

Multi-Visual-Inertial Sensor Fusion: Algorithm and Analysis

Abstract—In this paper, we propose a multi-visual-inertial sensor fusion algorithm to fuse multiple asynchronous inertial sensors (e.g., inertial measurement units (IMU) or gyroscopes (Gyro)) and visual cameras within a nonlinear least squares optimization. We are especially interested in the full calibration (including IMU intrinsics, camera intrinsics, spatial-temporal calibration for IMU-CAM and IMU-IMU) of such kind of system. In the proposed system, we offer a method to fuse auxiliary IMU requiring only auxiliary velocity and biases in the state, alongside a base IMU sensor which keeps the full navigation states. The proposed multi-IMU measurement function, containing all the necessary calibration parameters, leverages underlying rigid body constraints to eliminate the requirement to estimate the auxiliary IMU poses. In addition, we, for the first time, investigate the observability for multi-visual-inertial sensor system and show that the classical 4 unobservable directions remain, no matter how many inertial sensors are involved. Degenerate motions for IMU spatial-temporal parameters and inertial intrinsics are identified. We also advocate for the use of degenerate motion simulation to verify system consistency as system inconsistencies might be hidden when undergoing general 3D motions. Extensive simulations with fully excited and degenerate motions are performed to verify the proposed system and analysis. In simulation, we also show an example multi-IMU measurement model that is consistent in general motion but inconsistent in degenerate motion. Finally, we use our own sensor rigs built with off-the-shelf sensors to evaluate two commonly used inertial sensors and verify the repeatability of the proposed algorithm.

I. INTRODUCTION AND RELATED WORK

The ubiquity of camera and inertial measurement units (IMUs) has become prevalent in robotic systems in the recent decade due to their decrease in cost and complementary sensing nature. This has lead to a vast landscape of visual-inertial navigation system (VINS) algorithms which have focused on efficient and accurate pose estimation [18]. While many works have shown accurate estimation for the minimal sensing case of a single camera and IMU [30, 2, 9, 33, 14], it is well known [32] that the inclusion of additional sensors can provide improved accuracy due to additional information and robustness to single sensor failure cases. There have been many works which have explored multi-visual-inertial systems (MVIS) [11, 24, 39, 32, 5, 37, 31, 3, 22] all which have shown this improvement.

While the inclusion of multiple cameras has shown a large amount of focus, there has been limited exploration of multi-IMU fusion methods within the visual-inertial literature [27, 34, 21, 48, 7, 8]. Specifically, Ma

et al. [27] fused an auxiliary gyroscope in estimation without leveraging any of the acceleration information of the auxiliary IMU. Rehder et al. [34], extended the continuous-time Kalibr framework [11], to calibrate the extrinsic and intrinsic parameters of auxiliary inertial sensors by formulating the angular velocity and linear accelerations as functions of the trajectory spline derivatives. Kim et al. [21] reformulated preintegration by transforming the inertial readings from the auxiliary IMU frame into the base frame, but did not incorporate the estimation of the calibration and relied on the computation of angular accelerations to perform this transformation. Zhang et al. [48], proposed the conversion of multiple IMUs into a single “virtual” synthetic IMU measurement with lower noise. While offering computational savings, it relies on having perfect calibration and the synchronization of the sensors. Eckenhoff et al. [7, 8] proposed estimating each auxiliary IMU state (containing orientation, position, velocity, and biases), and enforced relative pose constraints between sensors at fixed rates while also showing robustness to inertial sensor failures.

To enable optimal MVIS, the spatial-temporal (including camera-to-IMU, IMU-to-IMU, and relative time offsets) and intrinsics (camera and IMU models) parameters are crucial. While there have been a few works that have shown to estimate only (some of) these parameters, in an offline or online, decoupled or coupled fashion [26, 11, 8], in this paper we present a complete MVIS system to jointly estimate all these parameters within a batch nonlinear least squares (NLS) optimization framework. We advocate for the joint estimation of all parameters to improve parameter estimation accuracy by modeling the correlations between calibration parameters, and for the first time, provide a detailed observability analysis of the proposed system to identify degenerate motions.

The observability of single-IMU camera systems has been studied extensively in quite a few works [29, 19, 20, 15, 40, 45, 46, 43, 47, 44], with a recent extension into multi-camera showing that the degenerate directions remain [10]. In this work we investigate the observability and degenerate motions of the multi-IMU calibration states within an MVIS estimation framework, thus filling the gap of the current visual-inertial literature.

The works closest to the proposed are the multi-IMU methods presented in [21, 48] along with the MVIS methods presented in [34, 8]. As compared to the relative preintegration method presented in [21] and

synthetic measurement in [48], we propose a method which does not require angular acceleration, estimates all inertial spacial-temporal and intrinsic parameters, and tries to reduce computational costs of auxiliary IMUs by leveraging inertial preintegration. While our work is similar to the calibration framework presented in [11, 34], we perform joint optimization of both camera and IMU parameters and provide a complete observability analysis for our MVIS. We also show that all calibration parameters are observable given general fully excited motion.

Eckenhoff et al. [5, 7, 8] leverage the similar camera measurement model as the proposed method, while differing in how we define our inertial measurement model in order to enable the incorporation within a batch-based optimization framework and to reduce complexity. In their work, a full inertial state is propagated forward in time for each auxiliary IMU and at the end of each integration period a rigid body constraint enforces that the relative pose of each integrated IMU must be equal to the extrinsic calibration between them. This method has a drawback due to the required increases in the state vector by an additional 15 degrees-of-freedom (DoF) per IMU. In the proposed method, we only require the increase of the state vector by 9 DoF and additionally perform intrinsic estimation of for all inertial sensors within the NLS estimation framework.

In particular, the main contributions of this work include:

- We propose a NLS-based multi-visual-inertial (accelerometer and/or gyroscope) sensor calibration system which jointly estimates *all* spatial-temporal and intrinsic parameters for an arbitrary number of visual and inertial sensors.
- We provide a new auxiliary inertial preintegration method, based on ACI² [42], which both allows for the incorporation of additional auxiliary IMUs and the estimation of their spacial-temporal and intrinsic parameters.
- We perform an observability analysis and degenerate motion identification for the MVIS calibration. We also conduct extensive simulations to show that we are able to recover all parameters (both visual and inertial) in the general motion case and compare to an existing state-of-the-art work in the real world experiments.
- With simulation, we verify the identified unobservable directions for the proposed MVIS under degenerate motion. Additionally, we leverage a simplified auxiliary IMU model example to demonstrate that degenerate motion trajectories are necessary to fully evaluate estimator consistency and advocate for degeneracy as an additional test of consistent estimator design.

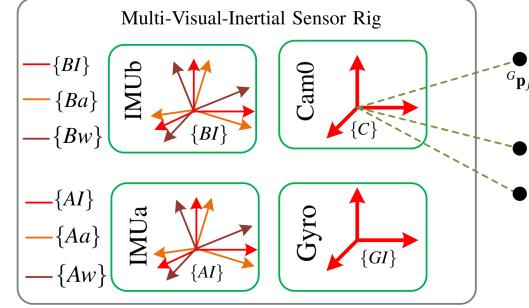


Fig. 1: The Multi-Visual-Inertial System (MVIS) sensor frames: base IMU (IMUb) sensor composed of accelerometer frame $\{Ba\}$ and gyroscope frame $\{Bw\}$, base “inertial” frame $\{BI\}$ is determined to coincide with gyroscope frame $\{Bw\}$, auxiliary IMU (IMUa) $\{AI\}$, auxiliary gyroscope (Gyro) $\{GI\}$, and camera (Cam0) $\{C\}$ frames. The system observes environmental landmarks ${}^G p_f$ through its cameras. Note that we only visualize unique sensor types but the system can contain arbitrary amounts.

II. MULTI-VISUAL-INERTIAL SYSTEM STATES

In this section, we show our choice of the full IMU model (containing scales, misalignment correction and gravity sensitivity). Then, the system state vector containing all the calibration parameters between IMUs and cameras is introduced. The overall NLS formulation is also defined.

A. IMU Intrinsic Model

Following [44], an IMU is assumed to consist of two separate frames: gyroscope frame $\{Bw\}$ and accelerometer frame $\{Ba\}$. The “inertial” frame $\{BI\}$ is chosen to coincide with either $\{Bw\}$ or $\{Ba\}$ (see Fig. 1). Different from the IMU models defined in [26, 44, 35], the raw angular velocity ${}^{Bw}\omega_m$ from the gyroscope and linear acceleration ${}^{Ba}\mathbf{a}_m$ from accelerometer can be written as:

$${}^{Bw}\omega_m = \mathbf{T}_w {}^{Bw}_{BI} \mathbf{R}^{BI} \omega + \mathbf{T}_g {}^{BI} \mathbf{a} + \mathbf{b}_g + \mathbf{n}_g \quad (1)$$

$${}^{Ba}\mathbf{a}_m = \mathbf{T}_a {}^{Ba}_{BI} \mathbf{R}^{BI} \mathbf{a} + \mathbf{b}_a + \mathbf{n}_a \quad (2)$$

where \mathbf{T}_w and \mathbf{T}_a denote the scale and axis misalignment for $\{Bw\}$ and $\{Ba\}$, respectively. \mathbf{T}_g represents the gravity sensitivity [26]. ${}^{Bw}_{BI} \mathbf{R}$ and ${}^{Ba}_{BI} \mathbf{R}$ denote the rotation from the gyroscope and acceleration frame to base “inertial” $\{BI\}$, respectively [44]. In this paper, we choose $\{BI\}$ coincides with $\{Bw\}$, thus ${}^{Bw}_{BI} \mathbf{R} = \mathbf{I}_3$.

Note that we use 6 parameters (indexed column-wise upper triangular matrix) to describe the \mathbf{T}_w , \mathbf{T}_a and \mathbf{T}_g . \mathbf{b}_g and \mathbf{b}_a are the gyroscope and accelerometer biases, which are both modeled as random walks, and \mathbf{n}_g and \mathbf{n}_a are the zero-mean Gaussian noises contaminating the measurements. We can write the desired (or corrected) angular velocity ${}^{BI}\omega$ and linear acceleration ${}^{BI}\mathbf{a}$ as:

$${}^{BI}\omega = {}^{BI}_{Bw} \mathbf{R} \mathbf{D}_w ({}^{Bw}\omega_m - \mathbf{b}_g - \mathbf{n}_g - \mathbf{T}_g {}^{BI} \mathbf{a}) \quad (3)$$

$${}^{BI}\mathbf{a} = {}^{BI}_{Ba} \mathbf{R} \mathbf{D}_a ({}^{Ba}\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a) \quad (4)$$

where $\mathbf{D}_w = \mathbf{T}_w^{-1}$ and $\mathbf{D}_a = \mathbf{T}_a^{-1}$. In practice, we treat \mathbf{D}_a , \mathbf{D}_w , \mathbf{T}_g and ${}^{BI}_{Ba} \mathbf{R}$ as a complete set of IMU intrinsic states.

B. State Vector

The state vector for MVIS contains the history of base IMU inertial states \mathcal{X}_{BI} , the history of auxiliary IMU states \mathcal{X}_{AI} , the history of auxiliary Gyro states \mathcal{X}_{GI} , all the inertial intrinsics \mathcal{X}_{In} , all the spatial-temporal calibration \mathcal{X}_{calib} and all the landmark features \mathcal{X}_F .

$$\mathbf{x} = [\mathcal{X}_{BI}^\top \ \mathcal{X}_{AI}^\top \ \mathcal{X}_{GI}^\top \ \mathcal{X}_{In}^\top \ \mathcal{X}_{calib}^\top \ \mathcal{X}_F^\top]^\top \quad (5)$$

$$\mathcal{X}_{BI} = [\mathbf{x}_{BI_1}^\top \ \cdots \ \mathbf{x}_{BI_m}^\top]^\top \quad (6)$$

$$\mathcal{X}_{AI} = [\mathbf{x}_{AI_1}^\top \ \cdots \ \mathbf{x}_{AI_m}^\top]^\top \quad (7)$$

$$\mathcal{X}_{GI} = [\mathbf{x}_{GI_1}^\top \ \cdots \ \mathbf{x}_{GI_m}^\top]^\top \quad (8)$$

$$\mathcal{X}_{In} = [\mathbf{x}_{Bn}^\top \ \mathbf{x}_{An}^\top \ \mathbf{x}_{Gn}^\top]^\top \quad (9)$$

$$\mathcal{X}_{calib} = [\mathbf{x}_{BA}^\top \ \mathbf{x}_{BG}^\top \ \mathbf{x}_{BC}^\top]^\top \quad (10)$$

$$\mathcal{X}_F = [{}^G\mathbf{p}_{f_1}^\top \ \cdots \ {}^G\mathbf{p}_{f_l}^\top]^\top \quad (11)$$

with:

$$\mathbf{x}_{BI} = [{}^G\bar{\mathbf{q}}^\top \ {}^G\mathbf{p}_{BI}^\top \ {}^G\mathbf{v}_{BI}^\top \ {}^{BI}\mathbf{b}_g^\top \ {}^{BI}\mathbf{b}_a^\top]^\top \quad (12)$$

$$\mathbf{x}_{AI} = [{}^G\mathbf{v}_{AI}^\top \ {}^{AI}\mathbf{b}_g^\top \ {}^{AI}\mathbf{b}_a^\top]^\top \quad (13)$$

$$\mathbf{x}_{GI} = {}^{GI}\mathbf{b}_g \quad (14)$$

$$\mathbf{x}_{Bn} = [\mathbf{x}_{Bw}^\top \ \mathbf{x}_{Ba}^\top \ \mathbf{x}_{Bg}^\top \ {}^{BI}\bar{\mathbf{q}}^\top]^\top \quad (15)$$

$$\mathbf{x}_{An} = [\mathbf{x}_{Aw}^\top \ \mathbf{x}_{Aa}^\top \ \mathbf{x}_{Ag}^\top \ {}^{AI}\bar{\mathbf{q}}^\top]^\top \quad (16)$$

$$\mathbf{x}_{Gn} = \mathbf{x}_{Gw} \quad (17)$$

$$\mathbf{x}_{BA} = [{}^{BI}\bar{\mathbf{q}}^\top \ {}^{BI}\mathbf{p}_{AI}^\top \ t_{AB}]^\top \quad (18)$$

$$\mathbf{x}_{BG} = [{}^{BI}\bar{\mathbf{q}}^\top \ t_{GB}]^\top \quad (19)$$

$$\mathbf{x}_{BC} = [{}^C\bar{\mathbf{q}}^\top \ {}^{BI}\mathbf{p}_C^\top \ t_{CB} \ \mathbf{x}_{Cn}^\top]^\top \quad (20)$$

where \mathbf{x}_{SI} denotes the inertial navigation state (S can be replaced by $B/A/G$ for base IMU, auxillary IMU, or Gyro, respectively). ${}^S\bar{\mathbf{q}}$ is in Hamilton quaternion from [42] and denotes the rotation ${}^S\mathbf{R}$ from the $\{SI\}$ sensor to global frame $\{G\}$. ${}^S\mathbf{p}_{SI}$ and ${}^S\mathbf{v}_{SI}$ represent the global position and velocity of $\{SI\}$ in $\{G\}$. ${}^S\mathbf{b}_g$ and ${}^S\mathbf{b}_a$ denote the gyro and accelerometer bias, respectively. \mathbf{x}_{Sn} denotes the inertial intrinsic states. As in [44], \mathbf{x}_{Sw} , \mathbf{x}_{Sa} and \mathbf{x}_{Sg} each contains 6 parameters from the upper triangular matrix \mathbf{D}_w , \mathbf{D}_a and \mathbf{T}_g , respectively. \mathbf{x}_{BS} represent the spatial-temporal calibration between the base and auxiliary inertial sensor with time offset defined as: $t_{SB} = t_{SI} - t_{BI}$, where t_{SI} and t_{BI} represent the auxiliary and base IMU measurement time, respectively. \mathbf{x}_{BC} denotes the complete calibration states between base IMU and camera with \mathbf{x}_{Cn} representing the camera intrinsics [14]. All the parameters that can be calibrated are summarized in Table I.

C. Nonlinear Least Squares Optimization

Given measurements from a sensor S , with additive white Gaussian noise, we have:

$$\mathbf{z}_S = \mathbf{h}_S(\mathbf{x}) + \mathbf{n}_S, \ \mathbf{n}_S \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_S) \quad (21)$$

TABLE I: All the calibration parameters estimated in the proposed MVIS.

Sensor	Extrinsics	Temporal	Intrinsics	Qty
Base IMU	${}^{BI}\mathbf{R}, {}^{BI}\mathbf{p}_{AI}$	—	\mathbf{x}_{Bn}	1
Aux IMU	${}^{AI}\mathbf{R}, {}^{BI}\mathbf{p}_{AI}$	t_{AB}	\mathbf{x}_{An}	≥ 1
Aux Gyro	${}^{BI}\mathbf{R}, {}^{GI}\mathbf{p}_{AI}$	t_{GB}	\mathbf{x}_{Gn}	≥ 1
Camera	${}^C\mathbf{R}, {}^{BI}\mathbf{p}_C$	t_{CB}	\mathbf{x}_{Cn}	≥ 1

Then, we can formulate the NLS problem as:

$$\min_{\mathbf{x}} \sum \|\mathbf{z}_S - \mathbf{h}_S(\mathbf{x})\|_{\mathbf{R}_S^{-1}}^2 \quad (22)$$

An initial guess $\hat{\mathbf{x}}^\ominus$ is needed to start the optimization. After computing the incremental state correction $\delta\mathbf{x}$, we can refine the state estimates by $\hat{\mathbf{x}}^\oplus = \hat{\mathbf{x}}^\ominus \boxplus \delta\mathbf{x}$, where \boxplus represents the state manifold update [1]. In summary, we have the following NLS which is equivalent to maximum likelihood estimation (MLE):

$$\min_{\mathbf{x}} \sum \mathbb{C}_{BI} + \sum \mathbb{C}_{AI} + \sum \mathbb{C}_{GI} + \sum \mathbb{C}_C \quad (23)$$

where \mathbb{C}_{BI} , \mathbb{C}_{AI} , \mathbb{C}_{GI} and \mathbb{C}_C denotes the cost for base IMU, auxiliary IMU, auxiliary Gyro, and camera, respectively. Solving this problem via iterative algorithms results in the optimal IMU states, visual point features, and all calibration parameters [see Eq. (5)]. We leverage GTSAM [4] to solve the NLS optimization problem.

D. Visual Cost

We define a visual cost with is a function of the base IMU states and related calibration parameters suitable for all asynchronous cameras. Assuming a visual feature that has been observed by a camera, we can write the visual-bearing measurements (i.e., pixel coordinates [14]) as the following:

$$\mathbf{z}_C = \mathbf{h}_d(\mathbf{h}_p(\mathbf{h}_t({}^G\mathbf{c}_{in}\mathbf{R}, {}^G\mathbf{p}_{Cin}, {}^G\mathbf{p}_f)), \mathbf{x}_{Cn}) + \mathbf{n}_C \quad (24)$$

$$\triangleq \mathbf{h}_C({}^G\mathbf{B}_{Iin}\mathbf{R}, {}^G\mathbf{p}_{BIin}, \mathbf{x}_{BC}, {}^G\mathbf{p}_f) + \mathbf{n}_C \quad (25)$$

where \mathbf{z}_C is the raw uv pixel coordinate, $\mathbf{n}_C \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_C)$ the raw pixel noise, ${}^G\mathbf{p}_f$ the estimated landmark position in $\{G\}$. The measurement functions \mathbf{h}_d , \mathbf{h}_p and \mathbf{h}_t (which follow the definitions in Open-VINS [13]) correspond to the intrinsic distortion, projection, and transformation functions and the corresponding measurement Jacobians can be computed through a simple chainrule. \mathbf{h}_C represents the combination of these functions \mathbf{h}_d , \mathbf{h}_p , and \mathbf{h}_t . We folowint the defintions

Since the state does not contain a IMU pose corresponding to the visual measurement time due to due to their time offsets and different frequencies, we leverage the pose interpolation in Appendix A to find the corresponding base IMU poses $\{{}^G\mathbf{B}_{Iin}\mathbf{R}, {}^G\mathbf{p}_{BIin}\}$ at the measurement collection time. Note that an additional chainrule needs to be taken in respect to the spacial and temporal sensor parameters this interpolated pose contain. Hence, the visual point cost can be formulated:

$$\mathbb{C}_C \triangleq \|\mathbf{z}_C - \mathbf{h}_C({}^G\mathbf{B}_{Iin}\mathbf{R}, {}^G\mathbf{p}_{BIin}, \mathbf{x}_{CI}, {}^G\mathbf{p}_f)\|_{\mathbf{Q}_C^{-1}}^2 \quad (26)$$

III. BASE INERTIAL COSTS

A. Base IMU Cost without Intrinsic

The IMU cost function is modeled after ACI² [42] with ${}^G\mathbf{g} = [0 \ 0 \ 9.81]^\top$. The IMU dynamic model can thus be defined as [38, 36]:

$$\begin{aligned} {}^{BI}\dot{\bar{q}} &= \frac{1}{2}\Omega({}^{BI}\boldsymbol{\omega})_{BI}^G\bar{q}, \quad {}^G\dot{\mathbf{p}}_{BI} = {}^G\mathbf{v}_{BI} \\ {}^G\dot{\mathbf{v}}_{BI} &= {}^G_{BI}\mathbf{R}^{BI}\mathbf{a} - {}^G\mathbf{g}, \quad {}^{BI}\dot{\mathbf{b}}_g = \mathbf{n}_{wg}, \quad {}^{BI}\dot{\mathbf{b}}_a = \mathbf{n}_{wa} \end{aligned} \quad (27)$$

where \mathbf{n}_{wg} and \mathbf{n}_{wa} are the white Gaussian noises driving the gyroscope and accelerometer biases. $\Omega(\boldsymbol{\omega}) := \begin{bmatrix} 0 & -\boldsymbol{\omega}^\top \\ \boldsymbol{\omega} & -[\boldsymbol{\omega}] \end{bmatrix}$ and $[\cdot]$ represents a skew symmetric matrix. Between two base camera times t_k and t_{k+1} we integrate the IMU measurements as:

$$\begin{aligned} {}^{BI}\Delta\bar{q} &\triangleq \mathbf{h}_q(\mathbf{x}_{BI_{k:k+1}}) = {}^G_{BI_k}\bar{q} \otimes {}^G_{BI_{k+1}}\bar{q} \\ {}^{BI}\Delta\mathbf{p} &\triangleq \mathbf{h}_p(\mathbf{x}_{BI_{k:k+1}}) \\ &= {}^G_{BI_k}\mathbf{R}^\top \left({}^G\mathbf{p}_{BI_{k+1}} - {}^G\mathbf{p}_{BI_k} - {}^G\mathbf{v}_{BI_k}\Delta t + \frac{1}{2}{}^G\mathbf{g}\Delta t^2 \right) \\ {}^{BI}\Delta\mathbf{v} &\triangleq \mathbf{h}_v(\mathbf{x}_{BI_{k:k+1}}) \\ &= {}^G_{BI_k}\mathbf{R}^\top \left({}^G\mathbf{v}_{BI_{k+1}} - {}^G\mathbf{v}_{BI_k} + {}^G\mathbf{g}\Delta t \right) \\ {}^{BI}\Delta\mathbf{b}_g &\triangleq {}^{BI}\mathbf{b}_{g_{k+1}} - {}^{BI}\mathbf{b}_{g_k} \\ {}^{BI}\Delta\mathbf{b}_a &\triangleq {}^{BI}\mathbf{b}_{a_{k+1}} - {}^{BI}\mathbf{b}_{a_k} \end{aligned}$$

where $\Delta t = t_{k+1} - t_k$. When doing the integration, the linearization for bias should be fixed. Hence, the IMU measurements should be corrected to account for the biases linearization change:

$${}^{BI}\Delta\bar{q} \simeq {}^{BI}\Delta\hat{q} \otimes \bar{q}(\mathbf{H}_{b_g}^q {}^{BI}\Delta\mathbf{b}_{g_k}) \quad (28)$$

$${}^{BI}\Delta\mathbf{p} \simeq {}^{BI}\Delta\hat{\mathbf{p}} + \mathbf{H}_b^p {}^{BI}\Delta\mathbf{b}_k \quad (29)$$

$${}^{BI}\Delta\mathbf{v} \simeq {}^{BI}\Delta\hat{\mathbf{v}} + \mathbf{H}_b^v {}^{BI}\Delta\mathbf{b}_k \quad (30)$$

where $\mathbf{b} = [\mathbf{b}_g^\top \ \mathbf{b}_a^\top]^\top$ and ${}^{BI}\Delta\mathbf{b}_k$ denotes the bias linearizaton changes [42]. Then, with these bias corrected pre-integrated measurements, we can construct:

$$\underbrace{\begin{bmatrix} \Delta\bar{q} \\ \Delta\mathbf{p} \\ \Delta\mathbf{v} \\ \Delta\mathbf{b} \end{bmatrix}}_{\mathbf{z}_{BI_{k+1}}} \simeq \underbrace{\begin{bmatrix} \mathbf{h}_q(\mathbf{x}_{BI_{k:k+1}}) \otimes \bar{q}^{-1}(\mathbf{H}_{b_g}^q {}^{BI}\tilde{\mathbf{b}}_{g_k}) \\ \mathbf{h}_p(\mathbf{x}_{BI_{k:k+1}}) - \mathbf{H}_b^p {}^{BI}\tilde{\mathbf{b}}_k \\ \mathbf{h}_v(\mathbf{x}_{BI_{k:k+1}}) - \mathbf{H}_b^v {}^{BI}\tilde{\mathbf{b}}_k \\ {}^{BI}\mathbf{b}_{k+1} - {}^{BI}\mathbf{b}_k \end{bmatrix}}_{\mathbf{h}_I(\mathbf{x}_{BI_{k:k+1}})} + \mathbf{n}_{BI_{k+1}} \quad (31)$$

where $\mathbf{n}_{BI_{k+1}} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{BI_{k+1}})$ represents the integrated IMU noise [42], ${}^{BI}\tilde{\mathbf{b}}_k$ is defined ${}^{BI}\mathbf{b}_k - {}^{BI}\hat{\mathbf{b}}_k$, $\mathbf{z}_{BI_{k+1}}$ is the integrated IMU measurements connecting state \mathbf{x}_{BI_k} and state $\mathbf{x}_{BI_{k+1}}$, and \mathbf{H}_y^z represents the Jacobians of state z to variable y . Finally, the corresponding base IMU cost function can be written as:

$$\mathbb{C}_{BI_{k+1}} \triangleq \|\mathbf{z}_{BI_{k+1}} - \mathbf{h}_I(\mathbf{x}_{BI_{k:k+1}})\|_{\mathbf{Q}_{BI_{k+1}}^{-1}}^2 \quad (32)$$

B. Base IMU Cost with Intrinsic

When incorporating IMU intrinsic, Xiao et al, [41] proposed treating them similar to the biases and fixed the IMU intrinsic linearization points during integration. However, in simulation we found that this operation can easily lead the intrinsic estimation to an undesirable local minimum.

Therefore, when incorporating the IMU intrinsic, we propose to relinearize bias and IMU intrinsic. For each iteration, the IMU measurements need to be re-integrated. Hence, the bias correction (29), (28) and (30) will not be needed and the IMU measurements following ACI² can be constructed as:

$$\underbrace{\begin{bmatrix} \Delta\bar{q} \\ \Delta\mathbf{p} \\ \Delta\mathbf{v} \\ \Delta\mathbf{b} \end{bmatrix}}_{\mathbf{z}_{BI_{k+1}}} \simeq \underbrace{\begin{bmatrix} \mathbf{h}_q(\mathbf{x}_{BI_{k:k+1}}) \otimes \bar{q}^{-1}(\mathbf{H}_{b_g}^q {}^{BI}\tilde{\mathbf{b}}_{g_k}) \otimes \bar{q}^{-1}(\mathbf{H}_{Bn}^q \tilde{\mathbf{x}}_{Bn}) \\ \mathbf{h}_p(\mathbf{x}_{BI_{k:k+1}}) - \mathbf{H}_b^p {}^{BI}\tilde{\mathbf{b}}_k - \mathbf{H}_{Bn}^p \tilde{\mathbf{x}}_{Bn} \\ \mathbf{h}_v(\mathbf{x}_{BI_{k:k+1}}) - \mathbf{H}_b^v {}^{BI}\tilde{\mathbf{b}}_k - \mathbf{H}_{Bn}^v \tilde{\mathbf{x}}_{Bn} \\ {}^{BI}\mathbf{b}_{k+1} - {}^{BI}\mathbf{b}_k \end{bmatrix}}_{\mathbf{h}_I(\mathbf{x}_{BI_{k:k+1}}, \mathbf{x}_{Bn})} + \mathbf{n}_{BI_{k+1}} \quad (33)$$

Then, the base IMU cost with intrinsic can be formulated similar to Eq. (32).

IV. AUXILIARY INERTIAL COSTS

A. Auxiliary IMU with Pose Constraints

With our auxiliary IMU measurements, which are independent from our base IMU and can be processed separately, we can also integrate the between the two base IMU timesteps:

$${}^{AI}\Delta\bar{q} = {}^G_{AI_k}\bar{q} \otimes {}^G_{AI_{k+1}}\bar{q}$$

$${}^{AI}\Delta\mathbf{p} = {}^G_{AI_k}\mathbf{R}^\top \left({}^G\mathbf{p}_{AI_{k+1}} - {}^G\mathbf{p}_{AI_k} - {}^G\mathbf{v}_{AI_k}\Delta t + \frac{1}{2}{}^G\mathbf{g}\Delta t^2 \right)$$

$${}^{AI}\Delta\mathbf{v} = {}^G_{AI_k}\mathbf{R}^\top \left({}^G\mathbf{v}_{AI_{k+1}} - {}^G\mathbf{v}_{AI_k} + {}^G\mathbf{g}\Delta t \right)$$

where $\Delta t = t_{k+1} - t_k$. The integrated measurements are all in the auxiliary IMU frame $\{AI\}$ and can be related to the base IMU state with a rigid body pose constraint:

$${}^G_{AI}\bar{q} = {}^G_{BI_{in}}\bar{q} \otimes {}^{AI}\bar{q} \quad (34)$$

$${}^G\mathbf{p}_{AI} = {}^G\mathbf{p}_{BI_{in}} + {}^G_{BI_{in}}\mathbf{R}^{BI}\mathbf{p}_{AI} \quad (35)$$

It is important to note, that we make this constraint a function of the interpolated pose $\{{}^G_{BI_{in}}\mathbf{R}, {}^G\mathbf{p}_{BI_{in}}\}$ in order to model the timeoffset error between the base and auxiliary IMUs. Hence, by combining the relationship of Eq. (34) and (35), we can construct the proposed auxiliary IMU measurement function:

$${}^{AI}\Delta\bar{q} \triangleq \mathbf{h}_q^{(AI)}(\mathbf{x}_{BI_{k:k+1}}, \mathbf{x}_{BA})$$

$${}^{AI}\Delta\mathbf{p} \triangleq \mathbf{h}_p^{(AI)}(\mathbf{x}_{BI_{k:k+1}}, \mathbf{x}_{BA})$$

$${}^{AI}\Delta\mathbf{v} \triangleq \mathbf{h}_v^{(AI)}(\mathbf{x}_{BI_{k:k+1}}, \mathbf{x}_{BA}, \mathbf{x}_{AI_{k:k+1}})$$

$${}^{AI}\Delta\mathbf{b} \triangleq {}^{AI}\mathbf{b}_{k+1} - {}^{AI}\mathbf{b}_k$$

where \mathbf{x}_{BA} is the spatial-temporal calibration. Then, following similar logic as Eq. (31) and (33), the auxiliary IMU cost functions \mathbb{C}_{AI} with or without IMU intrinsic \mathbf{x}_{An} can be defined.

B. Auxiliary Inertial Sensor Initialization

We need to initialize the initial state of \mathbf{x}_{AI} which contains the velocity and biases of the auxiliary IMU. The initial velocity of auxiliary IMU can be computed based on rigid body constraints:

$${}^G \mathbf{v}_{AI} = {}^G \mathbf{v}_{BI} + {}^G_{BI} \mathbf{R} [{}^{BI}_{AI} \mathbf{R} {}^{AI} \boldsymbol{\omega}] {}^{BI} \mathbf{p}_{AI} \quad (36)$$

For ${}^{AI} \hat{\mathbf{b}}_{g_k}$, we can first integrate the angular velocity readings with zero bias and solve the following:

$${}^{AI} \mathbf{H}_{b_g}^v {}^{AI} \hat{\mathbf{b}}_{g_k} = \log \left({}^{AI} \Delta \mathbf{R} {}^{G}_{AI_k} \mathbf{R} {}^{G}_{AI_{k+1}} \mathbf{R} \right) \quad (37)$$

where $\log(\cdot)$ denotes the *log* operation of SO(3) [1]. Then we use this newly computed ${}^{AI} \hat{\mathbf{b}}_{g_k}$ and zero accelerometer bias to compute ${}^{AI} \Delta \mathbf{v}$ and ${}^{AI} \Delta \mathbf{p}$. Finally, we can get ${}^{AI} \hat{\mathbf{b}}_{a_k}$ by solving:

$${}^{AI} \mathbf{H}_{b_a}^v {}^{AI} \hat{\mathbf{b}}_{a_k} = \mathbf{h}_v^{(AI)}(\hat{\mathbf{x}}_{BI_{k:k+1}}, \hat{\mathbf{x}}_{BA}) - {}^{AI} \Delta \mathbf{v} \quad (38)$$

$${}^{AI} \mathbf{H}_{b_a}^p {}^{AI} \hat{\mathbf{b}}_{a_k} = \mathbf{h}_p^{(AI)}(\hat{\mathbf{x}}_{BI_{k:k+1}}, \hat{\mathbf{x}}_{BA}) - {}^{AI} \Delta \mathbf{p} \quad (39)$$

C. Auxiliary Gyroscope Cost

As for a full auxiliary IMU, for an auxiliary gyroscope, integrate between two base IMU state times t_k and t_{k+1} :

$$\begin{aligned} {}^{GI} \Delta \bar{q} &= {}^{GI_k} \bar{q} \otimes {}^{G}_{GI_{k+1}} \bar{q} \\ {}^{GI} \Delta \mathbf{b}_g &\triangleq {}^{GI} \mathbf{b}_{g_{k+1}} - {}^{GI} \mathbf{b}_{g_k} \end{aligned}$$

We can relate this integration to the base IMU orientation:

$${}^{GI} \bar{q} = {}^{G}_{BI_{in}} \bar{q} \otimes {}^{BI} \bar{q} \quad (40)$$

where again we have defined our orientation ${}^{G}_{BI_{in}} \bar{q}$ as the interpolated base IMU orientation to account for the time offsets between the auxiliary Gyro and base IMU. Then, we formulate the measurement function as:

$$\begin{aligned} {}^{GI} \Delta \bar{q} &\triangleq \mathbf{h}_q^{(GI)}(\mathbf{x}_{BI_{k:k+1}}, \mathbf{x}_{BG}) \\ {}^{GI} \Delta \mathbf{b}_g &\triangleq {}^{GI} \mathbf{b}_{g_{k+1}} - {}^{GI} \mathbf{b}_{g_k} \end{aligned}$$

Similar to the base IMU case, we can then formulate the Gyro cost \mathbb{C}_{GI} (e.g., (31) and (33)) for without and with \mathbf{x}_{Gn} intrinsics.

D. Discussion

As mentioned before, when calibrating the IMU intrinsic parameters, reintegration of IMU measurements are needed to ensure accurate convergence. Hence, when performing the proposed MVIS with IMU intrinsic calibration, the optimization is computationally expensive, limiting its use for real-time applications. However, without inertial intrinsic calibration, the proposed base and auxiliary inertial constraints are conventional pre-integrated measurements and thus can be applied directly to current mature VIO algorithms (e.g., OKVIS [25], VINS-MONO [33], CPI [6]). Additionally, the spatial-temporal calibration between the base and auxiliary

IMUs is computationally efficient in nature and thus does not limit the real-time application of the estimator.

Note that we keep both the velocity and biases of the auxiliary IMUs in our state vector [see Eq. (5)]. While the model proposed in the previous sections is preferred, we also investigated a second model which does not require the estimation of the auxiliary IMU velocity, reducing the required additional states (see Appendix B). We later show that this second model has hidden inconsistencies that are not shown in the general 3D motion case, but can be detected under degenerate motion. We have a more detailed discussion in Section VII-3.

V. OBSERVABILITY ANALYSIS

Observability analysis plays an important role in state estimation for VINS [17, 28]. This analysis allows for determining the minimum measurements needed to uniquely determine the state and identify degenerate motions which can possibly hurt system performance by introducing additional unobservable directions for certain parameters [43, 40, 23, 44]. Since MVIS is becoming popular, an observability analysis for such a system with full calibration parameters is needed to better understand the underlying system property.

Without loss of generality, one base IMU, one auxiliary IMU, one auxiliary Gyro and one camera as unique sensors for the MVIS. We perform the observability analysis following [16, 47]. To simplify the analysis, we reorganize the state vector and assume the base IMU, auxiliary IMU and auxiliary Gyro are all kept as full states. All the states will be propagated together forward with time. The rigid body constraints and visual measurements will be used to update all these states.

With abuse of notation, we adjust the state vector for the observability analysis as:

$$\mathbf{x} = [\mathbf{x}_B^\top \ \mathbf{x}_A^\top \ \mathbf{x}_G^\top \ \mathbf{x}_{calib}^\top \ \mathbf{x}_F^\top]^\top \quad (41)$$

where:

$$\mathbf{x}_B = [\mathbf{x}_{BI}^\top \ \mathbf{x}_{Bn}^\top]^\top \quad (42)$$

$$\mathbf{x}_A = [\mathbf{x}_{AI}^\top \ \mathbf{x}_{An}^\top]^\top \quad (43)$$

$$\mathbf{x}_G = [\mathbf{x}_{GI}^\top \ \mathbf{x}_{Gn}^\top]^\top \quad (44)$$

$$\mathbf{x}_{calib} = [\mathbf{x}_{BA}^\top \ \mathbf{x}_{BG}^\top \ \mathbf{x}_{BC}^\top]^\top \quad (45)$$

$$\mathbf{x}_F = {}^G \mathbf{p}_f \quad (46)$$

Note that the auxiliary IMU and Gyro states are:

$$\mathbf{x}_{AI} = [{}^{G_{AI}} \bar{q}^\top \ {}^G \mathbf{p}_{AI}^\top \ {}^G \mathbf{v}_{AI}^\top \ {}^{AI} \mathbf{b}_g^\top \ {}^{AI} \mathbf{b}_a^\top]^\top \quad (47)$$

$$\mathbf{x}_{GI} = [{}^{G_{GI}} \bar{q}^\top \ {}^{GI} \mathbf{b}_g^\top]^\top \quad (48)$$

The observation constraints based on visual measurements and rigid body constraints can be listed as:

$$\mathbf{z} = [\mathbf{z}_C^\top \ \mathbf{z}_A^\top \ \mathbf{z}_G^\top]^\top \quad (49)$$

where \mathbf{z}_C denotes the visual cost, \mathbf{z}_A and \mathbf{z}_G represent the rigid body pose constraints between auxiliary and base inertial sensor. They are described in detail as:

$$\mathbf{z}_C = \begin{bmatrix} \frac{c_{p_x}}{c_{p_z}} & \frac{c_{p_y}}{c_{p_z}} \end{bmatrix}^\top, \quad {}^C\mathbf{p}_f = [c_{p_x} \ c_{p_y} \ c_{p_z}]^\top \quad (50)$$

$${}^C\mathbf{p}_f = {}^C_{BI}\mathbf{R}_G^{BI}\mathbf{R}({}^G\mathbf{p}_f - {}^G\mathbf{p}_{BI} - {}^G_{BI}\mathbf{R}^{BI}\mathbf{p}_C) \quad (51)$$

$$\mathbf{z}_A = \left[\log \left({}^G_{AI}\mathbf{R}^\top {}^G_{BI}\mathbf{R}^{BI}\mathbf{R} \right) \right] \quad (52)$$

$$\mathbf{z}_G = \log \left({}^G_{GI}\mathbf{R}^\top {}^G_{BI}\mathbf{R}^{BI}\mathbf{R} \right) \quad (53)$$

After we compute the state transition matrix and Jacobians (Appendix C and D), the block row of observability matrix \mathbf{M} can be written as:

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{CB} & \mathbf{0} & \mathbf{0} & \mathbf{M}_{CC} & \mathbf{M}_{CF} \\ \mathbf{M}_{AB} & \mathbf{M}_{AA} & \mathbf{0} & \mathbf{M}_{AC} & \mathbf{0} \\ \mathbf{M}_{GB} & \mathbf{0} & \mathbf{M}_{GG} & \mathbf{M}_{GC} & \mathbf{0} \end{bmatrix} \quad (54)$$

where \mathbf{M}_* can be found in Appendix E. By analyzing the observability matrix, we can have the following Lemma:

Lemma 1. *The proposed MVIS still keeps 4 unobservable directions as Eq. (55), which are related to global yaw rotation and global translation.*

$$\mathbf{N} = \begin{bmatrix} {}^{BI0}\mathbf{R}^G\mathbf{g} & \mathbf{0}_3 \\ [{}^G\mathbf{p}_{BI0}]^G\mathbf{g} & \mathbf{I}_3 \\ [{}^G\mathbf{v}_{BI0}]^G\mathbf{g} & \mathbf{0}_3 \\ \mathbf{0}_{6 \times 1} & \mathbf{0}_{6 \times 3} \\ \mathbf{0}_{21 \times 1} & \mathbf{0}_{21 \times 3} \\ {}^{AI0}\mathbf{R}^G\mathbf{g} & \mathbf{0}_3 \\ [{}^G\mathbf{p}_{AI0} - {}^G\mathbf{p}_{BI0}]^G\mathbf{g} & \mathbf{I}_3 \\ [{}^G\mathbf{v}_{AI0} - {}^G\mathbf{v}_{BI0}]^G\mathbf{g} & \mathbf{0}_3 \\ \mathbf{0}_{6 \times 1} & \mathbf{0}_{6 \times 3} \\ \mathbf{0}_{21 \times 1} & \mathbf{0}_{21 \times 3} \\ {}^{GI0}\mathbf{R}^G\mathbf{g} & \mathbf{0}_3 \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_3 \\ \mathbf{0}_{6 \times 1} & \mathbf{0}_{6 \times 3} \\ \mathbf{0}_{26 \times 1} & \mathbf{0}_{26 \times 3} \\ [{}^G\mathbf{p}_f]^G\mathbf{g} & \mathbf{I}_3 \end{bmatrix} \quad (55)$$

These four unobservable directions are similar to the 4 classical unobservable directions for a mono VINS system. From this lemma, we can conclude that the system won't lose unobservable directions by simply adding more inertial sensors (IMU or Gyros). It should also be pointed out that the velocity of IMU state will become unobservable if no visual measurements to static landmarks are available. Hence, naively adding auxiliary IMUs without cameras will not improve the system localization accuracy due to lack of global constraints to the base IMU velocity.

It can be observed that the calibration parameters, including \mathbf{x}_{Bn} , \mathbf{x}_{An} , \mathbf{x}_{Gn} and \mathbf{x}_{calib} , are highly related to the sensor motion. Under fully excited random motion, these parameters are observable, which can be seen from our simulation results (Section VII-1).

TABLE II: Degenerate motions for auxiliary IMU and Gyro extrinsics calibration.

Motion Types	Unobservable Parameters
No Rotation	${}^{BI}\mathbf{p}_{AI}$, ${}^{BI}_{GI}\mathbf{R}$ and t_{GB}
One-axis Rotation	${}^{BI}\mathbf{p}_{AI}$ and ${}^{BI}_{GI}\mathbf{R}$ along rotation axis

VI. DEGENERATE MOTION ANALYSIS

The degenerate motions for spatial-temporal parameters between IMU and camera has been studied in [43]. Therefore, in this paper, we study the degenerate motion for spatial-temporal and intrinsic parameters of auxiliary inertial sensors.

A. Spatial-Temporal of Auxiliary Inertial Sensors

We identified two degenerate motions for the spatial-temporal calibration between auxiliary inertial sensors and base IMU (see Tab. II).

1) *No Rotation*: If the system demonstrates 3D motion without rotation, the translation between auxiliary IMU and base IMU, the rotation and time offset between auxiliary Gyro and base IMU will all be unobservable. The unobservable directions are written as:

$$\mathbf{N}_{NoRot} = \begin{bmatrix} \mathbf{0}_{36 \times 3} & \mathbf{0}_{36 \times 3} & \mathbf{0}_{36 \times 1} \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 1} \\ \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{6 \times 3} & \mathbf{0}_{6 \times 3} & \mathbf{0}_{6 \times 1} \\ \mathbf{0}_{21 \times 3} & \mathbf{0}_{21 \times 3} & \mathbf{0}_{21 \times 1} \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{6 \times 3} & \mathbf{0}_{6 \times 3} & \mathbf{0}_{6 \times 1} \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & \mathbf{0} \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{1 \times 3} & \mathbf{0}_{1 \times 3} & \mathbf{1} \\ \mathbf{0}_{15 \times 3} & \mathbf{0}_{15 \times 3} & \mathbf{0}_{15 \times 1} \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 1} \end{bmatrix} \quad (56)$$

2) *One-Axis Rotation*: If the system undergoes 3D motion but with only one-axis rotation (similar to micro aerial vehicle (MAV)), the translation between auxiliary IMU and base IMU, the rotation between auxiliary Gyro and base IMU will be unobservable along the rotation

TABLE III: Summary of basic degenerate motions for IMUb, IMUa and Gyro intrinsics calibration. Any combinations of these unit motion primitives are degenerate.

Motion Types	Nullspace Dim.	Unobservable Parameters
constant $B^w\omega_1$	1	d_{Bw11}
constant $B^w\omega_2$	2	d_{Bw12}, d_{Bw22}
constant $B^w\omega_3$	3	$d_{Bw13}, d_{Bw23}, d_{Bw33}$
constant $B^a a_1$	3	d_{Ba11} , pitch and yaw of $B^I R^a$
constant $B^a a_2$	3	d_{Ba12}, d_{Ba22} , roll of $B^I R^a$
constant $B^a a_3$	3	$d_{Ba13}, d_{Ba23}, d_{Ba33}$
constant $A^w\omega_1$	1	d_{Aw11}
constant $A^w\omega_2$	2	d_{Aw12}, d_{Aw22}
constant $A^w\omega_3$	3	$d_{Aw13}, d_{Aw23}, d_{Aw33}$
constant $A^a a_1$	3	d_{Aa11} , pitch and yaw of $A^I R^a$
constant $A^a a_2$	3	d_{Aa12}, d_{Aa22} , roll of $A^I R^a$
constant $A^a a_3$	3	$d_{Aa13}, d_{Aa23}, d_{Aa33}$
constant $G^w\omega_1$	1	d_{Gw11}
constant $G^w\omega_2$	2	d_{Gw12}, d_{Gw22}
constant $G^w\omega_3$	3	$d_{Gw13}, d_{Gw23}, d_{Gw33}$

axis k. The unobservable directions can be described as:

$$\mathbf{N}_{1axis} = \begin{bmatrix} \mathbf{0}_{36 \times 1} & \mathbf{0}_{36 \times 1} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 1} \\ B^{IO} \mathbf{R}^{B^{IO}} \mathbf{k} & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{6 \times 1} & \mathbf{0}_{6 \times 1} \\ \mathbf{0}_{21 \times 1} & \mathbf{0}_{21 \times 1} \\ \mathbf{0}_{3 \times 1} & G^{IO} \mathbf{k} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{6 \times 1} & \mathbf{0}_{6 \times 1} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 1} \\ B^{IO} \mathbf{k} & \mathbf{0}_{3 \times 1} \\ 0 & 0 \\ \mathbf{0}_{3 \times 1} & G^{IO} \mathbf{k} \\ 0 & 0 \\ \mathbf{0}_{15 \times 1} & \mathbf{0}_{15 \times 1} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 1} \end{bmatrix} \quad (57)$$

Note that we, for the first time, explicitly found that one-axis rotation will cause the rotation calibration between auxiliary Gyro and base IMU to become unobservable. We verify this finding with simulations in Section VII-2.

B. Intrinsic for Base and Auxiliary Inertial Sensors

Unlike the work in [44] which does include gravity sensitivity in state vector, we consider calibrating multi-inertial sensors with a IMU model encapsulating the gravity sensitivity.

We analytically show that the 6 degenerate motion primitives presented in [44] still hold for IMU intrinsics even for our inertial model choice. The degenerate motions for base IMU, auxiliary IMU, and auxiliary Gyro are summarized in Tab. III. The detailed unobservable directions can be found in Appendix F.

Note that fully excited motions for the base IMU, all auxiliary IMUs, and auxiliary Gyros are needed in order to make all intrinsic parameters observable. We note that

TABLE IV: Simulation parameters and prior standard deviations that perturbations of measurements and initial states were drawn from.

Parameter	Value	Parameter	Value
IMU DW	0.003	IMU Da	0.003
Rot. atoI (rad)	0.003	IMU Tg	0.001
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)	1.0	Cam. Center (px)	1.0
d1 and d2	0.002	d3 and d4	0.002
Rot. CtoI (rad)	0.004	Pos. linC (m)	0.008
Pixel Proj. (px)	1	Cam-IMU Timeoff (s)	0.008
IMU-IMU Timeoff (s)	0.003	Gyro-IMU Timeoff (s)	0.003
Rot. AltoBI (rad)	0.003	Pos. AfinBI (m)	0.005
Cam Freq. (hz)	10 / 10 / 10	IMU Freq. (hz)	250 / 300 / 200

the gravity sensitivity is observable in the majority of the investigated motions.

VII. SIMULATION VALIDATIONS

The simulator provided within the OpenVINS project [14] along with the multi-IMU extension from [8] is leveraged to provide groundtruth measurements for verification of the observability of the proposed MVIS under different trajectory conditions. Specifically we transfer the acceleration from the groundtruth base IMU to the auxiliary IMU with the groundtruth angular velocity and acceleration, while the angular velocity of the secondary IMU can be simply rotated into the auxiliary frame. The basic configuration of the simulator is listed in Table IV and the two trajectories used are shown in Figure 5.

1) *Fully-Excited Motion*: We first evaluate the proposed system on a general 3D handheld trajectory, see Figure 5, which fully excites all 6 axes of the sensor platform. The sensor suite contains a base IMU, auxiliary IMU, gyroscope, and monocular camera. To save space we only report the parameters which relate to the inertial sensors and auxiliary IMU spacial parameters, but all parameters are perturbed and estimated during our simulation runs. The results shown in Figure 2 are of six different runs with different initial state perturbations. It is clear that all parameters are able to converge towards the true value within the first 20-40 seconds of the trajectory verifying that we are indeed able to perform calibration of all parameters for general motion trajectory, and thus verifies our analysis.

2) *Degenerate 1-Axis Motion*: We now perform a simulation where the trajectory only exhibits 1-axis rotation about the global z-axis to verify our identified degenerate motion. The sensor suite contains a base IMU, auxiliary IMU, gyroscope, and monocular camera. We generate a sinusoidal trajectory such that the roll and pitch of the orientation does not change over the trajectory (thus yaw only rotation), see Figure 5. Shown in Figure 3, multiple parameters are unable to converge, which match the parameters which we have identified as unobservable under this motion. We can see that the 3 parameters d_{w1} , d_{w2} and d_{w3} for both the base IMU, auxiliary IMU, and auxiliary gyroscope are unable to calibrate. Additionally, the camera to base IMU and

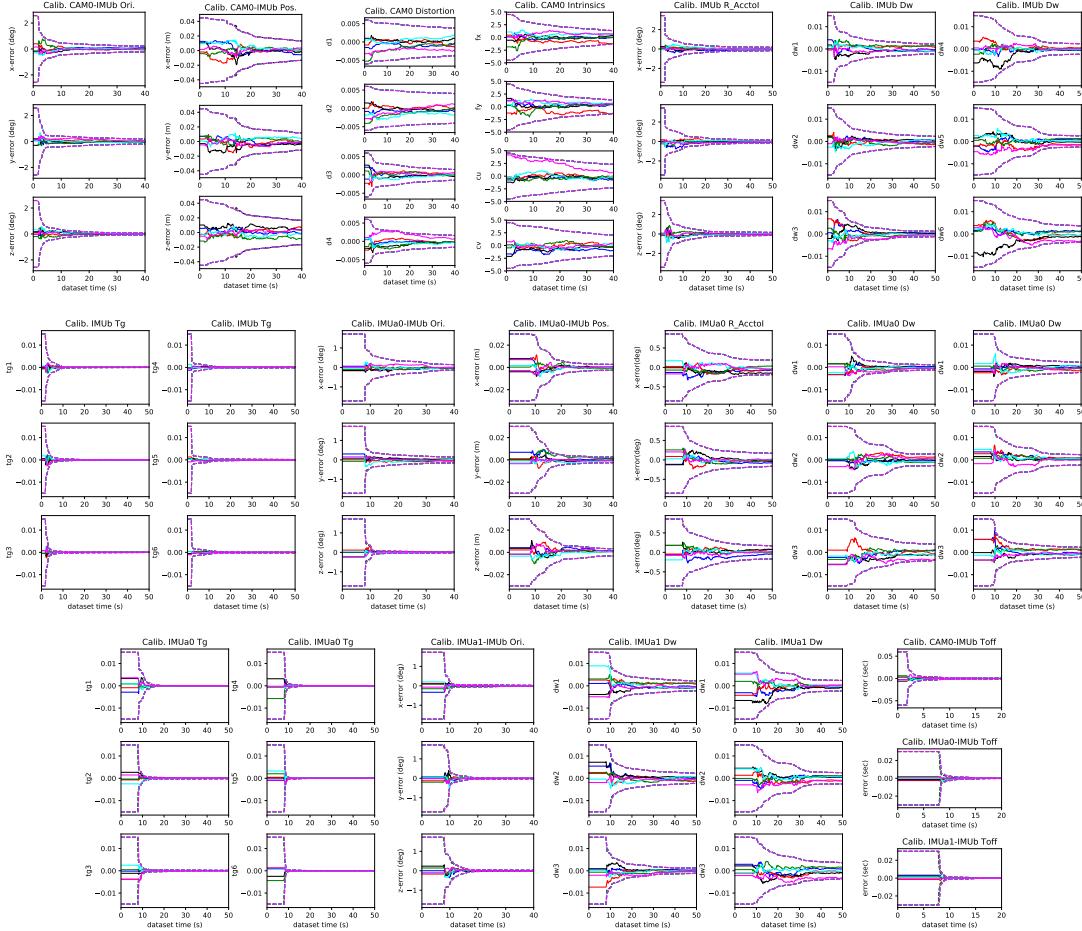


Fig. 2: Three sigma bounds (dashed line) and estimation errors (solid line) for six different runs (shown in different colors) with different initial calibration perturbations and realization of the measurement noise for *fully excited motion*. The sensor suite contains (from top to bottom, left to right) a single monocular camera (1-4), base IMU (5-9), auxiliary IMU (10-16), auxiliary gyroscope (17-19), and temporal time offsets. All the IMU intrinsic parameters converge nicely thus verifying that under general motion we are able to perform calibration of all parameters.

auxiliary IMU to base IMU position transform is unable to calibrate as expected. Furthermore it can be seen that we are unable to calibrate a portion of the relative rotation between the base IMU and auxiliary gyroscope which can be calibrated successfully in the full auxiliary IMU case.

3) *Consistency and Degeneracy Discussion:* During our investigation and verification of auxiliary IMU fusion algorithms, we stumbled upon an interesting by-product of degeneracy analysis and the simulation of degeneration trajectories to verify estimator consistency. We originally considered a simplified model which did not estimate the velocity of the auxiliary IMU and instead related it to the base IMU's velocity through the following (see Appendix B):

$${}^G \mathbf{v}_{AI} = {}^G \mathbf{v}_{BI} + {}^G \mathbf{R}_{|BI} {}^{|BI} \mathbf{R}_{|AI} {}^{|AI} \boldsymbol{\omega} {}^{|BI} \mathbf{p}_{AI} \quad (58)$$

This relation, if noise is not properly modeled for the ${}^{AI} \boldsymbol{\omega}$, will be inconsistent since the angular velocity will be taken as being true and thus inject incorrect

information into the system. Shown in Figure 4, when performing calibration on a general 3D trajectory, both systems (with and without auxiliary velocity in the state) are able to perform calibration accurately and consistently. Note, here we simulate a sensor suite that contain one base IMU, two auxiliary IMUs, and two cameras. On the other hand, if the systems are run on a degenerate motion trajectory then the first model correctly does not gain any information (sigma bound remains the same) in the unobservable direction. The second model, which has an underlying inconsistency and was able to perform consistent estimation during fully excited motion, is unable to calibrate and perform consistent calibration on the degenerate 1-axis motion trajectory. The complete results can be seen in Appendix G. Thus, we propose that future works also evaluate estimators on degenerate trajectories to further verify their estimator consistency, and to ensure that hidden inconsistencies in models come to light.

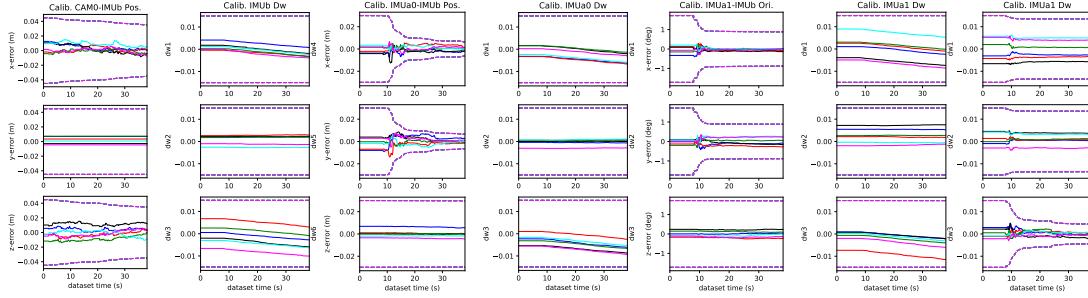


Fig. 3: Three sigma bounds (dashed line) and estimation errors (solid line) for six different runs (shown in different colors) with different initial calibration perturbations and realization of the measurement noise for *1-axis motion*. Only the calibration parameters which showed inability to converge (sigma bound does not decrease due to gain in information) are plotted for clarity. Also the Gyro IMUa1-IMUb global z-axis yaw orientation calibration is unable to converge as expected.

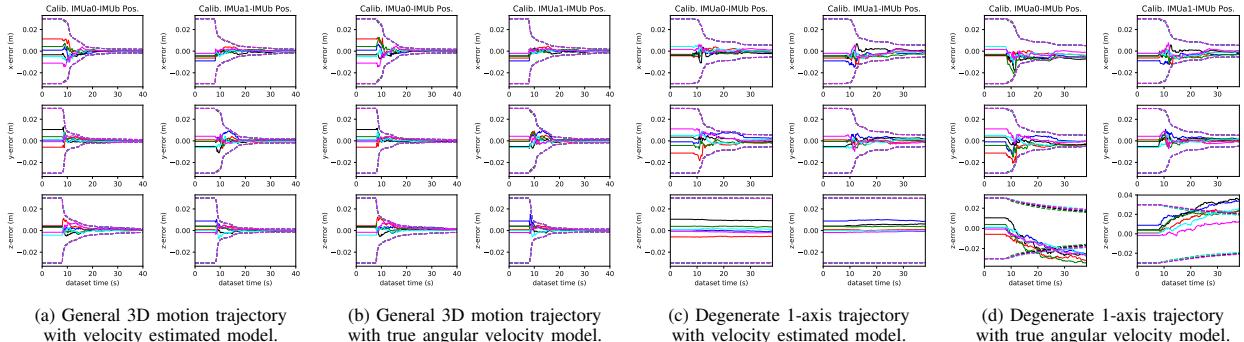


Fig. 4: Three sigma bounds (dashed line) and estimation errors (solid line) for six different runs (shown in different colors) with different initial calibration perturbations and realization of the measurement noise for the model with velocity in the state and one without on the general motion and degenerate trajectory. The sensor suite contains two auxiliary IMUs. It can be seen that in the general motion both models are able to consistently calibrate the relative position to the base IMU, while in the degenerate case only the first model which estimates velocity is able to remain consistent.

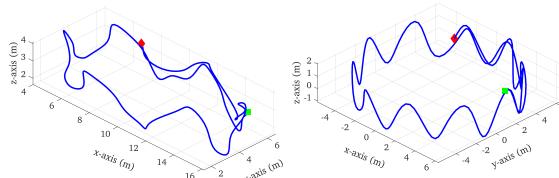


Fig. 5: Simulated trajectories, axes are in units of meters. General 3D trajectory (left) is 53 meters long, while the 1-axis degenerate trajectory (right) is 90 meters in length. Green square denotes the start and red diamond denotes the end.

4) MVIS Discussion: While the previous simulations have focuses on the single camera case, we have extensively verified the analysis in simulation for larger number of both camera and inertial sensors. The simulation presented here has focused on clarity, thus the plots of the multi-camera simulation have been left out (some can be seen in Appendix G) and instead we focus our real-world experiments on demonstrating the complete MVIS demonstration.

VIII. REAL-WORLD EXPERIMENTS

In addition to simulations, we also extensively examine the proposed MVIS on real-world datasets. We design two MVIS sensor rigs (Fig. 6) to collect datasets



Fig. 6: The self-assembled sensor rigs in real-world experiments. Left: VI-rigs containing 2 MircorStrain IMUs, 1 BlackFly camera and 1 IntelRealsense T265 tracking camera. Right: T265-rigs containing 2 Intel Realsense T265 cameras.

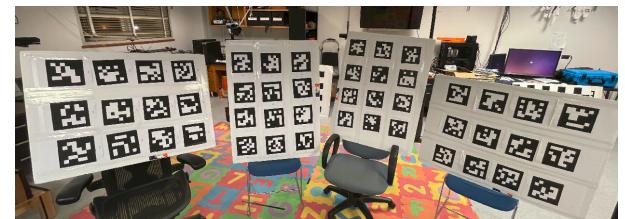


Fig. 7: ArUco tags for stable visual feature tracking.

and verify the proposed system. The VI-Rig (left of Fig. 6) contains 2 Mircosstrain IMUs¹, 1 BlackFly camera² and 1 Intel Realsense T265³. The T265-Rig (middle of Fig. 6) contains 2 Intel Realsense T265. The sensor

¹<https://www.microstrain.com/inertial-sensors/imu>

²<https://www.flir.com/products/blackfly-s-usb3/>

³<https://www.intelrealsense.com/tracking-camera-t265/>

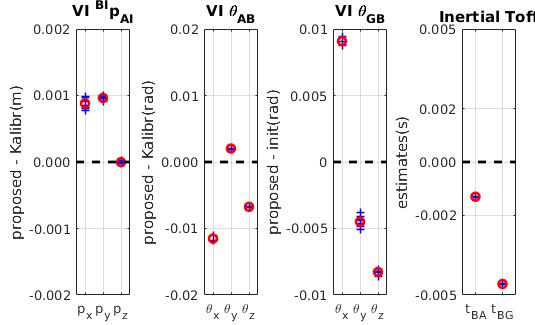


Fig. 8: The spatial-temporal calibration results for auxiliary IMU and gyro to base IMU in VI-Rig. Blue crosses represents the estimated values for each run. Red circle denotes the average value for all the runs. The black doted line is the value we get from Kalibr. It can be seen that the spatial-temporal parameters between these inertial sensors can converge nicely among different runs.

configuration are shown in Tab. V. In order to focus on examining the proposed MVIS fusion algorithm, we simplify the visual-tracking frontend by using ArUco Tags [12] for stable visual feature tracking (Fig. 7). One representative trajectory of a real-world experiment is shown on the right of Fig. 6.

With the VI-Rig, we collected 3 bags with all the sensors under fully excited motions. In order to test the repeatability of our proposed algorithm, we performed the experiments with 1, 2 and 3 cameras, respectively. Finally, we get 9 sets of results.

A. Spatial-Temporal Calibration

We also calibrate the VI-Rig with the well-known toolbox Kalibr [11] based on continuous-time batch optimization, and the results from Kalibr are also presented and coincide (or are close) with the calibration results from the proposed method. The calibration results can be shown in Fig. 8 and Fig. 9. Camera intrinsics are shown in Fig. 15. Note that the spatial-temporal parameter calibration are repeatable and very close to the values we get from Kalibr calibration, see Fig. 8.

B. IMU Intrinsics

Note we are unable to compare to Kalibr's IMU intrinsic model since our IMU intrinsic model is different with the one presented in Kalibr. Since we don't have the reference values for IMU intrinsics, we further evaluate our system with more datasets from T265-Rig. We collected 6 bags with the T265-Rig and performed experiments with 2 and 4 cameras, respectively. Finally, we get 12 sets of results. The intrinsics are shown in Fig. 10.

First, it is interesting to note that the IMU intrinsics of MicroStrain IMU are an order smaller than that of T265 IMU, especially for D_a , see Fig. 10. This is likely due to the cost differences of the two sensors, the quality of the default factory calibration, or that the T265 driver

interpolates its slower accelerometer from 62hz to 200hz to match the gyroscope.

Besides the high repeatability shown by the proposed MVIS calibration, the D_a , D_w , T_g for both MicroStrain IMUs in VI-Rig are quite similar, respectively, and the same case for the two T265 IMUs in T265-Rig. This is reasonable since the same type of IMUs should have similar IMU intrinsic calibration values (also shown in Fig. 10). This also supports the potential that better factory calibration can be performed to improve out-of-box sensor performance.

We also can see that while we are able to calibrate T_g repeatability, it is one order smaller in magnitude compared to that of D_a and D_w . Thus, this raises the question whether calibrating T_g will have a noticeable impact on MVIS performance.

IX. CONCLUSIONS AND FUTURE WORK

In this paper, we have developed a multi-visual-inertial sensor fusion (MVIS) algorithm, with a special focus on full calibration of all intrinsics, extrinsics, and time offset parameters between sensors. In particular, we proposed a novel auxiliary IMU measurement fusion based on ACI² preintegration to incorporate both spatial-temporal and intrinsic parameters. We, for the first time, have performed observability analysis for the MVIS system, proving that four canonical unobservable directions corresponding to global yaw and global translation remain, while the calibration parameters are observable under fully excited motion. Moreover, we have also identified the commonly seen degenerate motions that can cause certain calibration parameters unobservable. Extensive simulations have been performed to evaluate both the proposed system and verify the degenerate motion analysis. It is important to stress that degenerate motion can be used as a powerful tool to test system consistency and verify estimator designs. Two sensors rigs that consist of multiple IMUs and cameras were used for data collection and system evaluation in real-world experiments, where two widely used inertial sensors are calibrated and their need for intrinsic calibration was demonstrated. In the future, we will study in-depth the MVIS calibration under degenerate motions (e.g., on ground or micro-aerial vehicles), and will also investigate efficient marginalization techniques to enable the proposed batch optimization-based MVIS with online calibration be amenable for real-time navigaiton.

REFERENCES

- [1] T. D. Barfoot. *State estimation for robotics*. Cambridge University Press, 2017.
- [2] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart. Robust visual inertial odometry using a direct ekf-based approach. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 298–304. IEEE, 2015.

TABLE V: The sensor setup for VI-Rig and T265-Rig.

	IMUb	IMUa	Gyro	Cam0	Cam1	Cam2	Cam3
VI-Rig	MicroStrain0 500hz	MicroStrain1 100hz	T265 Gyro 200hz	T265 Left 30hz	Blackfly 25hz	T265 Right 30hz	-
T265-Rig	T265 IMU0 200hz	T265 IMU1 200hz	-	T265 Left0 30hz	T265 Right0 30hz	T265 Left1 30hz	T265 Right1 30hz

- [3] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós. Orb-slam3: An accurate open-source library for visual, visual-inertial and multi-map slam. *arXiv preprint arXiv:2007.11898*, 2020.
- [4] F. Dellaert. Factor graphs and gtsam: A hands-on introduction. Technical report, Georgia Institute of Technology, 2012.
- [5] K. Eckenhoff, P. Geneva, J. Bloecker, and G. Huang. Multi-camera visual-inertial navigation with online intrinsic and extrinsic calibration. In *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.
- [6] K. Eckenhoff, P. Geneva, and G. Huang. Closed-form preintegration methods for graph-based visual-inertial navigation. *International Journal of Robotics Research*, 38(5):563–586, 2019.
- [7] K. Eckenhoff, P. Geneva, and G. Huang. Sensor-failure-resilient multi-imu visual-inertial navigation. In *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.
- [8] K. Eckenhoff, P. Geneva, and G. Huang. MIMC-VINS: A versatile and resilient multi-imu multi-camera visual-inertial navigation system. *IEEE Transactions on Robotics*, Feb. 2021.
- [9] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza. On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics*, 33(1):1–21, 2016.
- [10] B. Fu, F. Han, Y. Wang, Y. Jiao, X. Ding, Q. Tan, L. Chen, M. Wang, and R. Xiong. High precision multi-camera assisted camera-imu calibration: Theory and method. *IEEE Transactions on Instrumentation and Measurement*, 2021.
- [11] P. Furgale, J. Rehder, and R. Siegwart. Unified temporal and spatial calibration for multi-sensor systems. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1280–1286. IEEE, 2013.
- [12] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and R. Medina-Carnicer. Generation of fiducial marker dictionaries using mixed integer linear programming. *Pattern Recognition*, 51:481–491, 2016.
- [13] P. Geneva, K. Eckenhoff, and G. Huang. A linear-complexity EKF for visual-inertial navigation with loop closures. In *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.
- [14] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang. Openvins: A research platform for visual-inertial estimation. In *Proc. of the IEEE International Conference on Robotics and Automation*, Paris, France, 2020.
- [15] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis. Consistency analysis and improvement of vision-aided inertial navigation. *IEEE Transactions on Robotics*, 30(1):158–176, 2013.
- [16] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis. Consistency analysis and improvement of vision-aided inertial navigation. *IEEE Transactions on Robotics*, 30(1):158–176, 2014.
- [17] G. Huang. *Improving the Consistency of Nonlinear Estimators: Analysis, Algorithms, and Applications*. PhD thesis, Department of Computer Science and Engineering, University of Minnesota, 2012.
- [18] G. Huang. Visual-inertial navigation: A concise review. In *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.
- [19] E. S. Jones and S. Soatto. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *The International Journal of Robotics Research*, 30(4):407–430, 2011.
- [20] J. Kelly and G. S. Sukhatme. Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration. *The International Journal of Robotics Research*, 30(1):56–79, 2011.
- [21] D. Kim, S. Shin, and I. S. Kweon. On-line initialization and extrinsic calibration of an inertial navigation system with a relative preintegration method on manifold. *IEEE Transactions on Automation Science and Engineering*, 15(3):1272–1285, 2017.
- [22] J. Kuo, M. Muglikar, Z. Zhang, and D. Scaramuzza. Redesigning slam for arbitrary multi-camera systems. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2116–2122. IEEE, 2020.
- [23] W. Lee, K. Eckenhoff, Y. Yang, P. Geneva, and G. Huang. Visual-inertial-wheel odometry with online calibration. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Las Vegas, NV, 2020.
- [24] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale. Keyframe-based visual-inertial odometry using nonlinear optimization. *The Inter-*

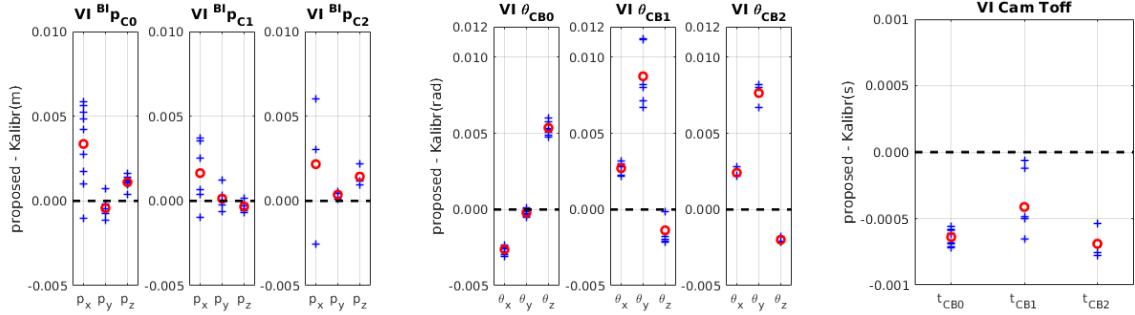


Fig. 9: The spatial-temporal calibration results for 3 cameras to base IMU in VI-Rig. Blue crosses represents the estimated values for each run. Red circle denotes the average value for all the runs. The black dotted line is the value from Kalibr. It can clearly be seen that the camera calibration results are very similar to that we get from Kalibr.

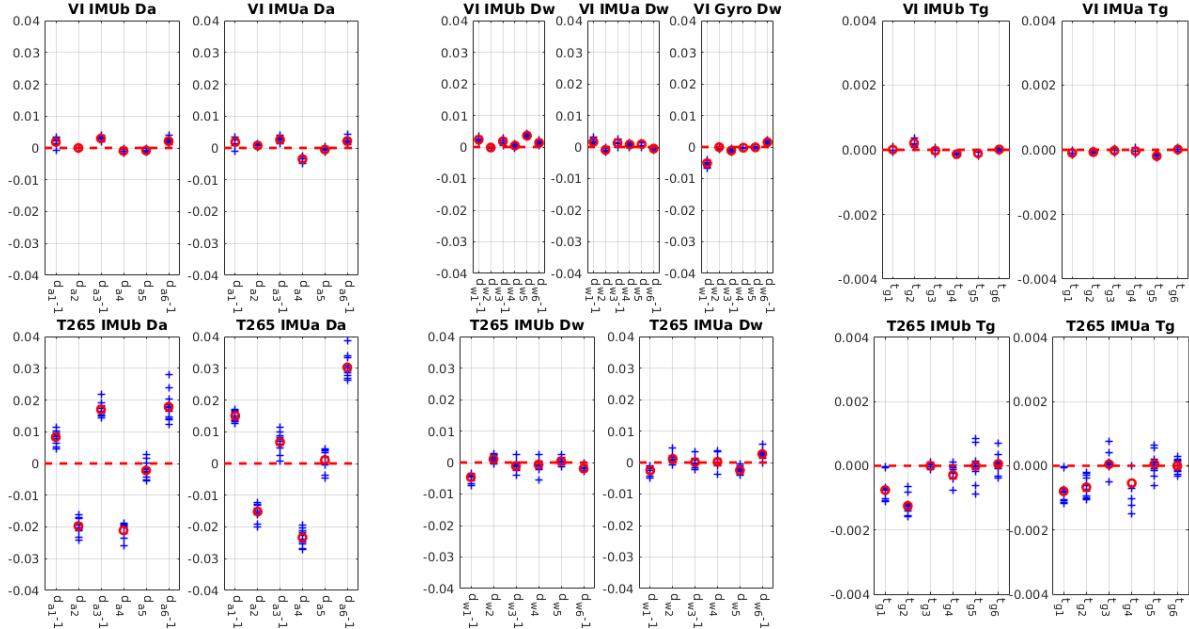


Fig. 10: The IMU intrinsics calibration results for VI-Rig and T265-Rig. Blue crosses represents the estimated values for each run. Red circle denotes the average value for all the runs. The red dotted line is the initial guess, which is also the ideal value we expected. The nearer the estimated value is to the red dotted line, the better the IMU is. To better demonstrate the calibration results, we use the same scale of y axis for both \mathbf{D}_a and \mathbf{D}_w . It can be seen that the Mirostrain IMU demonstrates a higher accuracy than the low-end Intel Realsense IMU no matter for \mathbf{D}_a , \mathbf{D}_w and \mathbf{T}_g .

- national Journal of Robotics Research*, 34(3):314–334, 2015.
- [25] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale. Keyframe-based visual–inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 34(3):314–334, 2015.
- [26] M. Li, H. Yu, X. Zheng, and A. I. Mourikis. High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 409–416. IEEE, 2014.
- [27] J. Ma, M. Bajracharya, S. Susca, L. Matthies, and M. Malchano. Real-time pose estimation of a dynamic quadruped in gps-denied environments for 24-hour operation. *The International Journal of Robotics Research*, 35(6):631–653, 2016.
- [28] A. Martinelli. Observability properties and deterministic algorithms in visual-inertial structure from motion. *Foundations and Trends in Robotics*, pages 1–75, 2013.
- [29] F. M. Mirzaei and S. I. Roumeliotis. A kalman filter-based algorithm for imu-camera calibration: Observability analysis and performance evaluation. *IEEE transactions on robotics*, 24(5):1143–1156, 2008.
- [30] A. I. Mourikis and S. I. Roumeliotis. A multi-state constraint kalman filter for vision-aided inertial navigation. In *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pages 3565–3572. IEEE, 2007.
- [31] M. K. Paul and S. I. Roumeliotis. Alternating-

- stereo vins: Observability analysis and performance evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4729–4737, 2018.
- [32] M. K. Paul, K. Wu, J. A. Hesch, E. D. Nerurkar, and S. I. Roumeliotis. A comparative analysis of tightly-coupled monocular, binocular, and stereo vins. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 165–172. IEEE, 2017.
- [33] T. Qin, P. Li, and S. Shen. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4):1004–1020, 2018.
- [34] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart. Extending kalibr: Calibrating the extrinsics of multiple imus and of individual axes. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4304–4311. IEEE, 2016.
- [35] T. Schneider, M. Li, C. Cadena, J. Nieto, and R. Siegwart. Observability-aware self-calibration of visual and inertial sensors for ego-motion estimation. *IEEE Sensors Journal*, 19(10):3846–3860, 2019.
- [36] J. Sola. Quaternion kinematics for the error-state kalman filter. *arXiv preprint arXiv:1711.02508*, 2017.
- [37] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, and V. Kumar. Robust stereo visual inertial odometry for fast autonomous flight. *IEEE Robotics and Automation Letters*, 3(2):965–972, 2018.
- [38] N. Trawny and S. I. Roumeliotis. Indirect kalman filter for 3d attitude estimation. *University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep*, 2:2005, 2005.
- [39] V. Usenko, J. Engel, J. Stückler, and D. Cremers. Direct visual-inertial odometry with stereo cameras. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1885–1892. IEEE, 2016.
- [40] K. J. Wu, C. X. Guo, G. Georgiou, and S. I. Roumeliotis. Vins on wheels. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5155–5162. IEEE, 2017.
- [41] Y. Xiao, X. Ruan, J. Chai, X. Zhang, and X. Zhu. Online imu self-calibration for visual-inertial systems. *Sensors*, 19(7):1624, 2019.
- [42] Y. Yang, B. P. W. Babu, C. Chen, G. Huang, and L. Ren. Analytic combined imu integrator for visual-inertial navigation. In *Proc. of the IEEE International Conference on Robotics and Automation*, Paris, France, 2020.
- [43] Y. Yang, P. Geneva, K. Eckenhoff, and G. Huang. Degenerate motion analysis for aided INS with online spatial and temporal calibration. *IEEE Robotics and Automation Letters (RA-L)*, 4(2):2070–2077, 2019.
- [44] Y. Yang, P. Geneva, X. Zuo, and G. Huang. Online imu intrinsic calibration: Is it necessary? In *Proc. of the Robotics: Science and Systems*, Paris, France, 2020.
- [45] Y. Yang and G. Huang. Aided inertial navigation with geometric features: Observability analysis. In *Proc. of the IEEE International Conference on Robotics and Automation*, Brisbane, Australia, May 21–25, 2018.
- [46] Y. Yang and G. Huang. Aided inertial navigation: Unified feature representations and observability analysis. In *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.
- [47] Y. Yang and G. Huang. Observability analysis of aided ins with heterogeneous features of points, lines and planes. *IEEE Transactions on Robotics*, 35(6):399–1418, Dec. 2019.
- [48] M. Zhang, X. Xu, Y. Chen, and M. Li. A lightweight and accurate localization algorithm using multiple inertial measurement units. *IEEE Robotics and Automation Letters*, 5(2):1508–1515, 2020.

APPENDIX A LINEAR POSE INTERPOLATION

A. Kinematic-based Interpolation

Since our state contains inertial states only at specific times, we need to be able to write our asynchronous measurements as a function of these poses. We can relate any time to the nearest inertial state by leveraging the IMU kinematic, ${}^G\mathbf{v}_{BI}$ and ${}^{BI}\boldsymbol{\omega}$, to directly compute the expected pose at our measurement time. For example, the measurement from sensor S at time $t_{S,in}$ requires getting the pose $\{{}^G_{S_{in}}\mathbf{R}, {}^G_{S_{in}}\mathbf{p}_S\}$. We can interpolate to this pose by first finding how much to interpolate by in the sensor clock frame $\delta t = t_{S,in} + t_{Sb} - t_{BI_k}$, where we have used the time offset between the measurement clock and the IMU clock frame which our inertial poses occur at. We can then define the following equations:

$${}^G_{S_{in}}\mathbf{R} = {}^G_{BI_k}\mathbf{R} \text{Exp}({}^{BI_k}\boldsymbol{\omega} \delta t) {}^B_I \mathbf{R} \quad (59)$$

$${}^G_{S_{in}}\mathbf{p}_S = {}^G_{BI_k}\mathbf{p}_S + {}^G_{BI_k}\mathbf{v}_{BI_k} \delta t + {}^G_{S_{in}}\mathbf{R} {}^B_I \mathbf{p}_S \quad (60)$$

where $\text{Exp}(\cdot)$ represents the exponential operation for SO(3). It is important to note here that this interpolation both allows us to find the derivative in respect to both the temporal, t_{SB} , and spacial, $\{{}^B_I \mathbf{R}, {}^B_I \mathbf{p}_S\}$ parameters. This interpolation is only a good approximation when near the pose we are interpolating from, thus we only leverage this for pose measurements which occur near the inertial state time (e.g., our base camera sensor measurements).

B. Pose-based Interpolation

Another more conventional pose interpolation scheme is to find a pose between two bounding inertial states. We leverage this non-base cameras which are more likely to have large amounts of interpolation. To find the sensor pose $\{\mathbf{S}_{in}^G \mathbf{R}, \mathbf{p}_{S_{in}}^G\}$ with the two bounding IMU states \mathbf{x}_{BI_k} and $\mathbf{x}_{BI_{k+1}}$ we can define the following:

$$\begin{aligned}\mathbf{S}_{in}^G \mathbf{R} &= \mathbf{B}_{I_k}^G \mathbf{R} \text{Exp}(\lambda \text{Log}(\mathbf{B}_{I_{k+1}}^G \mathbf{R})) \mathbf{S}_S^{BI} \mathbf{R} \\ \mathbf{p}_{S_{in}}^G &= \mathbf{p}_{BI_k}^G + \mathbf{B}_{I_k}^G \mathbf{R} (\lambda \mathbf{B}_{I_k}^{BI} \mathbf{p}_{BI_{k+1}}) + \mathbf{S}_{in}^G \mathbf{R}^{BI} \mathbf{p}_S \\ \lambda &= (t_{S,in} + t_{SB} - t_{BI_k}) / (t_{BI_{k+1}} - t_{BI_k})\end{aligned}$$

where $t_{S,in}$ and t_{SI} are the measurement time and sensor time offset, $\{\mathbf{S}_S^{BI} \mathbf{R}, \mathbf{p}_S^S\}$ represents the sensor rigid transformation to IMU frame, and $\text{Log}(\cdot)$ represents the log operation for SO(3). Same as before, this interpolation allows us to find derivatives in respect to these calibration parameters and thus perform calibration. This interpolation scheme is used for the majority of measurements which do not occur near the IMU pose time like the base camera.

APPENDIX B AUXILIARY IMU WITH POSE AND VELOCITY CONSTRAINTS

In the above constraints, we need to keep the velocity of auxiliary IMU in the state vector. However, we can also take advantage of the velocity constraints between the two IMUs. If so, we can get rid of the auxiliary velocity from the state vector, the state of auxiliary IMU state becomes:

$$\mathbf{x}_{AI} = [{}^{AI} \mathbf{b}_g^\top \quad {}^{AI} \mathbf{b}_a^\top]^\top \quad (61)$$

and have the rigid body kinematic constraints as:

$${}^{AI} \bar{q} = {}^{BI} \bar{q} \otimes {}^{AI} \bar{q} \quad (62)$$

$${}^G \mathbf{p}_{AI} = {}^G \mathbf{p}_{BI} + {}^{BI} \mathbf{R} {}^{BI} \mathbf{p}_{AI} \quad (63)$$

$${}^G \mathbf{v}_{AI} = {}^G \mathbf{v}_{BI} + {}^{BI} \mathbf{R} [{}^{AI} \mathbf{R} {}^{AI} \boldsymbol{\omega}] {}^{BI} \mathbf{p}_{AI} \quad (64)$$

With this, we can reformulate the IMU measurements as:

$$\begin{aligned}{}^{AI} \Delta \bar{q} &\triangleq \mathbf{h}_q^{(AI)}(\mathbf{x}_{BI_{k:k+1}}, \mathbf{x}_{BA}) \\ {}^{AI} \Delta \mathbf{p} &\triangleq \mathbf{h}_p^{(AI)}(\mathbf{x}_{BI_{k:k+1}}, \mathbf{x}_{BA}) \\ {}^{AI} \Delta \mathbf{v} &\triangleq \mathbf{h}_v^{(AI)}(\mathbf{x}_{BI_{k:k+1}}, \mathbf{x}_{BA}) \\ {}^{AI} \Delta \mathbf{b} &\triangleq {}^{AI} \mathbf{b}_{k+1} - {}^{AI} \mathbf{b}_k\end{aligned}$$

Then, we can follow similar way as (33) and (31) to defined auxiliary IMU cost functions \mathbb{C}_{AI} with or without IMU intrinsics \mathbf{x}_{An} .

APPENDIX C STATE TRANSITION MATRIX

The state transition matrix can be written as:

$$\Phi = \begin{bmatrix} \Phi_B & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Phi_A & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Phi_G & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \Phi_{calib} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \Phi_F \end{bmatrix} \quad (65)$$

where we have:

$$\Phi_B = \begin{bmatrix} \Phi_{BI} & \Phi_{Bn} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad (66)$$

$$\Phi_A = \begin{bmatrix} \Phi_{AI} & \Phi_{An} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad (67)$$

$$\Phi_G = \begin{bmatrix} \Phi_{GI} & \Phi_{Gn} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \quad (68)$$

$$\Phi_{calib} = \mathbf{I} \quad (69)$$

$$\Phi_F = \mathbf{I} \quad (70)$$

The state transition of Φ_B and Φ_A have the same structure. We can grab the gyro part of the state transition of Φ_B to get Φ_G . Hence, we only show Φ_B for clarity.

$$\Phi_{BI} = \begin{bmatrix} \Phi_{B11} & \mathbf{0}_3 & \mathbf{0}_3 & \Phi_{B14} & \Phi_{B15} \\ \Phi_{B21} & \mathbf{I}_3 & \mathbf{I}_3 \delta t & \Phi_{B24} & \Phi_{B25} \\ \Phi_{B31} & \mathbf{0}_3 & \mathbf{I}_3 & \Phi_{B34} & \Phi_{B35} \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \quad (71)$$

where we have:

$$\Phi_{B11} = \frac{{}^{BI_{k+1}} \hat{\mathbf{R}}}{{}^{BI_k}}$$

$$\Phi_{B21} = -[{}^G \hat{\mathbf{p}}_{BI_{k+1}} - {}^G \hat{\mathbf{p}}_{BI_k} - {}^G \hat{\mathbf{v}}_{BI_k} \delta t_k + \frac{1}{2} {}^G \mathbf{g} \delta t_k^2] \frac{{}^G \mathbf{R}}{{}^{BI_k}}$$

$$\Phi_{B31} = -[{}^G \hat{\mathbf{v}}_{BI_{k+1}} - {}^G \hat{\mathbf{v}}_{BI_k} + {}^G \mathbf{g} \delta t_k] \frac{{}^G \mathbf{R}}{{}^{BI_k}} \hat{\mathbf{R}}$$

$$\Phi_{B14} = -\mathbf{J}_r \delta t_k \frac{{}^{BI} \hat{\mathbf{R}}}{{}^{BW}} \hat{\mathbf{D}}_{BW}$$

$$\Phi_{B24} = \frac{{}^G \mathbf{R}}{{}^{BI_k}} \hat{\mathbf{R}} \mathbf{\Xi}_{4BW}^{BI} \hat{\mathbf{R}} \hat{\mathbf{D}}_{BW}$$

$$\Phi_{B34} = \frac{{}^G \mathbf{R}}{{}^{BI_k}} \hat{\mathbf{R}} \mathbf{\Xi}_{3BW}^{BI} \hat{\mathbf{R}} \hat{\mathbf{D}}_{BW}$$

$$\Phi_{B15} = -\mathbf{J}_r \delta t_k \frac{{}^{BI} \hat{\mathbf{R}}}{{}^{BW}} \hat{\mathbf{D}}_{BW} \hat{\mathbf{T}}_{BgBa} \frac{{}^{BI} \hat{\mathbf{R}}}{{}^{Ba}} \hat{\mathbf{D}}_{Ba}$$

$$\Phi_{B25} = -\mathbf{J}_r \delta t_k \frac{{}^{BI} \hat{\mathbf{R}}}{{}^{BW}} \hat{\mathbf{D}}_{BW} \hat{\mathbf{T}}_{BgBa} \frac{{}^{BI} \hat{\mathbf{R}}}{{}^{Ba}} \hat{\mathbf{D}}_{Ba}$$

$$\Phi_{Bn} = \begin{bmatrix} \Phi_{Bn11} & \Phi_{Bn12} & \Phi_{Bn13} & \Phi_{Bn14} \\ \Phi_{Bn21} & \Phi_{Bn22} & \Phi_{Bn23} & \Phi_{Bn24} \\ \Phi_{Bn31} & \Phi_{Bn23} & \Phi_{Bn33} & \Phi_{Bn34} \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \quad (72)$$

where we have:

$$\Phi_{Bn11} = \mathbf{J}_r \delta t_k \frac{{}^{BI} \hat{\mathbf{R}}}{{}^{BW}} \hat{\mathbf{H}}_{D_{BW}}$$

$$\Phi_{Bn21} = -\frac{{}^G \mathbf{R}}{{}^{BI_k}} \mathbf{\Xi}_{4BW}^{BI} \hat{\mathbf{R}} \hat{\mathbf{H}}_{D_{BW}}$$

$$\Phi_{Bn31} = -\frac{{}^G \mathbf{R}}{{}^{BI_k}} \mathbf{\Xi}_{3BW}^{BI} \hat{\mathbf{R}} \hat{\mathbf{H}}_{D_{BW}}$$

$$\Phi_{Bn12} = -\mathbf{J}_r \delta t_k \frac{{}^{BI} \hat{\mathbf{R}}}{{}^{BW}} \hat{\mathbf{D}}_{BW} \hat{\mathbf{T}}_{BgBa} \frac{{}^{BI} \hat{\mathbf{R}}}{{}^{Ba}} \hat{\mathbf{H}}_{D_{Ba}}$$

$$\Phi_{Bn22} = \frac{{}^G \mathbf{R}}{{}^{BI_k}} \hat{\mathbf{R}} \left(\mathbf{\Xi}_2 + \mathbf{\Xi}_{4BW}^{BI} \hat{\mathbf{R}} \hat{\mathbf{D}}_{BW} \hat{\mathbf{T}}_{Bg} \right) \frac{{}^{BI} \hat{\mathbf{R}}}{{}^{Ba}} \hat{\mathbf{H}}_{D_{Ba}}$$

$$\begin{aligned}
\Phi_{Bn32} &= {}^G_{BI_k} \hat{\mathbf{R}} \left(\Xi_1 + \Xi_3 {}^{BI}_{Bw} \hat{\mathbf{R}} \hat{\mathbf{D}}_{Bw} \hat{\mathbf{T}}_{Bg} \right) {}^{BI}_{Ba} \hat{\mathbf{R}} \mathbf{H}_{D_{Ba}} \\
\Phi_{Bn13} &= \mathbf{J}_r \delta t_k {}^{BI}_{Bw} \hat{\mathbf{R}} \hat{\mathbf{D}}_{Bw} \hat{\mathbf{T}}_{Bg} [{}^{BI} \hat{\mathbf{a}}] {}^{BI}_{Ba} \hat{\mathbf{R}} \\
\Phi_{Bn23} &= -{}^G_{BI_k} \hat{\mathbf{R}} \left(\Xi_2 + \Xi_4 {}^{BI}_{Bw} \hat{\mathbf{R}} \hat{\mathbf{D}}_{Bw} \hat{\mathbf{T}}_{Bg} \right) [{}^{BI} \hat{\mathbf{a}}] {}^{BI}_{Ba} \hat{\mathbf{R}} \\
\Phi_{Bn33} &= -{}^G_{BI_k} \hat{\mathbf{R}} \left(\Xi_1 + \Xi_3 {}^{BI}_{Bw} \hat{\mathbf{R}} \hat{\mathbf{D}}_{Bw} \hat{\mathbf{T}}_{Bg} \right) [{}^{BI} \hat{\mathbf{a}}] {}^{BI}_{Ba} \hat{\mathbf{R}} \\
\Phi_{Bn14} &= -\mathbf{J}_r \delta t_k {}^{BI}_{Bw} \hat{\mathbf{R}} \hat{\mathbf{D}}_{Bw} \mathbf{H}_{T_{Bg}} \\
\Phi_{Bn24} &= {}^G_{BI_k} \hat{\mathbf{R}} \Xi_4 {}^{BI}_{Bw} \hat{\mathbf{R}} \hat{\mathbf{D}}_{Bw} \mathbf{H}_{T_{Bg}} \\
\Phi_{Bn34} &= {}^G_{BI_k} \hat{\mathbf{R}} \Xi_3 {}^{BI}_{Bw} \hat{\mathbf{R}} \hat{\mathbf{D}}_{Bw} \mathbf{H}_{T_{Bg}}
\end{aligned}$$

APPENDIX D MEASUREMENT JACOBIANS

The corresponding Jacobians can be listed as:

$$\frac{\partial \tilde{\mathbf{x}}}{\partial \tilde{\mathbf{x}}} = \begin{bmatrix} \frac{\partial \tilde{\mathbf{z}}_C}{\partial \tilde{\mathbf{x}}} \\ \frac{\partial \tilde{\mathbf{z}}_A}{\partial \tilde{\mathbf{x}}} \\ \frac{\partial \tilde{\mathbf{z}}_G}{\partial \tilde{\mathbf{x}}} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{CB} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{CC} & \mathbf{H}_{CF} \\ \mathbf{H}_{AB} & \mathbf{H}_{AA} & \mathbf{0} & \mathbf{H}_{AC} & \mathbf{0} \\ \mathbf{H}_{GB} & \mathbf{0} & \mathbf{H}_{GG} & \mathbf{H}_{GC} & \mathbf{0} \end{bmatrix} \quad (73)$$

where:

$$\begin{aligned}
\mathbf{H}_{CB} &= \mathbf{H}_{proj_C} {}^{BI} \hat{\mathbf{R}} {}^{TG}_{BI} \hat{\mathbf{R}}^\top \times \\
&\quad [[{}^G \mathbf{p}_f - {}^G \mathbf{p}_{BI}] {}^{G}_{BI} \hat{\mathbf{R}} \quad -\mathbf{I}_3 \quad \mathbf{0}_{3 \times 30}] \\
\mathbf{H}_{CC} &= \mathbf{H}_{proj_C} {}^{BI} \hat{\mathbf{R}} {}^{TG}_{BI} \hat{\mathbf{R}}^\top \\
&\quad [\mathbf{0}_{3 \times 7} \quad \mathbf{0}_{3 \times 4} \quad \mathbf{H}_{CC3} \quad -{}^G_{BI} \hat{\mathbf{R}} \quad \mathbf{H}_{CC5} \quad \mathbf{0}_{3 \times 8}] \\
\mathbf{H}_{CC3} &= [{}^G \mathbf{p}_f - {}^G \mathbf{p}_{BI} - {}^G_{BI} \hat{\mathbf{R}} {}^{BI} \hat{\mathbf{p}}_C] \\
\mathbf{H}_{CC5} &= {}^G \mathbf{v}_{BI} - [{}^G \mathbf{p}_f - {}^G \mathbf{p}_{BI}] {}^{G}_{BI} \hat{\mathbf{R}} {}^{BI} \boldsymbol{\omega} \\
\mathbf{H}_{CF} &= {}^G_{BI} \hat{\mathbf{R}} {}^{TG}_{BI} \hat{\mathbf{R}} \\
\mathbf{H}_{AB} &= \begin{bmatrix} {}^{AI} \hat{\mathbf{R}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 21} \\ {}^G_{BI} \hat{\mathbf{R}} [{}^{BI} \hat{\mathbf{p}}_{AI}] & -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 21} \end{bmatrix} \\
\mathbf{H}_{AA} &= \begin{bmatrix} -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 21} \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 21} \end{bmatrix} \\
\mathbf{H}_{AC} &= \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 & & -{}^{AI} \boldsymbol{\omega} & & \mathbf{0}_{3 \times 19} \\ \mathbf{0}_3 & -{}^G_{BI} \hat{\mathbf{R}} & {}^G \mathbf{v}_{BI} + {}^G_{BI} \hat{\mathbf{R}} [{}^{BI} \boldsymbol{\omega}] {}^{BI} \mathbf{p}_{AI} & \mathbf{0}_{3 \times 19} & & \end{bmatrix} \\
\mathbf{H}_{GB} &= [{}^{GI} \hat{\mathbf{R}} \quad \mathbf{0}_3 \quad \mathbf{0}_3 \quad \mathbf{0}_3 \quad \mathbf{0}_3 \quad \mathbf{0}_{3 \times 6}] \\
\mathbf{H}_{GG} &= [-\mathbf{I}_3 \quad \mathbf{0}_3] \\
\mathbf{H}_{GC} &= [\mathbf{0}_{3 \times 7} \quad \mathbf{I}_3 \quad -{}^{GI} \hat{\boldsymbol{\omega}} \quad \mathbf{0}_{3 \times 7} \quad \mathbf{0}_{3 \times 8}]
\end{aligned}$$

APPENDIX E OBSERVABILITY MATRIX

The observability matrix can be written as:

$$\begin{aligned}
\mathbf{M} &= \begin{bmatrix} \frac{\partial \tilde{\mathbf{z}}_C}{\partial \tilde{\mathbf{x}}} \\ \frac{\partial \tilde{\mathbf{z}}_A}{\partial \tilde{\mathbf{x}}} \\ \frac{\partial \tilde{\mathbf{z}}_G}{\partial \tilde{\mathbf{x}}} \end{bmatrix} \Phi \quad (74) \\
&= \begin{bmatrix} \mathbf{M}_{CB} & \mathbf{0} & \mathbf{0} & \mathbf{M}_{CC} & \mathbf{M}_{CF} \\ \mathbf{M}_{AB} & \mathbf{M}_{AA} & \mathbf{0} & \mathbf{M}_{AC} & \mathbf{0} \\ \mathbf{M}_{GB} & \mathbf{0} & \mathbf{M}_{GG} & \mathbf{M}_{GC} & \mathbf{0} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{H}_{CB} \Phi_B & \mathbf{0} & \mathbf{0} & \mathbf{H}_{CC} \Phi_C & \mathbf{H}_{CF} \Phi_F \\ \mathbf{H}_{AB} \Phi_B & \mathbf{H}_{AA} \Phi_A & \mathbf{0} & \mathbf{H}_{AC} \Phi_C & \mathbf{0} \\ \mathbf{H}_{GB} \Phi_B & \mathbf{0} & \mathbf{H}_{GG} \Phi_G & \mathbf{H}_{GC} \Phi_C & \mathbf{0} \end{bmatrix}
\end{aligned}$$

APPENDIX F DEGENERATE MOTION FOR INERTIAL INTRINSICS

The degenerate motions for base IMU, auxiliary IMU and auxiliary Gyro are similar. Hence, for simplicity, we save the notation for frame reference and explain the degenerate motion for base IMU as an example.

1) ${}^w w_1$ constant: If ${}^w w_1$ is constant, d_{a11} will be unobservable with unobservable directions as:

$$\mathbf{N}_{w1} = [\mathbf{0}_{1 \times 9} \quad (\hat{\mathbf{D}}_w^{-1} \mathbf{e}_1)^\top {}^w w_1 \quad \mathbf{0}_{1 \times 3} \quad 1 \quad \mathbf{0}_{1 \times 91}]^\top \quad (75)$$

2) ${}^w w_2$ constant: If ${}^w w_2$ is constant, d_{a12} and d_{a22} will be unobservable with unobservable directions as:

$$\mathbf{N}_{w2} = \begin{bmatrix} \mathbf{0}_{1 \times 9} & (\hat{\mathbf{D}}_w^{-1} \mathbf{e}_1)^\top {}^w w_2 & \mathbf{0}_{1 \times 4} & 1 & \mathbf{0}_{1 \times 90} \\ \mathbf{0}_{1 \times 9} & (\hat{\mathbf{D}}_w^{-1} \mathbf{e}_2)^\top {}^w w_2 & \mathbf{0}_{1 \times 5} & 1 & \mathbf{0}_{1 \times 89} \end{bmatrix}^\top \quad (76)$$

3) ${}^w w_3$ constant: If ${}^w w_3$ is constant, d_{a13} , d_{a23} and d_{a33} are unobservable with unobservable directions as:

$$\mathbf{N}_{w3} = \begin{bmatrix} \mathbf{0}_{1 \times 9} & (\hat{\mathbf{D}}_w^{-1} \mathbf{e}_1)^\top {}^w w_3 & \mathbf{0}_{1 \times 6} & 1 & \mathbf{0}_{1 \times 88} \\ \mathbf{0}_{1 \times 9} & (\hat{\mathbf{D}}_w^{-1} \mathbf{e}_2)^\top {}^w w_3 & \mathbf{0}_{1 \times 7} & 1 & \mathbf{0}_{1 \times 87} \\ \mathbf{0}_{1 \times 9} & (\hat{\mathbf{D}}_w^{-1} \mathbf{e}_3)^\top {}^w w_3 & \mathbf{0}_{1 \times 8} & 1 & \mathbf{0}_{1 \times 86} \end{bmatrix}^\top \quad (77)$$

4) ${}^a a_1$ constant: If ${}^a a_1$ is constant, d_{a11} , pitch and yaw of ${}^I \hat{\mathbf{R}}$ are unobservable with unobservable directions as:

$$\mathbf{N}_{a1} = \begin{bmatrix} \mathbf{0}_{12 \times 1} & \mathbf{0}_{12 \times 1} & \mathbf{0}_{12 \times 1} \\ \hat{\mathbf{D}}_a^{-1} \mathbf{e}_1 {}^a a_1 & \hat{\mathbf{D}}_a^{-1} \mathbf{e}_2 {}^a d_{a11} {}^a a_1 & \hat{\mathbf{D}}_a^{-1} \mathbf{e}_3 {}^a d_{a11} {}^a d_{a22} {}^a a_1 \\ \mathbf{0}_{6 \times 1} & \mathbf{0}_{6 \times 1} & \mathbf{0}_{6 \times 1} \\ 1 & 0 & 0 \\ 0 & \hat{d}_{a22} & 0 \\ 0 & -\hat{d}_{a12} & 0 \\ 0 & \hat{d}_{a23} & -\hat{d}_{a33} \hat{d}_{a22} \\ 0 & -\hat{d}_{a13} & \hat{d}_{a12} \hat{d}_{a33} \\ 0 & 0 & \hat{d}_{a13} \hat{d}_{a22} - \hat{d}_{a12} \hat{d}_{a23} \\ \mathbf{0}_{3 \times 1} & {}^I \hat{\mathbf{R}} \mathbf{e}_3 & {}^I \hat{\mathbf{R}} (\mathbf{e}_1 \hat{d}_{a12} + \mathbf{e}_2 \hat{d}_{a22}) \\ \mathbf{0}_{77 \times 1} & \mathbf{0}_{77 \times 1} & \mathbf{0}_{77 \times 1} \end{bmatrix}$$

5) ${}^a a_2$ constant: If ${}^a a_2$ is constant, d_{a12} , d_{a22} and roll of ${}^I \hat{\mathbf{R}}$ are unobservable with unobservable directions as:

$$\mathbf{N}_{a2} = \begin{bmatrix} \mathbf{0}_{12 \times 1} & \mathbf{0}_{12 \times 1} & \mathbf{0}_{12 \times 1} \\ \hat{\mathbf{D}}_a^{-1} \mathbf{e}_1 {}^a a_2 & \hat{\mathbf{D}}_a^{-1} \mathbf{e}_2 {}^a a_2 & \hat{\mathbf{D}}_a^{-1} \mathbf{e}_3 {}^a d_{a22} {}^a a_2 \\ \mathbf{0}_{6 \times 1} & \mathbf{0}_{6 \times 1} & \mathbf{0}_{6 \times 1} \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \hat{d}_{a33} \\ 0 & 0 & -\hat{d}_{a23} \\ \mathbf{0}_{3 \times 1} & \mathbf{0}_{3 \times 1} & {}^I \hat{\mathbf{R}} \mathbf{e}_1 \\ \mathbf{0}_{77 \times 1} & \mathbf{0}_{77 \times 1} & \mathbf{0}_{77 \times 1} \end{bmatrix} \quad (78)$$

6) ${}^a a_3$ constant: If ${}^a a_3$ is constant, d_{a13} , d_{a23} and d_{a33} are unobservable with unobservable directions as:

$$\mathbf{N}_{a3} = \begin{bmatrix} \mathbf{0}_{12 \times 12} & (\hat{\mathbf{D}}_a^{-1} \mathbf{e}_1)^\top {}^a a_3 & \mathbf{0}_{1 \times 9} & 1 & \mathbf{0}_{1 \times 82} \\ \mathbf{0}_{1 \times 12} & (\hat{\mathbf{D}}_a^{-1} \mathbf{e}_2)^\top {}^a a_3 & \mathbf{0}_{1 \times 10} & 1 & \mathbf{0}_{1 \times 81} \\ \mathbf{0}_{1 \times 12} & (\hat{\mathbf{D}}_a^{-1} \mathbf{e}_3)^\top {}^a a_3 & \mathbf{0}_{1 \times 11} & 1 & \mathbf{0}_{1 \times 80} \end{bmatrix}^\top \quad (79)$$

APPENDIX G EXTENDED SIMULATION RESULTS

The full set of estimated calibration parameters for general 3D motion for the model which estimates the velocity of the auxiliary IMUs is shown in Figure 11, while the inconsistent model which does not estimate the velocity of the auxiliary IMUs is able to perform similarly is shown in Figure 12. For the 1-axis degenerate motion case the results are also shown for the proposed model, see Figure 13, and the inconsistent model, see Figure 14. The sensor rig used in these experiments is a two camera system with two auxiliary IMUs, and without estimation of the IMU intrinsic parameters.

APPENDIX H EXTENDED REAL WORLD RESULTS

More results for VI-Rig datasets of the camera intrinsics and IMU intrinsics ${}_a^I \mathbf{R}$ are shown in Fig. 9 and Fig. 15.

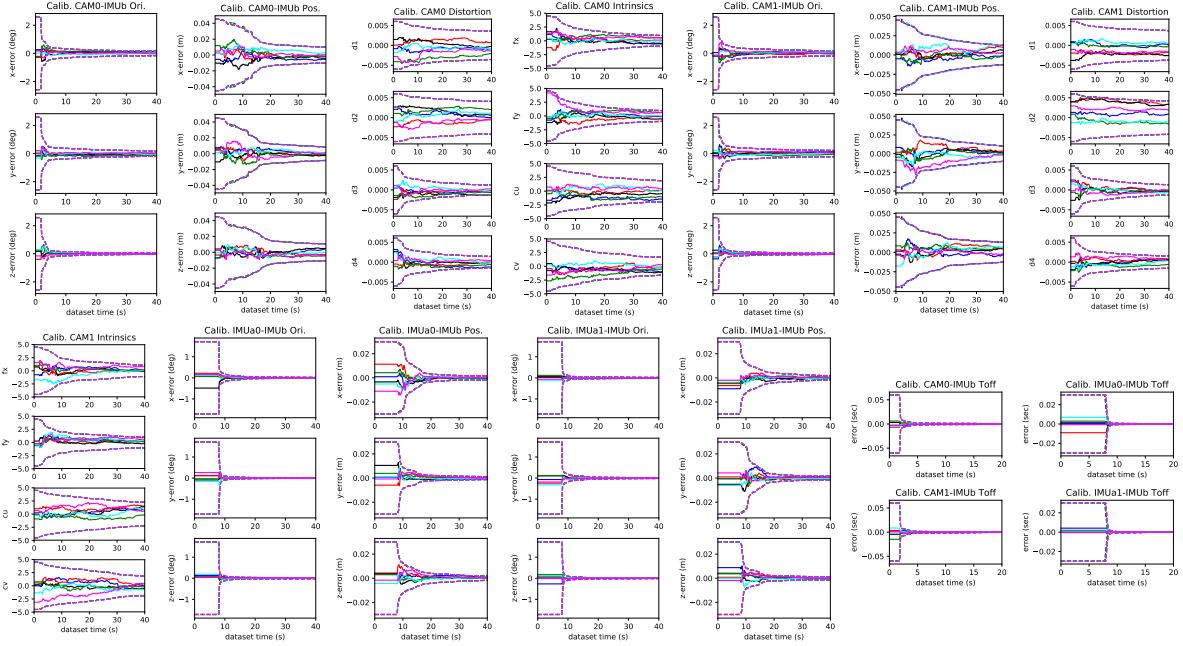


Fig. 11: Three sigma bounds and estimation errors for six different runs with different initial calibration perturbations and realization of the measurement noise for *fully excited motion*. The sensor suite contains (from top to bottom, left to right) a monocular camera 0 (1-4), monocular camera 1 (5-8), auxiliary IMU 0 (9-10), auxiliary IMU 1 (11-11), and temporal time offsets.

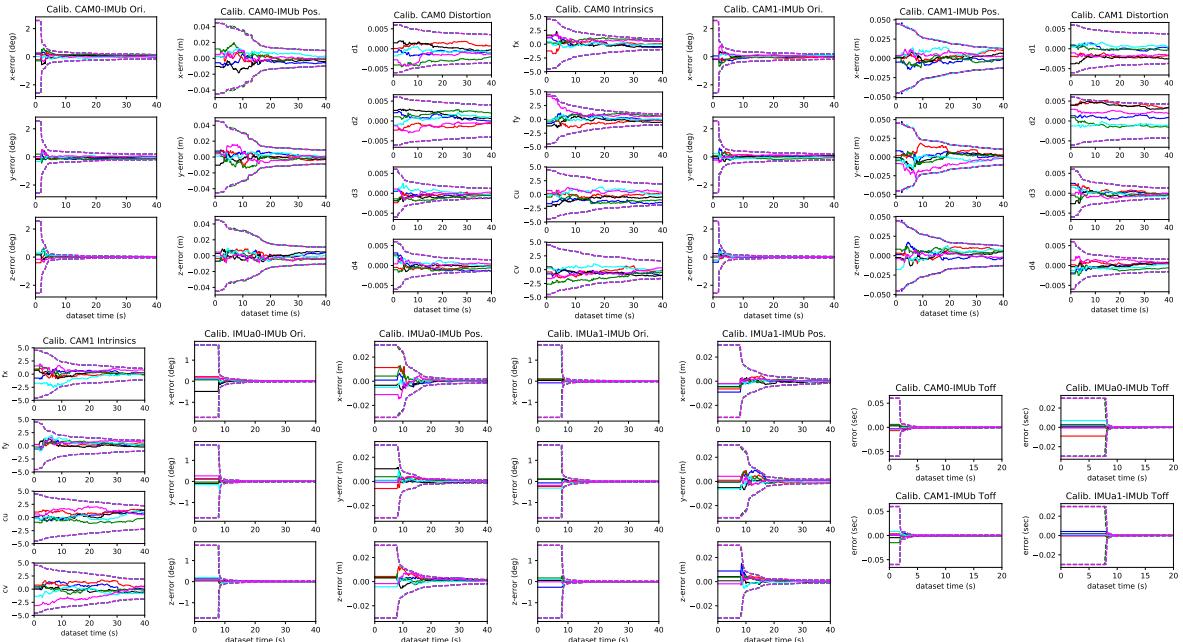


Fig. 12: Three sigma bounds (dashed line) and estimation errors (solid line) for six different runs (shown in different color) with different initial calibration perturbations and realization of the measurement noise for *fully excited motion* with the *inconsistent model* which does not estimate second velocity. The sensor suite contains (from top to bottom, left to right) a monocular camera 0 (1-4), monocular camera 1 (5-8), auxiliary IMU 0 (9-10), auxiliary IMU 1 (11-11), and temporal time offsets.

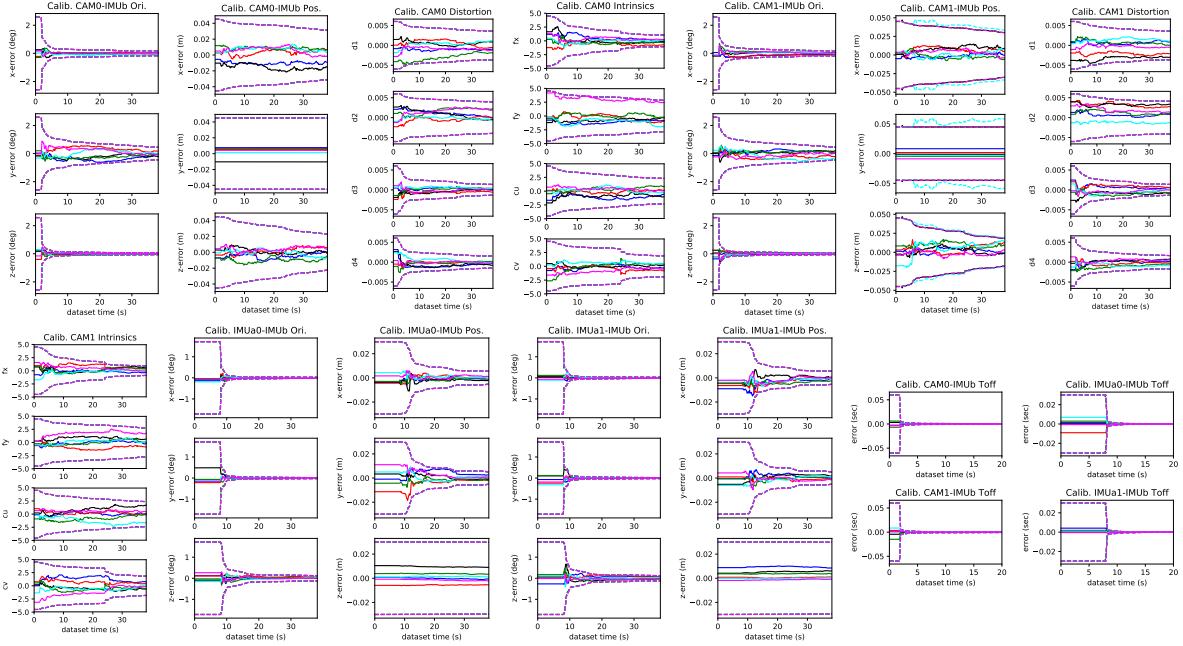


Fig. 13: Three sigma bounds and estimation errors for six different runs with different initial calibration perturbations and realization of the measurement noise for *1-axis degenerate motion*. The sensor suite contains (from top to bottom, left to right) a monocular camera 0 (1-4), monocular camera 1 (5-8), auxiliary IMU 0 (9-10), auxiliary IMU 1 (11-12), and temporal time offsets.

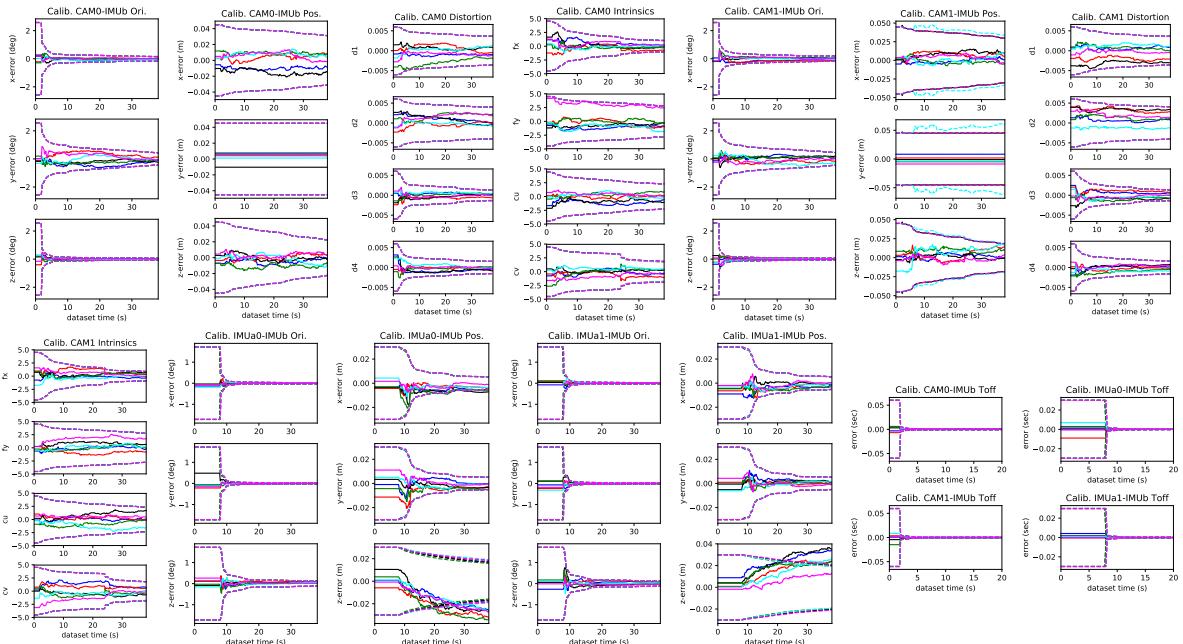


Fig. 14: Three sigma bounds (dashed line) and estimation errors (solid line) for six different runs (shown in different colors) with different initial calibration perturbations and realization of the measurement noise for *1-axis degenerate motion* with the *inconsistent model* which does not estimate second velocity. The sensor suite contains (from top to bottom, left to right) a monocular camera 0 (1-4), monocular camera 1 (5-8), auxiliary IMU 0 (9-10), auxiliary IMU 1 (11-12), and temporal time offsets.

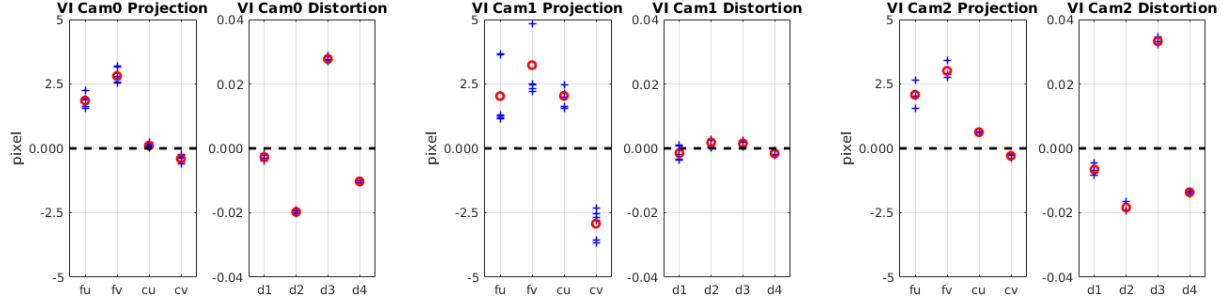


Fig. 15: The IMU intrinsics calibration results for VI-Rig and T265-Rig. Blue crosses represents the estimated values for each run. Red circle denotes the average value for all the runs. The black dotted line denotes the values we get from Kalibr. It can be seen that the our cam intrinsic calibration can converge to a value near that we get from Kalibr.

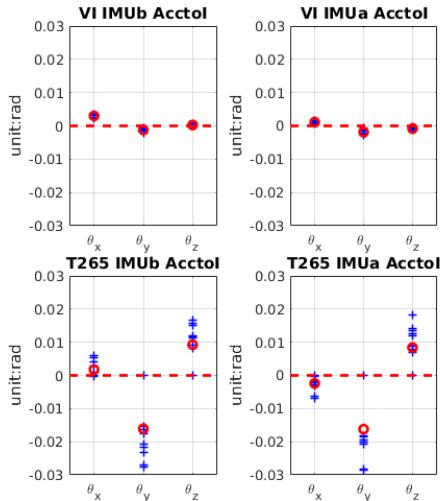


Fig. 16: The IMU intrinsics calibration results for VI-Rig and T265-Rig. Blue crosses represents the estimated values for each run. Red circle denotes the average value for all the runs. The red dotted line is the initial guess, which is also the ideal value we expected. The nearer the estimated value is to the red dotted line, the better the IMU is. To better demostrate the calibration results, we use the the same scale of y axis for both $B_a^I \mathbf{R}$ and $A_a^I \mathbf{R}$.