Visual-Inertial Ego-Motion Estimation using Rolling-Shutter Camera in autonomous driving

Wanlong Li, Yu Tang, Chao Ding, Xueshi Li and Feng Wen

Abstract— **Camera and IMU are widely used in autonomous driving cars. Fusing visual and inertial measurements can be used for ego-motion estimation. In real use, we need to consider two issues. These two sensors are out of sync and the camera is rolling-shutter, whose image is captured row-by-row. In this paper, we present an efficient method to interpolate the IMU pose between consecutive poses and set up a novel feature measurement error model to cover the time delay issues. In the real test, the algorithm achieves a high accuracy in autonomous driving application. Further, we analyze output results for time delay influence judgment, once the judgment under a threshold the VIO output can be used, which is very useful in the real system.**

I. Introduction

An autonomous driving system may contain various sensors, including wheel odometer, IMU, GNSS, camera, LIDAR, RADAR, ultrasonic etc. Considering factors of both precision and cost, IMU and camera are usually grouped together for positioning purposes, known as VINS, Visual-Inertial Navigation System.

VINS has been well studied and widely used over the years [1]. Over the past two decades, visual-inertial state estimation has been studied extensively by the research community and many methods and frameworks have been presented. The representative techniques are either filter-based framework [2], or batch optimization [3,4,13]. The above methods require accurate and up-to-date calibration of the equipped sensor model to achieve good estimation performance. However, we notice two problems arise when VINS is applied in our autonomous driving system, which may cause significant drift in trajectory.

One problem is the time offset between camera and IMU. The temporal misalignment between IMU and camera is a typical issue in low-cost and self-assembled devices, which is widely used in autonomous driving cars. Because of reasons such as different hardware time delay and different clock domain, sensor measurements received at the same time from camera and IMU may have an offset in generation time. This problem is also known as visual-inertial temporal calibration. The work can be divided into two types based on the markers difference. We first discuss the representative marked-based approaches that rely on external markers such as checkerboard patterns. The Kalibr calibration toolbox [5] solves the problem in an offline manner using continuous-time batch estimation. However, as the time offset changes over time, using results of offline calibration may suffer from deviation during operation and may not be applicable to an autonomous driving system. Different from target-based, the approaches like [6] rely on natural features to calculate the time offset without the need for external markers such as checkerboards. [7, 8] proposed visual-inertial online temporal calibration method in VINS systems. [7] takes the time offset as an additional state variable to be estimated in the state vector in a filter-based framework. [8] solves the problem by jointly estimating time offset in an optimization-based framework.

The other problem is rolling shutter effect, since rolling shutter cameras are used as sensors in our system due to certain constraints. Different from a global shutter camera, which captures the whole image in an instant. A rolling shutter camera captures each row of an image sequentially, which means features extracted in different rows are captured at slightly different time. Rolling shutter effect may cause trajectory drift if simply ignoring the time shift. [9] solves the problem by assuming the camera velocity and angular velocity are constant between two consecutive frames. [10] gives a more accurate modelling of rolling shutter cameras.

Both problems mentioned above are time related and can be solved with the same idea. [11, 12] consider both problems together. [11] expresses the residual using Tayler-series expansion supposing the readout time is small, which is roughly the same as done in [10]. Calibration parameters are estimated in state vector of filtering framework including time offset. [12] adds both the time offset between camera and IMU and rolling shutter readout time as variables to be estimated in the state vector of MSCKF. The poses of camera frames are estimated by interpolation. According to the analysis of [14], the time offset generally is observable.

For the time offset between the camera and IMU and the readout-time of the rolling shutter camera, we derived a novel feature measurement model with these time delay and adopt the optimization framework. The main contributions of this paper are:

- We introduce a novel feature measurement error model, using *SO(3)* to interpolate the IMU pose between consecutive poses corresponding to feature frame, considering all features the time delay between the IMU and camera;
- For the optimization result, we analyze the main factor that influence the output accuracy and give a real test result;
- We validate our algorithm in our real system in autonomous driving application;

The authors are from the Noah's ark Lab of Huawei at Beijing, China. { liwanlong,tangyu17,dingchao24, lixueshi1, wenfeng3}@huawei.com.

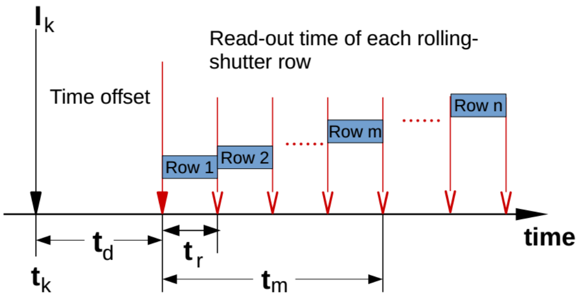


Fig.1. Time offset and rolling-shutter effect.

The rest of the paper is structured as follows. In Sect. II, we discuss the feature measurement model with time delay. The optimization algorithm is introduced in detail in Sect. III. The real-world experiment is shown in Sect. IV. Finally, in Sect V, we give a conclusion.

II. Feature Measurement Model With Time Delay

In this section, we will build a feature measurement model with a time difference between IMU and camera in an optimized VIO framework. This model also takes into account the effects of the rolling shutter camera, establish an alignment relationship with IMU measurements for different features.

Considering the characteristics of different sensors, the sensor devices on autonomous driving cars are mostly self-assembled. For low-cost and self-assembled visual inertial sensor groups, the camera and IMU are combined and there is no strict time synchronization. Due to the trigger delay, the transmission delay and the unsynchronized clock, the generated timestamp is not equal to the time at which the sample was measured. Therefore, there is usually a time offset between different measurements.

Also for rolling shutter cameras, the extra time offset introduced by the rolling-shutter effect should be considered. Specifically, the camera reads the imager row by row, so the time delay t_m of the pixel measurement in the m-th image-row, can be calculated as $t_m = mt_r$, where t_r is the read-out time of a single row.

As depicted in Fig.1, both the time offset t_d of the camera, as well as the rolling-shutter effect can be represented by a single time offset, corresponding to each row of pixels. In this paper, we consider the general case, where time offset t_d is a constant but unknown value.

Considering the effect above, the time delay of features on m-th image-row could be described as:

$$t = t_d + t_m. \tag{1}$$

To handle the time effect, we present the pose, I_{k+t} , corresponding to the feature j (see Fig.2) at image frame C_k,which is between two consecutive poses, I_k and I_{k+1}. For the three pose measurements, we apply a linear interpolation of position, and then interpolate the angle using the matrix exponential, to ensure that it remains a valid rotation matrix in *SO(3)*.

Defining { w } as the global frame of reference and a time-distance fraction λ between two consecutive poses as follows:

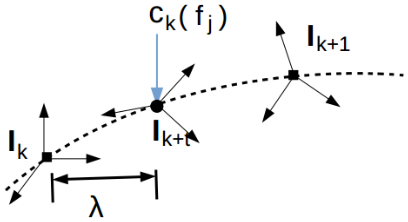


Fig.2. For the image sampling instant t_k, t_{k+1}, given two measurements I_k, I_{k+1} of IMU under the world coordinate system reference, we would like to interpolate to a new pose I_{k+t} according to the time offset of each feature.

$$\lambda = \frac{t}{t_{k+1}-t_k} \tag{2}$$

where t_k and t_{k+1} are the timestamps of the bounding measurements, and t is time we would like to interpolate to , can be refer to the (1).

We interpolate between the two pose readings, define two consecutive poses I_{k+1} as $\{R_w^{I_{k+1}}, p_{I_{k+1}}^w\}$ and I_k as$\{R_w^{I_k}, p_{I_k}^w\}$, which can be calculated by IMU propagation. The $p_{I_{k+t}}^w$ position can be easily express as:

$$p_{I_{k+t}}^w = (1-\lambda)p_{I_k}^w + \lambda p_{I_{k+1}}^w. \tag{3}$$

According to *SO(3)* matrix exponential and matrix logarithm:

$$R_w^{I_k} = \mathrm{Exp}(\theta_w^{I_k}) \tag{4}$$

$$Logv\left(R_w^{I_k}\right) = \theta_w^{I_k} \tag{5}$$

where $R_w^{I_k}, \theta_w^{I_k}$ are the rotation matrices and angles at the pose I_k, expressed in { w }.

Assuming a constant motion model, the rotation matric and angle at pose $I_{k+\mathrm{t}}$ can be expressed as follows:

$$\theta_{I_k}^{I_{k+t}} = \lambda\left[Logv\left(R_w^{I_{k+1}} {R_w^{I_k}}^T\right)\times\right] \tag{6}$$

$$R_w^{I_{k+t}} = \mathrm{Exp}(\theta_{I_k}^{I_{k+t}}) \tag{7}$$

$$R_w^{I_{k+t}} = \mathrm{Exp}\left(\lambda\left[Logv\left(R_w^{I_{k+1}} {R_w^{I_k}}^T\right)\times\right]\right) R_w^{I_k} \tag{8}$$

where $\{R_w^{I_{k+t}}, p_{I_{k+t}}^w\}$ is the IMU 3D pose of pose $I_{k+\mathrm{t}}$. The feature corresponding to the IMU can be calculated by Camera-IMU extrinsic parameters and feature global position.

III. Algorithm

A. Propagation

1)IMU Measurement Model: IMU measurement takes into account the force for countering gravity and are affected by acceleration bias b_a, gyroscope bias b_ω, and additive noise. The $\hat{a}_t$ and $\hat{\omega}_t$, which are the raw gyroscope and accelerometer measurements, are given by:

$$\hat{a}_t = a_t + R_w^t g^w + b_{at} + n_{at}$$

$$\hat{\omega}_t = \omega_t + b_{\omega t} + n_{\omega t} \tag{9}$$

where g^w is the gravitational acceleration, R_w^t is the rotation matrix between the world-frame and the body-frame, n_{at} and $n_{\omega t}$ are zero-mean white Gaussian noise processes. And the bias are modeled as random walk, whose derivatives are

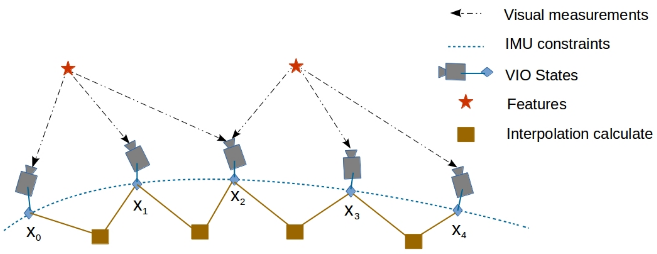


Fig.3. Illustration of the sliding window VIO. We maintain several camera frames and IMU measurements in a bundle and interpolate between IMU measurements corresponding to consecutive frames in the sliding window.

Gaussian white noise:

$$\dot{b}_{at} = n_{b_a} \text{ , } \dot{b}_{\omega t} = n_{b_\omega}. \tag{10}$$

2)Preintegration: Considering two consecutive frames c_k and c_{k+1}, whose corresponding time interval is $[t_k, t_{k+1}]$, there exists several inertial measurement. Given the bias estimation, we integrate them in the frame I_k as:

$$\alpha_{I_{k+1}}^{I_k} = \iint_{t\in[t_k,t_{k+1}]} R_t^{I_k}(\hat{a}_t - b_{at} - n_{at})\, dt^2$$

$$\beta_{I_{k+1}}^{I_k} = \int_{t\in[t_k,t_{k+1}]} R_t^{I_k}(\hat{a}_t - b_{at} - n_{at})\, dt$$

$$\gamma_{I_{k+1}}^{I_k} = \int_{t\in[t_k,t_{k+1}]} \frac{1}{2}\Omega(\widehat{\omega}_t - b_{\omega t} - n_{\omega t})\gamma_t^{I_k}\, dt \tag{11}$$

where

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

The covariance $P_{I_{k+1}}^{I_k}$ of α , β and γ also propagates accordingly. It can be seen that by using I_k as the reference frame for a given deviation, the preintegration term (11) can be obtained only by IMU measurements.

Given two time instants corresponding to image frames c_k and c_{k+1}, the position, velocity, and attitude states during the period, can be propagated by IMU measurements in the world coordinate system frame reference as follows:

$$p_{I_{k+1}}^w = p_{I_k}^w + v_{b_k}^w \Delta t_k - \frac{1}{2} g^w {\Delta t_k}^2 + R_{I_k}^w \alpha_{I_{k+1}}^{I_k}$$

$$v_{I_{k+1}}^w = v_{I_k}^w - g^w \Delta t_k + R_{I_k}^w \beta_{I_{k+1}}^{I_k}$$

$$q_w^{I_k} \otimes q_{I_{k+1}}^w = \gamma_{I_{k+1}}^{I_k} \tag{12}$$

where Δt_k is the duration between the time interval $[t_k, t_{k+1}]$.

B. Optimization

After a comprehensive analysis of the time delay, we focus on the establishment of the feature measurement model. Shown as Fig.2, we would like to interpolate a new pose I_{k+t} between IMU measurements on the image sampling instant t_k, t_{k+1}, so as to align feature measurements with time delay.

For the feature f_j observed on the m-th image-row, the corresponding measurement model is as follows:

$$z_k^j = h(p_{f_j}^{c_k}) + n_k^j,\ n_k^j \sim N(0, R_{k,j}) \tag{13}$$

where $p_{f_j}^{c_k}$ is the feature position expressed in the camera frame of reference at the exact time instant that the m-th image-row was read. The corresponding IMU measurement here, with a time offset t with respect to t_k, is calculated by interpolation. The specific expression is as follows:

$$p_{f_j}^{c_k} = R_I^c (R_w^{I_{k+t}} \left(p_{f_j}^w - p_{I_{k+t}}^w\right) - p_c^I). \tag{14}$$

To reduce the computational complexity of the optimization based VIO, marginalization is incorporated. As depicted in Fig.3, we maintain several camera frames and IMU measurements in a bundle and interpolate between IMU measurements corresponding to consecutive frames in the sliding window.

The residual corresponding to this measurement can be computed as:

$$r_k^j = z_k^j - h(\hat{p}_{f_j}^{c_k}). \tag{15}$$

We will add time compensation to optimize the measurement model based on a typical framework of visual inertia optimization. Visual-inertial odometry is formulated as a nonlinear optimization problem that tightly couples visual and inertial measurements. A local bundle adjustment (BA) jointly optimizes camera and IMU states, as well as feature locations.

The state vector augmented with time offset is defined as:

$$\mathcal{X} = [x_{I_0},\ x_{I_1}, \dots x_{I_n},\ p_{f_0}^w,\ p_{f_1}^w, \dots\ p_{f_l}^w,\ t_d]$$

$$x_{I_k} = [p_{I_k}^w,\ v_{I_k}^w,\ R_{I_k}^w,\ b_a,\ b_\omega] \tag{16}$$

where the k-th IMU state consists of the position $p_{I_k}^w$, velocity $v_{I_k}^w$, orientation $R_{I_k}^w$ in the world frame, and IMU bias b_a, b_ω in the local body frame. The feature $p_{f_l}^w$ is 3D position in the world frame.

The whole problem is formulated as a cost function containing IMU propagation factor, reprojection factor, and a certain prior factor:

$$\min_{\mathcal{X}} \left\{ \left\|r_p - H_p \mathcal{X}\right\|^2 + \Sigma_{k\in\mathcal{B}} \left\|r_{\mathcal{B}}\left(\hat{z}_{I_{k+1}}^{I_k}, \mathcal{X}\right)\right\|_{P_{I_{k+1}}^{I_k}}^2 + \right.$$

$$\left. \Sigma_{(l,j)\in\mathcal{C}} \left\|r_{\mathcal{C}}\left(\hat{z}_l^j, \mathcal{X}\right)\right\|_{P_l^j}^2 \right\} \tag{17}$$

where r_p and H_p are priors obtained via marginalization. $r_{\mathcal{B}}\left(\hat{z}_{I_{k+1}}^{I_k}, \mathcal{X}\right)$ is the error term from IMU propagation. $\mathcal{B}$ and $\mathcal{C}$ are the set of IMU and camera measurements. $r_C\left(\hat{z}_l^j, X\right)$ is the proposed visual reprojection error, which includes the time offset.

C. Analysis

Assuming the time offset is constant parameter, we put it in the system state for optimization. Once the time offset converged, the VIO output will be closed to the ground truth, which means time offset is influence less to the system output. At this moment, the system output can be used for the vehicle ego-motion estimate.

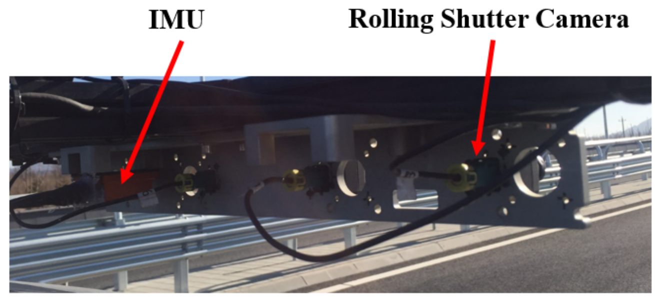


Fig.4. Self-assembled measurement equipment on the autonomous driving car, which contains IMU Xsens MTI-G-710 and Rolling shutter camera.

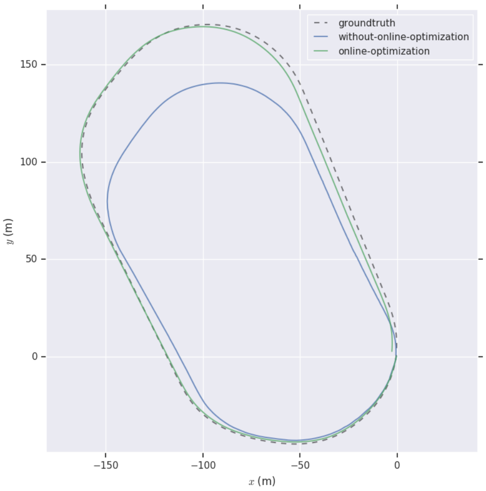


Fig.5. Trajectory of real-world experiment. The autonomous driving car is planned to travel around the playground.

We try to find a factor which can describe the time delay convergence. After every optimization, we check the time offset vary value, taking into account its influence to the feature reprojection error. We calculate the tracking features reprojection error before and after every optimization, record the average residual value. From the test result, we can find the reprojection error can be used for the time delay optimization judgment.

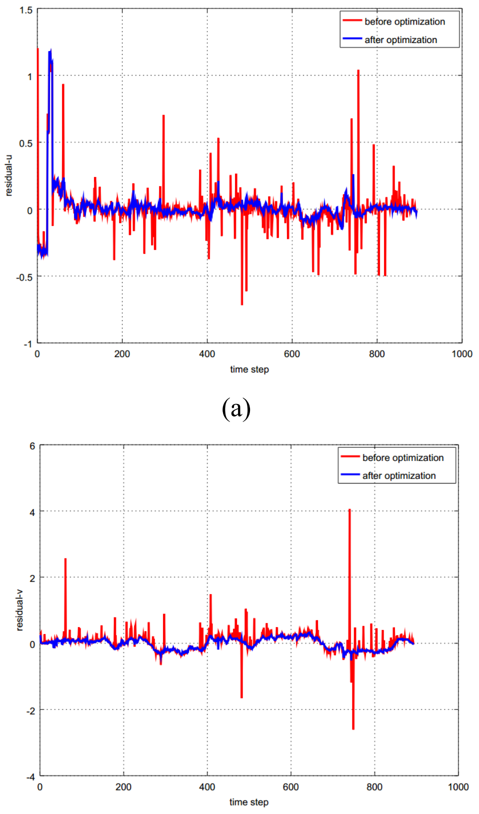
IV. Experiments

A. Real-world Experiments

In this section, we validated the performance of the proposed algorithm on an autonomous driving car equiped with rolling shutter camera and IMU as shown in Fig.5. The onboard computation resource is an Intel i9-8900K CPU running at 3.00 GHz .The equipped IMU is Xsens MTI-G-710, and the camera is Entroninc F001R100 rolling-shutter.

We use the visual-inertial optimization framework and the time delay influence is taking into the proposed visual factors (Sect. III-B).

We also considered the problem of degenerative motion in autonomous driving scenes. Owing to our vehicle motion must not be a constant angular velocity, so the time offset can be always observed during the experiment.



(a)

(b)

Fig.6. Reprojection residual before and after optimization, (a), (b) representing the residual of u and v coordinate directions on the pixel coordinate system.

In this experiment, we test the performance of autonomous trajectory tracking under state estimates from proposed algorithm. The autonomous driving car is planned to travel around the playground. The trajectory is as shown in Fig.5. In the absence of any time delay compensation, there is a scale difference between the real vehicle trajectory(blue line) and the ground truth(black dot line). Taking the time delay into account, the VIO output can converge quickly and the accuracy has significant improvement, while the result also shows our algorithm(green line) is closed to the ground truth.

B. Optimization Result Analysis

To analyze the VIO output, we compare the actual motion trajectory with the ground truth. We find the trajectory is closed to ground truth after the time delay is converged. During every optimization, we calculate the tracking features reprojection error before and after every optimization, record the average residual value.

The results shown in Fig.6, at the beginning, the reprojection error is large either before or after optimization. After a period of time (similar to time delay converged time), the reprojection error after optimization is also converged, although reprojection error has some jumps. Therefore, the reprojection error after optimization can be used for the judgment of the time delay convergence, while the VIO output can be used for ego-motion.

V. Conclusion

In this paper, we have presented a novel and efficient VIO ego-motion system with the time delay between IMU and rolling-shutter camera in autonomous driving application. Our experimental result shows the proposed approach can achieve high accuracy. We also analyze output result for time delay influence and find out one method to describe this influence. In the future work, we would consider the rolling shutter and IMU extrinsic online calibration influence, as in autonomous driving application the observability of spatial calibration is restricted.

References

[1] Huang, Guoquan. "Visual-inertial navigation: A concise review."*arXiv preprint arXiv:1906.02650* (2019).

[2] A. I. Mourikis and S. I. Roumeliotis, “A multi-state constraint Kalman filter for vision-aided inertial navigation,” in *Proc. of the IEEE Int. Conf. on Robot. and Autom.*, Roma, Italy, Apr. 2007, pp. 3565–3572.

[3] S. Shen, N. Michael, and V. Kumar, “Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft MAVs,” in *Proc. of the IEEE Int. Conf. on Robot. and Autom.*, Seattle, WA, May 2015

[4] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, “On-manifold preintegration for real-time visual–inertial odometry,” *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2017.

[5] Furgale, Paul, Joern Rehder, and Roland Siegwart. "Unified temporal and spatial calibration for multi-sensor systems." 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2013.

[6] A. Patron-Perez, S. Lovegrove, and G. Sibley, “A spline-based trajectory representation for sensor fusion and rolling shutter cameras,” *Int. J. Comput. Vis.*, vol. 113, no. 3, pp. 208–219, 2015.

[7] Li, Mingyang, and Anastasios I. Mourikis. "Online temporal calibration for camera–IMU systems: Theory and algorithms." *The International Journal of Robotics Research* 33.7 (2014): 947-964.

[8] Qin, T., & Shen, S. (2018, October). *Online temporal calibration for monocular visual-inertial systems. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 3662-3669). IEEE

[9] Li, Mingyang, Byung Hyung Kim, and Anastasios I. Mourikis. "Real-time motion tracking on a cellphone using inertial sensing and a rolling-shutter camera." *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013.

[10] Li, Mingyang, and Anastasios I. Mourikis. "Vision-aided inertial navigation with rolling-shutter cameras." *The International Journal of Robotics Research* 33.11 (2014): 1490-1507.

[11] Li, Mingyang, et al. "High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation." *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014.

[12] Guo, C. X., Kottas, D. G., DuToit, R., Ahmed, A., Li, R., & Roumeliotis, S. I. (2014, July). *Efficient Visual-Inertial Navigation using a Rolling-Shutter Camera with Inaccurate Timestamps. In Robotics*: Science and Systems.

[13] T. Qin, P. Li, and S. Shen, “Vins-mono: A robust and versatile monocular visual-inertial state estimator,” *arXiv preprint arXiv:1708.03852*,2017.

[14] Yang, Y., Geneva, P., Eckenhoff, K., & Huang, G. (2019). *Degenerate motion analysis for aided ins with online spatial and temporal sensor calibration.* IEEE Robotics and Automation Letters, 4(2), 2070-2077.