Radar and Event Camera Fusion for Agile Robot Ego-Motion Estimation

Yang Lyu, *Member, IEEE,*

***Abstract*—Achieving reliable ego motion estimation for agile robots, e.g. aerobatic aircraft, remains challenging due to most robot sensors fail to respond timely and clearly to highly dynamic robot motions, and often results in blurring, distortion and delays. In this paper, we consider deriving two types of instantaneous motion measurements, namely the translational velocity and angular velocity, from an event camera, and a Doppler radar respectively to estimate the motion of a highly dynamic robot platform. Based on the sensor setup, no sophisticated associations between measurement frames are needed, which will dramatically reduce the computational complexity. In the backend, we propose a continuous-time state-space model to fuse the asynchronous measurements and derive the ego-motion estimation. Finally, we validate our framework in self-collected experiment datasets, which demonstrate both the accuracy and efficiency of our proposed framework.**

***Index Terms*—Doppler radar, event camera, ego-motion estimation.**

I. Introduction

Reliable ego-motion estimation is one fundamental function featuring various autonomous robot platforms. Supporting techniques include early stage Global Navigation Satellite Systems (GNSS)/Inertial Navigation Systems (INS), and more recent Simultaneously Localization and Mapping (SLAM) which introduces more sensors, such as cameras, Light Detection and Ranging (LiDAR), Radars. Comparing the former approach, the SLAM-based approach do not rely on external satellite signals and are suitable for integrating more sensing modalities, which will undoubtedly expand its application scenarios. Successful deployment of the SLAM-base approaches on different platforms have been reported recently. Among them, various sensor modalities, such as cameras, Light Detection and Ranging (LiDAR), and intertial measurement units (IMUs) and combinations, are used based on the principle that all motion degrees of the platform can be sufficiently distinguished based on collected sensor measurements. However, for robot system with agile maneuverability, such as aerobatic UAVs and racing UGVs, the primary robot sensors mentioned above may fail to capture the motion of the platform.

As a matter of fact, cameras capture images over the exposure time, and most LiDARs achieves environment ranging serially over time. The motion during the sensing time period may not be neglected for aggressive moved platforms with high translational and rotational velocities, and will introduce blur and distortion to measurements. Comparing to the classical sensors mentioned above, recent developed event camera sensor can output pixel measurement asynchronously, responding to intensity changes by producing *events* according to the change of the light intensity, therefore can provide more motion-sensitive signal for subsequent ego-motion estimation. However, the event camera alone lacks metric information and cannot recover all motion degrees of a mobile platform.

In this paper, we focus on accurate agile robot ego-motion estimation based on event cameras. To fully recover the ego-motion, we additionally fuse doppler measurement from a 4D millimiter wave radar to provide extra metric information. One main advantage of above setup is that we can recover the instantaneous linear/angular velocity instead of most ego-motion works that estimate a finite posed displacement. In addition to that, with the measurement characteristics of both sensors, the computationally burdensome frame-to-frame feature matching/association can be avoided, therefore the setup further favors ego-motion estimation in high frequency. To handle the asynchronized measurement from the event camera and the radar, we fuse the information in a continuous-time SLAM framework to avoid measurement alignment. The contributions lies as below

- We develop a light-weight frontend to obtain translational and angular velocities from event camera and radar doppler measurement respectively, without frame-to-frame associations.
- We further formulate velocity factors and carry out the fusion of asynchronzied measurement in a continuous time fashion to support the ego-motion of an agile moving platform.
- We finally test our proposed method in various platforms, with comparisons to different setups, which can further support the advantage of our proposed platform.

The remaining of the paper is organized as follows. We provide related literature review in Section II. The registration-free velocity is derived in Section III. We formulate the continuous-time ego-motion estimator in Section IV. Validations of the proposed framework is provided in Section V. We conclude the paper in Section VI.

II. Related Works

A. Event-camera based ego-motion estimation

Visual SLAM (vSLAM) has drawn tremendous attention from various research communities and shown great potential in many applications due to the compelling characteristic, such

Yang Lyu was with the School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi, 710129 P.R. China e-mail: lyu.yang@nwpu.edu.cn.

This work was supported by the National Natural Science Foundation of China under Grant 62203358, Grant 62233014, and Grant 62073264. (*Corresponding author: Yang Lyu)*

as low cost and rich texture information of the vision sensors. Despite the advantages, ns, such as on a aggressively moving platform, or with complex light conditions. Until recently, the invention of the event camera may unlock the application of vSLAM in more challenging environments, by providing high temporal resolution, a high dynamic range (140dB vs. 60dB of standard cameras), and low power and bandwidth requirements. Different from a standard frame-base camera which output images at a fixed rate, each pixel from an event camera can operate independently and asynchronously respond to logarithmic scale of intensity change by producing 'events'. Due to sensing mechanism differences, the well researched vSLAM methods based on standard cameras cannot be integrated to event camera directly. Recently many works focus on taking advantage of the event-based data to achieve vSLAM in challenging environments.

Works can be categorized based on how event data is utilized [1]. Initial works basically follow the feature detection and tracking pipelines of typical vSLAM frontend but with special consideration on event data. [2] first synthesize an event frame within a spatio-temporal window from asynchronous events, and point features are detected and tracked from the event frame. Then a keyframe-based SLAM baseend is carried out. Simtraditional vision sensors suffers from low dynamic range, and are vulnerable to light changes or aggressive motions, which further hinder the application of vSLAM to more challenging conditioilar feature detection and tracking pipeline is also used in [3]. Besides, line features are adopted in [4], [5] to achieve reliable camera tracking. [6] fuses line features and inertial measurement to achieve high frequency velocity estimation. As one advantage, features can be more explicit tracked with the embedded dynamic information of the event data.

There are works also explore to recover the camera motion direct from the event besides the feature detection and tracking pipeline. A straightforward way to implement the direct vSLAM method is by generate an event frame data by temporal-spatial alignment [7]. One more complicated direct event-based odometry is proposed in [8] where the information from events and frames are tightly-coupled fused with the Event Generation Model (EGM) [9] and Photometric Bundle Adjustment (PBA).

Finally, with the remarkable success of deep learning in computer vision applications, using machine learning methods to achieve event camera SLAM becomes a promising research direction. Similarly, the asynchronous event data is first convert to event-frame representations similar to image frames to implement the convolution based deep neural networks. An unsupervised method is proposed in [10] which first represents the events in the form of a discretized volume that maintains the temporal distribution of the events, and then a unsupervised network is proposed to estimate the optical flow or/and egomotion and depth by evaluating the blur level of compensated event frame. In a supervised fashion, [11] proposed a novel network which can process asynchronous events and monocular image to predict depth and also egomotion in continuous time.

Although above works have demonstrated promising performance in robotics ego-motion, their inherent nature that using association between frames may still impose a computation burden to extreme limited computation resource of an onboard computer. On the other hand, due to the sensing mechanism, the event can be treated as a measure of instantaneous motion of the camera. Therefore in this paper we consider to obtain the instantaneous velocity from the event camera.

B. Radar ego-motion estimation

Usign Doppler radar to achieve ego motion estimation becomes a promising alternative solution to existing camera-based and LiDAR-based methods, mainly due to the success application of Doppler radar in automotive fields, which reveal its robustness against environment hazards and capability of providing reliable sensing results.

Doppler radar can provide two type of information to achieve ego-motion estimation. First, by registering point clouds between frames or between a frame and a map, relative transformation can be obtained. A radar SLAM framework is proposed in [12] which demonstrates all weather condition adaptation capability. The consecutive frame registration and pose estimation are carried out by represent pointclouds as image based 2D data. With similar radar data representation, an unsupervised-learning based method to process the feature detection and tracking is proposed in [13] which is followed by a odometry estimation framework. These point cloud registration, as the front-ends of above methods, often brings considerable computation overheads, which may challenge many robotics systems with limited computation resources.

Second, as a special feature of Doppler radar, the radial velocities of echoes can be treated as instantaneous measurement to derive the velocity of radar ego-motion. Instantaneous velocity of a car is obtained by directly analyzing the velocity profile from one or multiple radars, respectively in [14] and [15]. A more rigorous 3D velocity factor is constructed in [16] and integrated in a pose graph based ego-motion framework. These methods do not require pre-processing of point clouds, such as clustering, registration, therefore can dramatically reduce the computation and storage burden. As one drawback, these methods may be more prune to error accumulation, and also suffers from angular velocity un-observability.

In recent radar-based SLAM frameworks, both types of information are used to achieve reliable pose estimation. As the development of imaging 4D millimeter-wave radar, 3D ego-motion can be achieved, which draw a lot of recent research attention. A 4D radar SLAM framework is described in [17] which formulate the frame-to-map registration and Doppler velocity as between pose and velocity prior factors respectively. Similar frameworks are also used in [18] and [19] with different velocity integration formulation. A Doppler LiDAR based framework is proposed in [20] which integrates LiDAR odometry, IMU and velocity within a graph optimization framework.

In this paper, we aim at achieving ego-motion with limited computation and storage resources by taking advantage of the Doppler velocity. Specially, we consider to remove the angular unobervability by combing an event camera. With

the two instantaneous measurements, registration free 6-DOF odometry can be obtained directly.

C. Continous-time SLAM

The SLAM function consists of estimating the trajectory of moving sensors and the mapping the observation consistently. According to the representation of the trajectory, SLAM can be divided into two categories, namely the discrete time (DT) SLAM and the continuous-time (CT) SLAM. In the CT SLAM, the trajectory is usually approximately encodes as B-splines or Gaussian processes. As one advantage of the representation, state can be sampled at any time, which apparently support the fusion of asynchronous measurement. Moreover, the parameterization make the SLAM a bound-sized optimization problem regardless the rate of sensor measurements, which apparently support the fusion process of high rate sensors, such as the event camera. As one drawback, the CT representation embedded an assumption that the trajectory is smooth to some degree which often requires high order of parameterization when encounter complex motion.

Developing CT SLAM remains a hot topic in recent years as the growing requirement of pushing SLAM to more complex environments, such as high-dynamical scenarios, degenerate observations, and so on, which further promote the application, such as event camera, and multi-modal fusion frameworks. A early work by Furgale [21] first propose to represent the trajectory as Gaussian basis in continuous time in a SLAM framework. Afterwards, works demonstrate different sensor modalities are proposed. [22] describes a CT SLAM framework under an asynchronous stereo-inertial setup. With spline-based trajectory representation, rigorous IMU pre-integration is avoided in the fusion estimation process. A LiDAR only odometry is proposed in [23] based on CT trajectory formulation to achieve pose estimation in aggressive motions. Instead of scans, The CT formulation can handle measurement as high frequency streaming LiDAR points continuously captured at separate time therefore no motion compensations for radar scans are needed. Especial related to our work, [24] fuses high-rate and asynchronous event camera data and IMU in a CT-SLAM framework. Thanks to the continuous time formulation, high rate measurements can be handled in a bound-sized estimation problem, so as to achieve pose estimation in highly dynamic platform. Besides handling high rate sensor, researchers also turn to CT SLAM to handle multi-sensor fusion based SLAM. A CT-SLAM framework that handles LiDAR, camera and IMU is proposed in [25] with online time-offset estimation.

In this paper, we plane to achieve ego-motion estimation based on the fusion of an event camera and a Doppler radar, which output high-frequent and asynchronous measurements. Keen on the sensor setup, a CT SLAM is considered a proper representation.

D. Notations

In this paper, we uses lowcase and upcase bold letters to represent vectors and matrices respectively. Time in continuous- and discrete-time is denoted by $t \in \mathbb{R}^+ \cup \{0\}$ and $k \in \mathbb{Z}$. Specifically, we uses $(\cdot)(t)$ and $(\cdot)^k$ to represent variables in continuous-time domain and discrete time respectively. Further We use calligraphic font letters to denote variables in different frames. We uses $\mathcal{G}$ to represent the global frame, and $\mathcal{R}$ and $\mathcal{C}$ are the radar frame and camera frame respectively. For example, ${}^{\mathcal{G}}\boldsymbol{p}_r(t)$ denote the position of the radar in global frame at time t, and ${}^{\mathcal{R}}\boldsymbol{R}_r^k$ is the radar's rotation matrix at time instance k.

III. Velocity Front-ends

In this section, the ego-motion front-end is provided. Specifically, we derive linear and angular velocity from a 4D MMWR and an event camera respectively.

A. Translational Velocity from 4D Radar

The 3D velocity of a 4D millimeter-wave radar can be obtained based on the 3D point clouds and Doppler velocity of each point. Define the position of a point $i \in C^k$ in radar frame as ${}^{\mathcal{R}}\mathbf{p}_i^k \in \mathbb{R}^3$, and its radial velocity as $v_i^k \in \mathbb{R}$, where C^k defines the set of point in one scan at time instance k, then the following geometric relationship holds:

$$\frac{\left({}^{\mathcal{R}}\mathbf{p}_i^k\right)^\top}{\|{}^{\mathcal{R}}\mathbf{p}_i^k\|}{}^{\mathcal{R}}\boldsymbol{v}_{\mathrm{r}}^k = -v_i^k, \tag{1}$$

where ${}^{\mathcal{R}}\boldsymbol{v}_{\mathrm{r}}^k$ is the ego-motion velocity of the radar in its local frame. Considering all points in each scan, we have $n = |C|$ equations,

$$\begin{bmatrix} -v_1^k \\ -v_2^k \\ \vdots \\ -v_n^k \end{bmatrix} \triangleq \boldsymbol{v}_d^k = \begin{bmatrix} \frac{\left({}^{\mathcal{R}}\mathbf{p}_1^k\right)^\top}{\|{}^{\mathcal{R}}\mathbf{p}_1^k\|} \\ \frac{\left({}^{\mathcal{R}}\mathbf{p}_2^k\right)^\top}{\|{}^{\mathcal{R}}\mathbf{p}_2^k\|} \\ \vdots \\ \frac{\left({}^{\mathcal{R}}\mathbf{p}_n^k\right)^\top}{\|{}^{\mathcal{R}}\mathbf{p}_n^k\|} \end{bmatrix} {}^{\mathcal{R}}\boldsymbol{v}_{\mathrm{r}} \triangleq \mathbf{H}^k {}^{\mathcal{R}}\boldsymbol{v}_{\mathrm{r}}^k. \tag{2}$$

Then the velocity estimate ${}^{\mathcal{R}}\hat{\boldsymbol{v}}_{\mathrm{r}}^k$ can be obtained in a least-square manner as

$${}^{\mathcal{R}}\hat{\boldsymbol{v}}_{\mathrm{r}}^k = \left(\left(\mathbf{H}^k\right)^\top \mathbf{H}^k\right)^{-1} \left(\mathbf{H}^k\right)^\top \boldsymbol{v}_d. \tag{3}$$

Practically, there are outliers, from moving objects or clutter, that will dramatically affect the accuracy of the velocity estimation. To exclude the outliers, methods such as RANSAC [15], or introducing prior velocity from odometry or IMU [**?**], can be used.

After the outlier rejection, a measurement covariance can be calculated based on (3),

$$\mathbf{Q}_{\boldsymbol{v}_r} = \frac{\left(\boldsymbol{r}_{\boldsymbol{v}_d} \boldsymbol{r}_{\boldsymbol{v}_d}{}^\top\right) \left(\mathbf{H}^\top \mathbf{H}\right)^{-1}}{\dim(\boldsymbol{v}_d) - 3}, \tag{4}$$

where $\boldsymbol{r}_{\boldsymbol{v}_d}^k = \mathbf{H}^k \hat{\boldsymbol{v}}^k - \boldsymbol{v}_d^k$ denote the doppler measurement residuals.

B. Rotational Velocity from Event Camera

In this section, we propose to use event camera data to obtain the angular velocity of the event camera. First, we derive optical flow directly from asynchronous events, and then we derive the angular velocity based on continuous-time epipolar constraint.

1) Optical flow from events: To begin with, an event is defined as $e(\boldsymbol{x},t)$, where $\boldsymbol{x}=[x,y]^\top \in \mathbb{R}^2$ is the position of the event in pixel. The value of $e(\boldsymbol{m},t)$ could be 1 or -1 when the contrast change is positive or negative.

According to [26], the following relationship between event $e(x_j,y_j,t)$ and optical flow, which is defined as $\boldsymbol{u}(t)=[v_x(t),v_y(t)]^\top \in \mathbb{R}^2$ holds

$$\begin{aligned}&\left(\sum_{t-\Delta t}^{t} e\left(x_j,y_j,t\right)-\sum_{t-\Delta t}^{t} e\left(x_j-1,y_j,t\right)\right)v_x(t)\\&+\left(\sum_{t-\Delta t}^{t} e\left(x_j,y_j,t\right)-\sum_{t-\Delta t}^{t} e\left(x_j,y_j-1,t\right)\right)v_y(t)\\&=\frac{\sum_{t_1}^{t} e\left(x_j,y_j,t\right)}{t-t_1},\end{aligned}\tag{5}$$

where Δt is a small time interval used to accumulate events. $t_1<t$ is a time instance for calculate the time differentiation of e. Apparently (5) is ill-conditioned. Consider the smoothness of optical flow, we assume that the optical flow within a $n\times n$ small patch centered at $(x_j,y_j)$is constant, which further lead to n^2-1 more equations, and v_x,v_y can be calculated in a least-squared fashion similar to (3).

2) angular velocity from optical flow: Let us first define the camera pose as $\mathbf{T}(t)\in\mathbb{SE}(3)$, and its velocity vector in the camera frame is defined as $\left[{}^{\mathcal{E}}\boldsymbol{v}_e(t)^\top,{}^{\mathcal{E}}\boldsymbol{\omega}_e(t)^\top\right]^\top\in\mathfrak{se}(3)$. Given a landmark point in the camera frame as ${}^{\mathcal{E}}\boldsymbol{p}_l(t)\in\mathbb{R}^3$, the following equation holds:

$${}^{\mathcal{E}}\dot{\boldsymbol{p}}_l(t)=\left[{}^{\mathcal{E}}\boldsymbol{\omega}_e(t)\right]_\times {}^{\mathcal{E}}\boldsymbol{p}_l(t)+{}^{\mathcal{E}}\boldsymbol{v}_e(t),\tag{6}$$

where $\left[{}^{\mathcal{E}}\boldsymbol{\omega}_e\right]_\times$ is the skew-symmetric matrix associated with angular velocity vector ${}^{\mathcal{E}}\boldsymbol{\omega}_e\in\mathbb{R}^3$, and ${}^{\mathcal{E}}\boldsymbol{v}_e(t)\in\mathbb{R}^3$ denotes the translational velocity. We further define ${}^{\mathcal{E}}\boldsymbol{p}_l(t)=\lambda(t)\boldsymbol{x}(t)$, then ${}^{\mathcal{E}}\dot{\boldsymbol{p}}_l(t)=\dot{\lambda}(t)\boldsymbol{x}(t)+\lambda(t)\dot{\boldsymbol{x}}(t)$. Substitute it into (6), we have

$$\dot{\boldsymbol{x}}(t)=\left[{}^{\mathcal{E}}\boldsymbol{\omega}_e(t)\right]_\times\boldsymbol{x}(t)+\frac{1}{\lambda(t)}{}^{\mathcal{E}}\boldsymbol{v}_e(t)-\frac{\dot{\lambda}(t)}{\lambda(t)}\boldsymbol{x}(t).\tag{7}$$

Left multiply both sides of Eq. (7) with $\left([\boldsymbol{x}(t)]_\times {}^{\mathcal{E}}\boldsymbol{v}_e(t)\right)^\top$, it becomes

$$\left([\boldsymbol{x}(t)]_\times {}^{\mathcal{E}}\boldsymbol{v}_e(t)\right)^\top\dot{\boldsymbol{x}}(t)+\left([\boldsymbol{x}(t)]_\times {}^{\mathcal{E}}\boldsymbol{v}_e(t)\right)^\top[\boldsymbol{x}(t)]_\times {}^{\mathcal{E}}\boldsymbol{\omega}_e(t)=0.\tag{8}$$

Above equation defines the continuous epipolar constraint. Apparently, (8) do not dependent on the 3D position of any world point, and only on the 2D observations of the world point. Under the assumption of brightness constancy, $\dot{\boldsymbol{x}}(t)$ can be approximated by the optical flow $\boldsymbol{u}(t)$ obtained from the event camera.

The velocity of the camera in its own frame can be calculated from the radar velocity as

$${}^{\mathcal{E}}\boldsymbol{v}_e(t)=\boldsymbol{C}_{\mathcal{R}}^{\mathcal{E}}{}^{\mathcal{R}}\boldsymbol{v}_r(t)-\boldsymbol{C}_{\mathcal{B}}^{\mathcal{E}}{}^{\mathcal{B}}\boldsymbol{\omega}_b(t)\times{}^{\mathcal{B}}\boldsymbol{l}_{er},\tag{9}$$

where $\boldsymbol{C}_{\mathcal{R}}^{\mathcal{E}}$ and $\boldsymbol{C}_{\mathcal{B}}^{\mathcal{E}}$ are the rotation matrix from radar frame and body frame, to the event camera frame, respectively. ${}^{\mathcal{B}}\boldsymbol{l}_{er}$ denote the displacement from radar to the camera. When ${}^{\mathcal{B}}\boldsymbol{l}_{er}$ is small enough, the velocity can be approximately obtained as ${}^{\mathcal{E}}\boldsymbol{v}_e(t)=\boldsymbol{C}_{\mathcal{R}}^{\mathcal{E}}{}^{\mathcal{R}}\boldsymbol{v}_r(t)$, where ${}^{\mathcal{R}}\boldsymbol{v}_r(t)$ can be replaced with the closest radar ego-motion estimates in discrete time instance k.

Define $\boldsymbol{a}=\left([\boldsymbol{x}(t)]_\times\boldsymbol{C}_{\mathcal{R}}^{\mathcal{E}}{}^{\mathcal{R}}\boldsymbol{v}_r(t)\right)^\top[\boldsymbol{x}(t)]_\times$, and $b=\left([\boldsymbol{x}(t)]_\times\boldsymbol{C}_{\mathcal{R}}^{\mathcal{E}}{}^{\mathcal{R}}\boldsymbol{v}_r(t)\right)^\top\boldsymbol{u}(t)$, equation (8) is simplified as

$$\boldsymbol{a}^{\mathcal{E}}\boldsymbol{\omega}_e(t)=b.\tag{10}$$

Given n points satisfying constraint (8), we can have n equations. Consider that the angular velocity remains constant during very short time period, and there are n optical flow points,

$$\boldsymbol{A}^{\mathcal{E}}\boldsymbol{\omega}_e(t)=\boldsymbol{b}.\tag{11}$$

Then ${}^{\mathcal{E}}\boldsymbol{\omega}_e(t)$ can be obtained in a least square fashion.

IV. Continuous-time Estimation

In this section, the fusion of instantaneous velocity measurement is carry out to estimate the continuous-time odometry. The problem is formulated as a nonlinear sliding-windowed estimation problem and solved in a discrete-time manner.

A. B-splines based trajectory

To begin with, we represent the continuous-time trajectory as cumulative B-splines. We parameterize the trajectory similarly to [27] which splits the 6-degree pose to the body position part and the body rotation part in the global frame, which is represented by continuous-time function ${}^{\mathcal{G}}\boldsymbol{p}(t)_b\in\mathbb{R}^3$ and ${}^{\mathcal{G}}\boldsymbol{R}_b(t)\in\mathbb{SO}(3)$ respectively. The superscripts and subscripts are ignored and we use $\boldsymbol{p}(t)$ and $\boldsymbol{R}(t)$ when no ambiguities.

Considering the translational function $\boldsymbol{p}(t)$ over a time period is of order k, and controlled by points $\boldsymbol{p}_i,\boldsymbol{p}_{i+1},\cdots,\boldsymbol{p}_{i+k-1}$, then we have the following equation that represent $\boldsymbol{p}(t)$, which are

$$\boldsymbol{p}(t)=\boldsymbol{p}_i+\sum_{j=1}^{k-1}\lambda_j(t)\cdot\Delta\boldsymbol{p}_{ij},\tag{12}$$

where $\lambda_j(t)$ is constant coefficient which depends on the order k, and $\Delta\boldsymbol{p}_{ij}\triangleq\boldsymbol{p}_{i+j}-\boldsymbol{p}_{i+j-1}\in\mathbb{R}^3$. For the rotation part, the cumulative B-spline controlled by $\boldsymbol{R}_i,\boldsymbol{R}_{i+1},\cdots,\boldsymbol{R}_{i+k-1}$ is defined as

$$\boldsymbol{R}(t)=\boldsymbol{R}_i\cdot\prod_{j=1}^{k-1}\mathrm{Exp}\left(\lambda_j(t)\cdot\Delta\boldsymbol{R}_{ij}\right),\tag{13}$$

where $\Delta\boldsymbol{R}_{ij}\triangleq\log\left(\boldsymbol{R}_{i+j-1}^{-1}\boldsymbol{R}_{i+j}\right)\in\mathbb{R}^3$. Then we can define the 6D continuous trajectory in global frame as ${}^{\mathcal{G}}\boldsymbol{T}_b\triangleq\{\boldsymbol{p}(t),\boldsymbol{R}(t)\}$.

Given the continuous B-splines Eq.(12) and Eq.(13), we can obtain the velocity as

$${}^{\mathcal{G}}\boldsymbol{v}_b(t)={}^{\mathcal{G}}\dot{\boldsymbol{p}}_b(t)=\sum_{j=1}^{k-1}\dot{\lambda}_j(t)\cdot\Delta\boldsymbol{p}_{ij}\tag{14}$$

and

$${}^{\mathcal{G}}\boldsymbol{\omega}_b(t)=\left({}^{\mathcal{G}}\boldsymbol{R}_b^\top(t){}^{\mathcal{G}}\dot{\boldsymbol{R}}_b(t)\right)^\vee,\tag{15}$$

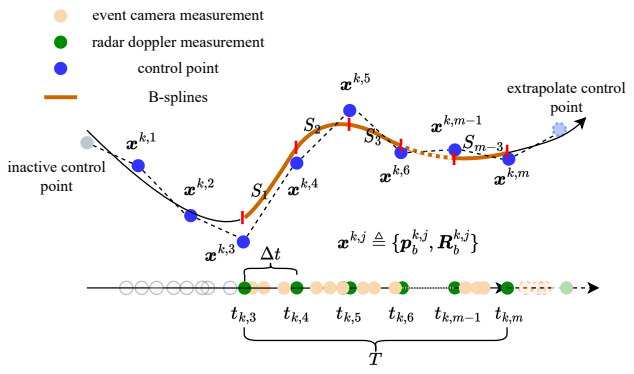

Fig. 1. The B-splines control points distribution and measurements

where

$$ {}^{\mathcal{G}}\dot{\boldsymbol{R}}_b(t) = {}^{\mathcal{G}}\boldsymbol{R}_i \sum_{j=1}^{k-1} \left(\prod_{l=1}^{j-1} \boldsymbol{A}_l(t) \right) \dot{\boldsymbol{A}}_j(t) \left(\prod_{l=j+1}^{k-1} \boldsymbol{A}_l(t) \right) $$

and $\boldsymbol{A}_j(t) \triangleq \mathrm{Exp}\left(\lambda_j(t) \cdot \Delta \boldsymbol{R}_{ij}\right)$.

Then we can derivative the transnational and rotational velocity in local frame as

$$ {}^{\mathcal{B}}\boldsymbol{v}_b(t) = {}^{\mathcal{G}}\boldsymbol{R}_b(t)^\top {}^{\mathcal{G}}\boldsymbol{v}_b(t) - {}^{\mathcal{G}}\boldsymbol{R}_b(t)^\top \left[{}^{\mathcal{G}}\boldsymbol{p}_b(t)\right]_\times {}^{\mathcal{G}}\boldsymbol{\omega}_b(t), \quad (16) $$

$$ {}^{\mathcal{B}}\boldsymbol{\omega}_b(t) = {}^{\mathcal{G}}\boldsymbol{R}_b(t)^\top {}^{\mathcal{G}}\boldsymbol{\omega}_b(t). \quad (17) $$

B. Sliding-windowed Optimization

In this paper, we formulate the trajectory estimation problem as a sliding-windowed optimization problem which is solved in a discrete time fashion. The objective function is to minimize the error between the predicted measurement extrapolated from the continuous-time trajectory function and the obtained instantaneous measurements. Specifically, we consider three types of information: the radar measurement, the event camera measurement, and the prior information from the last round of optimization. To begin with, we define the sliding-window state to be estimated at time instance k as follows:

$$ \mathcal{X}^k \triangleq \left\{ \boldsymbol{\Phi}^k, \boldsymbol{b}_{\boldsymbol{v}}^k, \boldsymbol{b}_{\boldsymbol{\omega}}^k, \Delta t_{re}^k \right\}, \quad (18) $$

where $\boldsymbol{\Phi}^k \triangleq \left\{ {}^{\mathcal{G}}\boldsymbol{p}_b^{k,j}, {}^{\mathcal{G}}\boldsymbol{R}_b^{k,j} \right\}_{j=1:m}$ denotes the control points of $m-3$ B-splines at time instance k, $\boldsymbol{b}_{\boldsymbol{v}}^k$ and $\boldsymbol{b}_{\boldsymbol{\omega}}^k$ are the biases of the translational and rotational velocity respectively, and Δt_{re}^k is the synchronous delays between the radar and event camera measurements.

Our sliding window-based optimization is illustrated in Fig.1. We consider the sliding window span T, and all control points are equally distributed with time intervals $\Delta t = T/(m-4)$. We consider using all measurements that fall into the period $[t_{k,3}, t_{k,m}]$. The objective function is formulated as

$$ \arg\min_{\mathcal{X}^k} \alpha_r(\mathcal{X}^k) + \alpha_e(\mathcal{X}^k) + \alpha_{\text{prior}}(\mathcal{X}^k), \quad (19) $$

where $\alpha_r \in \mathbb{R}$, $\alpha_e \in \mathbb{R}$, and $\alpha_{\text{prior}} \in \mathbb{R}$ are the radar, event camera, and prior information terms respectively.

C. Measurement Models

In this part, we present the translational velocity and angular velocity measurements obtained from a radar and an event camera respectively.

1) Doppler Velocity Measurement: We consider the radar measurement obtained from III-A, ${}^{\mathcal{R}}\hat{\boldsymbol{v}}_{\text{r}}$, includes the true linear velocity corrupted by a time-varying noise and a constant bias, namely

$$ {}^{\mathcal{R}}\hat{\boldsymbol{v}}_r^k = {}^{\mathcal{R}}\boldsymbol{v}_{r,true}^k + \boldsymbol{b}_v^k + \boldsymbol{\eta}_v^k, \quad (20) $$

where $\boldsymbol{b}_v^k$ denote the slow varying radar bias and $\boldsymbol{\eta}_v^k$ is the white noise. The velocity can be integrated into optimization in both loosely-coupled and tightly coupled fashion. In the loosely coupled fashion, all points within a scan are used to first calculate the ego-motion velocity, then the velocity error term can be calculated as

$$ \boldsymbol{r}_v^k = {}^{\mathcal{R}}\hat{\boldsymbol{v}}_r^k - {}^{\mathcal{R}}\bar{\boldsymbol{v}}_r^k - \hat{\boldsymbol{b}}_v^k, \quad (21) $$

where $\bar{\boldsymbol{\omega}}_b^k$ and ${}^{\mathcal{B}}\bar{\boldsymbol{v}}_b^k$ are the interpolated/extrapolated pseudo measurements based on current estimates, and ${}^{\mathcal{R}}\bar{\boldsymbol{v}}_r \triangleq {}^{\mathcal{R}}\boldsymbol{R}_b({}^{\mathcal{B}}\bar{\boldsymbol{\omega}}_b^k \times {}^{\mathcal{B}}\boldsymbol{l}_r + {}^{\mathcal{B}}\bar{\boldsymbol{v}}_b^k)$ are the predicted radar velocity based on pseudo measurements.

Then the cost term w.r.t. radar α_r is defined as

$$ \alpha_r = \sum_{j \in M_r^k} \left\| \boldsymbol{r}_v^j \right\|_{\boldsymbol{\Sigma}_v^{-1}}^2, \quad (22) $$

where M_r^k is the set of active radar velocity measurements within the sliding window, and $\boldsymbol{\Sigma}_v$ is the covariance matrix of the radar velocity measurement.

In addition to that, we also consider a tightly coupled fashion. In this fashion, an error term for each point Doppler measurement i is formulated as

$$ r_d^{k,i} = v_r^{k,i} - \frac{{}^{\mathcal{R}}\boldsymbol{p}_i^k}{\|{}^{\mathcal{R}}\boldsymbol{p}_i^k\|} \left({}^{\mathcal{R}}\bar{\boldsymbol{v}}_r^k + \hat{\boldsymbol{b}}_v^k \right). \quad (23) $$

Alternatively, we can define a tightly-coupled cost term w.r.t. radar α_j as

$$ \alpha_r = \sum_{j \in M_r^k} \sum_{i \in C_r^j} \frac{1}{\sigma_d^2} |r_d^{k,i}|^2, \quad (24) $$

where C_r^k is the set of points of the time instance j, σ_d is the standard deviation of the Doppler measurement of each point.

2) Optical Flow Measurement: Similar to the linear velocity, we consider the information from an event camera to be fused in both loosely coupled and tightly coupled fashions. In the loosely coupled fashion, an angular velocity measurement is first obtained based on 11. Without loss of generality, we consider the measurement to be corrupted by a time-varying noise and a constant bias, which are denoted as $\boldsymbol{\eta}_{\boldsymbol{\omega}}(t)$ and $\boldsymbol{b}_{\boldsymbol{\omega}}(t)$ respectively. Then the angular velocity measurement can be modeled as

$$ {}^{\mathcal{E}}\hat{\boldsymbol{\omega}}_{\text{e}}^{k'} = {}^{\mathcal{E}}\boldsymbol{\omega}_{e,true}^{k'} + \boldsymbol{b}_{\boldsymbol{\omega}}^{k'} + \boldsymbol{\eta}_{\boldsymbol{\omega}}^{k'}, \quad (25) $$

The superscript k' is used to denote the time instance when an angular velocity measurement is obtained. Note that we use a time instance variable k' to denote the time instance for the event camera due to its asynchronous sensing mechanism.

Then we can define a residual term based on the extrapolated pseudo measurements as

$$\boldsymbol{r}_{\omega} = {}^{\mathcal{E}}\hat{\boldsymbol{\omega}}_e^{k'} - {}^{\mathcal{E}}\bar{\boldsymbol{\omega}}_e^{k'} - \boldsymbol{b}_{\boldsymbol{\omega}}^{k'}, \tag{26}$$

where ${}^{\mathcal{E}}\bar{\boldsymbol{\omega}}_e^{k'} = {}^{\mathcal{B}}\boldsymbol{R}_e {}^{\mathcal{B}}\boldsymbol{\omega}_b$ is the pseudo angular velocity measurement based on the current estimates.

Then the cost term w.r.t. the angular velocity α_{ω} is defined as

$$\alpha_e = \sum_{j \in M_{\omega}^{k'}} \|\boldsymbol{r}_{\omega}\|^2_{\boldsymbol{\Sigma}_{\omega}^{-1}}, \tag{27}$$

where M_e^k is the set of active angular velocity measurements within the sliding window, and $\boldsymbol{\Sigma}_{\omega}$ is the covariance matrix of the angular velocity measurement.

We can also fuse the angular velocity information in a tightly coupled fashion. In this fashion, the error term for each event is defined as

$$\boldsymbol{r}_e^{k'} = \left(\left[\boldsymbol{x}^{k'}\right]_{\times} {}^{\mathcal{E}}\bar{\boldsymbol{v}}_e^{k'}\right)^{\top} \dot{\boldsymbol{x}}^{k'} + \left(\left[\boldsymbol{x}^{k'}\right]_{\times} {}^{\mathcal{E}}\bar{\boldsymbol{v}}_e^{k'}\right)^{\top} \left[\boldsymbol{x}^{k'}\right]_{\times} {}^{\mathcal{E}}\bar{\boldsymbol{\omega}}_e^{k} \tag{28}$$

where ${}^{\mathcal{E}}\bar{\boldsymbol{v}}_e^{k'} = {}^{\mathcal{E}}\boldsymbol{R}_b({}^{\mathcal{B}}\bar{\boldsymbol{\omega}}_b^{k'} \times {}^{\mathcal{B}}\boldsymbol{l}_e + {}^{\mathcal{B}}\bar{\boldsymbol{v}}_b^{k'})$ is the pseudo velocity measurement based on the current estimates. Apparently, the above error terms related to both translational velocity and angular velocity when using the tightly-coupled error term.

Alternatively, we can define a tightly-coupled cost term w.r.t. event camera α_{ω} as

$$\alpha_e = \sum_{j \in M_e^k} \|\boldsymbol{r}_e\|^2_{\boldsymbol{\Sigma}_e^{-1}}, \tag{29}$$

where M_e^k is the set of all active event measurements within the sliding window.

3) Prior information: Besides the above two factors, we also integrate prior information during each sliding-windowed optimization. Specifically, the prior factors constrain the common control points between two consecutive sliding windows. Besides, we consider the biases and time offset as slow-varying parameters. Therefore, the prior constraints on the following states:

$$\mathcal{X}_p^k \triangleq \{\boldsymbol{\Phi}^{k-1} \cap \boldsymbol{\Phi}^k, \boldsymbol{b}_{\boldsymbol{v}}^t, \boldsymbol{b}_{\boldsymbol{\omega}}^t, \delta_t\}. \tag{30}$$

Then the prior factor residual can be defined as

$$\boldsymbol{r}_p^k = \boldsymbol{H}_p \mathcal{X} - \mathcal{X}_p^k, \tag{31}$$

where $\boldsymbol{H}_p$ is the matrix to select overlapped terms of current states w.r.t. prior information. Then the prior constraint term α_p

V. OBSERVABILITY

We present the observability of 6-DOF pose estimation

VI. EXPERIMENTS ANALYSIS

We carry out the validation of the proposed method in both a hand-hold sensor package and platform-carrying packages. Specifically, we use both a vehicle and an airborne UAV.

APPENDIX A
PROOF OF THE FIRST ZONKLAR EQUATION

Appendix one text goes here.

APPENDIX B

Appendix two text goes here.

ACKNOWLEDGMENT

The authors would like to thank...

REFERENCES

[1] K. Huang, S. Zhang, J. Zhang, and D. Tao, "Event-based simultaneous localization and mapping: A comprehensive survey," *arXiv preprint arXiv:2304.09793*, 2023.

[2] H. Rebecq, T. Horstschaefer, and D. Scaramuzza, "Real-time visual-inertial odometry for event cameras using keyframe-based nonlinear optimization," 2017.

[3] 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.

[4] W. O. Chamorro Hernández, J. Andrade-Cetto, and J. Solà Ortega, "High-speed event camera tracking," in *Proceedings of the The 31st British Machine Vision Virtual Conference*, 2020, pp. 1–12.

[5] W. Chamorro, J. Sola, and J. Andrade-Cetto, "Event-based line slam in real-time," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 8146–8153, 2022.

[6] W. Xu, X. Peng, and L. Kneip, "Tight fusion of events and inertial measurements for direct velocity estimation," *IEEE Transactions on Robotics*, 2023.

[7] H. Rebecq, T. Horstschäfer, G. Gallego, and D. Scaramuzza, "Evo: A geometric approach to event-based 6-dof parallel tracking and mapping in real time," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 593–600, 2016.

[8] J. Hidalgo-Carrió, G. Gallego, and D. Scaramuzza, "Event-aided direct sparse odometry," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5781–5790.

[9] H. Kim, S. Leutenegger, and A. J. Davison, "Real-time 3d reconstruction and 6-dof tracking with an event camera," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14*. Springer, 2016, pp. 349–364.

[10] A. Z. Zhu, L. Yuan, K. Chaney, and K. Daniilidis, "Unsupervised event-based learning of optical flow, depth, and egomotion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 989–997.

[11] D. Gehrig, M. Rüegg, M. Gehrig, J. Hidalgo-Carrió, and D. Scaramuzza, "Combining events and frames using recurrent asynchronous multimodal networks for monocular depth prediction," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2822–2829, 2021.

[12] Z. Hong, Y. Petillot, A. Wallace, and S. Wang, "Radarslam: A robust simultaneous localization and mapping system for all weather conditions," *The International Journal of Robotics Research*, vol. 41, no. 5, pp. 519–542, 2022.

[13] D. Barnes and I. Posner, "Under the radar: Learning to predict robust keypoints for odometry estimation and metric localisation in radar," in *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020, pp. 9484–9490.

[14] D. Kellner, M. Barjenbruch, J. Klappstein, J. Dickmann, and K. Dietmayer, "Instantaneous ego-motion estimation using doppler radar," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*. IEEE, 2013, pp. 869–874.

[15] ——, "Instantaneous ego-motion estimation using multiple doppler radars," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 1592–1597.

[16] Y. S. Park, Y.-S. Shin, J. Kim, and A. Kim, "3d ego-motion estimation using low-cost mmwave radars via radar velocity factor for pose-graph slam," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7691–7698, 2021.

[17] Y. Zhuang, B. Wang, J. Huai, and M. Li, "4d iriom: 4d imaging radar inertial odometry and mapping," *IEEE Robotics and Automation Letters*, 2023.

[18] J. Zhang, H. Zhuge, Z. Wu, G. Peng, M. Wen, Y. Liu, and D. Wang, "4dradarslam: A 4d imaging radar slam system for large-scale environments based on pose graph optimization," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 8333–8340.

[19] X. Li, H. Zhang, and W. Chen, "4d radar-based pose graph slam with ego-velocity pre-integration factor," *IEEE Robotics and Automation Letters*, 2023.

[20] M. Nissov, S. Khattak, J. A. Edlund, C. Padgett, K. Alexis, and P. Spieler, "Roamer: Robust offroad autonomy using multimodal state estimation with radar velocity integration," *arXiv preprint arXiv:2401.17404*, 2024.
[21] P. Furgale, T. D. Barfoot, and G. Sibley, "Continuous-time batch estimation using temporal basis functions," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 2088–2095.
[22] D. Hug, P. Bänninger, I. Alzugaray, and M. Chli, "Continuous-time stereo-inertial odometry," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6455–6462, 2022.
[23] X. Zheng and J. Zhu, "Traj-lo: In defense of lidar-only odometry using an effective continuous-time trajectory," *IEEE Robotics and Automation Letters*, 2024.
[24] E. Mueggler, G. Gallego, H. Rebecq, and D. Scaramuzza, "Continuous-time visual-inertial odometry for event cameras," *IEEE Transactions on Robotics*, vol. 34, no. 6, pp. 1425–1440, 2018.
[25] J. Lv, X. Lang, J. Xu, M. Wang, Y. Liu, and X. Zuo, "Continuous-time fixed-lag smoothing for lidar-inertial-camera slam," *IEEE/ASME Transactions on Mechatronics*, 2023.
[26] R. Benosman, S.-H. Ieng, C. Clercq, C. Bartolozzi, and M. Srinivasan, "Asynchronous frameless event-based optical flow," *Neural Networks*, vol. 27, pp. 32–37, 2012.
[27] D. Hug and M. Chli, "Hyperslam: A generic and modular approach to sensor fusion and simultaneous localization and mapping in continuous-time," in *2020 International Conference on 3D Vision (3DV)*. IEEE, 2020, pp. 978–986.

PLACE PHOTO HERE

Michael Shell Biography text here.

John Doe Biography text here.

Jane Doe Biography text here.