Strong Robust LIO System Based on Event Camera Assistance

Xueli Guo[1], Xuanxuan Zhang[1], Zhichao Wen[1], Tianxiang Zhang[1], Tingting Lei[1,2], You Li[1]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing,
Wuhan University, Wuhan, China - (guoxueli, xuanxuanzhang,zhichaowen, cyberkona, liyou)@whu.edu.cn
[2] TUM School of Engineering and Design, Technical University of Munich, Munich, Germany - tingting.lei@tum.de

Keywords: EVIO, LIO, Event Camera, Degenerate, Navigation.

Abstract

At present, the LiDAR system can achieve high precision and is extensively used for indoor and outdoor mobile positioning and mapping. However, LiDAR systems still face issues in cluttered environments where strong features are absent, leading to a degradation of the LiDAR-based solution. When the carrier movement involves high-speed or prolonged exposure to the mirror wall, it can cause severe degradation issues or even positioning failures in the laser slam system. Event cameras are vision sensors inspired by biology that exhibit strong robustness in high dynamic and low texture environments, potentially leading to better performance in such environments. However, some issues with event cameras remain unresolved. In this paper, a multi-source fusion method based on EVIO, LIO and IMU trajectory layer post-processing method is proposed, which will fully consider the robustness of event camera in high dynamic environment and the high precision advantage of Lidar in conventional environment, and use an algorithm based on normalized uncertainty. The elastic multi-source fusion of event camera and LiDAR is realized and tested in realde environment. Experimental results show that the proposed algorithm can effectively improve the accuracy of event camera and LiDAR. Compared to the current more advanced algorithms, the proposed algorithm effectively addresses the problem of LIO in degraded environments. Additionally, it mitigates the scale inaccuracy and divergence of the EVIO trajectory to some extent. Compared to LIO, the algorithm can reduce the maximum position error by approximately 30% and increase the overall position accuracy by 32%. Additionally, it can significantly constrain the divergence of errors in the Y direction, improving its accuracy by about 75% and 65% compared to the LIO and EVIO algorithms, respectively.

1. Introduction

As an emerging field, simultaneous localization and mapping (SLAM) technology can utilize monocular, stereo, RGB-D cameras and Light Detection and Ranging (LiDAR) to estimate the the self movement. It has captured the research community's attention in robotics, autonomous driving, and computer vision. Numerous academics have conducted extensive research on this subject. Currently, LiDAR SLAM is the most widely used and mature SLAM system due to its cost-effectiveness, lightweight design, and high performance (Binas et al., 2017).

However, there are still some unresolved issues with LiDAR SLAM. LiDAR measurements typically rely on geometrical structures in the environment to identify feature points. However, these methods often fail in structure-less environments, such as long corridors or flat open fields (Shan et al., n.d.b). In cluttered environments where no strong features are present, the LiDAR-based solution can easily degenerate. This phenomenon becomes more severe when the carrier is in high-speed motion or faces the mirror wall for an extended period, leading to significant degradation issues or even positioning failures in the LiDAR SLAM system. Therefore, it is an urgent problem to improve the performance of LiDAR system in high dynamic and low texture environment.

Therefore, in order to completely solve the robustness and improve the accuracy of SLAM navigation algorithm in complex lighting, high dynamic, low texture environment and other challenging environments, a more robust sensor is needed.

Event cameras, such as the dynamic vision sensor (DVS) (Lichtsteiner et al., 2008, Binas et al., 2017), are biologically inspired vision sensors that offer several advantages over in high dynamic and low light environment. For example, event cameras are immune to motion blur due to the independence of individual pixels (Rebecq et al., n.d.), and they can transmit data at a speed of 1 MHz with a dynamic range of 140 dB (Mahlknecht et al., 2022), which effectively reduces motion blur and enables the device to work in challenging conditions such as HDR, low light, and fast motion scenarios (Vidal et al., 2018, Sun et al., 2021).

Their pixels function independently and transmit information (known as "events") only when the scene changes in brightness. Their output shows significant differences compared to the intensity frames generated by conventional cameras. Event cameras offer a continuous stream of asynchronous events instead of intensity frames (Gehrig et al., 2020), as demonstrated in Figure 1. These events exclusively communicate valuable details concerning local pixel-level brightness alterations (Kong and Fang, 2021), while each pixel of the event camera functions independently. Due to their capability to avoid the impact of illumination and fast motion compared to conventional cameras, event cameras have recently received increased attention.
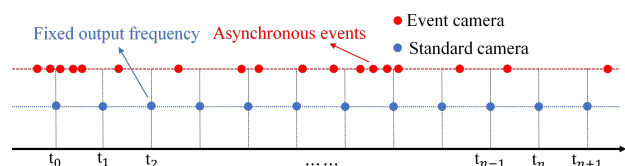

Figure 1. Schematic diagram of event camera and standard camera imaging.

However, event camera also has some drawbacks and problems that are difficult to solve. Current event camera have low resolution and produce high noise. Besides, integrating event cameras into SLAM is challenging due to the asynchronous nature of event streams that fundamentally differ from synchronous intensity images (Guan and Lu, n.d.). The main SLAM algorithm that employs event cameras uses the original polarity data of the camera, based on time surface (TS), and accumulates images based on time or number of event points. However, due to the binary imaging mode of the TS-based event camera, it can only present edge information while disregarding internal texture features. Compared with LiDAR-inertial odometry (LIO) algorithm, this will lead to the event camera has a certain precision limitation in normal environment, and has the disadvantages of low precision and high cost.

Therefore, it is crucial to integrate event cameras and LiDAR systems to achieve a robust slam system under multiple motion conditions. In this paper, the integrated algorithm of two sensors, LiDAR and event camera, is proposed for the first time, and its accuracy is verified by collecting multiple sets of data under natural scenes. A plug-and-play extended Kalman filter algorithm for event camera and LiDAR systems based on adaptive weights of normalized uncertainty is proposed. The algorithm consists of three parts, 1: event camera estimation system based on spatio-temporal TS and back-end optimization; 2: LiDAR SLAM system based on EKF; 3: an adaptive elastic weighted event camera and LiDAR multi-source system framework based on normalized uncertainty. The contributions of this research can be summarized in three key areas. First, the data from event camera and LiDAR were collected for event-visual-inertial odometry (EVIO) and LIO in various environments, under different dynamics and illumination conditions, within a wide range of natural scenes. Secondly, the LIO degradation detection scheme based on LIO normalized uncertainty is implemented, and the elastic weight determination scheme of LIO and EVIO is implemented based on this. Finally, a plug-and-play multi-source fusion algorithm framework based on EVIO and LIO trajectory post-processing based on EKF is implemented.

The main structure of this paper is as follows. Section II discusses the development of LIO algorithms and event-based visual algorithms. Section III introduces the TS-based EVIO algorithm of the event camera and the LIO algorithm framework, as well as the plug-and-play multi-source fusion algorithm framework based on EVIO and LIO trajectory post-processing based on EKF. Section IV presents the experimental component of this study. The performance of EVIO and LIO multi-source fusion algorithms based on EFK is demonstrated in detail using data collected from self-developed equipment. The conclusions of this experimentation are then outlined in Section V.

2. Related Works

This section will provide a detailed overview of the research progress related to event camera, LiDAR SLAM, and LiDAR SLAM algorithm in the field of anti-degradation.

2.1 Event Camera

Event cameras show significant differences compared to the data produced by conventional cameras. Event cameras provide a continuous stream of asynchronous events instead of absolute intensity frames (Gehrig et al., 2020). Each event point of the event camera works independently, and these event points only transmit information related to the relative brightness change of local pixels. A large number of asynchronous event streams containing event information are output, which represents the light change of the pixel point in the logarithmic field (Fu et al., 2023). Therefore, compared with traditional cameras, event camera, which can avoid the influence of light and motion blur, has been paid more and more attention by researchers in recent years.

Serrano-Gotarredona et al. (Serrano-Gotarredona et al., 1999) introduced the event cameras firstly in 1999, and the related products and the principle of event cameras were systematically introduced by the University of Zurich team (Lichtsteiner et al., 2008). The team from Technical University of Munich has researched event cameras utilizing particle filtering algorithms (Weikersdorfer and Conradt, n.d.). However, it is essential to note that this study primarily focuses on planar motion, and its scope is limited. Scaramuzza's team (Censi and Scaramuzza, n.d.) introduced the first visual odometry system that combines a DVS with a standard CMOS camera to provide absolute brightness values, while translation is challenging due to the sparse event generation caused by minimal apparent motion. In 2016, Kueng et al. (Kueng et al., n.d.) proposed a method to estimate camera motion using both intensity frames from traditional cameras and event stream data from event cameras, based on the event-based feature tracking algorithm. Although this method utilizes the high frame rate of the event camera to track fast-moving feature points, it still relies on the traditional intensity image frame in the feature extraction stage. Therefore, it is not possible to completely eliminate the impact of motion blur on image feature extraction. In 2018, the Scaramuzza team released the Event+Frame+IMU algorithm (Vidal et al., 2018, Rebecq et al., n.d.) Ultimate SLAM. This algorithm fused event stream data, standard frame data, and inertial measurement data in a tightly coupled manner for the first time. However, the system is required to remain static and it has not been tested on large-scale outdoor data. This paper reveals serious scale and operational efficiency issues during actual large-scale scene tests. Guan et al. (Guan and Lu, n.d.) conducted a study of EVIO in a large environment for the first time. However, the author also acknowledges the limitations of event cameras in low-texture environments and suggests improving global robustness and accuracy through multi-source fusion with sensors such as lidar.

2.2 LiDAR SLAM

The LiDAR SLAM can be divided into pure LiDAR odometer (LO) and LIO, depending on whether it is combined with IMU. LiDAR can directly measure distance and provide precise spatial position and shape information of objects, allowing for the construction of high-precision maps that improve the accuracy and stability of SLAM systems over long periods of time. Therefore, LiDAR SLAM is widely used in indoor navigation, 3D reconstruction, and autonomous driving.

The LOAM (Lidar Odometry and Mapping) algorithm, proposed by Zhang in 2014 (Shan and Englot, n.d.), is a pure lidar odometry algorithm that has inspired subsequent laser SLAM algorithms. The algorithm utilizes the feature point method to compute the curvature of points within a local range. It classifies the feature points into two categories: edge feature points and plane feature points based on the magnitude of curvature. The algorithm assumes uniform motion of LiDAR and compensates for the motion using the linear interpolation method.

However, the algorithm may drift in large-scale scenarios and scenarios with many loop-closure detection due to the lack of back-end optimization and loop-closure detection module, resulting in a significant reduction in accuracy. In 2018, Shan et al. added the loopback detection function to the open-source LeGO-LOAM (Shan et al., n.d.a) algorithm, which is based on LOAM. The algorithm introduced the concept of key frames and used them, along with their local data frames, to form Loop-Submap. This greatly reduced the amount of computation and filtered redundant data. The algorithm detects loop-closure by matching the current data frame with the Loop-Submap. However, the algorithm places high demands on the ground environment. Although the author fused the IMU sensor with loose coupling in the algorithm later, the improvement is limited compared with the tight coupling algorithm. The LIO-SAM (Wisth et al., 2021) algorithm is a classical tightly coupled laser-inertial odometer based on smoothing and mapping. The algorithm's odometer component builds upon the LOAM concept, utilizing IMU data to rectify point cloud distortion and establish the initial pose transformation value between data frames. At the back end, a factor graph is used to optimize the architecture, eliminate cumulative errors, and perform global optimization. However, the LIO-SAM algorithm is based on the feature point method, which has limitations inherent to the LiDAR sensor. As a result, the algorithm may fail in degenerate or unstructured scenes. The FAST-LIO series LIO algorithm (Xu and Zhang, 2021) utilizes an iterative extended Kalman filter that is tightly coupled to fuse lidar feature points with IMU data. This approach achieves robust navigation in fast-moving, noisy, or cluttered environments.

2.3 LIO anti-degradation

In order to address the issue of serious degradation of LIO in challenging scenarios, some scholars have conducted research on the degradation problem of LiDAR by introducing new sensors. Currently, multi-source sensor fusion that combines camera and LiDAR is a popular research topic. This is because the camera can gather abundant semantic information and compensate for the limitations of LiDAR (Lin et al., 2021).

LVI-SAM (Shan et al., n.d.b) consists of a Lidar-Inertial System (LIS) and a Visual-Inertial System (VIS). LIS provides accurate depth information and improves the accuracy of VIS. LIS uses the initial pose estimation of VIS for scan matching. The algorithm's advantage lies in its ability to function normally even if one of VIS or LIS fails, thus making LVI-SAM highly robust in environments with few textures and lack of features. Lin et al. (Lin et al., 2021) from the University of Hong Kong propose R2LIVE, which utilizes LIO and error state iterative Kalman filtering to fuse measurements from three sensors for state estimation. The system is further optimized by a factor graph to improve overall accuracy. However, the vision system adopts the feature point method and may fail in unstructured environments. Zheng et al. (Zheng et al., n.d.) proposed the tight coupling laser-inertial-vision odometry (LIVO) algorithm FAST-LIVO, which couples the two sensors at the measurement level. This algorithm can still operate reliably in scenes with LiDAR degradation and drastic changes in light. However, in scenes of intense movement, the fusion of fuzzy images does not improve the accuracy of the odometer. Its advantages are better reflected in low-speed movement scenes.

Therefore, combining the event camera sensor, which has strong robustness in high dynamic and low light environments,

with the LiDAR sensor, which has high accuracy in normal motion scenes, can effectively enhance the robustness and position accuracy of the two sensors in different scenarios.

3. Methodology

In this section, the event camera, LiDAR and IMU multisource fusion integrated navigation algorithmic framework will be demonstrated in detail. The framework encompasses three key components: LiDAR inertial odometer part, event camera inertial odometer part, and EKF-based LiDAR and event camera track loose combined post-processing part.

3.1 Back-end error modeling of LIO algorithm based on filter

The LiDAR odometer part mainly refers to the FAST-LIO algorithm. The algorithm consists of two parts, the front-end and the back-end optimization part. The front-end mainly includes the pre-processing part of LiDAR data frame matching (scan-to-map), and the back-end of the algorithm is the error modeling part based on IMU and LiDAR. Here is a brief introduction to LIO back-end error modeling.

Assume that the LiDAR residual term is $\mathbf{r}_l(\cdot)$, a point in the point cloud frame observed under the LiDAR system is $^L\mathbf{p}_j$, and the current updated state is $\overset{\vee}{\mathbf{x}}_{k+1}$, then the form of the residual term is as follows:

$$\mathbf{r}_l(\overset{\vee}{\mathbf{x}}_{k+1}, {}^L\mathbf{p}_j) = \mathbf{u}_j^\top ({}^G\mathbf{p}_j - \mathbf{q}_j) \qquad (1)$$

According to the current pose estimate $\overset{\vee}{\mathbf{x}}_{k+1}$, the LiDAR system point $^L\mathbf{p}_j$ can be transferred to the world system to get $^G\mathbf{p}_j$, and then the neighboring point cloud cluster of $^G\mathbf{p}_j$ in the point cloud map can be searched to fit the plane, thus obtaining the normal vector $\mathbf{u}_j^\top$ of the fitting plane and $\mathbf{q}_j$ of a certain point on the fitting plane.

The noise contained in the LiDAR observation is mainly affected by the absolute ranging accuracy of the LiDAR point, and for a certain LiDAR point $^L\mathbf{p}_j$ in the LiDAR frame, it has

$$^L\mathbf{p}_j = {}^L\mathbf{p}_j^{\mathbf{gt}} + \mathbf{n}_j \qquad (2)$$

$$\mathbf{n}_j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{n}_j}) \qquad (3)$$

where $^L\mathbf{p}_j^{\mathbf{gt}}$ represents the true value of the point and $\mathbf{n}_j$ represents the observed noise of the point. When in the real state, the residual term constructed by the LiDAR point truth value is zero, and the residual term can be linearly expanded as follows:

$$\mathbf{0} = \mathbf{r}_l(\mathbf{x}_{k+1}, {}^L\mathbf{p}_j^{\mathbf{gt}}) = \mathbf{r}_l(\overset{\vee}{\mathbf{x}}_{k+1}, {}^L\mathbf{p}_j) + \mathbf{H}_j^l \delta\overset{\vee}{\mathbf{x}}_{k+1} + \alpha_j \quad (4)$$

$$\mathbf{H}_j^l = \frac{\partial\mathbf{r}_l(\overset{\vee}{\mathbf{x}}_{k+1}\,\delta\overset{\vee}{\mathbf{x}}_{k+1}, {}^L\mathbf{p}_j)}{\partial\delta\overset{\vee}{\mathbf{x}}_{k+1}}\bigg|_{\delta\overset{\vee}{\mathbf{x}}_{k+1}=\mathbf{0}} \qquad (5)$$

$$\alpha_j = \mathbf{F}_{\mathbf{p}_j}\mathbf{n}_j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\alpha_j}) \qquad (6)$$

$$\boldsymbol{\Sigma}_{\alpha_j} = \mathbf{F}_{\mathbf{p}_j}\boldsymbol{\Sigma}_{\mathbf{n}_j}\mathbf{F}_{\mathbf{p}_j}^\top \qquad (7)$$

$$\mathbf{F}_{\mathbf{p}_j} = \left(\frac{\partial\mathbf{r}_l(\overset{\vee}{\mathbf{x}}_{k+1}, {}^L\mathbf{p}_j)}{\partial^L\mathbf{p}_j}\right) = {}^G\mathbf{R}_{I_{k+1}}\,{}^i\mathbf{R}_L \qquad (8)$$

For the LiDAR residual measurement matrix $\mathbf{H}_j^l$, the form is as follows:

$$\mathbf{H}_j^l = \mathbf{u}_j^\top \left[-{}^G\overset{\vee}{\mathbf{R}}_{I_{k+1}} \left[{}^I\mathbf{R}_L\, {}^L\mathbf{p}_j + {}^I\mathbf{p}_L \right]_\times \quad \mathbf{I} \quad \mathbf{0}_{3\times 15} \right] \quad (9)$$

$$\mathbf{r}_l(\mathbf{x}_{k+1}^\vee,\, {}^L\mathbf{p}_j) + \mathbf{H}_j^l \delta\mathbf{x}_{k+1}^\vee \sim \mathcal{N}(\mathbf{0},\, \mathbf{F}_{\mathbf{p}_j}\mathbf{\Sigma}_{\mathbf{n}_j}\mathbf{F}_{\mathbf{p}_j}^\top) \quad (10)$$

Thus, the residual term for a single LiDAR point of the LiDAR frame can be obtained as described in the above equation 10.

3.2 EVIO algorithm based on TS and graph optimization

3.2.1 Reprojection error: Referring to the USLAM algorithm, we use the time surface-based event camera image reconstruction algorithm to convert the event stream data into a frame format that can be used by the visual VIO algorithm. Besides, the FAST corner detection method was employed to identify its features. Subsequently, the Lukas-Kanade optical flow algorithm was used to track these detected feature points continuously. The projection process can be effectively modeled as

$$\tilde{\mathcal{P}} = \pi_c(\mathbf{R}_b^c(\mathbf{R}_w^b L^w + \mathbf{p}_w^b) + \mathbf{p}_b^c) + \mathbf{n}_c \quad (11)$$

where $\mathcal{P} = [u,v]^T$ is the coordinate of the feature point on the pixel plane coordinate system and L^w represents the 3D landmark position in the local world frame. $\pi_c(\cdot)$ represents the camera projection function and $\mathbf{n}_c$ is the measurement noise. Therefore, the reprojection error linking these two frames can be formulated as follows: if a feature represented by a landmark l has an inverse depth of ρ_l in frame i, and this feature is subsequently observed again in frame j, then the reprojection error is

$$\mathbf{r}_\mathcal{C}(\tilde{Z}_l, \mathcal{X}) = \tilde{\mathcal{P}}_l^{c_{t_j}} - \pi_c(\hat{X}_l^{c_{t_j}}) \quad (12)$$

$$\hat{X}_l^{c_{t_j}} = \mathbf{R}_b^c(\mathbf{R}_w^{b_{t_j}}(\mathbf{R}_{b_{t_i}}^w(\mathbf{R}_c^b \frac{1}{\rho_l}\pi_c^{-1}(\tilde{\mathcal{P}}_l^{c_{t_i}}) + \mathbf{p}_c^b) + \mathbf{p}_{b_{t_i}}^w) + \mathbf{p}_w^{b_{t_j}}) + \mathbf{p}_b^c \quad (13)$$

where $\{\mathbf{R}_c^b, \mathbf{t}_c^b\}$ are the extrinsic parameters between the event camera and IMU.

3.2.2 IMU pre-integration: The IMU pre-integration technique (Lupton and Sukkarieh, 2011) has widely been used in VIO systems. This algorithm removes the need to re-integrate state dynamics at each optimization step. An IMU can measure specific force and angular rate in the IMU's body frame. The specific force ${}^b\hat{a}$ and angular rate ${}^b\hat{\omega}$ can be calculated as

$$ {}^b\hat{a} = {}^b a + R_w^b\, {}^w g + {}^b b_a + {}^b n_a \quad (14)$$

$$ {}^b\hat{\omega} = {}^b\omega + {}^b b_g + {}^b n_g \quad (15)$$

where ${}^b b_a$, ${}^b b_g$, ${}^b n_a$ and ${}^b n_g$ are the biases and noises from the accelerometer and the gyroscope in the IMU frame, respectively, g is the gravity vector in the world frame.

We assume that the additional noise in the accelerometer and gyroscope readings is Gaussian white noise (Qin et al., 2018), $n_a(0, \sigma_a^2)$ and $n_w(0, \sigma_w^2)$. Additionally, the acceleration bias and gyroscope bias are modeled as a random walk, whose derivatives are also Gaussian white noise and can be expressed as

$$\dot{b}_{a_t} = n_{b_a}(0, \sigma_{b_a}^2) \quad (16)$$

$$\dot{b}_{w_t} = n_{b_w}(0, \sigma_{b_w}^2) \quad (17)$$

The IMU measurement data at intervals $[t_k, t_{k+1}]$ was collected between two successive event frames b_k and b_{k+1}. While simultaneously accounting for the influence of IMU bias, this data was integrated into the local frame b_k as

$$\alpha_{b_{k+1}}^{b_k} = \iint_{t\in[t_k,t_{k+1}]} R_t^{b_k}(\hat{a}_t - b_{a_t})\, dt^2 \quad (18)$$

$$\beta_{b_{k+1}}^{b_k} = \int_{t\in[t_k,t_{k+1}]} R_t^{b_k}(\hat{a}_t - b_{a_t})\, dt \quad (19)$$

$$\gamma_{b_{k+1}}^{b_k} = \int_{t\in[t_k,t_{k+1}]} \frac{1}{2}\Omega(\hat{\omega}_t - b_{w_t})\gamma_t^{b_k}\, dt \quad (20)$$

where

$$\Omega(\omega) = \begin{bmatrix} -[\omega]_\times & \omega \\ -\omega^T & 0 \end{bmatrix} \quad [\omega]_\times = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \quad (21)$$

where $\{\alpha, \beta, \gamma\}$ represents the relative position, velocity, and rotation information between frames b_k and b_{k+1}. Finally, the relationship between the residual and the system state and the pre-integrated IMU measurements can be expressed as

$$\mathbf{r}_{pre}(\tilde{Z}_{b_{t_{k+1}}}^{b_{t_k}}, \mathcal{X}) = \begin{bmatrix} \delta\alpha_{b_{t_{k+1}}}^{b_{t_k}} \\ \delta\beta_{b_{t_{k+1}}}^{b_{t_k}} \\ \delta\gamma_{b_{t_{k+1}}}^{b_{t_k}} \\ \delta\mathbf{b}_a \\ \delta\mathbf{b}_g \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{R}_w^{b_{t_k}}(\mathbf{p}_{b_{t_{k+1}}}^w - \mathbf{p}_{b_{t_k}}^w + \frac{1}{2}\mathbf{g}^w \Delta t_k^2 - \mathbf{v}_{b_{t_k}}^w \Delta t_k) - \hat{\alpha}_{b_{t_{k+1}}}^{b_{t_k}} \\ \mathbf{R}_w^{b_{t_k}}(\mathbf{v}_{b_{t_{k+1}}}^w + \mathbf{g}^w \Delta t_k - \mathbf{v}_{b_{t_k}}^w) - \hat{\beta}_{b_{t_{k+1}}}^{b_{t_k}} \\ 2\left[\mathbf{q}_{b_{t_k}}^{w^{-1}} \otimes \mathbf{q}_{b_{t_{k+1}}}^w \otimes \left(\hat{\gamma}_{b_{t_{k+1}}}^{b_{t_k}}\right)^{-1} \right]_v \\ \mathbf{b}_{ab_{t_{k+1}}} - \mathbf{b}_{ab_{t_k}} \\ \mathbf{b}_{wb_{t_{k+1}}} - \mathbf{b}_{wb_{t_k}} \end{bmatrix} \quad (22)$$

where $\delta\alpha_{b_{t_{k+1}}}^{b_{t_k}}$, $\delta\beta_{b_{t_{k+1}}}^{b_{t_k}}$ and $\delta\gamma_{b_{t_{k+1}}}^{b_{t_k}}$ represent the relative error in 3D euclidean space and $[]_v$ represents the imaginary component of the quaternion.

3.2.3 Graph optimization based on event cameras and IMUs: The nonlinear optimization approach was adopted in the EVIO algorithm's back-end, which is widely adopted in current SLAM research (Leutenegger et al., 2013). This method incorporates three parts: visual reprojection error, IMU pre-integration error, and a keyframe marginalization strategy. The problem can be structured as a factor graph, with sensor measurements serving as a sequence of factors that constrain the states of the system (Figure 2).

During graph optimization, the state of the system in the sliding window can be defined as

$$\mathcal{X} = [x_0, x_1, \ldots, x_n, x_c^b, \delta_0, \delta_1, \ldots, \delta_l] \quad (23)$$

$$x_k = [p_{wb_k}^w, \mathbf{q}_{wb_k}^w, v_{wb_k}^w, \mathbf{b}_{g_k}, \mathbf{b}_{a_k}], k\in[0,n] \quad (24)$$
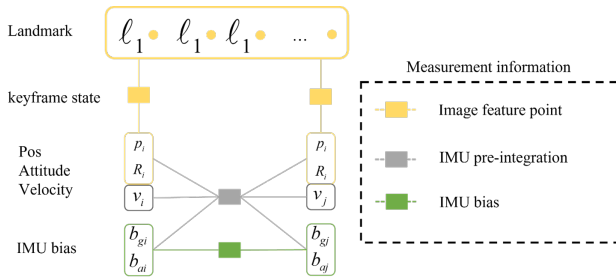
$$x_c^b = [p_{bc}^b, q_c^b] \quad (25)$$

Figure 2. Factor graph model during EVIO.

Figure 3. Algorithmic framework for EVLINS.

where x_k includes the position, attitude, velocity, and IMU bias. Additionally, x_c^b represents the extrinsic parameter between the event camera and the IMU, while δ refers to the inverse depth parameter of the landmark. We formulate the optimization problem as a maximum a posteriori estimation problem, where all observations are independent, and the noise is assumed to follow a zero-mean Gaussian white noise distribution and can be expressed as

$$
\begin{aligned}
\mathcal{X}^\star &= \arg\max_{\mathcal{X}} p(\mathcal{X}|\mathbf{z}) \\
&= \arg\max_{\mathcal{X}} p(\mathcal{X})p(\mathbf{z}|\mathcal{X}) \\
&= \arg\max_{\mathcal{X}} p(X)\prod_{i=1}^{n} p(\mathbf{z}_i|\mathcal{X}) \\
&= \arg\min_{\mathcal{X}} \left\{ \|\mathbf{r}_p - \mathbf{H}_p\mathcal{X}\|^2 + \sum_{i=1}^{n} \|\mathbf{r}(\mathbf{z}_i, \mathcal{X})\|_{\mathbf{P}_i}^2 \right\}
\end{aligned}
\tag{26}
$$

where z represents the combination of n independent sensor measurements, and $[\mathbf{r}_p, \mathbf{H}_p]$ embodies the prior information of the system state. The residual function of each measurement is denoted by $r()$ and $\|\|_P$ denotes the Mahalanobis norm.

Therefore, the maximum a posteriori estimation problem can be formulated as

$$
\begin{aligned}
\min_{\mathcal{X}} \Bigg\{ &\|\mathbf{r}_p - \mathbf{H}_p\mathcal{X}\|^2 + \sum_{k\in[1,n]} \left\|\mathbf{r}_{Pre}\begin{pmatrix} Pre \\ k-1,k \end{pmatrix}, \mathcal{X}\right\|_{\mathbf{\Sigma}_{k-1,k}^{Pre}}^2 \\
&+ \sum_{l\in\mathbf{L}} \left\|\mathbf{r}_V\begin{pmatrix} V_{i,j} \\ l \end{pmatrix}, \mathcal{X}\right\|_{\mathbf{\Sigma}_l^{V_{i,j}}}^2 \Bigg\}
\end{aligned}
\tag{27}
$$

where $\mathbf{r}_{Pre}$ is the residuals of the IMU pre-integration measurements, and $\mathbf{r}_V$ is the residuals of the visual measurements.

3.3 Integrated navigation of LIO+ EVIO based on EKF

The section describes a framework for a plug-and-play multi-source fusion algorithm based on EVIO and LIO trajectory post-processing using EKF. As illustrated in Figure 3, this paper proposes a post-processing algorithm framework. The EKF Kalman filter is used to fuse the processing results of LIO and EVIO, along with the raw data of IMU, achieving multi-source fusion.

It is worth noting that according to previous studies, the EVIO algorithm is highly robust and can function effectively in challenging environments, including those with high dynamics and low light. The LIO algorithm has high accuracy in conventional motion environments, but its pose accuracy fluctuates greatly in high dynamic motion, fast rotation, and low texture environments. However, even if the LIO position error fluctuates
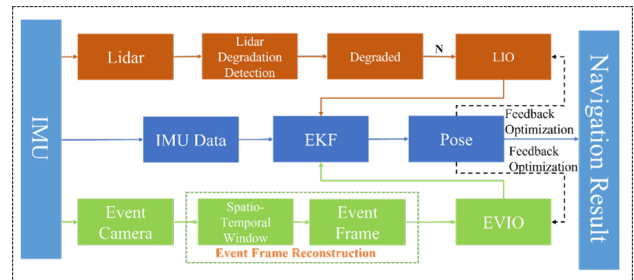
greatly, the absolute value of the position and attitude uncertainty of the output still remains in the range of 10^{-5} to 10^{-4}. Therefore, the error uncertainty is significantly different from the actual position error, and it cannot be used directly to determine whether the LIO trajectory is degraded.

This paper proposes using normalized LIO uncertainty as a criterion to determine whether LIO is degraded. If the normalized uncertainty of LIO exceeds a certain threshold, it is considered that LIO is in a degraded scene. At this time, the EVIO and IMU data are utilized as input for obtaining corresponding pose data through EKF fusion. This data is then used to correct the LIO trajectory and provide feedback for LIO data repair. On the contrary, When the uncertainty of LIO does not exceed a certain threshold, it is considered to be in a non-degraded environment. In this case, the results from EVIO are not used in integrated navigation. Instead, the result data from LIO is used as input data along with IMU data through EKF to obtain corresponding pose data. The feedback correction of the LIO data is a result of repairing the EVIO trajectory.

4. Accuracy evaluation of integrated navigation of LIO+ EVIO based on EKF

For data processing, the LIO results are obtained by the open-source algorithm FAST-LIO (Xu and Zhang, 2021), while the Event Camera algorithm is developed based on the classic EVIO algorithm (Vidal et al., 2018).

For all experiments conducted in this study, the DVXplorer, Intel RealSense, and Livox Avia lidar sensors are the data recording devices (Figure 4). Notably, the DVXplorer has a 640×480 pixels event camera and a 1 kHz IMU, while the RealSense sensor merely provides standard images for display purposes rather than feature tracking.

We tested the performance of our algorithm in an indoor office environment. During the collection process, there will be a variety of dynamics and environments such as facing the wall for a long time, turning quickly, and walking normally. It is used to fully simulate the applicable and challenging environments of event cameras and LiDAR, respectively, and run LIO and EVIO algorithms. In this paper, higher precision and more expensive IMUs, as well as high resolution industrial cameras and LiDAR sensors are used as truth values.

The result of LIO algorithm will output position, attitude and uncertainty information, but its error uncertainty is seriously untrustworthy. As shown in Figure 5, when the trajectory of LIO faces a smooth wall, its trajectory has begun to diverge seriously compared with the true value, but the position uncertainty

Figure 4. The equipment used in this paper.

is only 10^{-4} (Figure 6). If it is directly used as the weight information of LIO and EVIO multi-source fusion, it is obviously irrational and inconsistent with the real situation. Therefore, in order to make reasonable use of the uncertainty information, the three-axis uncertainty information is normalized to accurately describe its variation trend in the error uncertainty interval.
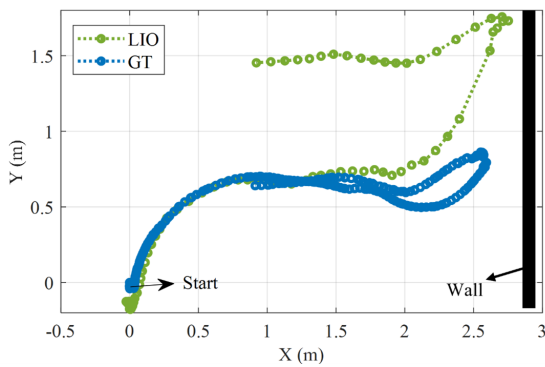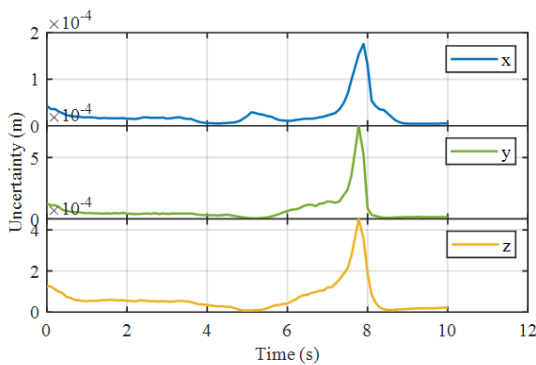

Figure 5. Comparison of LIO and ground truth.

Figure 6. LIO three-dimensional position uncertainty.

Figure 7 displays the normalized uncertainty. Additionally, Figure 8 shows the LIO trajectory along with the part that exceeds the normalized uncertainty threshold used to determine if LIO is degraded. The threshold value is obtained by synthesizing the uncertainty and error of the whole trajectory. Combined

with Figure 5, the divergence trend and divergence of error uncertainty are consistent. Before the movement, the computer screen briefly obstructs the data acquisition equipment, rendering it unable to obtain effective scene structure information, resulting in a brief degenerate state. During the initial stages of movement, LIO can gather detailed structural information from the surroundings, resulting in a trajectory that closely aligns with the actual value. However, between 6s and 8s, while walking towards a deteriorated environment and facing a wall that required a turn, the experimenter abruptly changed direction by 180 degrees. This sudden change caused the LIO trajectory to rapidly diverge, resulting in a brief peak in its normalized uncertainty that exceeded the threshold.
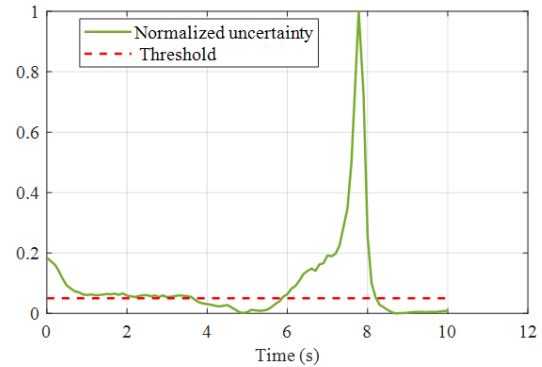

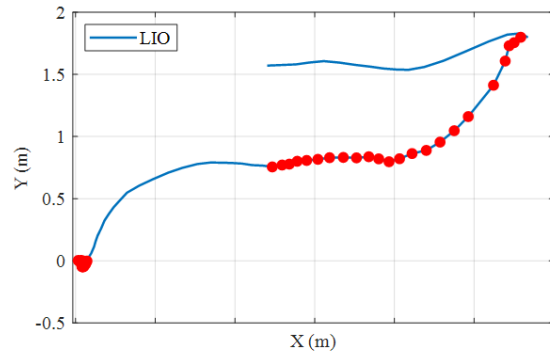Figure 7. LIO position normalization uncertainty and threshold.

Figure 8. Comparison of LIO trajectories and uncertainty.

In contrast, the event camera has good robustness under high-speed motion turns in the same scene due to its feature of event camera without motion blur, and its divergence on the trajectory is smaller than the LIO result (Figure 9). However, event cameras also have scaling issues. This is because event cameras can only generate corresponding event information when the change of light exceeds a certain threshold. Therefore, in the normal motion state, fewer event points are triggered compared to the high dynamic state, resulting in less texture information being formed and a significant scale problem.

Therefore, it is necessary to combine the characteristics of high accuracy of LIO trajectories under normal motion conditions and the characteristics of robustness under EVIO challenge scenarios. This paper proposes EVLINS, a post-processing algorithm framework for fusing LIO and EVIO trajectories based on EKF. The framework includes a method for judging whether
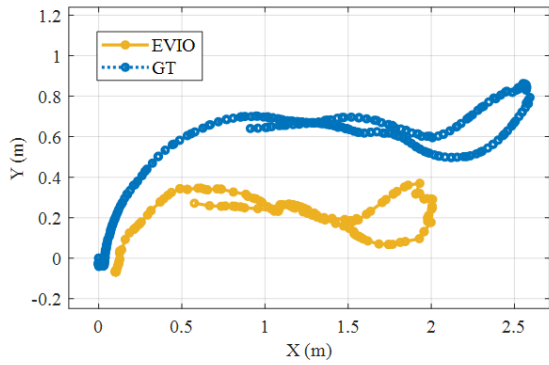
Figure 9. Comparison of EVIO and ground truth.

LIO is degraded by checking if the normalized uncertainty of LIO exceeds a certain threshold. By setting the threshold for LIO normalized uncertainty, LIO and EVIO tracks are fused flexibly. The integrated navigation results, as shown in Figure 3, are used to perform feedback correction on the tracks that do not participate in the fusion. The results are presented in Figure 10. It is evident that the fusion of LIO and EVIO has resulted in a significant improvement in the trajectory. The algorithm proposed in this paper is capable of timely switching to the more robust EVIO system when the LIO is in a degraded environment, and carrying out continuous trajectory repair for the subsequent divergent LIO trajectory. When the system detects that LIO is no longer in a degraded environment, it takes over the entire system again. Compared to the unused integrated navigation, its trajectory is closer to the true value, effectively addressing the problem of LIO in degraded environments. Furthermore, it improves the inaccuracy and divergence of EVIO's trajectory to a certain extent. In comparison to EVIO, it addresses the issue of scale inaccuracy that has been present since the beginning of the movement. It is possible that the entire system is now running more robustly.
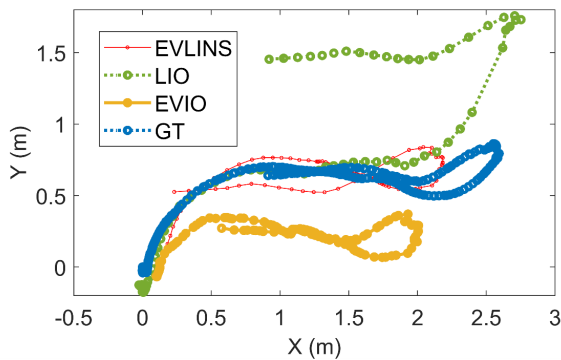

Figure 10. EVLINS integrated navigation results based on EKF.

The positional errors of the three algorithms are calculated separately, as presented in Table 1. It is evident that EVLINS has better position accuracy than the original algorithm. In comparison to LIO, EVLINS can reduce the maximum position error by approximately 30% and increase the overall position accuracy by 32%. However, upon comparing the position errors of the X-axis and Y-axis (refer to Figure 11), it becomes evident that the EVLINS algorithm performs poorly in the X-axis direction due to the significant scale influence of the EVIO algorithm. While the algorithm shows some improvement in

the Y-axis direction when compared to LIO, it effectively constrains the divergence of its error in the Y direction. Compared to the LIO and EVIO algorithms, the EVLINS algorithm can improve two-dimensional position accuracy by approximately 75% and 65%, respectively.

Algorithm	max (m)	rmse (m)	std (m)
EVLINS	0.68	0.35	0.21
LIO	0.98	0.43	0.31
EVIO	0.84	0.43	0.25

Table 1. MargComparison of 2D position error between the EVLINS algorithm and other algorithms.
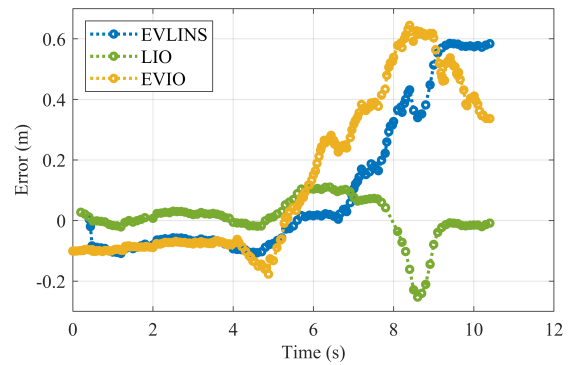

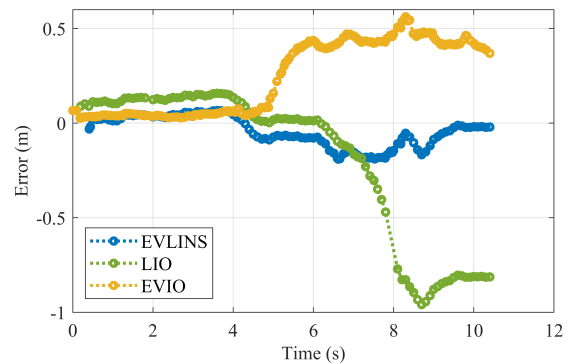Figure 11. X-axis error for LIO, EVIO, and EVLINS algorithms.

Figure 12. Y-axis error for LIO, EVIO, and EVLINS algorithms.

5. Conclusion

This paper presents a framework for a plug-and-play multi-source fusion algorithm based on EVIO and LIO trajectory post-processing using EKF. The aim is to combine LIO's high accuracy in conventional motion environments with EVIO's strong robustness in highly dynamic environments. The EKF Kalman filter is used to fuse the processing results of LIO and EVIO, along with the raw data of IMU, achieving multi-source fusion and provide feedback for the LIO and EVIO. The experimental results indicate that the proposed algorithm effectively addresses the problem of LIO in degraded environments. Additionally, it mitigates the scale inaccuracy and divergence of the EVIO trajectory to some extent. Compared to LIO, the algorithm can reduce the maximum position error by approximately 30% and increase the overall position accuracy by 32%.

Additionally, it can significantly constrain the divergence of errors in the Y direction, improving its accuracy by about 75% and 65% compared to the LIO and EVIO algorithms, respectively.

References

Binas, J., Neil, D., Liu, S.-C., Delbruck, T., 2017. DDD17: End-to-end DAVIS driving dataset. *arXiv preprint arXiv:1711.01458.*

Censi, A., Scaramuzza, D., n.d. Low-latency event-based visual odometry. *2014 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 703–710.

Fu, J., Yu, L., Yang, W., Lu, X., 2023. Event-based continuous optical flow estimation. *Acta Automatica Sinica*, 49(09), 1-12. https://kns.cnki.net/kcms/detail/11.2109.TP.20211130.1600.015.html.

Gehrig, D., Rebecq, H., Gallego, G., Scaramuzza, D., 2020. EKLT: Asynchronous photometric feature tracking using events and frames. *International Journal of Computer Vision*, 128(3), 601-618.

Guan, W., Lu, P., n.d. Monocular event visual inertial odometry based on event-corner using sliding windows graph-based optimization. *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2438–2445.

Kong, D., Fang, Z., 2021. A review of event-based vision sensors and their applications. *Information and Control*, 50(01), 1-19. https://kns.cnki.net/kcms/detail/21.1138.TP.20210218.1206.010.html.

Kueng, B., Mueggler, E., Gallego, G., Scaramuzza, D., n.d. Low-latency visual odometry using event-based feature tracks. *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 16–23.

Leutenegger, S., Furgale, P., Rabaud, V., Chli, M., Konolige, K., Siegwart, R., 2013. Keyframe-based visual-inertial slam using nonlinear optimization. *Proceedings of Robotis Science and Systems (RSS) 2013*.

Lichtsteiner, P., Posch, C., Delbruck, T., 2008. A 128x128 120dB 15s Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits*, 43(2), 566-576.

Lin, J., Zheng, C., Xu, W., Zhang, F., 2021. R^2 LIVE: A Robust, Real-Time, LiDAR-Inertial-Visual Tightly-Coupled State Estimator and Mapping. *IEEE Robotics and Automation Letters*, 6(4), 7469-7476.

Lupton, T., Sukkarieh, S., 2011. Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions. *IEEE Transactions on Robotics*, 28(1), 61-76.

Mahlknecht, F., Gehrig, D., Nash, J., Rockenbauer, F. M., Morrell, B., Delaune, J., Scaramuzza, D., 2022. Exploring event camera-based odometry for planetary robots. *IEEE Robotics and Automation Letters*, 7(4), 8651-8658.

Qin, T., Li, P., Shen, S., 2018. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4), 1004-1020.

Rebecq, H., Horstschaefer, T., Scaramuzza, D., n.d. Real-time visual-inertial odometry for event cameras using keyframe-based nonlinear optimization. *British Machine Vision Conference.*

Serrano-Gotarredona, T., Andreou, A. G., Linares-Barranco, B., 1999. AER image filtering architecture for vision-processing systems. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 46(9), 1064-1071.

Shan, T., Englot, B., n.d. Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 4758–4765.

Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., Rus, D., n.d.a. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE, 5135–5142.

Shan, T., Englot, B., Ratti, C., Rus, D., n.d.b. Lvi-sam: Tightly-coupled lidar-visual-inertial odometry via smoothing and mapping. *2021 IEEE international conference on robotics and automation (ICRA)*, IEEE, 5692–5698.

Sun, S., Cioffi, G., De Visser, C., Scaramuzza, D., 2021. Autonomous quadrotor flight despite rotor failure with onboard vision sensors: Frames vs. events. *IEEE Robotics and Automation Letters*, 6(2), 580-587.

Vidal, A. R., Rebecq, H., Horstschaefer, T., Scaramuzza, D., 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), 994-1001.

Weikersdorfer, D., Conradt, J., n.d. Event-based particle filtering for robot self-localization. *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, IEEE, 866–870.

Wisth, D., Camurri, M., Das, S., Fallon, M., 2021. Unified multi-modal landmark tracking for tightly coupled lidar-visual-inertial odometry. *IEEE Robotics and Automation Letters*, 6(2), 1004-1011.

Xu, W., Zhang, F., 2021. Fast-lio: A fast, robust lidar-inertial odometry package by tightly-coupled iterated kalman filter. *IEEE Robotics and Automation Letters*, 6(2), 3317-3324.

Zheng, C., Zhu, Q., Xu, W., Liu, X., Guo, Q., Zhang, F., n.d. Fast-livo: Fast and tightly-coupled sparse-direct lidar-inertial-visual odometry. *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 4003–4009.