Leg Kinematic Inertial Navigation

Chinmay Burgul - cmburgul@udel.edu
Patrick Geneva - pgeneva@udel.edu
Woosik Lee - woosik@udel.edu
Guoquan Huang - ghuang@udel.edu

Department of Mechanical Engineering
University of Delaware, Delaware, USA

Robot Perception and Navigation Group (RPNG)
Tech Report - RPNG-2022-Kinematic-inertial-navigation
Last Updated - July 25, 2024

Contents

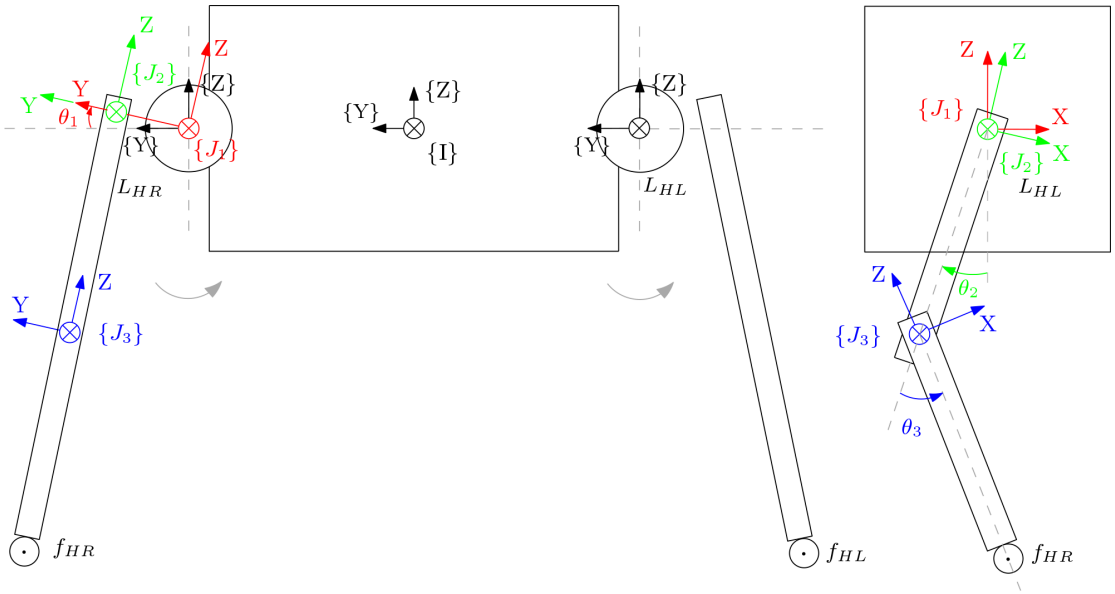
1 Methodology **1**
1.1 The State Vector 1
1.2 Leg Kinematic Model 2
1.2.1 Forward Kinematics 3
1.2.2 Kinematic Parameters Determination 3

2 Determining Kinematic Parameters **3**
2.0.1 VINS-based Motion Tracking of Floating Base 3
2.0.2 Estimating Legged Kinematics 4

3 Simulation Results **9**

4 Experimental Results **9**

References **16**



Front View Side View

1 Methodology

In this report, we explain how we can determine leg kinematics parameters by fusing Leg Kinematic measurements to the Visual-Inertial Navigation System. The Leg kinematic parameters includes the IMU-Leg spatial-temporal parameters and leg link lengths.

1.1 The State Vector

At time t_k, the state vector consists of current inertial state $\mathbf{x_{I_k}}$ and n historical IMU pose clones $\mathbf{x_{C_k}}$ and the position of the toe if they have a static ground contact as $\mathbf{x_T}$ represented in the global frame $\{\mathbf{G}\}$.

In the case of quadrupeds, if they stand still, all feet are in contact with the ground. While in a walking gait, either two or three feet are in contact with the ground. For simplicity, we will describe only one toe.

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{x}_{I_k}^\top & \mathbf{x}_{C_k}^\top & \mathbf{x}_T^\top & \mathbf{x}_\Theta^\top \end{bmatrix}^\top \tag{1}$$

$$\mathbf{x}_{I_k} = \begin{bmatrix} {}^{I_k}_{G}\bar{q}^\top & {}^G\mathbf{p}_{I_k}^\top & {}^G\mathbf{v}_{I_k}^\top & \mathbf{b}_g^\top & \mathbf{b}_a^\top \end{bmatrix}^\top \tag{2}$$

$$\mathbf{x}_{C_k} = \begin{bmatrix} {}^{I_{k-1}}_{G}\bar{q}^\top & {}^G\mathbf{p}_{I_{k-1}}^\top & \cdots & {}^{I_{k-n}}_{G}\bar{q}^\top & {}^G\mathbf{p}_{I_{k-n}}^\top \end{bmatrix}^\top \tag{3}$$

$$\mathbf{x}_T = \begin{bmatrix} {}^G\mathbf{p}_{T_1} & \cdots & {}^G\mathbf{p}_{T_w} \end{bmatrix} \tag{4}$$

$$\mathbf{x}_L = \begin{bmatrix} {}^L_I\mathbf{R} & {}^L\mathbf{p}_I & l_1 & l_2 & l_3 \end{bmatrix} \tag{5}$$

$$\mathbf{x}_\Theta = \begin{bmatrix} \mathbf{x}_{L_1} & \cdots & \mathbf{x}_{L_w} & {}^L t_I \end{bmatrix} \tag{6}$$

Where,
${}^I_G\bar{\mathbf{q}}$: is the JPL unit quaternion corresponding to the rotation ${}^I_G\mathbf{R}$ from Global $\{G\}$ to IMU frame $\{I\}$
${}^G\mathbf{p}_{I_k}$: is the position of IMU frame from $\{G\}$ to $\{I\}$
${}^G\mathbf{v}_{I_k}$: is the velocity of IMU frame from $\{G\}$ to $\{I\}$

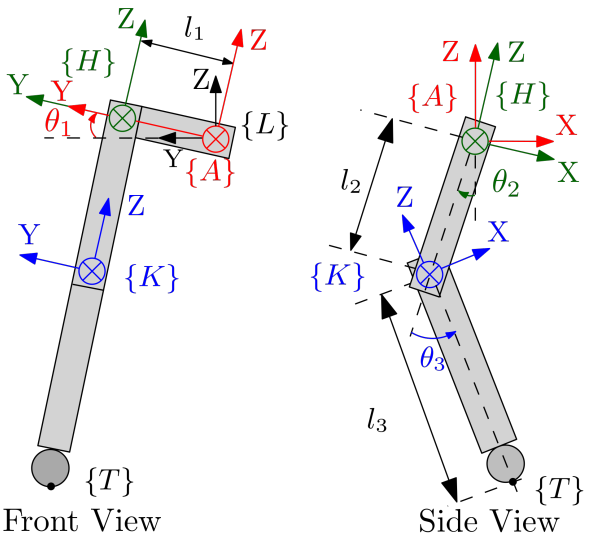


Figure 1: Kinematics of a general robot leg with four joint frames: abduction {A}, hip {H}, knee {K}, and toe {T}, and three links: hip (l_1), thigh (l_2) and shank (l_3).

$\mathbf{b}_g$, $\mathbf{b}_a$: are biases of the gyroscope and accelerometer
$\mathbf{x}_T$: are the w static toe landmark positions
${}^G\mathbf{p}_T$: is the position of toe T in global frame $\{G\}$
$\mathbf{x}_L$: are the IMU-LEG spatial parameters and link length parameters for a leg
$\mathbf{x}_\Theta$: are the kinematic parameters of w Legs along with time offset.

We define $\mathbf{x} = \hat{\mathbf{x}} \boxplus \tilde{\mathbf{x}}$, where $\mathbf{x}$ is the true state, $\hat{\mathbf{x}}$ as its estimate, $\tilde{\mathbf{x}}$ is the error state, and the operation $\boxplus$ which maps the error state vector to its corresponding manifold [1].

By using the Multi-State Constraint Kalman Filter, we propagate with the IMU measurements and update the state with visual measurements and Leg encoder measurements. The IMU propagation and the visual sensor update are not the main scope of this paper, so we will describe in detail how we fuse the Leg kinematic measurements and determine the Kinematic parameters within the following chapters.

1.2 Leg Kinematic Model

We present a general kinematic model of a single leg, while it can be extended to any legged robot with an arbitrary number of legs, such as quadrupeds or hexapods. Each leg has three joints, each joint consisting of an actuator and encoder. Encoders measure absolute angular readings between two links corrupted with a zero mean white-Gaussian noise.

Fig. 1 illustrates the standard kinematic chain and frame of references of a robot leg. The leg frame $\{L\}$ is rigidly connected to the robot's body frame of reference $\{I\}$ where the high-level motion planner, low-level locomotion controller or state estimator is typically operating, which clearly necessitates finding the leg kinematic relationships for legged robot's estimation and control. The leg $\{L\}$ often aligns with the abduction joint $\{A\}$ whose orientation is altered by activating the first joint. As seen from Fig. 1, the three links connect the four joint frames of abduction {A}, hip {H}, knee {K}, and toe {T}. The top three joints are actuated with joint encoders to provide absolute joint angle measurements:

$$\theta_{m,i} = \theta_i + n_{\theta_i}, \quad \forall i \in \{A, H, K\} \tag{7}$$

where θ_i is the true angle and n_{θ_i} is the white Gaussian noise. Stacking all of them yields the measurement vector $\boldsymbol{\theta}_m$.

1.2.1 Forward Kinematics

Foot placement is critical for legged locomotion, which manifestly requires the kinematic chain to determine the time-varying position of the foot (toe) $\{T\}$ with respect to the leg $\{L\}$ and thus the body $\{I\}$:

$$ {}^{I}\mathbf{p}_T = {}^{I}\mathbf{p}_L + {}^{L}_{I}\mathbf{R}^\top \left[{}^{L}\mathbf{p}_A + {}^{L}_{A}\mathbf{R}({}^{A}\mathbf{p}_H + {}^{A}_{H}\mathbf{R}({}^{H}\mathbf{p}_K + {}^{H}_{K}\mathbf{R}\,{}^{K}\mathbf{p}_T)) \right] \tag{8} $$

where the 6DOF rigid transformation between the leg and body frames $\{{}^{I}\mathbf{p}_L, {}^{L}_{I}\mathbf{R}\}$ needs to be estimated, e.g., by leveraging visual-inertial systems, and the other transformations between the associated joint frames are given by:

$$ {}^{L}\mathbf{p}_A = \mathbf{0}_{3\times 1}, \qquad {}^{L}_{A}\mathbf{R} = \mathrm{Exp}(\theta_A \mathbf{k}_A) \tag{9} $$

$$ {}^{A}\mathbf{p}_H = l_1 \mathbf{k}_1 \,, \qquad {}^{A}_{H}\mathbf{R} = \mathrm{Exp}(\theta_H \mathbf{k}_H) \tag{10} $$

$$ {}^{H}\mathbf{p}_K = l_2 \mathbf{k}_2 \,, \qquad {}^{H}_{K}\mathbf{R} = \mathrm{Exp}(\theta_K \mathbf{k}_K) \tag{11} $$

$$ {}^{K}\mathbf{p}_T = l_3 \mathbf{k}_3 \tag{12} $$

In the above expressions, $\mathbf{k}_A$, $\mathbf{k}_H$, and $\mathbf{k}_K$ are the rotation axis of the joints. $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$ are the unit direction vector between the joints. $\mathrm{Exp}(\cdot)$ is the $SO(3)$ matrix exponential function to represent the rotation [2].

1.2.2 Kinematic Parameters Determination

It becomes evident from the leg forward kinematic chain (8) that the rigid transformation $\{{}^{I}\mathbf{p}_L, {}^{L}_{I}\mathbf{R}\}$ is one of the key parameters required to calibrate accurately. Moreover, note that the leg is assumed to be aligned with the abduction actuator/encoder, which could be temporally (not only spatially) different from the body's IMU due to improper hardware triggering, transmission delays, or clock synchronization errors. This necessitates to also estimate the timeline misalignment (time offset) between the body's IMU and the leg's encoders. Note also that this temporal disparity may vary over time due to sensing latency or communication delays stemming from computational overload.

We thus model this varying time offset ${}^{I}t_L$ as:

$$ {}^{I}t_k = {}^{L}t_k + {}^{L}t_I \tag{13} $$

where ${}^{I}t_k$ and ${}^{L}t_k$ are the timestamps when measurement was stamped in the body IMU and leg encoder's clocks.

On the other hand, we do assume the leg encoders are all synchronized, which is a reasonable and common assumption in practice in order to synchronize all actuators. However, we do need to determine the lengths of three-link segments $\mathbf{x}_{LI} = [l_1 \; l_2 \; l_3]^\top$, as they are required to capitalize the leg kinematics (8) and may undergo variations due to contact-induced deformation during operations (e.g., walking).

2 Determining Kinematic Parameters

2.0.1 VINS-based Motion Tracking of Floating Base

As a common practice, we assume the legged robot system is represented by the IMU states on the robot's body or floating base, to which the locomotion control is applied. We leverage our

MSCKF-based VINS [3] to track its motion efficiently. Specifically, the IMU state (2) evolves over time with the IMU kinematics, using incoming linear acceleration ${}^{I}\mathbf{a}_k$, and angular velocity ${}^{I}\boldsymbol{\omega}_k$, which is given by (see [4]):

$$\mathbf{x}_{I_{k+1}} = \mathbf{f}_I(\mathbf{x}_{I_k}, {}^{I}\mathbf{a}_k, {}^{I}\boldsymbol{\omega}_k, \mathbf{n}_I) \tag{14}$$

Where $\mathbf{n}_I = [\mathbf{n}_a^\top \ \mathbf{n}_g^\top \ \mathbf{n}_{wa}^\top \ \mathbf{n}_{wg}^\top]^\top$ contains the zero-mean white Gaussian noises and random walks of the biases. We detect and track static environmental features over images with optical flow. The bearing measurement $\mathbf{z}_{C_k}$ of a feature $\mathbf{p}_f$ at timestep k is the perspective projection of its 3D global position ${}^{G}\mathbf{p}_f$ onto the image plane:

$$\mathbf{z}_{C_k} = \mathbf{h}_c(\mathbf{x}_{C_k}, {}^{C_k}\mathbf{p}_f) + \mathbf{n}_{C_k} \tag{15}$$

$${}^{C_k}\mathbf{p}_f = {}^{C}_{I}\mathbf{R}\,{}^{I_k}_{G}\mathbf{R}({}^{G}\mathbf{p}_f - {}^{G}\mathbf{p}_{I_k}) + {}^{C}\mathbf{p}_I \tag{16}$$

where $\{{}^{C}_{I}\mathbf{R}, {}^{C}\mathbf{p}_I\}$ is the IMU-camera extrinsic transformation which is assumed to be known otherwise can be calibrated either online or offline [5], and $\mathbf{n}_{C_k}$ is the white Gaussian measurement noise. We now employ the MSCKF to efficiently update the state with these visual measurements [3].

2.0.2 Estimating Legged Kinematics

Provided the motion tracking (i.e., "odometry") of the robot's body as described in the preceding section and given the legged kinematic constraint in Eq. (8), one may simply build a decoupled SLAM estimator by including the kinematic parameters and the static toes (as landmarks) as part of the state vector along with the robot body/IMU states in order to identify the kinematics. However, as decoupling the legged kinematics from the body's motion tracking may cause information loss, we propose to tightly couple the kinematic parameters into the VINS estimator so that the body's motion tracking can also benefit from the legged kinematics.

Specifically, the forward legged kinematics (8) reveals the relationship of the toe position in the body frame, ${}^{I}\mathbf{p}_{T_k}$, with the joint encoder measurements and the kinematic parameters (which are included in our state vector (1)). On the other hand, we can also represent the toe position in terms of the body pose and the global toe position (both of which are part of the state vector (1)) as follows:

$${}^{I}\mathbf{p}_{T_k} = {}^{I_k}_{G}\mathbf{R}\,{}^{G}\mathbf{p}_T - {}^{I_k}_{G}\mathbf{R}\,{}^{G}\mathbf{p}_{I_k} \tag{17}$$

To infer a legged kinematic measurement model that can be used in the proposed tightly-coupled MSCKF update, we equate (8) and (17) and build the following implicit measurement (note that we drop off the time index for simplicity):

$$\begin{aligned}\mathbf{z}_\ell := \mathbf{0} = \ & {}^{I}_{G}\mathbf{R}\,{}^{G}\mathbf{p}_T - {}^{I}_{G}\mathbf{R}\,{}^{G}\mathbf{p}_I - {}^{I}\mathbf{p}_L - \\ & {}^{L}_{I}\mathbf{R}^\top \left[{}^{L}\mathbf{p}_A + {}^{L}_{A}\mathbf{R}({}^{A}\mathbf{p}_H + {}^{A}_{H}\mathbf{R}({}^{H}\mathbf{p}_K + {}^{H}_{K}\mathbf{R}\,{}^{K}\mathbf{p}_T))\right] =: \mathbf{h}_\ell(\mathbf{x}, \boldsymbol{\theta}_m)\end{aligned} \tag{18}$$

where $\boldsymbol{\theta}_m$ denotes the stacked joint encoder measurements which are corrupted by zero-mean white Gaussian noise $\mathbf{n}_\ell = [n_{\theta_1} \ n_{\theta_2} \ n_{\theta_3}]^\top$ (see (7)). The residual of this inferred legged measurement (18) can be obtained by linearizing it with respect to the state $\mathbf{x}$ and encoder measurement noise $\mathbf{n}_\ell$ as:

$$\mathbf{r}_\ell = \mathbf{z}_\ell - \hat{\mathbf{z}}_\ell = -\hat{\mathbf{z}}_\ell = \mathbf{H}_x\tilde{\mathbf{x}} + \mathbf{G}_\mathbf{n}\mathbf{n}_\ell \tag{19}$$

where the two Jacobians are respectively defined by $\mathbf{H}_x = \left[\frac{\partial \mathbf{h}_\ell}{\partial {}^I_G\theta} \; \frac{\partial \mathbf{h}_\ell}{\partial {}^G\mathbf{p}_I} \; \frac{\partial \mathbf{h}_\ell}{\partial {}^G\mathbf{p}_T} \; \frac{\partial \mathbf{h}_\ell}{\partial {}^I t_L} \; \frac{\partial \mathbf{h}_\ell}{\partial {}^L_I\theta} \; \frac{\partial \mathbf{h}_\ell}{\partial {}^L\mathbf{p}_I} \; \frac{\partial \mathbf{h}_\ell}{\partial l_1} \; \frac{\partial \mathbf{h}_\ell}{\partial l_2} \; \frac{\partial \mathbf{h}_\ell}{\partial l_3}\right]$ and $\mathbf{G}_n = \left[\frac{\partial \mathbf{h}_\ell}{\partial n_{\theta_A}} \; \frac{\partial \mathbf{h}_\ell}{\partial n_{\theta_H}} \; \frac{\partial \mathbf{h}_\ell}{\partial n_{\theta_K}}\right]$, and are given by:

$$\frac{\partial \mathbf{h}_\ell}{\partial {}^I_G\boldsymbol{\theta}} = \lfloor {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I)\rfloor \tag{20}$$

$$\frac{\partial \mathbf{h}_\ell}{\partial {}^G\mathbf{p}_I} = -{}^I_G\mathbf{R} \; , \; \frac{\partial \mathbf{h}_\ell}{\partial {}^G\mathbf{p}_T} = {}^I_G\mathbf{R} \tag{21}$$

$$\frac{\partial \mathbf{h}_\ell}{\partial {}^I t_L} = -\lfloor {}^{I_k}_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_{I_k})\rfloor {}^{I_k}\boldsymbol{\omega} + {}^{I_k}_G\mathbf{R}\,{}^G\mathbf{v}_{I_k} \tag{22}$$

$$\frac{\partial \mathbf{h}_\ell}{\partial {}^L_I\theta} = -{}^L_I\mathbf{R}^\top \lfloor ({}^L\mathbf{p}_I - {}^L\mathbf{p}_T)\rfloor \; , \; \frac{\partial \mathbf{h}_\ell}{\partial {}^L\mathbf{p}_I} = {}^L_I\mathbf{R}^\top \tag{23}$$

$$\frac{\partial \mathbf{h}_\ell}{\partial l_1} = -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}\mathbf{k}_2, \; \frac{\partial \mathbf{h}_\ell}{\partial l_2} = -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}\,{}^A_H\mathbf{R}\mathbf{k}_3, \; \frac{\partial \mathbf{h}_\ell}{\partial l_3} = -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}\,{}^A_H\mathbf{R}\,{}^H_K\mathbf{R}\mathbf{k}_3 \tag{24}$$

$$\frac{\partial \mathbf{h}_\ell}{\partial n_{\theta_A}} = -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}\lfloor {}^A\mathbf{p}_H + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}\,{}^K\mathbf{p}_T)\rfloor \mathbf{J}_r(\theta_A\mathbf{k}_A)\mathbf{k}_A \tag{25}$$

$$\frac{\partial \mathbf{h}_\ell}{\partial n_{\theta_H}} = -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}\,{}^A_H\mathbf{R}\lfloor {}^H\mathbf{p}_K + {}^H_K\mathbf{R}\,{}^K\mathbf{p}_T\rfloor \mathbf{J}_\mathbf{r}(\theta_H\mathbf{k}_H)\; n_{\theta_H}\mathbf{k}_H \tag{26}$$

$$\frac{\partial \mathbf{h}_\ell}{\partial n_{\theta_K}} = -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}\,{}^A_H\mathbf{R}\,{}^H_K\mathbf{R}\lfloor {}^K\mathbf{p}_T\rfloor \mathbf{J}_\mathbf{r}(\theta_K\mathbf{k}_K) n_{\theta_K}\mathbf{k}_K \tag{27}$$

Where $\lfloor\cdot\rfloor$ is the skew-symmetric matrix. These Jacobians are critical to ensure accurate and consistent estimation, and their detailed derivations are shown below:

$$\mathbf{h}_\ell = {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^I_L\mathbf{R}\,{}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top \left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}\,{}^K\mathbf{p}_T))\right]$$
$$\mathbf{h}_\ell = {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T)$$

For eq. (20)

$$\begin{aligned}
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &\approx (\mathbf{I_{3x3}} - \lfloor {}^I_G\delta\theta\rfloor) {}^I_G\hat{\mathbf{R}}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &= {}^I_G\hat{\mathbf{R}}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) - \lfloor {}^I_G\delta\theta\rfloor {}^I_G\hat{\mathbf{R}}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) \\
\tilde{\mathbf{h}}_\ell &= \lfloor {}^I_G\hat{\mathbf{R}}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I)\rfloor {}^I_G\delta\boldsymbol{\theta} \\
\frac{\partial \mathbf{h}_\ell}{\partial {}^I_G\boldsymbol{\theta}} &= \lfloor {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I)\rfloor
\end{aligned}$$

For eq. (21)

$$\begin{aligned}
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &\approx {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\hat{\mathbf{p}}_I - {}^G\tilde{\mathbf{p}}_\mathbf{I}) + {}^L_I\mathbf{R}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\hat{\mathbf{p}}_I) + {}^L_I\mathbf{R}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) - {}^I_G\mathbf{R}\,{}^G\tilde{\mathbf{p}}_\mathbf{I} \\
\tilde{\mathbf{h}}_\ell &= -{}^I_G\mathbf{R}\,{}^G\tilde{\mathbf{p}}_\mathbf{I} \\
\tilde{\mathbf{h}}_\ell &= -{}^L_I\mathbf{R}\,{}^I_G\mathbf{R}\,{}^G\tilde{\mathbf{p}}_\mathbf{I} \\
\frac{\partial \mathbf{h}_\ell}{\partial {}^G\mathbf{p}_I} &= -{}^I_G\mathbf{R}
\end{aligned}$$

$$
\begin{aligned}
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &\approx {}^I_G\mathbf{R}({}^G\hat{\mathbf{p}}_T + {}^G\tilde{\mathbf{p}}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &= {}^I_G\mathbf{R}({}^G\hat{\mathbf{p}}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) + {}^I_G\mathbf{R}{}^G\tilde{\mathbf{p}}_T \\
\tilde{\mathbf{h}}_\ell &= {}^I_G\mathbf{R}{}^G\tilde{\mathbf{p}}_\mathbf{T} \\
\frac{\partial \mathbf{h}_\ell}{\partial {}^G\mathbf{p}_T} &= {}^I_G\mathbf{R}
\end{aligned}
$$

For eq.(23)

$$
\begin{aligned}
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &\approx {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + ((\mathbf{I_3} - \lfloor {}^L_I\delta\theta \rfloor){}^L_I\hat{\mathbf{R}})^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\hat{\mathbf{R}}^\top(\mathbf{I_3} + \lfloor {}^L_I\delta\theta \rfloor)({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\hat{\mathbf{R}}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) + {}^L_I\hat{\mathbf{R}}^\top\lfloor {}^L_I\delta\theta \rfloor({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) \\
\tilde{\mathbf{h}}_\ell &= {}^L_I\hat{\mathbf{R}}^\top\lfloor {}^L_I\delta\boldsymbol{\theta} \rfloor({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) \\
\frac{\partial \mathbf{h}_\ell}{\partial {}^L_I\boldsymbol{\theta}} &= -{}^L_I\mathbf{R}^\top\lfloor({}^L\mathbf{p}_I - {}^L\mathbf{p}_T)\rfloor
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top({}^L\mathbf{p}_I - {}^L\mathbf{p}_T) \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &\approx {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top({}^L\hat{\mathbf{p}}_I + {}^L\tilde{\mathbf{p}}_I - {}^L\mathbf{p}_T) \\
\tilde{\mathbf{h}}_\ell &= {}^L_I\mathbf{R}^\top{}^L\tilde{\mathbf{p}}_I \\
\frac{\partial \mathbf{h}_\ell}{\partial {}^I\mathbf{p}_L} &= {}^L_I\mathbf{R}^\top
\end{aligned}
$$

For eq.(24)

$$
\begin{aligned}
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top{}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top\left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}{}^K\mathbf{p}_T))\right] \\
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top{}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top\left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}(l_1\mathbf{k_2} + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}{}^K\mathbf{p}_T))\right] \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &\approx {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top{}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top\left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}((\hat{l}_1 + \tilde{l}_1)\mathbf{k_2} + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}{}^K\mathbf{p}_T))\right] \\
\tilde{\mathbf{h}}_\ell &= -{}^L_I\mathbf{R}^\top{}^L_A\mathbf{R}\mathbf{k}_2\tilde{l}_1 \\
\frac{\partial \mathbf{h}_\ell}{\partial l_1} &= -{}^L_I\mathbf{R}^\top{}^L_A\mathbf{R}\mathbf{k}_2
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top{}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top\left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}{}^K\mathbf{p}_T))\right] \\
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top{}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top\left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}(l_2\mathbf{k_3} + {}^H_K\mathbf{R}{}^K\mathbf{p}_T))\right] \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &\approx {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top{}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top\left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}((\hat{l}_2 + \tilde{l}_2)\mathbf{k_3} + {}^H_K\mathbf{R}{}^K\mathbf{p}_T))\right] \\
\tilde{\mathbf{h}}_\ell &= -{}^L_I\mathbf{R}^\top{}^L_A\mathbf{R}{}^A_H\mathbf{R}\mathbf{k}_3\tilde{l}_2 \\
\frac{\partial \mathbf{h}_\ell}{\partial l_2} &= -{}^L_I\mathbf{R}^\top{}^L_A\mathbf{R}{}^A_H\mathbf{R}\mathbf{k}_3
\end{aligned}
$$

$$
\mathbf{h}_\ell = {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top{}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top\left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}{}^K\mathbf{p}_T))\right]
$$

$$
\begin{aligned}
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top {}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top \left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}l_3\mathbf{k}_3))\right] \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &\approx {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top {}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top \left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}(\hat{l}_3 + \tilde{l}_3)\mathbf{k}_3))\right] \\
\tilde{\mathbf{h}}_\ell &= -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}{}^A_H\mathbf{R}{}^H_K\mathbf{R}\mathbf{k}_3\tilde{l}_3 \\
\frac{\partial \mathbf{h}_\ell}{\partial l_3} &= -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}{}^A_H\mathbf{R}{}^H_K\mathbf{R}\mathbf{k}_3
\end{aligned}
$$

For eq.(25)

$$
\begin{aligned}
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top {}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top \left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}{}^K\mathbf{p}_T))\right] \\
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top {}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top \left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}{}^A\mathbf{p}_T\right] \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &\approx {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top {}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top \left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}(\mathbf{I_3} + \lfloor -\mathbf{J_r}(\theta_A\mathbf{k}_A)\ n_{\theta_A}\mathbf{k}_A\rfloor){}^A\mathbf{p}_T\right] \\
\tilde{\mathbf{h}}_\ell &= -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}\lfloor {}^A\mathbf{p}_T\rfloor(\mathbf{J_r}(\theta_A\mathbf{k}_A)\ n_{\theta_A}\mathbf{k}_A) \\
\frac{\partial \mathbf{h}_\ell}{\partial n_{\theta_A}} &= -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}\lfloor {}^A\mathbf{p}_H + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}{}^K\mathbf{p}_T)\rfloor\mathbf{J_r}(\theta_A\mathbf{k}_A)\mathbf{k}_A
\end{aligned}
$$

For eq.(26)

$$
\begin{aligned}
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top {}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top \left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}{}^K\mathbf{p}_T))\right] \\
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top {}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top \left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}{}^H\mathbf{p}_T)\right] \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &\approx {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top {}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top \left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}(\mathbf{I_3} + \lfloor -\mathbf{J_r}(\theta_H\mathbf{k}_H)\ n_{\theta_H}\mathbf{k}_H\rfloor){}^H\mathbf{p}_T)\right] \\
\tilde{\mathbf{h}}_\ell &= -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}{}^A_H\mathbf{R}\lfloor {}^H\mathbf{p}_T\rfloor(\mathbf{J_r}(\theta_H\mathbf{k}_H)\ n_{\theta_H}\mathbf{k}_H) \\
\frac{\partial \mathbf{h}_\ell}{\partial n_{\theta_H}} &= -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}{}^A_H\mathbf{R}\lfloor {}^H\mathbf{p}_K + {}^H_K\mathbf{R}{}^K\mathbf{p}_T\rfloor\mathbf{J_r}(\theta_H\mathbf{k}_H)\ n_{\theta_H}\mathbf{k}_H
\end{aligned}
$$

For eq.(27)

$$
\begin{aligned}
\mathbf{h}_\ell &= {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top {}^L\mathbf{p}_I - {}^L_I\mathbf{R}^\top \left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}{}^K\mathbf{p}_T))\right] \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &\approx {}^I_G\mathbf{R}({}^G\mathbf{p}_T - {}^G\mathbf{p}_I) + {}^L_I\mathbf{R}^\top {}^L\mathbf{p}_I \\
&\quad - {}^L_I\mathbf{R}^\top \left[{}^L\mathbf{p}_A + {}^L_A\mathbf{R}({}^A\mathbf{p}_H + {}^A_H\mathbf{R}({}^H\mathbf{p}_K + {}^H_K\mathbf{R}(\mathbf{I_3} + \lfloor -\mathbf{J_r}(\theta_K\mathbf{k}_K)n_{\theta_K}\mathbf{k}_K\rfloor){}^K\mathbf{p}_T))\right] \\
\tilde{\mathbf{h}}_\ell &= -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}{}^A_H\mathbf{R}{}^H_K\mathbf{R}\lfloor {}^K\mathbf{p}_T\rfloor\mathbf{J_r}(\theta_K\mathbf{k}_K)n_{\theta_K}\mathbf{k}_K \\
\frac{\partial \mathbf{h}_\ell}{\partial n_{\theta_K}} &= -{}^L_I\mathbf{R}^\top {}^L_A\mathbf{R}{}^A_H\mathbf{R}{}^H_K\mathbf{R}\lfloor {}^K\mathbf{p}_T\rfloor\mathbf{J_r}(\theta_K\mathbf{k}_K)n_{\theta_K}\mathbf{k}_K
\end{aligned}
\tag{28}
$$

For the derivatives of parameter in the $SO(3)$, we use the Baker-Campbell-Hausdorff formula for approximations for small angles $\boldsymbol{\psi}$

$$
\begin{aligned}
Exp(\boldsymbol{\theta} + \boldsymbol{\psi}) &\approx Exp(\mathbf{J}_l(\boldsymbol{\theta})\boldsymbol{\psi})Exp(\boldsymbol{\theta}) \\
&\approx Exp(\boldsymbol{\theta})Exp(\mathbf{J}_r(\boldsymbol{\theta})\boldsymbol{\psi}) = \mathbf{R}(\mathbf{I}_3 - \lfloor \mathbf{J}_r(\boldsymbol{\theta})\boldsymbol{\psi}\rfloor)
\end{aligned}
\tag{29}
$$

where the derivation of Jacobians of $\mathbf{J}_l(\cdot)$ and $\mathbf{J}_l(\cdot)$ are the following:

$$
\begin{aligned}
\mathbf{J}_l(\boldsymbol{\phi}) &= \mathbf{I} + \frac{1-\cos(\|\boldsymbol{\phi}\|)}{\|\boldsymbol{\phi}\|^2}\lfloor\boldsymbol{\phi}\rfloor + \frac{\|\boldsymbol{\phi}\| - \sin(\|\boldsymbol{\phi}\|)}{\|\boldsymbol{\phi}\|^3}\lfloor\boldsymbol{\phi}\rfloor^2 \\
\mathbf{J}_r(\boldsymbol{\phi}) &= \mathbf{I} - \frac{1-\cos(\|\boldsymbol{\phi}\|)}{\|\boldsymbol{\phi}\|^2}\lfloor\boldsymbol{\phi}\rfloor + \frac{\|\boldsymbol{\phi}\| - \sin(\|\boldsymbol{\phi}\|)}{\|\boldsymbol{\phi}\|^3}\lfloor\boldsymbol{\phi}\rfloor^2
\end{aligned}
$$

Time Offset: The encoder measurement received at time ${}^{L}t_k$ in Leg clock time was in fact occurred in IMU clock time ${}^{I}t_k$ and can be written as ${}^{I}t_k = {}^{L}t_k + {}^{L}t_I$. Using the current best estimate of the time offset ${}^{L}\hat{t}_I$, we can write the Leg clock time in corresponding IMU clock time as ${}^{I}t'_k = {}^{L}t_k + {}^{L}\hat{t}_I$ and considering the time offset error we can write as ${}^{I}t'_k = {}^{I}t_k + {}^{L}\tilde{t}_I$. The Jacobians are derived as follows:

(30)

We model this time offset in IMU pose estimate with the following first-order approximation by accounting for the time-offset estimation error:

$$
\begin{aligned}
{}^{I({}^{I}t'_k)}_{G}\mathbf{R} &= {}^{I({}^{I}t_k + {}^{L}\tilde{t}_I)}_{G}\mathbf{R} \\
&\approx (\mathbf{I} - \lfloor {}^{I({}^{I}t_k)}\boldsymbol{\omega}({}^{L}\tilde{t}_I)\rfloor)\,{}^{I({}^{I}t_k)}_{G}\mathbf{R} \\
&= (\mathbf{I} - \lfloor {}^{I_k}\boldsymbol{\omega}({}^{L}\tilde{t}_I)\rfloor)\,{}^{I_k}_{G}\mathbf{R} \\
{}^{G}\mathbf{p}_I &= {}^{G}\mathbf{p}_{I({}^{I}t_k + {}^{L}\tilde{t}_I)} \\
&\approx {}^{G}\mathbf{p}_{I({}^{I}t_k)} + {}^{G}\mathbf{v}_{I({}^{I}t_k)}\,{}^{L}\tilde{t}_I \\
&= {}^{G}\mathbf{p}_{I_k} + {}^{G}\mathbf{v}_{I_k}\,{}^{L}\tilde{t}_I
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{h}_\ell &= {}^{I}_{G}\mathbf{R}({}^{G}\mathbf{p}_T - {}^{G}\mathbf{p}_I) + {}^{L}_{I}\mathbf{R}^\top({}^{L}\mathbf{p}_I - {}^{L}\mathbf{p}_T) \\
\hat{\mathbf{h}}_\ell + \tilde{\mathbf{h}}_\ell &\approx (\mathbf{I} - \lfloor {}^{I_k}\boldsymbol{\omega}({}^{L}\tilde{t}_I)\rfloor)\,{}^{I_k}_{G}\mathbf{R}({}^{G}\mathbf{p}_T - {}^{G}\mathbf{p}_{I_k} + {}^{G}\mathbf{v}_{I_k}\,{}^{L}\tilde{t}_I) + {}^{L}_{I}\mathbf{R}^\top({}^{L}\mathbf{p}_I - {}^{L}\mathbf{p}_T) \\
&= {}^{I_k}_{G}\mathbf{R}({}^{G}\mathbf{p}_T - {}^{G}\mathbf{p}_{I_k}) + {}^{L}_{I}\mathbf{R}^\top({}^{L}\mathbf{p}_I - {}^{L}\mathbf{p}_T) - \lfloor {}^{I_k}\boldsymbol{\omega}({}^{L}\tilde{t}_I)\rfloor\,{}^{I_k}_{G}\mathbf{R}({}^{G}\mathbf{p}_T - {}^{G}\mathbf{p}_{I_k}) \\
&\quad + {}^{I_k}_{G}\mathbf{R}\,{}^{G}\mathbf{v}_{I_k}\,{}^{L}\tilde{t}_I + \lfloor {}^{I_k}\boldsymbol{\omega}({}^{L}\tilde{t}_I)\rfloor\,{}^{I_k}_{G}\mathbf{R}\,{}^{G}\mathbf{v}_{I_k}\,{}^{L}\tilde{t}_I \\
\tilde{\mathbf{h}}_\ell &= -\lfloor {}^{I_k}_{G}\mathbf{R}({}^{G}\mathbf{p}_T - {}^{G}\mathbf{p}_{I_k})\rfloor\,{}^{I_k}\boldsymbol{\omega}\,{}^{L}\tilde{t}_I + {}^{I_k}_{G}\mathbf{R}\,{}^{G}\mathbf{v}_{I_k}\,{}^{L}\tilde{t}_I \\
\frac{\partial \mathbf{h}_\ell}{\partial {}^{I}t_L} &= -\lfloor {}^{I_k}_{G}\mathbf{R}({}^{G}\mathbf{p}_T - {}^{G}\mathbf{p}_{I_k})\rfloor\,{}^{I_k}\boldsymbol{\omega} + {}^{I_k}_{G}\mathbf{R}\,{}^{G}\mathbf{v}_{I_k}
\end{aligned}
$$

A crucial aspect of this linearization is to properly compute Jacobian $\mathbf{G_n}$ with respect to the encoder noise. It is important to note that when evaluating the above linearization, we employ the first estimates Jacobian (FEJ) methodology [6, 7, 8], of the toe landmarks ${}^{G}\mathbf{p}_T$, to ensure estimation consistency.

In general scenarios, at this point, we are ready to use the above-legged measurement residual (19), together with the visual measurement residual as in VINS (see (15)), to perform tightly-coupled EKF update of both the body's motion states and the leg's kinematic parameters. In the following, we take special care for toe landmarks when a legged robot performs different motion gaits, which are of practical significance.

Note that the above EKF update with the leg measurements is only valid while the toe is in contact at the same point; otherwise, it could hurt our estimation performance. This reveals the fact that robust and accurate contact detection during walking is crucial for fusing leg information, with various contact detection methods being investigated [9, 10]. However, we take advantage of the proper legged kinematic model and consistent covariance estimates available from our MSCKF, and adopt the Mahalanobis distance test. Specifically, we perform the following threshold check to see if the toe is in the same position:

$$
\tilde{\mathbf{r}}_{\ell_k}^\top (\mathbf{H}_k\ \mathbf{P}_{k|k}\ \mathbf{H}_k^\top + \mathbf{G_n}\mathbf{R}'\mathbf{G_n}^\top)^{-1}\ \tilde{\mathbf{r}}_{\ell_k} < \chi^2 \tag{31}
$$

Table 1: Simulation parameters and prior single standard deviations were drawn from perturbations of measurements and initial states.

Parameter	Value	Parameter	Value
Cam Freq. (Hz)	10	IMU Freq. (Hz)	400
Leg Freq. (Hz)	50	Num. Clones	11
Pixel Proj. (px)	1	Leg White Noise (m)	2.0e-02
Gyro. White Noise	5.4e-04	Gyro. Rand. Walk	1.6e-05
Accel. White Noise	7.3e-03	Accel. Rand. Walk	6.6e-04
Num. Legs	4	Leg Toff. Ptrb. (sec)	8.0e-03
Leg Ext (Ori). Ptrb. (rad)	0.015	Leg Ext (Pos). Ptrb. (m)	3.0e-02

Table 2: *Jueying Lite 2* quadruped real-world experiment parameters, and prior single standard deviations were drawn from perturbations of measurements and initial states.

Parameter	Value	Parameter	Value
Cam Freq. (Hz)	25	IMU Freq. (Hz)	200
Leg Freq. (Hz)	50	Num. Clones	11
Pixel Proj. (px)	1	Leg encoder Noise (rad)	0.005
Link Length. Ptrb. (m)	0.010	Leg Toff. Ptrb. (sec)	0.003
Leg Ext (Ori). Ptrb. (rad)	0.017	Leg Ext (Pos). Ptrb. (m)	0.010

where $\mathbf{P}_{k|k}$ is the covariance of the augmented state, $\mathbf{H}_k$ consists of the Jacobians, and χ is the threshold for the test. When the test fails, we consider the toe is lifted from the ground and marginalize ${}^{G}\mathbf{p}_{T}$ from the state. On the other hand, the measurement will be used to update the state as a contact constraint via the legged kinematics.

3 Simulation Results

To verify the proposed online kinematic determination, we extended the visual-inertial simulator based on OpenVINS [3] to simulate quadruped motion and generate leg measurements. Note that we here focus on evaluating only the kinematic parameters of the spatial rigid transformation between the robot's body and leg and the time offset between the body IMU and leg encoders, which are body-leg spatiotemporal parameters. During the simulation, the position of the toe in the leg frame, ${}^{L}\mathbf{p}_{T_k}$, is directly simulated with additive white noise perturbations. We include our initial perturbation of spatiotemporal parameters to get the kinematic measurements as in (8). We performed 25 Monte-Carlo runs with different initial perturbations, see Tab. 1, and have shown a single representative leg result in Fig. 2. It is clear that the proposed method is able to accurately recover the spatiotemporal parameters within 20 seconds, and the estimated uncertainty (dashed lines are the estimator's 3σ bound) captures the true error distribution.

4 Experimental Results

We have also evaluated on two quadruped robots, *Jueying Lite 2* and *Ghost Vision 60*, which are equipped with a stereo camera and an IMU, along with four 3-link chain legs (see Fig.1). Two different motion scenarios are captured: (1) *Dance* in which toe contacts remain for the whole

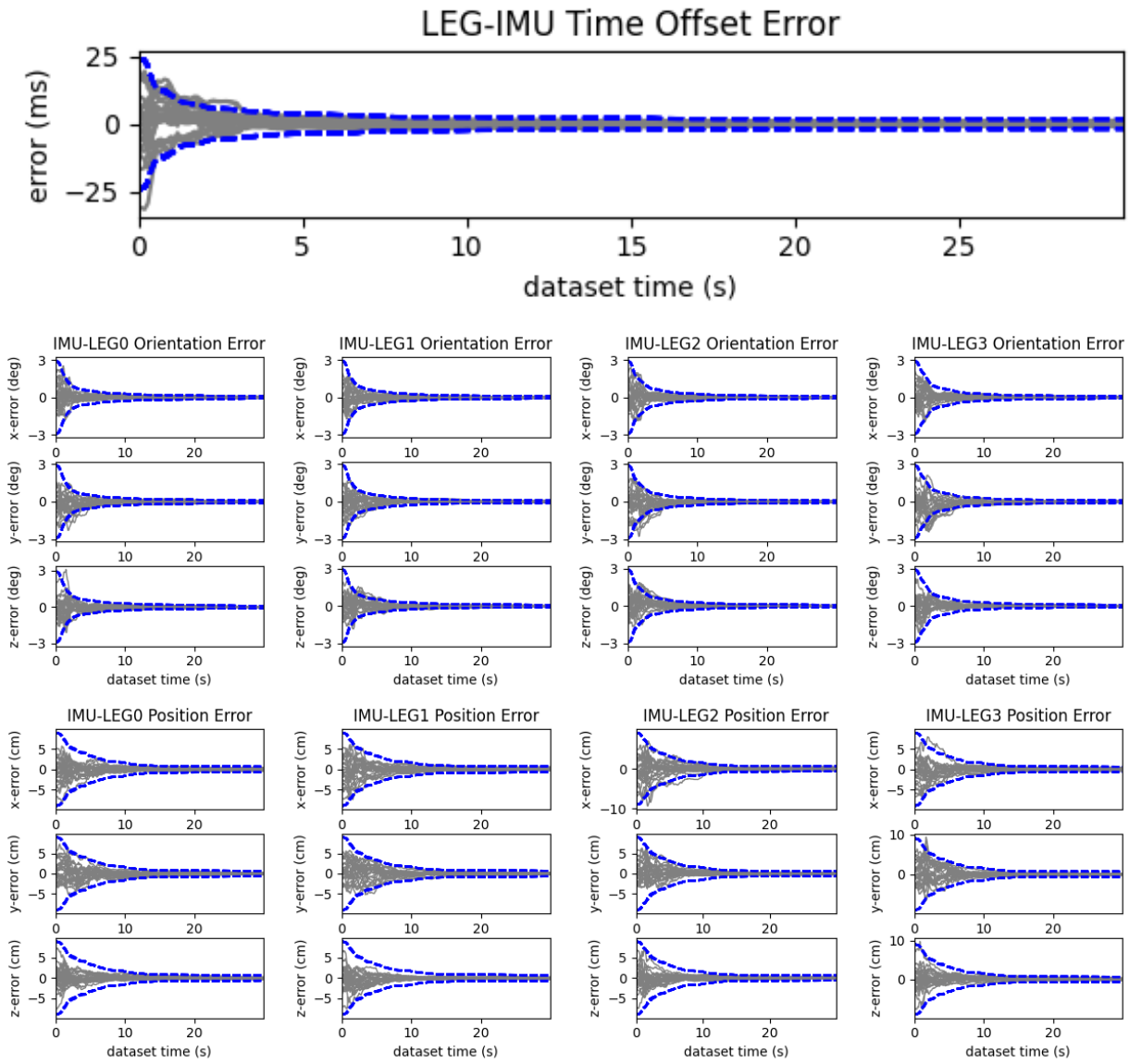


Figure 2: Extrinsic calibration errors (solid) and 3σ bound (dotted) for 25 different runs under random motion. Each solid line denotes a run with a different realization of the measurement noise and the initial values.

Table 3: Parameter repeatability over 25 runs with different initial conditions (see Table 2 for initial distributions) for *Jueying Lite 2* quadruped. All values reported are a single standard deviation after a 30-second dance for "Dance" dataset and a 30-second dance followed by two minutes of walk in "Dance_walk" dataset.

	Leg0 Rot (deg)	**Leg0 Pos (cm)**	**Leg0 Intrinsic (cm)**	**Toff (ms)**
Dance 0	$0.18 \pm 0.20, 0.87 \pm 0.11, -1.10 \pm 0.22$	$7.74 \pm 0.11, -10.02 \pm 0.13, 7.89 \pm 0.07$	$7.30 \pm 0.06, 19.00 \pm 0.09, 21.00 \pm 0.08$	-1.35 ± 1.55
Dance 1	$1.25 \pm 0.11, 1.01 \pm 0.09, -0.34 \pm 0.16$	$8.23 \pm 0.09, -10.01 \pm 0.16, 8.18 \pm 0.08$	$7.00 \pm 0.10, 19.00 \pm 0.09, 22.00 \pm 0.14$	-2.0 ± 0.77
Dance 2	$0.09 \pm 0.05, 0.94 \pm 0.07, -1.58 \pm 0.06$	$8.97 \pm 0.08, -10.29 \pm 0.15, 8.32 \pm 0.06$	$8.33 \pm 0.14, 19.00 \pm 0.06, 22.00 \pm 0.06$	-1.21 ± 1.14
Dance Walk 1	$-1.19 \pm 0.31, -2.9 \pm 0.12, -1.18 \pm 0.29$	$9.82 \pm 0.33, -10.44 \pm 0.26, 6.86 \pm 0.27$	$10.0 \pm 0.30, 21.0 \pm 0.16, 24.0 \pm 0.15$	-0.50 ± 0.64
Dance walk 2	$-0.54 \pm 0.32, -3.93 \pm 0.19, -1.21 \pm 0.23$	$9.40 \pm 0.23, -9.44 \pm 0.23, 5.50 \pm 0.17$	$9.20 \pm 0.18, 20.09 \pm 0.20, 23.00 \pm 0.28$	0.09 ± 0.73

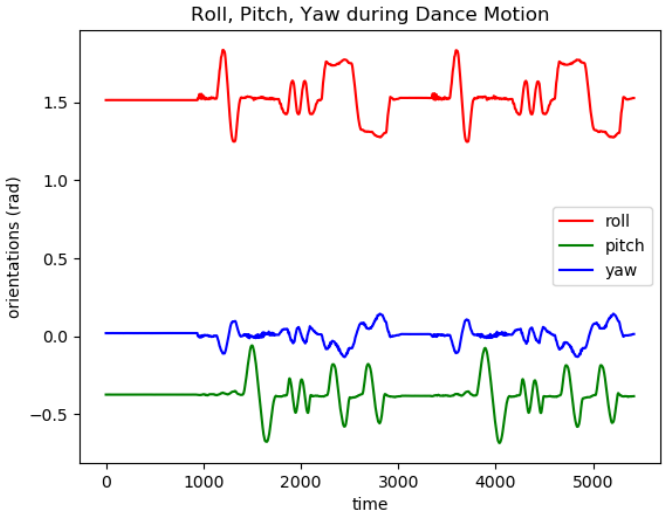


Figure 3: Roll, Pitch, Yaw motion of robot's base during Dance motion of *Jueying Lite 2* Quadruped

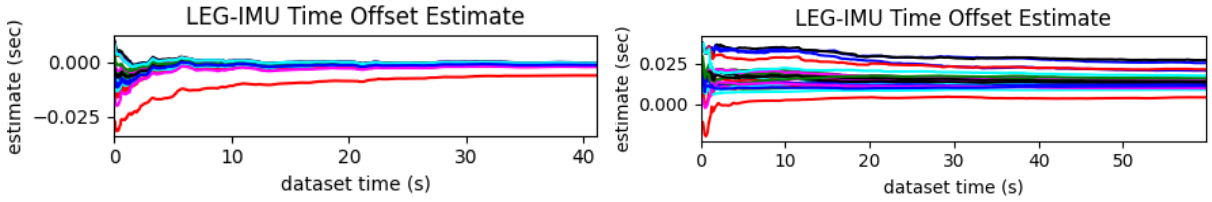


Figure 4: Real world Estimation results of the time offset parameter for the Dance dataset on *Left*: **JYE Robot** and *Right*: **Ghost Robot**.

Table 4: Absolute Trajectory Error (ATE) of each algorithm on *Jueying Lite 2*. The best results are highlighted in bold font.

	Leg-VINS wo. Kine	**Leg-VINS w. Kine**
Dance Walk 1	6.596 / **0.387**	**5.466** / 0.437
Dance Walk 2	7.258 / 0.681	**6.314** / **0.495**
Dance Walk 3	13.317 / 1.158	**10.497** / **0.822**
Dance Walk 4	6.736 / 0.550	**6.509** / **0.433**
Average	8.476 / 0.694	**7.196** / **0.547**

Table 5: *Ghost Vision 60* quadruped real-world experiment parameters, and prior single standard deviations were drawn from perturbations of measurements and initial states.

Parameter	**Value**	**Parameter**	**Value**
Cam Freq. (Hz)	30	IMU Freq. (Hz)	500
Leg Freq. (Hz)	230	Num. Clones	11
Pixel Proj. (px)	1	Leg encoder Noise (rad)	0.006
Link Length. Ptrb. (m)	0.04	Leg Toff. Ptrb. (sec)	0.004
Leg Ext (Ori). Ptrb. (rad)	0.043	Leg Ext (Pos). Ptrb. (m)	0.03

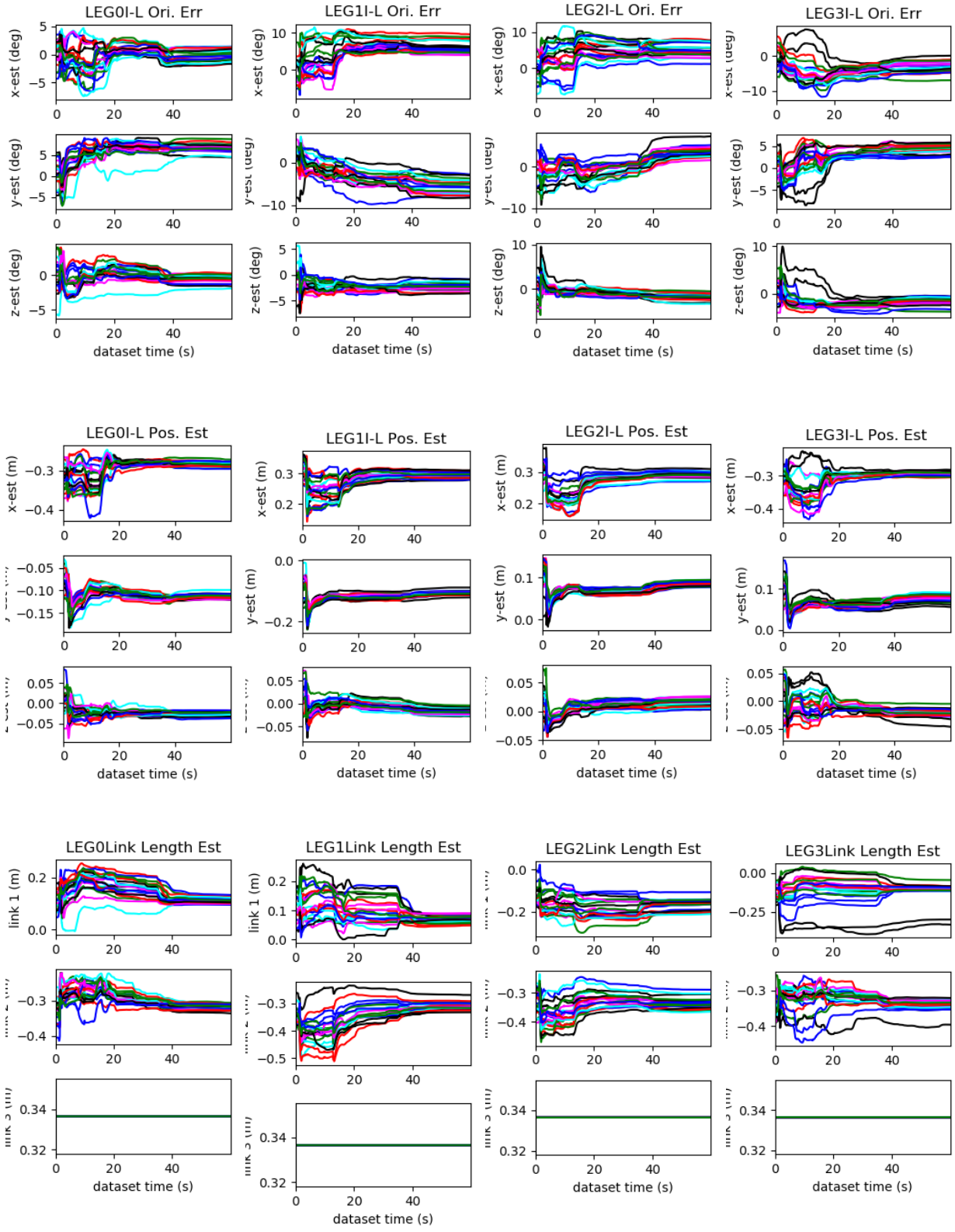


Figure 5: Real world Estimation results of the legged kinematic parameters for the Dance dataset on **Ghost Vision 60 Robot**. Each color denotes runs with different initial guesses of the calibration parameters (a total of 25 runs). *Top row*: Orientation results, *Mid row*: Position results, *Bottom row*: Link lengths Calibration results.

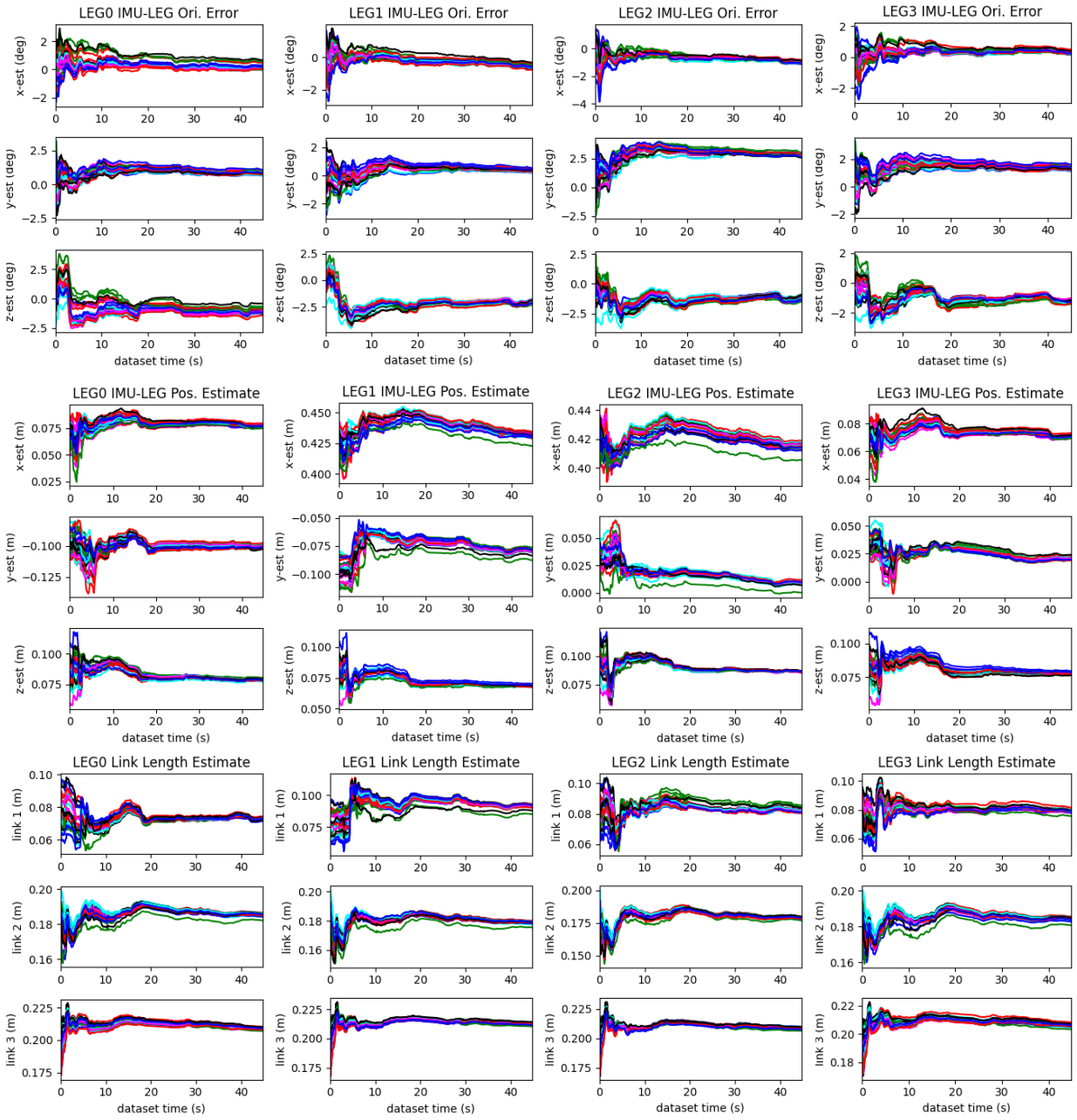


Figure 6: Real world Estimation results of the legged kinematic parameters for the Dance dataset on **JYE Robot**. Each color denotes runs with different initial guesses of the calibration parameters (a total of 25 runs). *Top row*: Orientation results, *Mid row*: Position results, *Bottom row*: Link lengths Calibration results.

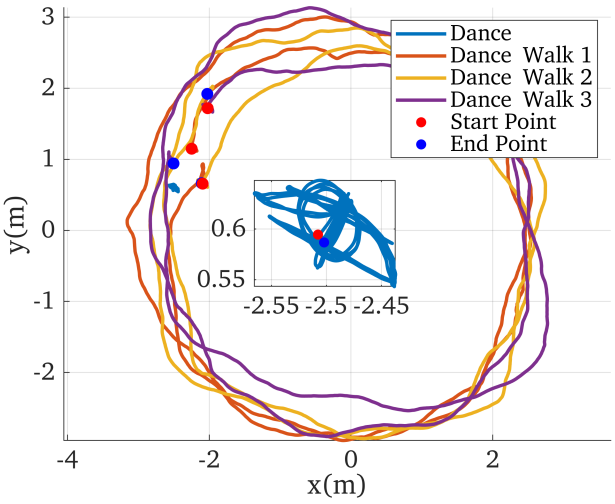


Figure 7: Trajectories of Dance and Dance Walk datasets of *JYE Robot* are shown in this figure.

Table 6: Parameter repeatability over 25 runs with different initial conditions (see Table 5 for initial distributions) for *Ghost Vision 60* quadruped. All values reported are a single standard deviation after a 30-second dance for "Dance" dataset and a 30-second dance followed by two minutes of walk in "Dance_walk" dataset.

	Leg0 Rot (deg)	**Leg0 Pos (cm)**	**Leg0 Intrinsic (cm)**	**Toff (ms)**
Dance 2	$2.67 \pm 1.63, -5.86 \pm 1.57, -0.97 \pm 0.69$	$29.57 \pm 0.99, -11.48 \pm 0.58, -2.35 \pm 0.42$	$10.34 \pm 0.01, -32.67 \pm 0.02, 33.64 \pm 000$	8.54 ± 0.87
Dance 3	$1.18 \pm 1.17, -5.38 \pm 1.28, -0.23 \pm 0.47$	$28.71 \pm 0.82, -9.87 \pm 0.55, -3.19 \pm 0.71$	$11.41 \pm 0.02, -32.49 \pm 0.01, 33.64 \pm 000$	16.003 ± 2.41
Dance 4	$2.59 \pm 2.47, -8.86 \pm 1.02, -1.06 \pm 0.84$	$28.68 \pm 0.53, -9.68 \pm 0.47, -2.42 \pm 0.51$	$9.71 \pm 0.02, -31.95 \pm 0.012, 33.64 \pm 000$	15.01 ± 1.43
Dance Walk 1	$2.03 \pm 6.71, -13.56 \pm 0.82, 0.66 \pm 1.39$	$30.87 \pm 1.09, -13.16 \pm 0.80, -1.47 \pm 2.43$	$9.46 \pm 0.06, -39.48 \pm 0.01, 33.64 \pm 000$	-3.58 ± 0.44
Dance walk 2	$-1.24 \pm 3.25, -6.57 \pm 0.94, -0.40 \pm 0.76$	$28.69 \pm 0.71, -10.21 \pm 0.40, -3.37 \pm 0.61$	$11.99 \pm 0.03, -30.65 \pm 0.01, 33.64 \pm 000$	3.44 ± 0.65

duration and enables evaluation of the calibration repeatability, and (2) *Dance-Walk* where after a dancing motion, the robot walks within a motion capture room so that we can quantify the impact of good and bad calibration of localization performance after parameter identification. For Dancing motion, we provided different combinations of roll, pitch, and yaw commands while keeping their feet stationary at all times. There is no particular dancing motion required, any combination that thoroughly excites the joints and body will suffice. A reference plot of robot's bodies roll, pitch, yaw during dance motion is shown with reference to measurements in Fig. 3. All *Dance* sequences are about 25-40 seconds and *Dance-Walk* sequences are about 2 minutes; their trajectory plots can be seen in Fig. 7.

We evaluated the proposed method with *Dance* sequence to first see the performance without being affected by the contact detection problem. The kinematic results are shown in Fig. 6 and Fig.

Table 7: Absolute Trajectory Error (ATE) of each algorithm on *Ghost Vision 60*. The best results are highlighted in bold font.

	Leg-VINS wo. Kine	**Leg-VINS w. Kine**
Dance Walk 2	125.302 / 1.690	125.096 / 1.719
Dance Walk 4	120.378 / 1.887	121.463 / 1.920
Dance Walk 5	136.779 / 2.153	133.740 / 2.128
Dance Walk 6	124.580 / 2.170	124.030 / 2.171
Average	126.760 / 1.975	126.083 / 1.984

Table 8: Parameter repeatability over 25 runs with different initial conditions (see Table 2 for initial distributions) for *Jueying Lite 2* quadruped. All values reported are a single standard deviation after a 30-second dance for "Dance" dataset and a 30-second dance followed by two minutes of walk in "Dance_walk" dataset.

Jueying Lite 2	**Leg0 Rot (deg)**	**Leg0 Pos (cm)**	**Leg0 Intrinsic (cm)**	**Toff (ms)**
Dance 0	$0.18 \pm 0.20, 0.87 \pm 0.11, -1.10 \pm 0.22$	$7.74 \pm 0.11, -10.02 \pm 0.13, 7.89 \pm 0.07$	$7.30 \pm 0.06, 19.00 \pm 0.09, 21.00 \pm 0.08$	-1.35 ± 1.55
Dance 1	$1.25 \pm 0.11, 1.01 \pm 0.09, -0.34 \pm 0.16$	$8.23 \pm 0.09, -10.01 \pm 0.16, 8.18 \pm 0.08$	$7.00 \pm 0.10, 19.00 \pm 0.09, 22.00 \pm 0.14$	-2.0 ± 0.77
Dance 2	$0.09 \pm 0.05, 0.94 \pm 0.07, -1.58 \pm 0.06$	$8.97 \pm 0.08, -10.29 \pm 0.15, 8.32 \pm 0.06$	$8.33 \pm 0.14, 19.00 \pm 0.06, 22.00 \pm 0.06$	-1.21 ± 1.14
Dance Walk 1	$-1.19 \pm 0.31, -2.9 \pm 0.12, -1.18 \pm 0.29$	$9.82 \pm 0.33, -10.44 \pm 0.26, 6.86 \pm 0.27$	$10.0 \pm 0.30, 21.0 \pm 0.16, 24.0 \pm 0.15$	-0.50 ± 0.64
Dance walk 2	$-0.54 \pm 0.32, -3.93 \pm 0.19, -1.21 \pm 0.23$	$9.40 \pm 0.23, -9.44 \pm 0.23, 5.50 \pm 0.17$	$9.20 \pm 0.18, 20.09 \pm 0.20, 23.00 \pm 0.28$	0.09 ± 0.73
Ghost Vision 60	**Leg0 Rot (deg)**	**Leg0 Pos (cm)**	**Leg0 Intrinsic (cm)**	**Toff (ms)**
Dance 2	$2.67 \pm 1.63, -5.86 \pm 1.57, -0.97 \pm 0.69$	$29.57 \pm 0.99, -11.48 \pm 0.58, -2.35 \pm 0.42$	$10.34 \pm 0.01, -32.67 \pm 0.02, 33.64 \pm 000$	8.54 ± 0.87
Dance 3	$1.18 \pm 1.17, -5.38 \pm 1.28, -0.23 \pm 0.47$	$28.71 \pm 0.82, -9.87 \pm 0.55, -3.19 \pm 0.71$	$11.41 \pm 0.02, -32.49 \pm 0.01, 33.64 \pm 000$	16.003 ± 2.41
Dance 4	$2.59 \pm 2.47, -8.86 \pm 1.02, -1.06 \pm 0.84$	$28.68 \pm 0.53, -9.68 \pm 0.47, -2.42 \pm 0.51$	$9.71 \pm 0.02, -31.95 \pm 0.012, 33.64 \pm 000$	15.01 ± 1.43
Dance Walk 1	$2.03 \pm 6.71, -13.56 \pm 0.82, 0.66 \pm 1.39$	$30.87 \pm 1.09, -13.16 \pm 0.80, -1.47 \pm 2.43$	$9.46 \pm 0.06, -39.48 \pm 0.01, 33.64 \pm 000$	-3.58 ± 0.44
Dance walk 2	$-1.24 \pm 3.25, -6.57 \pm 0.94, -0.40 \pm 0.76$	$28.69 \pm 0.71, -10.21 \pm 0.40, -3.37 \pm 0.61$	$11.99 \pm 0.03, -30.65 \pm 0.01, 33.64 \pm 000$	3.44 ± 0.65

5 for *Jueying Lite 2* and *Ghost Vision 60* respectively and the timeoffset results are shown in Fig. 4. For each run, we perturb the manufacturer-provided calibration (URDF) by the values in Tab. 2 and Tab. 5, and can see that most parameters quickly converged after 25 seconds of calibration start. The converged values are near our best estimate and have a low variance compared to prior perturbation variance, indicating the proposed system's fast and reliable identification performance. To quantify the repeatability, we look at the variance of the final estimated value in Tab. 8 for a representative leg on each dataset. This shows that we are able to calibrate within sub-degree, sub-centimeter, and 1-2ms repeatability using the proposed estimator. The *Dance Walk* datasets do have higher variance, which we attribute to evaluation after walking, which is affected by the contact detection accuracy. We did find that the time offset varied over each dataset, which is consistent with the observation by Yang et al. [11] within the visual-inertial temporal calibration context for the Realsense T265 stereo camera.

We further evaluate the estimation performance on all sequences (*Dance Walk 1-4*), in terms of localization accuracy, by comparing estimation with and without fusing legged kinematics (tagged Leg-VINS w. Kine and Leg-VINS wo. Kine, respectively). Note that we set the same initial kinematic values for all compared methods. The Absolute Trajectory Errors (ATE) [12] are shown in Tab. 7, which clearly show that the proposed Leg-VINS w. Kine improves the localization accuracy by properly incorporating the legged kinematic constraints.

References

[1] Christoph Hertzberg, René Wagner, Udo Frese, and Lutz Schröder. "Integrating generic sensor fusion algorithms with sound state representations through encapsulation of manifolds". In: *Information Fusion* 14.1 (2013), pp. 57–77.

[2] Gregory Chirikjian. *Stochastic Models, Information Theory, and Lie Groups, Volume 2: Analytic Methods and Modern Applications*. Vol. 2. Springer Science & Business Media, 2011.

[3] Patrick Geneva, Kevin Eckenhoff, Woosik Lee, Yulin Yang, and Guoquan Huang. "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.

[4] Averil Burton Chatfield. *Fundamentals of high accuracy inertial navigation*. Vol. 174. Aiaa, 1997.

[5] Yulin Yang, Patrick Geneva, Xingxing Zuo, and Guoquan Huang. "Online Self-Calibration for Visual-Inertial Navigation: Models, Analysis, and Degeneracy". In: *IEEE Transactions on Robotics* (2023).

[6] Guoquan Huang, Anastasios I. Mourikis, and Stergios I. Roumeliotis. "Analysis and Improvement of the Consistency of Extended Kalman Filter-Based SLAM". In: *Proc. of the IEEE International Conference on Robotics and Automation*. Pasadena, CA, 2008.

[7] Guoquan Huang, Anastasios I. Mourikis, and Stergios I. Roumeliotis. "A First-Estimates Jacobian EKF for Improving SLAM Consistency". In: *Proc. of the 11th International Symposium on Experimental Robotics*. Athens, Greece, July 2008.

[8] Guoquan Huang, Anastasios I. Mourikis, and Stergios I. Roumeliotis. "Observability-based Rules for Designing Consistent EKF SLAM Estimators". In: *International Journal of Robotics Research* (2010).

[9] Marco Camurri, Maurice Fallon, Stéphane Bazeille, Andreea Radulescu, Victor Barasuol, Darwin G Caldwell, and Claudio Semini. "Probabilistic contact estimation and impact detection for state estimation of quadruped robots". In: *IEEE Robotics and Automation Letters (RA-L)* (2017).

[10] Qingyu Liu, Bing Yuan, and Yang Wang. "Online learning for foot contact detection of legged robot based on data stream clustering". In: *Frontiers in Bioengineering and Biotechnology* (2022).

[11] Yulin Yang, Patrick Geneva, and Guoquan Huang. "Multi-visual-inertial system: Analysis, calibration and estimation". In: *arXiv preprint arXiv:2308.05303* (2023).

[12] Zichao Zhang and Davide Scaramuzza. "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2018.