A Spatiotemporal Hand-Eye Calibration for Trajectory Alignment in Visual(-Inertial) Odometry Evaluation

Zichao Shu[1], Lijun Li[1], Rui Wang[2] and Zetao Chen[1]

***Abstract*—A common prerequisite for evaluating a visual(-inertial) odometry (VO/VIO) algorithm is to align the timestamps and the reference frame of its estimated trajectory with a reference ground-truth derived from a system of superior precision, such as a motion capture system. The trajectory-based alignment, typically modeled as a classic hand-eye calibration, significantly influences the accuracy of evaluation metrics. However, traditional calibration methods are susceptible to the quality of the input poses. Few studies have taken this into account when evaluating VO/VIO trajectories that usually suffer from noise and drift. To fill this gap, we propose a novel spatiotemporal hand-eye calibration algorithm that fully leverages multiple constraints from screw theory for enhanced accuracy and robustness. Experimental results show that our algorithm has better performance and is less noise-prone than state-of-the-art methods.**

***Index Terms*—Calibration and Identification, Performance Evaluation and Benchmarking, Visual-Inertial SLAM.**

I. INTRODUCTION

VISUAL(-INERTIAL) odometry (VO/VIO) is known to provide state estimation of motion devices and has a wide range of application domains, such as robotics, extended reality, and autonomous driving. The performance evaluation is a fundamental task in VO/VIO research and application, where metrics are typically quantified by evaluating the estimated trajectory from VO/VIO with respect to the ground-truth. Commonly, the ground-truth trajectory can be obtained by tracking the motion device simultaneously with a system of superior precision, e.g., using a motion capture (MoCap) system, laser tracker, etc [1]–[3]. There are two main problems when comparing the estimated trajectory against the ground-truth: the trajectory pair is usually on different clock domains (thus with non-corresponding timestamps) and expressed in different global and local reference frames. While well-known methods such as the Umeyama algorithm [4] can align the global frames, the spatiotemporal alignment, which calculates the offsets of timestamps and local frames of the trajectory pair, still needs meticulous handling.

The spatiotemporal alignment problem above can be modeled as a classic hand-eye calibration problem: given the local frames of the ground-truth and estimated trajectory as the hand and the eye respectively, calculate the timestamp offset and estimate the homogeneous transformation between them. While the essence of hand-eye calibration problem has been well addressed in numerous studies [5]–[10], the accuracy and robustness may still be compromised in practical applications. The error introduced in this step will affect the transformation of the ground-truth trajectory, thereby exerting a substantial influence on the subsequent evaluation metrics.

A. Motivation

In this work, we consider the scenario in which only the trajectory information is available. This is common among commercial consumer devices, such as extended reality headsets or home robots, where the original raw sensor data used to derive the device trajectories are not accessible by users. Existing hand-eye calibration algorithms can be categorized into two distinct approaches: tightly-coupled and loosely-coupled. The former typically joints raw data from the sensors such as images with information of the calibration boards [6]–[8] or IMU measurements [1], [2], and optimizes the result in a maximum likelihood estimation (MLE) framework. The latter, on the other hand, directly calculates the offset between the hand and the eye based on their independently estimated poses [9], [10]. While tightly-coupled approaches can theoretically achieve higher accuracy and are used in well-known benchmarks such as EuRoC [1] and TUM-VI [2], they are not applicable in cases where only the trajectory information is available. Loosely-coupled approaches can perform calibration in the pose-only condition, but due to the ubiquitous noise and accumulated error in the VO/VIO estimation, the accuracy and robustness of existing methods are generally insufficient.

B. Contribution

In this letter, we propose a novel loosely-coupled spatiotemporal hand-eye calibration method tailored for VO/VIO evaluation. This method demonstrates robustness against noise and accumulated error in the input trajectories. For time alignment, we improve the correlation analysis of the screw invariant and obtain synchronized trajectories. For spatial calibration, we construct linear equations using local relative poses based on rotational constraint to fully utilize the motion information,

Manuscript received: December 12, 2023; Revised: March 7, 2024; Accepted: March 30, 2024.

This paper was recommended for publication by Editor Sven Behnke upon evaluation of the Associate Editor and Reviewers' comments.

[1]Zichao Shu, Lijun Li and Zetao Chen are with the Advanced Display and Sensing Research Center, Yongjiang Laboratory, China. {zichao-shu; lijun-li; zetao-chen}@ylab.ac.cn

[2]Rui Wang is with the Mixed Reality & AI Lab-Zürich, Microsoft, Switzerland. {wangr@microsoft.com}

Digital Object Identifier (DOI): see top of this page.

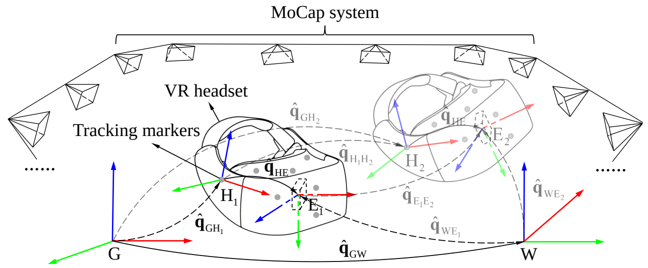
MoCap system
VR headset
Tracking markers
......
......
$\hat{\mathbf{q}}_{\mathrm{GH}_2}$
$\mathbf{q}_{\mathrm{HE}}$
$\mathbf{q}_{\mathrm{HE}}$
$\hat{\mathbf{q}}_{\mathrm{H}_1\mathrm{H}_2}$
H_2
E_2
H_1
E_1
$\hat{\mathbf{q}}_{\mathrm{E}_1\mathrm{E}_2}$
$\hat{\mathbf{q}}_{\mathrm{WE}_2}$
$\hat{\mathbf{q}}_{\mathrm{GH}_1}$
$\hat{\mathbf{q}}_{\mathrm{WE}_1}$
$\hat{\mathbf{q}}_{\mathrm{GW}}$
G
W

Fig. 1. Our spatiotemporal hand-eye calibration platform and the convention of the reference frames. The global frame for MoCap trajectory is denoted as G, and the local frame is referenced to a specific tracking marker indicated as H. The global frame for VIO trajectory is denoted as W, and the local frame E coincides with the IMU body frame. During this process, dashed lines represent transformations that change over time, while solid lines indicate static offsets.

rather than the naive global or inter-frame strategies [8]–[10]. Additionally, we introduce a well-designed robust kernel based on the screw theory to stabilize the linear solution. These operations are iteratively completed within the random sample consensus (RANSAC) framework to recover inlier data. Finally, we design a nonlinear optimization tool to jointly refine the time offset and the linear extrinsic solution. To validate the effectiveness of our algorithm, we conduct experiments on public and simulated datasets, as well as our own datasets collected by a virtual reality (VR) headset with VIO capability and a MoCap system (see in Fig. 1).

The rest of this letter is organized as follows. Section II reviews related work on spatiotemporal hand-eye calibration. The new method is described in Section III and its performance is evaluated in Section IV. Section V concludes the letter.

II. RELATED WORK

Spatiotemporal hand-eye calibration based on different strategies is a widely studied area. In our review, we briefly discuss the related work that shares the same strategy as the proposed method, i.e., the loosely-coupled methods, and motivate the design adapted in our work.

A. Spatial Hand-Eye Calibration

In our application scenario, the loosely-coupled hand-eye calibration can be formulated as $\mathbf{AX} = \mathbf{XB}$ [11], where $\mathbf{A}$ and $\mathbf{B}$ are the hand and the eye poses between two frames respectively, and $\mathbf{X}$ is the unknown homogeneous transformation between the hand and the eye. The solution for $\mathbf{X}$ can be categorized into two approaches: either separately or simultaneously solving the rotation and translation parts for the transformation.

One of the earliest separated approaches was presented by Shiu and Ahmad [11]. They represented rotation using angle-axis and proposed a closed-form solution for the $\mathbf{AX} = \mathbf{XB}$ formulation. Tsai and Lenz [12] proposed a similar but simplified method to improve computational efficiency, which has been widely used to this day (e.g. in OpenCV). Later methods focused on utilizing various parameterizations of the rotation, such as angle-axis [13], quaternion [14], Lie algebra [15], and Kronecker product [16], to achieve more efficient solution. Although separated methods are computationally efficient, they are error-prone due to the independence assumption between the rotation and translation, which are actually nonlinearly coupled [17].

In contrast, the simultaneous methods calibrate the rotation and translation parts jointly. Representative methods include screw motion [18], Kronecker product [19], dual quaternion [20] and dual tensor [21], which use alternative analytical parameterizations to express the complete homogeneous transformation. Additionally, there are also algorithms based on numerical optimization [22]–[24]. While the computational cost may increase, they are generally more accurate than the separated methods. In addition to improving accuracy, recent researchers have also focused on enhancing the robustness of algorithms across various applications [9], [10], [25], [26]. However, these efforts were mostly focused on scenarios characterized by comparatively high-precision pose data. This differs from our use case where the trajectories are substantially affected by noise or accumulated error. In this paper, we utilize a dual quaternion scheme similar to [10], [20] and construct a robust linear solving system using multiple constraints from screw theory to address the challenges from the VO/VIO trajectory.

B. Temporal Alignment

Given that the hand and the eye sensors usually operate on different clocks, the temporal correspondence between $\mathbf{A}$ and $\mathbf{B}$ is usually unknown. The time alignment prior to spatial calibration is thus necessary. Kelly and Sukhatme [27] considered the problem to be a registration task and solved it by utilizing the iterative closest point algorithm. Based on the discrete Fourier transform (DFT) theory, the time alignment of trajectories can be converted to the correlation analysis between two invariant signals derived from screw motion. This simple and effective method has been widely used in data synchronization for hand-eye calibration [10], [28]–[30]. However, the precision of this method is limited by the temporal resolution of the correlation function, with low-frequency data resulting in reduced time alignment accuracy. Alternatively, the TUM-VI benchmark [2] achieves time alignment by utilizing information from an error function, which is calculated through a grid search between the motion invariants. Additionally, a parabolic fitting is applied to the error function to enhance the precision of the time alignment. Unlike the grid search method in the time domain employed in TUM-VI, our approach builds upon the more commonly used DFT-based approach. To overcome the limitation posed by data frequency, we adopt a similar technique inspired by TUM-VI.

In some studies of continuous-time state estimation, by parameterizing the state variables as continuous-time functions, it is possible to achieve simultaneous spatiotemporal multi-sensor calibration within a MLE framework [31]–[33]. These high-precision methods can be easily extended to the pose-only scenario, but they are sensitive to the initial condition. In our work, we use the result of our linear calibration as the initial guess and perform a further refinement.

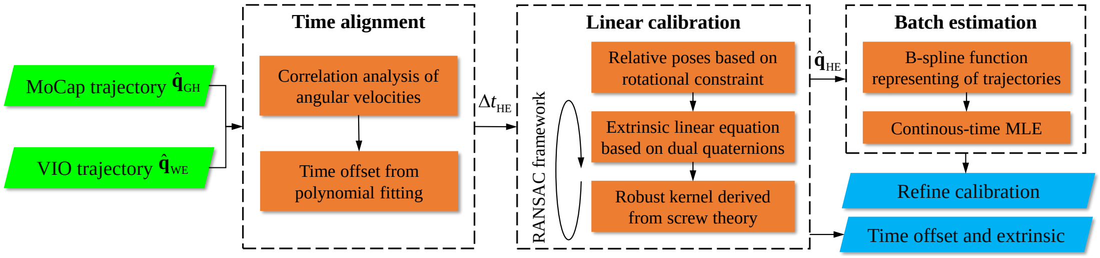
MoCap trajectory $\hat{\mathbf{q}}_{\mathrm{GH}}$
VIO trajectory $\hat{\mathbf{q}}_{\mathrm{WE}}$
Time alignment
Correlation analysis of angular velocities
Time offset from polynomial fitting
Δt_{HE}
Linear calibration
RANSAC framework
Relative poses based on rotational constraint
Extrinsic linear equation based on dual quaternions
Robust kernel derived from screw theory
$\hat{\mathbf{q}}_{\mathrm{HE}}$
Batch estimation
B-spline function representing of trajectories
Continous-time MLE
Refine calibration
Time offset and extrinsic

Fig. 2. Flowchart of the proposed spatiotemporal hand-eye calibration, where the green and blue parallelograms represent the inputs and outputs respectively, and the orange rectangles represent the critical processing steps.

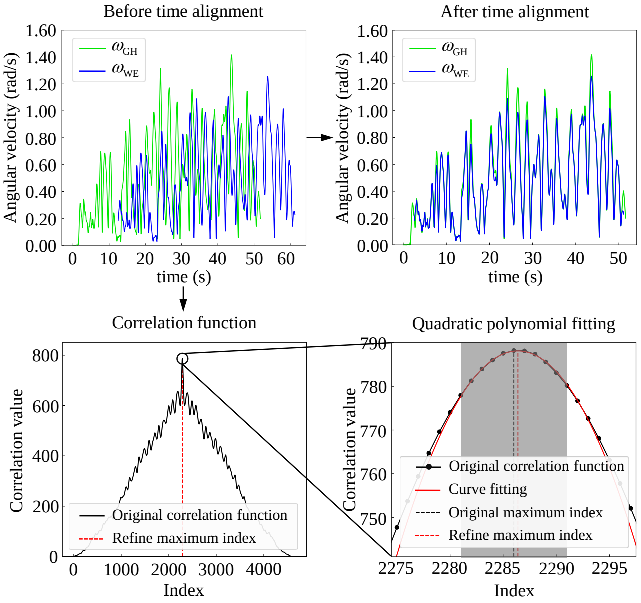

Fig. 3. Illustration of time alignment. The time offset can be determined by performing a quadratic polynomial curve fitting around the maximum (highlighted in gray) of the correlation function and obtaining the index of the maximum. This method enables synchronization of angular velocity at a finer granularity.

III. Methodology

To better illustrate our algorithm, we take our hand-eye calibration platform as an example (see in Fig. 1). The unit dual quaternions $\hat{\mathbf{q}} \in \mathbb{DQ}$ is used to represent the homogeneous transformation. A dual quaternion $\hat{\mathbf{q}}$ has the form $\hat{\mathbf{q}} = \mathbf{q} + \epsilon\mathbf{q}' = (\mathbf{q}, \mathbf{q}')$, where Hamiltonian quaternions $\mathbf{q}$ and $\mathbf{q}'$ are the standard part and the dual part of $\hat{\mathbf{q}}$ respectively, and ϵ is the infinitesimal unit satisfying $\epsilon^2 = 0$. Our goal is to estimate the time offset Δt_{HE} and the extrinsic $\hat{\mathbf{q}}_{\mathrm{HE}}$ between the MoCap (hand) trajectory $\hat{\mathbf{q}}_{\mathrm{GH}}$ and the estimated VIO (eye) trajectory $\hat{\mathbf{q}}_{\mathrm{WE}}$. Fig. 2 provides an overview of the proposed algorithm, which comprises three modules. The main steps of the algorithm will be described in the remaining of this section.

A. Time Alignment

In order to process poses from sensors with different clocks, the first crucial step is to align the two sets of timestamps. To achieve this, we can leverage the constraint from screw motion, i.e., based on the equality of the angular velocities ω_{GH} and ω_{WE} of trajectories $\hat{\mathbf{q}}_{\mathrm{GH}}$ and $\hat{\mathbf{q}}_{\mathrm{WE}}$, which is independent of calibration parameters. This simplifies the time alignment to the synchronization of angular velocity signals. Based on theory of DFT, the correlation between two time domain signals is greater when they exhibit higher similarity. Therefore, the synchronization involves finding the time shift, τ_{shift}, where the correlation function reaches its maximum:

$$\tau_{\text{shift}} = \underset{\text{index}}{\operatorname{argmax}} \left(\operatorname{Corr} \left(\omega_{\mathrm{GH}}, \omega_{\mathrm{WE}} \right) \right), \tag{1}$$

where $\operatorname{Corr}(\cdot)$ is the correlation function.

Given the discrete character of the angular velocity signals, the correlation function also appears discrete in the time domain, with the precision of the obtained time shift depending on the temporal resolution of the function. Assuming the data around the maximum follows a quadratic polynomial distribution, we perform curve fitting to refine our time shift similar to [2]. As illustrated in Fig. 3, this approach allows us to accurately determine the index of maximum correlation and trace it back to the corresponding time offset, Δt_{HE}.

B. Linear Calibration

Given the time aligned trajectories $\hat{\mathbf{q}}_{\mathrm{GH}}$ and $\hat{\mathbf{q}}_{\mathrm{WE}}$, we can perform spatial hand-eye calibration. As shown in Fig. 1, for any hand-eye motion from i to j in the trajectories, we have:

$$\hat{\mathbf{q}}_{\mathrm{GH}_i}\hat{\mathbf{q}}_{\mathrm{HE}}\hat{\mathbf{q}}_{\mathrm{WE}_i}^{-1} = \hat{\mathbf{q}}_{\mathrm{GH}_j}\hat{\mathbf{q}}_{\mathrm{HE}}\hat{\mathbf{q}}_{\mathrm{WE}_j}^{-1}. \tag{2}$$

Using the relative transformation between two poses, i.e., $\hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j} = \hat{\mathbf{q}}_{\mathrm{GH}_i}^{-1}\hat{\mathbf{q}}_{\mathrm{GH}_j}$ and $\hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j} = \hat{\mathbf{q}}_{\mathrm{WE}_i}^{-1}\hat{\mathbf{q}}_{\mathrm{WE}_j}$, (2) can be rewritten in the form of **AX = XB** as $\hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j}\hat{\mathbf{q}}_{\mathrm{HE}} = \hat{\mathbf{q}}_{\mathrm{HE}}\hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}$. This fundamental equation can be divided into the standard and dual parts, yielding:

$$\begin{aligned} \mathbf{q}_{\mathrm{H}_i\mathrm{H}_j}\mathbf{q}_{\mathrm{HE}} - \mathbf{q}_{\mathrm{HE}}\mathbf{q}_{\mathrm{E}_i\mathrm{E}_j} &= 0, \\ \mathbf{q}'_{\mathrm{H}_i\mathrm{H}_j}\mathbf{q}_{\mathrm{HE}} - \mathbf{q}_{\mathrm{HE}}\mathbf{q}'_{\mathrm{E}_i\mathrm{E}_j} + \mathbf{q}_{\mathrm{H}_i\mathrm{H}_j}\mathbf{q}'_{\mathrm{HE}} - \mathbf{q}'_{\mathrm{HE}}\mathbf{q}_{\mathrm{E}_i\mathrm{E}_j} &= 0. \end{aligned} \tag{3}$$

Due to the redundancy of the scalar part of the dual quaternion, we set $\mathbf{r} = \left(\mathbf{q}_{\mathrm{H}_i\mathrm{H}_j}\right)_v$, $\mathbf{r}' = \left(\mathbf{q}'_{\mathrm{H}_i\mathrm{H}_j}\right)_v$, $\mathbf{s} = \left(\mathbf{q}_{\mathrm{E}_i\mathrm{E}_j}\right)_v$ and $\mathbf{s}' = \left(\mathbf{q}'_{\mathrm{E}_i\mathrm{E}_j}\right)_v$, where $(\cdot)_v$ denotes the vector part of the quaternion. The linear equation for extrinsic calibration derived from a single motion can be written as:

$$\begin{bmatrix} \mathbf{r}-\mathbf{s} & (\mathbf{r}+\mathbf{s})^\wedge & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} \\ \mathbf{r}'-\mathbf{s}' & (\mathbf{r}'+\mathbf{s}')^\wedge & \mathbf{r}-\mathbf{s} & (\mathbf{r}+\mathbf{s})^\wedge \end{bmatrix} \begin{bmatrix} \mathbf{q}_{\mathrm{HE}} \\ \mathbf{q}'_{\mathrm{HE}} \end{bmatrix} = 0, \tag{4}$$

where $(\cdot)^\wedge$ denotes the antisymmetric matrix of a vector, and the coefficient matrix with dimensions of 6×8 will be denoted

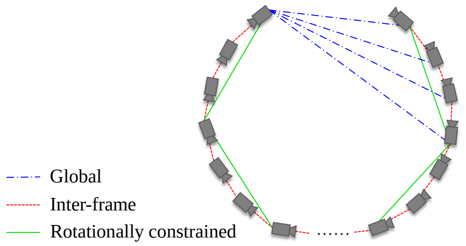

Fig. 4. Illustration of relative poses construction, different methods are represented by three different lines, and ours shown as the solid green line.

as $\mathbf{S}$. With $n \geq 2$ motions, we can stack $\mathbf{S}$ to obtain a $6n \times 8$ matrix with rank 6 in the noise-free case as:

$$\mathbf{M} = \left[\mathbf{S}_1^{\mathrm{T}}, \mathbf{S}_2^{\mathrm{T}}, \ldots, \mathbf{S}_n^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{5}$$

The singular value decomposition (SVD) algorithm is then used to find the linear least squares solution of the extrinsic with the constraint of the unit dual quaternion. For more detailed information about the fundamental principles of dual quaternion-based hand-eye calibration, please refer to [20].

SVD is known to be sensitive to noise and outliers. In the following, we will propose three strategies to enhance the accuracy and robustness of the algorithm.

1) Relative poses construction: As shown in (4), the equation entirely relies on the relative poses of the hand-eye motion, therefore, the relative poses construction method will significantly affect the quality of the solution. Conventional methods either use a global or an inter-frame strategy. The former fixes a certain frame and calculates the relative poses of the remaining frames with respect to it. However, this method is prone to coupling the trajectory drift in the VO/VIO scenario. The latter calculates the relative poses between two successive frames, but may suffer from the noise caused by insufficient motion. Moreover, for a relative pose with pure translation, the matrix in (4) will degenerate and cannot constrain the dual part of the extrinsic.

We use a rotationally constrained approach to select the hand-eye keyframes to construct the relative poses, aiming to mitigate the error coupling and solution degeneracy. Specifically, one can keep searching forward from a frame $\hat{\mathbf{q}}_i$ in the hand or eye trajectory, and build the relative pose when a frame $\hat{\mathbf{q}}_j$ satisfies the constraint:

$$2 \arccos\left(\left(\hat{\mathbf{q}}_i^{-1} \hat{\mathbf{q}}_j\right)_w\right) \geq \eta, \tag{6}$$

where $(\cdot)_w$ denotes the scalar part of the standard component of the dual quaternion, and η is an adjustable threshold which we set to 5 degrees. Fig. 4 illustrates the construction of relative poses using both conventional methods and our proposed method for comparison.

2) Robust kernel: Despite the effort to construct high-quality relative poses, they still contain varying degrees of noise. To quantify the weights of different relative poses, we propose a robust kernel in the linear system construction.

For a unit dual quaternion $\hat{\mathbf{q}}$, its scalar part is defined as:

$$\operatorname{Scalar}(\hat{\mathbf{q}}) = \frac{\left(\hat{\mathbf{q}} + \hat{\mathbf{q}}^{-1}\right)}{2}, \tag{7}$$

which can be expressed in the form of a vector, written as:

$$\begin{aligned} \operatorname{Scalar}(\hat{\mathbf{q}}) &= \left[\omega, \mathbf{0}_{1\times 3}, \omega', \mathbf{0}_{1\times 3}\right]^{\mathrm{T}} \\ &= \left[\cos\frac{\theta}{2}, \mathbf{0}_{1\times 3}, -\frac{d}{2}\sin\frac{\theta}{2}, \mathbf{0}_{1\times 3}\right]^{\mathrm{T}}, \end{aligned} \tag{8}$$

where ω and ω' are the scalar parts of the standard part and the dual part of the dual quaternion respectively, θ denotes the rotation angle, and d represents the translation norm of the screw motion.

The screw concatenation between the hand and eye can be further transformed from the form of $\mathbf{AX} = \mathbf{XB}$ to $\hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j} = \hat{\mathbf{q}}_{\mathrm{HE}} \hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j} \hat{\mathbf{q}}_{\mathrm{HE}}^{-1}$. Based on the definition of the scalar part in (7), we have:

$$\begin{aligned} \operatorname{Scalar}\left(\hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j}\right) &= \frac{1}{2}\left(\hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j} + \hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j}^{-1}\right) \\ &= \frac{1}{2}\left(\hat{\mathbf{q}}_{\mathrm{HE}} \hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j} \hat{\mathbf{q}}_{\mathrm{HE}}^{-1} + \hat{\mathbf{q}}_{\mathrm{HE}} \hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}^{-1} \hat{\mathbf{q}}_{\mathrm{HE}}^{-1}\right) \\ &= \frac{1}{2}\hat{\mathbf{q}}_{\mathrm{HE}}\left(\hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j} + \hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}^{-1}\right)\hat{\mathbf{q}}_{\mathrm{HE}}^{-1} \\ &= \hat{\mathbf{q}}_{\mathrm{HE}} \operatorname{Scalar}\left(\hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}\right) \hat{\mathbf{q}}_{\mathrm{HE}}^{-1} \\ &= \operatorname{Scalar}\left(\hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}\right) \hat{\mathbf{q}}_{\mathrm{HE}} \hat{\mathbf{q}}_{\mathrm{HE}}^{-1} \\ &= \operatorname{Scalar}\left(\hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}\right), \end{aligned} \tag{9}$$

which demonstrates that the scalar parts of the hand-eye relative pose pair are completely equal in the absence of noise. According to (8), this constraint can be transformed into the equality of rotation angles in local frames and translation norms along the principal axes of rotation, known as the screw congruence theorem.

We design the function as (10) to evaluate the quality of the hand-eye relative poses based on the scalar part of the dual quaternion in (8). The function yields a result of 1 when the screw motion constraint in (9) is strictly satisfied, while deviating from 1 as the noise increases.

$$\begin{aligned} & E_i\left(\hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j}, \hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}\right) \\ =& E_i'\left(\operatorname{Scalar}\left(\hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j}\right), \operatorname{Scalar}\left(\hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}\right)\right) \\ =& \frac{1}{2}\left(\frac{\max\left(\left|\omega_{\mathrm{H}_i\mathrm{H}_j}\right|, \left|\omega_{\mathrm{E}_i\mathrm{E}_j}\right|\right)}{\min\left(\left|\omega_{\mathrm{H}_i\mathrm{H}_j}\right|, \left|\omega_{\mathrm{E}_i\mathrm{E}_j}\right|\right)} + \frac{\max\left(\left|\omega'_{\mathrm{H}_i\mathrm{H}_j}\right|, \left|\omega'_{\mathrm{E}_i\mathrm{E}_j}\right|\right)}{\min\left(\left|\omega'_{\mathrm{H}_i\mathrm{H}_j}\right|, \left|\omega'_{\mathrm{E}_i\mathrm{E}_j}\right|\right)}\right). \end{aligned} \tag{10}$$

Based on (10), we can define the robust kernel W as:

$$W_i\left(\hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j}, \hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}\right) = \exp\left(\mu\left(1 - E_i\left(\hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j}, \hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}\right)^2\right)\right), \tag{11}$$

where the parameter μ is an adjustable magnification factor, which we set to 5. Hence, the linear calibration matrix in (5) can be robustified for better numerical stability:

$$\mathbf{M}_{\mathrm{r}} = \left[W_1\mathbf{S}_1^{\mathrm{T}}, W_2\mathbf{S}_2^{\mathrm{T}}, \ldots, W_n\mathbf{S}_n^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{12}$$

3) Outlier elimination: During the calibration process, it is also important to identify and discard outliers that exhibit significant error. We incorporate the RANSAC algorithm which utilizes an iterative sampling strategy, and has the ability to recover the inliers from noisy data.

Note that at least two pairs of relative poses are required to construct the linear calibration system. For each iteration, we

Algorithm 1: Robust linear hand-eye calibration

```
Input: Time aligned hand-eye trajectories $\hat{\mathbf{q}}_{\mathrm{GH}}$, $\hat{\mathbf{q}}_{\mathrm{WE}}$.
Output: Hand-eye extrinsic $\hat{\mathbf{q}}^{*}_{\mathrm{HE}}$.
$\{\hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}, \hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j}\} \in \mathcal{C} \leftarrow$ Construct the relative poses
    based on the rotational constraint in (6);
while not reached the iteration limit do
    Subset $\mathcal{D} \leftarrow$ Randomly sampling from $\mathcal{C}$;
    Construct $\mathbf{M}_{12\times 8}$ based on $\mathcal{D}$ using (5);
    $\hat{\mathbf{q}}^{\mathrm{init}}_{\mathrm{HE}} \leftarrow$ Apply SVD to M;
    foreach pose pair $\in \mathcal{C}$ do
        $\mathcal{G} \leftarrow$ Select inliers using (13) with $\hat{\mathbf{q}}^{\mathrm{init}}_{\mathrm{HE}}$;
    end
    Construct $\mathbf{M}'_{2n\times 8}$ with robust kernel based on $\mathcal{G}$
        using (12);
    $\hat{\mathbf{q}}^{\mathrm{refine}}_{\mathrm{HE}} \leftarrow$ Apply SVD to $\mathbf{M}'$;
    $\hat{\mathbf{q}}^{*}_{\mathrm{HE}} \leftarrow \hat{\mathbf{q}}^{\mathrm{refine}}_{\mathrm{HE}}$ with the smallest $\frac{\sigma_7}{\sigma_6}$ value;
end
```

randomly sample a pose pair subset consisting of two members and determine inliers by using the quantitative criterion as:

$$
\begin{aligned}
e_1 &= \left(\hat{\mathbf{q}}_{\mathrm{HE}}\hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}\hat{\mathbf{q}}_{\mathrm{HE}}^{-1}\hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j}^{-1}\right)_{\mathrm{rot}} < \varphi, \\
e_2 &= \left(\hat{\mathbf{q}}_{\mathrm{HE}}\hat{\mathbf{q}}_{\mathrm{E}_i\mathrm{E}_j}\hat{\mathbf{q}}_{\mathrm{HE}}^{-1}\hat{\mathbf{q}}_{\mathrm{H}_i\mathrm{H}_j}^{-1}\right)_{\mathrm{trans}} < \psi,
\end{aligned}
\tag{13}
$$

where $(\cdot)_{\mathrm{rot}}$ and $(\cdot)_{\mathrm{trans}}$ are the rotation angle and translation norm of the dual quaternion, respectively. Inliers are required to have both indicators in (13) less than the thresholds φ and ψ, which we set to 0.5 degrees and 0.02 m experimentally.

Additionally, we need to assess the quality of the solutions during the iteration. Given that the matrix $\mathbf{M}_{\mathrm{r}}$ in (12) possesses a two-dimensional null space, the singular values, σ_7 and σ_8, which are the last two in the descending diagonal matrix Σ from the SVD result $\mathbf{M}_{\mathrm{r}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}$, are expected to be zero in the absence of noise. However, when noise is present, these two singular values, corresponding to the noise, disproportionately increase compared to the remaining singular values. The ratio of σ_7 to the third-to-last singular value, σ_6, can serve as an indicator of this disproportionality, and a lower ratio suggests a reduced impact of noise. Consequently, we can establish a direct metric for quantifying the quality of the solutions, rather than relying on traditional measures such as the number of inliers or the root mean square error (RMSE) of the solutions with respect to inliers. The specific implementation of the linear spatial hand-eye calibration within the RANSAC framework is detailed in Algorithm 1.

C. Batch Estimation

Despite the fact that our linear calibration method achieves good accuracy and robustness in the experiments of Section IV, it can be further improved by introducing the correlation between the temporal and spatial calibration parameters. In this section we follow the continuous-time batch estimation methods proposed in [32], [33] and provide an estimator within the rigorous theoretical framework of MLE, to jointly optimize the time offset and the spatial transformation. Since the estimator is hard to converge globally, we used the results derived in Section III-A and III-B as the initial guesses.

Specifically, the original hand trajectory is parameterized using a B-spline functions $\mathbf{T}_{\mathrm{GH}}(t)$, with the translation part represented by a B-spline in three-dimensional vector space, and the rotation part parameterized by a B-spline on $SO(3)$. For the B-spline $\mathbf{q}(t)$ on $SO(3)$ with order ξ, knots $\{t_i | i \in \{1,2,3,\ldots,N\}\}$, and satisfying $N \geq 2\xi$, the function can be defined in each subinterval $\{t \in [t_i, t_{i+1}) \,|\, \xi \leq i \leq N-\xi\}$ as:

$$
\mathbf{q}(t) = \mathbf{q}_{l(i)} \prod_{j=1}^{\xi-1} \mathbf{EXP}\left(\Big(\sum_{k=\eta}^{i} f_k\Big)\mathbf{LOG}\left(\mathbf{q}_{\eta-1}^{-1}\mathbf{q}_{\eta}\right)\right), \tag{14}
$$

where $l(i) = i - \xi + 1$ and $\eta = l(i) + j$, f and $\mathbf{q}$ represent the B-spline basis functions and the control vertices in unit quaternion form, respectively. $\mathbf{EXP}$ denotes the mapping from the Lie algebra to the Lie group, while $\mathbf{LOG}$ represents the inverse process.

The parameters determined by our estimator include $\mathbf{T}_{\mathrm{GH}}(t)$, homogeneous transformation $\mathbf{T}_{\mathrm{HE}}$, and time offset Δt_{HE}. With the observations of hand-eye trajectories, we minimize the negative log-likelihood function as:

$$
\begin{aligned}
g = &\sum_{h=1}^{H-1} \rho\left(\left\| d\left(\mathbf{T}_{\mathrm{GH}}(t_h), \mathbf{T}_{\mathrm{GH}}(t_{h+1}), \mathbf{T}_{\mathrm{GH}_h}, \mathbf{T}_{\mathrm{GH}_{h+1}}\right)\right\|^2_{\Sigma_{\mathrm{H}}}\right) \\
&+ \sum_{e=1}^{E-1} \rho\left(\left\| d\left(\mathbf{T}_{\mathrm{WE}}(t_e), \mathbf{T}_{\mathrm{WE}}(t_{e+1}), \mathbf{T}_{\mathrm{WE}_e}, \mathbf{T}_{\mathrm{WE}_{e+1}}\right)\right\|^2_{\Sigma_{\mathrm{E}}}\right),
\end{aligned}
\tag{15}
$$

where H and E denote the number of pose observations of the hand and the eye respectively, and $\rho(\cdot)$ is the Huber loss function. Due to the rigid connection between the hand and the eye, we have $\mathbf{T}_{\mathrm{WE}}(t) = \mathbf{T}_{\mathrm{WG}}\mathbf{T}_{\mathrm{GH}}(t - \Delta t_{\mathrm{HE}})\mathbf{T}_{\mathrm{HE}}$. Meanwhile, we define the residual function in a relative form to eliminate the influence of the unknown $\mathbf{T}_{\mathrm{WG}}$ as:

$$
d\left(\mathbf{T}_i, \mathbf{T}_{i+1}, \mathbf{T}'_i, \mathbf{T}'_{i+1}\right) = \mathbf{LOG}\Big(\mathbf{T}_i^{-1}\mathbf{T}_{i+1}\left(\mathbf{T}'^{-1}_i\mathbf{T}'_{i+1}\right)^{-1}\Big). \tag{16}
$$

For the above cost function, the Levenberg–Marquardt algorithm is used to obtain the refined calibration result.

IV. EXPERIMENTAL RESULTS

A. Ablation Studies

To validate the effectiveness of the key improvement strategies proposed in our methodology, we conduct comparative experiments using the simulated datasets. Specifically, to obtain the required data, we model a real motion trajectory using the B-spline. Subsequently, we extract two trajectories from the model at 100Hz and 20Hz to simulate MoCap and VO/VIO trajectories for hand-eye calibration. Throughout the process, we introduce frame-wise cumulative error into each eye pose to simulate the noise and drift typically present in the VO/VIO trajectory. The standard deviations for translation and rotation errors are divided into ten levels, ranging from 0 to 5 mm and 0 to 0.2 degrees, respectively.

We first test the time alignment algorithm proposed in Section III-A to verify the enhancement provided by the quadratic

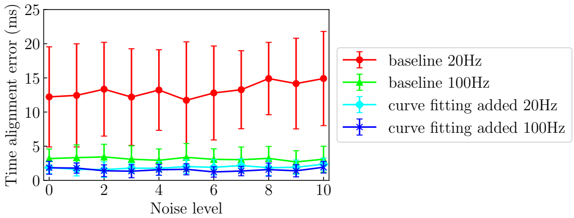

Fig. 5. Performance comparison of different time alignment methods under various noise levels. We report the mean and standard deviation of the time alignment error.

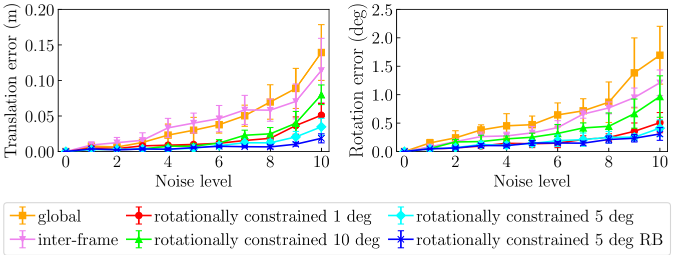

Fig. 6. Performance comparison of different strategies used in linear calibration under various noise levels. We calculate the translation error and the rotation error separately, and report the mean and standard deviation.

polynomial curve fitting in the correlation analysis compared to the baseline approach used in [10]. In this experiment, we introduce random delays into the eye trajectories derived from the B-spline to simulate the time offsets. To facilitate the correlation function computation, standardizing the frequencies of the trajectory pair is necessary. Specifically, we adjust the frequencies to match those of hand and eye trajectories, which are respectively 100Hz and 20Hz. As shown in Fig. 5, correlation analysis-based time alignment methods exhibit minimal sensitivity to trajectory noise. However, limited by the temporal resolution of the correlation function, time alignment error of the baseline method is essentially determined by the trajectory frequency. Acquiring the high-frequency trajectory is challenging and requires significant computational effort. By implementing the curve fitting strategy, we can determine the maximum index of the correlation function with greater accuracy, leading to higher precision in time alignment, even at low frequency.

Additionally, we evaluate the effectiveness of the strategies for spatial linear calibration proposed in Section III-B. The three relative poses construction methods, namely global, inter-frame, and rotationally constrained, are comparatively validated, and different thresholds in (6) are explored within the proposed rotationally constrained approach. Meanwhile, we test the impact of the robust kernel (RB). In this experiment, the calibration error is used as a quantitative metric for assessment. Fig. 6 indicates that the calibration errors for all scenarios trend to increase as noise levels rise. Nonetheless, our method, which constructs rotationally constrained relative poses yields smaller error due to its ability to derive the linear equation with higher signal-to-noise ratio. At the same time, the rotational constraint threshold also affects the accuracy of the solution, with 5 degrees proving to be optimal in our tests. Furthermore, our robust kernel strategy demonstrates superior robustness in the presence of increasing noise, as indicated by the relatively small rise in calibration error and standard deviation.

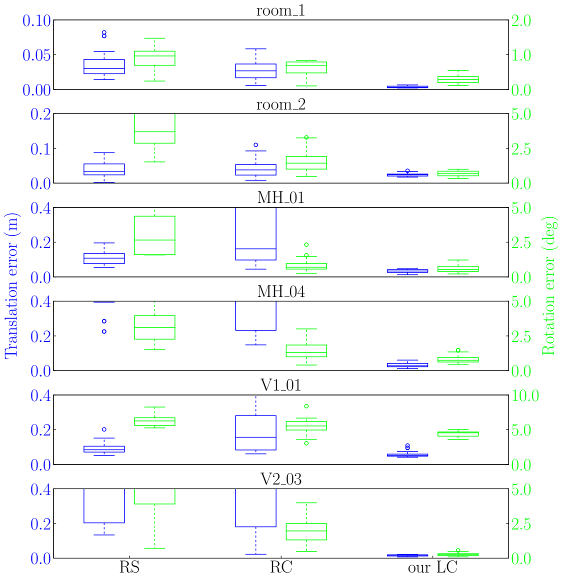

Fig. 7. Performance comparison of different linear calibration methods on public datasets. We select two sequences in each scenario and count the distribution of the translation error and the rotation error separately.

B. Overall Evaluation

We evaluate the performance of our hand-eye calibration algorithm using real-world datasets, including public VIO datasets, and datasets collected by our system (see in Fig. 1). For comparison, we also evaluate two other state-of-the-art (SOTA) linear hand-eye calibration algorithms based on dual quaternion and random sampling, as implemented in [10], namely RANSAC scalar-based inlier check (RS) and RANSAC classic (RC). We use the well-known VIO algorithm OpenVINS [34] to estimate the trajectory of the eye, while the trajectory of the hand is derived from the ground-truth system.

We first test on the TUM-VI room scenario [2] as well as machine hall (MH) and Vicon room (V) scenarios provided by EuRoC benchmark [1]. All three scenarios provide complete raw data of ground-truth trajectories, and achieve high-precision hand-eye calibration through the tightly-coupled method. We align the estimated VIO trajectories with the raw ground-truth trajectories using different hand-eye calibration algorithms. For evaluation metrics, we compare with the calibration results provided by the public datasets, and calculate the translation error and the rotation error. As shown in Fig. 7, we first evaluate our linear calibration (LC). The comparison results demonstrate that our algorithm achieves the highest accuracy and repeatability in all sequences. In challenging sequences such as MH_04 and V2_03, where the VIO algorithm struggles to deliver high-quality trajectories, traditional calibration algorithms are prone to fail, but our algorithm handles these situations effectively.

Table I details the performance of our algorithm in each sequence, including the time alignment errors and the results

TABLE I
DETAILED PERFORMANCE OF THE PROPOSED METHODS AND THE TRAJECTORY METRICS ON PUBLIC DATASETS

Sequences	Error of our LC			Error of our BE			Original metrics		Our metrics	
	Time (ms)	Trans (m)	Rot (deg)	Time (ms)	Trans (m)	Rot (deg)	APE (m)	ARE (deg)	APE (m)	ARE (deg)
Room_1	1.393	0.003	0.284	0.198	0.004	0.106	0.059	1.593	0.059	1.593
Room_6	1.266	0.024	0.675	0.296	0.016	0.199	0.085	1.686	0.085	1.683
MH_01	3.430	0.032	0.596	0.883	0.018	0.641	0.156	1.884	0.157	1.683
MH_04	5.404	0.031	0.927	3.691	0.018	0.749	0.161	0.950	0.157	0.561
V1_01	3.718	0.053	4.444	1.982	0.049	4.479	0.061	5.520	0.047	1.064
V2_03	3.092	0.015	0.280	2.135	0.007	0.118	0.095	1.157	0.095	1.139

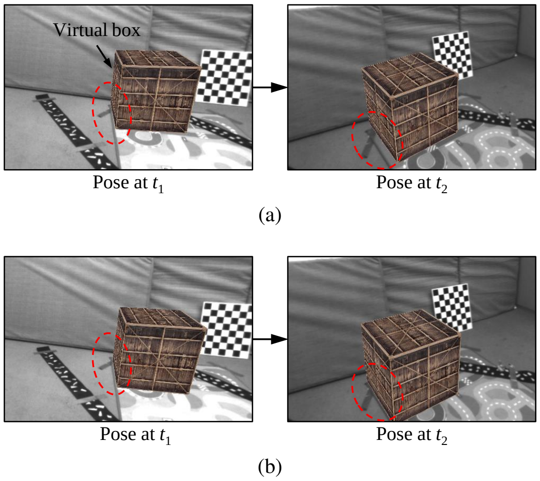

Fig. 8. AR visualization of the EuRoC V1_01 sequence, featuring a virtual box rendered in the scene. This visualization utilizes (a) transformed ground-truth poses provided by the EuRoC benchmark and (b) transformed ground-truth poses based on the raw pre-calibrated trajectory and our calibration result. The red dashed circle highlights the difference in consistency between the virtual and real elements when using different poses.

of our batch estimation (BE). Additionally, to analyze the impact of hand-eye calibration on VIO trajectory evaluation, we use our calibration results to transform the raw ground-truth trajectories and calculate the absolute positional error (APE) and absolute rotational error (ARE) [3], which are important metrics in VIO evaluation. As a comparison, we also calculate the original metrics using the transformed ground-truth provided by the public datasets. The obtained results demonstrate that our LC effectively accomplishes spatiotemporal hand-eye calibration, and our BE can further optimizes the result to achieve higher accuracy. In most of the sequences, our calibration algorithm provides accurate transformations for ground-truths with minimal impact on evaluation metrics. However, it is interesting to note that in some sequences, particularly V1_01, our algorithm exhibits some errors. This may stem from the inherent inaccuracies in the calibration results of the benchmarks, corroborated by a similar conclusion in [34]. To provide a more intuitive illustration, we render a virtual box on the images of the V1_01 sequence using two different transformed ground-truth trajectories. The first is provided by the EuRoC benchmark and the second is obtained based on our hand-eye calibration result. Inaccurate hand-eye calibration can lead to misalignment between the virtual and the real elements in this augmented reality (AR) application. As shown in the comparison results in Fig. 8, the virtual box is more consistent with the real world in our AR result, i.e., implying higher calibration accuracy.

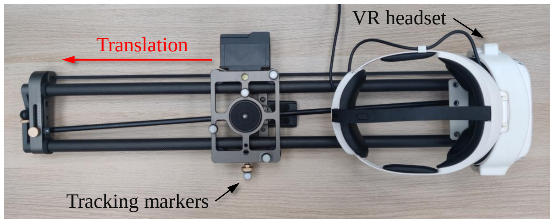

Fig. 9. Experimental setup for evaluating the hand-eye calibration using a relative translation method. The VR headset remains stationary, and the tracking markers can be translated along the red arrow.

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT CALIBRATION METHODS ON SELF-COLLECTED DATASETS. WE COUNT THE TRANSLATION ERROR AT DIFFERENT RELATIVE TRANSLATION DISTANCES

Sequences	Error of RS (m)	Error of RC (m)	Error of our_LC (m)	Error of our_BE (m)
0.1m_easy	0.035	0.028	0.007	0.003
0.2m_easy	0.026	0.022	0.005	0.003
0.3m_easy	0.049	0.034	0.007	0.002
0.1m_difficult	0.041	0.045	0.007	0.004
0.2m_difficult	0.056	0.038	0.006	0.004
0.3m_difficult	0.062	0.047	0.010	0.004

In real-world system, evaluating hand-eye calibration is inherently difficult, as the ground-truth of offset is not available. To address this, we employ a relative approach using self-collected data. The experimental setup, as illustrated in Fig. 9, involved mounting a VR headset and tracking markers on the same object for motion. After obtaining a set of hand-eye trajectories, we translate the tracking markers in a specified direction to gather additional calibration data. We compute the norm of the relative translation using the extrinsics derived from the calibrations performed before and after the translation. Calibration error is then determined by comparing this norm against the high-precision, directly measured result.

Table II presents the errors obtained from different hand-eye calibration methods when the tracking markers are translated by 0.1 m, 0.2 m, and 0.3 m, respectively. We categorize the scenarios into easy (texture-rich and slow-moving) and difficult (texture-less and fast-moving) in the context of VIO, to test the robustness of the calibration algorithms. The experimental results show that our algorithm performs best in all sequences. Additionally, both our LC and BE achieve millimeter-level calibration accuracy and are less affected by trajectory error.

V. CONCLUSIONS

In this letter, we propose an improved spatiotemporal hand-eye calibration algorithm for trajectory alignment in VO/VIO evaluation. Aiming to optimize for VO/VIO scenarios, we have designed multiple strategies based on screw theory to enhance both the accuracy and robustness of our proposed algorithm. The validation experiments demonstrate that our algorithm outperforms SOTA methods, exhibiting superior accuracy while effectively mitigating the influence of noise. Our method is well poised to be applied in the evaluation of modern VO/VIO algorithms. Nevertheless, our algorithm is less effective over an extended period. This limitation arises partly because our calibration algorithm processes the entire trajectory, leading to inefficiency with long sequences. Furthermore, our algorithm presupposes a constant time offset, an assumption unsuitable over a long time. In future work, we will focus on spatiotemporal hand-eye calibration tailored for long-duration trajectories. The aim is to develop an algorithm that can process extensive data efficiently and address the problem of time offset drift over an extended period.

REFERENCES

[1] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.

[2] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stückler, and D. Cremers, "The TUM VI benchmark for evaluating visual-inertial odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2018, pp. 1680–1687.

[3] J. Li, B. Yang, D. Chen, N. Wang, G. Zhang, and H. Bao, "Survey and evaluation of monocular visual-inertial SLAM algorithms for augmented reality," *Virtual Reality & Intelligent Hardware*, vol. 1, no. 4, pp. 386–410, 2019.

[4] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 13, no. 4, pp. 376–380, 1991.

[5] J. Jiang, X. Luo, Q. Luo, L. Qiao, and M. Li, "An overview of hand-eye calibration," *The International Journal of Advanced Manufacturing Technology*, vol. 119, no. 1-2, pp. 77–97, 2022.

[6] E. Pedrosa, M. Oliveira, N. Lau, and V. Santos, "A general approach to hand-eye calibration through the optimization of atomic transformations," *IEEE Transactions on Robotics*, vol. 37, no. 5, pp. 1619–1633, 2021.

[7] K. Koide and E. Menegatti, "General hand-eye calibration based on reprojection error minimization," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1021–1028, 2019.

[8] I. Ali, O. Suominen, A. Gotchev, and E. R. Morales, "Methods for simultaneous robot-world-hand-eye calibration: A comparative study," *Sensors*, vol. 19, no. 12, p. 2837, 2019.

[9] S. Sarabandi, J. M. Porta, and F. Thomas, "Hand-eye calibration made easy through a closed-form two-stage method," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 3679–3686, 2022.

[10] F. Furrer, M. Fehr, T. Novkovic, H. Sommer, I. Gilitschenski, and R. Siegwart, "Evaluation of combined time-offset estimation and hand-eye calibration on robotic datasets," in *Field and Service Robotics: Results of the 11th International Conference*. Springer, 2018, pp. 145–159.

[11] Y. C. Shiu and S. Ahmad, "Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX = XB," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 1, pp. 16–29, 1989.

[12] R. Y. Tsai and R. K. Lenz, "Real time versatile robotics hand/eye calibration using 3D machine vision," in *1988 IEEE International Conference on Robotics and Automation*. IEEE, 1988, pp. 554–561.

[13] C. C. Wang, "Extrinsic calibration of a vision sensor mounted on a robot," *IEEE Transactions on Robotics and Automation*, vol. 8, no. 2, pp. 161–175, 1992.

[14] J. C. Chou and M. Kamel, "Finding the position and orientation of a sensor on a robot manipulator using quaternions," *The International Journal of Robotics Research*, vol. 10, no. 3, pp. 240–254, 1991.

[15] F. C. Park and B. J. Martin, "Robot sensor calibration: Solving AX = XB on the Euclidean group," *IEEE Transactions on Robotics and Automation*, vol. 10, no. 5, pp. 717–721, 1994.

[16] R. Liang and J. Mao, "Hand-eye calibration with a new linear decomposition algorithm," *Journal of Zhejiang University-SCIENCE A*, vol. 9, no. 10, pp. 1363–1368, 2008.

[17] H. Nguyen and Q. C. Pham, "On the covariance of X in AX = XB," *IEEE Transactions on Robotics*, vol. 34, no. 6, pp. 1651–1658, 2018.

[18] H. H. Chen, "A screw motion approach to uniqueness analysis of head-eye geometry," in *1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 1991, pp. 145–146.

[19] N. Andreff, R. Horaud, and B. Espiau, "On-line hand-eye calibration," in *Second International Conference on 3-D Digital Imaging and Modeling*. IEEE, 1999, pp. 430–436.

[20] K. Daniilidis, "Hand-eye calibration using dual quaternions," *The International Journal of Robotics Research*, vol. 18, no. 3, pp. 286–298, 1999.

[21] D. Condurache and A. Burlacu, "Orthogonal dual tensor method for solving the AX = XB sensor calibration problem," *Mechanism and Machine Theory*, vol. 104, pp. 382–404, 2016.

[22] K. H. Strobl and G. Hirzinger, "Optimal hand-eye calibration," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006, pp. 4647–4653.

[23] J. Heller, D. Henrion, and T. Pajdla, "Hand-eye and robot-world calibration by global polynomial optimization," in *2014 IEEE International Conference on Robotics and Automation*. IEEE, 2014, pp. 3157–3164.

[24] Z. Zhao, "Simultaneous robot-world and hand-eye calibration by the alternative linear programming," *Pattern Recognition Letters*, vol. 127, pp. 174–180, 2019.

[25] J. Schmidt, F. Vogt, and H. Niemann, "Robust hand-eye calibration of an endoscopic surgery robot using dual quaternions," in *Pattern Recognition: 25th DAGM Symposium*. Springer, 2003, pp. 548–556.

[26] J. Wu, Y. Sun, M. Wang, and M. Liu, "Hand-eye calibration: 4-D procrustes analysis approach," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 2966–2981, 2019.

[27] J. Kelly and G. S. Sukhatme, "A general framework for temporal calibration of multiple proprioceptive and exteroceptive sensors," in *Experimental Robotics: The 12th International Symposium on Experimental Robotics*. Springer, 2014, pp. 195–209.

[28] M. K. Ackerman, A. Cheng, B. Shiffman, E. Boctor, and G. Chirikjian, "Sensor calibration with unknown correspondence: Solving AX = XB using Euclidean-group invariants," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 1308–1313.

[29] H. Li, Q. Ma, T. Wang, and G. S. Chirikjian, "Simultaneous hand-eye and robot-world calibration by solving the AX = YB problem without correspondence," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 145–152, 2015.

[30] K. Pachtrachai, F. Vasconcelos, G. Dwyer, V. Pawar, S. Hailes, and D. Stoyanov, "Chess—Calibrating the hand-eye matrix with screw constraints and synchronization," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2000–2007, 2018.

[31] 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.

[32] J. Rehder, R. Siegwart, and P. Furgale, "A general approach to spatiotemporal calibration in multisensor systems," *IEEE Transactions on Robotics*, vol. 32, no. 2, pp. 383–398, 2016.

[33] H. Sommer, J. R. Forbes, R. Siegwart, and P. Furgale, "Continuous-time estimation of attitude using B-splines on Lie groups," *Journal of Guidance, Control, and Dynamics*, vol. 39, no. 2, pp. 242–261, 2016.

[34] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "OpenVINS: A research platform for visual-inertial estimation," in *2020 IEEE International Conference on Robotics and Automation*. IEEE, 2020, pp. 4666–4672.