IEEE Robotics and Automation Letters (RAL) paper presented at
2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

Spatio-Temporal Calibration for Omni-Directional Vehicle-Mounted Event Cameras

Xiao Li, Yi Zhou, Ruibin Guo, Xin Peng, Zongtan Zhou, and Huimin Lu

***Abstract*—We present a solution to the problem of spatio-temporal calibration for event cameras mounted on an onmi-directional vehicle. Different from traditional methods that typically determine the camera's pose with respect to the vehicle's body frame using alignment of trajectories, our approach leverages the kinematic correlation of two sets of linear velocity estimates from event data and wheel odometers, respectively. The overall calibration task consists of estimating the underlying temporal offset between the two heterogeneous sensors, and furthermore, recovering the extrinsic rotation that defines the linear relationship between the two sets of velocity estimates. The optimal temporal offset is determined by maximizing a correlation measurement invariant to arbitrary linear transformation. Once the temporal offset is compensated, the extrinsic rotation can be worked out with an iterative closed-form solver that incrementally registers associated linear velocity estimates. The proposed algorithm is proved effective on both synthetic data and real data, outperforming traditional methods based on alignment of trajectories.**

***Index Terms*—Calibration and identification, SLAM, event-based vision.**

I. Introduction

EXTRINSIC calibration is a prerequisite to almost any mobile robot application, because measurements from different sensor modalities are sometimes processed and even fused into a unified coordinate system, such as a robot's body frame. For an autonomous ground vehicle equipped with cameras, extrinsic calibration refers to the operation that determines each camera's mounting position and orientation with respect to the vehicle's body frame. Existing solutions designed for standard cameras typically run a pipeline based on alignment of trajectories, which estimates the extrinsic parameters by registering two trajectories recovered from vision information and wheel odometers, respectively. This widely used approach is, however, inapplicable when the recovered trajectory from the vision end is unreliable.

Different from its standard counterpart, an event camera is a biologically-inspired novel sensor which reports only brightness changes asynchronously. This unique characteristic leads to better performance in terms of temporal resolution and dynamic range. Thus, event cameras are suitable for dealing with robotic perception [1], [2], [3], [4], localization [5], [6], [7], [8] and control [9], [10] tasks involving aggressive motion and high-dynamic-range (HDR) illumination conditions. However, existing image processing techniques designed for standard vision cannot be applied straightforwardly to event data due to the special output format. Specifically, mature feature detection and matching techniques, based on which reliable and long-term event data association is to be established and maintained, are lacking. Hence, techniques that can effectively eliminate accumulated errors in the recovered trajectory, e.g., local bundle adjustment (BA) [11], are still not available in event-based visual odometry. Consequently, the drifted trajectory will lead to inaccurate extrinsic calibration results when applying trajectory alignment based pipelines. This issue poses a challenge when trying to extrinsically calibrate an event camera to an acceptable level.

In this letter, we look into the problem of extrinsic calibration of event cameras mounted on an omni-directional mobile platform, as shown in Fig. 1. Omni-directional vehicles, including non-holonomic all-wheel-steering vehicles and holonomic omni-directional vehicles, are mobility systems that are capable of moving in arbitrary direction. Such a flexible and potentially aggressive maneuverability leads to a great need for event-based vision. To circumvent the issue of inaccurate trajectory estimates from an event-based visual odometry, we propose a novel calibration method that recovers the temporal offset and the spatial extrinsics by exploiting the kinematic correlation between linear velocity estimates obtained from event data and wheel odometers, respectively. The core of our method is a Canonical Correlation Analysis (CCA) process, which evaluates the underlying linear correlation between the two sets of linear velocity estimates. The applied trace correlation measurement is invariant to arbitrary linear transformation, and thus, the temporal offset can be recovered by maximizing the correlation measurement between the two sets of spatially non-aligned data. With the temporally-aligned data, the extrinsic rotation can be furthermore worked out with an iterative closed-form solver

Manuscript received 6 October 2023; accepted 6 January 2024. Date of publication 18 January 2024; date of current version 29 January 2024. This letter was recommended for publication by Associate Editor L. Zhang and Editor J. Civera upon evaluation of the reviewers' comments. This work was supported in part by the National Science Foundation of China under Grants U1913202, U22A2059, and 62203460, and in part by the Major Project of Natural Science Foundation of Hunan Province under Grant 2021JC0004. *(Xiao Li and Yi Zhou contributed equally to this work.) (Corresponding author: Huimin Lu.)*

Xiao Li, Ruibin Guo, Zongtan Zhou, and Huimin Lu are with the College of Intelligence Science and Technology, National University of Defense Technology, Changsha, Hunan 410073, China (e-mail: lx852357@outlook.com; guoruibin08@nudt.edu.cn; narcz@nudt.edu.cn; lhmnew@nudt.edu.cn).

Yi Zhou is with Neuromorphic Automation and Intelligence Lab (NAIL), School of Robotics, Hunan University, Changsha, Hunan 410012, China (e-mail: eeyzhou@hnu.edu.cn).

Xin Peng is with Motovis Intelligent Technologies (Shanghai) Company Ltd, Shanghai 51100, China (e-mail: pengxin1@shanghaitech.edu.cn).

Multimedia Material Code: https://github.com/esheroe/EvCalib.git.

Digital Object Identifier 10.1109/LRA.2024.3355765

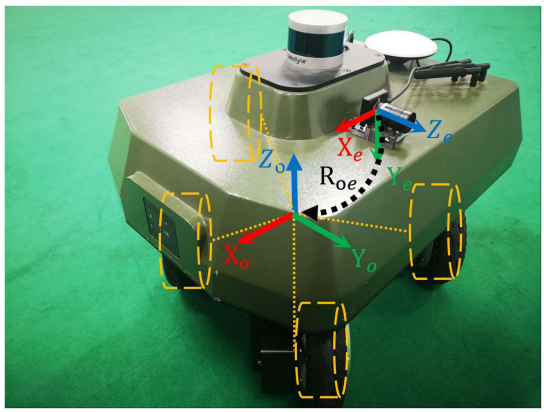


Fig. 1. Illustration of the geometry of the calibration problem. The goal is to recover the relative orientation ($\mathtt{R}_{oe}$) of the event camera with respect to the mobile platform body frame, while compensating the temporal offset t_d between the two heterogeneous sensors, i.e., the odometers and the event camera. The yellow dashed cylinders illustrate the four wheels. The coordinate system $\{X_oY_oZ_o\}$ and $\{X_eY_eZ_e\}$ refer to the body frame and the event camera frame, respectively. The GPS and LIDAR shown are not used in this project.

that incrementally registers associated linear velocity estimates. It is worth mentioning that our method does not rely on any specific tools, such as a LED array [12] and a chessboard [13], [14], indicating an application potential of online operation. The contribution of this letter consists of the following aspects:

- An efficient and robust method that determines the direction of linear velocity of a moving event camera, using as input the short-term feature correspondences obtained from the proposed speed-invariant image representation of event data.
- A rotation-free solution to the problem of spatio-temporal calibration for omni-directional vehicle mounted event cameras, with special consideration to the mobility system's kinematic characteristics and the limited accuracy of event-based visual odometry. We leverage the kinematic correlation between the linear velocity estimates from two heterogeneous sources, and build the solver on top of a CCA scheme.
- An extensive evaluation of the proposed method using both synthetic and real data, and an open-source implementation of our method.

The letter is organized as follows. A literature review on methods of cameras' extrinsic calibration is provided in Section II. The proposed method is detailed in Section III, followed with experimental evaluation and relevant analysis in Section IV. We draw the conclusion in Section V.

II. Related Work

A. Trajectory Alignment Based Methods

The main category of extrinsic calibration methods is hand-eye calibration [15], [16], [17], which is based on aligning multiple pose estimates (i.e., trajectories) from two independent sensors. The hand-eye calibration method establishes the extrinsic constraint using a closed loop of local transformations:

$$\mathbf{AX} = \mathbf{XB}, \tag{1}$$

where $\mathbf{A}$ and $\mathbf{B}$ represent, respectively, the poses of the two involved sensors at the same time instant, and $\mathbf{X}$ the unknown extrinsic parameters, all expressed in homogeneous form $\begin{pmatrix} \mathtt{R} & \mathbf{t} \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix}$. The hand-eye calibration pipeline typically consists of two steps: 1) Obtaining two sets of time-synchronized pose estimates; 2) Looking for the optimal extrinsic parameters that align maximally the two trajectories. Given time-synchronized poses $\xi_\mathtt{A} = \{\mathbf{A}_0, \mathbf{A}_1, \ldots, \mathbf{A}_N\}$ and $\xi_\mathtt{B} = \{\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_N\}$, the trajectory alignment task is accomplished by solving a non-linear optimization problem with an objective function:

$$\mathtt{R}_\mathtt{E}^*, \mathbf{t}_\mathtt{E}^* = \arg\min_{\mathtt{R}_\mathtt{E}, \mathbf{t}_\mathtt{E}} \sum_{i=0}^{N} \|\mathbf{A}_i\mathbf{X} - \mathbf{XB}_i\|, \tag{2}$$

where $\mathtt{R}_\mathtt{E}$ and $\mathbf{t}_\mathtt{E}$ represent the under estimated extrinsic parameters.

The hand-eye calibration method and its variants haven been witnessed in a lot of work that estimates the extrinsics between two heterogeneous sensors, especially those recovering the pose of an exteroceptive sensor with respect to the body frame of a robot. Censi et al. [18] utilize hand-eye calibration to calculate the extrinsic pose of a range finder with respect to a differential-drive robot's body frame. Guo et al. [19] unlock the limitation that original hand-eye calibration cannot recover all the three degrees of freedom (DoF) in the camera's orientation. Their method formulates a least-squares problem to estimate a subset of the odometer-camera rotation parameters, and furthermore, uses these parameters to formulate a second least-squares problem for estimating the remaining unknown parameters of the odometer-camera transformation. Besides, some globally optimal solutions based on hand-eye calibration, such as [20], [21], [22], are developed with special consideration in parametrization of motion parameters.

The accuracy of hand-eye calibration and its variants is largely up to the quality of the input sensor poses in terms of temporal synchronicity and spatial consistency. To improve temporal synchronicity between input signals, the underlying temporal offset between the two sensors needs to be considered as an additional variable to be optimized in (2). Consequently, the temporal offset and the extrinsic parameters are solved jointly, and thus, a proper initial value is needed. However, hand-eye calibration is not suitable for spatio-temporal calibration of event cameras, because state-of-the-art event-based visual odometry methods [5], [6], [7] cannot suppress drifts in the recovered trajectory effectively.

B. Motion Correlation Based Methods

Different from trajectory alignment based methods, motion correlation based methods can decouple the overall calibration task into sub problems of recovering temporal offset and spatial transformation. The sub problem of recovering temporal offset can be solved using cross correlation, which is widely used in signal processing [23], [24]. The optimal temporal offset δt^* can

be found by maximizing the following objective function:

$$\delta t^* = \arg\max_{\delta t} \sum_{i=0}^{N} {}^{a}\zeta(t_i + \delta t) \cdot {}^{b}\zeta(t_i), \tag{3}$$

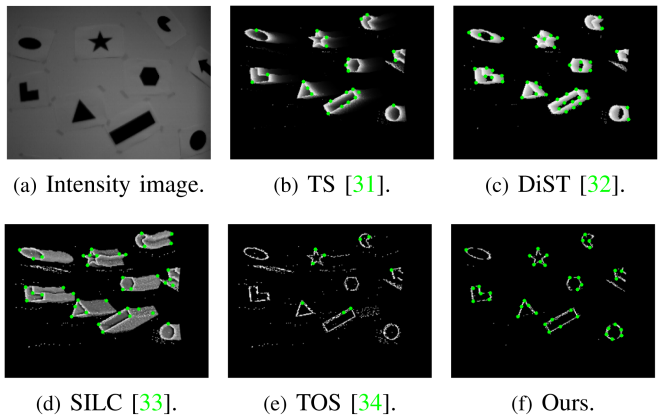
where ${}^{a}\zeta$ and ${}^{b}\zeta$ represent the identical 1-D signal measured by sensor a and b, respectively. This cross correlation can be established on any kinematic measurements (or estimates) independent of coordinate system, such as the absolute angular velocity, which has been used to solve the task of IMU-to-Camera temporal calibration [25]. The limitation of the cross-correlation pipeline (3) is that it is hardly extended to high-dimensional data that are frame dependent. Consequently, the recovered temporal offset is likely to be inaccurate when data are biased in the missing degrees of freedom. To overcome this limitation, Qiu et al. [26] propose a unified calibration framework based on 3-D motion correlation, which can efficiently work out the temporal offset and extrinsic rotation between two heterogeneous sensors. The method can evaluate the correlation of two sets of multivariate random vectors (e.g., 3-D angular velocity), because the applied CCA technique [27] is invariant to the underlying (unknown) linear transformation between the two kinematic signals. Although 3-D angular velocity can be estimated from event data [28], the 3-D motion correlation on angular velocity estimates is, however, inapplicable to our task, because one DoF of the 3D orientation is never observable. Fortunately, we could leverage the omni-directional locomotion property of holonomic vehicles that can generate pure translation in arbitrary directions (within the ground plane). Therefore, our method takes advantage of the correlation between linear velocity estimates from heterogeneous sensors and it is also built on top of the CCA scheme. To estimate the direction of linear velocity (heading direction) from event data, we determine the epipolar geometry using as input the feature correspondences obtained on a novel speed-invariant image representation of event data.

III. Methodology

Given two sets of linear velocity estimates from an event camera and odometers respectively, the goal is to recover the temporal offset between the two heterogeneous sensors, and furthermore, determine the orientation of the event camera with respect to the vehicle's body frame. In this section, we first discuss our method for estimating the direction of linear velocity from pure event data (Section III-A). Second, we demonstrate the way of acquiring body-frame velocity according to the kinematic model of the omni-directional vehicle used (Section III-B). Finally, we disclose the method that recovers the temporal offset and spatial extrinsics by maximizing the kinematic correlation between linear velocities estimates from heterogeneous sources (Section III-C).

A. Determining Heading Direction From Event Data

Assuming the ground is a perfect 2D horizontal plane, the vehicle can move ideally in a straight line under a constant steering command. Thus, the heading direction (i.e., direction of linear velocity) can be well approximated by the normalized translation vector between any two successive time instants. Although there exist several ways for relative pose estimation using event data as input, such as by fitting a homography [29], [30], they are typically computational expensive and require the camera to observe a planar scene. Instead, we leverage a feature-based method for relative pose estimation in standard vision, which calculates the essential matrix from a set of feature correspondences.



(a) Intensity image. (b) TS [31]. (c) DiST [32]. (d) SILC [33]. (e) TOS [34]. (f) Ours.

Fig. 2. Various image-like representations of event data [35] and results of corner detection [36]. (a) is the intensity image, which is only for visualization. (b)–(e) are state-of-the-art representations of event data. (f) shows our design.

Detection and matching of event-based features (e.g., corners) have to be carried out on image-like representation of event data. Assuming the photometric pattern of the environment is stable, the output of an event-based camera is dependent on the relative speed with respect to the scene. Thus, an ideal representation, on one hand, is supposed to be invariant to speed, and, on the other hand, can be rendered (refreshed) efficiently. To discover the optimal representation, we investigate a number of existing representations, including Time Surface (TS) [31], Discounted Sorted Timestamp Image (DiST) [32], Speed Invariant Time Surface (SILC) [33], and Threshold-Ordinal Surface (TOS) [34].

TS is a 2D representation where each pixel stores a single time value, e.g., the timestamp of the most recent event at that pixel [37]. With an additional exponential decay kernel [31] applied on the last spiking time t_{last} at each pixel coordinate $\mathbf{x} = (u, v)^{\mathrm{T}}$, a TS map at time t is defined as $\mathcal{T}_{\text{TS}}(\mathbf{x}, t) \doteq \exp(-\frac{t - t_{\text{last}}(\mathbf{x})}{\delta})$, where δ, the decay rate parameter, is a small constant number. As illustrated in Fig. 2(b), a TS map emphasizes recent events over past events and has been proved discriminative in pattern recognition tasks. Similarly, DiST [32] (Fig. 2(c)) aims at preserving semantic profiles for object recognition under the camera's motion, and it does a great job in noise suppression. These two representations are, however, not invariant to the event camera's speed. The appearance similarity, even at two close views, will not be preserved in the presence of speed variation, and thus, no guarantee for the success of short-baseline feature matching. To circumvent this issue, some hand-crafted representations (e.g., SILC [33] (Fig. 2(d)) and TOS [34] (Fig. 2(e))) are designed to keep the 2D spatial gradient of moving edges constant under a variation of camera's speed.

This is basically achieved by constantly assigning the most recently firing coordinates to the maximum value, and then reducing the magnitude of the adjacent area by either a distance-related quantity [33] or a threshold-based truncation [34]. Due to the limited space, we recommend interested readers find details of the rendering logic of the SILC and TOS in [33], [34]. The main issue of these two speed-invariant representations is about the relatively low signal-to-noise ratio, which is largely due to the fact that historical information is not recycled in time.

To hold the speed-invariant property while maintaining a good signal-to-noise ratio, we propose a novel representation by combining TS and TOS. Given a TS and a TOS rendered by the same time, our representation is specifically obtained by performing a logical AND operation on corresponding pixels of the two maps. As seen in Fig. 2(f), the resulting representation inherits the speed-invariant property from TOS; meanwhile it is much cleaner than the original TOS due to the exponential decay kernel used in the TS. We apply Arc* [36] for corner detection and BRIEF [38] for feature description and matching. As seen from Fig. 2(b) to (f), more true-positive corners can be detected on the resulting representation, and thus, it is beneficial to solving the essential matrix in the following.

To calculate the essential matrix, we implement a five-point algorithm [39] inside a RANSAC scheme. We notice that the essential matrix is skew symmetric due to the absence of rotation, and therefore, the resulting translation vector can be straightforwardly retrieved from it. Finally, the resulting heading direction from event data, denoted by $\mathbf{v}_e$, can be approximated by the normalized translation vector.

B. Acquiring Body-Frame Velocity From Kinematic Model

The ground vehicle used in this work is an all-wheel-steering mobility system as illustrated in Fig. 1. Different from commonly seen non-holonomic counterparts (e.g., Ackermann mobility systems), an all-wheel-steering vehicle is capable of moving in any direction by simply steering all wheels into a certain angle. Therefore, the direction of the linear velocity is not always along the X-axis of the body frame, and such a kinematic property enables us to obtain linear velocity measurements in various directions in the vehicle's body frame.

The vehicle's body frame is defined as the coordinate system in Fig. 3. Note that the location of the origin can be set anywhere, which does not affect the extrinsic rotation result. What matters is the definition of the body frame's orientation. The X-axis is defined to be with the longitudinal direction, and the Y-axis is the lateral direction. The remaining Z-axis can be obtained using the right-hand rule. The four wheels are actuated independently, and the corresponding odometer can report two states, i.e., the steering angle θ_i and the wheel speed v_{wi}. The linear velocity of the vehicle can be simply derived from

$$\mathbf{v}_o = \begin{bmatrix} \cos(\bar{\theta}) & -\sin(\bar{\theta}) & 0 \\ \sin(\bar{\theta}) & \cos(\bar{\theta}) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{v}_w \\ 0 \\ 0 \end{bmatrix}, \tag{4}$$

where $\mathbf{v}_o$ denotes the linear velocity (in metric scale) represented in the body frame o, and $\bar{v}_w$ and $\bar{\theta}$ the average speed and steering angle of the four wheels. Note that we use only the direction of linear velocity in the following extrinsic calibration. With a slight abuse of mathematical notation, for brevity, we hereafter denote the direction by $\mathbf{v}_o$.

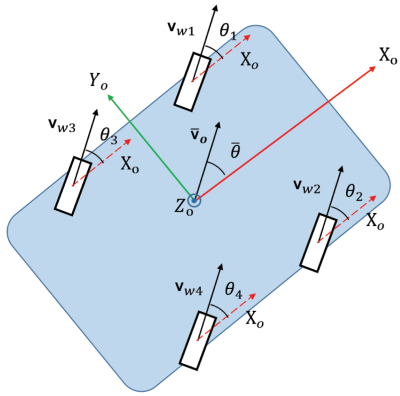


Fig. 3. Kinematic model of the all-wheel-steering mobile platform. Pure linear motion at arbitrary direction can be realized by controlling all wheels with same steering angle and identical rotating speed.

C. Extrinsic Calibration Via Correlation Maximization

The above velocity direction estimates, $\mathbf{v}_e$ and $\mathbf{v}_o$, are linearly correlated through the extrinsic rotation $\mathtt{R}_{oe}$, as

$$\mathbf{v}_o = \mathtt{R}_{oe}\mathbf{v}_e. \tag{5}$$

Obviously, the minimal problem to solve $\mathtt{R}_{oe}$ requires two pairs of velocity direction estimates, which must be not parallel. In the absence of temporal offset, the extrinsic rotation calibration can be simply solved as a 3-D point registration problem using the closed-form solution in [40]. However, simply neglecting the temporal offset may lead to inaccurate spatial calibration. Thus, we need a method to determine the temporal offset in the presence of unknown extrinsic rotation.

CCA is an effective tool for evaluating the linear correlation between two random vectors (i.e., discrete samples of two "synchronized" signals). It is typically used to recover the underlying linear transformation from the signals' statistical profiles, such as covariances. Assuming that we have collected two sets of velocity direction estimates, denoted by $\mathcal{V}_o \doteq \{\mathbf{v}_o(t_i)\}_{i=1}^N$ and $\mathcal{V}_e \doteq \{\mathbf{v}_e(t_i)\}_{i=1}^N$, the cross-covariance and auto-covariance can be approximately calculated, in the presence of temporal offset t_d, by

$$\begin{aligned}
\Sigma_{v_o v_e}(t_d) &\approx \frac{1}{N-1}\sum_{i=0}^{N-1}(\mathbf{v}_o(t_i)-\bar{\mathbf{v}}_o)(\mathbf{v}_e(t_i+t_d)-\bar{\mathbf{v}}'_e)^{\mathrm{T}}, \\
\Sigma_{v_o v_o}(t_d) &\approx \frac{1}{N-1}\sum_{i=0}^{N-1}(\mathbf{v}_o(t_i)-\bar{\mathbf{v}}_o)(\mathbf{v}_o(t_i+t_d)-\bar{\mathbf{v}}'_o)^{\mathrm{T}}, \\
\Sigma_{v_e v_e}(t_d) &\approx \frac{1}{N-1}\sum_{i=0}^{N-1}(\mathbf{v}_e(t_i)-\bar{\mathbf{v}}_e)(\mathbf{v}_e(t_i+t_d)-\bar{\mathbf{v}}'_e)^{\mathrm{T}},
\end{aligned} \tag{6}$$

where t_i refers to the sampling time, N the number of samples, $\bar{(\cdot)}$ the mean of original samples, and $\bar{(\cdot)}'$ the mean of temporal-offset compensated samples.

For multi-variable random vectors, the underlying linear transformation is made from several linear combination pairs. In our case, the linear combination pairs are $\mathbf{s}_i \leftrightarrow \mathbf{r}_i$ $(i = \{1, 2, 3\})$, which represent respectively the basis vectors of the 3-by-3 identity matrix and the transpose of the extrinsic rotation matrix, namely $\mathbf{I}_{3\times3} = [\mathbf{s}_1|\mathbf{s}_2|\mathbf{s}_3]$ and $\mathtt{R}_{oe}^{\mathtt{T}} = [\mathbf{r}_1|\mathbf{r}_2|\mathbf{r}_3]$. Each column vector $\mathbf{r}_i$ can be determined by maximizing the correlation coefficient

$$\begin{aligned}\rho_i &\doteq Corr(\mathbf{s}_i^{\mathtt{T}}\mathbf{v}_o, \mathbf{r}_i^{\mathtt{T}}\mathbf{v}_e)\\ &= \frac{\mathbf{s}_i^{\mathtt{T}}\Sigma_{v_ov_e}\mathbf{r}_i}{\sqrt{\mathbf{s}_i^{\mathtt{T}}\Sigma_{v_ov_o}\mathbf{s}_i}\sqrt{\mathbf{r}_i^{\mathtt{T}}\Sigma_{v_ev_e}\mathbf{r}_i}}, i \in \{1, 2, 3\}. \end{aligned} \tag{7}$$

The goal is identical to solving the following constrained optimization problem:

$$\begin{aligned}&\arg\min_{\mathbf{s},\mathbf{r}} \ -\mathbf{s}^{\mathtt{T}}\Sigma_{v_ov_e}\mathbf{r}\\ &s.t. \ \ \mathbf{s}^{\mathtt{T}}\Sigma_{v_ov_o}\mathbf{s} = 1, \mathbf{r}^{\mathtt{T}}\Sigma_{v_ev_e}\mathbf{r} = 1. \end{aligned} \tag{8}$$

By introducing non-negative Lagrange multipliers that add the constraints to the original objective function, we obtain its Lagrangian:

$$\mathcal{L}(\mathbf{s}, \mathbf{r}) = -\mathbf{s}^{\mathtt{T}}\Sigma_{v_ov_e}\mathbf{r} + \lambda(\mathbf{s}^{\mathtt{T}}\Sigma_{v_ov_o}\mathbf{s} - 1) + \theta(\mathbf{r}^{\mathtt{T}}\Sigma_{v_ev_e}\mathbf{r} - 1), \tag{9}$$

where λ and θ are referred to as Lagrange multipliers. Let the partial derivative of (9) w.r.t $\mathbf{s}$ and $\mathbf{r}$ be equal to zero, respectively,

$$\frac{\partial\mathcal{L}(\mathbf{s}, \mathbf{r})}{\partial\mathbf{s}} = \Sigma_{v_ov_e}\mathbf{r} - \lambda\Sigma_{v_ov_o}\mathbf{s} = 0, \tag{10a}$$

$$\frac{\partial\mathcal{L}(\mathbf{s}, \mathbf{r})}{\partial\mathbf{r}} = \Sigma_{v_ov_e}\mathbf{s} - \theta\Sigma_{v_ev_e}\mathbf{r} = 0, \tag{10b}$$

and also take into account the two constraints in (8), we can obtain

$$\lambda = \theta^{\mathtt{T}} = \mathbf{s}^{\mathtt{T}}\Sigma_{v_ov_e}\mathbf{r}. \tag{11}$$

Substituting (11) into (10), and further eliminating $\mathbf{r}$, we have

$$\Sigma_{v_ov_o}^{-1}\Sigma_{v_ov_e}\Sigma_{v_ev_e}^{-1}\Sigma_{v_ev_o}\mathbf{s} = \lambda^2\mathbf{s}. \tag{12}$$

Since λ^2 is the eigenvalue of (12), and also because of $\lambda_i = \rho_i$, we obtain

$$\sum_{i=1}^{3}\rho_i^2 = \sum_{i=1}^{3}\lambda_i^2 = \mathrm{Tr}(\Sigma_{v_ov_o}^{-1}\Sigma_{v_ov_e}\Sigma_{v_ev_e}^{-1}\Sigma_{v_ev_o}), \tag{13}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of an input matrix. With the three canonical correlation coefficients, the so-called trace correlation between the two set of estimates is further defined as

$$r(\mathcal{V}_o, \mathcal{V}_e) \doteq \sqrt{\frac{1}{3}\sum\nolimits_{i=1}^{3}\rho_i^2} = \sqrt{\frac{1}{3}\mathrm{Tr}(\Sigma_{v_ov_o}^{-1}\Sigma_{v_ov_e}\Sigma_{v_ev_e}^{-1}\Sigma_{v_ev_o})}. \tag{14}$$

It is also a normalized measurement that evaluates the correlation between two signals. One of the great properties of trace correlation is that it is invariant to underling linear transformation between two signals, such as scaling, rotation and translation. In our case, hence, we have $r(\mathcal{V}_o, \mathtt{R}_{oe}\mathcal{V}_e) = r(\mathcal{V}_o, \mathcal{V}_e)$. This property decouples the impact of the temporal offset on the trace correlation from the unknown spatial extrinics. Therefore, the optimal t_d can be independently obtained by maximizing (14). Due to the low computation complexity, we solve it through brute-force enumeration within a proper range. As shown in IV-C, it is very efficient and can preserve notably good precision.

Once the temporal offset is compensated in the data, the extrinsic rotation $\mathtt{R}_{oe}$ can be simply worked out as a by-product of the above CCA process [26]. Using the covariance matrices obtained in the CCA process, the extrinsic rotation can be calculated by

$$\mathtt{R}_{oe}^{\mathtt{T}} = \mathbf{U}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & \det(\mathbf{U}\mathbf{V}^{\mathtt{T}})\end{bmatrix}\mathbf{V}^{\mathtt{T}}, \tag{15}$$

where $\mathbf{U}$ and $\mathbf{V}$ are obtained from the following singular value decomposition (SVD)

$$\Sigma_{v_ev_e}^{-1}\Sigma_{v_ev_o} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathtt{T}}. \tag{16}$$

The determinant operation in (15) is to guarantee that the resulting rotation matrix is not a reflection. However, this quick solver cannot be straightforwardly applied to our case. The kinematic characteristics of the ground vehicle gives rise to zero-velocity measurements in z axis of the body frame. A big condition number is witnessed for $\Sigma_{v_ev_e}^{-1}\Sigma_{v_ev_o}$, and thus, the decomposition result is always numerically unstable. Although this issue can be more or less mitigated by adding an additional perturbation in the non-stimulated dimension, accurate results are still not guaranteed as seen in our experiments.

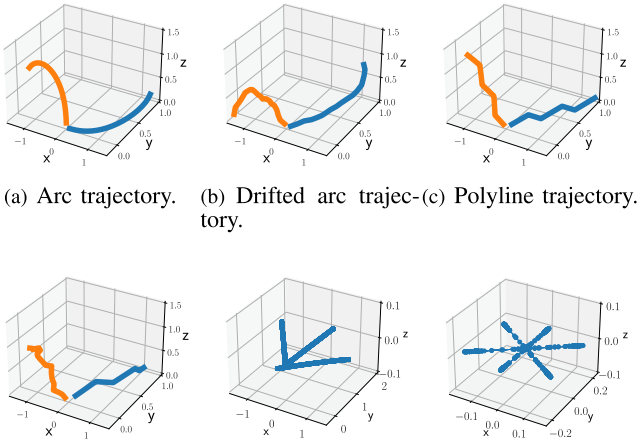
To this end, we regard the two sets of linear velocity estimates as general "point cloud" on a unit sphere, and assort to a least-squares method [40] for registering two point sets. Consequently, the extrinsic rotation can be calculated by maximizing

$$\begin{aligned}\mathcal{C} &= \sum_{i=1}^{N}\mathbf{v}_{o,t_i}^{T}\mathtt{R}_{oe}\mathbf{v}_{e,t_i}\\ &= \mathrm{Tr}\left(\sum_{i=1}^{N}\mathtt{R}_{oe}\mathbf{v}_{e,t_i}\mathbf{v}_{o,t_i}^{T}\right) = \mathrm{Tr}(\mathtt{R}_{oe}\mathbf{H}), \end{aligned} \tag{17}$$

where $\mathbf{v}_{o,t_i} \doteq \mathbf{v}_o(t_i)$, $\mathbf{v}_{e,t_i} \doteq \mathbf{v}_e(t_i)$, and $\mathbf{H} := \sum_{i=1}^{N}\mathbf{v}_{e,t_i}\mathbf{v}_{o,t_i}^{T}$. Let the SVD of $\mathbf{H}$ be $\mathbf{U}\mathbf{\Lambda}\mathbf{V}^{T}$. The resulting extrinsic rotation matrix is given as $\mathtt{R}_{oe} = \mathbf{V}\mathbf{U}^{T}$. To deal with noise and outliers, we replaces the original sum of squares with sum of absolute values. The resulting problem can be solved efficiently using an iteratively re-weighted least square method.

IV. Experiments

In this section, we evaluate the proposed calibration method. We first show that how synthetic data and real data for evaluation are generated and collected in Section IV-A and IV-B, respectively. Then we quantitatively evaluate the proposed method and compare against trajectory-alignment based pipelines in



(a) Arc trajectory. (b) Drifted arc trajectory. (c) Polyline trajectory.

(d) Drifted polyline trajectory. (e) Finger-shape trajectory. (f) Linear velocities

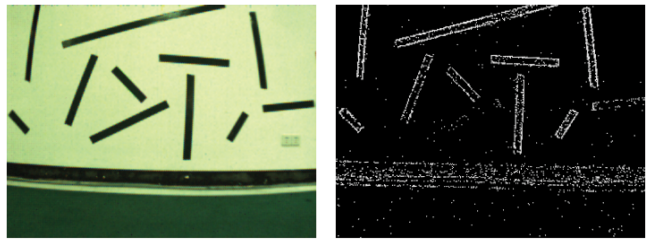
Fig. 4. *Illustration of the synthetic and real data.* (a)–(d) are simulation of perfect trajectories and corresponding drifted ones. Trajectories from the two heterogeneous sensors are expressed in a unified coordinate system. Different colors are used to distinguish trajectories from heterogeneous sources. (e) and (f) Show an example of finger-shape trajectories carried out in the collection of real data. (e) The trajectory of three "fingers." (f) Corresponding linear velocity measurements.

Section IV-C, demonstrating the advantage and efficacy of our method.

A. Generation of Synthetic Data

In order to demonstrate the negative impact of drifted trajectories on trajectory-alignment based calibration methods, we generate two types of input trajectories to be aligned, which include an arc-shape trajectory (Fig. 4(a)) and a polyline-shape trajectory (Fig. 4(c)). The perfect arc-shape trajectory can be determined solely by one parameter, i.e., the radius of the big circle. The direction of instantaneous linear velocity at any way point on the arc can be determined as the tangential direction, while its magnitude is determined by the instantaneous angular velocity. Without loss of generality, we simply use a constant angular velocity in each trial. As for the perfect polyline-shape trajectory, the design parameters consist of the turning angle and the length of each segment. Still, we assume these two parameters are constant for simplicity. Therefore, the direction of instantaneous linear velocity at any way point (except for those corners) is identical to that of the corresponding segment, and its magnitude is set with a constant number. To simulate spatial drifts in the trajectories to be aligned (see Fig. 4(b) and (d)), we add Gaussian noises to the perfect linear velocities and obtain the drifted trajectories by dead reckoning according to the following simple kinematics:

$$\begin{aligned}
\hat{\mathbf{v}}_{o,i} &= \mathbf{v}_{o,i} + \boldsymbol{\epsilon}_o, \\
\hat{\mathbf{v}}_{e,i} &= \mathtt{R}_{\mathtt{E}}\mathbf{v}_{o,i} + \boldsymbol{\epsilon}_e, \\
\mathbf{t}_{o,i} &= \mathbf{t}_{o,i-1} + \hat{\mathbf{v}}_{o,i-1} \cdot dt, \\
\mathbf{t}_{e,i} &= \mathbf{t}_{e,i-1} + \hat{\mathbf{v}}_{e,i-1} \cdot dt,
\end{aligned} \tag{18}$$



(a) Image of the scene. (b) Image-like representation.

Fig. 5. Illustration of visual information in the real data. (a) A sample RGB image of the scene for visualization only. (b) Our image-like representation rendered by the same time using events collected.

where $\boldsymbol{\epsilon}_o$ and $\boldsymbol{\epsilon}_e$ denote the added Gaussian noises, and $\mathtt{R}_{\mathtt{E}}$ the groundtruth extrinsic rotation. To simulate temporal offset between data, we furthermore slide temporally one drifted trajectory by a certain time length with respect to the other. As a result, we have temporally non-aligned and spatially perturbed kinematic information (i.e., linear velocities and way points) from two heterogeneous sources. For extensive evaluation, we generate six groups of data for each type of trajectory with different design parameters, and in each group we propose 100 trials with random temporal offset and extrinsic rotation.

B. Collection of Real Data

We also collect real data using an event camera mounted on a non-holonomic all-wheel-steering mobile platform. By periodically adjusting the steering angle of all the wheels, we can obtain finger-shape trajectories which are friendly to our calibration method. An example of three-finger-shape trajectory and the corresponding linear velocity information are illustrated in Fig. 4(e) and (f). The measurement of linear velocity (i.e., $\mathbf{v}_o$) can be obtained according to (4) using as input the control commands of steering angle and wheel speed. We collect four groups of data with various number of "fingers", aiming to show that the more directions the data contain the more accurate the calibration result. An illustration of visual data is shown in Fig 5. As for the groundtruth extrinsic rotation, we manage to set the event camera at a desired orientation with the help of two independent laser level meters.

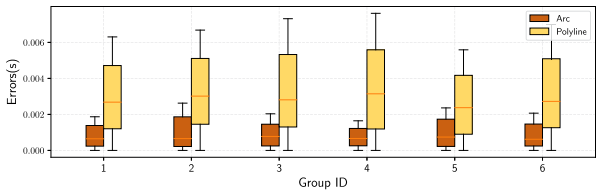
C. Evaluation Metrics and Results

We evaluate our method using the above datasets, and compare against existing pipelines listed in the following:

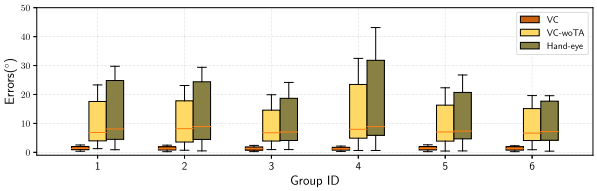
- CGOC: A globally optimal solution method using quadratically constrained quadratic programs (QCQPs) proposed in [22].
- Hand-eye: An implementation based on (2).
- VC: The proposed velocity-correlation based method.
- VC-woTA: The proposed method without temporal alignment.

The metric used for quantitative evaluation of extrinsic rotation is defined as

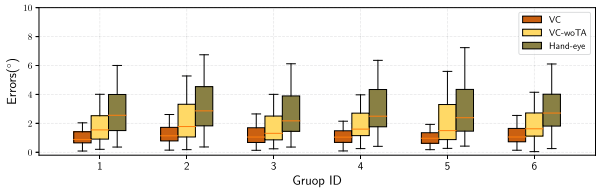
$$e = arccos((\mathrm{Tr}(\mathtt{R}_g^T\mathtt{R}_{est}) - 1)/2) * 180/\pi, \tag{19}$$



(a) Temporal-offset estimation results on simulated data.



(b) Results on simulated arc-shape trajectory.



(c) Results on simulated polyline-shape trajectory.

Fig. 6. Evaluation on synthetic data.

where $\mathtt{R}_{\mathtt{est}}$ and $\mathtt{R}_g$ are the estimation result and the groundtruth orientation, respectively.

Using the synthetic dataset, we first independently evaluate our temporal calibration method. As seen in Fig. 6(a), our method is successful on all sequences, achieving millisecond-level accuracy. To assess the computational performance of the temporal calibration, we run the algorithm (implemented with Python) using a PC equipped with an Intel i7-6700HQ CPU. The input data consists of 1 k pairs of velocity estimates, and the searching is performed within a proper range, e.g., $[-0.5, 0.5]$ s. With a step length of 1 ms, the average run time is 1 s, demonstrating that the temporal calibration can be performed online.

Then we calculate the spatial calibration error according to (19) and illustrate the error's statistics using box plot. As seen in Fig. 6(b) and (c), our methods (both VC and VC-woTA) outperform the Hand-eye calibration method in terms of accuracy. This result justifies our point: Trajectory-alignment based methods suffer from drifted trajectories while methods based on correlation of instantaneous first-order kinematics (e.g., linear velocity in our case) can work. Moreover, we clearly observe that VC outperforms VC-woTA. This indicates that temporal calibration is vital in the calibration task involving heterogeneous sensors. Besides, we find that the calibration results of the polyline-shape dataset are typically more accurate than those of the arc-shape dataset. This is because the temporal offset has less effect on the polyline-shape dataset than on the arc-shape dataset.

We also evaluate our method on real data. The direction of linear velocity from event data is obtained using our method

TABLE I
CALIBRATION ERRORS (°) ON REAL DATA

#Fingers	CGOC [22]	Hand-eye	VC-woTA	VC
2	86.22	35.59	6.29	**5.97**
3	38.98	5.82	2.02	**1.98**
4	7.96	9.19	2.11	**2.01**
5	6.21	7.20	1.82	**1.68**

TABLE II
CALIBRATION ERRORS (°) OF OUR METHOD USING DIFFERENT EVENT REPRESENTATIONS AS INPUT

#Fingers	TS	SILC	TOS	DisT	Ours
5	2.97	3.20	fail	4.76	**1.68**

discussed in Section III-A. Meanwhile, the linear velocity from odometers is obtained as shown in Section III-B, and the trajectory is obtained by dead reckoning. Since the estimates from the heterogeneous sensors are typically not synchronized, we obtain data association (synchronized and temporally-non-aligned) by data interpolation. Compared to the evaluation on synthetic dataset, we introduce another trajectory-alignment based solution [22] as a counterpart, which can return globally optimal extrinsic estimates. The trajectory used is estimated using EVO [6], a state-of-the-art visual odometry pipeline for a monocular event camera. As seen in Table I, the best results are highlighted in bold, and the conclusion is consistent with that in the evaluation on synthetic data. Our methods (VC and VC-woTA) outperform trajectory-alignment based methods, and more specifically, VC outperforms VC-woTA. Besides, we see a clear trend in the last column of Table I that the diversity of linear velocity direction (i.e., the number of "fingers") would promote the accuracy of calibration results. This is due to the fact that the more directions are considered, the more distinctive the kinematic profile becomes.

Furthermore, an ablation study on the choice of event representation is performed. To compare previously-mentioned five event representations, we use them as input respectively and run our method on the last sequence of the real data. As seen in Table II, using TOS alone leads to calibration failure. This is due to the unsatisfactory result of feature detection and matching, which in turn leads to inaccurate heading direction estimation result. Our choice (TOS+TS) outperforms the others in terms of calibration error.

V. CONCLUSION

This letter provides a novel solution to the problem of spatio-temporal calibration for omni-directional vehicle mounted event cameras. We argue that trajectory-alignment based methods suffer from drifts in any of the input trajectories. To this end, we propose a two-step method that establishes correlation on first-order dynamics, namely instantaneous linear velocity. In the first step, the optimal temporal offset is estimated by maximizing a correlation measurement invariant to the unknown extrinsic rotation. In the second step, we regard directions of linear velocity estimates as general point clouds on a unit sphere and

calculate the extrinsic rotation matrix via a point cloud registration process, which is accurately and robustly solved using an iteratively re-weighted least-squares method. Experiments on both synthetic data and real data demonstrate the efficacy of the proposed calibration method. Finally, we hope this work inspires new research in the topic of multi-sensor calibration that involves event-based cameras.

REFERENCES

[1] H. Rebecq, G. Gallego, E. Mueggler, and D. Scaramuzza, "EMVS: Event-based multi-view stereo-3D reconstruction with an event camera in real-time," *Int. J. Comput. Vis.*, vol. 126, no. 12, pp. 1394–1414, Dec. 2018.

[2] Y. Zhou, G. Gallego, H. Rebecq, L. Kneip, H. Li, and D. Scaramuzza, "Semi-dense 3D reconstruction with a stereo event camera," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 242–258.

[3] T. Stoffregen, G. Gallego, T. Drummond, L. Kleeman, and D. Scaramuzza, "Event-based motion segmentation by motion compensation," in *Proc. Int. Conf. Comput. Vis.*, 2019, pp. 7244–7253.

[4] Y. Zhou, G. Gallego, X. Lu, S. Liu, and S. Shen, "Event-based motion segmentation with spatio-temporal graph cuts," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 8, pp. 4868–4880, Aug. 2023.

[5] H. Kim, S. Leutenegger, and A. J. Davison, "Real-time 3D reconstruction and 6-DoF tracking with an event camera," in *Proc. 14th Eur. Conf. Comput. Vis.*, 2016, pp. 349–364.

[6] 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 Robot. Autom. Lett.*, vol. 2, no. 2, pp. 593–600, Apr. 2017.

[7] Y. Zhou, G. Gallego, and S. Shen, "Event-based stereo visual odometry," *IEEE Trans. Robot.*, vol. 37, no. 5, pp. 1433–1450, Oct. 2021.

[8] A. Safa et al., "Fusing event-based camera and radar for slam using spiking neural networks with continual STDP learning," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2023, pp. 2782–2788.

[9] J. Conradt, M. Cook, R. Berner, P. Lichtsteiner, R. J. Douglas, and T. Delbruck, "A pencil balancing robot using a pair of AER dynamic vision sensors," in *Proc. IEEE Int. Symp. Circuits Syst*, 2009, pp. 781–784.

[10] T. Delbruck and M. Lang, "Robotic goalie with 3 ms reaction time at 4% CPU load using event-based dynamic vision sensor," *Front. Neurosci.*, vol. 7, 2013, Art. no. 223.

[11] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment–a modern synthesis," in *Proc. Vis. Algorithms Theory Pract. Int. Workshop Vis. Algorithms Corfu*, 2000, pp. 298–372.

[12] M. J. Domínguez-Morales, Á Jiménez-Fernández, G. Jiménez-Moreno, C. Conde, E. Cabello, and A. Linares-Barranco, "Bio-inspired stereo vision calibration for dynamic vision sensors," *IEEE Access*, vol. 7, pp. 138415–138425, 2019.

[13] K. Huang, Y. Wang, and L. Kneip, "Dynamic event camera calibration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2021, pp. 7021–7028.

[14] M. Muglikar, M. Gehrig, D. Gehrig, and D. Scaramuzza, "How to calibrate your event camera," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 1403–1409.

[15] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration," *IEEE Trans. Robot.*, vol. 5, no. 3, pp. 345–358, Jun. 1989.

[16] R. Horaud and F. Dornaika, "Hand-eye calibration," *Int. J. Robot. Res.*, vol. 14, no. 3, pp. 195–210, 1995.

[17] K. H. Strobl and G. Hirzinger, "Optimal hand-eye calibration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2006, pp. 4647–4653.

[18] A. Censi, A. Franchi, L. Marchionni, and G. Oriolo, "Simultaneous calibration of odometry and sensor parameters for mobile robots," *IEEE Trans. Robot.*, vol. 29, no. 2, pp. 475–492, Apr. 2013.

[19] C. X. Guo, F. M. Mirzaei, and S. I. Roumeliotis, "An analytical least-squares solution to the odometer-camera extrinsic calibration problem," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2012, pp. 3962–3968.

[20] T. Wodtko, M. Horn, M. Buchholz, and K. Dietmayer, "Globally optimal multi-scale monocular hand-eye calibration using dual quaternions," in *Proc. IEEE 2021 Int. Conf. 3D Vis.*, 2021, pp. 249–257.

[21] M. Horn, T. Wodtko, M. Buchholz, and K. Dietmayer, "Online extrinsic calibration based on per-sensor ego-motion using dual quaternions," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 982–989, Apr. 2021.

[22] M. Giamou, Z. Ma, V. Peretroukhin, and J. Kelly, "Certifiably globally optimal extrinsic calibration from per-sensor egomotion," *IEEE Robot. Automat. Lett.*, vol. 4, no. 2, pp. 367–374, Apr. 2019.

[23] M. Azaria and D. Hertz, "Time delay estimation by generalized cross correlation methods," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 2, pp. 280–285, Apr. 1984.

[24] A. Fertner and A. Sjolund, "Comparison of various time delay estimation methods by computer simulation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1329–1330, Oct. 1986.

[25] E. Mair, M. Fleps, M. Suppa, and D. Burschka, "Spatio-temporal initialization for IMU to camera registration," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, 2011, pp. 557–564.

[26] K. Qiu, T. Qin, J. Pan, S. Liu, and S. Shen, "Real-time temporal and rotational calibration of heterogeneous sensors using motion correlation analysis," *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 587–602, Apr. 2021.

[27] B. Thompson, "Canonical correlation analysis," in *Encyclopedia Statistics Behavioral Science*, vol. 1, Wiley, 2005, pp. 192–196.

[28] G. Gallego and D. Scaramuzza, "Accurate angular velocity estimation with an event camera," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 632–639, Apr. 2017.

[29] G. Gallego, H. Rebecq, and D. Scaramuzza, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3867–3876.

[30] X. Huang, Y. Zhang, and Z. Xiong, "Progressive spatio-temporal alignment for efficient event-based motion estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 1537–1546.

[31] X. Lagorce, G. Orchard, F. Gallupi, B. E. Shi, and R. Benosman, "HOTS: A hierarchy of event-based time-surfaces for pattern recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1346–1359, Jul. 2017.

[32] J. Kim, J. Bae, G. Park, D. Zhang, and Y. M. Kim, "N-ImageNet: Towards robust, fine-grained object recognition with event cameras," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 2146–2156.

[33] J. Manderscheid, A. Sironi, N. Bourdis, D. Migliore, and V. Lepetit, "Speed invariant time surface for learning to detect corner points with event-based cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10237–10246.

[34] A. Glover, A. Dinale, L. D. S. Rosa, S. Bamford, and C. Bartolozzi, "luvHarris: A practical corner detector for event-cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 10087–10098, Dec. 2022.

[35] 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," *Int. J. Robot. Res.*, vol. 36, no. 2, pp. 142–149, 2017.

[36] I. Alzugaray and M. Chli, "Asynchronous corner detection and tracking for event cameras in real time," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3177–3184, Oct. 2018.

[37] T. Delbruck, "Frame-free dynamic digital vision," in *Proc. Int. Symp. Secure-Life Electron.*, 2008, pp. 21–26.

[38] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *Proc. 11th Eur. Conf. Comput. Vis. Comput. Vis.*, 2010, pp. 778–792.

[39] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 756–770, Jun. 2004.

[40] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 5, pp. 698–700, Sep. 1987.