The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-8-2025
46th Canadian Symposium on Remote Sensing (CSRS)
"From Mountains to Kitchens: Remote Sensing Innovations for Water, Food & Security", 16–19 June 2025, Lethbridge, Canada

# State Covariance Based Spatio-Temporal Trajectory Alignment for VIO Systems

Zelin Zhou[1], Hongzhou Yang[1*]

[1] Z. Zhou & H. Yang, University of Calgary - (zelin.zhou1, honyang)@ucalgary.ca

**Keywords:** Trajectory alignment, state covariance, visual-inertial odometry, spatio-temporal alignment, pose transformation

**Abstract**

With the growing interest in robotics and autonomous vehicles, visual-inertial odometry (VIO) in SLAM techniques have become increasingly important for estimating a robot's trajectory using visual and inertial data. Evaluating the accuracy of a VIO trajectory estimate typically requires aligning it with a reference trajectory, where the goal is to find a transformation that minimizes the discrepancy between estimated and reference poses. However, existing methods often overlook the state covariance of the estimated trajectory or rely solely on manufacturer specifications, without considering the covariance of VIO estimation. As a result, the accuracy of the aligned trajectory may not be truly reflected, as the alignment process minimizes global discrepancies rather than prioritizing state pairs with higher confidence. This paper investigates the impact of state covariance of estimated trajectory on spatio-temporal trajectory alignment. To validate our approach, we use various open-source datasets that exhibit different state covariance behaviors and conduct comprehensive statistical analyses. The results indicate that the proposed method improves the precision and internal reliability in estimating alignment parameters. Additionally, the resulting a-posteriori variance factor of unit weight from the proposed method reflects a better-calibrated stochastic model and can serve as an indicator of state covariance estimation quality in VIO systems.

## 1. Introduction

With the growing interest in robotics and autonomous vehicles, visual-inertial odometry (VIO) in SLAM techniques have become increasingly important for estimating a robot's trajectory from visual and inertial data. The accuracy of VIO is commonly evaluated by comparing the estimated trajectory against a reference trajectory. However, without global positioning from GNSS, the VIO or SLAM trajectory cannot be directly compared with the reference trajectory, as they are typically expressed in different reference frames. Consequently, a preliminary alignment is required, which is often carried out using either a similarity or rigid-body transformation, depending on the measurement and sensor configuration. The parameters of this transformation can be estimated using all the states from trajectories or only the initial states, with the distinctions and implications discussed by Zhang et al. (2018).

Horn's method (1987) is widely used to estimate the rotation and translation parameters of the alignment transformation, assuming that temporal correspondence between the estimated and reference poses is already established. It solves a closed-loop least squares problem by minimizing the translational errors between the aligned estimates and their corresponding reference translations. Umeyama's method (1991) extends this approach by incorporating the estimation of scale parameters, making the alignment applicable for the trajectory with scale ambiguity such as the trajectory from monocular visual odometry. However, these two approaches treat all estimated states as having identical uncertainty—which is generally not the case—and use only the translational component of the poses, presuming any rotation error will manifest in the sequential translation. To investigate the effect of explicitly including rotation in the alignment, Salas et al. (2015) used both simulated data with added Gaussian noise and real data to show that accounting for rotation can reduce the absolute rotation error, albeit at the cost of increased absolute translation error. Nevertheless, this method continues to assume identical state covariance across the estimated poses, an approximation that may not fully reflect real-world conditions.

In addition, without precise GPS timing, VIO systems can suffer from time synchronization errors, which introduce inaccuracies in pose timestamps. Different output rates for the estimation and groundtruth systems further complicate alignment, as poses are usually matched by the closest timestamps. Manually measuring the lever arms between sensors adds another potential source of error. To address these challenges, Tombrink et al. (2024) introduced a more sophisticated spatio-temporal alignment that jointly estimates the trajectory transformation parameters, the time offset between sensors, and the lever arms in the sensor frame (assuming the lever arm errors are only present in translation component). However, instead of using the state covariance computed by the estimator, they relied on the localization error specifications provided by the manufacturer. This approach implicitly assumes that real-world uncertainties are consistent with nominal manufacturer specifications—an assumption that may not hold if unexpected errors or noise arise in practice.

For a least squares based estimator, the state covariance can be described as the inverse of the Hessian of the cost function after the solution has converged. It encodes the strength of the constraints imposed by the observations on each parameter (Moritz, 1972). All the previously mentioned methods neglect state covariance after pose estimation, despite its potential benefits for trajectory alignment and subsequent evaluation. Additionally, the accuracy of state covariance estimated from VIO systems remains uncertain, and its quality assessment has received little research attention. However, state covariance from VIO is crucial for applications such as loosely coupled integration (Liu et al., 2025; Sirtkaya et al., 2013; Zhang et al., 2024) and solution credibility evaluation (Niu et al., 2023; Xu et al., 2024).

By incorporating VIO-estimated state covariance, we can achieve a more intuitive trajectory alignment that reflects the

quality of pose estimation from real data. Furthermore, the uncertainties of estimated poses can be propagated through the alignment process and captured in the resulting state standard deviations, cross-state correlations, and a-posteriori variance of the estimated alignment parameters. This provides valuable insights into the precision and internal reliability of alignment parameter estimation, as well as the quality of state covariance estimation in VIO systems.

Consequently, this paper investigates how state covariance of estimated pose influences spatio-temporal trajectory alignment for VIO systems and discuss an assessment approach for the state covariance estimation from VIO using the alignment statistics. Specifically, we integrate the estimated state covariance from Open-VINS (Geneva et al., 2020), a state-of-the-art (SOTA) filtering-based VIO estimator into the alignment, considering all estimated poses and weighing them according to their respective variances. The four degrees-of-freedom yaw-only rigid body transformation includes a 3D translation and a yaw-only rotation. These parameters (which are sufficient to align a VIO trajectory as discussed by Kelly & Sukhatme (2011)), along with a time offset between the estimated and reference trajectories, are estimated to align the VIO trajectory. To investigate the impact of fluctuations in VIO-estimated state covariance on alignment parameter estimation, we evaluate our method using four data sequences from two widely used open-source visual-inertial datasets: EuROC MAV (Burri et al., 2016) and TUM (Schubert et al., 2018). Each data sequence represents varying dynamics and levels of feature-tracking difficulty.

The remainder of the paper is structured as follows. First, we provide a brief introduction to the workflow of the proposed algorithm. Next, we discuss the state covariance based trajectory alignment and an assessment approach for VIO state covariance estimation using statistical results after alignment. Subsequently, we elaborate on the alignment configuration, validation datasets, and their characteristics. Finally, we present and analyze the performance of the alignment parameter estimation and the evaluation of the aligned trajectory.

## 2. Methodology

The implementation of the proposed method is illustrated in Figure 1. The inputs for state covariance based trajectory alignment are highlighted in green, while the system's output or trajectory evaluation results are marked in red. These values are also used for comparison with benchmark results.

The VIO provides two primary outputs: the estimated trajectory, consisting of a sequence of estimated poses with their corresponding timestamps, and the associated state covariance, which characterizes the uncertainty in the estimated states. The inputs for alignment parameter estimation include the reference trajectory along with its accuracy specifications, derived from the sensor manufacturer's datasheet, as well as the estimated trajectory and its associated state covariance. A least squares estimation is used to determine the spatio-temporal transformation, represented by the alignment parameters. These parameters consist of three translations along the x-, y-, and z-axes, a rotation around the z-axis, and a time offset between the sensors of the estimated trajectory and the reference trajectory. Furthermore, estimation statistics are computed to assess the performance of the alignment parameter estimation. These include the standard deviations of the estimated alignment parameters, the a-posteriori variance factor of unit weight, and the correlation matrix of the alignment parameters. Once the spatio-temporal transformation is estimated, it is applied to the estimated trajectory to express it in the reference trajectory frame. With both trajectories aligned in a common frame, their consistency is quantitatively evaluated using the Absolute Pose Error (APE) metric.

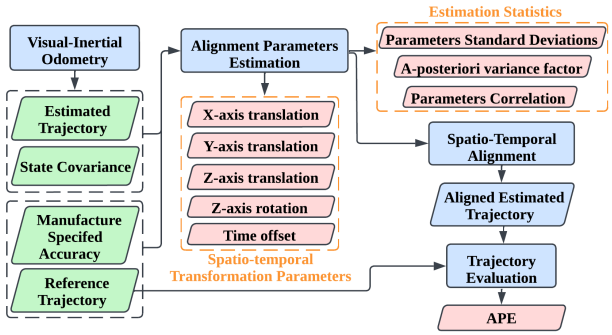



Figure 1. Overview of the state covariance based trajectory alignment.

### 2.1 State Covariance based Trajectory Alignment

Given both the estimated and reference trajectories, temporal correspondence can be established by matching poses with their nearest timestamps. Let the translation components of the matched poses be denoted as $\{\hat{\mathbf{p}}_i\}_{i=0}^{N-1}$ and $\{\mathbf{p}_i\}_{i=0}^{N-1}$, for estimated and reference poses respectively. The goal of general trajectory alignment is to find a transformation $g'(\cdot)$ that satisfies:

$$g' = \arg\min \sum_{i=0}^{N-1} \|\mathbf{p}_i - g(\hat{\mathbf{p}}_i)\|^2 \tag{1}$$

However, this nearest-timestamp matching does not guarantee perfect temporal correspondence between estimated and reference poses. Even when two poses share identical timestamps, they may not be truly synchronized due to time synchronization errors, which can arise in VIO systems lacking precise timing services (e.g., GPS). Consequently, inaccuracies in temporal correspondence introduce additional errors into trajectory alignment, which should be accounted for in the alignment process. Moreover, the least squares estimation in Equation (1) assumes uniform uncertainty across translation states, which does not accurately reflect real-world state estimation uncertainties.

Therefore, a weighted spatio-temporal trajectory alignment is required to address these issues. In this process, we aim to estimate the following alignment parameters:

$$\mathbf{x} = [t'_x, t'_y, t'_z, R'_z, \Delta_t] \tag{2}$$

Given the observations for the least squares problem:

$$\mathbf{l}_i = [p_x, p_y, p_z, \hat{p}_x, \hat{p}_y, \hat{p}_z, v_x, v_y, v_z] \tag{3}$$

And Equation (1) can be explicitly reformulated as Equation (4):

$$g' = \underset{g=\{\mathbf{R}_z,\mathbf{t},\mathbf{\Delta}_t\}}{\arg\min} \sum_{i=0}^{N-1} \mathbf{r}_i^T \mathbf{\Sigma}_i^{-1} \mathbf{r}_i \tag{4}$$

$$\mathbf{r}_i = \mathbf{p}_i - \mathbf{R}_z(\hat{\mathbf{p}}_i + \mathbf{v}_i\mathbf{\Delta}_t) - \mathbf{t} \tag{5}$$

where $\mathbf{R}_z$ represents the rotation around the z-axis, and $\mathbf{t} = [t_x, t_y, t_z]^T$ denotes the 3D translations. The term $\mathbf{\Delta}_t$ corresponds to the time offset between the estimated and reference trajectories. The residual $\mathbf{r}_i$ is explicitly defined in Equation (5), while $\Sigma_i^{-1}$ represents the inverse of the observation covariance, which is constructed based on the state covariance of estimated trajectory and reference trajectory, as shown in Equation (6). For estimated trajectory covariance, filtering-based VIO is chosen over graph optimization-based VIO as it inherently maintains and propagates state covariances in a sequential manner, providing direct uncertainty estimates for each estimated state. For the covariance of the reference trajectory, which is also an observation in this least squares problem, we assign its covariance based on the manufacturer specifications, which are typically much smaller than those of the estimated trajectory. The velocity $\mathbf{v}_i = [v_x, v_y, v_z]$ is computed in the estimated trajectory frame as the gradients of translation with respect to time. Notably, the sensors used for both the estimated and reference trajectories are mounted on the same vehicle, meaning their velocities are identical but expressed in different frames.

$$\mathbf{\Sigma}_i = \mathbf{\Sigma}_{est,i} + \mathbf{\Sigma}_{ref,i} \tag{6}$$

To solve the least squares problem with an implicit functional model, the Gauss-Helmert model (Koch, 2014) is employed to estimate the alignment parameters. For fast and reliable convergence, the initial estimate of the yaw-only rigid body transformation is computed using the traditional Umeyama's method. Subsequently, the initial value for $\mathbf{\Delta}_t$ is determined separately.

Once the solution has converged, least squares statistics are used to evaluate the estimation performance. These statistics include the standard deviation and correlation of the estimated alignment parameters, as well as the a-posteriori variance factor of the unit weight, which will be discussed in detail in the next section.

$$\hat{\mathbf{p}}_i' = \mathbf{R}_z'(\hat{\mathbf{p}}_i + \mathbf{v}_i\mathbf{\Delta}_t) + \mathbf{t}', \quad \hat{\mathbf{R}}_i' = \mathbf{R}_z'\hat{\mathbf{R}}_i \tag{7}$$

To evaluate the trajectory, the aligned trajectory, represented by Equation (7), is compared against the reference trajectory.

### 2.2 Assessment of State Covariance Estimation

In section 2.1, the alignment parameters estimation can be formulated as a weighted least squares problem. The state covariance estimated by the VIO systems are now served as the covariance of observations that account for the uncertainty of the estimated poses. After the least squares solution converges, the a-posteriori variance factor of unit weight can be computed as:

$$\hat{\sigma}_\mathbf{o}^\mathbf{2} = \frac{1}{n-m} \sum_i \mathbf{r}_i^T \mathbf{\Sigma}_i^{-1} \mathbf{r}_i \tag{8}$$

where $n$ and $m$ are the number of observations and estimates, respectively. Its difference represents the degree of freedom in the least square estimation. $\mathbf{r}_i$ refers to the residual defined by Equation (5) and $\mathbf{\Sigma}_i$ denotes the covariance of the observations.

The value of $\hat{\sigma}_\mathbf{o}^\mathbf{2}$ provides a direct gauge of how well the pose covariances $\mathbf{\Sigma}_i$ reflect the actual statistical spread of the alignment residuals. If $\hat{\sigma}_\mathbf{o}^\mathbf{2}$ is close to '1', the residuals are consistent with the assumed covariances (Teunissen, 2003). when $\hat{\sigma}_\mathbf{o}^\mathbf{2} < 1$, it indicates that the observation noises has been overestimated (i.e., the actual alignment errors are smaller than predicted). Conversely, if $\hat{\sigma}_\mathbf{o}^\mathbf{2} > 1$, it implies the covariances were too optimistic (the errors are larger than predicted).

Since the reference trajectory is usually generated by a high-precision groundtruth sensor (e.g., a motion capture system or a high-end survey-grade device), its error stochastic are accurately known and characterized by $\mathbf{\Sigma}_{ref,i}$. Consequently, the main source of uncertainty in the alignment problem comes from the estimated poses' covariances $\mathbf{\Sigma}_{est,i}$. Therefore, once the weighted least squares solution converges, the $\hat{\sigma}_\mathbf{o}^\mathbf{2}$ primarily indicates whether $\mathbf{\Sigma}_{est,i}$ has been modeled properly.

## 3. Experiments

To validate the proposed method, we conduct four experiments using data sequences from the datasets: EuROC MAV and TUM. Both datasets provide synchronized visual-inertial indoor data with high-quality ground truth. The visual data consists of monochrome stereo images. For simplicity, we refer to the data sequences used in the experiments as Dataset I, II, III, and IV, respectively. Each dataset exhibits unique characteristics and presents different levels of difficulty for VIO, leading to varying state covariance behavior.

We use Open-VINS to estimate the trajectory and its state covariance for all four datasets. Since these datasets have been thoroughly evaluated in their original studies, we adopt their estimation configurations to ensure optimal trajectory estimation performance.

Both weighted and unweighted trajectory alignments are performed and compared. In the weighted alignment, the estimated state covariance from VIO serves as the stochastic model. In the unweighted approach, the standard deviation of 3D translations is fixed at $1\ m$, following the assumption in Umeyama's method, which implicitly treats all observations as having uniform and isotropic uncertainty.

Since the state covariance of the reference trajectory is also required as an input for estimation, we use the translational accuracy reported by the dataset publishers—approximately $1\ mm$—for both EuROC MAV and TUM-based data.

### 3.1 Dataset Characteristics

In general, Datasets I, II, III, and IV are arranged in increasing order of estimation difficulty. Their state covariances are plotted in Figure 2.

Dataset I is based on the TUM Room 3 data sequence, which exhibits overall small covariances for both rotational and translational components. The data features smooth motion and consistent visual clarity, allowing VIO to perform optimally. However, a significant amount of circular motion (as seen in Figure 4) and near-static orientation variations introduce additional

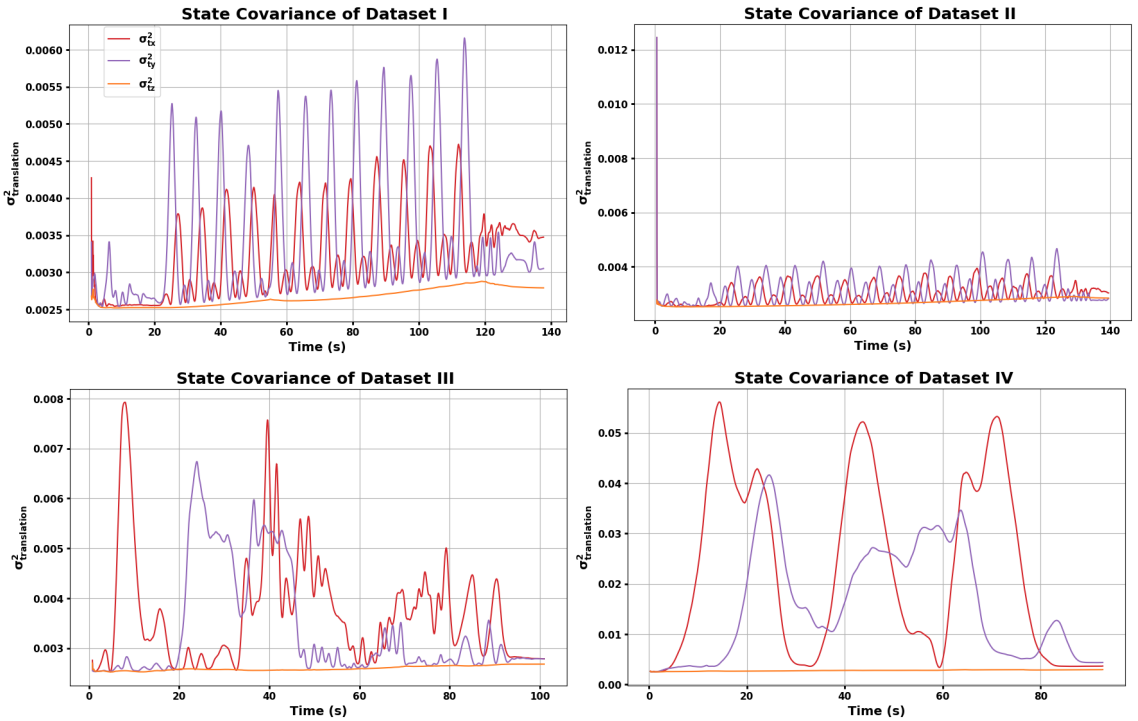



Figure 2. State covariance of estimated trajectory from Dataset I, II, III, IV.

challenges to the alignment process, as discussed in (Tombrink et al., 2024).

Dataset II corresponds to the TUM Room 2 sequence, which shows large covariance values for the translation components at the beginning of the estimated trajectory. This can be attributed to the handheld dynamic initialization, which makes IMU-related initialization more challenging.

Dataset III comes from the EuROC MAV Vicon Room 01-03 sequence and exhibits increased covariance in both rotation and translation. In this dataset, independent automatic exposure control is enabled for each camera, leading to different shutter times and image brightness variations. These factors make stereo matching and feature tracking more challenging.

Dataset IV is based on the EuROC MAV Machine Hall 05 sequence, which introduces much more rapid motion compared to Dataset III. This significantly increases the difficulty of VIO estimation due to additional error sources, including motion blur in visual data, abrupt IMU error accumulation, and exacerbated time synchronization errors. These challenges are reflected in the much higher translational covariances shown in Figure 2.

## 4. Results

### 4.1 Estimation of Alignment Parameters

After performing the least squares estimation for the alignment parameters, their values and associated statistical metrics are computed. Table 1 presents the estimated alignment parameters, their standard deviations, and the a-posteriori variance factor of unit weight upon solution convergence for both weighted and unweighted approaches. Standard deviations of alignment estimates that are smaller, as well as a-posteriori variance factor values closer to 1, are highlighted in bold.

From Table 1, we observe that the weighted approach yields significantly smaller standard deviations for the estimates. This can be attributed to the fact that the VIO-estimated state covariance in the weighted approach is much smaller than in the unweighted method, which assumes a constant covariance of $1\ m^2$ for the estimated trajectory. Additionally, its a-posteriori variance factor is much closer to 1. Notably, in Dataset II, the weighted method produces an a-posteriori variance factor of 0.967, compared to $1.472 \cdot 10^{-3}$ for the unweighted method. By incorporating the VIO-based state covariance as the observation uncertainty model, a more reasonable stochastic model is established, as reflected in the a-posteriori variance factor.

Notably, Dataset IV shows the most significant differences in alignment parameter estimations between the unweighted and weighted methods. The weighted and unweighted estimations yield $t_x$ values of 4.322 and 4.299, $t_y$ values of -1.855 and -1.888, $r_z$ values of 37.690 and 37.851, and $\Delta_t$ values of 0.012 and -0.033, respectively. This variation is likely due to the large state covariance of the estimated trajectory in Dataset IV. However, its a-posteriori variance factor of 1.960 still suggests that the VIO-estimated state covariance remains underestimated.

Among all datasets, Dataset II provides the a-posteriori variance factor closest to 1, indicating a strong consistency between the VIO-estimated state covariance and the residuals of the alignment parameter estimation. This suggests a well-modeled state uncertainty in the VIO system.

$$\rho_{i,j} = \frac{Cov(x_i, x_j)}{\sigma_{x_i}\sigma_{x_j}} \tag{9}$$

Figure 3 presents heatmaps depicting the correlation between alignment estimates for both weighted and unweighted methods, with values computed using Equation (9). The state covariance-based trajectory alignment generally results in

| Dataset | Weighted | $t_x \pm \sigma_{t_x}$ [m] | $t_y \pm \sigma_{t_y}$ [m] | $t_z \pm \sigma_{t_z}$ [m] | $r_z \pm \sigma_{r_z}$ [deg.] | $\Delta_t \pm \sigma_{\Delta_t}$ [s] | $\hat{\sigma}_o^2$ |
|---|---|---|---|---|---|---|---|
| I | no | 1.301 ± 0.030 | -0.433 ± 0.037 | 1.171 ± 0.029 | -174.422 ± 1.359 | 0.004 ± 0.028 | $9.637 \cdot 10^{-4}$ |
| | yes | 1.299 ± **0.001** | -0.434 ± **0.001** | 1.173 ± **0.001** | -174.535 ± **0.058** | 0.005 ± **0.001** | **0.643** |
| II | no | 0.716 ± 0.030 | -0.312 ± 0.032 | 1.267 ± 0.028 | -177.989 ± 2.211 | 0.030 ± 0.045 | $1.472 \cdot 10^{-3}$ |
| | yes | 0.712 ± **0.001** | -0.313 ± **0.001** | 1.267 ± **0.001** | -178.117 ± **0.087** | 0.030 ± **0.002** | **0.967** |
| III | no | 0.944 ± 0.028 | 2.064 ± 0.027 | 0.972 ± 0.022 | -6.479 ± 0.849 | -0.006 ± 0.026 | $9.483 \cdot 10^{-4}$ |
| | yes | 0.945 ± **0.002** | 2.063 ± **0.002** | 0.972 ± **0.001** | -6.482 ± **0.054** | -0.006 ± **0.002** | **0.308** |
| IV | no | 4.299 ± 0.039 | -1.888 ± 0.035 | 0.544 ± 0.033 | 37.851 ± 0.263 | -0.033 ± 0.029 | 0.006 |
| | yes | 4.322 ± **0.003** | -1.855 ± **0.002** | 0.545 ± **0.001** | 37.690 ± **0.029** | 0.012 ± **0.002** | **1.960** |

Table 1. Alignment parameter estimates, their standard deviations and a-posteriori variance factor.

lower correlation values (i.e., $|\rho_{i,j}|$ closer to zero), particularly for correlations involving z-axis rotation. For instance, in Dataset IV, the correlation between z-axis rotation and $t_x$ is 0.46 for the weighted method and -0.54 for the unweighted method. Similarly, in Datasets II and III, the correlation between z-axis rotation and $t_y$ is 0.48 and 0.55 for the weighted and unweighted methods, respectively. These findings suggest that alignment parameter estimates are more independently constrained by the input trajectory data rather than being strongly coupled through their uncertainties. As a result, the weighted approach yields a better-conditioned estimation process.

### 4.2 Evaluation of Aligned Trajectory

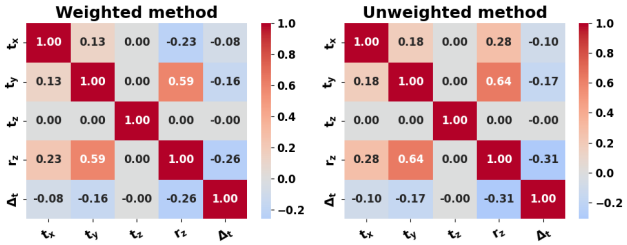
Once the alignment parameters are determined from the least squares estimation, they are applied to transform the estimated trajectory for alignment with the reference trajectory. The trajectory evaluation is then performed by computing trajectory errors using APE.

Table 2 presents the numerical values of APE for both unweighted and weighted approaches. Translational and rotational errors are reported in terms of their mean, standard deviation, and root-mean-square error (RMSE). The smaller value in each entry is bolded to highlight the contrast.

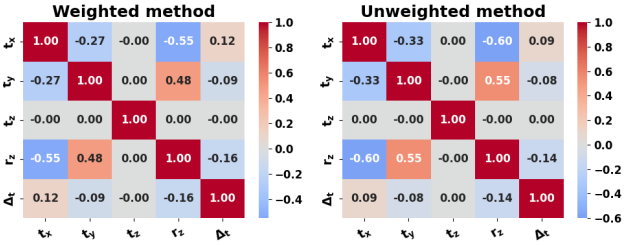
Overall, the difference in APE values between the two methods is not substantial. The results indicate that the weighted alignment yields smaller mean translational errors for Datasets I and II. Specifically, Dataset I shows mean translational errors of 0.0643 $m$ and 0.0645 $m$ for the weighted and unweighted methods, respectively, while Dataset II reports 0.0817 $m$ and 0.0818 $m$. Dataset III achieves lower mean rotational and RMSE values, with mean rotational errors of 2.6131° and 2.6157° for the weighted and unweighted methods, respectively, and RMS rotational errors of 2.6483° and 2.6508°. Additionally, Dataset IV exhibits a smaller standard deviation in translational error, measuring 0.0609 $m$ for the weighted method and 0.0742 $m$ for the unweighted method.

Dataset II exhibits the highest values across all rotational error metrics, likely due to unstable VIO initialization and the presence of substantial circular motion, which complicates orientation alignment and amplifies orientation errors.

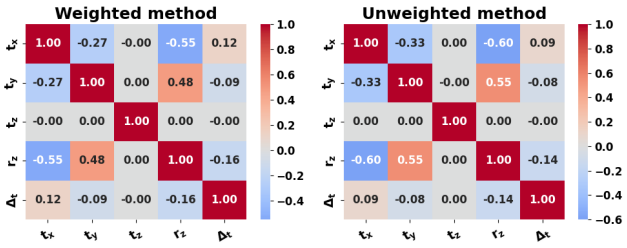
The largest difference in both translational and rotational errors between the unweighted and weighted methods is observed in Dataset IV, which corresponds to its large state covariance. Additionally, Dataset IV exhibits the highest translational error,



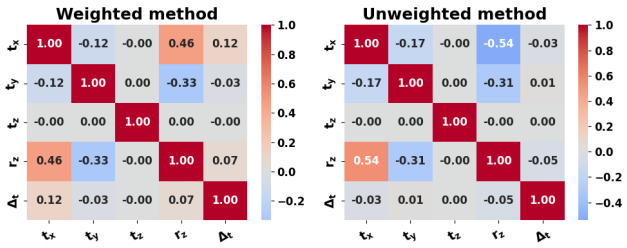

(a) Dataset I




(b) Dataset II




(c) Dataset III



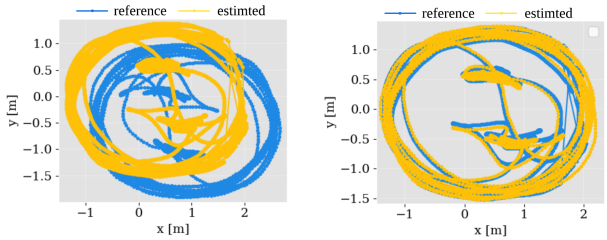

(d) Dataset IV

Figure 3. Heatmap comparison of alignment parameter correlations for Dataset I, II, III, IV.

aligning with its large state covariance in the estimated trajectory and the corresponding a-posteriori variance factor.

| | | Translational error [m] | | Rotational error [deg.] | |
|---|---|---|---|---|---|
| Dataset | Configuration | Mean $\pm$ $\sigma_t$ | RMS | Mean $\pm$ $\sigma_r$ | RMS |
| I | unweighted | 0.0645 + **0.0402** | **0.0760** | **1.1505** + **0.7347** | **1.3651** |
| | weighted | **0.0643** + 0.0407 | 0.0761 | 1.1602 + 0.7765 | 1.3961 |
| II | unweighted | 0.0818 + **0.0463** | **0.0940** | **3.1679** + **2.0795** | **3.7910** |
| | weighted | **0.0817** + 0.0467 | 0.0941 | 3.1914 + 2.1436 | 3.8445 |
| III | unweighted | 0.0489 + 0.0213 | 0.0533 | 2.6157 + 0.4302 | 2.6508 |
| | weighted | 0.0489 + 0.0213 | 0.0533 | **2.6131** + 0.4302 | **2.6483** |
| IV | unweighted | **0.1712** + 0.0742 | **0.1866** | **0.7258**+ **0.3451** | **0.8037** |
| | weighted | 0.1899 + **0.0609** | 0.1994 | 0.7630 + 0.4000 | 0.8615 |

Table 2. Comparison of translation and rotation errors across datasets.




(a) Prior-alignment (b) Post-alignment

Figure 4. 2D visualization of the estimated and reference trajectories for Dataset I, shown before and after alignment.

## 5. Conclusions

In this paper, we apply state covariance-based trajectory alignment using the estimated state covariance from a SOTA filtering-based VIO system, Open-VINS. Additionally, we discuss an approach for assessing the quality of VIO state covariance estimation based on the a-posteriori variance factor of unit weight, which is computed after the convergence of weighted alignment parameter estimation.

Both the alignment parameter estimation results and the trajectory evaluation outcomes are compared against those obtained from the unweighted alignment approach. Statistical analysis shows that the weighted method yields smaller standard deviations in alignment state, generally lower alignment state correlations, and a better a-posteriori variance factor compared to the unweighted approach.

While the APE values from both methods do not show significant differences, state covariance based trajectory alignment demonstrates smaller mean translational and rotational errors, as well as reduced standard deviations, for certain datasets.

## Acknowledgment

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC). Additionally, we extend our gratitude to Qiaozhuang Xu, Haoqing Li, and Ananya Vishwanath for their valuable technical discussions on this topic.

## References

Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M.W. and Siegwart, R., 2016. The EuRoC micro aerial vehicle datasets. The International Journal of Robotics Research, 35(10), pp.1157-1163.

Geneva, P., Eckenhoff, K., Lee, W., Yang, Y. and Huang, G., 2020, May. Openvins: A research platform for visual-inertial estimation. In 2020 IEEE International Conference on Robotics and Automation (ICRA) (pp. 4666-4672). IEEE.

Horn, B.K., 1987. Closed-form solution of absolute orientation using unit quaternions. Journal of the Optical Society of America A, 4(4), pp.629-642.

Kelly, J. and Sukhatme, G.S., 2011. Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration. The International Journal of Robotics Research, 30(1), pp.56-79.

Koch, K.R., 2014. Robust estimations for the nonlinear Gauss-Helmert model by the expectation maximization algorithm. Journal of Geodesy, 88, pp.263-271.

Liu, C., Wang, T., Li, Z. and Tian, P., 2025. A Novel Real-Time Autonomous Localization Algorithm Based on Weighted Loosely Coupled Visual–Inertial Data of the Velocity Layer. Applied Sciences, 15(2), p.989.

Moritz, H., 1972. Advanced least-squares methods. Ohio State University. Division of Geodetic Science.

Niu, X., Dai, Y., Liu, T., Chen, Q. and Zhang, Q., 2023. Feature-based GNSS positioning error consistency optimization for GNSS/INS integrated system. GPS Solutions, 27(2), p.89.

Salas, M., Latif, Y., Reid, I.D. and Montiel, J., 2015. Trajectory alignment and evaluation in SLAM: Horn's method vs alignment on the manifold. In Robotics: Science and Systems Workshop: The Problem of Mobile Sensors (pp. 1-3). SN.

Schubert, D., Goll, T., Demmel, N., Usenko, V., Stückler, J. and Cremers, D., 2018, October. The TUM VI benchmark for evaluating visual-inertial odometry. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 1680-1687). IEEE.

Sirtkaya, S., Seymen, B. and Alatan, A.A., 2013, July. Loosely coupled Kalman filtering for fusion of visual odometry and inertial navigation. In Proceedings of the 16th International Conference on Information Fusion (pp. 219-226). IEEE.

Teunissen, P.J., 2003. Adjustment theory. Netherlands: VSSD.

Tombrink, G., Dreier, A., Klingbeil, L. and Kuhlmann, H., 2024. Spatio-temporal trajectory alignment for trajectory evaluation. Journal of Applied Geodesy, (0).

Umeyama, S., 1991. Least-squares estimation of transformation parameters between two point patterns. IEEE Transactions on Pattern Analysis & Machine Intelligence, 13(04), pp.376-380.

Xu, Q., Gao, Z., Yang, C., Yang, H. and Wang, L., 2024. Credible positioning of BDS RTK/INS integration based on multi-information cross-validation. IEEE Transactions on Intelligent Transportation Systems.

Zhang, L., Ye, W., Yan, J., Zhang, H., Betz, J. and Yin, H., 2024. Loosely coupled stereo VINS based on point-line feature tracking with feedback loops. IEEE Transactions on Vehicular Technology.

Zhang, Z. and Scaramuzza, D., 2018, October. A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 7244-7251). IEEE.