

Online Self-Calibration for Visual-Inertial Navigation Systems: Models, Analysis and Degeneracy

Journal Title
XX(X):1–31
©The Author(s) 2021

Yulin Yang¹, Patrick Geneva¹, Xingxing Zuo² and Guoquan Huang¹

Abstract

In this paper, we study in-depth the problem of online self-calibration for robust and accurate visual-inertial state estimation. In particular, we first perform a complete observability analysis for visual-inertial navigation systems (VINS) with full calibration of sensing parameters, including IMU and camera intrinsics and IMU-camera spatial-temporal extrinsic calibration, along with readout time of rolling shutter (RS) cameras (if used). We investigate different inertial model variants containing IMU intrinsic parameters that encompass most commonly used models for low-cost inertial sensors. With these models, the state transition matrix and visual measurement Jacobians are analytically derived and the observability analysis of linearized VINS with full sensor calibration is performed. The analysis results prove that, as intuitively assumed in the literature, VINS with full sensor calibration has four unobservable directions, corresponding to the system's global yaw and translation, while all sensor calibration parameters are observable given fully-excited 6-axis motion. Moreover, we, for the first time, identify primitive degenerate motions for IMU and camera intrinsic calibration, which, when combined, may produce complex degenerate motions. This result holds true for the different inertial model variants investigated in this work and has significant impacts on practical applications of online self-calibration to many robotic platforms. Extensive Monte-Carlo simulations and real-world experiments are performed to validate both the observability analysis and identified degenerate motions, showing that online self-calibration improves system accuracy and robustness to calibration inaccuracies. We compare the proposed *online* self-calibration on commonly-used IMUs against the state-of-art *offline* calibration toolbox Kalibr, and show that the proposed system achieves better consistency and repeatability. As sensor calibration plays an important role in making real VINS work well in practice, based on our analysis and experimental evaluations, we also provide practical guidelines for how to perform online IMU-camera sensor self-calibration.

Keywords

Sensor calibration, visualinertial systems, state estimation, observability analysis, degenerate motions, Monte Carlo analysis

1 Introduction

Due to the decreasing cost of integrated inertial/visual sensor rigs, visual-inertial navigation system (VINS), which fuses high-rate inertial readings from an IMU and images of the surrounding environment from a camera, has gained great popularity in 6 degree-of-freedom (DoF) motion tracking for mobile devices and autonomous robots – such as micro aerial vehicles (MAV) (Shen et al. 2014), self-driving cars (Heng et al. 2019), unmanned ground vehicles (UGV) (Zhang et al. 2021; Lee et al. 2020) and smart phones (Li and Mourikis 2013a; Guo et al. 2014) – during the past decades (Huang 2019). Many efficient and robust VINS algorithms based on filtering (Mourikis and Roumeliotis 2007; Li and Mourikis 2013a; Wu et al. 2015; Geneva et al. 2020; Eckenhoff et al. 2021) or batch least squares solver (Leutenegger et al. 2015; Forster et al. 2017; Qin et al. 2018; Usenko et al. 2020; Campos et al. 2021) techniques have been developed in recent years to address this pose estimation problem.

There are many factors which attribute to VINS performance, such as visual feature tracking/triangulation, velocity/biases initialization and sensor calibration. Among them, robust and accurate sensor calibration – including the rigid transformation between sensors (spatial calibration),

time offset between IMU-camera (temporal calibration), image line readout time for rolling shutter (RS) cameras, and IMU/camera intrinsics – is crucial, especially when plug-and-play visual-inertial sensor rigs with widely available off-the-shelf low-cost IMUs and rolling shutter cameras are deployed. In addition, sensor calibration itself can vary due to extended usage, sensor failure replacement and environmental effects such as varying temperature, humidity, vibrations, non-rigid mounting, and among others. For example, IMU biases and intrinsics suffer from the temperature and humidity changes (Li et al. 2014), and rigid transformation between IMU and camera can vary if the sensor is replaced or subjected to vibration. As such, online sensor self-calibration in VINS has attracted significant attentions and research efforts (Li and Mourikis 2013a; Guo

¹ Robot Perception and Navigation Group, University of Delaware, Newark, DE 19716, USA. ² Department of Informatics, Technical University of Munich, Garching b. Munich, 85748, Germany.

Corresponding author:

Yulin Yang, Department of Mechanical Engineering, University of Delaware, 126 Spencer Lab, Newark, DE 19716, USA.
Email: yuyang@udel.edu

et al. 2014; Yang et al. 2019; Eckenhoff et al. 2021; Yang et al. 2020b) in recent years, due to its potential to handle poor prior calibration or calibration changes, which can degrade the state estimate accuracy in the case that these calibrations are treated to be true.

System observability analysis for VINS with online IMU-camera (Mirzaei and Roumeliotis 2008; Kelly and Sukhatme 2011; Guo and Roumeliotis 2013) or IMU/camera intrinsic (Tsao and Jan 2019; Yang et al. 2020b) calibration has also been carried out to show that these calibration parameters can be identified given fully excited motions. However, complete analysis for VINS with full calibration parameters – including IMU/camera intrinsics, IMU-camera rigid transformation, temporal time offset, and camera RS readout time – is still absent from the existing literature.

Blindly performing online calibration is risky, as in most cases domain knowledge on specific motions and prior distribution choices are needed to ensure calibration can converge consistently (Schneider et al. 2019). In the meantime, existing research efforts (Li and Mourikis 2014; Yang et al. 2019, 2020b) have also identified several basic motion profiles, termed degenerate motions, that cause online sensor self-calibration failures. In this work, we investigate degenerate motions which impact the deployment of VINS on mobile robots, which typically have constrained motions, when jointly estimating IMU/camera intrinsics, IMU-camera spatial-temporal calibration, and RS readout time. For example, aerial and ground vehicles can only perform a few motion profiles due to their under-actuation, and can easily “fall” into degenerate conditions for calibration during typical deployment. As compared to our investigation into degenerate motions when performing full-parameter self-calibration, and their impacts especially for under-actuated autonomous robots, most approaches on VINS sensor self-calibration are limited to either handheld or trajectory segments involving rich motion information (Li et al. 2014; Schneider et al. 2019).

In this paper, we build an accurate and robust monocular VINS estimator with full self-calibration. We also investigate in-depth the observability analysis for visual-inertial self-calibration and perform degenerate motion analysis for *all* calibration parameters. In particular, the main contributions of this work include:

- An efficient filter-based visual-inertial estimator capable of performing self-calibration for all spatial-temporal extrinsic and intrinsic parameters.
- We perform a complete observability and degeneracy analysis for the proposed visual-inertial models and, for the first time, identify the degenerate motions that cause IMU and camera intrinsic parameters to become unobservable.
- Extensive simulations and real-world experiments are performed to verify the parameter convergence of the estimator with online self-calibration under fully-excited 6DoF motion and a series of identified degenerate motions of practical significance.
- We show that under general motion, self-calibration is necessary to achieve accurate pose estimation in a robust manner for consumer grade sensors, which continue to become more prevalent, with only minimal computational impact. Additionally, we show that

degenerate motions can and do have a significant negative impact on the performance of the estimator, leading to a series of guideline recommendations.

The rest of the paper is organized as follows: After reviewing the related work in Section 2 and estimation preliminaries in Section 3, we present the sensing models including inertial and camera models in Section 4. In Section 5, we derive the linearized system dynamics and measurement model of the VINS with full sensor calibration. Based on that, we perform the observability analysis and degenerate motion identification in Sections 6 and 7, while the proposed estimator is presented in Section 8. In Sections 9, 10, 11 and 12, we extensively validate our analysis and estimator through both simulations and real-world experiments. Finally, we offer discussions and final remarks in Sections 13 and 14.

2 Related Works

Extensive works have studied online or offline IMU and camera calibration for VINS. However, the joint self-calibration of all the calibration parameters for visual and inertial sensors (including IMU/camera intrinsics, IMU-camera spatial-temporal calibration and RS readout time) was not investigated sufficiently, and the observability analysis and degenerate motion identification for the complete VINS with calibration are still missing from the literature. In terms of the complete parameter calibration for VINS, the related works can be divided into the following four categories:

2.1 Camera Calibration

Visual-only offline calibration of camera intrinsic parameters is well studied (Hartley and Zisserman 2004). For example, Faugeras et al. (1992) demonstrated camera self-calibration without a pattern, based on which Qian and Chellappa (2004) and Civera et al. (2009) proposed to use sum of Gaussian filters. Recently, Agudo (2021) extended the above works to use RGB video with objects of non-rigid shape for camera self-calibration.

Many works have also focused on RS camera calibration, especially for the image line readout time. For instance, after formulating the geometric models of RS cameras and investigating how RS affects the image generation, Meingast et al. (2005) leveraged flashing LED lights to calibrate the RS readout time. Similarly, Oth et al. (2013) used continuous time trajectory representation to model RS effects and performed calibration with an April tag pattern (Olson 2011). However, these methods assume known camera intrinsics (including distortion parameters) when calibrating the RS readout time.

Nguyen and Lhuillier (2017) investigated the self-calibration of multiple omnidirectional RS cameras. They first initialized the rigid transformation and time offset between cameras based on the structure-from-motion (SFM) trajectories generated from each monocular camera with global shutter (GS) assumption. Then, all the camera related parameters (i.e., the rigid transformation, time offset, camera intrinsics, and RS readout time for each camera) are refined by a bundle adjustment (BA) with all the camera

measurements. Kukelova et al. (2020) presented the first minimal solution to absolute pose estimation of a RS camera with unknown focal length and unknown radial distortion, from seven point correspondences. With the proposed solvers they can achieve accurate solutions for camera poses, RS readout time, focal length and radial distortion. Note that both Kukelova et al. (2020) and Nguyen and Lhuillier (2017) are offline calibration algorithms and rely only on visual sensors. Karpenko et al. (2011) proposed to combine a gyroscope for RS camera image correction with natural scenes. However, it requires the pre-calibration of rotation between gyroscope and RS camera and the motion of the camera is restricted to rotation only. Tsao and Jan (2019) investigated online camera intrinsic calibration (only focal length and principal points) within a VINS framework. In contrast to the above works only focusing on camera calibration, our proposed method fuses the measurements of an IMU and a monocular camera, and provides online calibration of all the camera parameters including camera intrinsics and RS readout time. Both radial-tangential (*radtan*) and equivalent-distant (*equidist*) distortion (Furgale et al. 2013) are supported.

2.2 IMU Intrinsic Calibration

Generally, the gyroscope and acceleration biases are needed for accurate inertial modeling. They are both modeled as random walks and estimated as part of IMU states. It is a common practice to estimate biases online in VINS such as Jones and Soatto (2011), Kelly and Sukhatme (2011), Leutenegger et al. (2015) and Mourikis and Roumeliotis (2007).

Besides these biases, the IMU intrinsic parameters – including the scale correction and axis misalignment for gyroscope and accelerometer, the rotation from gyroscope or accelerometer frame to IMU frame, and the gravity sensitivity – also need to be calibrated offline or online, especially for low-cost inertial sensors. Bristeau et al. (2011) calibrated the IMU’s scale correction and axis misalignment for aerial vehicles by solving a least squares problem with known special sensor motions. Xiao et al. (2019) improved the IMU pre-integration (Forster et al. 2017) to incorporate the IMU intrinsic parameters in a keyframe based VINS algorithm for online self-calibration. Jung et al. (2020) studied IMU intrinsic calibration within multi-state constrained Kalman filter (MSCKF) by using a stereo camera and an IMU sensor, where they also examined the inertial calibration results under planar and random motions.

Building upon our prior work (Yang et al. 2020b), in which we have investigated online IMU intrinsic calibration with the minimal sensor configuration of a single IMU and a monocular camera and compared the performance of four different IMU intrinsic model variants in VINS, in this work, we study 18 different IMU intrinsic model variants which can encompass or be equivalent to most published IMU models for inertial navigation and perform online self-calibration. Comprehensive degenerate motion analysis, which can cause online self-calibration to fail, is also provided.

2.3 Joint IMU-Camera Self Calibration

Since VINS fuses IMU measurements and camera images, the joint calibration of IMU-camera parameters is preferred to improve system accuracy and robustness. Extensive works have studied joint sensor calibration in VINS. For instance, Mirzaei and Roumeliotis (2008) proposed to use an extended Kalman filter (EKF) for the spatial calibration (i.e. the rigid transformation between the camera and IMU) of VINS and performed an observability analysis. They showed that the rigid transformation is not fully observable under one-axis rotation. Zachariah and Jansson (2010) proposed to use the recursive Sigma-Point Kalman filter to estimate IMU intrinsics and IMU-camera spatial parameters with measurements from an IMU and a monocular camera. However, Mirzaei and Roumeliotis (2008) and Zachariah and Jansson (2010) did not calibrate the camera intrinsics or IMU-camera time offset and both relied on calibration chessboards.

Furgale et al. (2013) developed the well-known calibration toolbox: Kalibr, a continuous-time spline-based batch estimator, for IMU-camera extrinsics, time offset and camera intrinsics calibration. Rehder et al. (2016) extended Kalibr to incorporate and estimate IMU intrinsics (including scaling parameters, axis misalignments, and gravity sensitivity). Huai et al. (2021) further extended the above work to calibrate readout time for RS cameras. Nikolic et al. (2016) formulated a maximum likelihood estimation problem based on discrete IMU poses to calibrate the IMU-camera spatial-temporal and IMU intrinsic parameters. The above mentioned works are all offline methods and need calibration targets. In addition, they do not support full-parameter joint optimization of camera intrinsics with other calibration parameters.

Schneider et al. (2019) reduced IMU-camera calibration optimization complexity by selecting the most informative trajectory segments for calibration. The selection is based on the information matrix of the measurements from the trajectory segments. Although this work does not need a calibration board, the temporal calibration between IMU and camera is not included.

Many recent VINS algorithms perform online IMU-camera joint calibration. Qin et al. (2018) and Qin and Shen (2018) is able to perform online IMU-camera extrinsic and time offset calibration with natural scene to improve the system robustness and accuracy. Guo et al. (2014) proposed to use linear pose interpolation to model RS effects and calibrate readout time. Eckenhoff et al. (2019) proposed a multi-camera aided VINS with online IMU-camera spatial-temporal and camera intrinsic calibration. Eckenhoff et al. (2021) further proposed a generalized polynomial based pose interpolation for readout time calibration of RS cameras. However, the IMU intrinsics were not considered in the above systems. The closest work to ours is by Li et al. (2014) which included IMU-camera extrinsics, time offset, rolling-shutter readout time, camera and IMU intrinsics into the state vector within MSCKF (Mourikis and Roumeliotis 2007) based visual-inertial odometry and successfully calibrated all these parameters. They showed that these parameters can converge in simulation with fully excited motions and verified the system using a real-world experiment. However, system observability and degenerate motion analysis are

still missing, which are the focus of our work along with more extensive multi-run statistical validations. In addition, we also compared different IMU model variants which have appeared in literature to investigate their estimation performances within VINS.

2.4 Observability and Degeneracy

Observability analysis plays an important role in state estimation (Huang et al. 2010; Martinelli 2012; Hesch et al. 2013), especially when the system incorporates calibration parameters (Martinelli 2011; Li and Mourikis 2014; Yang et al. 2019, 2020b). We wish to identify whether these calibration parameters can be calibrated with visual and inertial measurements, and also identify degenerate motions, which might cause calibration to fail. In addition, observability properties can be leveraged for consistent estimator design (Huang 2012; Hesch et al. 2013; Wu et al. 2017a). Mirzaei and Roumeliotis (2008) performed the observability analysis for VINS with IMU-camera spatial calibration and showed that the spatial calibration is observable given fully excited motion of the IMU. They also found that that one-axis rotation is degenerate for the calibration of IMU-camera translation. Kelly and Sukhatme (2011) studied the IMU-camera self-calibration and performed nonlinear observability analysis using Lie derivative to show that the rigid transformation between IMU-camera is observable given random motions. Guo and Roumeliotis (2013) simplified the proof and analytically showed that the spatial calibration between the IMU and RGBD camera is observable. Li and Mourikis (2014) analyzed the identifiability for IMU-camera temporal calibration given the measurements of a single IMU and a monocular camera and identified a degenerate motion that can cause the IMU-camera time offset to become unobservable. Jung et al. (2020) studied the observability of stereo VINS with IMU intrinsics also based on Lie derivative to build the observability matrix and showed that the IMU intrinsics (including scale correction and axis misalignment for gyroscope and accelerometer, respectively) is observable given fully excited motions. Tsao and Jan (2019) built the observability matrix for VINS using linearized system model and showed that the camera intrinsics (only including focal length and principal points in their work) is observable. However, none of the above mentioned works ever performed and verified the observability analysis with full-parameter calibration for VINS.

In our previous work (Yang et al. 2019), we built the observability matrix for VINS using the linearized system with IMU-camera spatial-temporal calibration and showed that given fully excited motions all these calibration parameters are observable. We have also, for first time, identified four degenerate motions that can cause these calibration to become unobservable. In our recent work (Yang et al. 2020b), we performed observability analysis for monocular VINS with IMU intrinsic calibration (including scale and axis-misalignment for gyroscope and accelerometer, the rotation from gyroscope or accelerometer to IMU frame), and identified the degenerate motions for the IMU intrinsics. Building upon these prior works, in this work, we perform full-parameter calibration – including IMU intrinsics with gravity sensitivity, camera

intrinsics and the IMU-camera spatial-temporal calibration with RS readout time – for VINS with a single IMU and a monocular RS camera. Comprehensive observability analysis and degenerate motion identification are performed for these calibration parameters. Both simulations and real world experiments are also leveraged to verify our analysis.

3 Estimation Preliminaries

The state \mathbf{x} is propagated forward from timestep $k-1$ to k using incoming system control inputs \mathbf{u}_{k-1} based on the following generic nonlinear function:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}) \quad (1)$$

where $\mathbf{w}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1})$ denotes the white Gaussian noise of the control input. The state estimate at k can be predicted from the state estimate at $k-1$ with the nonlinear system:

$$\hat{\mathbf{x}}_{k|k-1} = f(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_{k-1}, \mathbf{0}) \quad (2)$$

where the subscript $i|j$ denotes the estimate at time i given the measurements up to time j ; $\hat{\mathbf{x}}$ denotes the estimated value for state \mathbf{x} and $\tilde{\mathbf{x}} = \mathbf{x} \boxminus \hat{\mathbf{x}}$ represents the error states, where “ \boxminus ” can be defined within a manifold (Barfoot 2017). The inverse operation of “ \boxminus ” can be defined as $\mathbf{x} = \hat{\mathbf{x}} \boxplus \tilde{\mathbf{x}}$, accordingly. The state covariance can be defined on the error states as $\tilde{\mathbf{x}} \sim \mathcal{N}(\mathbf{0}, \mathbf{P})$.

After linearizing the nonlinear function [see Eq. (1)] at current state estimate $\hat{\mathbf{x}}_{k-1|k-1}$, the propagated state covariance $\mathbf{P}_{k|k-1}$ for state estimate $\hat{\mathbf{x}}_{k|k-1}$ can be computed as:

$$\tilde{\mathbf{x}}_{k|k-1} \simeq \Phi_{k-1|k-1} \tilde{\mathbf{x}}_{k-1|k-1} + \mathbf{G}_{k-1} \mathbf{w}_{k-1} \quad (3)$$

$$\mathbf{P}_{k|k-1} = \Phi_{k-1} \mathbf{P}_{k-1|k-1} \Phi_{k-1}^\top + \mathbf{G}_{k-1} \mathbf{Q}_{k-1} \mathbf{G}_{k-1}^\top \quad (4)$$

where Φ_{k-1} and \mathbf{G}_{k-1} are the state transition matrix and noise Jacobians, respectively. A nonlinear measurement function can be described as:

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{n}_k \quad (5)$$

where $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ is white Gaussian noise. For the EKF update, we need to first linearize the above equation at the current state estimate $\hat{\mathbf{x}}_{k|k-1}$ as:

$$\mathbf{z}_k = \mathbf{h}(\hat{\mathbf{x}}_{k|k-1} \boxplus \tilde{\mathbf{x}}_{k|k-1}) + \mathbf{n}_k \quad (6)$$

$$\simeq \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) + \mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} + \mathbf{n}_k \quad (7)$$

$$\Rightarrow \tilde{\mathbf{z}}_k \triangleq \mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1}) \simeq \mathbf{H}_k \tilde{\mathbf{x}}_{k|k-1} + \mathbf{n}_k \quad (8)$$

where \mathbf{H}_k is the measurement Jacobian and $\tilde{\mathbf{z}}_k$ is the measurement residual. With these, we can now perform an EKF update to refine state estimates and covariance at time step k (Maybeck 1979):

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} \boxplus \mathbf{K}_k (\mathbf{z}_k - \mathbf{h}(\hat{\mathbf{x}}_{k|k-1})) \quad (9)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_{k|k-1} \quad (10)$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^\top (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k)^{-1} \quad (11)$$

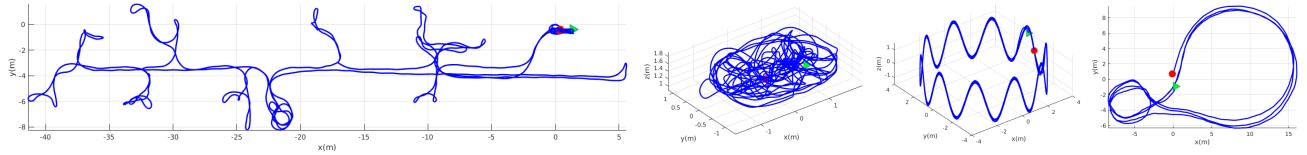


Figure 3. Simulated trajectories for Monte-Carlo simulations. Left: *tum_corridor* with fully excited 3D motion; Middle left: *tum_room* with 1 axis rotation and 3D translation; Middle right: *sine_3d* with constant acceleration along x direction; Right: *udel_gore* planar motion with constant z and only yaw rotation. The green triangle and red circle denote the beginning and ending of these trajectories, respectively.

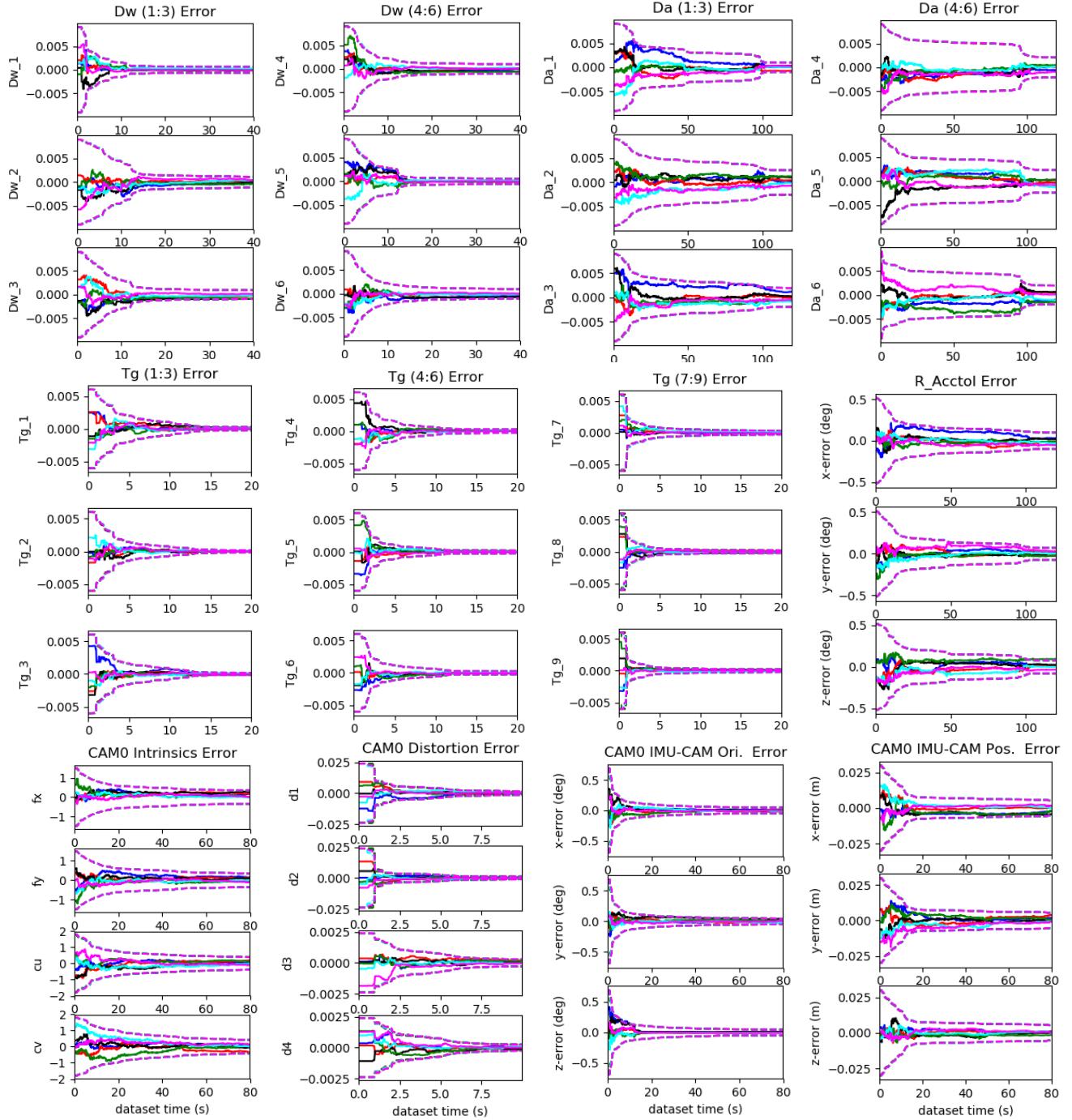


Figure 4. Calibration results for proposed system evaluated on *tum_corridor* with fully excited motion (using *imu22* and *radtan*). 3σ bounds (dotted lines) and estimation errors (solid lines) for six different runs (different colors) with different realization of the measurement noise and initial perturbations. All the calibration parameters converge nicely.

timestamp along the trajectory. The true angular velocity and linear accelerations can be directly found and corrupted

using the random walk biases and white noises. The basic configurations for our simulator are listed in Table 4.

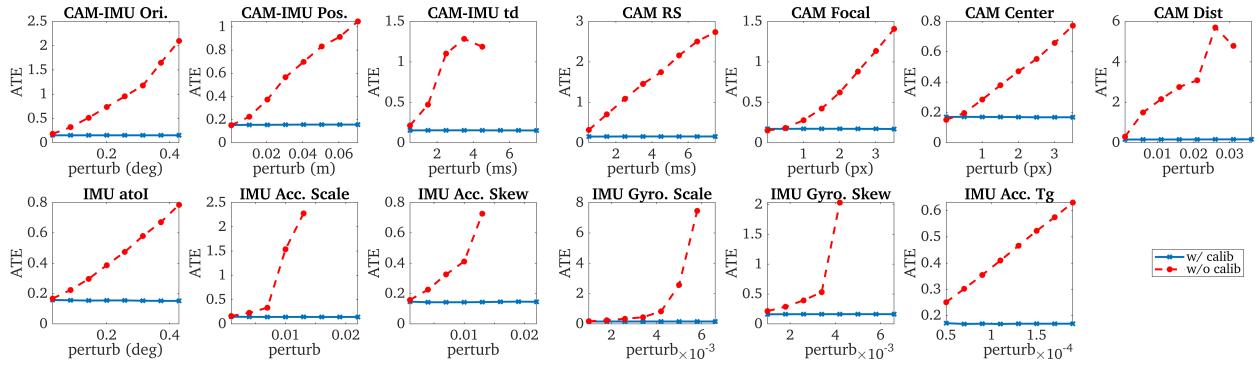


Figure 5. Absolute trajectory errors (ATE) using *imu22* and *radian* on the *tum_corridor* with full 3D motion given different levels of perturbation. Each parameter calibrated was only perturbed while other parameters were initialized to their true values and not estimated. ATE above eight meters were not reported in the figures and can be considered diverged.

Table 4. Simulation parameters and prior standard deviations that perturbations of measurements and initial states were drawn from.

Parameter	Value	Parameter	Value
IMU Scale	0.003	IMU Skew	0.003
Rot. atol (rad)	0.003	Rot. wtol (rad)	0.003
Gyro. White Noise	1.6968e-04	Gyro. Rand. Walk	1.9393e-05
Accel. White Noise	2.0000e-3	Accel. Rand. Walk	3.0000e-3
Focal Len. (px/m)	0.50	Cam. Center (px)	0.60
d1 and d2	0.008	d3 and d4	0.002
Rot. CtoI (rad)	0.004	Pos. InIC (m)	0.010
Readout Time (ms)	0.5	Timeoff (s)	0.005
Cam Freq. (hz)	20	IMU Freq. (hz)	400
Avg. Feats	100	Num. SLAM	50
Num. Clones	20	Feat. Rep.	GLOBAL

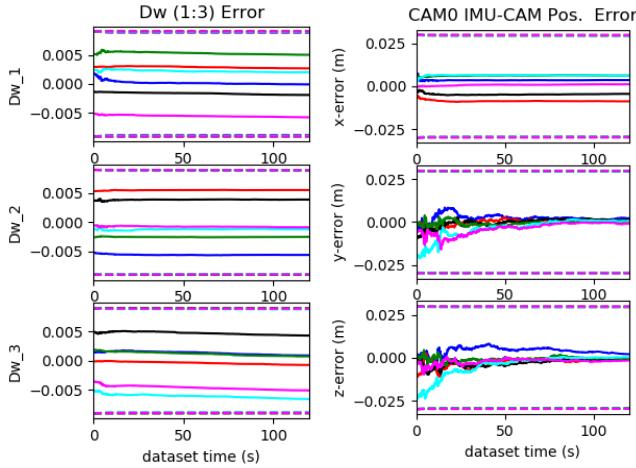


Figure 6. Calibration results for the proposed system evaluated on *tum_room* with one-axis rotation using *imu22* and *radian*. Note that the estimation errors and 3σ bounds for d_{w1} , d_{w2} , d_{w3} and the IMU-camera position calibration along the rotation axis can not converge.

All simulation convergence figures show the 3σ bounds (dotted lines) and estimation errors (solid lines) for six different runs (different colors) with different realization of the measurement noises and initial calibration state perturbations.

To simulate RS visual bearing measurements, we follow the logic presented by Li and Mourikis (2013a) and Eckenhoff et al. (2021). Specifically, static environmental

features are first generated along the length of the trajectory at random depths and bearings. Then, for a given imaging time of features in view, we project each into the current image frame using the true camera intrinsic and distortion model to find the corresponding observation row. Given this projected row and image time, we can find the pose at which this row should have been (i.e., the pose at which that RS row should have been exposed). We can then re-project this feature into the new pose and iterate until the projected row does not change (which typically requires 2-3 iterations). We now have a feature measurement which occurred at the correct pose for its given RS row. This measurement is then corrupted with white noise. The imaging timestamp corresponding to the starting row is then shifted by the true IMU-camera time offset t_d to simulate cross-sensor delay.

9.1 Simulation with Fully-Excited Motion

We first perform a general trajectory simulation, for which we perform full calibration of the IMU-camera extrinsics, time offset, RS readout time, camera intrinsics with *radian* model and IMU intrinsics with *imu22*. The trajectory, shown in the left of Figure 3, is designed based on *tum_corridor* sequence of TUM visual-inertial dataset with full excitation of all 6 axes and provides a realistic 3D hand-held motion (Schubert et al. 2018). From the results shown in Figure 4 and 9, the estimation errors and 3σ bounds for all the calibration parameters (including *imu22* and *radian*) can converge quite nicely, verifying that the analysis for general motions holds true. We plot results from six different realizations of the initial calibration guesses based on the specified priors, and it is clear that the estimates for all these calibration parameters are able to converge from different initial guesses to near the ground truth. Each parameter is able to “gain” information since their 3σ bounds shrink. These results verify our Lemma 1 that all these online calibration parameters are observable given a fully-excited motion.

9.2 Sensitivities to Perturbations

The next natural question is how robust the system is to the initial perturbations and whether the use of online sensor calibration enables improvements in robustness and accuracy. Shown in Figure 5, for each of the different calibration parameters we perturb it with different levels

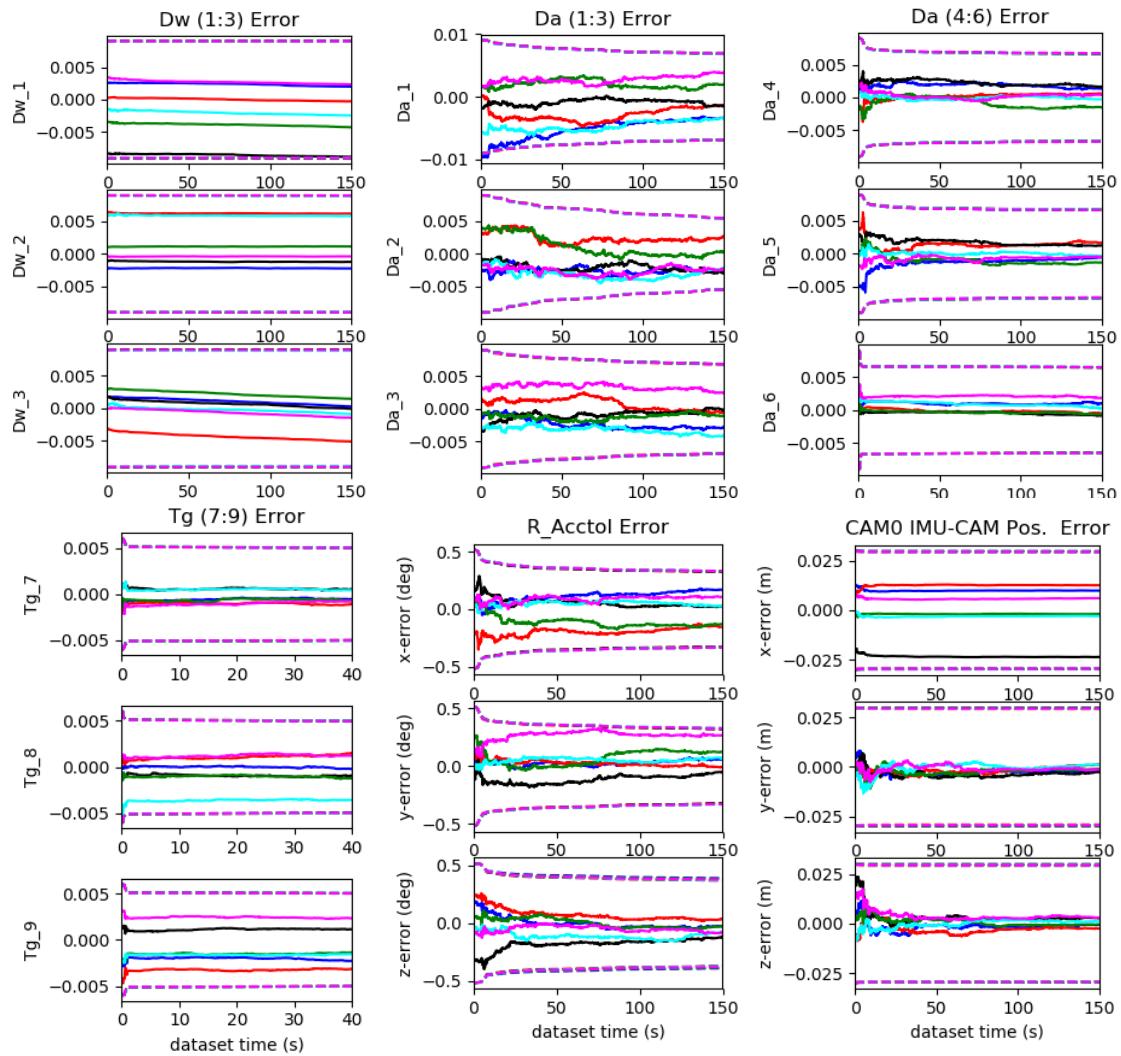


Figure 8. Calibration results of the proposed system evaluated on *udel_gore* with planar motion using *imu22* and *radtan*. With planar motion, the estimation errors and 3σ bounds of $d_{w1}, d_{w2}, d_{w3}, t_{g7}, t_{g8}, t_{g9}$ and the IMU-camera position cannot converge. Due to lack of motion excitation, the parameters of \mathbf{D}_a and ${}^I_a\mathbf{R}$ converge much slower than the other motion cases.

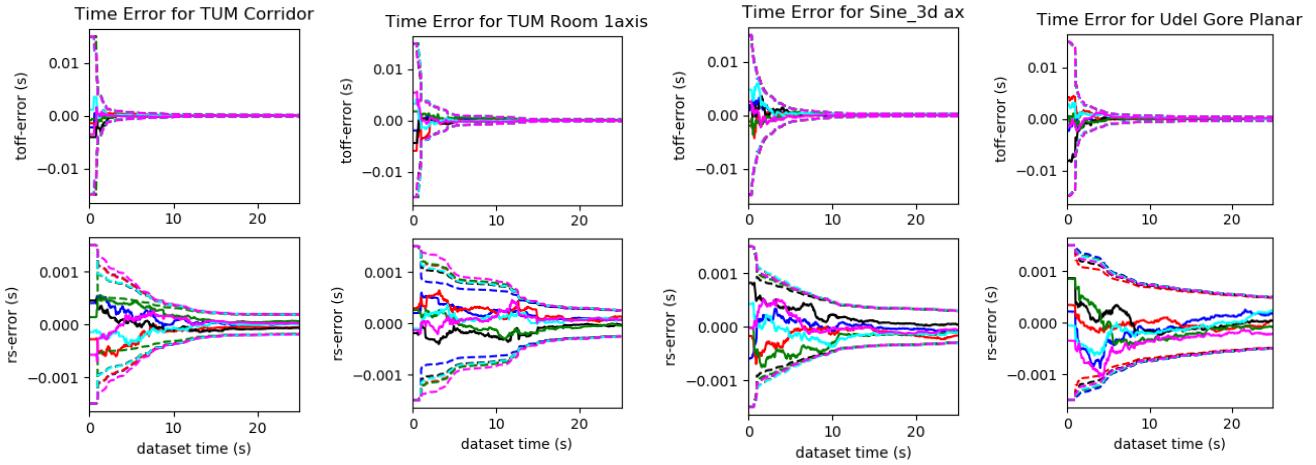


Figure 9. Camera temporal and read out time calibration results of the proposed system (using *imu22* and *radtan*) for different trajectories. The temporal parameters can finally converge in all the 4 motion types. Note that the readout time converges slower in the planar motion case probably due to the lack of motion in the beginning.

9.4 Degenerate Motion Verification

We now verify the identified degenerate motions and present results for three special motions. In all simulations, we perform full-parameter calibration to fully test our system

and present the complete results in Appendix 21. The trajectories shown in Figure 3 are created as follows:

- One-axis rotation with a modified *tum_room* trajectory, see middle left, which has its roll and pitch orientation

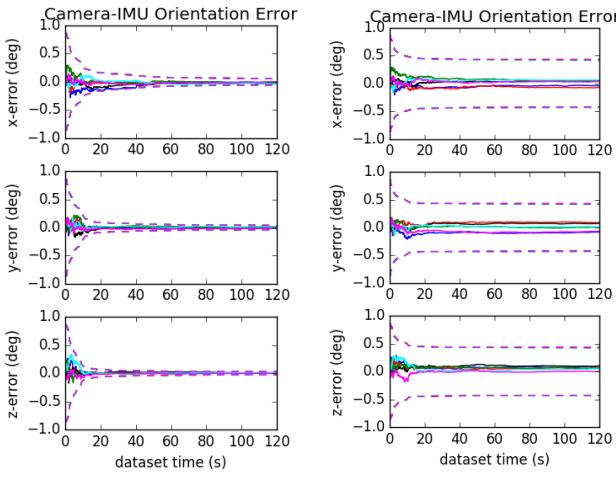


Figure 10. Camera to IMU orientation errors when using IMU *imu2* (left) and the over parameterized *imu5* (right). Note that only the IMU intrinsics and relative pose between IMU and camera were online calibrated.

changes removed to create a yaw and 3D translation only dataset.

- Constant local a_x with modified *sine_3d*, see middle right, for which we have a constant pitch and make the current yaw angle tangent to the trajectory in the x-y plane (gives constant local acceleration along local x-axis).
- Planar motion with modified *udel_gore*, see right, which has its roll and pitch orientation removed and all poses are projected to x-y plane (planar motion in the global x-y plane).

9.4.1 One-axis rotation motion. Shown in Figure 6, the first 3 parameters (d_{w1} , d_{w2} and d_{w3}) for \mathbf{D}_w do not converge at all (the 3σ bounds are almost straight lines), which matches our analysis, see Table 2. These parameters should be unobservable in the case of one-axis rotation with ${}^w w_x$ (roll) and ${}^w w_y$ (pitch) are constant. Additionally, the translation between IMU and camera does not converge either. The x-error of the IMU-camera translation even diverges reinforcing the undesirability of degenerate motions and verifies the analysis presented in Table 3.

9.4.2 Constant local acceleration motion. The results shown in Figure 7, where we have enforced that the local acceleration along the x-axis, a_x , is constant. The d_{a1} , and pitch and yaw of ${}^I \mathbf{R}$ does not converge, thus validating our analysis shown in Table 2. Note that in the simulation, we have set ${}^I \mathbf{R} \simeq \mathbf{I}_3$ and $\mathbf{D}_a \simeq \mathbf{I}_3$. Hence, ${}^a \dot{\mathbf{a}} \simeq {}^I \mathbf{a}$ and ${}^I a_x$ is also near constant. Therefore, three terms of gravity sensitivity (t_{g1} , t_{g2} and t_{g3}) are also unobservable and converge much slower than other terms.

9.4.3 Planar motion. Shown in Figure 8, with one-axis rotation (yaw axis) for planar motion the d_{w1} , d_{w2} and d_{w3} for \mathbf{D}_w and the IMU-camera translation are unobservable and does not converge. Since the ${}^I \mathbf{a}_z$ is constant, the last three terms of gravity sensitivity (t_{g7} , t_{g8} and t_{g9}) become unobservable and cannot converge. Both these results verify our analysis shown in Tables 2 and 3. Additionally, this trajectory is quite smooth with small excitation of linear

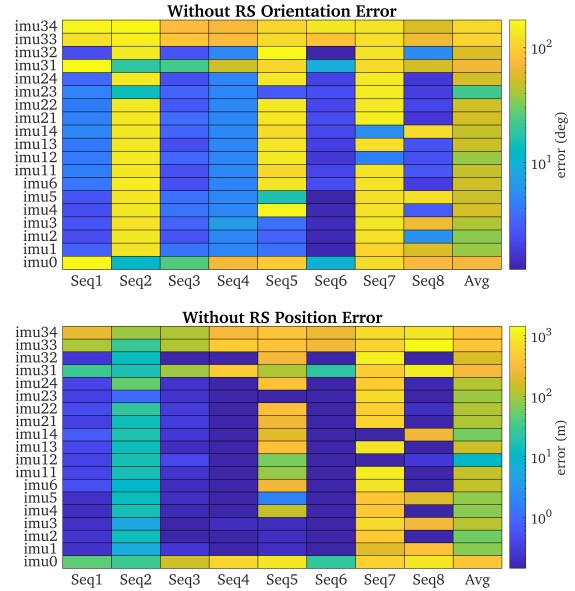


Figure 11. Results on TUM Rolling Shutter VIO Dataset without rolling shutter readout time calibration, with different IMU intrinsic models. The averaged absolute trajectory errors (ATE) of 5 runs in degree (top) and meters (bottom) are provided. Note that the camera intrinsics, and IMU-camera spatial-temporal calibration.

acceleration, hence, the terms of \mathbf{D}_a and ${}^I \mathbf{R}$ in general converge much slower than the fully excited motion case.

9.5 Simulated Over Parametrization

We now look to investigate the impact of poor choice of calibration parameters which *over parameterizes* the IMU intrinsics. The *imu5* model, see Table 1, is an over parametrization since we calibrate both 9 parameters for gyroscope and accelerometer, which causes the IMU-camera orientation to be affected since the intermediate inertial frame $\{I\}$ is not constrained. If we change the relative rotation from $\{I\}$ to $\{C\}$, then this perturbed rotation can be absorbed into the $\{a\}$ to $\{I\}$ and $\{w\}$ to $\{I\}$ terms. Thus, it means we have an extra 3 degrees of freedom (DoF) for rotation not constrained by our measurements. We compare this *imu5* model to its close equivalent *imu2* model in Figure 10. We can see that even though the trajectory fully excites the sensor platform, the convergence of ${}^I \mathbf{R}$ becomes much worse if we calibrate IMU-camera extrinsics and all 18 parameters for the IMU model *imu5* even when the *same* priors and measurements are used. This further motivates the use of the minimal calibration parameters to ensure fast and robust convergence of all state parameters.

10 Real-World TUM RS VIO Datasets

We first evaluate our proposed algorithms on the TUM RS VIO Dataset which contains a time-synchronized stereo pair of two uEye UI-3241LE-M-GL cameras (left: global-shutter and right: rolling-shutter) and a Bosch BMI160 IMU (Schubert et al. 2019). When collecting data, the cameras were operated at 20Hz while the IMU operated at 200Hz and an OptiTrack system captured the ground-truth motion. The dataset is provided in both “raw” and

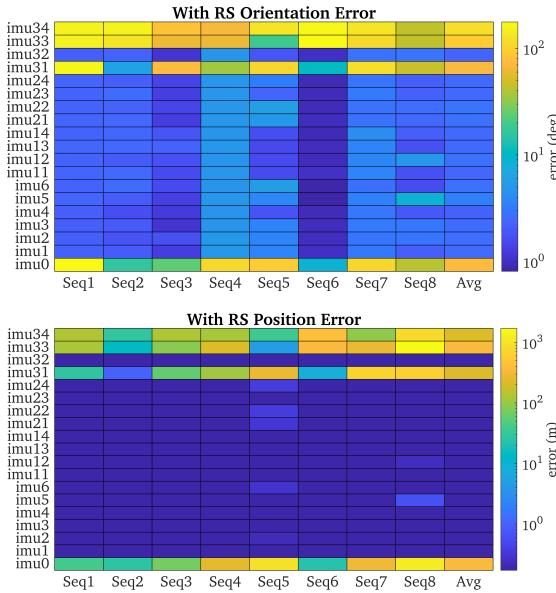


Figure 12. Results on TUM Rolling Shutter VIO Dataset with rolling shutter readout time calibration, with different IMU intrinsic models. The averaged absolute trajectory errors (ATE) of 5 runs in degree (top) and meters (bottom) are provided. Note that the camera intrinsics, and IMU-camera spatial-temporal calibration.

Table 6. Averaged absolute trajectory errors (ATE) of 5 runs over all 8 sequences of the TUM Rolling Shutter VIO Dataset with rolling shutter, camera intrinsics, and IMU-CAM spatial-temporal calibration.

IMU Model	ATE (deg)	ATE (m)
<i>imu0</i>	72.994	363.610
<i>imu1</i>	2.574	0.092
<i>imu2</i>	2.679	0.094
<i>imu3</i>	2.590	0.093
<i>imu4</i>	2.205	0.076
<i>imu5</i>	3.418	0.149
<i>imu32</i>	2.368	0.079

“calibrated” formats. The “calibrate” dataset has had IMU intrinsics calibrated from Kalibr (Furgale et al. 2013) pre-applied to the “raw” dataset along with some re-sampling. We evaluate our proposed system by using the right (RS) camera directly with the raw datasets, which have very noisy measurements with varying sensing rates. Hence, the raw datasets are more challenging compared to the calibrated datasets. We re-calibrated the camera intrinsics and IMU-camera spacial-temporal parameters using the raw calibration datasets as the provided calibration parameters were only for the calibrated datasets. We directly use Eq. (30) since the dataset timestamps correspond to the first row of the image. Note that we set the initial values for \mathbf{D}_a , \mathbf{D}_w , ${}^I_w\mathbf{R}$ and ${}^I_w\mathbf{R}$ as identity and \mathbf{T}_g as zeros, while the initial readout time for the whole RS image is set to 20ms as prior calibration. All IMU intrinsic models listed in Table 1 were run with and without RS calibration. The results are presented in the following sections.

10.1 RS Self-Calibration

The results are shown in Figure 11 and 12, with and without RS readout calibration, respectively. It is clear that the systems without RS readout time calibration and without IMU intrinsic calibration (*imu0* and *imu31-imu34*) are unstable and diverges to large orientation and positional errors. With IMU intrinsic calibration, the system still fails for certain datasets (*imu1-imu24*) and thus online readout time calibration will greatly improve the system robustness for RS cameras. The finally estimated RS readout time for each image is around 30ms, which means given the image resolution of 1280×1024 , the row readout time should be around 29us.

10.2 IMU Intrinsic Self-Calibration

We focus on the results in Figure 12 which has RS enabled. It is clear from the performance of *imu0* that without IMU intrinsic calibration the BMI160 IMU will cause large trajectory errors, with the models which do perform intrinsic calibration being an order of magnitude more accurate. Table 6 shows the average error over all sequences for the first 6 IMU models. It can be seen that the *imu5* model which over parameterizes the intrinsics has worst accuracy in both orientation and position trajectory estimates, while the accuracy of the other *imu1 - imu4* models is comparable to each other (similar accuracy level). We further do an ablation study with models *imu31 - imu34* to find the individual impact of each of the IMU intrinsic parameters. We can see that the *imu32* model which estimates \mathbf{D}_{w9} has large accuracy gains over the other four. This indicates that the readings from gyroscope of BMI160 are very noisy. The calibration of \mathbf{D}_{w9} dominates the performance of this VINS system, and just the calibration of it can achieve similar results as full IMU model calibration (see bottom of Table 6). Through all these we show that online IMU intrinsic calibration can enhance both the system robustness and accuracy.

10.3 Comparison to Kalibr Calibration

We run Kalibr’s offline calibration with *scale-misalignment** IMU model on 5 calibration datasets provided for the Bosch BMI160 IMU and treat these results as reference values to compare to the proposed *online* calibration results. The Kalibr calibration datasets were collected with the stereo camera pair both operating with a global shutter mode along with an April tag board (Furgale et al. 2013). For the results evaluation, we directly report $\mathbf{T}'_w = (\mathbf{D}'_w)^{-1}$, $\mathbf{T}'_a = (\mathbf{D}'_a)^{-1}$ and ${}^w_I\mathbf{R} = {}^I_w\mathbf{R}^\top$. By contrast, the proposed system is run on one RS camera with only temporal environmental feature tracks on the 8 data sequences using *imu6*, which is equivalent to the *scale-misalignment* IMU model of Kalibr.

As shown in the boxplots in Figure 13, even though we run on more challenging datasets in real-time, our proposed system can still achieve reasonable calibration results for \mathbf{D}'_w , \mathbf{D}'_a , \mathbf{T}_g and ${}^I_w\mathbf{R}$, which are close to the values from baseline Kalibr. Additionally we can see that the values of

*<https://github.com/ethz-asl/kalibr/wiki/Multi-IMU-and-IMU-intrinsic-calibration>

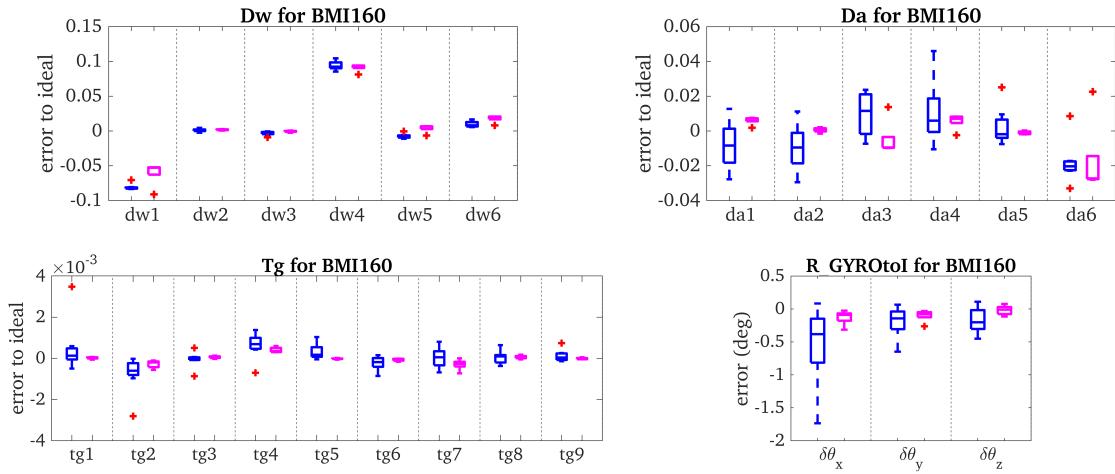


Figure 13. IMU intrinsic evaluation of Bosch BMI160 IMU used in TUM Rolling Shutter VIO datasets using the proposed method with *imu6* and Kalibr relative to the “ideal” sensor intrinsics. The boxplots show the final converged value of both methods. Kalibr (magenta, right in each group) was run with two global shutter cameras and a Bosch BMI160 IMU available over 5 calibration datasets using April tag board, while the proposed system (blue, left) was run with only one rolling shutter camera and the same IMU on the 8 data sequences without any tags. Note that *imu6* is equivalent to the *scale-misalignment* IMU model of Kalibr.



Figure 14. Visual-Inertial Sensor Rig contains a MicroStrain GX5-25 IMU, MicroStrain GX5-35, Xsens MTi 100, FLIR Blackfly camera and RealSense T265 tracking camera. The RealSense T265 tracking camera contains an integrated IMU and a fisheye stereo camera.

gravity sensitivity \mathbf{T}_g of the BMI160 IMU are generally one or two orders smaller than the other IMU intrinsics. This matches the results presented in Figure 12, for which the estimation errors of *imu1* - *imu4* (without gravity sensitivity) are similarly to those of *imu11* - *imu14* (with gravity sensitivity of 6 parameters) and *imu21* - *imu24* (with gravity sensitivity of 9 parameters). This means that the proposed system performance is less sensitive to gravity sensitivity, no matter 0, 6 or 9 parameters are used.

We can also see that the terms in scale-misalignment for gyroscope \mathbf{D}'_w for the BMI160 IMU are much larger than \mathbf{D}'_a and \mathbf{T}_g . This indicates that the readings from gyroscope of BMI160 are very noisy and the calibration of \mathbf{D}'_w dominates the performance of this VINS system. This fact is again confirmed by model *imu32*, which only calibrates the \mathbf{D}'_w and can achieve similar results as full IMU model calibration (see bottom of Table 6).

11 More Validation and Investigation of Camera and IMU Impacts

We further evaluate the proposed self-calibration system on a custom made visual-inertial sensor rig (VI-Rig, shown in Figure 14) which contains multiple IMU and camera sensors to facilitate an investigation into how they individually impact overall performance. Specifically, it contains a MicroStrain GX5-25, MicroStrain GX5-35, Xsens MTi 100, FLIR Blackfly camera and RealSense T265 tracking camera which contains an integrated BMI055 IMU along with a fisheye stereo camera. Here we note that all cameras used are not rolling shutter to ensure fair comparison against the baseline Kalibr (Furgale et al. 2013) which only supports IMU-camera calibration with global shutter cameras. In total 10 datasets were collected of an April tag board on which both the proposed system and the Kalibr calibration toolbox were run to report repeatability statistics and expected real-world performance of both systems. During data collection, all 6-axis motion of VI-Rig were excited to avoid degenerate motions for calibration parameters.

11.1 Visual Front-End and Initial Conditions

To provide a fair comparison, we modified the front-end of the proposed system to directly use the same April tag detection as Kalibr and to only use such tags during estimation. Additionally, while the proposed system was only run with one of the four IMUs and either the Blackfly or left T265 Realsense camera, Kalibr used all the available sensors to ensure the highest and most consistent performance (4 IMUs and 3 cameras). The *imu6* model is used during evaluation, which is equivalent to the *scale-misalignment* IMU model of Kalibr. We define the “ideal” IMU sensor intrinsics as $\mathbf{D}'_w = \mathbf{D}'_a = \mathbf{I}_3$, ${}^I_w\mathbf{R} = {}^I_a\mathbf{R} = \mathbf{I}_3$ and $\mathbf{T}_g = \mathbf{0}_3$ if factory or offline calibration has been pre-applied. Generally, these values are what the users expect for a ready-to-use IMU, and are the initial values that the proposed estimator starts from. The quality of each IMU can

Table 7. Average processing time for each image (including propagation and update) for the proposed system with (w/) and without (w/o) online calibration (unit: second) on the 10 datasets collected with VI-Rig. The time increase (0.0036s in average) for online calibration is negligible compared to no calibration.

Algorithm	Data-1	Data-2	Data-3	Data-4	Data-5	Data-6	Data-7	Data-8	Data-9	Data-10	Avg.
w/ online calib	0.0227	0.0231	0.0228	0.0223	0.0229	0.0220	0.0214	0.0228	0.0223	0.0215	0.0224
w/o online calib	0.0187	0.0191	0.0189	0.0190	0.0190	0.0185	0.0186	0.0186	0.0189	0.0190	0.0188

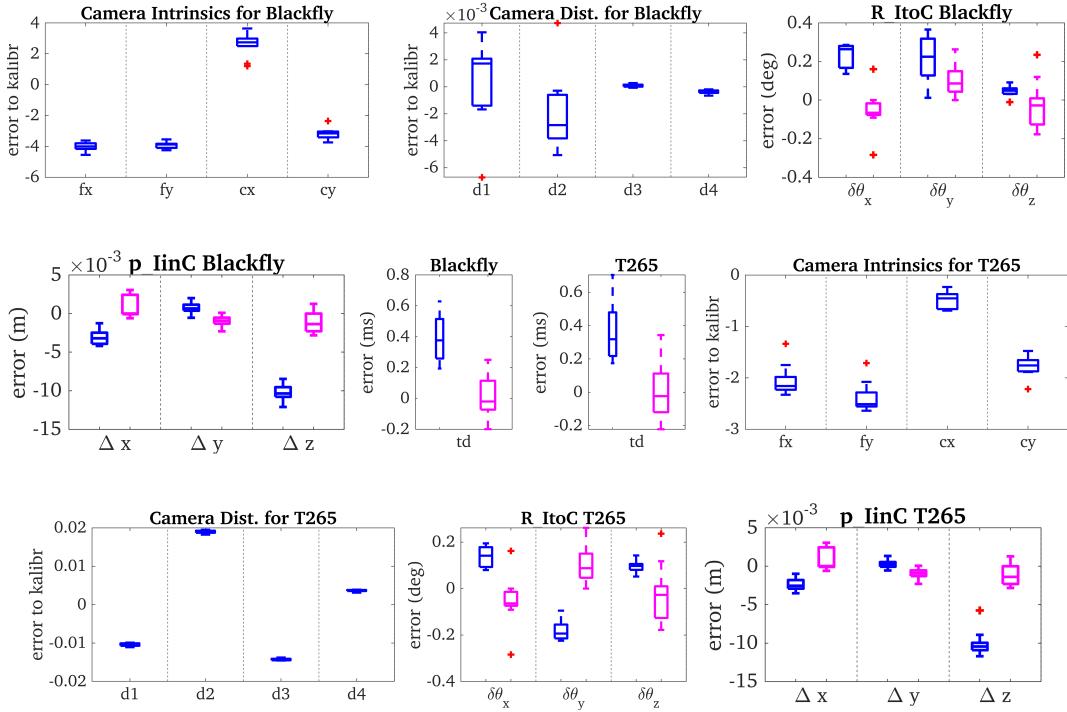


Figure 15. Comparison of the proposed method and Kalibr relative to a baseline Kalibr value. The boxplots show the final converged value of both methods, while for camera intrinsic only the proposed is reported since Kalibr fixes this during optimization. Kalibr (magenta, right in each group) was run with all cameras and IMUs available over 10 datasets, while the proposed system (blue, left) was run with either the Blackfly camera or left T265 fisheye and the MicroStrain GX-25 IMU resulting in 10 runs for each.

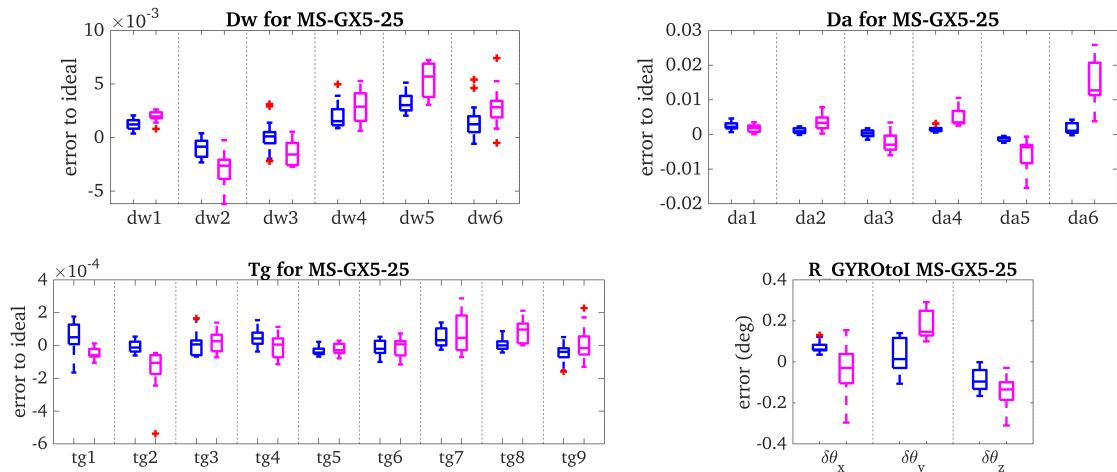


Figure 16. Comparison of the proposed method with *imu6* and Kalibr relative to the “ideal” sensor intrinsics. The boxplots show the final converged value of both methods. Kalibr (magenta, right in each group) was run with all cameras and IMUs available over 10 datasets, while the proposed system (blue, left) was run with either the Blackfly camera or left T265 fisheye resulting in 20 runs.

be evaluated by how close the converged calibrated values are to these “ideal” values.

11.2 IMU-Camera Spatiotemporal Extrinsics and Intrinsic

We first investigate the convergence of the IMU-camera extrinsics and temporal parameters along with the camera

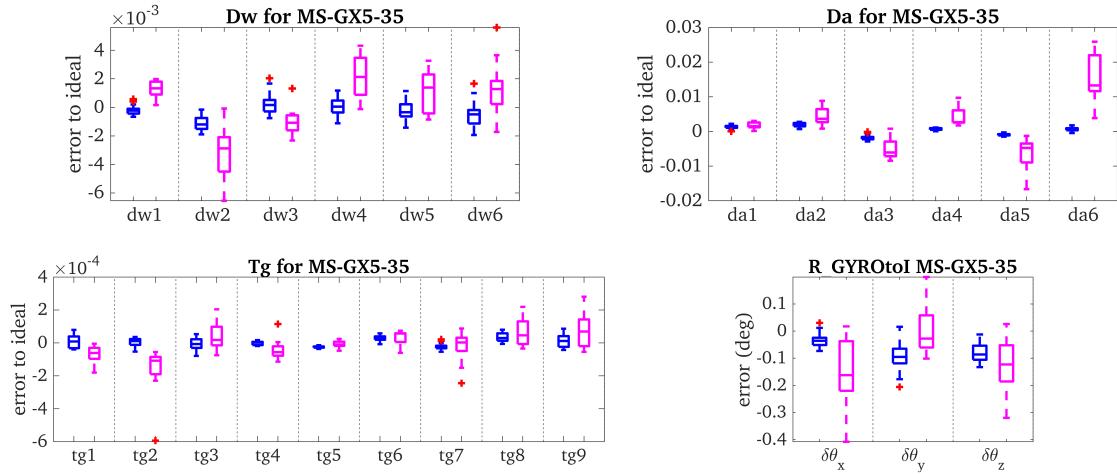


Figure 17. Comparison of the proposed method with *imu6* and Kalibr relative to the “ideal” sensor intrinsics. The boxplots show the final converged value of both methods. Kalibr (magenta, right in each group) was run with all cameras and IMUs available over 10 datasets, while the proposed system (blue, left) was run with either the Blackfly camera or left T265 fisheye resulting in 20 runs.

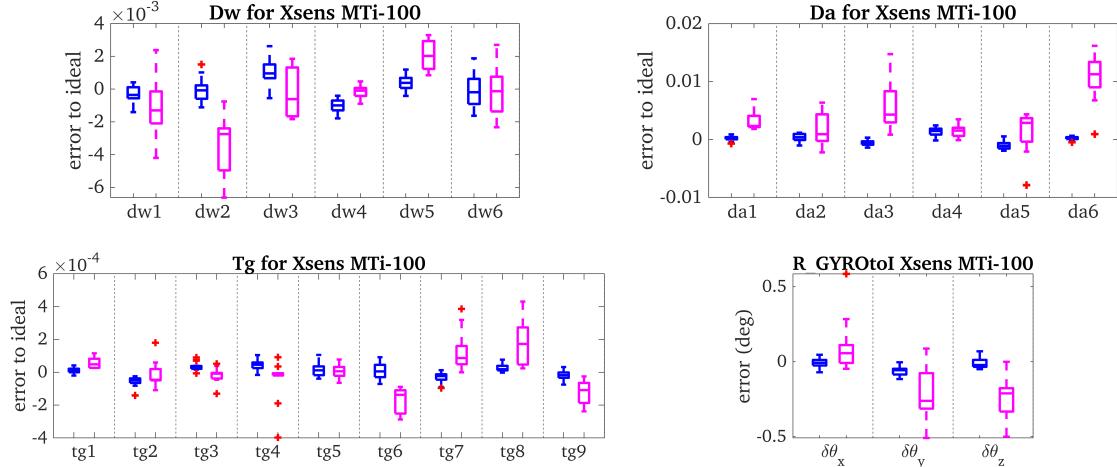


Figure 18. Comparison of the proposed method with *imu6* and Kalibr relative to the “ideal” sensor intrinsics. The boxplots show the final converged value of both methods. Kalibr (magenta, right in each group) was run with all cameras and IMUs available over 10 datasets, while the proposed system (blue, left) was run with either the Blackfly camera or left T265 fisheye resulting in 20 runs.

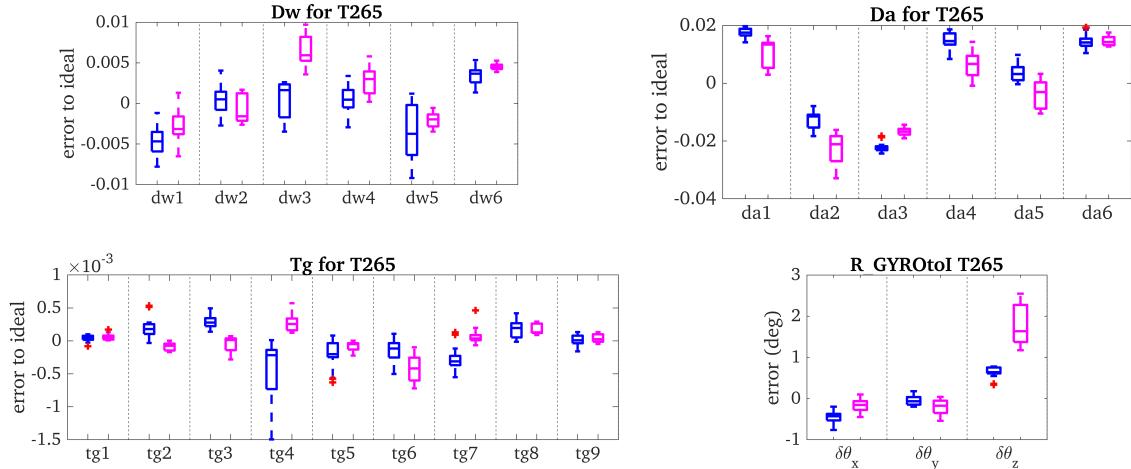


Figure 19. Comparison of the proposed method with *imu6* and Kalibr relative to the “identity” sensor intrinsics. The boxplots show the final converged value of both methods. Kalibr (magenta, right in each group) was run with all cameras and IMUs available over 10 datasets, while the proposed system (blue, left) was run with either the Blackfly camera or left T265 fisheye resulting in 20 runs.

intrinsics of the proposed system. The results shown in Figure 15 demonstrate that the proposed system is able to calibrate the spatial and temporal parameters with both high repeatability and accuracy relative to the offline Kalibr calibration baseline. Additionally shown is the convergence of camera intrinsics estimated by the proposed algorithm relative to the Kalibr static calibration results which are fixed during their IMU-camera calibration. Although the camera intrinsic estimates of blackfly and T265 camera have a few deviations compared to the reference values from Kalibr, the proposed system has very good convergence and high repeatability (the groundtruth is not known here).

11.3 IMU Intrinsic Parameters

The calibration results are summarized as boxplots shown in Figure 16 - 19 for the MicroStrain MS-GX5-25, MS-GX5-35, Xsens MTi-100 and T265 IMU, respectively. As shown, the calibration errors of the proposed system are quite close to the results of Kalibr, and the estimate differences of \mathbf{D}'_w , \mathbf{D}'_a and \mathbf{T}_g are around 1e-3, 1e-1 and 1e-4, respectively, for MS-GX5-25, MX-GX5-35 and Xsens MTi-100. Additionally, our proposed algorithm demonstrates better repeatability as the calibration errors have smaller variances and less outliers. In general, we can discuss the following results concerning the IMUs presented throughout the paper (see Figure 13 and 16 - 19):

- The MicroStrain MS-GX5-25, MS-GX5-35 and Xsens MTi-100 IMU are more close to “ideal” IMU than T265 IMU and BMI160 IMU. This is reasonable since both the MicroStrain and Xsens IMU are more expensive high-end IMUs with likely more sophisticated out-of-factory calibration than T265 IMU and BMI 160 IMU.
- For each IMU, the gravity sensitivity terms are, in general, one or two orders smaller than the other terms of the IMU intrinsic model. This suggests that the gravity sensitivity should not have significant effects on system performance. This is likely due to the handheld motion of the platform and levels of achievable acceleration magnitudes.
- The BMI160 IMU (Figure 13), has a much more significant gyroscope calibration, \mathbf{D}'_w , compared to its accelerometer calibration and other IMUs. Thus the BMI160 can see large accuracy gains from only calibrating \mathbf{D}'_w , while for other IMUs, the calibration of \mathbf{D}'_a should be more impactful.

Hence, these results validate the accuracy and consistency of the IMU intrinsic calibration of the proposed *online real-time* algorithm, which outputs comparable calibration results to Kalibr’s offline calibration procedures.

11.4 Timing Evaluation

We also evaluate the running time for the proposed system with and without online sensor calibration shown in Table 7. We use the 10 datasets recorded with the VI-Rig and only use the measurements from MicroStrain GX5-25 IMU and the left camera of T265 for evaluation. In order to get more realistic timing evaluation, no April tags are detected and only the natural features tracked from

images are used. We track 200 features from each image and keep at most 30 SLAM point features in the state vector with a sliding window of 20 clones. The averaged execution time for processing each coming image (including propagation and update) is recorded (shown in Table 7). The average execution time of the proposed system with online calibration is 0.0224s, which shows negligible increases than 0.0188s, which is the average running time without online calibration.

12 Real-World Degenerate Motion Demonstration and Analysis

Next, we evaluate the proposed system on a collection of real-world datasets which exhibit varying degrees of degenerate motions. Specifically, we evaluate on the EuRoc MAV dataset (Burri et al. 2016) which has under-actuated MAV with weakly excited acceleration and approximately 1-axis rotation. We also evaluate on the KAIST complex urban dataset (Jeong et al. 2019) which is a planar autonomous vehicle dataset with relatively constant velocity throughout. We note that we do not estimate the gravity sensitivity since it has been demonstrated in the preceding sections that it is not significant for VINS performance, and test only the *imu1* - *imu4* models.

12.1 EuRoC MAV: Under-Actuated Motion

The EuRoC MAV dataset (Burri et al. 2016) contains a series of trajectories from a MAV and provides 20 Hz grayscale stereo images, 200 Hz inertial readings, and an external groundtruth pose from a motion capture system. The proposed estimator is run with just the left camera on each of the Vicon room datasets and report the results in Table 8. Here we can see that not performing IMU intrinsic calibration, *imu0* model, outperforms the methods which additionally estimate the IMU intrinsics. This makes sense since the IMU intrinsics suffer from a large number of degenerate motions caused with constant local angular velocity and linear acceleration which can be expected for the MAV platform. Additionally, we believe that this is specifically caused by the MAV being unable to fully excite its 6DoF motion for a given small time interval and thus undergoes (nearly) degenerate motions locally throughout the whole trajectory, hurting the sliding-window filter. We can see that for more dynamic datasets, such as V2_03_difficult, there are still some partial improvements in accuracy possibly due to the more dynamic motion exhibited.

In order to verify our reasoning that the accuracy loss is caused by degenerate motions, we use the groundtruth trajectories of *tum_room1* with full 6DoF motion and *EuRoc V1_02* to simulate synthetic inertial and feature bearing measurements (see Section 9 on how we perform simulation) and evaluate our system with these simulated data. Figure 20 and 21 shows four different runs with estimation errors and 3σ bounds for \mathbf{D}_a and ${}^I_a\mathbf{R}$. It is clear that the motion of sensor on the *EuRoc V1_02* trajectory (right) is mildly excited and the acceleration readings are varying very slowly within the local window, causing poor convergence of the \mathbf{D}_a and ${}^I_a\mathbf{R}$ with relatively flat 3σ bounds as compared to the *tum_room1* (left). This verifies that the online IMU intrinsic calibration will benefit VINS with fully-excited motion (e.g.,

Table 8. Absolute Trajectory Error (ATE) on EuRoC MAV Vicon room sequences (with units degrees/meters).

IMU Model	V1.01_easy	V1.02_medium	V1.03_difficult	V2.01_easy	V2.02_medium	V2.03_difficult	Average
imu0	0.657 / 0.043	1.805 / 0.060	2.437 / 0.069	0.869 / 0.109	1.373 / 0.080	1.277 / 0.180	1.403 / 0.090
imu1	0.601 / 0.055	1.924 / 0.065	2.334 / 0.073	1.201 / 0.115	1.342 / 0.086	1.710 / 0.168	1.519 / 0.094
imu2	0.552 / 0.054	1.990 / 0.062	2.197 / 0.083	0.960 / 0.107	1.453 / 0.085	1.666 / 0.216	1.470 / 0.101
imu3	0.606 / 0.055	1.905 / 0.065	2.359 / 0.073	1.180 / 0.114	1.335 / 0.088	1.640 / 0.167	1.504 / 0.094
imu4	0.569 / 0.056	1.969 / 0.069	2.165 / 0.076	0.846 / 0.127	1.636 / 0.094	1.577 / 0.195	1.461 / 0.103

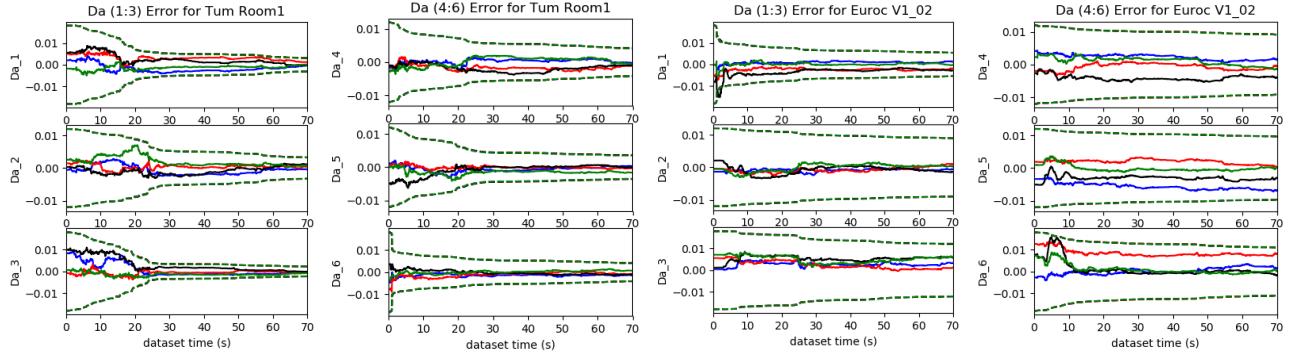


Figure 20. Simulation results for \mathbf{D}_a of the proposed system evaluated with groundtruth *tum_room1* (left) and *EuRoc V1_02* (right) trajectories using *imu2* and *radtan*. 3 sigma bounds (dotted lines) and estimation errors (solid lines) for four different runs (different colors) with different realization of the measurement noise and initial perturbations are drawn. The *tum_room1* estimation errors and 3σ bounds converge nicely, while due to lack of motion excitation, the convergence of \mathbf{D}_a , especially d_{a4} , d_{a5} and d_{a6} , for the *EuRoc V1_02* is poor.

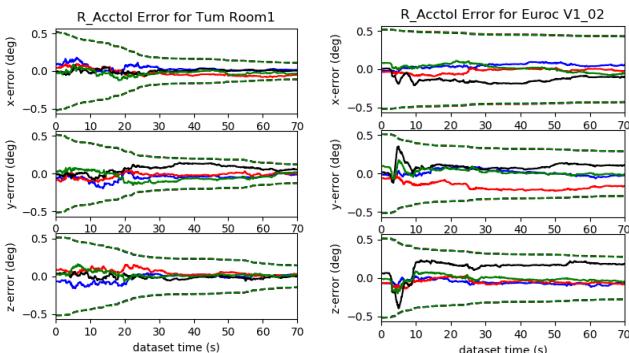


Figure 21. Simulation results for ${}^I\mathbf{R}$ of the proposed system evaluated with groundtruth *tum_room1* and *EuRoc V1_02* trajectories using *imu2* and *radtan*. With *tum_room1* trajectory, the estimation errors and 3σ bounds of ${}^I\mathbf{R}$ converge nicely, while the convergence is poor on the *EuRoc V1_02*.

the *tum_room1* trajectory) and might not be a good option for under-actuated motions such as the *EuRoc V1_02* trajectory.

12.2 KAIST Complex Urban: Planar Motion

Next we evaluate on the KAIST Complex Urban dataset (Jeong et al. 2019), which provides stereo grayscale images at 10 Hz, 200 Hz inertial readings, and a pseudo-groundtruth from an offline optimization method with the RTK GPS. As in the Table 2 and 3, this dataset contains planar motion and, in the case of *imu2*, 6 parameters are unobservable. We use the stereo camera pair along with the IMU to remove the scale ambiguity for monocular VINS caused by constant local acceleration (Wu et al. 2017b; Yang and Huang 2019).

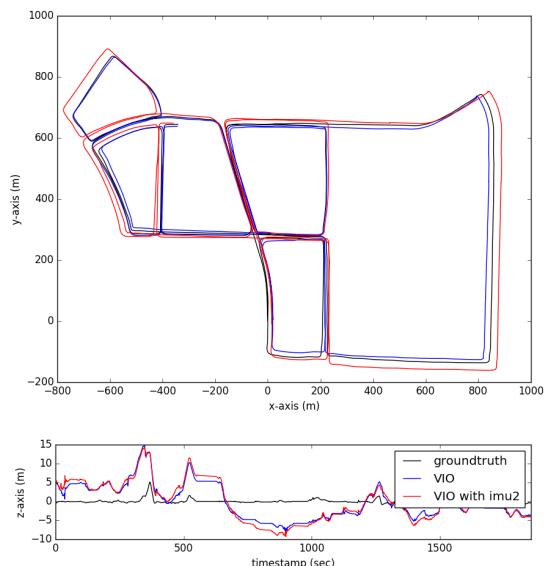


Figure 22. Trajectory plots for the KAIST Urban 39 dataset 10km in total length. Figure is best seen in color.

We plot Urban 39 trajectory estimates of the proposed system with and without IMU intrinsic calibration in Figure 22. It can be seen that the system with IMU intrinsic calibration has larger drift compared to the one without. When looking at the absolute trajectory error (ATE) in respect to the dataset's groundtruth, the error of the standard VIO, is 1.58 degrees with 13.03 meters (0.12%), while with *imu2* online IMU intrinsic calibration it is 1.41 degrees with 23.13 meters (0.22%). We propose that this larger error is due to the

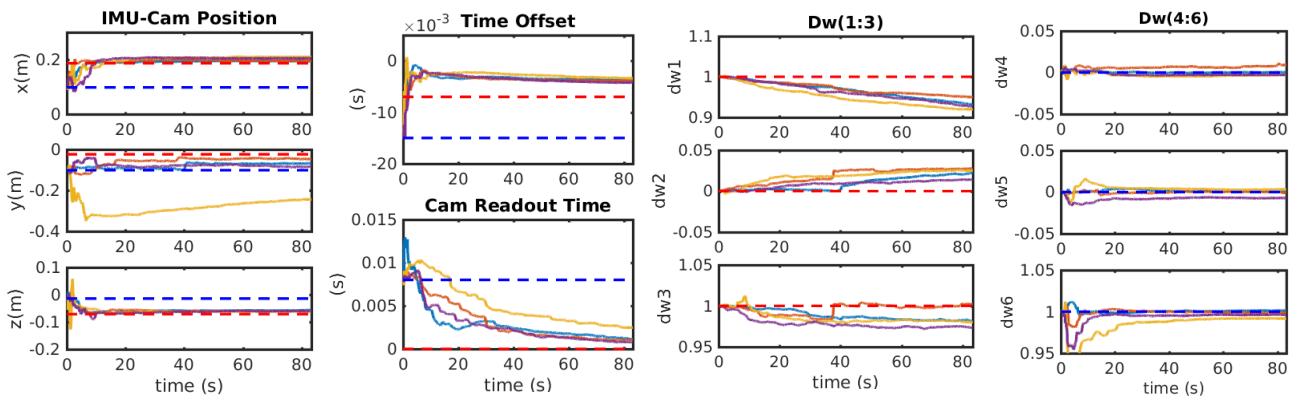


Figure 23. Calibration results (colored solid lines) for ${}^C\mathbf{p}_I$, t_d , t_r and \mathbf{D}_w of the proposed system evaluated with four VI-Rig planar motion datasets using *imu2* and *equi-dist*. Red and blue dotted lines denote the reference value from Kalibr and initial (perturbed) values, respectively. Colored solid line represents the estimated calibration parameters during online calibration for each dataset. All the temporal calibration, dw_4 , dw_5 and dw_6 can converge well to the reference values, while the y component of ${}^C\mathbf{p}_I$, dw_1 , dw_2 and dw_3 diverges.

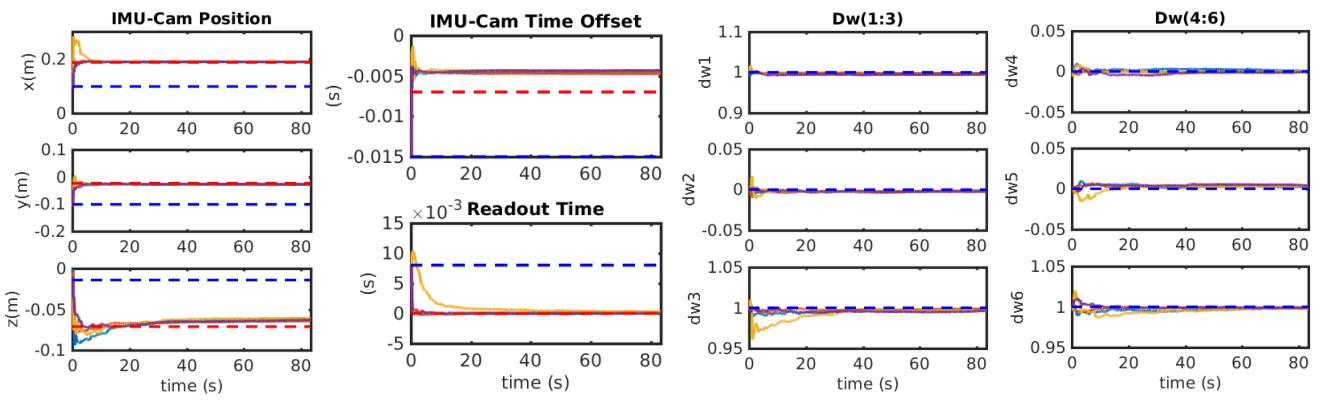


Figure 24. Calibration results (colored solid lines) for ${}^C\mathbf{p}_I$, t_d , t_r and \mathbf{D}_w of the proposed system evaluated with four 3D-motion datasets from Section 11 using *imu2* and *equidist*. Red and blue dotted lines denote the reference value from Kalibr and initial (perturbed) values, respectively. Colored solid line represents the estimated calibration parameters during online calibration for each dataset. All the temporal calibration, ${}^C\mathbf{p}_I$, and \mathbf{D}_w can converge well to the reference values.

introduced unobservable directions in online IMU intrinsic calibration when experiencing planar motion.

12.3 VI-Rig Planar Motion Datasets

Lastly, we evaluate on 4 datasets collected with VI-Rig (shown in Figure 14) under planar motion. In this evaluation, Microstrain GX5-25 and the left camera of T265 are used. When collecting data, we put the VI-Vig on a chair and only move the chairs in the ground plane to make sure VI-Vig is performing planar motion with global yaw as rotation axis, which is also the y-axis (pointing downward) of the camera. We calibrate all calibration parameters using *imu2* and *equidist* when running the system. Since T265 is a global shutter camera, the readout time is 0.00s.

The calibration results for the translation parameters ${}^C\mathbf{p}_I$, time offset t_d , readout time t_r and the \mathbf{D}_w are shown in Figure 23. All the temporal calibration can converge well to the reference values based on offline calibration results of Kalibr. Note that t_r converges to almost 0s as expected and t_d converges from 0.015s to 0.005s with reference values as 0.007s. The final estimation errors are around 0.002s, which is pretty small. While the x and z components of ${}^C\mathbf{p}_I$ can also converge well to the reference values with small standard deviations (smaller than 0.4cm), the y component

diverges with estimation errors more than 5cm and the standard deviation reach 3cm since it is along the rotation axis of the camera and hence, unobservable. Since the system has only yaw rotation for the IMU sensor, the dw_1 , dw_2 and dw_3 are also unobservable (see Table 2), and their calibration results diverge a lot compared to those of dw_4 , dw_5 and dw_6 . This result verifies our degenerate motion analysis for IMU-camera and IMU intrinsic calibration. As comparison, we also plot the online calibration results of the proposed system running on another four datasets from Section 11 with fully excited motions in Figure 24. We use the same scale to plot the results for both Figure 23 and 24. It is clear that all these calibration parameters (t_r , t_d , ${}^C\mathbf{p}_I$ and \mathbf{D}_w) can converge better in fully excited motions than planar motion.

13 Discussion: Online Self-Calibration?

As learnt from the preceding extensive Monte-Carlo simulations and real world experiments, we highly recommend online self-calibration for VINS especially in the following scenarios:

- Poor calibration priors are provided.
- Low-end IMUs or cameras are used.
- RS cameras are used.

- The system undergoes fully-excited motions.

Specifically, as shown in Table 5 (simulation) and Figures 11 and 12 (TUM RS VIO Datasets), if the system starts with imperfect calibration, the system without online self-calibration is highly likely to fail, clearly demonstrated by Figure 5. But online calibration can greatly improve the system robustness and accuracy. From Figure 12, we see that online calibration for the low-end IMU (Bosch BMI160) and RS readout time is necessary, which can improve the system performance greatly. Based on the results from the EuRoC MAV dataset (Table 8) and KAIST datasets (Figure 22), online calibration, especially IMU intrinsic calibration, can hurt the system performance when the system undergoes underactuated motions.

Based on our analysis on the degenerate motions for these calibration parameters, we can recommend what calibration parameters should be calibrated with different motion types:

- Fully excited 3D motion: all calibration parameters.
- Mildly excited 3D motion: IMU-camera spatial-temporal calibration, readout time, and camera intrinsics.
- Under-actuated motions: readout time and camera intrinsics.

As shown in our degenerate motion analysis, there are a large number of motion types that prohibit accurate calibration of the IMU intrinsics and IMU-camera spatial calibration, while the camera intrinsics and IMU-camera temporal calibration are more robust to different motions. More importantly, in the most commonly-seen motion cases of aerial and ground vehicles, there is usually at least one unobservable direction due to these robots traveling with either underactuated 3D or planar motion. The impact on performance was shown with the EuRoC MAV and KAIST Urban datasets (Table 8 and Figure 22) where the use of online IMU calibration may hurt the estimator.

Due to the high likelihood of experiencing degenerate motions for some periods of time, solely based on our analysis and results, we do *not* recommend performing online IMU intrinsic and IMU-camera spatial calibration during real-time operations for most underactuated motions (e.g., planar motion and one-axis rotation for most ground vehicles). The exception to this is the handheld cases (e.g., mobile AR/VR), which often exhibit full 6DoF motions and thus is recommended to perform online calibration to improve estimation accuracy, especially when low-end IMUs or RS cameras are used. For both of these applications, we do recommend using an offline batch optimization to obtain an accurate initial calibration guess for the filter and/or treat the calibration parameters (especially intrinsics) as “true” if one knows they are going to experience degenerate motions. For online IMU intrinsic calibration, it is not necessary to calibrate the full IMU model and instead one may calibrate only the dominating parameters in the inertial models (e.g., D_w for BMI160 IMU or D_a for MicroStrain, Xsens and T265 IMUs).

14 Conclusions and Future Work

In this paper, we address the problem of online full-parameter self-calibration for visual-inertial navigation

to achieve accurate and robust performance. We first investigate different IMU intrinsic model variants which contain scale correction, axis misalignment and gravity sensitivity. These model variants can cover most used inertial models in practice. We also introduce the full visual measurement model which contains IMU-camera spatial-temporal parameters including rolling shutter readout time. After computing the state transition matrix and measurements Jacobian regarding the state containing full calibration parameters, we perform observability analysis based on the linearized VINS system and show that with full-parameter calibration it still has only 4 unobservable directions, which relate to global yaw and global translation. All the calibration parameters of VINS are observable given fully excited motions.

Based on the observability analysis, we, for the first time, have identified basic degenerate motion patterns for IMU/camera intrinsics, and any combination of these degenerate motions will still be unobservable directions. Extensive validation on simulated and real-world datasets are performed to verify both the observability and degenerate motion analysis. We also show that online self-calibration can improve the robustness and accuracy of VINS. As shown through our experiments, online IMU intrinsic calibration is risky due to its dependence on the motion profile to ensure observability. In the case of autonomous (ground) vehicles, most trajectories have degenerate motions, thus resulting in *not* recommending online calibration of IMU intrinsics for robots with underactuated motions. In the case of handheld motion, however we found that the estimation of calibration parameters improved performance as expected.

In the future, we will investigate a complete degenerate motion analysis for multi-visual-inertial system. In addition, robust algorithms to perform online calibration under degenerate motions will also be studied.

References

- Agudo A (2021) Total estimation from rgb video: On-line camera self-calibration, non-rigid shape and motion. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 8140–8147. DOI:10.1109/ICPR48806.2021.9412923.
- Barfoot TD (2017) *State estimation for robotics*. Cambridge University Press.
- Bristeau PJ, Callou F, Vissire D and Petit N (2011) The navigation and control technology inside the ar.drone micro uav. *IFAC Proceedings Volumes* 44(1): 1477–1484. DOI:<https://doi.org/10.3182/20110828-6-IT-1002.02327>. 18th IFAC World Congress.
- Burri M, Nikolic J, Gohl P, Schneider T, Rehder J, Omari S, Achtelik MW and Siegwart R (2016) The euroc micro aerial vehicle datasets. *The International Journal of Robotics Research* 35(10): 1157–1163.
- Campos C, Elvira R, Rodriguez JJG, M Montiel JM and D Tards J (2021) Orb-slam3: An accurate open-source library for visual, visualinertial, and multimap slam. *IEEE Transactions on Robotics* : 1–17 DOI:10.1109/TRO.2021.3075644.
- Chatfield AB (1997) *Fundamentals of High Accuracy Inertial Navigation*. American Institute of Aeronautics and Astronautics.
- Chirikjian G (2011) *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern*

- Applications*, volume 2. Springer Science & Business Media.
- Civera J, Bueno DR, Davison AJ and Montiel JMM (2009) Camera self-calibration for sequential bayesian structure from motion. In: *2009 IEEE International Conference on Robotics and Automation*. pp. 403–408. DOI:10.1109/ROBOT.2009.5152719.
- Eckenhoff K, Geneva P, Bloecker J and Huang G (2019) Multi-camera visual-inertial navigation with online intrinsic and extrinsic calibration. In: *Proc. International Conference on Robotics and Automation*. Montreal, Canada.
- Eckenhoff K, Geneva P and Huang G (2021) Mimc-vins: A versatile and resilient multi-imu multi-camera visual-inertial navigation system. *IEEE Transactions on Robotics* 37(5): 1360–1380. DOI:10.1109/TRO.2021.3049445.
- Faugeras OD, Luong QT and Maybank SJ (1992) Camera self-calibration: Theory and experiments. In: Sandini G (ed.) *Computer Vision — ECCV'92*. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-540-47069-4, pp. 321–334.
- Forster C, Carlone L, Dellaert F and Scaramuzza D (2017) On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics* 33(1): 1–21.
- Furgale P, Rehder J and Siegwart R (2013) Unified temporal and spatial calibration for multi-sensor systems. In: *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp. 1280–1286.
- Geneva P, Eckenhoff K, Lee W, Yang Y and Huang G (2020) Openvins: A research platform for visual-inertial estimation. In: *Proc. of the IEEE International Conference on Robotics and Automation*. Paris, France. URL https://github.com/rpng/open_vins.
- Guo C, Kottas D, DuToit R, Ahmed A, Li R and Roumeliotis S (2014) Efficient visual-inertial navigation using a rolling-shutter camera with inaccurate timestamps. In: *Proc. of the Robotics: Science and Systems Conference*. Berkeley, CA.
- Guo C and Roumeliotis S (2013) IMU-RGBD camera 3D pose estimation and extrinsic calibration: Observability analysis and consistency improvement. In: *Proc. of the IEEE International Conference on Robotics and Automation*. Karlsruhe, Germany.
- Hartley R and Zisserman A (2004) *Multiple View Geometry in Computer Vision*. Cambridge University Press.
- Heng L, Choi B, Cui Z, Geppert M, Hu S, Kuan B, Liu P, Nguyen R, Yeo YC, Geiger A et al. (2019) Project autovision: Localization and 3d scene perception for an autonomous vehicle with a multi-camera system. In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 4695–4702.
- Hesch J, Kottas D, Bowman S and Roumeliotis S (2013) Consistency analysis and improvement of vision-aided inertial navigation. *IEEE Transactions on Robotics* 30(1): 158–176.
- Hesch JA, Kottas DG, Bowman SL and Roumeliotis SI (2014) Consistency analysis and improvement of vision-aided inertial navigation. *IEEE Transactions on Robotics* 30(1): 158–176. DOI:10.1109/TRO.2013.2277549.
- Huai J, Zhuang Y, Yuan Q and Lin Y (2021) Continuous-time spatiotemporal calibration of a rolling shutter camera-imu system. *CoRR* abs/2108.07200.
- Huang G (2012) *Improving the Consistency of Nonlinear Estimators: Analysis, Algorithms, and Applications*. PhD Thesis, Department of Computer Science and Engineering, University of Minnesota. URL <https://conservancy.umn.edu/handle/11299/146717>.
- Huang G (2019) Visual-inertial navigation: A concise review. In: *Proc. International Conference on Robotics and Automation*. Montreal, Canada.
- Huang G, Mourikis AI and Roumeliotis SI (2010) Observability-based rules for designing consistent EKF SLAM estimators. *International Journal of Robotics Research* 29(5): 502–528. DOI:10.1177/0278364909353640.
- Jeong J, Cho Y, Shin YS, Roh H and Kim A (2019) Complex urban dataset with multi-level sensors from highly diverse urban environments. *The International Journal of Robotics Research* 38(6): 642–657.
- Jones ES and Soatto S (2011) Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *International Journal of Robotics Research* 30(4): 407–430.
- Jung JH, Heo S and Park CG (2020) Observability analysis of imu intrinsic parameters in stereo visualinertial odometry. *IEEE Transactions on Instrumentation and Measurement* 69(10): 7530–7541. DOI:10.1109/TIM.2020.2985174.
- Karpenko A, Jacobs D, Baek J and Levoy M (2011) Digital video stabilization and rolling shutter correction using gyroscopes. *Stanford University Computer Science Tech Report CSTR 2011-03* 1(2): 13.
- Kelly J and Sukhatme GS (2011) Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration. *International Journal of Robotics Research* 30(1): 56–79.
- Kukelova Z, Albl C, Sugimoto A, Schindler K and Pajdla T (2020) Minimal rolling shutter absolute pose with unknown focal length and radial distortion. In: Vedaldi A, Bischof H, Brox T and Frahm JM (eds.) *Computer Vision – ECCV 2020*. Cham: Springer International Publishing. ISBN 978-3-030-58558-7, pp. 698–714.
- Lee W, Eckenhoff K, Yang Y, Geneva P and Huang G (2020) Visual-inertial-wheel odometry with online calibration. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.
- Leutenegger S, Lynen S, Bosse M, Siegwart R and Furgale P (2015) Keyframe-based visual-inertial odometry using nonlinear optimization. *International Journal of Robotics Research* 34(3): 314–334.
- Li M and Mourikis A (2013a) High-precision, consistent EKF-based visual-inertial odometry. *International Journal of Robotics Research* 32(6): 690–711.
- Li M and Mourikis AI (2013b) Optimization-based estimator design for vision-aided inertial navigation. In: *Robotics: Science and Systems*. Berlin, Germany, pp. 241–248.
- Li M and Mourikis AI (2014) Online temporal calibration for Camera-IMU systems: Theory and algorithms. *International Journal of Robotics Research* 33(7): 947–964.
- Li M, Yu H, Zheng X and Mourikis AI (2014) High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation. In: *IEEE International Conference on Robotics and Automation (ICRA)*. pp. 409–416.
- Martinelli A (2011) State estimation based on the concept of continuous symmetry and observability analysis: The case of calibration. *IEEE Transactions on Robotics* 27(2): 239–255.
- Martinelli A (2012) Vision and IMU data fusion: Closed-form solutions for attitude, speed, absolute scale, and bias determination. *IEEE Transactions on Robotics* 28(1): 44–60.
- Martinelli A (2014) Observability properties and deterministic algorithms in visual-inertial structure from motion. *Found.*

- Trends Robot* 3(3): 139209. DOI:10.1561/2300000030.
- Maybeck PS (1979) *Stochastic Models, Estimation, and Control*, volume 1. London: Academic Press.
- Meingast M, Geyer C and Sastry SS (2005) Geometric models of rolling-shutter cameras. *The 6th Workshop on Omnidirectional Vision abs/cs/0503076*: 4.
- Mirzaei FM and Roumeliotis SI (2008) A Kalman filter-based algorithm for IMU-camera calibration: Observability analysis and performance evaluation. *IEEE Transactions on Robotics* 24(5): 1143–1156.
- Mourikis AI and Roumeliotis SI (2007) A multi-state constraint Kalman filter for vision-aided inertial navigation. In: *Proceedings of the IEEE International Conference on Robotics and Automation*. Rome, Italy, pp. 3565–3572.
- Nguyen TT and Lhuillier M (2017) Self-calibration of omnidirectional multi-cameras including synchronization and rolling shutter. *Computer Vision and Image Understanding* 162: 166–184. DOI:10.1016/j.cviu.2017.08.010.
- Nikolic J, Burri M, Gilitschenski I, Nieto J and Siegwart R (2016) Non-parametric extrinsic and intrinsic calibration of visual-inertial sensor systems. *IEEE Sensors Journal* 16(13): 5433–5443. DOI:10.1109/JSEN.2016.2556662.
- Olson E (2011) Apriltag: A robust and flexible visual fiducial system. In: *Proc. of the IEEE International Conference on Robotics and Automation*. Shanghai, China, pp. 3400–3407.
- OpenCV Developers Team (2021) Open source computer vision (OpenCV) library. Available: <http://opencv.org>.
- Oth L, Furgale P, Kneip L and Siegwart R (2013) Rolling shutter camera calibration. In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1360–1367. DOI:10.1109/CVPR.2013.179.
- Qian G and Chellappa R (2004) Bayesian self-calibration of a moving camera. *Computer Vision Image Understanding* 95(3): 287316. DOI:10.1016/j.cviu.2004.03.009.
- Qin T, Li P and Shen S (2018) VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics* 34(4): 1004–1020.
- Qin T and Shen S (2018) Online temporal calibration for monocular visual-inertial systems. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 3662–3669. DOI:10.1109/IROS.2018.8593603.
- Rehder J, Nikolic J, Schneider T, Hinzmann T and Siegwart R (2016) Extending kalibr: Calibrating the extrinsics of multiple imus and of individual axes. In: *IEEE International Conference on Robotics and Automation (ICRA)*. pp. 4304–4311. DOI:10.1109/ICRA.2016.7487628.
- Rehder J and Siegwart R (2017) Camera/imu calibration revisited. *IEEE Sensors Journal* 17(11): 3257–3268. DOI:10.1109/JSEN.2017.2674307.
- Schneider T, Li M, Cadena C, Nieto J and Siegwart R (2019) Observability-aware self-calibration of visual and inertial sensors for ego-motion estimation. *IEEE Sensors Journal* 19(10): 3846–3860.
- Schneider T, Li M, Cadena C, Nieto J and Siegwart R (2019) Observability-aware self-calibration of visual and inertial sensors for ego-motion estimation. *IEEE Sensors Journal* 19(10): 3846–3860. DOI:10.1109/JSEN.2019.2893809.
- Schubert D, Demmel N, von Stumberg L, Usenko V and Cremers D (2019) Rolling-shutter modelling for visual-inertial odometry. In: *International Conference on Intelligent Robots and Systems (IROS)*.
- Schubert D, Goll T, Demmel N, Usenko V, Stückler J and Cremers D (2018) The tum vi benchmark for evaluating visual-inertial odometry. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1680–1687.
- Shen S, Mulgaonkar Y, Michael N and Kumar V (2014) Multi-sensor fusion for robust autonomous flight in indoor and outdoor environments with a rotorcraft MAV. In: *Proc. of the IEEE International Conference on Robotics and Automation*. Hong Kong, China, pp. 4974–4981.
- Trawny N and Roumeliotis SI (2005) Indirect Kalman filter for 3D attitude estimation. Technical report, University of Minnesota, Dept. of Comp. Sci. & Eng. URL <http://mars.cs.umn.edu/tr/reports/Trawny05b.pdf>.
- Tsao SH and Jan SS (2019) Observability analysis and performance evaluation of ekf-based visual-inertial odometry with online intrinsic camera parameter calibration. *IEEE Sensors Journal* 19(7): 2695–2703. DOI:10.1109/JSEN.2018.2886764.
- Usenko VC, Demmel N, Schubert D, Stückler J and Cremers D (2020) Visual-inertial mapping with non-linear factor recovery. *IEEE Robotics and Automation Letters* 5: 422–429.
- Wu K, Zhang T, Su D, Huang S and Dissanayake G (2017a) An invariant-ekf vins algorithm for improving consistency. In: *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp. 1578–1585.
- Wu KJ, Ahmed AM, Georgiou GA and Roumeliotis SI (2015) A square root inverse filter for efficient vision-aided inertial navigation on mobile devices. In: *Robotics: Science and Systems Conference (RSS)*.
- Wu KJ, Guo CX, Georgiou G and Roumeliotis SI (2017b) VINS on wheels. In: *Proc. of the IEEE International Conference on Robotics and Automation*. pp. 5155–5162.
- Xiao Y, Ruan X, Chai J, Zhang X and Zhu X (2019) Online imu self-calibration for visual-inertial systems. *Sensors* 19(7). DOI:10.3390/s19071624.
- Yang Y, Babu BPW, Chen C, Huang G and Ren L (2020a) Analytic combined imu integration for visual-inertial navigation. In: *Proc. of the IEEE International Conference on Robotics and Automation*. Paris, France.
- Yang Y, Geneva P, Eckenhoff K and Huang G (2019) Degenerate motion analysis for aided INS with online spatial and temporal calibration. *IEEE Robotics and Automation Letters (RA-L)* 4(2): 2070–2077.
- Yang Y, Geneva P, Zuo X and Huang G (2020b) Online imu intrinsic calibration: Is it necessary? In: *Proc. of Robotics: Science and Systems (RSS)*. Corvallis, Or.
- Yang Y and Huang G (2019) Observability analysis of aided ins with heterogeneous features of points, lines and planes. *IEEE Transactions on Robotics* 35(6): 399–1418.
- Yang Y, Maley J and Huang G (2017) Null-space-based marginalization: Analysis and algorithm. In: *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*. Vancouver, Canada, pp. 6749–6755.
- Zachariah D and Jansson M (2010) Joint calibration of an inertial measurement unit and coordinate transformation parameters using a monocular camera. In: *2010 International Conference on Indoor Positioning and Indoor Navigation*. pp. 1–7. DOI:10.1109/IPIN.2010.5646840.

$$= \begin{bmatrix} f_u (u_n + 3k_1 u_n r^2 + 5k_2 u_n r^4 + 4p_1 u_n v_n + 2p_2(r^2 + 2u_n^2)) \\ f_v (v_n + 3k_1 v_n r^2 + 5k_2 v_n r^4 + 2p_1(r^2 + 2u_n^2) + 4p_2 u_n v_n) \end{bmatrix}$$

At the same time, with one-axis rotation assumption, we have:

$$\begin{aligned} \mathbf{M}_f {}^G \mathbf{k} &= \mathbf{H}_{\mathbf{p}_f} {}^G \hat{\mathbf{R}}_{\mathbf{G}}^{I_k} \hat{\mathbf{R}} \cdot {}^G \hat{\mathbf{R}}_{\mathbf{C}}^I \hat{\mathbf{R}} {}^C \mathbf{k} {}^C z_f \\ &= \mathbf{H}_{\mathbf{p}_f} \mathbf{e}_3 {}^C z_f \\ &= \frac{\partial \tilde{\mathbf{z}}_C}{\partial \tilde{\mathbf{z}}_n} \begin{bmatrix} 1 & 0 & -u_n \\ 0 & 1 & -v_n \end{bmatrix} \mathbf{e}_3 \\ &= -\frac{\partial \tilde{\mathbf{z}}_C}{\partial \tilde{\mathbf{z}}_n} \begin{bmatrix} u_n \\ v_n \end{bmatrix} \end{aligned}$$

Note that ${}^G \hat{\mathbf{R}} = {}_{I_1} \hat{\mathbf{R}} {}_G^{I_1} \hat{\mathbf{R}}$ and ${}_{I_1} \hat{\mathbf{R}} {}^I \mathbf{k} = {}^I \mathbf{k}$ (due to one-axis rotation). We can easily verify that:

$$\begin{aligned} \frac{\partial \tilde{\mathbf{z}}_C}{\partial \tilde{\mathbf{z}}_n} \begin{bmatrix} u_n \\ v_n \end{bmatrix} &= \\ \begin{bmatrix} f_u (u_n + 3k_1 u_n r^2 + 5k_2 u_n r^4 + 4p_1 u_n v_n + 2p_2(r^2 + 2u_n^2)) \\ f_v (v_n + 3k_1 v_n r^2 + 5k_2 v_n r^4 + 2p_1(r^2 + 2u_n^2) + 4p_2 u_n v_n) \end{bmatrix} & \end{aligned}$$

Therefore, the verification of \mathbf{N}_{Cin} can be written as:

$$\begin{aligned} &\Leftrightarrow \mathcal{O}_k \mathbf{N}_{Cin} = \mathbf{0} \\ &\Leftrightarrow \mathbf{M}_{Cin} \times [f_u \ f_v \ 0 \ 0 \ 2k_1 \ 4k_2 \ p_1 \ p_2]^\top + \mathbf{M}_f {}^G \mathbf{k} = \mathbf{0} \end{aligned}$$

20 Appendix F: Interpolation Jacobians

We perturb ${}_{I_{ci-1}}^G \mathbf{R}$ and ${}_{I_{ci}}^G \mathbf{R}$ as:

$${}_{I_{ci-1}}^G \mathbf{R} = {}_{I_{ci-1}}^G \hat{\mathbf{R}} \exp(\delta \boldsymbol{\theta}_{I_{ci-1}}) \quad (150)$$

$${}_{I_{ci}}^G \mathbf{R} = {}_{I_{ci}}^G \hat{\mathbf{R}} \exp(\delta \boldsymbol{\theta}_{I_{ci}}) \quad (151)$$

By representing $\boldsymbol{\theta}_{i-1,i} = \log({}^G_{I_{ci-1}} \mathbf{R} {}_{I_{ci}}^G \mathbf{R})$, we have the linearization for the interpolation as:

$$\begin{aligned} \delta \boldsymbol{\theta}_{I(t)} &\simeq \left(\exp(-\hat{\lambda} \hat{\boldsymbol{\theta}}_{i-1,i}) - \mathbf{J}_r(\hat{\lambda} \hat{\boldsymbol{\theta}}_{i-1,i}) \mathbf{J}_l^{-1}(\hat{\boldsymbol{\theta}}_{i-1,i}) \right) \delta \boldsymbol{\theta}_{I_{ci-1}} \\ &+ \mathbf{J}_r(\hat{\lambda} \hat{\boldsymbol{\theta}}_{i-1,i}) \mathbf{J}_r^{-1}(\hat{\boldsymbol{\theta}}_{i-1,i}) \delta \boldsymbol{\theta}_{I_{ci}} + \frac{m}{M} \boldsymbol{\omega}_{i-1,i} \tilde{t}_r \\ {}^G \tilde{\mathbf{p}}_{I(t)} &\simeq (1 - \hat{\lambda}) {}^G \tilde{\mathbf{p}}_{I_{ci-1}} + \hat{\lambda} {}^G \tilde{\mathbf{p}}_{I_{ci}} + \frac{m}{M} \mathbf{v}_{i-1,i} \tilde{t}_r \end{aligned}$$

where:

$$\hat{\boldsymbol{\omega}}_{i-1,i} = \frac{\hat{\boldsymbol{\theta}}_{i-1,i}}{t_{ci} - t_{ci-1}} \quad (152)$$

$$\hat{\mathbf{v}}_{i-1,i} = \frac{{}^G \tilde{\mathbf{p}}_{I_{ci}} - {}^G \tilde{\mathbf{p}}_{I_{ci-1}}}{t_{ci} - t_{ci-1}} \quad (153)$$

21 Appendix G: Degenerate Motion Simulation Results

The complete calibration plots for IMU/camera intrinsic and IMU-camera spatial calibration under one-axis rotation, constant a_x acceleration and planar motion are shown in Figure 25, 26 and 27. All the temporal calibration results for IMU-camera time offset and rolling shutter readout time on the four simulated trajectories are shown in Figure 9.

====

Acknowledgements

This work was partially supported by the University of Delaware (UD) College of Engineering, the NSF (IIS-1924897), and Google ARCore. Geneva was partially supported by the University Doctoral Fellowship. Thanks Dr. Kun Fu for 3D printing of the VI-Rig. Thanks Dr. Kevin Eckenhoff for the rolling shutter implementation from MIMC-VINS project (Eckenhoff et al. 2021). Thanks David Schubert and Nikolaus Demmel from TUM for the helpful discussion on rolling-shutter visual-inertial datasets (Schubert et al. 2019).

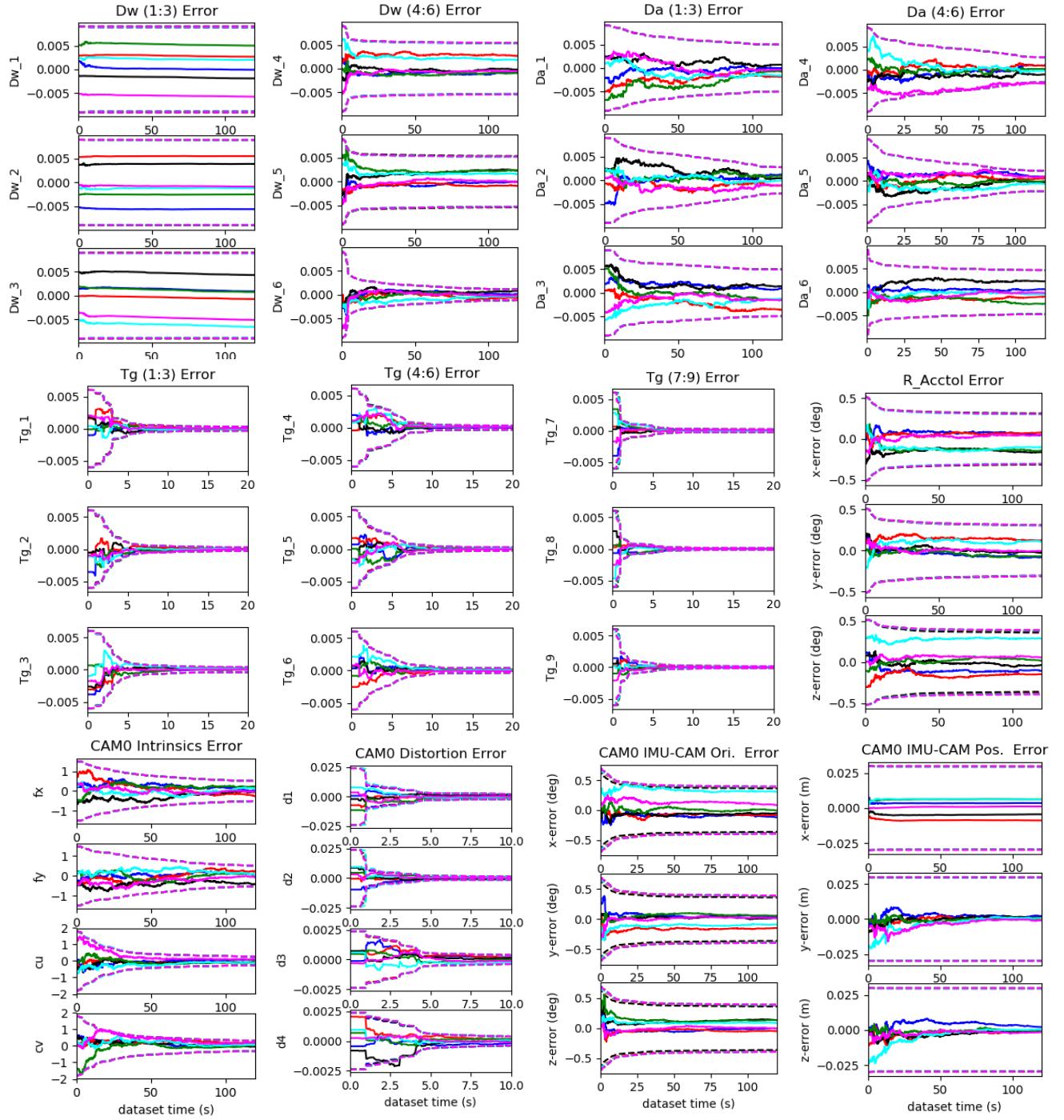


Figure 25. Calibration results for the proposed system evaluated on *tum_room* with one-axis rotation using *imu22* and *radian*. 3 sigma bounds (dotted lines) and estimation errors (solid lines) for six different runs (different colors) with different realization of the measurement noise and initial perturbations. Note that the estimation errors and 3σ bounds for d_{w1} , d_{w2} , d_{w3} and the IMU-CAM position calibration along the rotation axis can not converge.

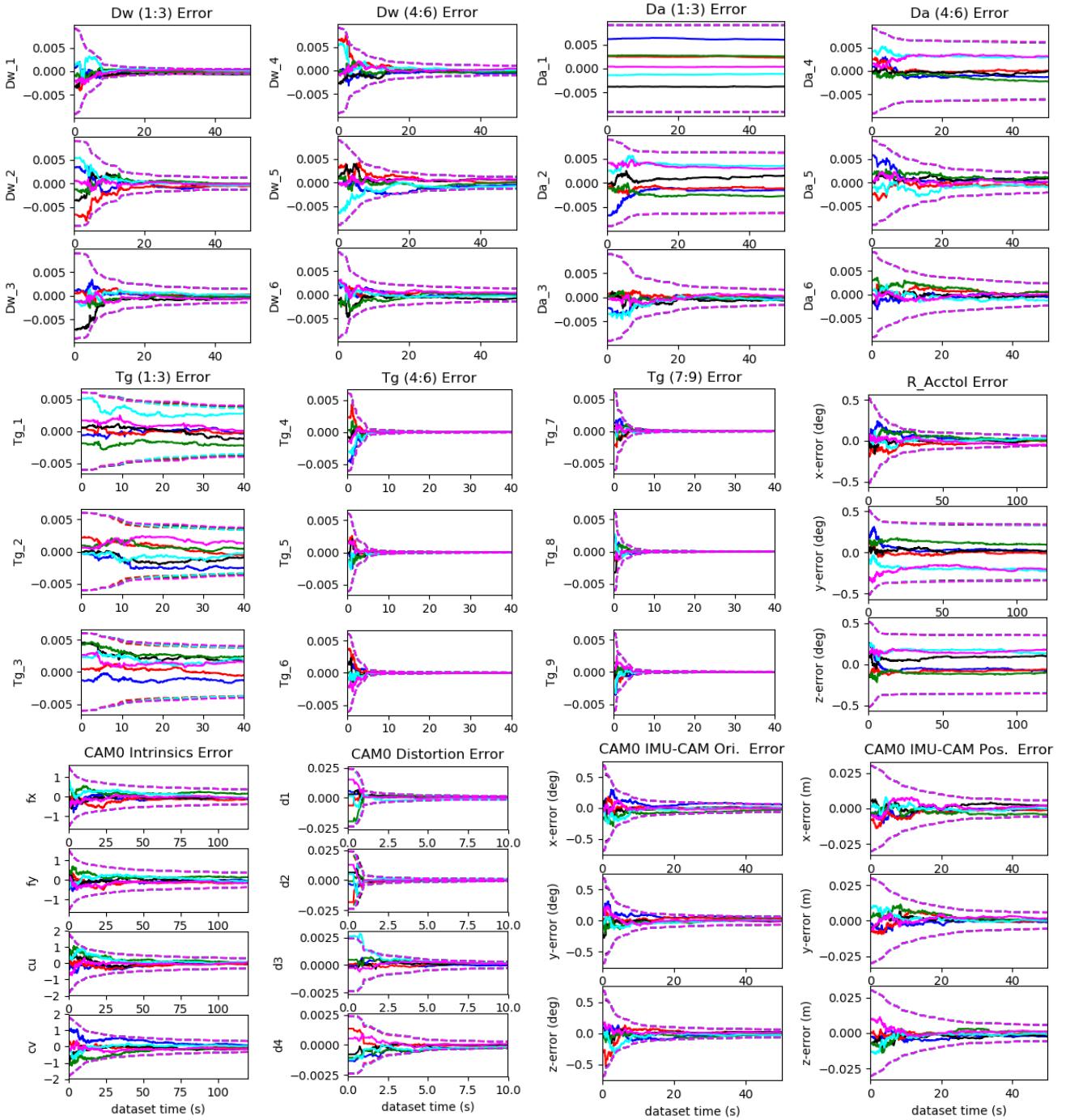


Figure 26. Calibration results for the proposed system evaluated the *sine.3d* with constant acceleration along x direction using *imu22* and *radtan*. 3 sigma bounds (dotted lines) and estimation errors (solid lines) for six different runs (different colors) with different realization of the measurement noise and initial perturbations. The estimation errors and 3σ bounds for d_{a1} , pitch and yaw of \mathbf{R}^I_a cannot converge. Note that t_{g1} , t_{g2} and t_{g3} are also unobservable.

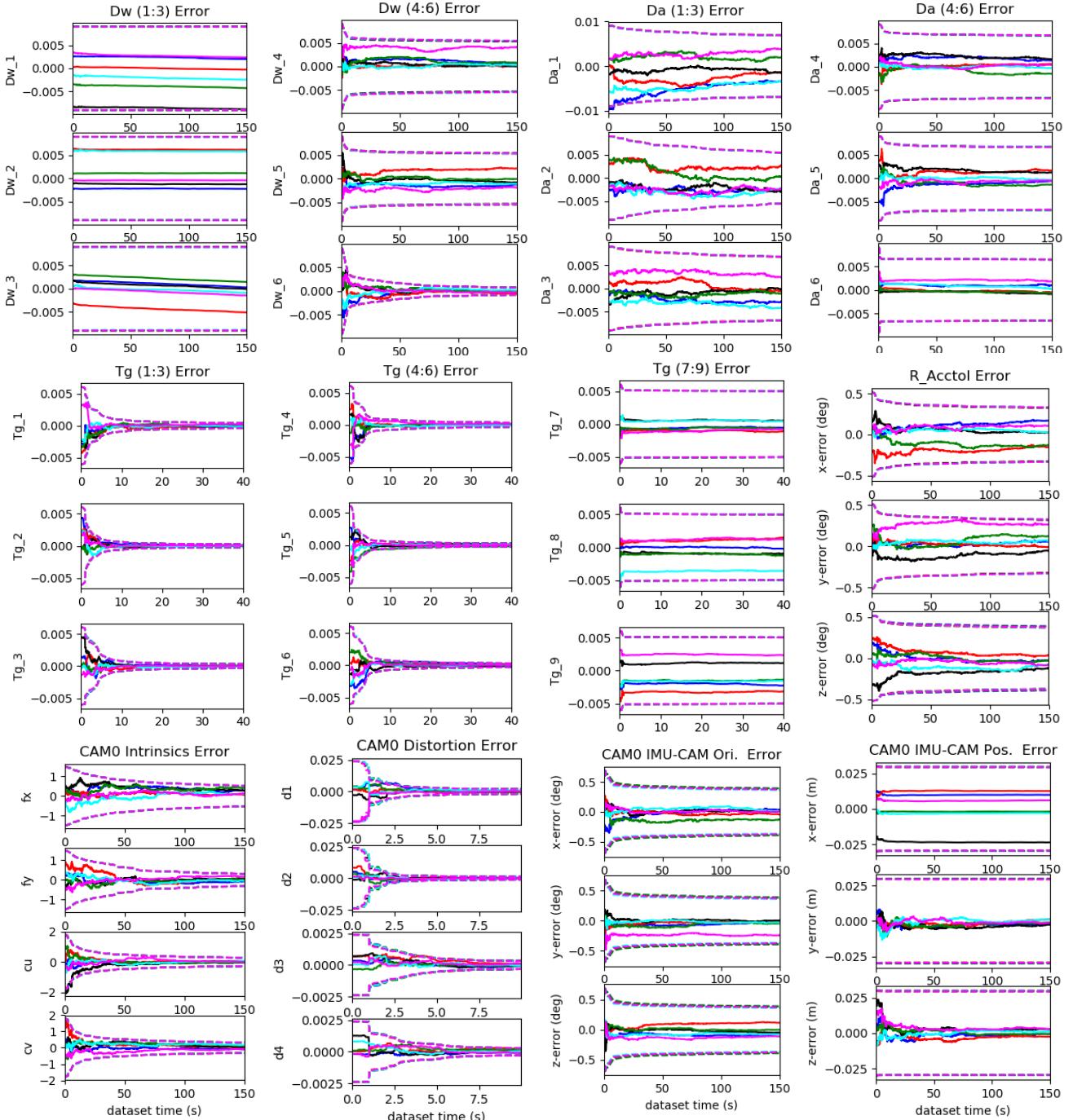


Figure 27. Calibration results of the proposed system evaluated on *udel_gore* with planar motion using *imu22* and *radian*. 3 sigma bounds (dotted lines) and estimation errors (solid lines) for six different runs (different colors) with different realization of the measurement noise and initial perturbations. With planar motion, the estimation errors and 3σ bounds of d_{w1} , d_{w2} , d_{w3} , t_{g7} , t_{g8} , t_{g9} and the IMU-CAM position cannot converge. Due to lack of motion excitation, the parameters of D_a and R_a converge much slower than the other motion cases.