise and maintains accurate state estimates over time.

By incorporating these constraints, the sliding window nonlinear optimization framework in TEVIO effectively fuses thermal event data, visual data, and inertial measurements to provide reliable odometry in challenging environments.

The back-end optimization's per-iteration complexity is approximated as $\mathcal{O}(N_{\text{IMU}} + N_{\text{evt}} + p)$, where N_{IMU} is the number of IMU factors, N_{evt} is the number of reprojected features, and p is the dimension of the retained state. Empirical tests show that our optimization step completes in tens to hundreds of milliseconds, which is feasible for most real-time applications. For synchronization, we utilize the natural timing property of event data, creating "hard-synchronized" data packets for the front-end thread. The back-end uses event timestamps as an anchor for accurate time alignment across all sensor data, enabling reliable sensor fusion without complex hardware triggers or time-offset estimations.

IV. Experiments

In this section, we evaluate the performance of our proposed algorithm through three sets of experiments. First, we evaluate the accuracy of our proposed event-only tracking method, which is equipped with the ETSM feature and other event-only algorithms, on the public dataset. Then, we conduct an outdoor experiment using a vehicle equipped with our system to verify the performance in complex conditions day and night, providing insights into its practical applicability and reliability in real-world vehicular contexts.

A. Dataset Experiments: Our EVO versus Other Algorithms

We evaluated our method using datasets from human-made environments captured with the DAVIS 240C camera models [37], including eight sequences with rapid motions and high-dynamic-range scenes. The DAVIS-240C features a 240×180 pixel resolution, a 1 kHz IMU, and provides standard frames at 24 Hz. These datasets, referred to as HDR boxes, HDR Poster, Shapes 6DOF, Shapes translation, Poster 6DOF, and Poster translation, varying depths, camera speeds, and medium-textured surfaces. Furthermore, our approach was tested on the Dynamic 6DOF and Dynamic translation sequence, which is considered the most representative human-made scenario.

To demonstrate the robustness of our EVO in the highly dynamic and low-texture environment on other monocular algorithms, we conducted comparative tests with Ultimate-SLAM, ORBSlam3, and SVO. EVO was bootstrapped with SVO in the test sequences. TableI shows the RMSE(root mean squared error) comparison between the 8 test sequences.

We can observe in Fig. 6 that the consistency of feature tracking and coherent mapping are maintained even in rapidly changing scenes. In Fig. 7, the RMSE box plots compare ETSM-EVO, ORB3, EVO-L, and Ult-SLAM across eight event-camera datasets (HDR-Boxes, Shapes-6DoF, etc.). Our ETSM-EVO maintains the smallest or near-smallest median RMSE in nearly all scenarios, indicating lower average error and reduced variance. In high-speed or high-dynamic-range sequences (e.g., HDR-Boxes, Poster-6DoF), ETSM-EVO's interquartile ranges are tighter, reflecting greater robustness compared to the RMSE distributions of ORB3, EVO-L, and Ult-SLAM. Additionally, we focus on the results affirming our EVO's robustness and effectiveness in dynamic and low-texture environments while highlighting the limitations of the odometry of monocular standard frame-based cameras. Fig. 8 also shows our ETSM-EVO maintains a trajectory closer to the ground truth and experiences minimum drift under rapid camera motions. These consistent improvements affirm that adaptive event-driven feature extraction enhances pose tracking and reduces drift, especially in challenging motion and lighting conditions, thus improving overall odometry accuracy.

The experimental results demonstrate that our EVO algorithm consistently outperforms other state-of-the-art monocular visual odometry algorithms in challenging scenarios involving high dynamic range, rapid motion, and complex 3D environments. EVO's robustness and accuracy make it suitable for real-time applications in dynamic and low-texture environments.

B. Outdoor Environment

In our outdoor vehicle experiment, we utilized an Agilex Bunker vehicle equipped with our multi-modal sensor system to evaluate TEVIO's performance. The sensor suite comprised three primary components: a DVXplorer Mini event camera for capturing raw event data, an Optris PI 450 thermal camera for thermal imaging, and an Intel RealSense D455 for RGBD data acquisition to facilitate system bootstrapping. The Agilex Bunker was chosen for its stability and maneuverability across various terrains, making it ideal for diverse environmental testing.

Data collection and state estimation were performed using an industrial computer with an Intel i5-12400 CPU running Ubuntu 20.04 and the Robot Operating System (ROS). This setup provided sufficient computational power to handle the real-time processing demands of our multimodal odometry system. We employed a Ublox F9P RTK (Real-Time Kinematic) module to establish an accurate ground truth for performance evaluation. This high-precision GNSS system offers centimeter-level accuracy, allowing for reliable benchmarking of our TEVIO algorithm against a robust reference trajectory, the whole experimental vehicle system is shown in Fig. 9.

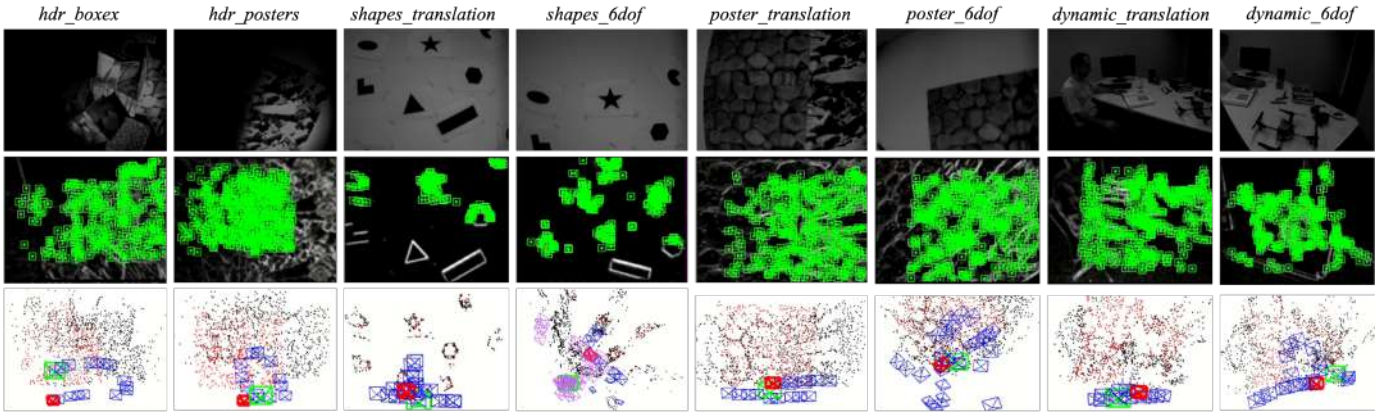


Fig. 6. Visual representation of EVO tracking and mapping performance. Each row represents the intensity images from the input sequence, feature tracking results, camera trajectory estimation, and semi-dense mapping output.

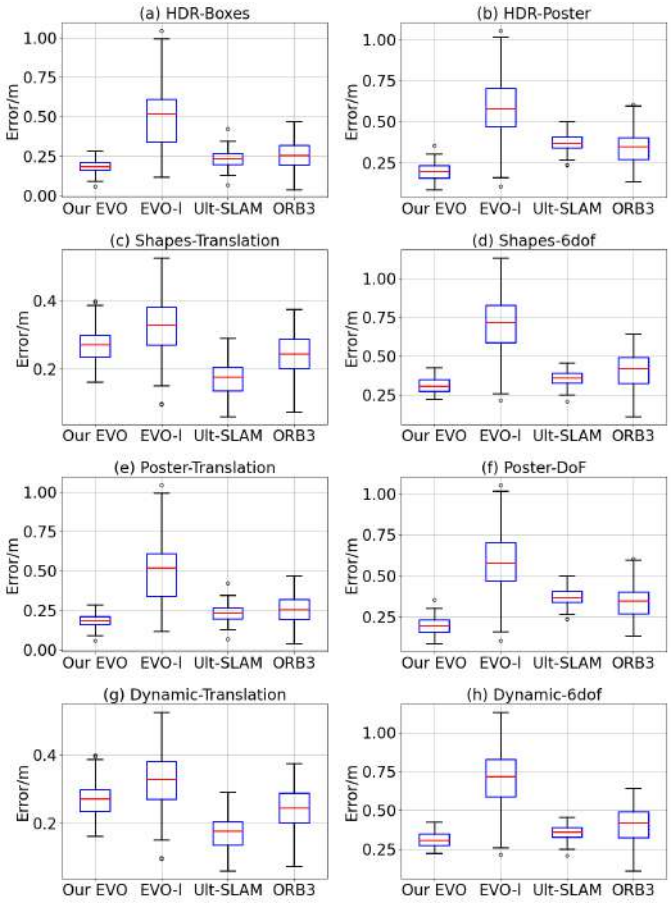


Fig. 7. Box plots showing the Root Mean Square Error (RMSE) results for various methods across different datasets. The methods evaluated include Our ETSM-EVO, ORB3, EVO-L, and Ult-SLAM. Each subplot represents a different dataset scenario: (a) HDR-Boxes, (b) HDR-Poster, (c) Shapes-Translation, (d) Shapes-6DoF, (e) Poster-Translation, (f) Poster-DoF, (g) Dynamic-Translation, and (h) Dynamic-6DoF. Lower RMSE values indicate better performance.

We conducted several outdoor vehicle experiments on the HKPolyU campus to validate the performance of TEVIO. For comparison, the high-performance event-inertial odometry EVIO and the widely applied odometry ORB-SLAM3 were chosen. We tested these algorithms across five sequences, each presenting a unique environment combination regarding illumination, time duration, feature amount, and motion dynamics. The details of these sequences are presented in the Table II. It needs to be emphasized that the **ROMI** and **Runway** sequences provide a long-term nighttime test, demonstrating TEVIO's robust performance in extreme illuminance and highly dynamic environments over extended periods as Fig. 10 shows.

TABLE II
ENVIRONMENT SETTING FOR EVALUATION SEQUENCES

Sequence	Time of day	Duration	Property
Triangle	Day (1pm)	42.4s	High dynamic
Circle	Day (1pm)	52.5s	Low features
Runway	Night (8pm)	265.2s	Long-term
Square	Midnight (11pm)	56s	Low features
ROMI	Midnight (1am)	138.2s	Highly dynamic

These sequences are particularly challenging for visual-inertial odometry systems and provide a comprehensive evaluation of long-term stability and accuracy. Fig. 11 presents the estimated trajectories of TEVIO, EVIO, and ORB-SLAM3 compared to the ground truth for each sequence, along with their respective position errors. Our TEVIO intuitively shows the best results in each sequence.

To quantitatively assess the performance of TEVIO against EVIO and ORB-SLAM3, we employed three standard metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Maximum Error (Max Error). RMSE measures the standard deviation of prediction errors, giving higher weight to significant errors. MAE represents the average magnitude of errors without considering their direction, providing a linear score. Max Error indicates the most significant discrepancy between estimated and ground truth positions, highlighting worst-case performance. Table III summarizes the quantitative results across all sequences and metrics. These results reveal that TEVIO consistently outperforms both EVIO and ORB-SLAM3 across all sequences regarding RMSE and MAE, demonstrating its superior accuracy and robustness in diverse environments. In the **Runway** sequence, TEVIO

TABLE III
EVALUATION OF DIFFERENT SEQUENCES

	RMSE			MAE			Max Error		
	TEVIO	EVIO	ORB	TEVIO	EVIO	ORB	TEVIO	EVIO	ORB
triangle	**0.3210**	0.6204	1.2732	**0.3706**	0.5319	1.0184	2.3216	**1.1656**	2.1542
circle	**0.6470**	0.8628	2.8312	**0.5598**	0.7714	2.6538	**1.3600**	1.4425	4.0803
runway	**0.1073**	0.2370	18.0607	**0.0944**	0.2111	16.1248	**0.3574**	0.5975	22.0043
square	**1.1107**	2.0799	5.5376	**1.0320**	2.0100	5.1949	**1.8305**	3.0752	7.0822
romi	**4.2689**	4.3870	6.8403	**3.7413**	3.9191	5.9087	**8.0824**	8.5559	13.8787

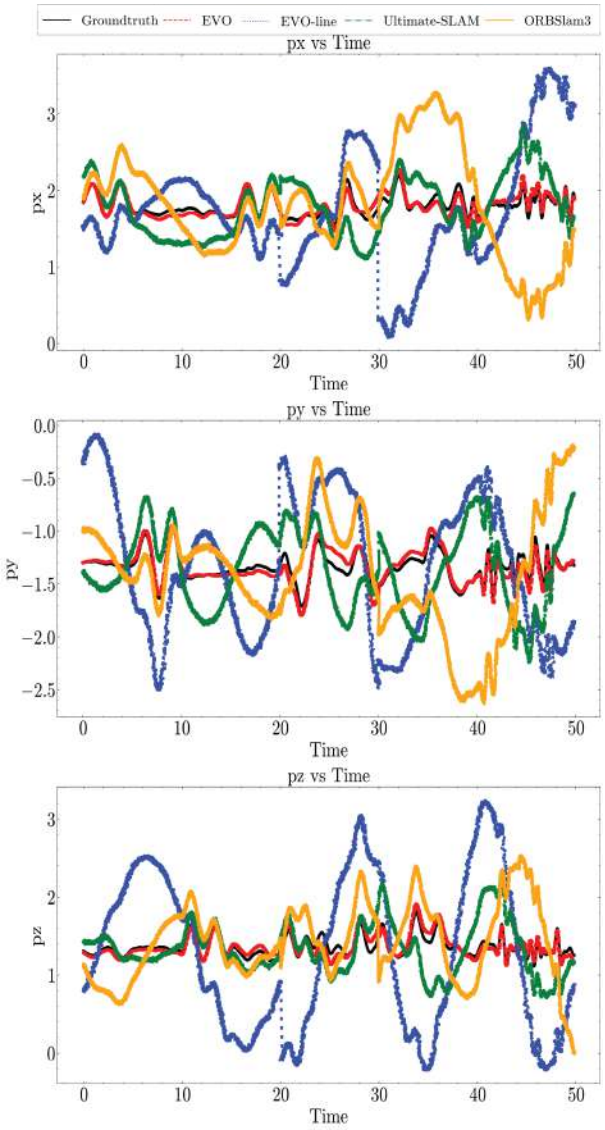


Fig. 8. The position estimation of *dynamic-6dof* sequence.

achieves remarkable accuracy with an RMSE of 0.1073m, MAE of 0.0944m, and Max Error of 0.3574m. This performance significantly surpasses EVIO (RMSE: 0.2370m) and ORB-SLAM3 (RMSE: 18.0607m), underscoring TEVIO's exceptional long-term stability in challenging nighttime conditions. The ROMI sequence also showcases TEVIO's ability to handle relatively long-ter,m highly dynamic midnight scenarios.

Fig. 9. Experimental vehicle system(**left**) : (1) Dvxplorer mini, (2) Optris 450PI, (3) Realsense D455, (4) CMP10A, (5) Intel i5-12400 CPU computer and environment (**right**).

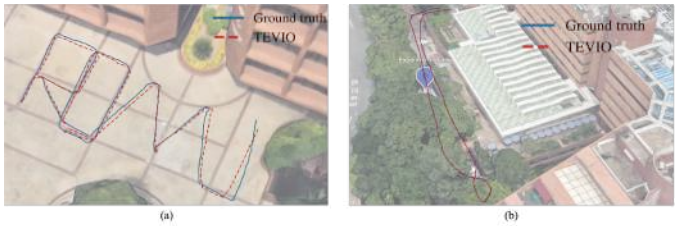


Fig. 10. TEVIO tracking results on both ROMI and Runway sequences referring to the satellite image show that TEVIO has good tracking results in relatively long-term experiments.

The integration of event-based motion compensation and thermal information significantly enhances TEVIO's resilience to rapid scene changes and varying illumination, as demonstrated by the consistent performance across all test sequences.

V. CONCLUSION

This paper presented TEVIO, a Thermal Event Visual Inertial Odometry system that integrates thermal imaging, event cameras, and inertial measurements to achieve robust and accurate state estimation in challenging environments. Our contributions include developing a multi-modal visual-inertial odometry system for low-light conditions, applying an enhanced time surface map (ETSM) for event cameras, and the thermal-guided motion compensation algorithm for event data to improve motion estimation accuracy. The experimental results demonstrate the effectiveness of our approach, showing superior performance in both high-dynamic event datasets and real-world outdoor scenarios. TEVIO consistently outperforms existing monocular methods, particularly in low-light, high-motion, and low-texture environments. It proves its potential for applications in autonomous navigation, search and rescue

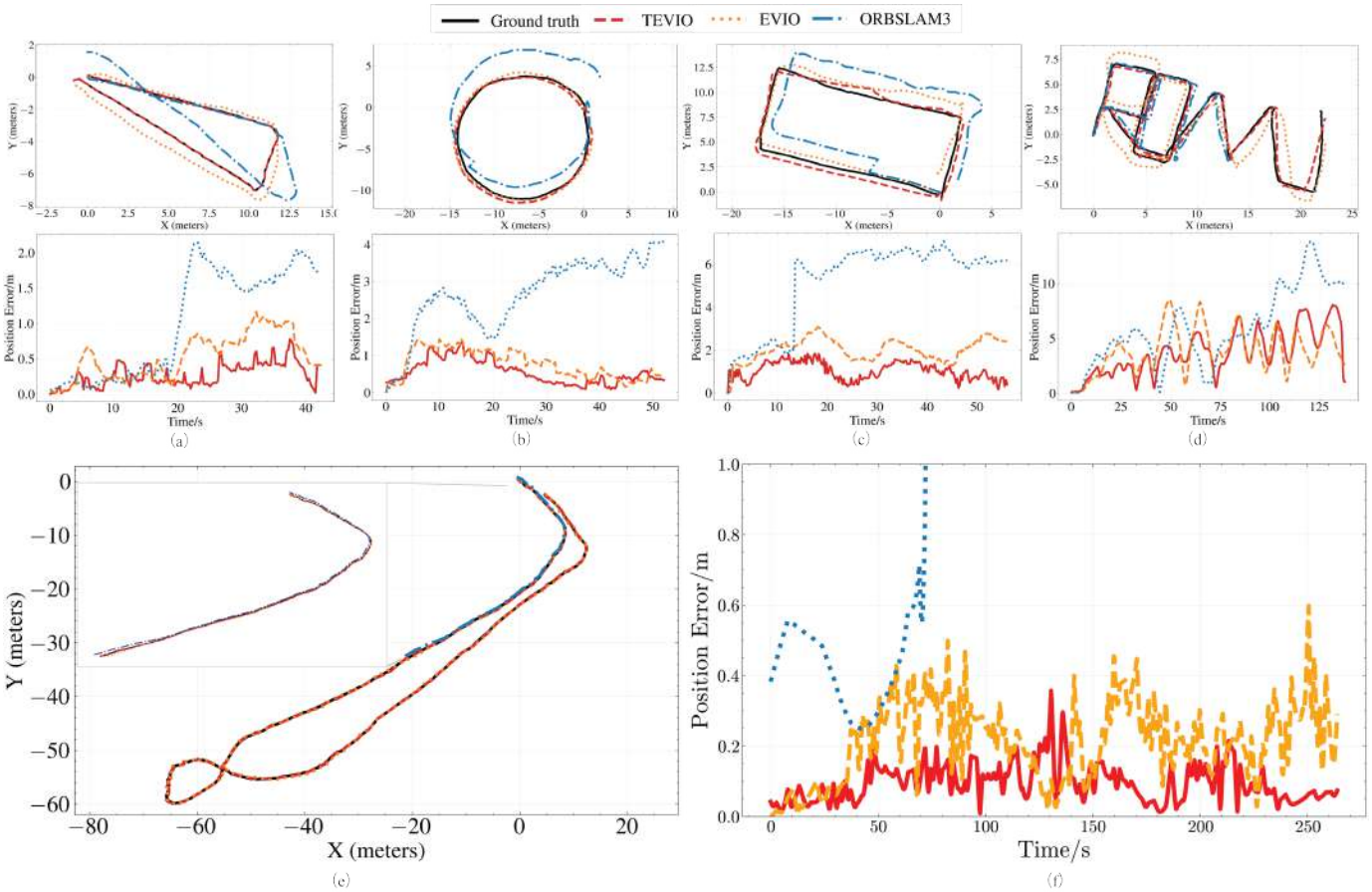


Fig. 11. Trajectories and position errors of TEVIO, EVIO and ORBSlam3 tested on the sequences **triangle**, **circle**, **square**, **ROMI** and **Runway**. Our TEVIO shows accuracy and robustness in all the sequences. In **Runway** trajectory figure (e) and position error figure (f), our TEVIO shows dramatic tracking performance compared to the EVIO and the drifted Orbslam3 in this long-term sequence.

missions, and other fields requiring reliable odometry under challenging conditions.

Future work will focus onhigh-precision,d rendered 3D mapping of multiple modalities with event-centricmapping and exploring additional sensor modalities to enhanceTEVIO'se robustness and versatilitO. The promising results indicate that the fusion of thermal, event-based, and inertial data holds significant potential for advancing the stability of state estimation in various complicated environments.

REFERENCES

[1] A. Alekseev, E. Goshin, N. Davydov, N. Ivliev, and A. Nikonorov, "Visual-inertial odometry algorithms on the base of thermal camera," vol. 2416, pp. 183–188, 2019.

[2] J. Delaune, R. Hewitt, L. Lytle, C. Sorice, R. Thakker, and L. Matthies, "Thermal-inertial odometry for autonomous flight throughout the night," pp. 1122–1128, 2019.

[3] P. V. K. Borges and S. Vidas, "Practical infrared visual odometry," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 8, pp. 2205–2213, 2016.

[4] B. R. van Manen, V. Sluiter, and A. Y. Mersha, "Firebotslam: thermal slam to increase situational awareness in smoke-filled environments," *Sensors*, vol. 23, no. 17, p. 7611, 2023.

[5] R. Edlinger, G. Himmelbauer, G. Zauner, and A. Nüchter, "Visual odometry and mapping under poor visibility conditions using a stereo infrared thermal imaging system," *Electronic Imaging*, vol. 35, pp. 1–7, 2023.

[6] V. Polizzi, R. Hewitt, J. Hidalgo-Carrió, J. Delaune, and D. Scaramuzza, "Data-efficient collaborative decentralized thermal-inertial odometry," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 681–10 688, 2022.

[7] C. Chu and S. Yang, "Keyframe-based rgb-d visual-inertial odometry and camera extrinsic calibration using extended kalman filter," *IEEE Sensors Journal*, vol. 20, pp. 6130–6138, 2020.

[8] C. Ryan, A. Elrasad, W. Shariff, J. Lemley, P. Kielty, P. A. Hurney, and P. Corcoran, "Real-time multi-task facial analytics with event cameras," *IEEE Access*, vol. 11, pp. 76 964–76 976, 2023.

[9] T. Liu, B. Li, G. Chen, L. Yang, J. Qiao, and W. Chen, "Tightly coupled integration of gnss/uwb/vio for reliable and seamless positioning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, pp. 2116–2128, 2024.

[10] Z. Liu, D. Shi, R. Li, and S. Yang, "Esvio: event-based stereo visual-inertial odometry," *Sensors*, vol. 23, no. 4, p. 1998, 2023.

[11] P. S. Chib and P. Singh, "Recent advancements in end-to-end autonomous driving using deep learning: A survey," *IEEE Transactions on Intelligent Vehicles*, 2023.

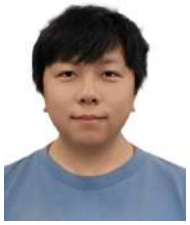
[12] M. Á. Sotelo, R. P. Garcia, I. Parra, D. Fernández, M. Gavilán, S. Alvarez, and J. E. Naranjo, "Visual odometry for road vehicles—feasibility analysis," *Journal of Zhejiang University-SCIENCE A*, vol. 8, pp. 2017–2020, 2007.

[13] K. Xiao, P. Li, G. Wang, Z. Li, Y. Chen, Y. Xie, and Y. Fang, "A preliminary research on space situational awareness based on event cameras," in *2022 13th International Conference on Mechanical and Aerospace Engineering (ICMAE)*. IEEE, 2022, pp. 390–395.

[14] J. Ribeiro-Gomes, J. Gaspar, and A. Bernardino, "Event-based feature tracking in a visual inertial odometry framework," *Frontiers in Robotics and AI*, vol. 10, p. 994488, 2023.

[15] M. Planamente, C. Plizzari, M. Cannici, M. Ciccone, F. Strada, A. Bottino, M. Matteucci, and B. Caputo, "Da4event: towards bridging the sim-

to-real gap for event cameras using domain adaptation," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6616–6623, 2021.
[16] J. P. Rodríguez-Gómez, R. Tapia, A. G. Eguíluz, J. R. M. de Dios, A. Ollero, and G. Meeting, "Uav human teleoperation using event-based and frame-based cameras," *2021 Aerial Robotic Systems Physically Interacting with the Environment (AIRPHARO)*, pp. 1–5, 2021.
[17] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis *et al.*, "Event-based vision: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 1, pp. 154–180, 2020.
[18] X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, and R. B. Benosman, "Hots: a hierarchy of event-based time-surfaces for pattern recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 7, pp. 1346–1359, 2016.
[19] B. Ramachandra, P. Nawathe, J. Monroe, K. Han, Y. Ham, and R. R. Vatsavai, "Real-time energy audit of built environments: Simultaneous localization and thermal mapping," *Journal of Infrastructure Systems*, vol. 24, no. 3, p. 04018013, 2018.
[20] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol. 23, no. 1, pp. 3–20, 2006.
[21] Y. Cheng, M. Maimone, and L. Matthies, "Visual odometry on the mars exploration rovers," in *2005 IEEE International Conference on Systems, Man and Cybernetics*, vol. 1. IEEE, 2005, pp. 903–910.
[22] J. Jeong, Y. Cho, Y.-S. Shin, H. Roh, and A. Kim, "Complex urban dataset with multi-level sensors from highly diverse urban environments," *The International Journal of Robotics Research*, vol. 38, no. 6, pp. 642–657, 2019.
[23] J. Lin, C. Zheng, W. Xu, and F. Zhang, "R3LIVE: A robust, real-time, RGB-colored, LiDAR-Inertial-Visual tightly-coupled state Estimation and mapping package," in *Proceedings of the 2022 International Conference on Robotics and Automation (ICRA)*, Philadelphia, PA, USA, 2022, pp. 10672–10678.
[24] L. Luo, F. Peng, and L. Dong, "Improved Multi-Sensor Fusion Dynamic Odometry Based on Neural Networks," *Sensors*, vol. 24, no. 19, p. 6193, 2024.
[25] C. Jiang, Y. Xu, S. Zhang, Y. Gu, and X. Chen, "Fusion of Visual-Inertial Odometry and UWB for Robust Indoor Localization," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–13, 2023.
[26] T. Qin, S. Cao, J. Pan, and S. Shen, "VINS-Fusion: A Tightly-Coupled Visual-Inertial-GPS Fusion System for Robust and Accurate Localization," *IEEE Transactions on Robotics*, vol. 40, no. 1, pp. 258–274, 2024.
[27] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
[28] T. Zhang, L. Hu, Y. Sun, L. Li, and D. Navarro-Alarcon, "Computing thermal point clouds by fusing rgb-d and infrared images: From dense object reconstruction to environment mapping," in *2022 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2022, pp. 1707–1714.
[29] A. Censi and D. Scaramuzza, "Low-latency event-based visual odometry," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 703–710.
[30] A. Zihao Zhu, N. Atanasov, and K. Daniilidis, "Event-based visual inertial odometry," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5391–5399.
[31] H. Rebecq, T. Horstschaefer, and D. Scaramuzza, "Real-time visual-inertial odometry for event cameras using keyframe-based nonlinear optimization," 2017.
[32] Y. Zhou, G. Gallego, and S. Shen, "Event-based stereo visual odometry," *IEEE Transactions on Robotics*, vol. 37, no. 5, pp. 1433–1450, 2021.
[33] Y. Wang, J. Yang, X. Peng, P. Wu, L. Gao, K. Huang, J. Chen, and L. Kneip, "Visual odometry with an event camera using continuous ray warping and volumetric contrast maximization," *Sensors*, vol. 22, no. 15, p. 5687, 2022.
[34] D. Liu, A. Parra, and T.-J. Chin, "Spatiotemporal registration for event-based visual odometry," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 4937–4946.
[35] A. R. Vidal, H. Rebecq, T. Horstschaefer, and D. Scaramuzza, "Ultimate slam? combining events, images, and imu for robust visual slam in hdr and high-speed scenarios," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 994–1001, 2018.
[36] W. Chamorro, J. Solà, and J. Andrade-Cetto, "Event-IMU Fusion Strategies for Faster-Than-IMU Estimation Throughput," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, pp. 3975–3984.
[37] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam," *The International Journal of Robotics Research*, vol. 36, no. 2, pp. 142–149, 2017.



Gu Gong received the M.Sc. degree in control science and technology from the Harbin Institute of Technology (HIT), Harbin, China, in 2019. He is pursuing a Ph.D. in mechanical engineering at The Hong Kong Polytechnic University and a Ph.D. in control science and engineering at the HIT. His research interests focus on event-based vision, multi-sensor fusion, and SLAM.

Fuji Hu received his M.Sc. degree in mechanical engineering from The Hong Kong Polytechnic University, Hong Kong, in 2024, and he is currently pursuing the Ph.D. degree in information and construction technology at the same university. His research interest includes multi-sensor fusion, teleoperation, and robotic control.

Fangyuan Wang received the M.Sc. degree in software engineering from Zhejiang Sci-Tech University, China, in 2022. He is currently pursuing the Ph.D. in mechanical engineering at The Hong Kong Polytechnic University, Hong Kong. His research interests focus on reinforcement learning, multi-agent systems, and robotic manipulation.

Muhammad Muddassir received a Bachelor of Science degree in Electrical and Electronics Engineering from the National University of Computer and Emerging Sciences, Karachi, Pakistan, in 2015 and a Master of Science degree in Control Science and Technology from Beijing Institute of Technology, Beijing, China, in 2017. He earned his Ph.D. in Mechanical Engineering from The Hong Kong Polytechnic University in 2022. Since 2022, he has been a Postdoctoral Fellow at the Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong. His research interests include robotics, skin biophysics, intelligent control systems, and automation.

Peng Zhou received his Ph.D. degree in robotics from PolyU, Hong Kong, in 2022. In 2021, he was a visiting Ph.D. student at KTH Royal Institute of Technology, Stockholm, Sweden. He is currently a Research Officer at the Centre for Transformative Garment Production and a Postdoctoral Research Fellow at The University of Hong Kong. His research interests include deformable object manipulation, robot reasoning and learning, and task and motion planning.

Lu Li received her Diploma and her Ph.D. degrees from Hefei University of Technology, Hefei, China, in 2003 and 2010 respectively, both in Mechanical Engineering. From 2008 to 2009, she was a visiting Ph.D. student at Stuttgart University, Germany. She is currently a research fellow at Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, China. Her research interests include design and control of robotic systems, legged robots, and man-machine interactive systems.

Qiang Wang (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in control science and engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 1998, 2000, and 2004, respectively. Since 2008, he has been a professor at the Department of Control Science and Engineering, HIT. His research interests include hyperspectral image denoising, signal/image processing, multi-sensor data fusion, wireless sensor networks, and intelligent detection technology.

Zhen He received the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2000. From 1997 to 1998, she was a visiting Ph.D. student at the Mita Laboratory of Tokyo Institute of Technology, Japan.

She has been with Harbin Institute of Technology since 2000 and is currently a Professor with the Department of Control Science and Engineering. Her research interests include robust control, optimal control, H-infinity control, and generalized systems.

David Navarro-Alarcon (Senior Member, IEEE) received the Ph.D. degree in mechanical and automation engineering from The Chinese University of Hong Kong (CUHK), in 2014. From 2014 to 2017, he worked as a Postdoctoral Fellow and then as a Research Assistant Professor at the T Stone Robotics Institute of CUHK. Since 2017, he has been with The Hong Kong Polytechnic University (PolyU), where he is currently an Associate Professor in the Department of Mechanical Engineering, and the Principal Investigator of the Robotics and Machine Intelligence Laboratory (ROMI-Lab). His current research interests include perceptual robotics and control theory. Dr. Navarro-Alarcon currently serves as an Associate Editor of IEEE TRANSACTIONS ON ROBOTICS (T-RO).