Multi-Visual-Inertial Sensor Calibration: Algorithm and Analysis

Yulin Yang - yuyang@udel.edu
Patrick Geneva - pgeneva@udel.edu
Guoquan Huang - ghuang@udel.edu

Department of Mechanical Engineering
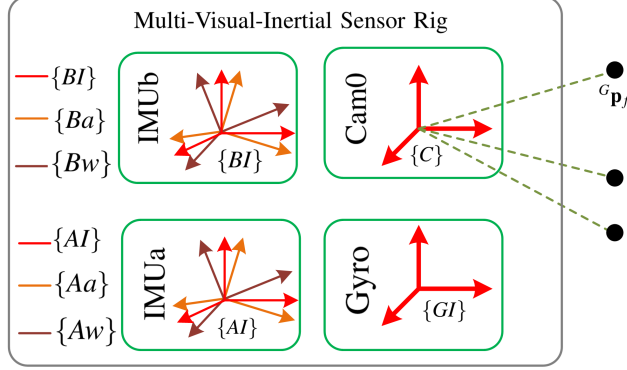University of Delaware, Delaware, USA

RPNG

Contents

Figure 1: The Multi-Visual-Inertial System (MVIS) sensor frames: base IMU (IMUb) sensor composed of accelerometer frame $\{Ba\}$ and gyroscope $\{Bw\}$, 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 $^G\mathbf{p}_f$ through its cameras. Note that we only visualize unique sensor types but the system can contain arbitrary amounts.

1 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.

1.1 IMU Intrinsic Model

Following [1], 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 [2, 1, 3], the raw angular velocity $^{Bw}\boldsymbol{\omega}_m$ from the gyroscope and linear acceleration $^{Ba}\mathbf{a}_m$ from accelerometer can be written as:

$$^{Bw}\boldsymbol{\omega}_m = \mathbf{T}_w{}^{Bw}_{BI}\mathbf{R}^{BI}\boldsymbol{\omega} + \mathbf{T}_g{}^{BI}\mathbf{a} + \mathbf{b}_g + \mathbf{n}_g \tag{1}$$

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

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 [2]. $^{Bw}_{BI}\mathbf{R}$ and $^{Ba}_{BI}\mathbf{R}$ denote the rotation from the gyroscope and acceleration frame to base "inertial" $\{BI\}$, respectively [1]. 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 $^{BI}\boldsymbol{\omega}$ and linear acceleration $^{BI}\mathbf{a}$ as:

$$^{BI}\boldsymbol{\omega} = {}^{BI}_{Bw}\mathbf{R}\mathbf{D}_w\left(^{Bw}\boldsymbol{\omega}_m - \mathbf{b}_g - \mathbf{n}_g - \mathbf{T}_g{}^{BI}\mathbf{a}\right) \tag{3}$$

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

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.

1.2 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 $\mathcal{X}_F$.

$$\mathbf{x} = \begin{bmatrix} \mathcal{X}_{BI}^\top & \mathcal{X}_{AI}^\top & \mathcal{X}_{GI}^\top & \mathcal{X}_{In}^\top & \mathcal{X}_{calib}^\top & \mathcal{X}_F^\top \end{bmatrix}^\top \tag{5}$$

$$\mathcal{X}_{BI} = \begin{bmatrix} \mathbf{x}_{BI_1}^\top & \cdots & \mathbf{x}_{BI_m}^\top \end{bmatrix}^\top \tag{6}$$

$$\mathcal{X}_{AI} = \begin{bmatrix} \mathbf{x}_{AI_1}^\top & \cdots & \mathbf{x}_{AI_m}^\top \end{bmatrix}^\top \tag{7}$$

$$\mathcal{X}_{GI} = \begin{bmatrix} \mathbf{x}_{GI_1}^\top & \cdots & \mathbf{x}_{GI_m}^\top \end{bmatrix}^\top \tag{8}$$

$$\mathcal{X}_{In} = \begin{bmatrix} \mathbf{x}_{Bn}^\top & \mathbf{x}_{An}^\top & \mathbf{x}_{Gn}^\top \end{bmatrix}^\top \tag{9}$$

$$\mathcal{X}_{calib} = \begin{bmatrix} \mathbf{x}_{BA}^\top & \mathbf{x}_{BG}^\top & \mathbf{x}_{BC}^\top \end{bmatrix}^\top \tag{10}$$

$$\mathcal{X}_F = \begin{bmatrix} {}^G\mathbf{p}_{f_1}^\top \cdots {}^G\mathbf{p}_{f_l}^\top \end{bmatrix}^\top \tag{11}$$

with:

$$\mathbf{x}_{BI} = \begin{bmatrix} {}^G_{BI}\bar{q}^\top & {}^G\mathbf{p}_{BI}^\top & {}^G\mathbf{v}_{BI}^\top & {}^{BI}\mathbf{b}_g^\top & {}^{BI}\mathbf{b}_a^\top \end{bmatrix}^\top \tag{12}$$

$$\mathbf{x}_{AI} = \begin{bmatrix} {}^G\mathbf{v}_{AI}^\top & {}^{AI}\mathbf{b}_g^\top & {}^{AI}\mathbf{b}_a^\top \end{bmatrix}^\top \tag{13}$$

$$\mathbf{x}_{GI} = {}^{GI}\mathbf{b}_g \tag{14}$$

$$\mathbf{x}_{Bn} = \begin{bmatrix} \mathbf{x}_{Bw}^\top & \mathbf{x}_{Ba}^\top & \mathbf{x}_{Bg}^\top & {}^{BI}_{Ba}\bar{q}^\top \end{bmatrix}^\top \tag{15}$$

$$\mathbf{x}_{An} = \begin{bmatrix} \mathbf{x}_{Aw}^\top & \mathbf{x}_{Aa}^\top & \mathbf{x}_{Ag}^\top & {}^{AI}_{Aa}\bar{q}^\top \end{bmatrix}^\top \tag{16}$$

$$\mathbf{x}_{Gn} = \mathbf{x}_{Gw} \tag{17}$$

$$\mathbf{x}_{BA} = \begin{bmatrix} {}^{BI}_{AI}\bar{q}^\top & {}^{BI}\mathbf{p}_{AI}^\top & t_{AB} \end{bmatrix}^\top \tag{18}$$

$$\mathbf{x}_{BG} = \begin{bmatrix} {}^{BI}_{GI}\bar{q}^\top & t_{GB} \end{bmatrix}^\top \tag{19}$$

$$\mathbf{x}_{BC} = \begin{bmatrix} {}^{BI}_{C}\bar{q}^\top & {}^{BI}\mathbf{p}_C^\top & t_{CB} & \mathbf{x}_{Cn}^\top \end{bmatrix}^\top \tag{20}$$

where $\mathbf{x}_{SI}$ denotes the inertial navigation state (S can be replaced by $B/A/G$ for base IMU, auxiliary IMU, or Gyro, respectively). ${}^G_{SI}\bar{q}$ is in Hamilton quaternion from [4] and denotes the rotation ${}^G_{SI}\mathbf{R}$ from the $\{SI\}$ sensor to global frame $\{G\}$. ${}^G\mathbf{p}_{SI}$ and ${}^G\mathbf{v}_{SI}$ represent the global position and velocity of $\{SI\}$ in $\{G\}$. ${}^{SI}\mathbf{b}_g$ and ${}^{SI}\mathbf{b}_a$ denote the gyro and accelerometer bias, respectively. $\mathbf{x}_{Sn}$ denotes the inertial intrinsic states. As in [1], $\mathbf{x}_{Sw}$, $\mathbf{x}_{Sa}$ 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 [5]. All the parameters that can be calibrated are summarized in Table 1.

1.3 Nonlinear Least Squares Optimization

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

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

Table 1: All the calibration parameters estimated in the proposed MVIS.

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

Then, we can formulate the NLS problem as:

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

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 [6]. In summary, we have the following NLS which is equivalent to maximum likelihood estimation (MLE):

$$\min_{\mathbf{x}} \sum \mathbb{C}_{BI} + \sum \mathbb{C}_{AI} + \sum \mathbb{C}_{GI} + \sum \mathbb{C}_{C} \tag{23}$$

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 [7] to solve the NLS optimization problem.

1.4 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 [5]) as the following:

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

$$\triangleq \mathbf{h}_C({}^{G}_{BI_{in}}\mathbf{R}, {}^{G}\mathbf{p}_{BI_{in}}, \mathbf{x}_{BC}, {}^{G}\mathbf{p}_f) + \mathbf{n}_C \tag{25}$$

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 OpenVINS [8]) 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 $\{{}^{G}_{BI_{in}}\mathbf{R}, {}^{G}\mathbf{p}_{BI_{in}}\}$ 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:

$$\mathbb{C}_C \triangleq \left\| \mathbf{z}_C - \mathbf{h}_C({}^{G}_{BI_{in}}\mathbf{R}, {}^{G}\mathbf{p}_{BI_{in}}, \mathbf{x}_{BC}, {}^{G}\mathbf{p}_f) \right\|^2_{\mathbf{Q}_C^{-1}} \tag{26}$$

2 Base Inertial Costs

2.1 Base IMU Cost without Intrinsics

The IMU cost function is modeled after ACI^2 [4] with $^G\mathbf{g} = [0\ 0\ 9.81]^\top$. The IMU dynamic model can thus be defined as [9, 10]:

$$^G_{BI}\dot{\bar{q}} = \frac{1}{2}\mathbf{\Omega}(^{BI}\boldsymbol{\omega})^G_{BI}\bar{q} \tag{27}$$

$$^G\dot{\mathbf{p}}_{BI} = {}^G\mathbf{v}_{BI} \tag{28}$$

$$^G\dot{\mathbf{v}}_{BI} = {}^G_{BI}\mathbf{R}^{BI}\mathbf{a} - {}^G\mathbf{g} \tag{29}$$

$$^{BI}\dot{\mathbf{b}}_g = \mathbf{n}_{wg} \tag{30}$$

$$^{BI}\dot{\mathbf{b}}_a = \mathbf{n}_{wa} \tag{31}$$

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

$$^{BI}\Delta\bar{q} \triangleq \mathbf{h}_q(\mathbf{x}_{BI_{k:k+1}}) = {}^{BI_k}_G\bar{q} \otimes {}^G_{BI_{k+1}}\bar{q}$$

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

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

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

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

$$^{BI}\Delta\mathbf{b}_g \triangleq {}^{BI}\mathbf{b}_{g_{k+1}} - {}^{BI}\mathbf{b}_{g_k}$$

$$^{BI}\Delta\mathbf{b}_a \triangleq {}^{BI}\mathbf{b}_{a_{k+1}} - {}^{BI}\mathbf{b}_{a_k}$$

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:

$$^{BI}\Delta\bar{q} \simeq {}^{BI}\Delta\hat{\bar{q}} \otimes \bar{q}(\mathbf{H}^q_{b_g}{}^{BI}\Delta\mathbf{b}_{g_k}) \tag{32}$$

$$^{BI}\Delta\mathbf{p} \simeq {}^{BI}\Delta\hat{\mathbf{p}} + \mathbf{H}^p_b{}^{BI}\Delta\mathbf{b}_k \tag{33}$$

$$^{BI}\Delta\mathbf{v} \simeq {}^{BI}\Delta\hat{\mathbf{v}} + \mathbf{H}^v_b{}^{BI}\Delta\mathbf{b}_k \tag{34}$$

where $\mathbf{b} = \begin{bmatrix} \mathbf{b}_g^\top & \mathbf{b}_a^\top \end{bmatrix}^\top$ and $^{BI}\Delta\mathbf{b}_k$ denotes the bias linearizaton changes [4]. Then, with these bias corrected pre-integrated measurements, we can construct:

$$\underbrace{\begin{bmatrix} \Delta\bar{q} \\ \Delta\mathbf{p} \\ \Delta\mathbf{v} \\ \Delta\mathbf{b} \end{bmatrix}}_{\mathbf{z}_{BI_{k+1}}} \simeq \underbrace{\begin{bmatrix} \mathbf{h}_q(\mathbf{x}_{BI_{k:k+1}}) \otimes \bar{q}^{-1}(\mathbf{H}^q_{b_g}{}^{BI}\tilde{\mathbf{b}}_{g_k}) \\ \mathbf{h}_p(\mathbf{x}_{BI_{k:k+1}}) - \mathbf{H}^p_b{}^{BI}\tilde{\mathbf{b}}_k \\ \mathbf{h}_v(\mathbf{x}_{BI_{k:k+1}}) - \mathbf{H}^v_b{}^{BI}\tilde{\mathbf{b}}_k \\ {}^{BI}\mathbf{b}_{k+1} - {}^{BI}\mathbf{b}_k \end{bmatrix}}_{\mathbf{h}_I(\mathbf{x}_{BI_{k:k+1}})} + \mathbf{n}_{BI_{k+1}} \tag{35}$$

where $\mathbf{n}_{BI_{k+1}} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{BI_{k+1}})$ represents the integrated IMU noise [4], $^{BI}\tilde{\mathbf{b}}_k$ 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 $\mathbf{H}^z_y$

represents the Jacobians of state z to variable y. Finally, the corresponding base IMU cost function can be written as:

$$\mathbb{C}_{BI_{k+1}} \triangleq \left\| \mathbf{z}_{BI_{k+1}} - \mathbf{h}_I(\mathbf{x}_{BI_{k:k+1}}) \right\|_{\mathbf{Q}_{BI_{k+1}}^{-1}}^2 \tag{36}$$

2.2 Base IMU Cost with Intrinsics

When incorporating IMU intrinsics, Xiao et al, [11] 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 intrinsics, we propose to relinearize bias and IMU intrinsics. For each iteration, the IMU measurements need to be re-integrated. Hence, the bias correction, Eq. (33), (32) and (34), will not be needed and the IMU measurements following ACI² can be constructed as:

$$\underbrace{\begin{bmatrix} \Delta\bar{q} \\ \Delta\mathbf{p} \\ \Delta\mathbf{v} \\ \Delta\mathbf{b} \end{bmatrix}}_{\mathbf{z}_{BI_{k+1}}} \simeq \underbrace{\begin{bmatrix} \mathbf{h}_q(\mathbf{x}_{BI_{k:k+1}}) \otimes \bar{q}^{-1}(\mathbf{H}_{b_g}^{q\ BI}\tilde{\mathbf{b}}_{g_k}) \otimes \bar{q}^{-1}(\mathbf{H}_{Bn}^q\tilde{\mathbf{x}}_{Bn}) \\ \mathbf{h}_p(\mathbf{x}_{BI_{k:k+1}}) - \mathbf{H}_b^{pBI}\tilde{\mathbf{b}}_k - \mathbf{H}_{Bn}^p\tilde{\mathbf{x}}_{Bn} \\ \mathbf{h}_v(\mathbf{x}_{BI_{k:k+1}}) - \mathbf{H}_b^{vBI}\tilde{\mathbf{b}}_k - \mathbf{H}_{Bn}^v\tilde{\mathbf{x}}_{Bn} \\ {}^{BI}\mathbf{b}_{k+1} - {}^{BI}\mathbf{b}_k \end{bmatrix}}_{\mathbf{h}_I(\mathbf{x}_{BI_{k:k+1}}, \mathbf{x}_{Bn})} + \mathbf{n}_{BI_{k+1}} \tag{37}$$

Then, the base IMU cost with intrinsics can be formulated similar to Eq. (36).

3 Auxiliary Inertial Costs

3.1 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:

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

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

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

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:

$$ {}^G_{AI}\bar{q} = {}^G_{BI_{in}}\bar{q} \otimes {}^{BI}_{AI}\bar{q} \tag{38} $$

$$ {}^G\mathbf{p}_{AI} = {}^G\mathbf{p}_{BI_{in}} + {}^G_{BI_{in}}\mathbf{R}^{BI}\mathbf{p}_{AI} \tag{39} $$

It is important to note, that we make this constaint 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. (38) and (39), 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. (35) and (37), the auxiliary IMU cost functions $\mathbb{C}_{AI}$ with or without IMU intrinsics $\mathbf{x}_{An}$ can be defined.

3.2 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:

$$^{G}\mathbf{v}_{AI} = {}^{G}\mathbf{v}_{BI} + {}^{G}_{BI}\mathbf{R} \lfloor {}^{BI}_{AI}\mathbf{R}^{AI}\boldsymbol{\omega} \rfloor {}^{BI}\mathbf{p}_{AI} \tag{40}$$

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

$$^{AI}\mathbf{H}^{v}_{b_g}{}^{AI}\hat{\mathbf{b}}_{g_k} = \mathbf{log}\left({}^{AI}\Delta\mathbf{R}^{\top G}_{AI_k}\mathbf{R}^{\top G}_{AI_{k+1}}\mathbf{R} \right) \tag{41}$$

where $\mathbf{log}\left(\cdot\right)$ denotes the *log* operation of SO(3) [6]. 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:

$$^{AI}\mathbf{H}^{v}_{b_a}{}^{AI}\hat{\mathbf{b}}_{a_k} = \mathbf{h}^{(AI)}_{v}(\hat{\mathbf{x}}_{BI_{k:k+1}}, \hat{\mathbf{x}}_{BA}) - {}^{AI}\Delta\mathbf{v} \tag{42}$$

$$^{AI}\mathbf{H}^{p}_{b_a}{}^{AI}\hat{\mathbf{b}}_{a_k} = \mathbf{h}^{(AI)}_{p}(\hat{\mathbf{x}}_{BI_{k:k+1}}, \hat{\mathbf{x}}_{BA}) - {}^{AI}\Delta\mathbf{p} \tag{43}$$

3.3 Auxiliary Gyroscope Cost

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

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

We can relate this integration to the base IMU orientation:

$$^{G}_{GI}\bar{q} = {}^{G}_{BI_{in}}\bar{q} \otimes {}^{BI}_{GI}\bar{q} \tag{44}$$

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:

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

Similar to the base IMU case, we can then formulate the Gyro cost $\mathbb{C}_{GI}$ (e.g., Eq. (35) and (37)) for without and with $\mathbf{x}_{Gn}$ intrinsics.

3.4 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 [12], VINS-MONO [13], CPI [14]). 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 6.3.

4 Observability Analysis

Observability analysis plays an important role in state estimation for VINS [15, 16]. 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 [17, 18, 19, 1]. 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.

4.1 Reduced State Vector

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 [20, 21]. To simplify the analysis, we reorganize the state vector and assume the base IMU, auxiliary IMU and auxiliary Gryo 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:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_B^\top & \mathbf{x}_A^\top & \mathbf{x}_G^\top & \mathbf{x}_{calib}^\top & \mathbf{x}_F^\top \end{bmatrix}^\top \tag{45}$$

where:

$$\mathbf{x}_B = \begin{bmatrix} \mathbf{x}_{BI}^\top & \mathbf{x}_{Bn}^\top \end{bmatrix}^\top \tag{46}$$

$$\mathbf{x}_A = \begin{bmatrix} \mathbf{x}_{AI}^\top & \mathbf{x}_{An}^\top \end{bmatrix}^\top \tag{47}$$

$$\mathbf{x}_G = \begin{bmatrix} \mathbf{x}_{GI}^\top & \mathbf{x}_{Gn}^\top \end{bmatrix}^\top \tag{48}$$

$$\mathbf{x}_{calib} = \begin{bmatrix} \mathbf{x}_{BA}^\top & \mathbf{x}_{BG}^\top & \mathbf{x}_{BC}^\top \end{bmatrix}^\top \tag{49}$$

$$\mathbf{x}_F = {}^G\mathbf{p}_f \tag{50}$$

Note that the auxiliary IMU and Gyro states are:

$$\mathbf{x}_{AI} = \begin{bmatrix} {}^G_{AI}\bar{q}^\top & {}^G\mathbf{p}_{AI}^\top & {}^G\mathbf{v}_{AI}^\top & {}^{AI}\mathbf{b}_g^\top & {}^{AI}\mathbf{b}_a^\top \end{bmatrix}^\top \tag{51}$$

$$\mathbf{x}_{GI} = \begin{bmatrix} {}^G_{GI}\bar{q}^\top & {}^{GI}\mathbf{b}_g^\top \end{bmatrix}^\top \tag{52}$$

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

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_C^\top & \mathbf{z}_A^\top & \mathbf{z}_G^\top \end{bmatrix}^\top \tag{53}$$

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:

$$\mathbf{z}_C = \begin{bmatrix} \frac{C_{p_x}}{C_{p_z}} & \frac{C_{p_y}}{C_{p_z}} \end{bmatrix}^\top, \ ^C\mathbf{p}_f = \begin{bmatrix} ^C p_x & ^C p_y & ^C p_z \end{bmatrix}^\top \tag{54}$$

$$^C\mathbf{p}_f = {}_{BI}^C\mathbf{R} {}_G^{BI}\mathbf{R} \left({}^G\mathbf{p}_f - {}^G\mathbf{p}_{BI} - {}_{BI}^G\mathbf{R}^{BI}\mathbf{p}_C \right) \tag{55}$$

$$\mathbf{z}_A = \begin{bmatrix} \log \left({}_{AI}^G\mathbf{R}^\top {}_{BI}^G\mathbf{R} {}_{AI}^{BI}\mathbf{R} \right) \\ {}^G\mathbf{p}_{AI} - {}^G\mathbf{p}_{BI} - {}_{BI}^G\mathbf{R}^{BI}\mathbf{p}_{AI} \end{bmatrix} \tag{56}$$

$$\mathbf{z}_G = \log \left({}_{GI}^G\mathbf{R}^\top {}_{BI}^G\mathbf{R} {}_{GI}^{BI}\mathbf{R} \right) \tag{57}$$

4.2   State Transition Matrix

The state transition matrix of the reduced state vector can be written as:

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_B & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}_A & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi}_G & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi}_{calib} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi}_F \end{bmatrix} \tag{58}$$

where we have:

$$\boldsymbol{\Phi}_B = \begin{bmatrix} \boldsymbol{\Phi}_{BI} & \boldsymbol{\Phi}_{Bn} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{59}$$

$$\boldsymbol{\Phi}_A = \begin{bmatrix} \boldsymbol{\Phi}_{AI} & \boldsymbol{\Phi}_{An} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{60}$$

$$\boldsymbol{\Phi}_G = \begin{bmatrix} \boldsymbol{\Phi}_{GI} & \boldsymbol{\Phi}_{Gn} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{61}$$

$$\boldsymbol{\Phi}_{calib} = \mathbf{I} \tag{62}$$

$$\boldsymbol{\Phi}_F = \mathbf{I} \tag{63}$$

The state transition of $\boldsymbol{\Phi}_B$ and $\boldsymbol{\Phi}_A$ have the same structure. We can grab the gyro part of the state transition of $\boldsymbol{\Phi}_B$ to get $\boldsymbol{\Phi}_G$. Hence, we only show $\boldsymbol{\Phi}_B$ for clarity.

$$\boldsymbol{\Phi}_{BI} = \begin{bmatrix} \boldsymbol{\Phi}_{B11} & \mathbf{0}_3 & \mathbf{0}_3 & \boldsymbol{\Phi}_{B14} & \boldsymbol{\Phi}_{B15} \\ \boldsymbol{\Phi}_{B21} & \mathbf{I}_3 & \mathbf{I}_3\delta t & \boldsymbol{\Phi}_{B24} & \boldsymbol{\Phi}_{B25} \\ \boldsymbol{\Phi}_{B31} & \mathbf{0}_3 & \mathbf{I}_3 & \boldsymbol{\Phi}_{B34} & \boldsymbol{\Phi}_{B35} \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \tag{64}$$

where we have:

$$\mathbf{\Phi}_{B11} = {}^{BI_{k+1}}_{BI_k}\hat{\mathbf{R}}$$

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

$$\mathbf{\Phi}_{B31} = -\lfloor {}^{G}\hat{\mathbf{v}}_{BI_{k+1}} - {}^{G}\hat{\mathbf{v}}_{BI_k} + {}^{G}\mathbf{g}\delta t_k \rfloor {}^{G}_{BI_k}\hat{\mathbf{R}}$$

$$\mathbf{\Phi}_{B14} = -\mathbf{J}_r\delta t_k {}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}$$

$$\mathbf{\Phi}_{B24} = {}^{G}_{BI_k}\hat{\mathbf{R}}\mathbf{\Xi}_4{}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}$$

$$\mathbf{\Phi}_{B34} = {}^{G}_{BI_k}\hat{\mathbf{R}}\mathbf{\Xi}_3{}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}$$

$$\mathbf{\Phi}_{B15} = -\mathbf{J}_r\delta t_k {}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}\hat{\mathbf{T}}_{Bg}{}^{BI}_{Ba}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Ba}$$

$$\mathbf{\Phi}_{B25} = -\mathbf{J}_r\delta t_k {}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}\hat{\mathbf{T}}_{Bg}{}^{BI}_{Ba}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Ba}$$

$$\mathbf{\Phi}_{Bn} = \begin{bmatrix} \mathbf{\Phi}_{Bn11} & \mathbf{\Phi}_{Bn12} & \mathbf{\Phi}_{Bn13} & \mathbf{\Phi}_{Bn14} \\ \mathbf{\Phi}_{Bn21} & \mathbf{\Phi}_{Bn22} & \mathbf{\Phi}_{Bn23} & \mathbf{\Phi}_{Bn24} \\ \mathbf{\Phi}_{Bn31} & \mathbf{\Phi}_{Bn23} & \mathbf{\Phi}_{Bn33} & \mathbf{\Phi}_{Bn34} \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \tag{65}$$

where we have:

$$\mathbf{\Phi}_{Bn11} = \mathbf{J}_r\delta t_k {}^{BI}_{Bw}\hat{\mathbf{R}}\mathbf{H}_{D_{Bw}}$$

$$\mathbf{\Phi}_{Bn21} = -{}^{G}_{BI_k}\mathbf{\Xi}_4{}^{BI}_{Bw}\hat{\mathbf{R}}\mathbf{H}_{D_{Bw}}$$

$$\mathbf{\Phi}_{Bn31} = -{}^{G}_{BI_k}\mathbf{\Xi}_3{}^{BI}_{Bw}\hat{\mathbf{R}}\mathbf{H}_{D_{Bw}}$$

$$\mathbf{\Phi}_{Bn12} = -\mathbf{J}_r\delta t_k {}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}\hat{\mathbf{T}}_{Bg}{}^{BI}_{Ba}\hat{\mathbf{R}}\mathbf{H}_{D_{Ba}}$$

$$\mathbf{\Phi}_{Bn22} = {}^{G}_{BI_k}\hat{\mathbf{R}}\left(\mathbf{\Xi}_2 + \mathbf{\Xi}_4{}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}\hat{\mathbf{T}}_{Bg}\right){}^{BI}_{Ba}\hat{\mathbf{R}}\mathbf{H}_{D_{Ba}}$$

$$\mathbf{\Phi}_{Bn32} = {}^{G}_{BI_k}\hat{\mathbf{R}}\left(\mathbf{\Xi}_1 + \mathbf{\Xi}_3{}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}\hat{\mathbf{T}}_{Bg}\right){}^{BI}_{Ba}\hat{\mathbf{R}}\mathbf{H}_{D_{Ba}}$$

$$\mathbf{\Phi}_{Bn13} = \mathbf{J}_r\delta t_k {}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}\hat{\mathbf{T}}_{Bg}\lfloor {}^{BI}\hat{\mathbf{a}}\rfloor {}^{BI}_{Ba}\hat{\mathbf{R}}$$

$$\mathbf{\Phi}_{Bn23} = -{}^{G}_{BI_k}\hat{\mathbf{R}}\left(\mathbf{\Xi}_2 + \mathbf{\Xi}_4{}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}\hat{\mathbf{T}}_{Bg}\right)\lfloor {}^{BI}\hat{\mathbf{a}}\rfloor {}^{BI}_{Ba}\hat{\mathbf{R}}$$

$$\mathbf{\Phi}_{Bn33} = -{}^{G}_{BI_k}\hat{\mathbf{R}}\left(\mathbf{\Xi}_1 + \mathbf{\Xi}_3{}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}\hat{\mathbf{T}}_{Bg}\right)\lfloor {}^{BI}\hat{\mathbf{a}}\rfloor {}^{BI}_{Ba}\hat{\mathbf{R}}$$

$$\mathbf{\Phi}_{Bn14} = -\mathbf{J}_r\delta t_k {}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}\mathbf{H}_{T_{Bg}}$$

$$\mathbf{\Phi}_{Bn24} = {}^{G}_{BI_k}\hat{\mathbf{R}}\mathbf{\Xi}_4{}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}\mathbf{H}_{T_{Bg}}$$

$$\mathbf{\Phi}_{Bn34} = {}^{G}_{BI_k}\hat{\mathbf{R}}\mathbf{\Xi}_3{}^{BI}_{Bw}\hat{\mathbf{R}}\hat{\mathbf{D}}_{Bw}\mathbf{H}_{T_{Bg}}$$

4.3 Measurement Jacobians

The corresponding Jacobians can be listed as:

$$\frac{\partial\tilde{\mathbf{z}}}{\partial\tilde{\mathbf{x}}} = \begin{bmatrix} \frac{\partial\tilde{\mathbf{z}}_C}{\partial\tilde{\mathbf{x}}} \\ \frac{\partial\tilde{\mathbf{z}}_A}{\partial\tilde{\mathbf{x}}} \\ \frac{\partial\tilde{\mathbf{z}}_G}{\partial\tilde{\mathbf{x}}} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{CB} & \mathbf{0} & \mathbf{0} & \mathbf{H}_{CC} & \mathbf{H}_{CF} \\ \mathbf{H}_{AB} & \mathbf{H}_{AA} & \mathbf{0} & \mathbf{H}_{AC} & \mathbf{0} \\ \mathbf{H}_{GB} & \mathbf{0} & \mathbf{H}_{GG} & \mathbf{H}_{GC} & \mathbf{0} \end{bmatrix} \tag{66}$$

where:

$$\mathbf{H}_{CB} = \mathbf{H}_{proj}{}_C^{BI}\hat{\mathbf{R}}^{\top}{}_{BI}^{G}\hat{\mathbf{R}}^{\top} \times$$
$$\begin{bmatrix} \lfloor{}^G\mathbf{p}_f - {}^G\mathbf{p}_{BI}\rfloor_{BI}^{G}\hat{\mathbf{R}} & -\mathbf{I}_3 & \mathbf{0}_{3\times 30} \end{bmatrix}$$

$$\mathbf{H}_{CC} = \mathbf{H}_{proj}{}_C^{BI}\hat{\mathbf{R}}^{\top}{}_{BI}^{G}\hat{\mathbf{R}}^{\top}$$
$$\begin{bmatrix} \mathbf{0}_{3\times 7} & \mathbf{0}_{3\times 4} & \mathbf{H}_{CC3} & -{}_{BI}^{G}\hat{\mathbf{R}} & \mathbf{H}_{CC5} & \mathbf{0}_{3\times 8} \end{bmatrix}$$

$$\mathbf{H}_{CC3} = \lfloor {}^G\mathbf{p}_f - {}^G\mathbf{p}_{BI} - {}_{BI}^{G}\mathbf{R}^{BI}\hat{\mathbf{p}}_C \rfloor$$

$$\mathbf{H}_{CC5} = {}^G\mathbf{v}_{BI} - \lfloor {}^G\mathbf{p}_f - {}^G\mathbf{p}_{BI}\rfloor_{BI}^{G}\hat{\mathbf{R}}^{BI}\boldsymbol{\omega}$$

$$\mathbf{H}_{CF} = {}_C^{BI}\hat{\mathbf{R}}^{\top}{}_{BI}^{G}\hat{\mathbf{R}}$$

$$\mathbf{H}_{AB} = \begin{bmatrix} {}_{BI}^{AI}\hat{\mathbf{R}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times 21} \\ {}_{BI}^{G}\hat{\mathbf{R}}\lfloor {}^{BI}\hat{\mathbf{p}}_{AI}\rfloor & -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times 21} \end{bmatrix}$$

$$\mathbf{H}_{AA} = \begin{bmatrix} -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times 21} \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times 21} \end{bmatrix}$$

$$\mathbf{H}_{AC} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 & -{}^{AI}\boldsymbol{\omega} & \mathbf{0}_{3\times 19} \\ \mathbf{0}_3 & -{}_{BI}^{G}\hat{\mathbf{R}} & {}^G\mathbf{v}_{BI} + {}_{BI}^{G}\hat{\mathbf{R}}\lfloor {}^{BI}\boldsymbol{\omega}\rfloor^{BI}\mathbf{p}_{AI} & \mathbf{0}_{3\times 19} \end{bmatrix}$$

$$\mathbf{H}_{GB} = \begin{bmatrix} {}_{BI}^{GI}\hat{\mathbf{R}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times 6} \end{bmatrix}$$

$$\mathbf{H}_{GG} = \begin{bmatrix} -\mathbf{I}_3 & \mathbf{0}_3 \end{bmatrix}$$

$$\mathbf{H}_{GC} = \begin{bmatrix} \mathbf{0}_{3\times 7} & \mathbf{I}_3 & -{}^{GI}\hat{\boldsymbol{\omega}} & \mathbf{0}_{3\times 7} & \mathbf{0}_{3\times 8} \end{bmatrix}$$

4.4 Observability Matrix

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

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{CB} & \mathbf{0} & \mathbf{0} & \mathbf{M}_{CC} & \mathbf{M}_{CF} \\ \mathbf{M}_{AB} & \mathbf{M}_{AA} & \mathbf{0} & \mathbf{M}_{AC} & \mathbf{0} \\ \mathbf{M}_{GB} & \mathbf{0} & \mathbf{M}_{GG} & \mathbf{M}_{GC} & \mathbf{0} \end{bmatrix} \tag{67}$$

where $\mathbf{M}_{**}$ is as follows:

$$\mathbf{M} = \begin{bmatrix} \frac{\partial \tilde{\mathbf{z}}_C}{\partial \tilde{\mathbf{x}}} \\ \frac{\partial \tilde{\mathbf{z}}_A}{\partial \tilde{\mathbf{x}}} \\ \frac{\partial \tilde{\mathbf{z}}_G}{\partial \tilde{\mathbf{x}}} \end{bmatrix} \boldsymbol{\Phi} \tag{68}$$

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

$$= \begin{bmatrix} \mathbf{H}_{CB}\boldsymbol{\Phi}_B & \mathbf{0} & \mathbf{0} & \mathbf{H}_{CC}\boldsymbol{\Phi}_C & \mathbf{H}_{CF}\boldsymbol{\Phi}_F \\ \mathbf{H}_{AB}\boldsymbol{\Phi}_B & \mathbf{H}_{AA}\boldsymbol{\Phi}_A & \mathbf{0} & \mathbf{H}_{AC}\boldsymbol{\Phi}_C & \mathbf{0} \\ \mathbf{H}_{GB}\boldsymbol{\Phi}_B & \mathbf{0} & \mathbf{H}_{GG}\boldsymbol{\Phi}_G & \mathbf{H}_{GC}\boldsymbol{\Phi}_C & \mathbf{0} \end{bmatrix}$$

By analyzing the observability matrix, we can have the following Lemma:

Lemma 1. *The proposed MVIS still keeps 4 unobservable directions as Eq. (69), which are related*

to global yaw rotation and global translation.

$$\mathbf{N} = \begin{bmatrix} {}^{BI0}_{G}\mathbf{R}^{G}\mathbf{g} & \mathbf{0}_3 \\ \lfloor {}^{G}\mathbf{p}_{BI0} \rfloor {}^{G}\mathbf{g} & \mathbf{I}_3 \\ \lfloor {}^{G}\mathbf{v}_{BI0} \rfloor {}^{G}\mathbf{g} & \mathbf{0}_3 \\ \mathbf{0}_{6\times1} & \mathbf{0}_{6\times3} \\ \mathbf{0}_{21\times1} & \mathbf{0}_{21\times3} \\ {}^{AI0}_{G}\mathbf{R}^{G}\mathbf{g} & \mathbf{0}_3 \\ \lfloor {}^{G}\mathbf{p}_{AI0} - {}^{G}\mathbf{p}_{BI0} \rfloor {}^{G}\mathbf{g} & \mathbf{I}_3 \\ \lfloor {}^{G}\mathbf{v}_{AI0} - {}^{G}\mathbf{v}_{BI0} \rfloor {}^{G}\mathbf{g} & \mathbf{0}_3 \\ \mathbf{0}_{6\times1} & \mathbf{0}_{6\times3} \\ \mathbf{0}_{21\times1} & \mathbf{0}_{21\times3} \\ {}^{GI0}_{G}\mathbf{R}^{G}\mathbf{g} & \mathbf{0}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_3 \\ \mathbf{0}_{6\times1} & \mathbf{0}_{6\times3} \\ \mathbf{0}_{26\times1} & \mathbf{0}_{26\times3} \\ \lfloor {}^{G}\mathbf{p}_{f} \rfloor {}^{G}\mathbf{g} & \mathbf{I}_3 \end{bmatrix} \tag{69}$$

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 6.1).

5 Degenerate Motion Analysis

The degenerate motions for spatial-temporal parameters between IMU and camera has been studied in [17]. Therefore, in this paper, we study the degenerate motion for spatial-temporal and intrinsic parameters of auxiliary inertial sensors.

5.1 Spatial-Temporal of Auxiliary Inertial Sensors

Table 2: 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

We identified two degenerate motions for the spatial-temporal calibration between auxiliary inertial sensors and base IMU (see Tab. 2).

5.1.1 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:

$$
\mathbf{N}_{NoRot} = \begin{bmatrix}
\mathbf{0}_{36\times3} & \mathbf{0}_{36\times3} & \mathbf{0}_{36\times1} \\
\mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times1} \\
\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times1} \\
\mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times1} \\
\mathbf{0}_{6\times3} & \mathbf{0}_{6\times3} & \mathbf{0}_{6\times1} \\
\mathbf{0}_{21\times3} & \mathbf{0}_{21\times3} & \mathbf{0}_{21\times1} \\
\mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_{3\times1} \\
\mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times1} \\
\mathbf{0}_{6\times3} & \mathbf{0}_{6\times3} & \mathbf{0}_{6\times1} \\
\mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times1} \\
{}_{G}^{BI0}\mathbf{R} & \mathbf{0}_3 & \mathbf{0}_{3\times1} \\
\mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & 0 \\
\mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_{3\times1} \\
\mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & 1 \\
\mathbf{0}_{15\times3} & \mathbf{0}_{15\times3} & \mathbf{0}_{15\times1} \\
\mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_{3\times1}
\end{bmatrix} \tag{70}
$$

5.1.2 One-Axis Rotation

If the system undergoes 3D motion but with only one-axis rotation (similar to micro areal vehicle (MAV)), the translation between auxiliary IMU and base IMU, the rotation between auxiliary Gyro and base IMU will be unobservable along the rotation axis $\mathbf{k}$. The unobservable directions can be described as:

$$
\mathbf{N}_{1axis} = \begin{bmatrix}
\mathbf{0}_{36\times1} & \mathbf{0}_{36\times1} \\
\mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\
{}_{BI0}^{G}\mathbf{R}^{BI0}\mathbf{k} & \mathbf{0}_{3\times1} \\
\mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\
\mathbf{0}_{6\times1} & \mathbf{0}_{6\times1} \\
\mathbf{0}_{21\times1} & \mathbf{0}_{21\times1} \\
\mathbf{0}_{3\times1} & {}^{GI0}\mathbf{k} \\
\mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\
\mathbf{0}_{6\times1} & \mathbf{0}_{6\times1} \\
\mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\
{}^{BI0}\mathbf{k} & \mathbf{0}_{3\times1} \\
0 & 0 \\
\mathbf{0}_{3\times1} & {}^{GI0}\mathbf{k} \\
0 & 0 \\
\mathbf{0}_{15\times1} & \mathbf{0}_{15\times1} \\
\mathbf{0}_{3\times1} & \mathbf{0}_{3\times1}
\end{bmatrix} \tag{71}
$$

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 6.2.

Table 3: 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 $^{Bw}\omega_1$	1	d_{Bw11}
constant $^{Bw}\omega_2$	2	d_{Bw12}, d_{Bw22}
constant $^{Bw}\omega_3$	3	$d_{Bw13}, d_{Bw23}, d_{Bw33}$
constant $^{Ba}a_1$	3	d_{Ba11}, pitch and yaw of $^{BI}_{Ba}\mathbf{R}$
constant $^{Ba}a_2$	3	d_{Ba12}, d_{Ba22}, roll of $^{BI}_{Ba}\mathbf{R}$
constant $^{Ba}a_3$	3	$d_{Ba13}, d_{Ba23}, d_{Ba33}$
constant $^{Aw}\omega_1$	1	d_{Aw11}
constant $^{Aw}\omega_2$	2	d_{Aw12}, d_{Aw22}
constant $^{Aw}\omega_3$	3	$d_{Aw13}, d_{Aw23}, d_{Aw33}$
constant $^{Aa}a_1$	3	d_{Aa11}, pitch and yaw of $^{AI}_{Aa}\mathbf{R}$
constant $^{Aa}a_2$	3	d_{Aa12}, d_{Aa22}, roll of $^{AI}_{Aa}\mathbf{R}$
constant $^{Aa}a_3$	3	$d_{Aa13}, d_{Aa23}, d_{Aa33}$
constant $^{Gw}\omega_1$	1	d_{Gw11}
constant $^{Gw}\omega_2$	2	d_{Gw12}, d_{Gw22}
constant $^{Gw}\omega_3$	3	$d_{Gw13}, d_{Gw23}, d_{Gw33}$

Table 4: 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. IinC (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. AItoBI (rad)	0.003	Pos. AIinBI (m)	0.005
Cam Freq. (hz)	10 / 10 /10	IMU Freq. (hz)	250 / 300 / 200

5.2   Intrinsics for Base and Auxiliary Inertial Sensors

Unlike the work in [1] 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 [1] still hold for IMU intrinsics even for our inertial model choice. The degenerate motions for base IMU, auxiliary IMU, and auxiliary Gryo are summarized in Tab. 3. The detailed unobservable directions can be found in Appendix C.

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 the gravity sensitivity is observable in the majority of the investigated motions.
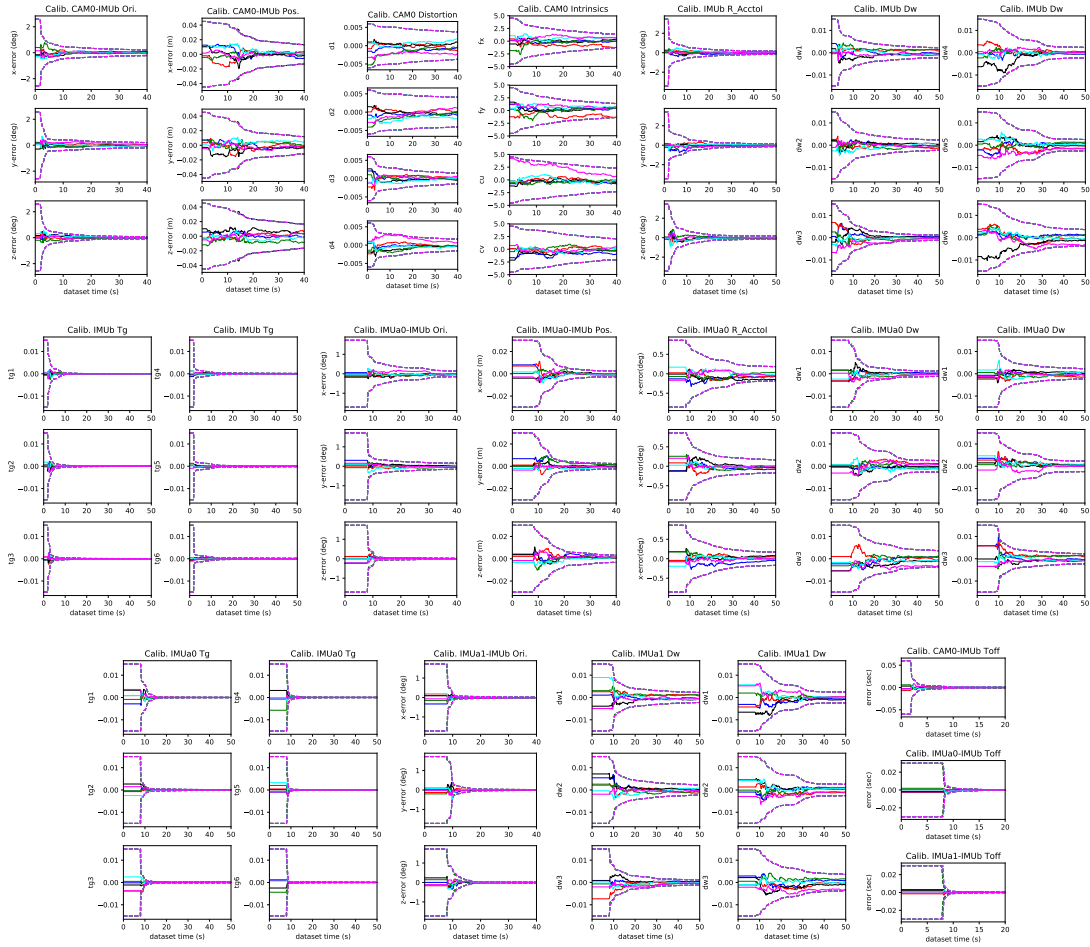
Figure 2: Three sigma bounds (dashed line) and estimation errors (solid line) for six different runs (shown in different colors) with different initial calibration perturbations and realization of the measurement noise for *fully excited motion.* The sensor suite contains (from top to bottom, left to 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.
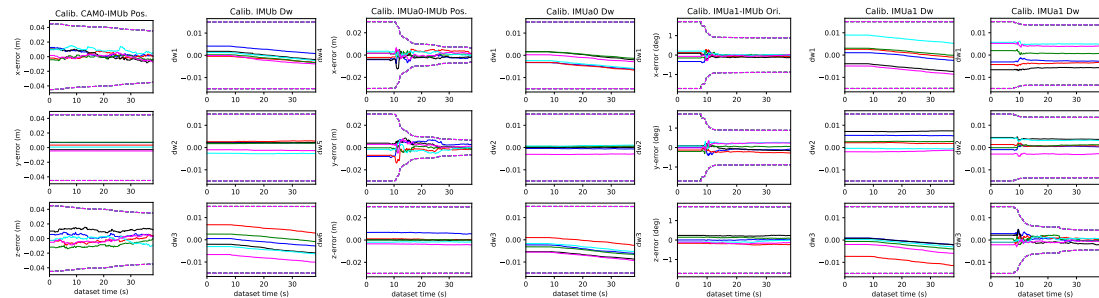

Figure 3: Three sigma bounds (dashed line) and estimation errors (solid line) for six different runs (shown in different colors) with different initial calibration perturbations and realization of the measurement noise for *1-axis motion.* Only the calibration parameters which showed inability to converge (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.
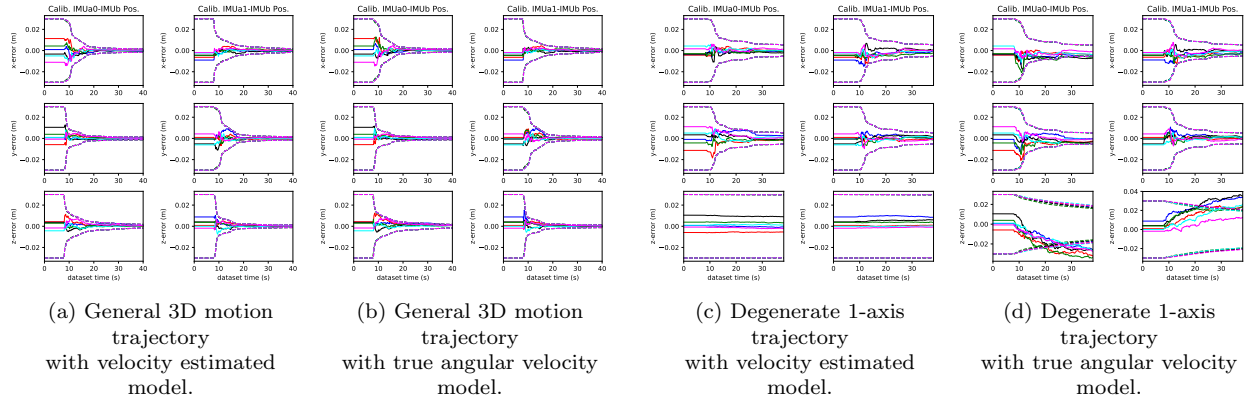
(a) General 3D motion
trajectory
with velocity estimated
model.

(b) General 3D motion
trajectory
with true angular velocity
model.

(c) Degenerate 1-axis
trajectory
with velocity estimated
model.

(d) Degenerate 1-axis
trajectory
with true angular velocity
model.

Figure 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.
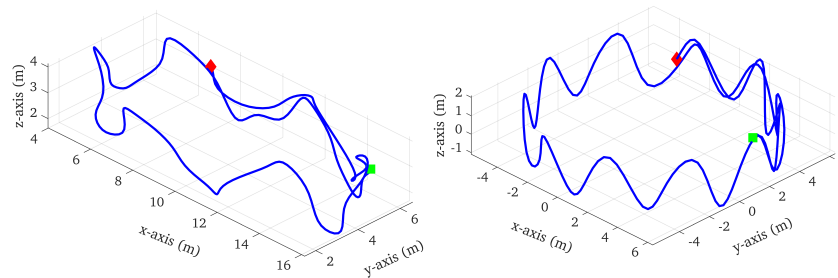

Figure 5: Simulated trajectories, axes are in units of meters. General 3D trajectory (left) is 53 meters long, while the 1-axis degenerate trajectory (right) is 90 meters in length. Green square denotes the start and red diamond denotes the end.

6   Simulation Validations

The simulator provided within the OpenVINS project [5] along with the multi-IMU extension from [22] 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 4 and the two trajectories used are shown in Figure 5.

6.1   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.

6.2   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 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 calibrate successfully in the full auxiliary IMU case.

6.3   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):

$$^{G}\mathbf{v}_{AI} = {}^{G}\mathbf{v}_{BI} + {}_{BI}^{G}\mathbf{R}\lfloor_{AI}^{BI}\mathbf{R}^{AI}\boldsymbol{\omega}\rfloor^{BI}\mathbf{p}_{AI} \tag{72}$$

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.

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

Figure 7: ArUco tags for stable visual feature tracking.

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 D. 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.

6.4 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 D) and instead we focus our real-world experiments on demonstrating the complete MVIS demonstration.

Table 5: The sensor setup for VI-Rig and T265-Rig.

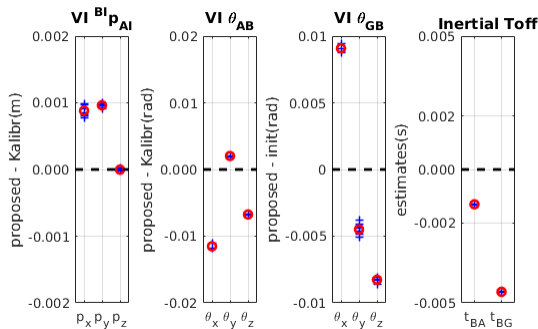|	IMUb	IMUa	Gyro	Cam0	Cam1	Cam2	Cam3
VI-Rig	MicroStrain0 500hz	MicroStrain1 100hz	T265 Gyro 200hz	T265 Left 30hz	Blackfly 25hz	T265 Right 30hz	-
T265-Rig	T265 IMU0 200hz	T265 IMU1 200hz	-	T265 Left0 30hz	T265 Right0 30hz	T265 Left1 30hz	T265 Right1 30hz

Figure 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.

7 Real-World Experiments

In addition to simulations, we also extensively examine the proposed MVIS on real-world datasets. We design two MVIS sensor rigs (Figure 6) to collect datasets and verify the proposed system. The VI-Rig (left of Figure 6) contains 2 Mircostrain IMUs[1], 1 BlackFly camera[2] and 1 Intel Realsense T265[3]. The T265-Rig (middle of Figure 6) contains 2 Intel Realsense T265. The sensor configuration are shown in Tab. 5. In order to focus on examining the proposed MVIS fusion algorithm, we simplify the visual-tracking frontend by using ArUco Tags [23] for stable visual feature tracking (Figure 7). One representative trajectory of a real-world experiment is shown on the right of Figure 6.

We collected 3 datasets with VI-Rig under fully excited motions. When using the proposed algorithms for calibration, we use 1, 2 and 3 camera variations for each dataset. Finally, we get 9 sets of calibration trials. Our calibration results are very close to the reference values we obtained from Kalibr. Note that we are unable to evaluate IMU intrinsic calibration directly since our IMU intrinsic model is different with the one presented in Kalibr. To verify the IMU intrinsic calibration, we further evaluate our system with 6 datasets collected using T265-Rig. Similar to VI-Rig, we calibrated these datasets by using 2 and 4 cameras, respectively. Finally, we get 12 sets of calibration trials for the T265-Rig.

---

[1] https://www.microstrain.com/inertial-sensors/imu

[2] https://www.flir.com/products/blackfly-s-usb3/

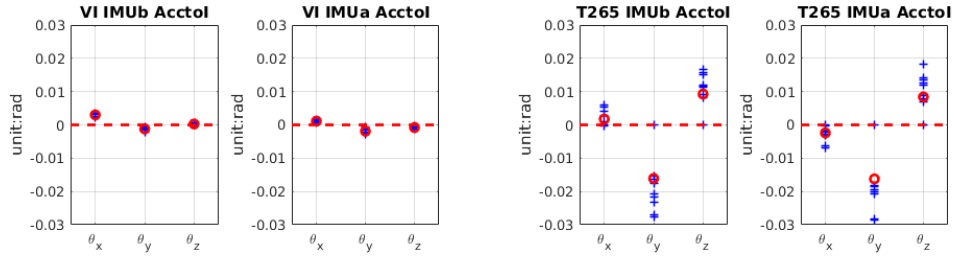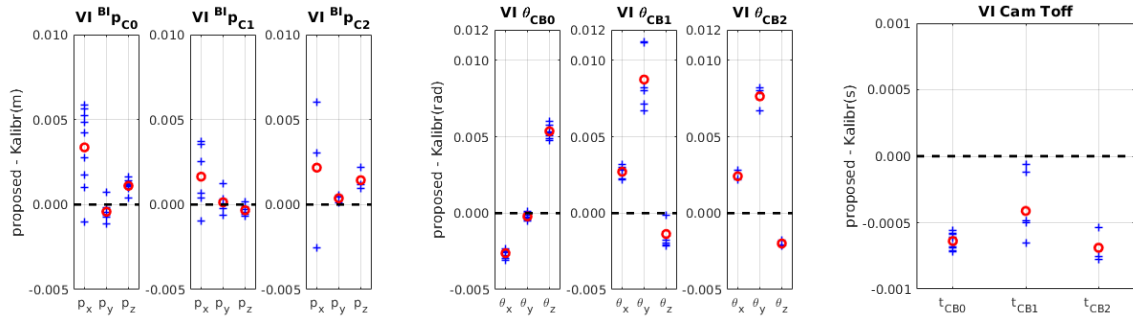[3] https://www.intelrealsense.com/tracking-camera-t265/

Figure 9: 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 doted 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 the same scale of y axis for both $^{BI}_{Ba}\mathbf{R}$ and $^{AI}_{Aa}\mathbf{R}$.

Figure 10: 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 doted 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.
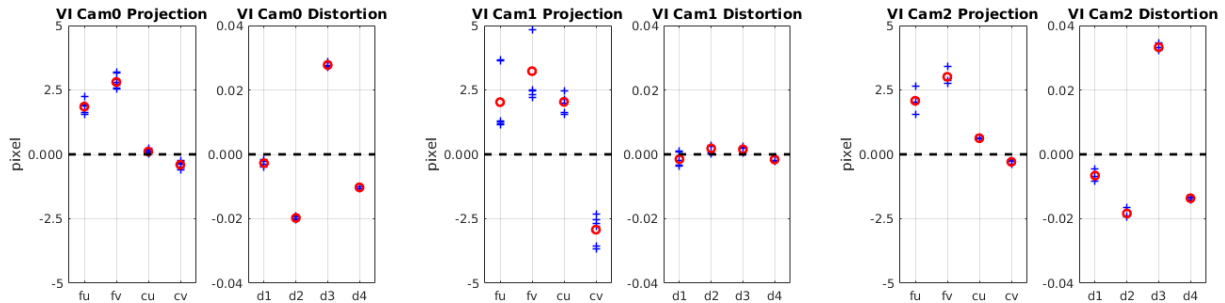

Figure 11: The camera intrinsics calibration results for VI-Rig and T265-Rig. Blue crosses represents the estimated values for each run. Red circle denotes the average value for all the runs. The black doted line denotes the values we get from Kalibr. It can be seen that the our cam intrinsic calibration can converge to a value near that we get from Kalibr.
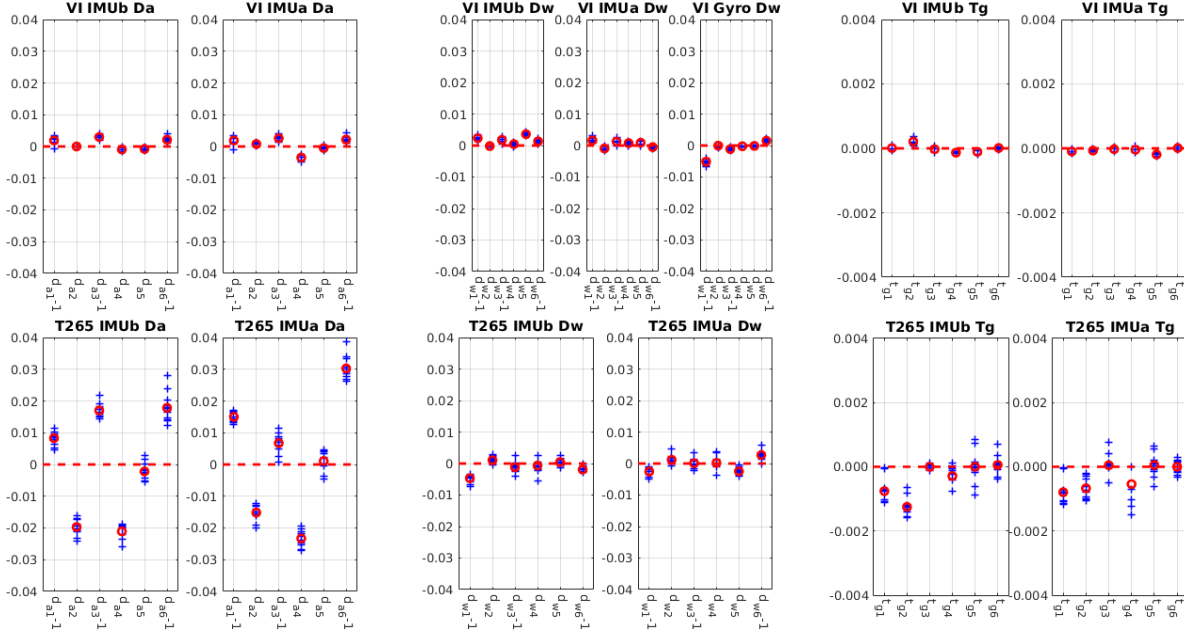
Figure 12: The IMU intrinsics calibration results for VI-Rig and T265-Rig. Blue crosses represents the estimated values for each run. Red circle denotes the average value for all the runs. The red doted line is the initial guess, which is also the ideal value we expected. The nearer the estimated value is to the red dotted line, the better the IMU is. To better demonstrate the calibration results, we use the 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 $\mathbf{T}_g$.

7.1 Spatial-Temporal Calibration Evaluation

We also calibrate the VI-Rig with the well-known toolbox Kalibr [24] 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 Figure 8 and Figure 10. Camera intrinsics are shown in Figure 11. Note that the spatial-temporal parameter calibration are repeatable and very close to the values we get from Kalibr calibration, see Figure 8.

7.2 IMU Intrinsics Evaluation

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 Figure 12.

First, it is interesting to note that the IMU intrinsics of MicroStrain IMU are an order smaller than that of T265 IMU, especially for $\mathbf{D}_a$, see Figure 12. 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 $\mathbf{D}_a$, $\mathbf{D}_w$, $\mathbf{T}_g$ 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 Figure 12). 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 $\mathbf{T}_g$ repeatability, it is one order smaller in magnitude compared to that of $\mathbf{D}_a$ and $\mathbf{D}_w$. Thus, this raises the question whether calibrating $\mathbf{T}_g$ will have a noticeable impact on MVIS performance.

Appendix A: Linear Pose Interpolation

A.1: 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 related 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 $\{^G_{S_{in}}\mathbf{R}, ^G\mathbf{p}_{S_{in}}\}$. 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:

$$^G_{S_{in}}\mathbf{R} = {}^G_{BI_k}\mathbf{R}\mathbf{Exp}(^{BI_k}\boldsymbol{\omega}\delta t)^{BI}_S\mathbf{R} \tag{73}$$

$$^G\mathbf{p}_{S_{in}} = {}^G\mathbf{p}_{BI_k} + {}^G\mathbf{v}_{BI_k}\delta t + {}^G_{S_{in}}\mathbf{R}^{BI}\mathbf{p}_S \tag{74}$$

where $\mathbf{Exp}(\cdot)$ 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, $\{^{BI}_S\mathbf{R}, ^{BI}\mathbf{p}_S\}$ 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).

A.2: 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 $\{^G_{S_{in}}\mathbf{R}, ^G\mathbf{p}_{S_{in}}\}$ with the two bounding IMU states $\mathbf{x}_{BI_k}$ and $\mathbf{x}_{BI_{k+1}}$ we can define the following:

$$^G_{S_{in}}\mathbf{R} = {}^G_{BI_k}\mathbf{R}\mathbf{Exp}(\lambda\mathbf{Log}(^{BI_k}_{BI_{k+1}}\mathbf{R}))^{BI}_S\mathbf{R}$$

$$^G\mathbf{p}_{S_{in}} = {}^G\mathbf{p}_{BI_k} + {}^G_{BI_k}\mathbf{R}(\lambda^{BI_k}\mathbf{p}_{BI_{k+1}}) + {}^G_{S_{in}}\mathbf{R}^{BI}\mathbf{p}_S$$

$$\lambda = (t_{S,in} + t_{SB} - t_{BI_k})/(t_{BI_{k+1}} - t_{BI_k})$$

where $t_{S,in}$ and t_{SI} are the measurement time and sensor time offset, $\{^{BI}_S\mathbf{R}, ^{BI}\mathbf{p}_S\}$ represents the sensor rigid transformation to IMU frame, and $\mathbf{Log}(\cdot)$ 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:

$$\mathbf{x}_{AI} = \begin{bmatrix} ^{AI}\mathbf{b}_g^\top & ^{AI}\mathbf{b}_a^\top \end{bmatrix}^\top \tag{75}$$

and have the rigid body kinematic constraints as:

$$
{}^{G}_{AI}\bar{q} = {}^{G}_{BI}\bar{q} \otimes {}^{BI}_{AI}\bar{q} \tag{76}
$$

$$
{}^{G}\mathbf{p}_{AI} = {}^{G}\mathbf{p}_{BI} + {}^{G}_{BI}\mathbf{R}^{BI}\mathbf{p}_{AI} \tag{77}
$$

$$
{}^{G}\mathbf{v}_{AI} = {}^{G}\mathbf{v}_{BI} + {}^{G}_{BI}\mathbf{R}\lfloor {}^{BI}_{AI}\mathbf{R}^{AI}\boldsymbol{\omega}\rfloor {}^{BI}\mathbf{p}_{AI} \tag{78}
$$

With this, we can reformulate the IMU measurements as:

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

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

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

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

Then, we can follow similar way as (37) and (35) to defined auxiliary IMU cost functions $\mathbb{C}_{AI}$ with or without IMU intrinsics $\mathbf{x}_{An}$.

Appendix C: 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.

C.1: ${}^{w}w_1$ constant

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

$$
\mathbf{N}_{w1} = \begin{bmatrix} \mathbf{0}_{1\times9} & (\hat{\mathbf{D}}_w^{-1}\mathbf{e}_1)^{\top}{}^{w}w_1 & \mathbf{0}_{1\times3} & 1 & \mathbf{0}_{1\times91} \end{bmatrix}^{\top} \tag{79}
$$

C.2: ${}^{w}w_2$ constant

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

$$
\mathbf{N}_{w2} = \begin{bmatrix} \mathbf{0}_{1\times9} & (\hat{\mathbf{D}}_w^{-1}\mathbf{e}_1)^{\top}{}^{w}w_2 & \mathbf{0}_{1\times4} & 1 & \mathbf{0}_{1\times90} \\ \mathbf{0}_{1\times9} & (\hat{\mathbf{D}}_w^{-1}\mathbf{e}_2)^{\top}{}^{w}w_2 & \mathbf{0}_{1\times5} & 1 & \mathbf{0}_{1\times89} \end{bmatrix}^{\top} \tag{80}
$$

C.3: ${}^{w}w_3$ constant

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

$$
\mathbf{N}_{w3} = \begin{bmatrix} \mathbf{0}_{1\times9} & (\hat{\mathbf{D}}_w^{-1}\mathbf{e}_1)^{\top}{}^{w}w_3 & \mathbf{0}_{1\times6} & 1 & \mathbf{0}_{1\times88} \\ \mathbf{0}_{1\times9} & (\hat{\mathbf{D}}_w^{-1}\mathbf{e}_2)^{\top}{}^{w}w_3 & \mathbf{0}_{1\times7} & 1 & \mathbf{0}_{1\times87} \\ \mathbf{0}_{1\times9} & (\hat{\mathbf{D}}_w^{-1}\mathbf{e}_3)^{\top}{}^{w}w_3 & \mathbf{0}_{1\times8} & 1 & \mathbf{0}_{1\times86} \end{bmatrix}^{\top} \tag{81}
$$

C.4: $^aa_1$ constant

If $^aa_1$ is constant, d_{a11}, pitch and yaw of $^I_a\mathbf{R}$ are unobservable with unobservable directions as:

$$
\mathbf{N}_{a1} = \begin{bmatrix}
\mathbf{0}_{12\times1} & \mathbf{0}_{12\times1} & \mathbf{0}_{12\times1} \\
\hat{\mathbf{D}}_a^{-1}\mathbf{e}_1{}^aa_1 & \hat{\mathbf{D}}_a^{-1}\mathbf{e}_2\hat{d}_{a11}{}^aa_1 & \hat{\mathbf{D}}_a^{-1}\mathbf{e}_3\hat{d}_{a11}\hat{d}_{a22}{}^aa_1 \\
\mathbf{0}_{6\times1} & \mathbf{0}_{6\times1} & \mathbf{0}_{6\times1} \\
1 & 0 & 0 \\
0 & \hat{d}_{a22} & 0 \\
0 & -\hat{d}_{a12} & 0 \\
0 & \hat{d}_{a23} & -\hat{d}_{a33}\hat{d}_{a22} \\
0 & -\hat{d}_{a13} & \hat{d}_{a12}\hat{d}_{a33} \\
0 & 0 & \hat{d}_{a13}\hat{d}_{a22} - \hat{d}_{a12}\hat{d}_{a23} \\
\mathbf{0}_{3\times1} & {}^I_a\hat{\mathbf{R}}\mathbf{e}_3 & {}^I_a\hat{\mathbf{R}}(\mathbf{e}_1\hat{d}_{a12} + \mathbf{e}_2\hat{d}_{a22}) \\
\mathbf{0}_{77\times1} & \mathbf{0}_{77\times1} & \mathbf{0}_{77\times1}
\end{bmatrix}
$$

C.5: $^aa_2$ constant

If $^aa_2$ is constant, d_{a12}, d_{a22} and roll of $^I_a\mathbf{R}$ are unobservable with unobservable directions as:

$$
\mathbf{N}_{a2} = \begin{bmatrix}
\mathbf{0}_{12\times1} & \mathbf{0}_{12\times1} & \mathbf{0}_{12\times1} \\
\hat{\mathbf{D}}_a^{-1}\mathbf{e}_1{}^aa_2 & \hat{\mathbf{D}}_a^{-1}\mathbf{e}_2{}^aa_2 & \hat{\mathbf{D}}_a^{-1}\mathbf{e}_3\hat{d}_{a22}{}^aa_2 \\
\mathbf{0}_{6\times1} & \mathbf{0}_{6\times1} & \mathbf{0}_{6\times1} \\
0 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0 \\
0 & 0 & \hat{d}_{a33} \\
0 & 0 & -\hat{d}_{a23} \\
\mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & {}^I_a\hat{\mathbf{R}}\mathbf{e}_1 \\
\mathbf{0}_{77\times1} & \mathbf{0}_{77\times1} & \mathbf{0}_{77\times1}
\end{bmatrix}
\tag{82}
$$

C.6: $^aa_3$ constant

If $^aa_3$ is constant, d_{a13}, d_{a23} and d_{a33} are unobservable with unobservable directions as:

$$
\mathbf{N}_{a3} = \begin{bmatrix}
\mathbf{0}_{1\times12} & (\hat{\mathbf{D}}_a^{-1}\mathbf{e}_1)^\top{}^aa_3 & \mathbf{0}_{1\times9} & 1 & \mathbf{0}_{1\times82} \\
\mathbf{0}_{1\times12} & (\hat{\mathbf{D}}_a^{-1}\mathbf{e}_2)^\top{}^aa_3 & \mathbf{0}_{1\times10} & 1 & \mathbf{0}_{1\times81} \\
\mathbf{0}_{1\times12} & (\hat{\mathbf{D}}_a^{-1}\mathbf{e}_3)^\top{}^aa_3 & \mathbf{0}_{1\times11} & 1 & \mathbf{0}_{1\times80}
\end{bmatrix}^\top
\tag{83}
$$

Appendix D: 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 13, while the inconsistent model which does not estimate the velocity of the auxiliary IMUs is able to perform similarly is shown in Figure 14. For the 1-axis degenerate motion case the results are also shown for the proposed model, see Figure 15, and the inconsistent model, see Figure 16. 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 E: Extended Real World Results

More results for VI-Rig datasets of the camera intrinsics and IMU intrinsics $^I_a\mathbf{R}$ are shown in Fig. 10, Fig. 9 and Fig. 11.
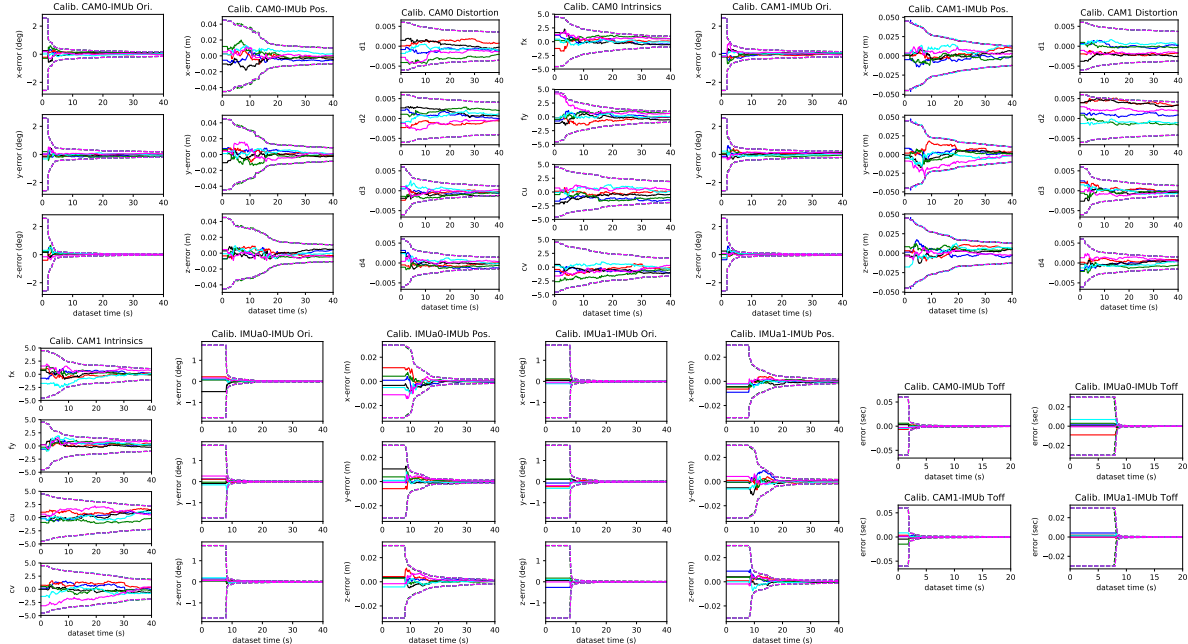
Figure 13: Three sigma bounds and estimation errors for six different runs with different initial calibration perturbations and realization of the measurement noise for *fully excited motion*. The sensor suite contains (from top to bottom, left to right) a monocular camera 0 (1-4), monocular camera 1 (5-8), auxiliary IMU 0 (9-10), auxiliary IMU 1 (11-11), and temporal time offsets.
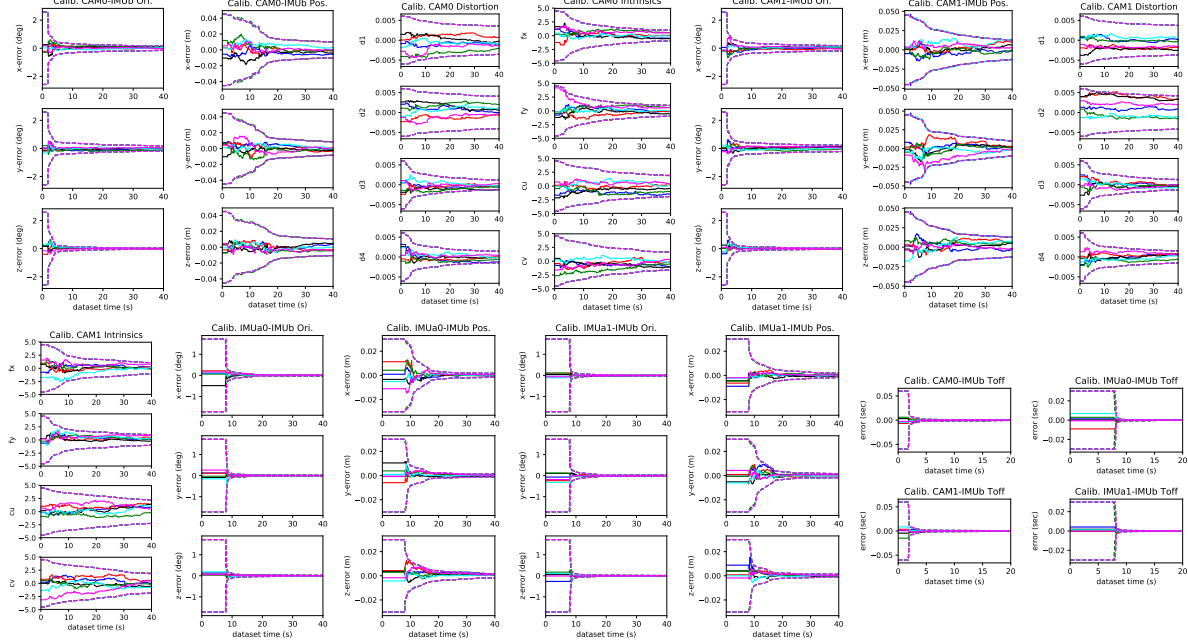

Figure 14: Three sigma bounds (dashed line) and estimation errors (solid line) for six different runs (shown in different color) with different initial calibration perturbations and realization of the measurement noise for *fully excited motion* with the *inconsistent model* which does not estimate second 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.
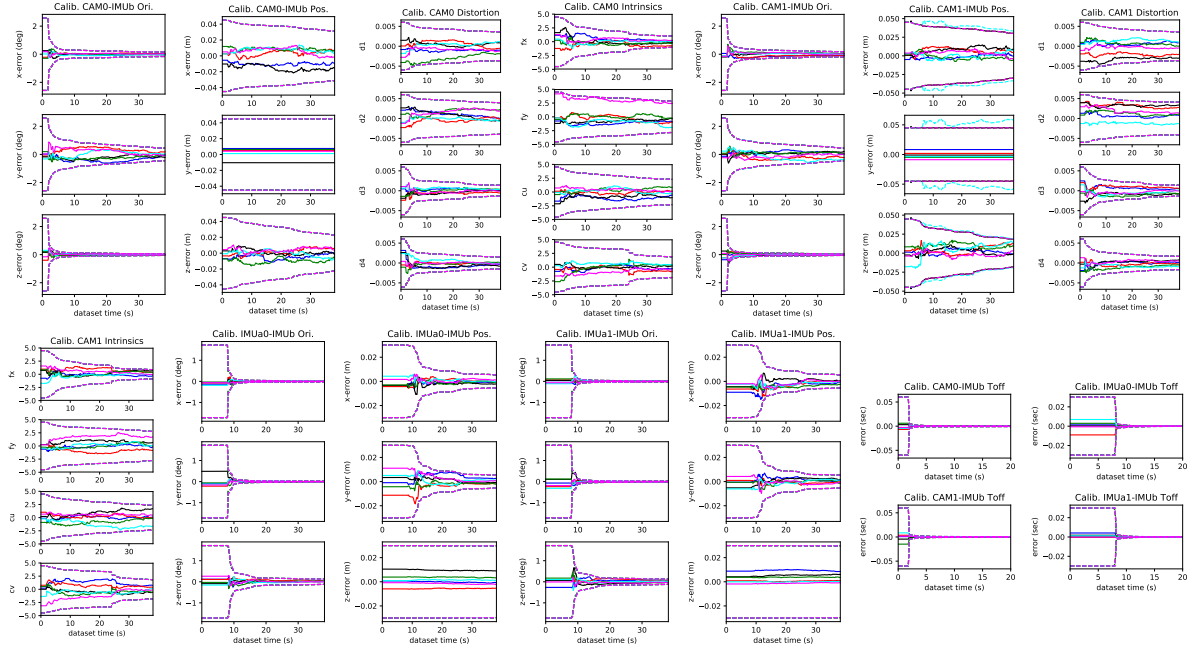
Figure 15: Three sigma bounds and estimation errors for six different runs with different initial calibration perturbations and realization of the measurement noise for *1-axis degenerate motion*. The sensor suite contains (from top to bottom, left to right) a monocular camera 0 (1-4), monocular camera 1 (5-8), auxiliary IMU 0 (9-10), auxiliary IMU 1 (11-11), and temporal time offsets.
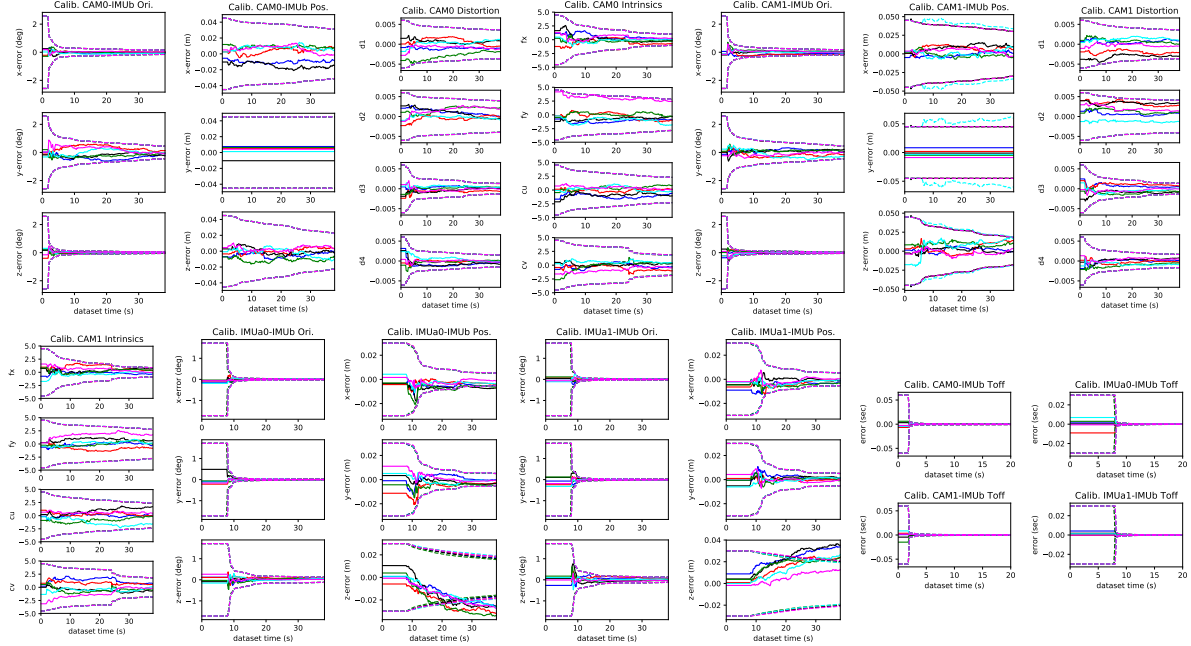

Figure 16: Three sigma bounds (dashed line) and estimation errors (solid line) for six different runs (shown in different colors) with different initial calibration perturbations and realization of the measurement noise for *1-axis degenerate motion* with the *inconsistent model* which does not estimate 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 (11-11), and temporal time offsets.

References

[1] Yulin Yang et al. "Online IMU Intrinsic Calibration: Is It Necessary?" In: *Proc. of the Robotics: Science and Systems*. Paris, France, 2020.

[2] Mingyang Li et al. "High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation". In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2014, pp. 409–416.

[3] Thomas Schneider et al. "Observability-aware self-calibration of visual and inertial sensors for ego-motion estimation". In: *IEEE Sensors Journal* 19.10 (2019), pp. 3846–3860.

[4] Yulin Yang et al. "Analytic Combined IMU Integrator for Visual-Inertial Navigation". In: *Proc. of the IEEE International Conference on Robotics and Automation*. Paris, France, 2020.

[5] Patrick Geneva et al. "OpenVINS: A Research Platform for Visual-Inertial Estimation". In: *Proc. of the IEEE International Conference on Robotics and Automation*. Paris, France, 2020. URL: https://github.com/rpng/open_vins.

[6] Timothy D Barfoot. *State estimation for robotics*. Cambridge University Press, 2017.

[7] Frank Dellaert. *Factor graphs and GTSAM: A hands-on introduction*. Tech. rep. Georgia Institute of Technology, 2012.

[8] Patrick Geneva, Kevin Eckenhoff, and Guoquan Huang. "A Linear-Complexity EKF for Visual-Inertial Navigation with Loop Closures". In: *Proc. International Conference on Robotics and Automation*. Montreal, Canada, May 2019.

[9] Nikolas Trawny and Stergios I Roumeliotis. "Indirect Kalman filter for 3D attitude estimation". In: *University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep* 2 (2005), p. 2005.

[10] Joan Sola. "Quaternion kinematics for the error-state Kalman filter". In: *arXiv preprint arXiv:1711.02508* (2017).

[11] Yao Xiao et al. "Online IMU self-calibration for visual-inertial systems". In: *Sensors* 19.7 (2019), p. 1624.

[12] Stefan Leutenegger et al. "Keyframe-based visual–inertial odometry using nonlinear optimization". In: *The International Journal of Robotics Research* 34.3 (2015), pp. 314–334.

[13] Tong Qin, Peiliang Li, and Shaojie Shen. "Vins-mono: A robust and versatile monocular visual-inertial state estimator". In: *IEEE Transactions on Robotics* 34.4 (2018), pp. 1004–1020.

[14] Kevin Eckenhoff, Patrick Geneva, and Guoquan Huang. "Closed-form Preintegration Methods for Graph-based Visual-Inertial Navigation". In: *International Journal of Robotics Research* 38.5 (2019), pp. 563–586.

[15] Guoquan Huang. "Improving the Consistency of Nonlinear Estimators: Analysis, Algorithms, and Applications". PhD thesis. Department of Computer Science and Engineering, University of Minnesota, 2012. URL: https://conservancy.umn.edu/handle/11299/146717.

[16] Agostino Martinelli. "Observabilty properties and deterministic algorithms in visual-inertial structure from motion". In: *Foundations and Trends in Robotics* (2013), pp. 1–75.

[17] Yulin Yang et al. "Degenerate Motion Analysis for Aided INS with Online Spatial and Temporal Calibration". In: *IEEE Robotics and Automation Letters (RA-L)* 4.2 (2019), pp. 2070–2077.

[18] Kejian J Wu et al. "Vins on wheels". In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2017, pp. 5155–5162.

[19] Woosik Lee et al. "Visual-Inertial-Wheel Odometry with Online Calibration". In: *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. Las Vegas, NV, 2020.

[20] J. A. Hesch et al. "Consistency Analysis and Improvement of Vision-aided Inertial Navigation". In: *IEEE Transactions on Robotics* 30.1 (2014), pp. 158–176.

[21] Yulin Yang and Guoquan Huang. "Observability Analysis of Aided INS with Heterogeneous Features of Points, Lines and Planes". In: *IEEE Transactions on Robotics* 35.6 (Dec. 2019), pp. 399–1418.

[22] Kevin Eckenhoff, Patrick Geneva, and Guoquan Huang. "MIMC-VINS: A Versatile and Resilient Multi-IMU Multi-Camera Visual-Inertial Navigation System". In: *IEEE Transactions on Robotics* (Feb. 2021). URL: https://arxiv.org/abs/2006.15699.

[23] S. Garrido-Jurado et al. "Generation of fiducial marker dictionaries using Mixed Integer Linear Programming". In: *Pattern Recognition* 51 (2016), pp. 481–491. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2015.09.023. URL: https://www.sciencedirect.com/science/article/pii/S0031320315003544.

[24] Paul Furgale, Joern Rehder, and Roland Siegwart. "Unified temporal and spatial calibration for multi-sensor systems". In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2013, pp. 1280–1286.