
Robocentric Visual-Inertial Odometry

Zheng Huai - zhuai@udel.edu
Guoquan Huang - ghuang@udel.edu

Department of Mechanical Engineering
University of Delaware, Delaware, USA

RPNG

Robot Perception and Navigation Group (RPNG)
Tech Report - RPNG-2018-RVIO
Last Updated - April 4, 2018

Contents

1	Introduction	1
2	Robocentric VIO	2
2.1	State Vector	2
2.2	Propagation	3
2.2.1	Continuous Time Error-State Model	3
2.2.2	Discrete Time State Propagation	6
2.2.3	Discrete Time Covariance Propagation	7
2.3	Update	8
2.3.1	Bearing-only Model	8
2.3.2	Inverse Depth Bearing-only Model	10
2.3.3	EKF Update	13
2.3.4	State Augmentation	14
2.4	Composition	14
2.5	Observability Analysis	15
2.5.1	Analytic State Transition Matrix	16
2.5.2	Measurement Jacobian	17
2.5.3	Propagation and Update	17
2.5.4	Composition	18
2.6	Initialization	20
3	Simulation Results	20
4	Experimental Results	20
4.1	Highly Dynamic Capability	22
4.2	Long-Term Sustainability	22
4.3	Versatility	22
5	Conclusion	22
Appendix A	Analytic State Transition Matrix	25
Appendix B	Bundle Adjustment with Euclidean Coordinates	27
Appendix C	Bundle Adjustment with Inverse Depth Parameters	30
References		31

Abstract

In this report, we propose a novel *robocentric* formulation of visual-inertial navigation systems (VINS) within a multi-state constraint Kalman filter (MSCKF) framework and develop an efficient, lightweight, *robocentric visual-inertial odometry* (R-VIO) algorithm for consistent localization in challenging environments using only monocular vision. The key idea of the proposed approach is to deliberately reformulate the 3D VINS with respect to a moving local frame (i.e., robocentric), rather than a fixed global frame of reference as in the standard world-centric VINS, and instead utilize high-accuracy relative motion estimates for global pose update. As an immediate advantage of using this robocentric formulation, the proposed R-VIO can start from an arbitrary pose, *without* the need to align its orientation with the global gravity vector. More importantly, we analytically show that the proposed robocentric EKF-based VINS does *not* undergo the observability mismatch issue as in the standard world-centric framework which was identified as the main cause of estimation inconsistency. The proposed R-VIO is extensively tested through both Monte Carlo simulations and the real-world experiments with different sensor platforms in different environments, and shown to achieve a competitive performance with the state-of-the-art VINS algorithms in terms of consistency and accuracy.

1 Introduction

Enabling high-precision 3D navigation on mobile devices and robots with minimal sensing of low-cost sensors holds potentially huge implications in the real applications ranging from mobile augmented reality to autonomous driving. To this end, inertial navigation offers a classical 3D localization solution which utilizes an inertial measurement unit (IMU) measuring the 3 degree-of-freedom (DOF) angular velocity and linear acceleration of the sensor platform on which it is rigidly attached. Typically, IMU works at a high frequency (e.g., 100Hz~1000Hz) that enable it to sense highly dynamic motion, however, due to the corrupting sensor noise and bias, purely integrating IMU measurements can easily result in unusable motion estimates. This necessitates to fuse the aiding information from *at least* a single camera to reduce the accumulated inertial navigation drifts, which comes into the visual-inertial navigation systems (VINS).

In the past decade, we have witnessed significant progress on VINS, including both visual-inertial SLAM (VI-SLAM) and visual-inertial odometry (VIO), and many different algorithms have been developed (e.g., [1, 2, 3, 4, 5, 6, 7, 8] and references therein). However, almost all these algorithms are based on the standard *world-centric* formulation – that is, to estimate the absolute motion with respect to a fixed global frame of reference, such as the earth-centered earth-fixed (ECEF) or the north-east-down (NED) frame. In order to achieve accurate localization, such world-centric VINS often require a particular initialization procedure to estimate the starting pose in the global frame of reference, which, however, is hard to guarantee the accuracy in some cases (e.g., quick start, or poor vision). While an extended Kalman filter (EKF)-based world-centric VINS algorithm has the advantage of lower computational cost [1, 4] in comparing to the batch optimization-based ones (which incur high computation due to performing iterated linearization [5, 6]), it may become *inconsistent*, primarily due to the fact that the EKF linearized systems have different observability properties from the corresponding underlying nonlinear systems [9, 10, 4]. The remedies for mitigating this issue include enforcing the correct observability [4, 11, 12] or employing an invariant error representation [13]. Therefore, one may ask: *Do we have to formulate VINS in the standard world-centric form?* The answer is *no*. Intuitively, inspired by how humans navigate – we may not remember the starting pose after traveling a long distance while knowing well the relative motion within a recent, short time interval – we might relax the fixed global frame of VINS, instead, choose a moving local frame to better estimate relative motion which can be used later for global pose update.

Note that, this sensor-centered idea for localization can be traced back to 2D laser-based robocentric mapping [14], where the global frame is treated as a feature being observed from the moving robot frame and the relative motion measurements from an odometer are fused for pose update, while the composition step makes it possible to shift the local frame of reference during the motion. Following the similar idea, in [15] a camera-centered formulation shows the potential to fuse the visual information with the measurements from proprioceptive sensors (e.g., angular and linear velocity measurements). Both methods had been applied to the EKF-based SLAM while performing mapping in the local frames, thus limiting the global uncertainty and improving the estimation consistency. It should also be noted that a robust VINS algorithm using a different robocentric formulation and sensor-fusion scheme was recently introduced [16]. In particular, its

state vector includes both IMU states and the features which are reformulated with respect to the local IMU frame, while the camera and IMU measurements are fused in a *direct* fashion. Additionally, in contrast to [14, 15], it employs the *iterated* EKF to perform update, without the composition step to shift the local frame of reference.

In this report, following the idea [14, 15], we reformulate the VINS in 3D with respect to a local IMU frame, while in contrast to [14, 15, 16] which keep features in the state vector and have to concern the increasing computational cost as more features are observed and included, we focus on an EKF-based visual-inertial odometry (EKF-VIO) framework, i.e., multi-state constraint Kalman filter (MSCKF) [1], whose stochastic cloning enables the VINS to process hundreds of features while only keep a small number of robot poses from which the features have been observed, in the state vector, thus significantly reducing the computational cost. However, as studied in [4], the world-centric MSCKF is inconsistent. To enable consistent MSCKF-based 3D localization, a novel lightweight, robocentric VIO algorithm (R-VIO) is proposed in this report with the following keypoints:

- The global frame has been treated as a feature which involves the *gravity* effect, while the local frame of reference is shifted at every image time through a *composition* step.
- The relative motion estimates used for global updates are obtained by *tightly* fusing the camera and IMU measurements in a local frame of reference, for which instead of the features, a sliding *relative pose* window is included in the state vector to aid the estimation.
- A tailored *inverse-depth* measurement model is developed, which allows for fusing bearing information provided by the distant features, especially at motionless state.
- A *constant* unobservable subspace is analytically shown with the proposed robocentric formulation, which is independent of the linearization points while possessing correct dimensions and desired unobservable directions.

We perform extensive tests on both Monte Carlo simulations and the real-world experiments running real data of different sensor platforms from the micro aerial vehicle (MAV) flying indoor to the hand-held device navigating outdoor. All real-time results validate the superior performance of the proposed R-VIO algorithm.

2 Robocentric VIO

2.1 State Vector

In the proposed R-VIO, the robot frame $\{R\}$ is chosen as the immediate frame of reference for state estimation. For robot equipped with an IMU, the corresponding frame $\{I\}$ is usually considered to be aligned with the robot frame. As a result, the global frame $\{G\}$ ($=\{R_0\}$, the first robot frame) turns into a "moving" object with respect to $\{R\}$, and as compared to the world-centric case (e.g. [1]), we have a sequential frames of reference as the robot is moving.

At time-step $\tau \in [t_k, t_{k+1}]$ the robocentric state with respect to $\{R_k\}$ (set at time-step k) consists of the global state and the IMU state, where the global state maintains the information of the start pose, while the IMU state characterizes the motion from $\{R_k\}$ to the current IMU pose, which is defined by the vector: ¹

$${}^{R_k} \mathbf{x}_\tau = \begin{bmatrix} {}^{R_k} \mathbf{x}_G^\top & {}^{R_k} \mathbf{x}_{I_\tau}^\top \end{bmatrix}^\top \quad (1)$$

$$= \begin{bmatrix} {}^G \bar{q}^\top & {}^{R_k} \mathbf{p}_G^\top & {}^{R_k} \mathbf{g}^\top & | & {}^{\tau} \bar{q}^\top & {}^{R_k} \mathbf{p}_{I_\tau}^\top & \mathbf{v}_{I_\tau}^\top & \mathbf{b}_{g_\tau}^\top & \mathbf{b}_{a_\tau}^\top \end{bmatrix}^\top \quad (2)$$

¹Throughout this report, $k, k+1, \dots$ indicate the image time-steps, while $\tau, \tau+1, \dots$ are the IMU time-steps between every two consecutive images. $\{I\}$ and $\{C\}$ denote the IMU frame and camera frame, respectively, $\{R\}$ is the robocentric frame of reference which is selected with the corresponding IMU frame at every image time-step. The subscript $\ell|i$ refers to the estimate of a quantity at time-step ℓ , after all measurements up to time-step i have been processed. \hat{x} is used to denote the estimate of a random variable x , while $\tilde{x} = x - \hat{x}$ is the additive error in this estimate. \mathbf{I}_n and $\mathbf{0}_n$ are the $n \times n$ identity and zero matrices, respectively. Finally, the left superscript denotes the frame of reference with respect to which the vector is expressed.

where ${}^k_G\bar{q}$ is the unit quaternion [17] describing the rotation from $\{R_k\}$ to $\{G\}$, ${}^{R_k}\mathbf{p}_G$ is the position of $\{G\}$ in $\{R_k\}$, ${}^\tau_k\bar{q}$ and ${}^{R_k}\dot{\mathbf{p}}_{I_\tau}$ are the relative rotation and translation of robot from $\{R_k\}$ to $\{I_\tau\}$, and \mathbf{v}_{I_τ} is the robot velocity in the frame $\{I_\tau\}$. In particular, the local gravity vector is included which implicitly encodes the orientation of $\{R_k\}$. Following Eq. (2), the error-state vector is defined as:

$${}^{R_k}\tilde{\mathbf{x}}_\tau = \begin{bmatrix} {}^{R_k}\tilde{\mathbf{x}}_G^\top & {}^{R_k}\tilde{\mathbf{x}}_{I_\tau}^\top \end{bmatrix}^\top \quad (3)$$

$$= \begin{bmatrix} \delta\boldsymbol{\theta}_G^\top & {}^{R_k}\tilde{\mathbf{p}}_G^\top & {}^{R_k}\tilde{\mathbf{g}}^\top & | & \delta\boldsymbol{\theta}_\tau^\top & {}^{R_k}\tilde{\mathbf{p}}_{I_\tau}^\top & \tilde{\mathbf{v}}_{I_\tau}^\top & \tilde{\mathbf{b}}_{g_\tau}^\top & \tilde{\mathbf{b}}_{a_\tau}^\top \end{bmatrix}^\top \quad (4)$$

where, in particular, if the quaternion estimate is \hat{q} , then the error quaternion $\delta\bar{q}$ is defined by:

$$\bar{q} = \delta\bar{q} \otimes \hat{q} \quad (5)$$

$$\delta\bar{q} \simeq \left[\frac{1}{2}\delta\boldsymbol{\theta}^\top \quad 1 \right]^\top, \quad \mathbf{C}(\delta\bar{q}) = \mathbf{I}_{3 \times 3} - [\delta\boldsymbol{\theta} \times] \quad (6)$$

where \otimes denotes the quaternion multiplication, $\delta\boldsymbol{\theta}$ is 3 degree-of-freedom (DOF) error angle associated to the error quaternion, and $\mathbf{C}(\cdot)$ represents a 3×3 rotation matrix with $[\cdot \times]$ the skew-symmetric operator.

Assuming that at time-step k when $\{I_k\}$ is set to be the frame of reference (i.e., $\{R_k\}$), a window of the relative poses between the last N robocentric frames of reference is included in the state vector:

$$\hat{\mathbf{x}}_k = \begin{bmatrix} {}^{R_k}\hat{\mathbf{x}}_k^\top & \tilde{\mathbf{w}}_k^\top \end{bmatrix}^\top \quad (7)$$

$$= \begin{bmatrix} {}^{R_k}\hat{\mathbf{x}}_G^\top & {}^{R_k}\hat{\mathbf{x}}_{R_k}^\top & | & {}^2\hat{q}^\top & {}^{R_1}\hat{\mathbf{p}}_{R_2}^\top & \dots & {}^i\hat{q}^\top & {}^{R_{i-1}}\hat{\mathbf{p}}_{R_i}^\top & \dots & {}^N\hat{q}^\top & {}^{R_{N-1}}\hat{\mathbf{p}}_{R_N}^\top \end{bmatrix}^\top \quad (8)$$

where ${}^i\hat{q}$ and ${}^{R_{i-1}}\hat{\mathbf{p}}_{R_i}$ are cloned from the IMU state estimate ${}^{R_{i-1}}\hat{\mathbf{x}}_{R_i}$, $i \in \{2, \dots, N\}$ (cf. Section 2.3.4). The error-state vector is defined accordingly:

$$\tilde{\mathbf{x}}_k = \begin{bmatrix} {}^{R_k}\tilde{\mathbf{x}}_k^\top & \tilde{\mathbf{w}}_k^\top \end{bmatrix}^\top \quad (9)$$

$$= \begin{bmatrix} {}^{R_k}\tilde{\mathbf{x}}_G^\top & {}^{R_k}\tilde{\mathbf{x}}_{R_k}^\top & | & \delta\boldsymbol{\theta}_2^\top & {}^{R_1}\tilde{\mathbf{p}}_{R_2}^\top & \dots & \delta\boldsymbol{\theta}_i^\top & {}^{R_{i-1}}\tilde{\mathbf{p}}_{R_i}^\top & \dots & \delta\boldsymbol{\theta}_N^\top & {}^{R_{N-1}}\tilde{\mathbf{p}}_{R_N}^\top \end{bmatrix}^\top \quad (10)$$

Specifically, if $i \in \{k-N+1, k-N+2, \dots, k\}$, then this window represents a short memory of the system motion. With such interpretation, everything happens at time-step k could be correlated with the previous time-steps in the window, which in next is shown in the design of the measurement model (cf. Section 2.3). In our implementation, we manage this in a sliding-window fashion, i.e., delete the oldest one when a new relative pose is cloned, to save the computational cost.

2.2 Propagation

2.2.1 Continuous Time Error-State Model

The typical measurements of IMU consist of the gyroscope and accelerometer measurements, $\boldsymbol{\omega}_m$ and \mathbf{a}_m , which are respectively given as:

$$\boldsymbol{\omega}_m = \boldsymbol{\omega} + \mathbf{b}_g + \mathbf{n}_g \quad (11)$$

$$\mathbf{a}_m = {}^I\mathbf{a} + {}^I\mathbf{g} + \mathbf{b}_a + \mathbf{n}_a \quad (12)$$

where $\boldsymbol{\omega}$ and ${}^I\mathbf{a}$ are the angular velocity and linear acceleration in the IMU frame, \mathbf{n}_g and \mathbf{n}_a are the zero-mean white Gaussian sensor noise, \mathbf{b}_g and \mathbf{b}_a are the gyroscope and accelerometer biases which are modeled as random walk processes, driven by the zero-mean white Gaussian noise \mathbf{n}_{wg} and \mathbf{n}_{wa} , respectively, and ${}^I\mathbf{g}$ denotes the local gravity vector, whose global counterpart is constant (e.g., ${}^G\mathbf{g} = [0, 0, 9.8]^\top$, if $\{G\}$ is an earth-centered frame).

The continuous time IMU model with respect to the frame of reference $\{R_k\}$ is shown as follows:

$${}^\tau_k\dot{\bar{q}} = \frac{1}{2}\boldsymbol{\Omega}(\boldsymbol{\omega})^\tau_k\bar{q} \quad (13)$$

$${}^{R_k}\dot{\mathbf{p}}_{I_\tau} = \mathbf{C}({}^\tau_k\bar{q})^\top \mathbf{v}_{I_\tau} \quad (14)$$

$$\dot{\mathbf{v}}_{I_\tau} = {}^\tau \mathbf{a} - [\boldsymbol{\omega} \times] \mathbf{v}_{I_\tau} \quad (15)$$

$$\dot{\mathbf{b}}_g = \mathbf{n}_{wg} \quad (16)$$

$$\dot{\mathbf{b}}_a = \mathbf{n}_{wa} \quad (17)$$

and for $\boldsymbol{\omega} = [\omega_x, \omega_y, \omega_z]^\top$, we have:

$$\boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} -[\boldsymbol{\omega} \times] & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^\top & 1 \end{bmatrix}, \quad [\boldsymbol{\omega} \times] = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \quad (18)$$

Especially, the derivation for local velocity (cf. Eq. (15)) is:

$$\mathbf{v}_{I_\tau} = \mathbf{C}({}^\tau_k \bar{q})^{R_k} \mathbf{v}_{I_\tau} \quad (19)$$

$$\Rightarrow \dot{\mathbf{v}}_{I_\tau} = \mathbf{C}({}^\tau_k \bar{q})^{R_k} \dot{\mathbf{v}}_{I_\tau} + \dot{\mathbf{C}}({}^\tau_k \bar{q})^{R_k} \mathbf{v}_{I_\tau} = \mathbf{C}({}^\tau_k \bar{q})^{R_k} \mathbf{a}_{I_\tau} - [\boldsymbol{\omega} \times] \mathbf{C}({}^\tau_k \bar{q})^{R_k} \mathbf{v}_{I_\tau} = {}^\tau \mathbf{a} - [\boldsymbol{\omega} \times] \mathbf{v}_{I_\tau} \quad (20)$$

The continuous time IMU state propagation with respect to $\{R_k\}$ is accordingly expressed as:

$${}^\tau_k \dot{\hat{q}} = \frac{1}{2} \boldsymbol{\Omega}(\hat{\boldsymbol{\omega}}) {}^\tau_k \hat{q} \quad (21)$$

$${}^{R_k} \dot{\hat{\mathbf{p}}}_{I_\tau} = {}^\tau_k \mathbf{C}_{\hat{q}}^\top \hat{\mathbf{v}}_{I_\tau} \quad (22)$$

$$\dot{\hat{\mathbf{v}}}_{I_\tau} = \hat{\mathbf{a}} - {}^\tau_k \mathbf{C}_{\hat{q}}^{R_k} \hat{\mathbf{g}} - [\hat{\boldsymbol{\omega}} \times] \hat{\mathbf{v}}_{I_\tau} \quad (23)$$

$$\dot{\hat{\mathbf{b}}}_g = \mathbf{0}_{3 \times 1} \quad (24)$$

$$\dot{\hat{\mathbf{b}}}_a = \mathbf{0}_{3 \times 1} \quad (25)$$

where $\hat{\boldsymbol{\omega}} = \boldsymbol{\omega}_m - \hat{\mathbf{b}}_g$ and $\hat{\mathbf{a}} = \mathbf{a}_m - \hat{\mathbf{b}}_a$ by denoting $\boldsymbol{\omega} = \boldsymbol{\omega}_m - \mathbf{b}_g - \mathbf{n}_g$ and $\mathbf{a} = \mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a$, and for brevity we have denoted ${}^\tau_k \mathbf{C}_{\hat{q}} = \mathbf{C}({}^\tau_k \hat{q})$.

In EKF, the dynamics of IMU error-state is needed for which we start from the error quaternion. By the definition of Eq. (5):

$${}^\tau_k \bar{q} = \delta \bar{q} \otimes {}^\tau_k \hat{q} \quad \left| \frac{d}{dt} \right. \quad (26)$$

$${}^\tau_k \dot{\bar{q}} = \delta \dot{\bar{q}} \otimes {}^\tau_k \hat{q} + \delta \bar{q} \otimes {}^\tau_k \dot{\hat{q}} \quad (27)$$

Substituting Eq. (13) and Eq. (21) leads to:

$$\frac{1}{2} \begin{bmatrix} \boldsymbol{\omega} \\ 0 \end{bmatrix} \otimes {}^\tau_k \bar{q} = \delta \dot{\bar{q}} \otimes {}^\tau_k \hat{q} + \delta \bar{q} \otimes \left(\frac{1}{2} \begin{bmatrix} \hat{\boldsymbol{\omega}} \\ 0 \end{bmatrix} \otimes {}^\tau_k \hat{q} \right) \quad |\boldsymbol{\Omega}(\boldsymbol{\omega}) \bar{q} = \begin{bmatrix} \boldsymbol{\omega} \\ 0 \end{bmatrix} \otimes \bar{q} \quad (28)$$

By rearranging the terms in above equation, we have:

$$\delta \dot{\bar{q}} \otimes {}^\tau_k \hat{q} = \frac{1}{2} \left(\begin{bmatrix} \boldsymbol{\omega} \\ 0 \end{bmatrix} \otimes {}^\tau_k \bar{q} - \delta \bar{q} \otimes \begin{bmatrix} \hat{\boldsymbol{\omega}} \\ 0 \end{bmatrix} \otimes {}^\tau_k \hat{q} \right) \quad |\otimes {}^\tau_k \hat{q}^{-1} \quad (29)$$

$$\delta \dot{\bar{q}} = \frac{1}{2} \left(\begin{bmatrix} \boldsymbol{\omega} \\ 0 \end{bmatrix} \otimes \delta \bar{q} - \delta \bar{q} \otimes \begin{bmatrix} \hat{\boldsymbol{\omega}} \\ 0 \end{bmatrix} \right) \quad |\boldsymbol{\omega} = \hat{\boldsymbol{\omega}} - \tilde{\mathbf{b}}_g - \mathbf{n}_g \quad (30)$$

$$\delta \dot{\bar{q}} = \frac{1}{2} \left(\begin{bmatrix} \hat{\boldsymbol{\omega}} \\ 0 \end{bmatrix} \otimes \delta \bar{q} - \delta \bar{q} \otimes \begin{bmatrix} \hat{\boldsymbol{\omega}} \\ 0 \end{bmatrix} \right) - \frac{1}{2} \begin{bmatrix} \tilde{\mathbf{b}}_g + \mathbf{n}_g \\ 0 \end{bmatrix} \otimes \delta \bar{q} \quad (31)$$

$$= \frac{1}{2} \left(\begin{bmatrix} -[\hat{\boldsymbol{\omega}} \times] & \hat{\boldsymbol{\omega}} \\ -\hat{\boldsymbol{\omega}}^\top & 0 \end{bmatrix} \delta \bar{q} - \begin{bmatrix} [\hat{\boldsymbol{\omega}} \times] & \hat{\boldsymbol{\omega}} \\ -\hat{\boldsymbol{\omega}}^\top & 0 \end{bmatrix} \delta \bar{q} \right) - \frac{1}{2} \begin{bmatrix} -[(\tilde{\mathbf{b}}_g + \mathbf{n}_g) \times] & \tilde{\mathbf{b}}_g + \mathbf{n}_g \\ -(\tilde{\mathbf{b}}_g + \mathbf{n}_g)^\top & 0 \end{bmatrix} \delta \bar{q} \quad (32)$$

$$= \frac{1}{2} \begin{bmatrix} -2[\hat{\boldsymbol{\omega}} \times] + [(\tilde{\mathbf{b}}_g + \mathbf{n}_g) \times] & -(\tilde{\mathbf{b}}_g + \mathbf{n}_g) \\ (\tilde{\mathbf{b}}_g + \mathbf{n}_g)^\top & 0 \end{bmatrix} \delta \bar{q} \quad (33)$$

$$\approx \begin{bmatrix} -\frac{1}{2}[\hat{\omega} \times] \delta\theta_I - \frac{1}{2}(\tilde{\mathbf{b}}_g + \mathbf{n}_g) + \frac{1}{4}[(\tilde{\mathbf{b}}_g + \mathbf{n}_g) \times] \delta\theta_I \\ \frac{1}{4}(\tilde{\mathbf{b}}_g + \mathbf{n}_g)^T \delta\theta_I \end{bmatrix} \quad (34)$$

After neglecting the second order error terms in above equation, we get:

$$\dot{\delta\tilde{q}} \approx \begin{bmatrix} \frac{1}{2}\delta\dot{\theta}_I \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}[\hat{\omega} \times] \delta\theta_I - \frac{1}{2}(\tilde{\mathbf{b}}_g + \mathbf{n}_g) \\ 0 \end{bmatrix} \quad (35)$$

and the target linear equation is the upper part:

$$\dot{\delta\theta}_I = -[\hat{\omega} \times] \delta\theta_I - \tilde{\mathbf{b}}_g - \mathbf{n}_g \quad (36)$$

Next for the error position, we have:

$$\dot{\tilde{\mathbf{p}}}_I = {}^{R_k}\dot{\mathbf{p}}_{I_\tau} - {}^{R_k}\dot{\mathbf{p}}_{I_\tau} \quad (37)$$

$$= {}^k\mathbf{C}_{\hat{q}}^\top \mathbf{v}_{I_\tau} - {}^k\mathbf{C}_{\hat{q}}^\top \hat{\mathbf{v}}_{I_\tau} \quad (38)$$

$$= {}^k\mathbf{C}_{\hat{q}}^\top (\mathbf{I} + [\delta\theta_I \times]) (\hat{\mathbf{v}}_{I_\tau} + \tilde{\mathbf{v}}_{I_\tau}) - {}^k\mathbf{C}_{\hat{q}}^\top \hat{\mathbf{v}}_{I_\tau} \quad (39)$$

$$= {}^k\mathbf{C}_{\hat{q}}^\top [\delta\theta_I \times] \hat{\mathbf{v}}_{I_\tau} + {}^k\mathbf{C}_{\hat{q}}^\top \tilde{\mathbf{v}}_{I_\tau} + {}^k\mathbf{C}_{\hat{q}}^\top [\delta\theta_I \times] \tilde{\mathbf{v}}_{I_\tau} \quad (40)$$

After neglecting the second order error terms, the resulting linear equation is:

$$\dot{\tilde{\mathbf{p}}}_I = {}^k\mathbf{C}_{\hat{q}}^\top [\delta\theta_I \times] \hat{\mathbf{v}}_{I_\tau} + {}^k\mathbf{C}_{\hat{q}}^\top \tilde{\mathbf{v}}_{I_\tau} \quad (41)$$

$$= -{}^k\mathbf{C}_{\hat{q}}^\top [\hat{\mathbf{v}}_{I_\tau} \times] \delta\theta_I + {}^k\mathbf{C}_{\hat{q}}^\top \tilde{\mathbf{v}}_{I_\tau} \quad (42)$$

Similarly for the error velocity, we have:

$$\dot{\tilde{\mathbf{v}}}_I = \dot{\mathbf{v}}_{I_\tau} - \dot{\hat{\mathbf{v}}}_{I_\tau} \quad (43)$$

$$= \mathbf{a} - {}^k\mathbf{C}_{\hat{q}}^{R_k} \mathbf{g} - [\omega \times] \mathbf{v}_{I_\tau} - \hat{\mathbf{a}} + {}^k\mathbf{C}_{\hat{q}}^{R_k} \hat{\mathbf{g}} + [\hat{\omega} \times] \hat{\mathbf{v}}_{I_\tau} \quad (44)$$

$$= -(\tilde{\mathbf{b}}_a + \mathbf{n}_a) - (\mathbf{I} - [\delta\theta_I \times]) {}^k\mathbf{C}_{\hat{q}}^{(R_k \hat{\mathbf{g}} + {}^{R_k} \tilde{\mathbf{g}})} + {}^k\mathbf{C}_{\hat{q}}^{R_k} \hat{\mathbf{g}} \quad (45)$$

$$- [(\hat{\omega} - \tilde{\mathbf{b}}_g - \mathbf{n}_g) \times] (\hat{\mathbf{v}}_{I_\tau} + \tilde{\mathbf{v}}_{I_\tau}) + [\hat{\omega} \times] \hat{\mathbf{v}}_{I_\tau} \quad (45)$$

$$= -(\tilde{\mathbf{b}}_a + \mathbf{n}_a) - {}^k\mathbf{C}_{\hat{q}}^{R_k} \tilde{\mathbf{g}} + [\delta\theta_I \times] {}^k\mathbf{C}_{\hat{q}}^{R_k} \hat{\mathbf{g}} + [\delta\theta_I \times] {}^k\mathbf{C}_{\hat{q}}^{R_k} \tilde{\mathbf{g}} \quad (46)$$

$$- [\hat{\omega} \times] \tilde{\mathbf{v}}_{I_\tau} + [(\tilde{\mathbf{b}}_g + \mathbf{n}_g) \times] \hat{\mathbf{v}}_{I_\tau} + [(\tilde{\mathbf{b}}_g + \mathbf{n}_g) \times] \tilde{\mathbf{v}}_{I_\tau} \quad (46)$$

Again, after neglecting the second order error terms from above equation, we get:

$$\dot{\tilde{\mathbf{v}}}_I = -(\tilde{\mathbf{b}}_a + \mathbf{n}_a) - {}^k\mathbf{C}_{\hat{q}}^{R_k} \tilde{\mathbf{g}} + [\delta\theta_I \times] {}^k\mathbf{C}_{\hat{q}}^{R_k} \hat{\mathbf{g}} - [\hat{\omega} \times] \tilde{\mathbf{v}}_{I_\tau} + [(\tilde{\mathbf{b}}_g + \mathbf{n}_g) \times] \hat{\mathbf{v}}_{I_\tau} \quad (47)$$

$$= -{}^k\mathbf{C}_{\hat{q}}^{R_k} \tilde{\mathbf{g}} - [\hat{\mathbf{g}} \times] \delta\theta_I - [\hat{\omega} \times] \tilde{\mathbf{v}}_{I_\tau} - [\hat{\mathbf{v}}_{I_\tau} \times] (\tilde{\mathbf{b}}_g + \mathbf{n}_g) - (\tilde{\mathbf{b}}_a + \mathbf{n}_a) \quad (48)$$

where for brevity we have denoted $\hat{\mathbf{g}} = {}^k\mathbf{C}_{\hat{q}}^{R_k} \hat{\mathbf{g}}$. The remaining IMU error-states can be obtained as:

$$\dot{\tilde{\mathbf{b}}}_g = \dot{\mathbf{b}}_g - \dot{\hat{\mathbf{b}}}_g = \mathbf{n}_{wg} \quad (49)$$

$$\dot{\tilde{\mathbf{b}}}_a = \dot{\mathbf{b}}_a - \dot{\hat{\mathbf{b}}}_a = \mathbf{n}_{wa} \quad (50)$$

For the error propagation of ${}^{R_k} \mathbf{x}_G$, we consider it as a zero dynamic process, then:

$$\dot{\delta\theta}_G = \mathbf{0}_{3 \times 1} \quad (51)$$

$$\dot{\tilde{\mathbf{p}}}_G = {}^{R_k} \dot{\mathbf{p}}_G - {}^{R_k} \dot{\mathbf{p}}_G = \mathbf{0}_{3 \times 1} \quad (52)$$

$$\dot{\tilde{\mathbf{g}}} = {}^{R_k} \dot{\mathbf{g}} - {}^{R_k} \dot{\mathbf{g}} = \mathbf{0}_{3 \times 1} \quad (53)$$

After combining above results, we obtain the continuous time, linear model of the robocentric error-state as:

$$\begin{aligned} \begin{bmatrix} \dot{\delta\theta}_G \\ \dot{\tilde{\mathbf{p}}}_G \\ \dot{\tilde{\mathbf{g}}} \\ \dot{\delta\theta}_I \\ \dot{\tilde{\mathbf{p}}}_I \\ \dot{\tilde{\mathbf{v}}}_I \\ \dot{\tilde{\mathbf{b}}}_g \\ \dot{\tilde{\mathbf{b}}}_a \end{bmatrix} &= \begin{bmatrix} \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -[\hat{\omega} \times] & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -\tau_k^T \mathbf{C}_{\hat{q}}^\top [\hat{\mathbf{v}}_{I_\tau} \times] & \mathbf{0}_3 & \tau_k^T \mathbf{C}_{\hat{q}}^\top & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & -\tau_k^T \mathbf{C}_{\hat{q}}^\top & -[\tau \hat{\mathbf{g}} \times] & \mathbf{0}_3 & -[\hat{\omega} \times] & -[\hat{\mathbf{v}}_{I_\tau} \times] & -\mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \begin{bmatrix} \delta\theta_G \\ \tilde{\mathbf{p}}_G \\ \tilde{\mathbf{g}} \\ \delta\theta_I \\ \tilde{\mathbf{p}}_I \\ \tilde{\mathbf{v}}_I \\ \tilde{\mathbf{b}}_g \\ \tilde{\mathbf{b}}_a \end{bmatrix} \\ + \begin{bmatrix} \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ -[\hat{\mathbf{v}}_{I_\tau} \times] & \mathbf{0}_3 & -\mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \mathbf{n}_g \\ \mathbf{n}_{wg} \\ \mathbf{n}_a \\ \mathbf{n}_{wa} \end{bmatrix} \end{aligned} \quad (54)$$

$$\Rightarrow \dot{\tilde{\mathbf{x}}} = \mathbf{F}\tilde{\mathbf{x}} + \mathbf{G}\mathbf{n} \quad (55)$$

2.2.2 Discrete Time State Propagation

For an actual EKF implementation, a discrete time propagation model is needed. From time-step k to τ , $\tau_k^T \hat{q}$ can be obtained using the zeroth order quaternion integrator [18]:

$$\tau_k^T \hat{q} = \int_{t_k}^{t_\tau} s \dot{\hat{q}} ds \quad (56)$$

$$= \int_{t_k}^{t_\tau} \frac{1}{2} \Omega(\hat{\omega})_k^s \hat{q} ds \quad (57)$$

$$= \int_{t_k}^{t_\tau} \frac{1}{2} \Omega(\omega_{m,s} - \hat{\mathbf{b}}_g)_k^s \hat{q} ds \quad (58)$$

and $R_k \hat{\mathbf{p}}_{I_\tau}$ and $R_k \hat{\mathbf{v}}_{I_\tau}$ can be computed respectively using IMU preintegration:

$$R_k \hat{\mathbf{p}}_{I_\tau} = \hat{\mathbf{v}}_{I_k} \Delta t + \int_{t_k}^{t_\tau} \int_{t_k}^s \mu_k^{\mu} \mathbf{C}_{\hat{q}}^\top \mu \hat{\mathbf{a}} d\mu ds \quad (59)$$

$$= \hat{\mathbf{v}}_{I_k} \Delta t + \int_{t_k}^{t_\tau} \int_{t_k}^s \mu_k^{\mu} \mathbf{C}_{\hat{q}}^\top (\mathbf{a}_{m,\mu} - \hat{\mathbf{b}}_a - {}^\mu \hat{\mathbf{g}}) d\mu ds \quad (60)$$

$$= \hat{\mathbf{v}}_{I_k} \Delta t - \frac{1}{2} R_k \hat{\mathbf{g}} \Delta t^2 + \underbrace{\int_{t_k}^{t_\tau} \int_{t_k}^s \mu_k^{\mu} \mathbf{C}_{\hat{q}}^\top (\mathbf{a}_{m,\mu} - \hat{\mathbf{b}}_a) d\mu ds}_{\Delta \mathbf{p}_{k,\tau}} \quad (61)$$

$$R_k \hat{\mathbf{v}}_{I_\tau} = \hat{\mathbf{v}}_{I_k} + \int_{t_k}^{t_\tau} s_k^s \mathbf{C}_{\hat{q}}^\top s \hat{\mathbf{a}} ds \quad (62)$$

$$= \hat{\mathbf{v}}_{I_k} + \int_{t_k}^{t_\tau} s_k^s \mathbf{C}_{\hat{q}}^\top (\mathbf{a}_{m,s} - \hat{\mathbf{b}}_a - {}^s \hat{\mathbf{g}}) ds \quad (63)$$

$$= \hat{\mathbf{v}}_{I_k} - R_k \hat{\mathbf{g}} \Delta t + \underbrace{\int_{t_k}^{t_\tau} s_k^s \mathbf{C}_{\hat{q}}^\top (\mathbf{a}_{m,s} - \hat{\mathbf{b}}_a) ds}_{\Delta \mathbf{v}_{k,\tau}} \quad (64)$$

where $\hat{\mathbf{b}}_g = \hat{\mathbf{b}}_{g_k}$, $\hat{\mathbf{b}}_a = \hat{\mathbf{b}}_{a_k}$, and $\Delta t = t_\tau - t_k$. We can find that the preintegrated terms, $\Delta \mathbf{p}_{k,\tau}$ and $\Delta \mathbf{v}_{k,\tau}$, can be computed recursively with the IMU measurements [19]. Therefore, the estimate for the local velocity, $\hat{\mathbf{v}}_{I_\tau}$, can be obtained using the preceding results as:

$$\hat{\mathbf{v}}_{I_\tau} = {}^\tau_k \mathbf{C}_{\hat{\tilde{q}}}^{R_k} \hat{\mathbf{v}}_{I_\tau} \quad (65)$$

The bias estimates are assumed to be constant within the time interval $[t_k, t_\tau]$, which are

$$\hat{\mathbf{b}}_{g_\tau} = \hat{\mathbf{b}}_{g_k} \quad (66)$$

$$\hat{\mathbf{b}}_{a_\tau} = \hat{\mathbf{b}}_{a_k} \quad (67)$$

Thus, the discrete time state propagation from time-step τ to $\tau + 1$ can be described as:

$${}^{R_k} \hat{\mathbf{x}}_{G_{\tau+1}} = {}^{R_k} \hat{\mathbf{x}}_{G_\tau} \quad (68)$$

and for the IMU state

$${}^{\tau+1} \hat{\tilde{q}} = {}^\tau \hat{\tilde{q}} \otimes {}^\tau \hat{\tilde{q}} \quad (69)$$

$${}^{R_k} \hat{\mathbf{p}}_{I_{\tau+1}} = {}^{R_k} \hat{\mathbf{p}}_{I_\tau} + \Delta \mathbf{p}_{\tau, \tau+1} \quad (70)$$

$$\hat{\mathbf{v}}_{I_{\tau+1}} = {}^\tau \hat{\mathbf{C}}_{\hat{\tilde{q}}} \left({}^{R_k} \hat{\mathbf{v}}_{I_\tau} + \Delta \mathbf{v}_{\tau, \tau+1} \right) \quad (71)$$

$$\hat{\mathbf{b}}_{g_{\tau+1}} = \hat{\mathbf{b}}_{g_\tau} \quad (72)$$

$$\hat{\mathbf{b}}_{a_{\tau+1}} = \hat{\mathbf{b}}_{a_\tau} \quad (73)$$

In addition, we consider the estimate of the clone states, $\hat{\mathbf{w}}$, is static within $[t_k, t_{k+1}]$.

2.2.3 Discrete Time Covariance Propagation

In EKF, the covariance propagation from time-step τ to $\tau + 1$ with respect to $\{R_k\}$ can be expressed as:

$$\mathbf{P}_{\tau+1|k} = \Phi_{\tau+1,\tau} \mathbf{P}_{\tau|k} \Phi_{\tau+1,\tau}^\top + \mathbf{Q}_\tau \quad (74)$$

where $\mathbf{P}_{\tau|k}$ and $\mathbf{P}_{\tau+1|k}$ are the estimated and predicted covariance matrices, respectively, $\Phi_{\tau+1,\tau}$ is the state transition matrix from τ to $\tau + 1$, and \mathbf{Q}_τ is the noise covariance matrix. Assuming that $\hat{\omega}$ and $\hat{\mathbf{a}}$ are constant for time interval $[t_\tau, t_{\tau+1}]$ with $\delta t = t_{\tau+1} - t_\tau$, \mathbf{F} is constant and the state transition matrix can be found as:

$$\Phi(\tau + 1, \tau) = \exp(\mathbf{F}\delta t) = \mathbf{I}_{24} + \mathbf{F}\delta t + \frac{1}{2!}\mathbf{F}^2\delta t^2 + \dots \quad (75)$$

Considering that IMU has a very high frequency (e.g., δt is $10^{-3} \sim 10^{-2}$ s for 100 ~ 1000Hz), we can obtain the discrete time error-state transition matrix by just using the first order approximation of Eq. (75):

$$\Phi_{\tau+1,\tau} = \mathbf{I}_{24} + \mathbf{F}\delta t \quad (76)$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -[\hat{\omega} \times] & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -{}^\tau_k \mathbf{C}_{\hat{\tilde{q}}}^\top [\hat{\mathbf{v}}_{I_\tau} \times] & \mathbf{0}_3 & {}^\tau_k \mathbf{C}_{\hat{\tilde{q}}}^\top & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & -{}^\tau_k \mathbf{C}_{\hat{\tilde{q}}} & -[{}^\tau \hat{\mathbf{g}} \times] & \mathbf{0}_3 & -[\hat{\omega} \times] & -[\hat{\mathbf{v}}_{I_\tau} \times] & -\mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \quad (77)$$

In order to obtain the noise covarian matrix \mathbf{Q}_τ , we first show the discrete-time solution of Eq. (55):

$$\tilde{\mathbf{x}}_{\tau+1} = \Phi(\tau + 1, \tau) \tilde{\mathbf{x}}_\tau + \int_{t_\tau}^{t_{\tau+1}} \Phi(\tau + 1, s) \mathbf{G}(s) \mathbf{n} ds \quad (78)$$

$$\mathbf{G}(s) = \begin{bmatrix} \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ -[\hat{\mathbf{V}}_{I_s} \times] & \mathbf{0}_3 & -\mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix}, \quad \mathbf{n} = \begin{bmatrix} \mathbf{n}_g \\ \mathbf{n}_{wg} \\ \mathbf{n}_a \\ \mathbf{n}_{wa} \end{bmatrix} \quad (79)$$

If all matrices are constant for time interval $[t_\tau, t_{\tau+1}]$, the noise covariance for Eq. (78) can be computed as:

$$\begin{aligned} \mathbf{Q}_\tau &= \mathbb{E}[\Phi(\tau+1, \tau)\mathbf{G}(\tau)\mathbf{n}(\Phi(\tau+1, \tau)\mathbf{G}(\tau)\mathbf{n})^\top]\delta t \\ &= \Phi_{\tau+1, \tau}\mathbf{G}\mathbb{E}[\mathbf{n}\mathbf{n}^\top]\mathbf{G}^\top\Phi_{\tau+1, \tau}^\top\delta t \\ &= \Phi_{\tau+1, \tau}\mathbf{G}\Sigma\mathbf{G}^\top\Phi_{\tau+1, \tau}^\top\delta t \end{aligned} \quad (80)$$

$$\Sigma = \begin{bmatrix} \sigma_g^2\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \sigma_{wg}^2\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \sigma_a^2\mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \sigma_{wa}^2\mathbf{I}_3 \end{bmatrix} \quad (81)$$

where Σ is the continuous time noise covariance matrix. Substituting Eq. (76) into Eq. (80) and neglecting the second order terms of δt , we have the discrete time noise covariance matrix:

$$\mathbf{Q}_\tau = \mathbf{G}\Sigma\mathbf{G}^\top\delta t \quad (82)$$

We introduce the following partition for the augmented covariance matrix at time-step k :

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{P}_{\mathbf{x}\mathbf{x}_k} & \mathbf{P}_{\mathbf{x}\mathbf{w}_k} \\ \mathbf{P}_{\mathbf{x}\mathbf{w}_k}^\top & \mathbf{P}_{\mathbf{w}\mathbf{w}_k} \end{bmatrix} \quad (83)$$

where $\mathbf{P}_{\mathbf{x}\mathbf{x}_k}$ is the covariance matrix of the evolving robocentric state, $\mathbf{P}_{\mathbf{w}\mathbf{w}_k}$ is the covariance matrix of the states in the window, and $\mathbf{P}_{\mathbf{x}\mathbf{w}_k}$ is the correlation between the errors in the robocentric state and the states in the window. With this notation, at time-step $\tau+1 \in [t_k, t_{k+1}]$ the propagated covariance is given by:

$$\mathbf{P}_{\tau+1|k} = \begin{bmatrix} \mathbf{P}_{\mathbf{x}\mathbf{x}_{\tau+1|k}} & \Phi_{\tau+1,k}\mathbf{P}_{\mathbf{x}\mathbf{w}_k} \\ \mathbf{P}_{\mathbf{x}\mathbf{w}_k}^\top\Phi_{\tau+1,k}^\top & \mathbf{P}_{\mathbf{w}\mathbf{w}_k} \end{bmatrix} \quad (84)$$

where $\mathbf{P}_{\mathbf{x}\mathbf{x}_{\tau+1|k}}$ can be recursively computed using Eq. (74), and the evolving error-state transition matrix:

$$\Phi_{\tau+1,k} = \Phi_{\tau+1,\tau} \cdots \Phi_{k+\delta t,k} = \prod_{\ell=k}^{\tau} \Phi_{\ell+\delta t,\ell} \quad (85)$$

with $\Phi_{k,k} = \mathbf{I}_{24}$ at time-step k .

2.3 Update

2.3.1 Bearing-only Model

We present this measurement model by considering the case of a single landmark L_j which has been observed from a set of n_j robocentric frames of reference \mathcal{R}_j . For $i \in \mathcal{R}_j$, the observation of landmark L_j is described by the following bearing-only model in the xyz coordinates:

$$\mathbf{z}_{j,i} = \frac{1}{z_j^i} \begin{bmatrix} x_j^i \\ y_j^i \\ z_j^i \end{bmatrix} + \mathbf{n}_{j,i}, \quad {}^C_i \mathbf{p}_{L_j} = \begin{bmatrix} x_j^i \\ y_j^i \\ z_j^i \end{bmatrix} = {}_I^C \mathbf{C}_{\bar{q}} {}^{R_i} \mathbf{p}_{L_j} + {}^C \mathbf{p}_I \quad (86)$$

where $\mathbf{n}_{j,i} \sim \mathcal{N}(\mathbf{0}, \sigma_{im}^2 \mathbf{I}_2)$ represents the image noise, and ${}^C_i \mathbf{p}_{L_j}$ is the position of landmark expressed in the camera frame $\{C_i\}$. Assume that $\mathcal{R}_j = \{1, 2, \dots, n\}$ and $\{R_1\}$ is the first robocentric frame from which the landmark is observed, we begin to derive the expression for ${}^{R_i} \mathbf{p}_{L_j}$, $i \in \{\mathcal{R}_j \setminus 1\}$ following the steps:

1. If $i = 2$, then

$${}^{R_2} \mathbf{p}_{L_j} = {}^2_1 \mathbf{C}_{\bar{q}} ({}^{R_1} \mathbf{p}_{L_j} - {}^{R_1} \mathbf{p}_{R_2}) \quad (87)$$

$$= {}^2_1 \mathbf{C}_{\bar{q}} {}^{R_1} \mathbf{p}_{L_j} - {}^2_1 \mathbf{C}_{\bar{q}} {}^{R_1} \mathbf{p}_{R_2} \quad (88)$$

2. If $i = 3$, then

$${}^{R_3} \mathbf{p}_{L_j} = {}^3_1 \mathbf{C}_{\bar{q}} ({}^{R_1} \mathbf{p}_{L_j} - {}^{R_1} \mathbf{p}_{R_3}) \quad (89)$$

$$= {}^3_1 \mathbf{C}_{\bar{q}} {}^{R_1} \mathbf{p}_{L_j} - {}^3_2 \mathbf{C}_{\bar{q}} {}^{R_2} \mathbf{p}_{R_3} - {}^3_1 \mathbf{C}_{\bar{q}} {}^{R_1} \mathbf{p}_{R_2} \quad (90)$$

3. If $i = 4$, then

$${}^{R_4} \mathbf{p}_{L_j} = {}^4_1 \mathbf{C}_{\bar{q}} ({}^{R_1} \mathbf{p}_{L_j} - {}^{R_1} \mathbf{p}_{R_4}) \quad (91)$$

$$= {}^4_1 \mathbf{C}_{\bar{q}} {}^{R_1} \mathbf{p}_{L_j} - {}^4_3 \mathbf{C}_{\bar{q}} {}^{R_3} \mathbf{p}_{R_4} - {}^4_2 \mathbf{C}_{\bar{q}} {}^{R_2} \mathbf{p}_{R_3} - {}^4_1 \mathbf{C}_{\bar{q}} {}^{R_1} \mathbf{p}_{R_2} \quad (92)$$

4.

5. From above results, we can conclude that

$${}^{R_i} \mathbf{p}_{L_j} = {}^i_1 \mathbf{C}_{\bar{q}} ({}^{R_1} \mathbf{p}_{L_j} - {}^{R_1} \mathbf{p}_{R_i}) \quad (93)$$

$$= {}^i_1 \mathbf{C}_{\bar{q}} {}^{R_1} \mathbf{p}_{L_j} - {}^i_{i-1} \mathbf{C}_{\bar{q}} {}^{R_{i-1}} \mathbf{p}_{R_i} - {}^i_{i-2} \mathbf{C}_{\bar{q}} {}^{R_{i-2}} \mathbf{p}_{R_{i-1}} - \dots - {}^i_{\ell-1} \mathbf{C}_{\bar{q}} {}^{R_{\ell-1}} \mathbf{p}_{R_\ell} - \dots - {}^i_1 \mathbf{C}_{\bar{q}} {}^{R_1} \mathbf{p}_{R_2} \quad (94)$$

Note that in Eq. (94) all relative poses between $\{R_1\}$ and $\{R_i\}$, $i \in \{\mathcal{R}_j \setminus 1\}$ have been used to express the i -th observation of L_j . Then, the corresponding Jacobians with respect to the error-state are derived as:

1. If $i = 2$, then

$$\frac{\partial {}^{R_2} \tilde{\mathbf{p}}_{L_j}}{\partial \delta \boldsymbol{\theta}_2} = [{}^2_1 \mathbf{C}_{\hat{q}} ({}^{R_1} \hat{\mathbf{p}}_{L_j} - {}^{R_1} \hat{\mathbf{p}}_{R_2}) \times], \quad \frac{\partial {}^{R_2} \tilde{\mathbf{p}}_{L_j}}{\partial {}^{R_1} \tilde{\mathbf{p}}_{R_2}} = -{}^2_1 \mathbf{C}_{\hat{q}} \quad (95)$$

$$\frac{\partial {}^{R_2} \tilde{\mathbf{p}}_{L_j}}{\partial {}^{R_1} \tilde{\mathbf{p}}_{L_j}} = {}^2_1 \mathbf{C}_{\hat{q}} \quad (96)$$

2. If $i = 3$, then

$$\frac{\partial {}^{R_3} \tilde{\mathbf{p}}_{L_j}}{\partial \delta \boldsymbol{\theta}_3} = [{}^3_1 \mathbf{C}_{\hat{q}} ({}^{R_1} \hat{\mathbf{p}}_{L_j} - {}^{R_1} \hat{\mathbf{p}}_{R_3}) \times], \quad \frac{\partial {}^{R_3} \tilde{\mathbf{p}}_{L_j}}{\partial {}^{R_2} \tilde{\mathbf{p}}_{R_3}} = -{}^3_2 \mathbf{C}_{\hat{q}} \quad (97)$$

$$\frac{\partial {}^{R_3} \tilde{\mathbf{p}}_{L_j}}{\partial \delta \boldsymbol{\theta}_2} = {}^3_2 \mathbf{C}_{\hat{q}} [{}^2_1 \mathbf{C}_{\hat{q}} ({}^{R_1} \hat{\mathbf{p}}_{L_j} - {}^{R_1} \hat{\mathbf{p}}_{R_2}) \times], \quad \frac{\partial {}^{R_3} \tilde{\mathbf{p}}_{L_j}}{\partial {}^{R_1} \tilde{\mathbf{p}}_{R_2}} = -{}^3_1 \mathbf{C}_{\hat{q}} \quad (98)$$

$$\frac{\partial {}^{R_3} \tilde{\mathbf{p}}_{L_j}}{\partial {}^{R_1} \tilde{\mathbf{p}}_{L_j}} = {}^3_1 \mathbf{C}_{\hat{q}} \quad (99)$$

3. If $i = 4$, then

$$\frac{\partial {}^{R_4} \tilde{\mathbf{p}}_{L_j}}{\partial \delta \boldsymbol{\theta}_4} = [{}^4_1 \mathbf{C}_{\hat{q}} ({}^{R_1} \hat{\mathbf{p}}_{L_j} - {}^{R_1} \hat{\mathbf{p}}_{R_4}) \times], \quad \frac{\partial {}^{R_4} \tilde{\mathbf{p}}_{L_j}}{\partial {}^{R_3} \tilde{\mathbf{p}}_{R_4}} = -{}^4_3 \mathbf{C}_{\hat{q}} \quad (100)$$

$$\frac{\partial {}^{R_4} \tilde{\mathbf{p}}_{L_j}}{\partial \delta \boldsymbol{\theta}_3} = {}^4_3 \mathbf{C}_{\hat{q}} [{}^3_1 \mathbf{C}_{\hat{q}} ({}^{R_1} \hat{\mathbf{p}}_{L_j} - {}^{R_1} \hat{\mathbf{p}}_{R_3}) \times], \quad \frac{\partial {}^{R_4} \tilde{\mathbf{p}}_{L_j}}{\partial {}^{R_2} \tilde{\mathbf{p}}_{R_3}} = -{}^4_2 \mathbf{C}_{\hat{q}} \quad (101)$$

$$\frac{\partial {}^{R_4} \tilde{\mathbf{p}}_{L_j}}{\partial \delta \boldsymbol{\theta}_2} = {}^4_2 \mathbf{C}_{\hat{q}} [{}^2_1 \mathbf{C}_{\hat{q}} ({}^{R_1} \hat{\mathbf{p}}_{L_j} - {}^{R_1} \hat{\mathbf{p}}_{R_2}) \times], \quad \frac{\partial {}^{R_4} \tilde{\mathbf{p}}_{L_j}}{\partial {}^{R_1} \tilde{\mathbf{p}}_{R_2}} = -{}^4_1 \mathbf{C}_{\hat{q}} \quad (102)$$

$$\frac{\partial {}^{R_4} \tilde{\mathbf{p}}_{L_j}}{\partial {}^{R_1} \tilde{\mathbf{p}}_{L_j}} = {}^4_1 \mathbf{C}_{\hat{q}} \quad (103)$$

4.

5. In conclusion, we have the general form that, $\forall i \in \{\mathcal{R}_j \setminus 1\}$

$$\frac{\partial^{R_i} \tilde{\mathbf{p}}_{L_j}}{\partial \delta \theta_\ell} = {}^i \mathbf{C}_{\hat{q}} |_1^\ell \mathbf{C}_{\hat{q}} ({}^{R_1} \hat{\mathbf{p}}_{L_j} - {}^{R_1} \hat{\mathbf{p}}_{R_\ell}) \times |_1^\ell = {}^i \mathbf{C}_{\hat{q}} |({}^{R_1} \hat{\mathbf{p}}_{L_j} - {}^{R_1} \hat{\mathbf{p}}_{R_\ell}) \times |_1^\ell \mathbf{C}_{\hat{q}}^\top, \quad (104)$$

$$\frac{\partial^{R_i} \tilde{\mathbf{p}}_{L_j}}{\partial {}^{R_{\ell-1}} \tilde{\mathbf{p}}_{R_\ell}} = -{}^i_{\ell-1} \mathbf{C}_{\hat{q}}, \quad \forall \ell \in \{2, 3, \dots, i\} \quad (105)$$

$$\frac{\partial^{R_i} \tilde{\mathbf{p}}_{L_j}}{\partial {}^{R_1} \tilde{\mathbf{p}}_{L_j}} = {}^i \mathbf{C}_{\hat{q}} \quad (106)$$

From above, we can find that the landmark observation from one robocentric frame has been correlated to a *chain* of relative poses through its tracking history, which is contrast to the world-centric case (e.g., [1]) where the landmark observation was only correlated to the pose from which it was obtained.

Linearizing about the estimates of the robot state and landmark position, the residual of Eq. (86) can be approximated as:

$$\mathbf{r}_{j,i} = \mathbf{z}_{j,i} - \hat{\mathbf{z}}_{j,i} \simeq \mathbf{H}_{\mathbf{x}_{j,i}} \tilde{\mathbf{x}} + \mathbf{H}_{L_{j,i}} {}^{R_1} \tilde{\mathbf{p}}_{L_j} + \mathbf{n}_{j,i} \quad (107)$$

$$\begin{aligned} \mathbf{H}_{\mathbf{x}_{j,i}} &= \mathbf{H}_{\mathbf{p}_{j,i}} {}^C \mathbf{C}_{\hat{q}} |_1^i \mathbf{C}_{\hat{q}} \begin{bmatrix} \mathbf{0}_{3 \times 24} & \mathbf{0}_{3 \times 6} & \dots & \mathbf{H}_{\mathbf{w}_{j,i}} & \dots & \mathbf{0}_{3 \times 6} \end{bmatrix}, \quad \mathbf{H}_{L_{j,i}} = \mathbf{H}_{\mathbf{p}_{j,i}} {}^C \mathbf{C}_{\hat{q}} |_1^i \mathbf{C}_{\hat{q}}, \\ \mathbf{H}_{\mathbf{p}_{j,i}} &= \frac{1}{\hat{z}_j^i} \begin{bmatrix} 1 & 0 & -\frac{\hat{x}_j^i}{\hat{z}_j^i} \\ 0 & 1 & -\frac{\hat{y}_j^i}{\hat{z}_j^i} \end{bmatrix}, \\ \mathbf{H}_{\mathbf{w}_{j,i}} &= \left[|({}^{R_1} \hat{\mathbf{p}}_{L_j} - {}^{R_1} \hat{\mathbf{p}}_{R_2}) \times |_1^2 \mathbf{C}_{\hat{q}}^\top - \mathbf{I}_3 \dots |({}^{R_1} \hat{\mathbf{p}}_{L_j} - {}^{R_1} \hat{\mathbf{p}}_{R_i}) \times |_1^i \mathbf{C}_{\hat{q}}^\top - {}_1^{i-1} \mathbf{C}_{\hat{q}}^\top \right] \end{aligned} \quad (108)$$

where $\mathbf{H}_{\mathbf{x}_{j,i}}$ ($\mathbf{H}_{\mathbf{x}_{j,1}} = \mathbf{0}$) and $\mathbf{H}_{L_{j,i}}$ ($\mathbf{H}_{L_{j,1}} = \mathbf{H}_{\mathbf{p}_{j,1}} {}^C \mathbf{C}_{\hat{q}}$) are the Jacobians with respect to the errors of the state and landmark position, respectively. While ${}^{R_1} \hat{\mathbf{p}}_{L_j}$ is the position of landmark L_j in $\{R_1\}$ whose value is unknown in computation, for which we first solve a nonlinear least-squares minimization using the measurements $\mathbf{z}_{j,i}$, $i \in \mathcal{R}_j$ and the relative pose estimates $\hat{\mathbf{w}}$ (cf. Appendix B). We stack all $\mathbf{r}_{j,i}$, $i \in \mathcal{R}_j$ to obtain the following equation for landmark L_j :

$$\mathbf{r}_j \simeq \mathbf{H}_{\mathbf{x}_j} \tilde{\mathbf{x}} + \mathbf{H}_{L_j} {}^{R_1} \tilde{\mathbf{p}}_{L_j} + \mathbf{n}_j \quad (109)$$

As ${}^{R_1} \hat{\mathbf{p}}_{L_j}$ is computed using the state estimate $\tilde{\mathbf{x}}$ (precisely, $\hat{\mathbf{w}}$), ${}^{R_1} \tilde{\mathbf{p}}_{L_j}$ is correlated with $\tilde{\mathbf{x}}$. However, ${}^{R_1} \hat{\mathbf{p}}_{L_j}$ is not in the state vector, in order to find a valid residual for EKF update, we project \mathbf{r}_j to the left nullspace of \mathbf{H}_{L_j} . Let \mathbf{O}_{L_j} denotes the matrix consisting of the basis of the left nullspace of \mathbf{H}_{L_j} , we have:

$$\bar{\mathbf{r}}_j = \mathbf{O}_{L_j}^\top \mathbf{r}_j \simeq \mathbf{O}_{L_j}^\top \mathbf{H}_{\mathbf{x}_j} \tilde{\mathbf{x}} + \mathbf{O}_{L_j}^\top \mathbf{n}_j = \bar{\mathbf{H}}_{\mathbf{x}_j} \tilde{\mathbf{x}} + \bar{\mathbf{n}}_j \quad (110)$$

After rejecting the observations of the landmarks at infinity ($z_j^i \gg x_j^i, z_j^i \gg y_j^i, i \in \mathcal{R}_j$) and the observations near the principal point of image ($x_j^i \rightarrow 0, y_j^i \rightarrow 0, i \in \mathcal{R}_j$), we have the full column rank $2n_j \times 3$ matrix \mathbf{H}_{L_j} , of which the nullspace is of dimension $2n_j - 3$. In Eq. (110) without explicitly finding \mathbf{O}_{L_j} , the projection of \mathbf{r}_j and $\mathbf{H}_{\mathbf{x}_j}$ on the nullspace of \mathbf{H}_{L_j} is efficiently computed using *Givens rotation* [20] with complexity $O(n_j^2)$. Also, $\bar{\mathbf{n}}_j = \mathbf{O}_{L_j}^\top \mathbf{n}_j$ is the noise vector with the covariance $\bar{\mathbf{R}} = \mathbf{O}_{L_j}^\top \mathbb{E}[\mathbf{n}_j \mathbf{n}_j^\top] \mathbf{O}_{L_j} = \sigma_{im}^2 \mathbf{I}_{2n_j-3}$.

2.3.2 Inverse Depth Bearing-only Model

We still consider the bearing-only model and the case of a single landmark L_j which are used in the preceding section. However, instead of using the *xyz* coordinates, we introduce the following tailored *inverse depth* [21] parametrization for the landmark position:

$${}^{C_i} \mathbf{p}_{L_j} = {}^i \bar{\mathbf{C}}_{\bar{q}} {}^{C_1} \mathbf{p}_{L_j} + {}^i \bar{\mathbf{p}}_1 := \mathbf{f}_i(\phi, \psi, \rho) \quad (111)$$

$${}^{C_1} \mathbf{p}_{L_j} = \frac{1}{\rho} \mathbf{e}(\phi, \psi), \quad \mathbf{e} = \begin{bmatrix} \cos \phi \sin \psi \\ \sin \phi \\ \cos \phi \cos \psi \end{bmatrix} \quad (112)$$

where we use \mathbf{e} (i.e., the unit vector of ${}^C_1 \mathbf{p}_{L_j}$) and $1/\rho$ (i.e., the depth of L_j in $\{C_1\}$) to express ${}^C_1 \mathbf{p}_{L_j}$ with ϕ and ψ the elevation and azimuth expressed in $\{C_1\}$. Especially, the relative poses between $\{C_1\}$ and $\{C_i\}$, $i \in \{\mathcal{R}_j \setminus 1\}$ are expressed using the calibration result, $\{{}^C_I \mathbf{C}_{\bar{q}}, {}^C \mathbf{p}_I\}$, and the relative poses between the robocentric frames of reference, \mathbf{w} , as

$${}_1 \bar{\mathbf{C}}_{\bar{q}} = {}^C_I \mathbf{C}_{\bar{q}} {}^i \mathbf{C}_{\bar{q}} {}^I \mathbf{C}_{\bar{q}} \quad (113)$$

$${}^i \bar{\mathbf{p}}_1 = {}^C_I \mathbf{C}_{\bar{q}} {}^i \mathbf{C}_{\bar{q}} {}^I \mathbf{p}_C + {}^C_I \mathbf{C}_{\bar{q}} {}^{R_i} \mathbf{p}_{R_1} + {}^C \mathbf{p}_I \quad (114)$$

Considering the observation of landmark at infinity (i.e., $\rho \rightarrow 0$), we can normalize Eq. (111) with ρ to avoid numerical issues. Note that such operation dose not affect the bearing expression because the scale is always eliminated in the fraction form:

$$\mathbf{h}_i(\mathbf{w}, \phi, \psi, \rho) = \begin{bmatrix} h_{i,1}(\mathbf{w}, \phi, \psi, \rho) \\ h_{i,2}(\mathbf{w}, \phi, \psi, \rho) \\ h_{i,3}(\mathbf{w}, \phi, \psi, \rho) \end{bmatrix} =: {}^i \bar{\mathbf{C}}_{\bar{q}} \mathbf{e}(\phi, \psi) + \rho^i \bar{\mathbf{p}}_1 \quad (115)$$

This equation retains the geometry of Eq. (111), meanwhile compasses two degenerate cases: (i) observing the landmarks at infinity (i.e., $\rho \rightarrow 0$), and (ii) having low parallax between $\{C_1\}$ and $\{C_i\}$ (i.e., ${}^i \bar{\mathbf{p}}_1 \rightarrow 0$). For both cases, Eq. (115) can be approximated by $\mathbf{h}_i \simeq {}^i \bar{\mathbf{C}}_{\bar{q}} \mathbf{e}(\phi, \psi)$, and hence the corresponding measurements can still provide the information about the camera rotation. Thus, we introduce the following inverse depth measurement model for robocentric VIO:

$$\mathbf{z}_{j,i} = \frac{1}{h_{i,3}(\mathbf{w}, \phi, \psi, \rho)} \begin{bmatrix} h_{i,1}(\mathbf{w}, \phi, \psi, \rho) \\ h_{i,2}(\mathbf{w}, \phi, \psi, \rho) \end{bmatrix} + \mathbf{n}_{j,i} \quad (116)$$

Assume that $\mathcal{R}_j = \{1, 2, \dots, n\}$, and $\{R_1\}$ is the robocentric frame from which L_j is firstly observed, then:

$$\begin{aligned} {}^{C_i} \mathbf{p}_{L_j} &= {}^i \bar{\mathbf{C}}_{\bar{q}} {}^{C_1} \mathbf{p}_{L_j} + {}^i \bar{\mathbf{p}}_1 \\ &= {}^C_I \mathbf{C}_{\bar{q}} {}^i \mathbf{C}_{\bar{q}} {}^I \mathbf{C}_{\bar{q}} {}^{C_1} \mathbf{p}_{L_j} + {}^C_I \mathbf{C}_{\bar{q}} {}^i \mathbf{C}_{\bar{q}} {}^I \mathbf{p}_C + {}^C_I \mathbf{C}_{\bar{q}} {}^{R_i} \mathbf{p}_{R_1} + {}^C \mathbf{p}_I \\ &= {}^C_I \mathbf{C}_{\bar{q}} {}^i \mathbf{C}_{\bar{q}} ({}^I_C \mathbf{C}_{\bar{q}} {}^{C_1} \mathbf{p}_{L_j} + {}^I \mathbf{p}_C - {}^{R_1} \mathbf{p}_{R_i}) + {}^C \mathbf{p}_I \end{aligned} \quad (117)$$

$$\Rightarrow \mathbf{h}_i(\phi, \psi, \rho) = {}^C_I \mathbf{C}_{\bar{q}} {}^i \mathbf{C}_{\bar{q}} ({}^I_C \mathbf{C}_{\bar{q}} \mathbf{e}(\phi, \psi) + \rho {}^I \mathbf{p}_C - \rho {}^{R_1} \mathbf{p}_{R_i}) + \rho {}^C \mathbf{p}_I \quad (118)$$

1. If $i = 2$, then

$$\mathbf{h}_2 = {}^C_I \mathbf{C}_{\bar{q}} {}^2 \mathbf{C}_{\bar{q}} ({}^I_C \mathbf{C}_{\bar{q}} \mathbf{e}(\phi, \psi) + \rho {}^I \mathbf{p}_C - \rho {}^{R_1} \mathbf{p}_{R_2}) + \rho {}^C \mathbf{p}_I \quad (119)$$

$$= {}^C_I \mathbf{C}_{\bar{q}} {}^2 \mathbf{C}_{\bar{q}} ({}^I_C \mathbf{C}_{\bar{q}} \mathbf{e}(\phi, \psi) + \rho {}^I \mathbf{p}_C) - \rho {}^C_I \mathbf{C}_{\bar{q}} {}^2 \mathbf{C}_{\bar{q}} {}^{R_1} \mathbf{p}_{R_2} + \rho {}^C \mathbf{p}_I \quad (120)$$

2. If $i = 3$, then

$$\mathbf{h}_3 = {}^C_I \mathbf{C}_{\bar{q}} {}^3 \mathbf{C}_{\bar{q}} ({}^I_C \mathbf{C}_{\bar{q}} \mathbf{e}(\phi, \psi) + \rho {}^I \mathbf{p}_C - \rho {}^{R_1} \mathbf{p}_{R_3}) + \rho {}^C \mathbf{p}_I \quad (121)$$

$$= {}^C_I \mathbf{C}_{\bar{q}} {}^3 \mathbf{C}_{\bar{q}} ({}^I_C \mathbf{C}_{\bar{q}} \mathbf{e}(\phi, \psi) + \rho {}^I \mathbf{p}_C) - \rho {}^C_I \mathbf{C}_{\bar{q}} ({}^3_2 \mathbf{C}_{\bar{q}} {}^{R_2} \mathbf{p}_{R_3} + {}^3_1 \mathbf{C}_{\bar{q}} {}^{R_1} \mathbf{p}_{R_2}) + \rho {}^C \mathbf{p}_I \quad (122)$$

3. If $i = 4$, then

$$\mathbf{h}_4 = {}^C_I \mathbf{C}_{\bar{q}} {}^4 \mathbf{C}_{\bar{q}} ({}^I_C \mathbf{C}_{\bar{q}} \mathbf{e}(\phi, \psi) + \rho {}^I \mathbf{p}_C - \rho {}^{R_1} \mathbf{p}_{R_4}) + \rho {}^C \mathbf{p}_I \quad (123)$$

$$= {}^C_I \mathbf{C}_{\bar{q}} {}^4 \mathbf{C}_{\bar{q}} ({}^I_C \mathbf{C}_{\bar{q}} \mathbf{e}(\phi, \psi) + \rho {}^I \mathbf{p}_C) - \rho {}^C_I \mathbf{C}_{\bar{q}} ({}^4_3 \mathbf{C}_{\bar{q}} {}^{R_3} \mathbf{p}_{R_4} + {}^4_2 \mathbf{C}_{\bar{q}} {}^{R_2} \mathbf{p}_{R_3} + {}^4_1 \mathbf{C}_{\bar{q}} {}^{R_1} \mathbf{p}_{R_2}) + \rho {}^C \mathbf{p}_I \quad (124)$$

4.

5. Finally, we can conclude that

$$\mathbf{h}_i = {}^C_I \mathbf{C}_{\bar{q}} {}^i \mathbf{C}_{\bar{q}} ({}^I_C \mathbf{C}_{\bar{q}} \mathbf{e}(\phi, \psi) + \rho {}^I \mathbf{p}_C - \rho {}^{R_1} \mathbf{p}_{R_i}) + \rho {}^C \mathbf{p}_I \quad (125)$$

$$\begin{aligned} &= {}^C_I \mathbf{C}_{\bar{q}} {}^i \mathbf{C}_{\bar{q}} ({}^I_C \mathbf{C}_{\bar{q}} \mathbf{e}(\phi, \psi) + \rho {}^I \mathbf{p}_C) \\ &\quad - \rho {}^C_I \mathbf{C}_{\bar{q}} ({}^i_{i-1} \mathbf{C}_{\bar{q}} {}^{R_{i-1}} \mathbf{p}_{R_i} + {}^i_{i-2} \mathbf{C}_{\bar{q}} {}^{R_{i-2}} \mathbf{p}_{R_{i-1}} + \dots + {}^i_{\ell-1} \mathbf{C}_{\bar{q}} {}^{R_{\ell-1}} \mathbf{p}_{R_\ell} + \dots + {}^i_1 \mathbf{C}_{\bar{q}} {}^{R_1} \mathbf{p}_{R_2}) \\ &\quad + \rho {}^C \mathbf{p}_I \end{aligned} \quad (126)$$

Similarly, we can obtain the Jacobians of \mathbf{h}_i , $i \in \{\mathcal{R}_j \setminus 1\}$ with respect to the error-state and $[\phi, \psi, \rho]^\top$:

1. If $i = 2$, then

$$\frac{\partial \mathbf{h}_2}{\partial \delta \theta_2} = {}_I^C \mathbf{C}_{\bar{q}} |_1^2 \mathbf{C}_{\hat{q}} ({}_C^I \mathbf{C}_{\bar{q}} \hat{\mathbf{e}} + \hat{\rho}^I \mathbf{p}_C - \hat{\rho}^{R_1} \hat{\mathbf{p}}_{R_2}) \times |, \quad \frac{\partial \mathbf{h}_2}{\partial R_1 \tilde{\mathbf{p}}_{R_2}} = -\hat{\rho}_I^C \mathbf{C}_{\bar{q}1}^2 \mathbf{C}_{\hat{q}} \quad (127)$$

2. If $i = 3$, then

$$\frac{\partial \mathbf{h}_3}{\partial \delta \theta_3} = {}_I^C \mathbf{C}_{\bar{q}} |_1^3 \mathbf{C}_{\hat{q}} ({}_C^I \mathbf{C}_{\bar{q}} \hat{\mathbf{e}} + \hat{\rho}^I \mathbf{p}_C - \hat{\rho}^{R_1} \hat{\mathbf{p}}_{R_3}) \times |, \quad \frac{\partial \mathbf{h}_3}{\partial R_2 \tilde{\mathbf{p}}_{R_3}} = -\hat{\rho}_I^C \mathbf{C}_{\bar{q}2}^3 \mathbf{C}_{\hat{q}} \quad (128)$$

$$\frac{\partial \mathbf{h}_3}{\partial \delta \theta_2} = {}_I^C \mathbf{C}_{\bar{q}2}^3 \mathbf{C}_{\hat{q}} |_1^2 \mathbf{C}_{\hat{q}} ({}_C^I \mathbf{C}_{\bar{q}} \hat{\mathbf{e}} + \hat{\rho}^I \mathbf{p}_C - \hat{\rho}^{R_1} \hat{\mathbf{p}}_{R_2}) \times |, \quad \frac{\partial \mathbf{h}_3}{\partial R_1 \tilde{\mathbf{p}}_{R_2}} = -\hat{\rho}_I^C \mathbf{C}_{\bar{q}1}^3 \mathbf{C}_{\hat{q}} \quad (129)$$

3. If $i = 4$, then

$$\frac{\partial \mathbf{h}_4}{\partial \delta \theta_4} = {}_I^C \mathbf{C}_{\bar{q}} |_1^4 \mathbf{C}_{\hat{q}} ({}_C^I \mathbf{C}_{\bar{q}} \hat{\mathbf{e}} + \hat{\rho}^I \mathbf{p}_C - \hat{\rho}^{R_1} \hat{\mathbf{p}}_{R_4}) \times |, \quad \frac{\partial \mathbf{h}_4}{\partial R_3 \tilde{\mathbf{p}}_{R_4}} = -\hat{\rho}_I^C \mathbf{C}_{\bar{q}3}^4 \mathbf{C}_{\hat{q}} \quad (130)$$

$$\frac{\partial \mathbf{h}_4}{\partial \delta \theta_3} = {}_I^C \mathbf{C}_{\bar{q}3}^4 \mathbf{C}_{\hat{q}} |_1^3 \mathbf{C}_{\hat{q}} ({}_C^I \mathbf{C}_{\bar{q}} \hat{\mathbf{e}} + \hat{\rho}^I \mathbf{p}_C - \hat{\rho}^{R_1} \hat{\mathbf{p}}_{R_3}) \times |, \quad \frac{\partial \mathbf{h}_4}{\partial R_2 \tilde{\mathbf{p}}_{R_3}} = -\hat{\rho}_I^C \mathbf{C}_{\bar{q}2}^4 \mathbf{C}_{\hat{q}} \quad (131)$$

$$\frac{\partial \mathbf{h}_4}{\partial \delta \theta_2} = {}_I^C \mathbf{C}_{\bar{q}2}^4 \mathbf{C}_{\hat{q}} |_1^2 \mathbf{C}_{\hat{q}} ({}_C^I \mathbf{C}_{\bar{q}} \hat{\mathbf{e}} + \hat{\rho}^I \mathbf{p}_C - \hat{\rho}^{R_1} \hat{\mathbf{p}}_{R_2}) \times |, \quad \frac{\partial \mathbf{h}_4}{\partial R_1 \tilde{\mathbf{p}}_{R_2}} = -\hat{\rho}_I^C \mathbf{C}_{\bar{q}1}^4 \mathbf{C}_{\hat{q}} \quad (132)$$

4.

5. In conclusion, we have the general form as follows. Specifically, the Jacobians with respect to $[\phi, \psi, \rho]^\top$ are obtained from Eq. (115):

$$\frac{\partial \mathbf{h}_i}{\partial \delta \theta_\ell} = {}_I^C \mathbf{C}_{\bar{q}\ell}^i \mathbf{C}_{\hat{q}} |_1^\ell \mathbf{C}_{\hat{q}} ({}_C^I \mathbf{C}_{\bar{q}} \hat{\mathbf{e}} + \hat{\rho}^I \mathbf{p}_C - \hat{\rho}^{R_1} \hat{\mathbf{p}}_{R_\ell}) \times | \quad (133)$$

$$= {}_I^C \mathbf{C}_{\bar{q}1}^i \mathbf{C}_{\hat{q}} | ({}_C^I \mathbf{C}_{\bar{q}} \hat{\mathbf{e}} + \hat{\rho}^I \mathbf{p}_C - \hat{\rho}^{R_1} \hat{\mathbf{p}}_{R_\ell}) \times |_1^\ell \mathbf{C}_{\hat{q}}^\top, \quad (134)$$

$$\frac{\partial \mathbf{h}_i}{\partial R_{\ell-1} \tilde{\mathbf{p}}_{R_\ell}} = -\hat{\rho}_I^C \mathbf{C}_{\bar{q}\ell-1}^i \mathbf{C}_{\hat{q}}, \quad \forall \ell \in \{2, 3, \dots, i\} \quad (135)$$

$$\frac{\partial \mathbf{h}_i}{\partial [\phi, \psi, \rho]^\top} = \begin{bmatrix} \frac{\partial \mathbf{h}_i}{\partial [\phi, \psi]^\top} & \frac{\partial \mathbf{h}_i}{\partial \rho} \end{bmatrix} = \begin{bmatrix} {}_1^i \bar{\mathbf{C}}_{\hat{q}} \begin{bmatrix} -\sin \hat{\phi} \sin \hat{\psi} & \cos \hat{\phi} \cos \hat{\psi} \\ \cos \hat{\phi} & 0 \\ -\sin \hat{\phi} \cos \hat{\psi} & -\cos \hat{\phi} \sin \hat{\psi} \end{bmatrix} & {}_1^i \hat{\mathbf{p}}_1 \end{bmatrix} \quad (136)$$

Let us denote $\boldsymbol{\lambda} = [\phi, \psi, \rho]^\top$, then the residual of Eq. (116) by linearizing about the estimates of the robot state and $\boldsymbol{\lambda}$ can be approximated as:

$$\mathbf{r}_{j,i} = \mathbf{z}_{j,i} - \hat{\mathbf{z}}_{j,i} \simeq \mathbf{H}_{\mathbf{x}_{j,i}} \tilde{\mathbf{x}} + \mathbf{H}_{\boldsymbol{\lambda}_{j,i}} \tilde{\boldsymbol{\lambda}} + \mathbf{n}_{j,i} \quad (137)$$

$$\begin{aligned} \mathbf{H}_{\mathbf{x}_{j,i}} &= \mathbf{H}_{\mathbf{p}_{j,i}} {}_I^C \mathbf{C}_{\bar{q}1}^i \mathbf{C}_{\hat{q}} \begin{bmatrix} \mathbf{0}_{3 \times 24} & \mathbf{0}_{3 \times 6} & \dots & \mathbf{H}_{\mathbf{w}_{j,i}} & \dots & \mathbf{0}_{3 \times 6} \end{bmatrix}, \quad \mathbf{H}_{\boldsymbol{\lambda}_{j,i}} = \mathbf{H}_{\mathbf{p}_{j,i}} \mathbf{H}_{\text{inv}_{j,i}}, \\ \mathbf{H}_{\mathbf{p}_{j,i}} &= \frac{1}{\hat{h}_{i,3}} \begin{bmatrix} 1 & 0 & -\frac{\hat{h}_{i,1}}{\hat{h}_{i,3}} \\ 0 & 1 & -\frac{\hat{h}_{i,2}}{\hat{h}_{i,3}} \end{bmatrix}, \\ \mathbf{H}_{\text{inv}_{j,i}} &= \begin{bmatrix} {}_1^i \bar{\mathbf{C}}_{\hat{q}} \begin{bmatrix} -\sin \hat{\phi} \sin \hat{\psi} & \cos \hat{\phi} \cos \hat{\psi} \\ \cos \hat{\phi} & 0 \\ -\sin \hat{\phi} \cos \hat{\psi} & -\cos \hat{\phi} \sin \hat{\psi} \end{bmatrix} & {}_1^i \hat{\mathbf{p}}_1 \end{bmatrix}, \\ \mathbf{H}_{\mathbf{w}_{j,i}} &= \left[| ({}_C^I \mathbf{C}_{\bar{q}} \hat{\mathbf{e}} + \hat{\rho}^I \mathbf{p}_C - \hat{\rho}^{R_1} \hat{\mathbf{p}}_{R_2}) \times |_1^2 \mathbf{C}_{\hat{q}}^\top | \quad -\hat{\rho} \mathbf{I}_3 \quad \dots \quad | ({}_C^I \mathbf{C}_{\bar{q}} \hat{\mathbf{e}} + \hat{\rho}^I \mathbf{p}_C - \hat{\rho}^{R_1} \hat{\mathbf{p}}_{R_i}) \times |_1^i \mathbf{C}_{\hat{q}}^\top | \quad -\hat{\rho}_1^{i-1} \mathbf{C}_{\hat{q}}^\top \right] \end{aligned} \quad (138)$$

Here let us take a break to discuss the *rank deficiency* in $\mathbf{H}_{\lambda_{j,i}}$, which determines the way to get a nominal residual as using the Givens rotations (cf. Eq. (110)). We consider the rank deficiency happened in either $\mathbf{H}_{p_{j,i}}$ or $\mathbf{H}_{\text{inv}_{j,i}}$. For the former, we find two possible conditions for $\hat{\mathbf{h}}_i$: (i) $\hat{h}_{i,1} = \hat{h}_{i,2} = \hat{h}_{i,3}$, which means that the image size should be at least $2f \times 2f$ (f is the focal length), or (ii) $\hat{h}_{i,1} \rightarrow 0$ and $\hat{h}_{i,2} \rightarrow 0$, which means that the measurement of L_j is near the principal point of image; however, for the latter we can also find two conditions: (iii) $\cos \hat{\phi} \rightarrow 0$, which means that we have either infinitely small focal length or infinitely large image size for the camera so that $|\hat{\phi}| \rightarrow \pi/2$ can happen, or (iv) ${}^i\hat{\mathbf{p}}_1 \rightarrow 0$, which means a small parallax between $\{C_1\}$ and $\{C_i\}$. Among those conditions, (i) is about the selection for the lens (e.g., a wide-angle lens) which is restricted by the image size of the camera, and (iii) is too ideal to be realized in the real world. While (ii) and (iv) are common in the vision-aided navigation and can be conveniently detected from the pixel values of measurements and the relative poses, respectively. We can discard the measurements which meet (ii) when computing the Jacobians. While (iv) is the case that we can utilize, however, makes the last column of $\mathbf{H}_{\lambda_{j,i}}$ approaches zero, for which we perform Givens rotation on the first two columns to guarantee a valid nullspace projection numerically. Also, in Eq. (137) an estimate of λ is needed, for which a nonlinear least-squares minimization is solved at first (cf. Appendix C). To the end, we stack all $\bar{\mathbf{r}}_{j,i}$, $i \in \mathcal{R}_j$ to obtain the residual $\bar{\mathbf{r}}_j$ for EKF update (cf. Eq. (110)).

2.3.3 EKF Update

Assuming at time-step $k+1$, we have the observations of M landmarks to process, we can stack $\bar{\mathbf{r}}_j$, $j = 1 \dots M$ to have:

$$\bar{\mathbf{r}} = \bar{\mathbf{H}}_{\mathbf{x}} \tilde{\mathbf{x}} + \bar{\mathbf{n}} \quad (139)$$

of which the resulting row dimension is $d = \sum_{j=1}^M (2n_j - 3)$. However, in practice, d can be a large number even if M is small (e.g., $d = 170$, when 10 landmarks are observed from 10 robot poses). In order to reduce the computational complexity in EKF update, we perform QR decomposition on $\bar{\mathbf{H}}_{\mathbf{x}}$ [1] before stepping into the standard procedure. Inheriting Eq. (108), $\bar{\mathbf{H}}_{\mathbf{x}}$ is *rank deficient* due to the zero columns corresponding to the robocentric state, however, the rest of columns corresponding to the clone states are linearly independent (i.e., assuming independent observations). Therefore, we can perform QR decomposition as:

$$\bar{\mathbf{H}}_{\mathbf{x}} = [\mathbf{0}_{d \times 24} \quad \bar{\mathbf{T}}_{\mathbf{w}}] \quad (140)$$

$$= [\mathbf{0}_{d \times 24} \quad [\mathbf{Q}_1 \quad \mathbf{Q}_2] \left[\begin{array}{c} \bar{\mathbf{T}}_{\mathbf{w}} \\ \mathbf{0}_{(d-6(N-1)) \times 6(N-1)} \end{array} \right]] \quad (141)$$

$$= [\mathbf{Q}_1 \quad \mathbf{Q}_2] \left[\begin{array}{cc} \mathbf{0}_{6(N-1) \times 24} & \bar{\mathbf{T}}_{\mathbf{w}} \\ \mathbf{0}_{(d-6(N-1)) \times 24} & \mathbf{0}_{(d-6(N-1)) \times 6(N-1)} \end{array} \right] \quad (142)$$

where \mathbf{Q}_1 and \mathbf{Q}_2 are unitary matrices of dimension $d \times 6(N-1)$ and $d \times (d-6(N-1))$, respectively, and $\bar{\mathbf{T}}_{\mathbf{w}}$ is a $6(N-1) \times 6(N-1)$ upper triangular matrix. With this definition, Eq. (139) yields:

$$\bar{\mathbf{r}} = [\mathbf{Q}_1 \quad \mathbf{Q}_2] \left[\begin{array}{c} \mathbf{0} \quad \bar{\mathbf{T}}_{\mathbf{w}} \\ \mathbf{0} \quad \mathbf{0} \end{array} \right] \tilde{\mathbf{x}} + \bar{\mathbf{n}} \quad (143)$$

$$\Rightarrow \left[\begin{array}{c} \mathbf{Q}_1^\top \\ \mathbf{Q}_2^\top \end{array} \right] \bar{\mathbf{r}} = \left[\begin{array}{c} \mathbf{0} \quad \bar{\mathbf{T}}_{\mathbf{w}} \\ \mathbf{0} \quad \mathbf{0} \end{array} \right] \tilde{\mathbf{x}} + \left[\begin{array}{c} \mathbf{Q}_1^\top \\ \mathbf{Q}_2^\top \end{array} \right] \bar{\mathbf{n}} \quad (144)$$

For which after discarding the lower rows which are only about the measurement noise, we employ the upper $6(N-1)$ rows, instead of Eq. (139), as the residual for the following EKF update:

$$\check{\mathbf{r}} = \mathbf{Q}_1^\top \bar{\mathbf{r}} = [\mathbf{0} \quad \bar{\mathbf{T}}_{\mathbf{w}}] \tilde{\mathbf{x}} + \mathbf{Q}_1^\top \bar{\mathbf{n}} = \check{\mathbf{H}}_{\mathbf{x}} \tilde{\mathbf{x}} + \check{\mathbf{n}} \quad (145)$$

where $\check{\mathbf{n}} = \mathbf{Q}_1^\top \bar{\mathbf{n}}$ is the noise vector with the covariance $\check{\mathbf{R}} = \mathbf{Q}_1^\top \mathbb{E}[\bar{\mathbf{n}} \bar{\mathbf{n}}^\top] \mathbf{Q}_1 = \sigma_{im}^2 \mathbf{I}_{6(N-1)}$. In such a way, the dimension of measurement model is reduced to $6(N-1)$. Especially, when $d \gg 6(N-1)$ these can be done by Givens rotations with complexity $O(N^2 d)$. Next, the EKF update is conducted as:

$$\mathbf{K} = \mathbf{P} \check{\mathbf{H}}_{\mathbf{x}}^\top (\check{\mathbf{H}}_{\mathbf{x}} \mathbf{P} \check{\mathbf{H}}_{\mathbf{x}}^\top + \check{\mathbf{R}})^{-1} \quad (146)$$

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K} \check{\mathbf{r}} \quad (147)$$

$$\mathbf{P}_{k+1|k+1} = (\mathbf{I}_{24+6(N-1)} - \mathbf{K} \check{\mathbf{H}}_{\mathbf{x}}) \mathbf{P}_{k+1|k} (\mathbf{I}_{24+6(N-1)} - \mathbf{K} \check{\mathbf{H}}_{\mathbf{x}})^\top + \mathbf{K} \check{\mathbf{R}} \mathbf{K}^\top \quad (148)$$

2.3.4 State Augmentation

In contrast to [1] which clones the pose estimates at the time of receiving a new image, we perform a *delayed clone* after the EKF update. Assuming that at time-step $k+1$ the EKF update is done, we append the copy of the relative pose estimate, $\{\hat{q}_{k+1|k+1}, \hat{\mathbf{p}}_{R_{k+1|k+1}}\}$, to the end of the state vector $\hat{\mathbf{x}}_{k+1|k+1}$. Moreover, for covariance matrix we introduce the following Jacobian to facilitate matrix augmentation:

$$\mathbf{P}_{k+1|k+1} \leftarrow \begin{bmatrix} \mathbf{I}_{24+6(N-1)} \\ \mathbf{J} \end{bmatrix} \mathbf{P}_{k+1|k+1} \begin{bmatrix} \mathbf{I}_{24+6(N-1)} \\ \mathbf{J} \end{bmatrix}^\top \quad (149)$$

$$\mathbf{J} = \begin{bmatrix} \mathbf{0}_{3 \times 9} & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_{3 \times 9} & \mathbf{0}_{3 \times 6(N-1)} \\ \mathbf{0}_{3 \times 9} & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_{3 \times 9} & \mathbf{0}_{3 \times 6(N-1)} \end{bmatrix} \quad (150)$$

2.4 Composition

After the state augmentation we have the state vector in the form of:

$$\hat{\mathbf{x}}_{k+1|k+1} = \begin{bmatrix} R_k \hat{\mathbf{x}}_{k+1|k+1}^\top & \hat{\mathbf{w}}_{k+1|k+1}^\top \end{bmatrix}^\top \quad (151)$$

$$= \begin{bmatrix} R_k \hat{\mathbf{x}}_{G_{k+1|k+1}}^\top & R_k \hat{\mathbf{x}}_{I_{k+1|k+1}}^\top & \mid & {}^2 \hat{\mathbf{q}}^\top & {}^{R_1} \hat{\mathbf{p}}_{R_2}^\top & \dots & {}^N \hat{\mathbf{q}}^\top & {}^{R_N} \hat{\mathbf{p}}_{R_{N+1}}^\top \end{bmatrix}^\top \quad (152)$$

As we shift local frame of reference at every image time, the corresponding IMU frame, $\{I_{k+1}\}$, is selected as the frame of reference (i.e., $\{R_{k+1}\}$). The state estimates in the new frame of reference is obtained through:

$$\hat{\mathbf{x}}_{k+1} = \begin{bmatrix} R_{k+1} \hat{\mathbf{x}}_{k+1} \\ \hat{\mathbf{w}}_{k+1} \end{bmatrix} = \begin{bmatrix} R_k \hat{\mathbf{x}}_{k+1|k+1} \boxplus {}^{R_k} \hat{\mathbf{x}}_{I_{k+1|k+1}} \\ \hat{\mathbf{w}}_{k+1|k+1} \end{bmatrix} \quad (153)$$

$$\begin{bmatrix} {}^{k+1} \hat{\mathbf{q}} \\ {}^{R_{k+1}} \hat{\mathbf{p}}_G \\ {}^{R_{k+1}} \hat{\mathbf{g}} \\ {}^{k+1} \hat{\mathbf{q}} \\ {}^{R_{k+1}} \hat{\mathbf{p}}_{R_{k+1}} \\ \hat{\mathbf{v}}_{R_{k+1}} \\ \hat{\mathbf{b}}_{g_{k+1}} \\ \hat{\mathbf{b}}_{a_{k+1}} \\ \hat{\mathbf{w}}_{k+1} \end{bmatrix} = \begin{bmatrix} {}^{k+1} \hat{\mathbf{q}}_{k+1|k+1} \otimes {}^k \hat{\mathbf{q}}_{k+1|k+1} \\ ({}^k \hat{\mathbf{p}}_{G_{k+1|k+1}} - {}^k \hat{\mathbf{p}}_{I_{k+1|k+1}}) \\ {}^k \hat{\mathbf{g}}_{k+1|k+1} \\ \bar{\mathbf{q}}_0 \\ \mathbf{0}_{3 \times 1} \\ \hat{\mathbf{v}}_{I_{k+1|k+1}} \\ \hat{\mathbf{b}}_{g_{k+1|k+1}} \\ \hat{\mathbf{b}}_{a_{k+1|k+1}} \\ \hat{\mathbf{w}}_{k+1|k+1} \end{bmatrix} \quad (154)$$

where $\bar{\mathbf{q}}_0 = [0, 0, 0, 1]^\top$. Accordingly, the covariance matrix is transformed through:

$$\mathbf{P}_{k+1} = \mathbf{U}_{k+1} \mathbf{P}_{k+1|k+1} \mathbf{U}_{k+1}^\top \quad (155)$$

$$\mathbf{U}_{k+1} = \frac{\partial \tilde{\mathbf{x}}_{k+1}}{\partial \tilde{\mathbf{x}}_{k+1|k+1}} = \begin{bmatrix} \mathbf{V}_{k+1} & \mathbf{0}_{24 \times 6N} \\ \mathbf{0}_{6N \times 24} & \mathbf{I}_{6N} \end{bmatrix}, \quad (156)$$

$$\mathbf{V}_{k+1} = \frac{\partial R_{k+1} \tilde{\mathbf{x}}_{k+1}}{\partial R_k \tilde{\mathbf{x}}_{k+1|k+1}} = \begin{bmatrix} {}^{k+1} \mathbf{C}_{\hat{\mathbf{q}}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & {}^k \mathbf{C}_{\hat{\mathbf{q}}} & \mathbf{0}_3 & \lfloor {}^{R_{k+1}} \hat{\mathbf{p}}_G \times \rfloor & -{}^k \mathbf{C}_{\hat{\mathbf{q}}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & {}^{k+1} \mathbf{C}_{\hat{\mathbf{q}}} & \lfloor {}^{R_{k+1}} \hat{\mathbf{g}} \times \rfloor & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix}_{k+1|k+1} \quad (157)$$

Note that, the relative pose in IMU state is *reset* to the origin of the new frame of reference, $\{R_k\}$, however, the velocity and biases are evolving in the sensor frame (cf. Eq. (15), (16), and (17)), thus are not affected by the change of frame of reference. In particular, through (157) the covariance of the relative pose is also *reset* to zero, i.e., no uncertainties for robocentric frame of reference itself. We outline the proposed approach in Algorithm 1.

Algorithm 1 Robocentric Visual-Inertial Odometry

Input: camera images, and IMU measurements

Output: 6DOF real-time robot poses

R-VIO: Initialize the first frame of reference $\{R_0\}$ ($=\{G\}$) when the first IMU measurement(s) comes in. Then, when a new camera image comes in, do

- **Feature tracking:** extract corner features from the image, then perform KLT tracking and 2-point RANSAC outlier rejection. Record the tracking history for all inliers.
 - **Propagation:** propagate state and covariance matrix using preintegration with all IMU measurements starting from the last image time-step.
 - **Update:**
 - (i) *EKF update:* compute the inverse-depth measurement model matrices for the features whose track are complete (i.e., lost track, or reach the maximum tracking length). Use the features passed the Mahalanobis test for an EKF update.
 - (ii) *State augmentation:* augment state vector and covariance matrix with the *updated* relative pose estimates (state and covariance).
 - **Composition:** shift local frame of reference to current IMU frame, update *global* state and covariance using the relative pose estimates. After that, reset relative pose (state and covariance).
-

2.5 Observability Analysis

System observability reveals whether the information provided by the measurements is sufficient to estimate the state without ambiguity. Since the observability matrix describes the information available in the measurements, by studying its nullspace we gain insight about the directions in the state space from which the system acquires information. To this end, we perform the observability analysis of the proposed robocentric VIO within the EKF-SLAM framework that has the same observability properties as the EKF-VIO provided the same linearization points used [4]. For brevity of presentation, we employ the case that only a single landmark is included in the state vector, while the conclusion can be easily generalized to the case of multiple landmarks. In this case, the state vector at time-step k includes a landmark L :

$$\mathbf{x}_k = \begin{bmatrix} {}^{R_k} \mathbf{x}_G^\top & {}^{R_k} \mathbf{x}_{I_k}^\top & {}^{R_k} \mathbf{p}_L^\top \end{bmatrix}^\top \quad (158)$$

$$= \begin{bmatrix} {}^k \bar{q}^\top & {}^{R_k} \mathbf{p}_G^\top & {}^{R_k} \mathbf{g}^\top & | & {}^k \bar{q}^\top & {}^{R_k} \mathbf{p}_{I_k}^\top & \mathbf{v}_{I_k}^\top & \mathbf{b}_{g_k}^\top & \mathbf{b}_{a_k}^\top & | & {}^{R_k} \mathbf{p}_L^\top \end{bmatrix}^\top \quad (159)$$

and the observability matrix \mathbf{M} is computed as [22]:

$$\mathbf{M} = \begin{bmatrix} \mathbf{H}_k \\ \vdots \\ \mathbf{H}_\ell \Psi_{\ell,k} \\ \vdots \\ \mathbf{H}_{k+m} \Psi_{k+m,k} \end{bmatrix} \quad (160)$$

where $\Psi_{\ell,k}$ is the state transition matrix from time-step k to ℓ , and \mathbf{H}_ℓ is the measurement Jacobian for the landmark observation at time-step ℓ . Each row is evaluated at state, \mathbf{x}_i , for $i = k, \dots, \ell, \dots, k + m$. The nullspace of \mathbf{M} describes the directions of the state space for which no information is provided by the measurements, i.e., the unobservable state subspace. Since the proposed robocentric EKF includes three steps: propagation, update, and composition, and the composition step changes the frame of reference, we analyze the observability for a complete cycle of (i) propagation and update, and (ii) composition. We analytically show that the proposed robocentric VIO system has a constant unobservable subspace, and does *not* have inconsistency issue which is caused by observability mismatch [9, 11], thus improving consistency.

2.5.1 Analytic State Transition Matrix

The analytic form state transition matrix is used for the analysis, for which we integrate the following differential equation over time interval $[t_k, t_\ell]$:

$$\dot{\Psi}(\ell, k) = \check{\mathbf{F}}\Psi(\ell, k) \quad (161)$$

$$\check{\mathbf{F}} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -[\hat{\omega} \times] & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -\frac{\ell}{k}\mathbf{C}_{\hat{q}}^\top [\hat{\mathbf{v}}_{I_\ell} \times] & \mathbf{0}_3 & \frac{\ell}{k}\mathbf{C}_{\hat{q}}^\top & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & -\frac{\ell}{k}\mathbf{C}_{\hat{q}} & -[\hat{\mathbf{g}} \times] & \mathbf{0}_3 & -[\hat{\omega} \times] & -[\hat{\mathbf{v}}_{I_\ell} \times] & -\mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \quad (162)$$

where $\check{\mathbf{F}}$ is the augmented state transition matrix assuming zero dynamics for the landmark position in $\{R_k\}$. By checking Eq. (161) with initial condition $\Psi(k, k) = \mathbf{I}_{27}$, we can get the following structure of Ψ :

$$\Psi(\ell, k) = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \Psi_{44}(\ell, k) & \mathbf{0}_3 & \mathbf{0}_3 & \Psi_{47}(\ell, k) & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \Psi_{53}(\ell, k) & \Psi_{54}(\ell, k) & \mathbf{I}_3 & \Psi_{56}(\ell, k) & \Psi_{57}(\ell, k) & \Psi_{58}(\ell, k) & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \Psi_{63}(\ell, k) & \Psi_{64}(\ell, k) & \mathbf{0}_3 & \Psi_{66}(\ell, k) & \Psi_{67}(\ell, k) & \Psi_{68}(\ell, k) & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \quad (163)$$

We follow an analogous approach to derive the expressions for the block entries of $\Psi(\ell, k)$ (cf. Appendix A), and the results are listed as follows:

$$\Psi_{44}(\ell, k) = \frac{\ell}{k}\mathbf{C}_{\hat{q}} \quad (164)$$

$$\Psi_{47}(\ell, k) = -\frac{\ell}{k}\mathbf{C}_{\hat{q}} \int_{t_k}^{t_\ell} \frac{s}{k}\mathbf{C}_{\hat{q}}^\top ds = - \int_{t_k}^{t_\ell} \frac{\ell}{s}\mathbf{C}_{\hat{q}} ds \quad (165)$$

$$\Psi_{53}(\ell, k) = -\frac{1}{2}\mathbf{I}_3 \Delta t_{k,\ell}^2 \quad (166)$$

$$\Psi_{54}(\ell, k) = -[\left({}^{R_k} \hat{\mathbf{p}}_{I_\ell} + \frac{1}{2} {}^{R_k} \hat{\mathbf{g}} \Delta t_{k,\ell}^2 \right) \times] \quad (167)$$

$$\Psi_{56}(\ell, k) = \mathbf{I}_3 \Delta t_{k,\ell} \quad (168)$$

$$\begin{aligned} \Psi_{57}(\ell, k) &= \int_{t_k}^{t_\ell} [\frac{s}{k}\mathbf{C}_{\hat{q}}^\top \hat{\mathbf{v}}_{I_s} \times] \int_{t_k}^s \frac{\mu}{k}\mathbf{C}_{\hat{q}}^\top d\mu ds + [\left({}^{R_k} \hat{\mathbf{g}} \times \right)] \int_{t_k}^{t_\ell} \int_{t_k}^s \int_{t_k}^\mu \frac{\lambda}{k}\mathbf{C}_{\hat{q}}^\top d\lambda d\mu ds \\ &\quad - \int_{t_k}^{t_\ell} \int_{t_k}^s \frac{\mu}{k}\mathbf{C}_{\hat{q}}^\top [\hat{\mathbf{v}}_{I_\mu} \times] d\mu ds \end{aligned} \quad (169)$$

$$\Psi_{58}(\ell, k) = - \int_{t_k}^{t_\ell} \int_{t_k}^s \frac{\mu}{k}\mathbf{C}_{\hat{q}}^\top d\mu ds \quad (170)$$

$$\Psi_{63}(\ell, k) = -\frac{\ell}{k}\mathbf{C}_{\hat{q}} \Delta t_{k,\ell} \quad (171)$$

$$\Psi_{64}(\ell, k) = -\frac{\ell}{k}\mathbf{C}_{\hat{q}} [\left({}^{R_k} \hat{\mathbf{g}} \times \right) \Delta t_{k,\ell}] \quad (172)$$

$$\Psi_{66}(\ell, k) = \Psi_{44}(\ell, k) \quad (173)$$

$$\Psi_{67}(\ell, k) = \frac{\ell}{k}\mathbf{C}_{\hat{q}} [\left({}^{R_k} \hat{\mathbf{g}} \times \right)] \int_{t_k}^{t_\ell} \int_{t_k}^s \frac{\mu}{k}\mathbf{C}_{\hat{q}}^\top d\mu ds - \int_{t_k}^{t_\ell} \frac{\ell}{s}\mathbf{C}_{\hat{q}} [\left(\hat{\mathbf{v}}_{I_s} \times \right)] ds \quad (174)$$

$$\Psi_{68}(\ell, k) = -\int_{t_k}^{t_\ell} {}_k^s \mathbf{C}_{\hat{q}}^\top ds = -\int_{t_k}^{t_\ell} {}_s^k \mathbf{C}_{\hat{q}} ds \quad (175)$$

2.5.2 Measurement Jacobian

At time-step $\ell \in [t_k, t_{k+m})$ the estimate of landmark position in $\{I_\ell\}$ can be expressed as

$${}_{I_\ell} \hat{\mathbf{p}}_L = {}_k^{\ell} \mathbf{C}_{\hat{q}} ({}^{R_k} \hat{\mathbf{p}}_L - {}^{R_k} \hat{\mathbf{p}}_{I_\ell}) \quad (176)$$

According to Eq. (86), the bearing-only measurement model can be expressed as:

$$\mathbf{z}_\ell = \frac{1}{z} \begin{bmatrix} x \\ y \end{bmatrix} + \mathbf{n}_\ell, \quad {}^C \mathbf{p}_L = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = {}_I^C \mathbf{C}_{\bar{q}k} {}_k^{\ell} \mathbf{C}_{\hat{q}} ({}^{R_k} \mathbf{p}_L - {}^{R_k} \mathbf{p}_{I_\ell}) + {}^C \mathbf{p}_I \quad (177)$$

where $\{{}_I^C \bar{q}, {}^C \mathbf{p}_I\}$ is the result of camera-IMU calibration. The corresponding measurement Jacobian is:

$$\begin{aligned} \mathbf{H}_\ell &= \mathbf{H}_{\text{proj}} {}_I^C \mathbf{C}_{\bar{q}k} {}_k^{\ell} \mathbf{C}_{\hat{q}} [\mathbf{0}_3 \quad \mathbf{0}_3 \quad \mathbf{0}_3 \mid \mathbf{H}_{\theta_\ell} \quad -\mathbf{I}_3 \quad \mathbf{0}_{3 \times 9} \mid \mathbf{I}_3], \\ \mathbf{H}_{\text{proj}} &= \frac{1}{\hat{z}} \begin{bmatrix} 1 & 0 & -\frac{\hat{x}}{\hat{z}} \\ 0 & 1 & -\frac{\hat{y}}{\hat{z}} \end{bmatrix}, \\ \mathbf{H}_{\theta_\ell} &= [({}^{R_k} \hat{\mathbf{p}}_L - {}^{R_k} \hat{\mathbf{p}}_{I_\ell}) \times] {}_k^{\ell} \mathbf{C}_{\hat{q}}^\top \end{aligned} \quad (178)$$

2.5.3 Propagation and Update

Based on the results we have, we can write out the ℓ -th row, \mathbf{M}_ℓ , of \mathbf{M} as follows:

$$\mathbf{M}_\ell = \mathbf{H}_\ell \Psi_{\ell,k} \quad (179)$$

$$\mathbf{H}_\ell = \mathbf{H}_{\text{proj}} {}_I^C \mathbf{C}_{\bar{q}k} {}_k^{\ell} \mathbf{C}_{\hat{q}} [\mathbf{0}_3 \quad \mathbf{0}_3 \quad \mathbf{0}_3 \mid \mathbf{H}_{\theta_\ell} \quad -\mathbf{I}_3 \quad \mathbf{0}_{3 \times 9} \mid \mathbf{I}_3] \quad (180)$$

$$\mathbf{H}_{\theta_\ell} = [({}^{R_k} \hat{\mathbf{p}}_L - {}^{R_k} \hat{\mathbf{p}}_{I_\ell}) \times] {}_k^{\ell} \mathbf{C}_{\hat{q}}^\top \quad (181)$$

$$\Rightarrow \mathbf{M}_\ell = \Pi \Gamma = \Pi [\mathbf{0}_3 \quad \mathbf{0}_3 \quad \Gamma_1 \mid \Gamma_2 \quad -\mathbf{I}_3 \quad \Gamma_3 \quad \Gamma_4 \quad \Gamma_5 \mid \mathbf{I}_3] \quad (182)$$

where

$$\Pi = \mathbf{H}_{\text{proj}} {}_I^C \mathbf{C}_{\bar{q}k} {}_k^{\ell} \mathbf{C}_{\hat{q}} \quad (183)$$

$$\Gamma_1 = -\Psi_{53} = \frac{1}{2} \mathbf{I}_3 \Delta t_{k,\ell}^2 \quad (184)$$

$$\Gamma_2 = [({}^{R_k} \hat{\mathbf{p}}_L - {}^{R_k} \hat{\mathbf{p}}_{I_\ell}) \times] {}_k^{\ell} \mathbf{C}_{\hat{q}}^\top \Psi_{44} - \Psi_{54} = [{}^{R_k} \hat{\mathbf{p}}_L \times] + \frac{1}{2} [{}^{R_k} \hat{\mathbf{g}} \times] \Delta t_{k,\ell}^2 \quad (185)$$

$$\Gamma_3 = -\Psi_{56} = -\mathbf{I}_3 \Delta t_{k,\ell} \quad (186)$$

$$\Gamma_4 = [({}^{R_k} \hat{\mathbf{p}}_L - {}^{R_k} \hat{\mathbf{p}}_{I_\ell}) \times] {}_k^{\ell} \mathbf{C}_{\hat{q}}^\top \Psi_{47} - \Psi_{57} = -[({}^{R_k} \hat{\mathbf{p}}_L - {}^{R_k} \hat{\mathbf{p}}_{I_\ell}) \times] \int_{t_k}^{t_\ell} {}_s^k \mathbf{C}_{\hat{q}}^\top ds - \Psi_{57} \quad (187)$$

$$\Gamma_5 = -\Psi_{58} \quad (188)$$

We note that for generic motion, i.e., $\omega \neq \mathbf{0}_{3 \times 1}$ and $\mathbf{a} \neq \mathbf{0}_{3 \times 1}$, the values of Ψ_{57} and Ψ_{58} are time-varying, then Γ_4 and Γ_5 are linearly independent. Moreover, the value of $\Delta t_{k,\ell}$ is varying for different time interval, then the stacked Γ_1 , Γ_2 , and Γ_3 are linearly independent. Based on that, we perform the following Gaussian elimination on \mathbf{M}_ℓ to facilitate the search for the nullspace:

$$\mathbf{M}_\ell = \Pi [\underline{\mathbf{0}_3} \quad \underline{\mathbf{0}_3} \quad \Gamma_1 \mid \Gamma_2 \quad -\mathbf{I}_3 \quad \Gamma_3 \quad \Gamma_4 \quad \Gamma_5 \mid \mathbf{I}_3] \quad (189)$$

$$\Rightarrow \mathbf{M}_\ell \sim \Pi [\underline{\mathbf{0}_3} \quad \underline{\mathbf{0}_3} \quad \Gamma_1 \mid \Gamma_2 \quad -\mathbf{I}_3 \quad \Gamma_3 \quad \Gamma_4 \quad \Gamma_5 \mid \underline{\mathbf{0}_3}] \quad (190)$$

from which we can find that \mathbf{M}_ℓ is rank deficient by 9, and accordingly the nullspace is of rank 9. Specifically, $\forall \ell \geq k$, we can find that the nullspace of \mathbf{M} consists of the following *nine* directions:

$$\text{null}(\mathbf{M}) = \text{span}_{\text{col.}} \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \quad (191)$$

Remark 1. The first 6DOF correspond to the orientation (3) and position (3) of the global frame.

Remark 2. The last 3DOF belong to the same translation (3) simultaneously applied to the landmark(s) and sensor.

2.5.4 Composition

Assuming that the estimates of ${}^R_k \mathbf{x}_\ell$ and ${}^R_k \mathbf{p}_L$ are obtained at time-step ℓ , we have the following linear dynamic model from time-step k to ℓ including the state composition:

$$\tilde{\mathbf{x}}_\ell = \check{\mathbf{V}}_\ell \check{\Psi}_{\ell,k} \tilde{\mathbf{x}}_k = \check{\Psi}_{\ell,k} \tilde{\mathbf{x}}_k \quad (192)$$

$$\check{\mathbf{V}}_\ell = \begin{bmatrix} {}^\ell_k \mathbf{C}_{\hat{q}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & {}^\ell_k \mathbf{C}_{\hat{q}} & \mathbf{0}_3 & \lfloor {}^{R_\ell} \hat{\mathbf{p}}_G \times \rfloor & -{}^\ell_k \mathbf{C}_{\hat{q}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & {}^\ell_k \mathbf{C}_{\hat{q}} & \lfloor {}^{R_\ell} \hat{\mathbf{g}} \times \rfloor & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \lfloor {}^{R_\ell} \hat{\mathbf{p}}_L \times \rfloor & -{}^\ell_k \mathbf{C}_{\hat{q}} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & {}^\ell_k \mathbf{C}_{\hat{q}} \end{bmatrix} \quad (193)$$

$$\check{\Psi}_{\ell,k} = \begin{bmatrix} \check{\Psi}_{11} & \mathbf{0}_3 & \mathbf{0}_3 & \check{\Psi}_{14} & \mathbf{0}_3 & \mathbf{0}_3 & \check{\Psi}_{17} & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \check{\Psi}_{22} & \check{\Psi}_{23} & \check{\Psi}_{24} & \check{\Psi}_{25} & \check{\Psi}_{26} & \check{\Psi}_{27} & \check{\Psi}_{28} & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \check{\Psi}_{33} & \check{\Psi}_{34} & \mathbf{0}_3 & \mathbf{0}_3 & \check{\Psi}_{37} & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \check{\Psi}_{63} & \check{\Psi}_{64} & \mathbf{0}_3 & \check{\Psi}_{66} & \check{\Psi}_{67} & \check{\Psi}_{68} & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \check{\Psi}_{93} & \check{\Psi}_{94} & \check{\Psi}_{95} & \check{\Psi}_{96} & \check{\Psi}_{97} & \check{\Psi}_{98} & \check{\Psi}_{99} \end{bmatrix} \quad (194)$$

$$\check{\Psi}_{11} = {}^\ell_k \mathbf{C}_{\hat{q}} \quad (195)$$

$$\check{\Psi}_{14} = \Psi_{44} \quad (196)$$

$$\check{\Psi}_{17} = \Psi_{47} \quad (197)$$

$$\check{\Psi}_{22} = {}^\ell_k \mathbf{C}_{\hat{q}} \quad (198)$$

$$\check{\Psi}_{23} = -{}^\ell_k \mathbf{C}_{\hat{q}} \Psi_{53} \quad (199)$$

$$\check{\Psi}_{24} = \lfloor {}^{R_\ell} \hat{\mathbf{p}}_G \times \rfloor \Psi_{44} - {}^\ell_k \mathbf{C}_{\hat{q}} \Psi_{54} \quad (200)$$

$$\check{\Psi}_{25} = -{}^\ell_k \mathbf{C}_{\hat{q}} \quad (201)$$

$$\check{\Psi}_{26} = -\frac{\ell}{k} \mathbf{C}_{\hat{q}} \Psi_{56} \quad (202)$$

$$\check{\Psi}_{27} = \lfloor^{R_\ell} \hat{\mathbf{p}}_G \times \rfloor \Psi_{47} - \frac{\ell}{k} \mathbf{C}_{\hat{q}} \Psi_{57} \quad (203)$$

$$\check{\Psi}_{28} = -\frac{\ell}{k} \mathbf{C}_{\hat{q}} \Psi_{58} \quad (204)$$

$$\check{\Psi}_{33} = \frac{\ell}{k} \mathbf{C}_{\hat{q}} \quad (205)$$

$$\check{\Psi}_{34} = \lfloor^{R_\ell} \hat{\mathbf{g}} \times \rfloor \Psi_{44} \quad (206)$$

$$\check{\Psi}_{37} = \lfloor^{R_\ell} \hat{\mathbf{g}} \times \rfloor \Psi_{47} \quad (207)$$

$$\check{\Psi}_{63} = \Psi_{63} \quad (208)$$

$$\check{\Psi}_{64} = \Psi_{64} \quad (209)$$

$$\check{\Psi}_{66} = \Psi_{66} \quad (210)$$

$$\check{\Psi}_{67} = \Psi_{67} \quad (211)$$

$$\check{\Psi}_{68} = \Psi_{68} \quad (212)$$

$$\check{\Psi}_{93} = -\frac{\ell}{k} \mathbf{C}_{\hat{q}} \Psi_{53} \quad (213)$$

$$\check{\Psi}_{94} = \lfloor^{R_\ell} \hat{\mathbf{p}}_L \times \rfloor \Psi_{44} - \frac{\ell}{k} \mathbf{C}_{\hat{q}} \Psi_{54} \quad (214)$$

$$\check{\Psi}_{95} = -\frac{\ell}{k} \mathbf{C}_{\hat{q}} \quad (215)$$

$$\check{\Psi}_{96} = -\frac{\ell}{k} \mathbf{C}_{\hat{q}} \Psi_{56} \quad (216)$$

$$\check{\Psi}_{97} = \lfloor^{R_\ell} \hat{\mathbf{p}}_L \times \rfloor \Psi_{47} - \frac{\ell}{k} \mathbf{C}_{\hat{q}} \Psi_{57} \quad (217)$$

$$\check{\Psi}_{98} = -\frac{\ell}{k} \mathbf{C}_{\hat{q}} \Psi_{58} \quad (218)$$

$$\check{\Psi}_{99} = \frac{\ell}{k} \mathbf{C}_{\hat{q}} \quad (219)$$

Accordingly at time-step ℓ the measurement model of Eq. (177) becomes linear:

$$\mathbf{z}_\ell = {}^{R_\ell} \mathbf{p}_L + \mathbf{n}_\ell, \quad {}^{R_\ell} \mathbf{p}_L = \frac{\ell}{k} \mathbf{C}_{\hat{q}} \left({}^{R_k} \mathbf{p}_L - {}^{R_k} \mathbf{p}_{I_\ell} \right) \quad (220)$$

for which the measurement Jacobian with respect to $\tilde{\mathbf{x}}_\ell$ is in the form of:

$$\check{\mathbf{H}}_\ell = [\mathbf{0}_3 \quad \mathbf{0}_3 \quad \mathbf{0}_3 \mid \mathbf{0}_3 \quad \mathbf{0}_3 \quad \mathbf{0}_3 \quad \mathbf{0}_3 \quad \mathbf{0}_3 \mid \mathbf{I}_3] \quad (221)$$

Therefore, for time interval $[t_k, t_\ell]$, after composition we have the corresponding row \mathbf{M}_ℓ , of \mathbf{M} as follows:

$$\mathbf{M}_\ell = \check{\mathbf{H}}_\ell \check{\Psi}_{\ell,k} \quad (222)$$

$$= [\mathbf{0}_3 \quad \mathbf{0}_3 \quad \check{\Psi}_{93} \quad \check{\Psi}_{94} \quad \check{\Psi}_{95} \quad \check{\Psi}_{96} \quad \check{\Psi}_{97} \quad \check{\Psi}_{98} \quad \check{\Psi}_{99}] \quad (223)$$

$$= [\mathbf{0}_3 \quad \mathbf{0}_3 \quad \check{\Psi}_{93} \quad \check{\Psi}_{94} \quad -\frac{\ell}{k} \mathbf{C}_{\hat{q}} \quad \check{\Psi}_{96} \quad \check{\Psi}_{97} \quad \check{\Psi}_{98} \quad \frac{\ell}{k} \mathbf{C}_{\hat{q}}] \quad (224)$$

for which in the generic motion case, i.e., $\omega \neq \mathbf{0}_{3 \times 1}$, and $\mathbf{a} \neq \mathbf{0}_{3 \times 1}$, $\check{\Psi}_{93}$, $\check{\Psi}_{94}$, $\check{\Psi}_{96}$, $\check{\Psi}_{97}$, and $\check{\Psi}_{98}$ are linearly independent, and the same nullspace as in the propagation and update can be obtained:

$$\text{null}(\mathbf{M}) = \text{span}_{\text{col.}} \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \quad (225)$$

Remark 3. *Changing frame of reference does not change the unobservable subspace.*

Remark 4. *The proposed robocentric system has a constant unobservable subspace, i.e., independent of the linearization points, which not only guarantees the correct unobservable dimensions, but also the desired unobservable directions, thus improving consistency.*

2.6 Initialization

It is important to point out that using the proposed robocentric formulation, the filter initialization becomes much simpler, as the states, by design, are simply relative to the sensor's local frame of reference and typically start from zero *without* the need to align the initial pose with the fixed global frame. In particular, in our implementation, (i) the initial global pose and IMU relative pose are both set to $\{\bar{q}_0, \mathbf{0}_{3 \times 1}\}$, (ii) the initial local gravity is the average of first available accelerometer measurement(s) before moving, and (iii) the initial acceleration bias is obtained by removing gravity effects while the initial gyroscope bias is the average of the corresponding stationary measurements. Similarly, the corresponding uncertainties for the poses are set to *zero*, while for the local gravity and biases are set to: $\Sigma_g = \Delta T \sigma_a^2 \mathbf{I}_3$, $\Sigma_{b_g} = \Delta T \sigma_{wg}^2 \mathbf{I}_3$, and $\Sigma_{b_a} = \Delta T \sigma_{wa}^2 \mathbf{I}_3$, where ΔT is the time length of initialization.

3 Simulation Results

We conducted a series of Monte Carlo simulations under realistic conditions to validate the proposed algorithm. Two metrics are employed to evaluate the performance: (i) the root mean squared error (RMSE), and (ii) the normalized estimation error squared (NEES). The RMSE provides a concise metric of the accuracy of a given filter, while the NEES is a standard criterion for evaluating the consistency. We compared with two world-centric counterparts: the standard MSCKF [1], and the state-of-the-art state-transition observability constrained (STOC)-MSCKF [12] by enforcing correct observability to improve consistency. To ensure a fair comparison, we implemented the compared filters using the same parameters, such as the sliding-window size, and processed the same data (i.e., 50 trails for the scenario of Figure 1 under real MEMS sensor noise and bias level [12]). The comparison results are in Figure 2 (the calculation details refers to [9]) and Table 1 provides average RMSE and NEES values for all of the algorithms, which clearly show that the proposed R-VIO significantly outperforms the standard MSCKF as well as the STOC-MSCKF, in terms of both RMSE (accuracy) and NEES (consistency), attributed to the novel reformulation of the system. Note that in Figure 2c the orientation NEES of proposed method has a jump at the beginning which is primarily due to the small covariance we used for initialization (cf. Section 2.6), however, it can quickly recover and perform consistently only after a short period of time.

Table 1: Average RMSE and NEES results corresponding to Figure 2.

Algorithm	Orientation RMSE ($^\circ$)	Position RMSE (m)	Orientation NEES	Position NEES
Std-MSCKF	3.4700	0.4774	7.0487	5.8103
STOC-MSCKF	2.5232	0.4305	4.0964	3.7936
Proposed	0.6811	0.0715	2.4146	1.9061

4 Experimental Results

We further validate the proposed R-VIO experimentally in both indoor and outdoor environments, using both the benchmark public datasets and the datasets collected by our own sensor platform. As described in Algorithm 1, we implemented a C++ multithread framework. In *front end*, the visual tracking thread extracts features from the image using the Shi-Tomasi corner detector [23], then tracks them pairwise from the previous image using the KLT algorithm [24]. To remove the outliers from tracking result, we realized an IMU-aided two-point RANSAC algorithm [25]. Especially, for the feature lost track we use all its measurements within the sliding window for EKF update, while for the one reaching the maximum tracking length (e.g., the window size) we use its subset (e.g., 1/2) of measurements and maintain the rest for next update. Once the visual tracking is done, the *back end* processes the measurements using the proposed robocentric EKF. In particular, we present extensive comparison results against a state-of-the-art implementation, Okvis (i.e., open keyframe-based visual-inertial SLAM)², which does nonlinear optimization in the back end lit-

²<https://github.com/ethz-asl/okvis>

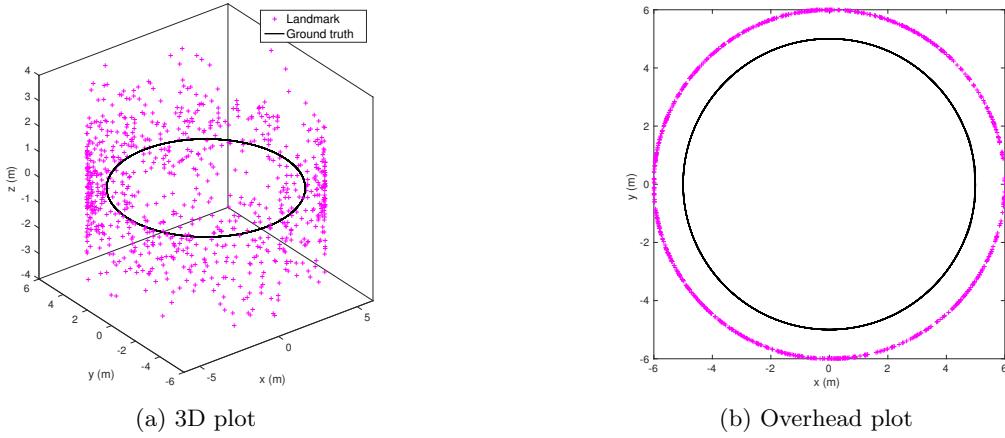


Figure 1: The scenario (i.e., trajectory and landmarks) in simulation. The camera/IMU pair moves along a circular path of radius 5m (**black**) at an average speed 1m/s. The camera with 45° field of view observes point features randomly distributed on a circumscribing cylinder of radius 6m (**pink**). The standard deviation of image noise was set to 1.5 pixels, while IMU measurements were modelled with MEMS sensor quality.

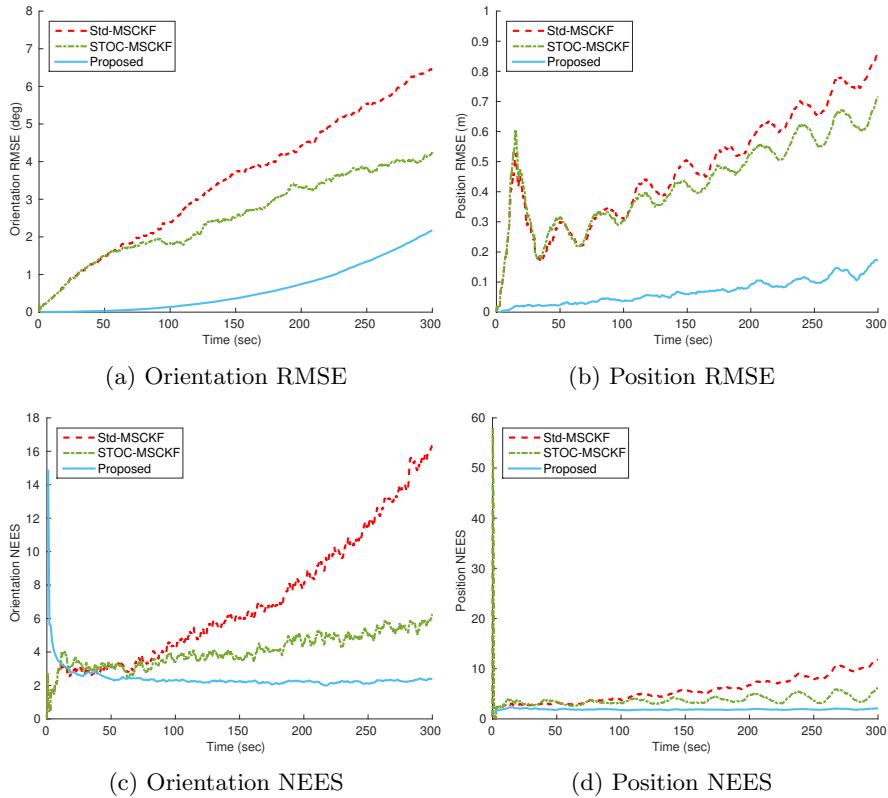


Figure 2: The statistical results of 50 Monte Carlo simulations.

erately incurring higher computational cost, however, taking the advantage of iterated linearization to limit linearization errors [5]. All the tests run *real time* on a Core i7-4710MQ @ 2.5GHz × 8, 16GB RAM laptop.

4.1 Highly Dynamic Capability

To show the robustness of the proposed R-VIO under highly dynamic motion, we consider the EuRoC micro aerial vehicle (MAV) dataset where the VI-sensor (ADIS16448 IMU @ 200Hz, MT9V034 cameras @ 20Hz) mounted on a FireFly hex-rotor helicopter was used for data collection [26]. During the tests, only the left camera (cam0) images are used as vision input. The Okvis was set with default parameters, while the R-VIO was set up to maintain 20 relative poses (i.e., about 1 second relative motion record), and uniformly extracted 200 corner features per image. Note that the computational complexity of R-VIO is linear to the number of features for EKF update (cf. Section 2.3), thus significantly reduces the computational cost. The estimated trajectories are shown in Figure 3 and we provide the corresponding pose error results in Figure 4. In general, the proposed R-VIO performs comparably to the Okvis, even *better* in some sequences.

4.2 Long-Term Sustainability

Aiming at long-term navigation, we further evaluated the accuracy and consistency of the proposed R-VIO with a *Long distance* dataset, which contains a 14 minutes, 1.2km circular path recorded by the VI-Sensor and has been employed in [5]. We compared the results with [5] which included the results of Okvis and an MSCKF implementation of [27]. Figure 5 shows the estimated trajectories after 6DOF pose alignment, where the z -drifts are also presented to illustrate the altitude performance.

4.3 Versatility

We have also validated the proposed R-VIO both indoor and outdoor, using our own sensor platform that consists of: a MicroStrain 3DM-GX3-35 IMU (@ 500Hz) and a PointGrey Chameleon3 camera (@ 30Hz), and is mounted onto the laptop. Both daytime and nighttime datasets were collected for the indoor test, where we traveled about 150m loop at an average speed of 0.539m/s, covering two floors in a building (with white wall, variant illumination, and strong glare in the hallway), while the outdoor dataset recorded about 360m loop at an average speed of 1.216m/s (with uneven terrain, and opportunistic moving objects). In both tests, due to the lack of the ground truth, to visualize the performance we overlay the results onto the floor plan and the Google Map, respectively (see Figure 6b,6d). The final position errors are about 0.349% (daytime) and 0.615% (nighttime) over the travelled distance in the indoor tests, and about 1.173% in the outdoor test.

5 Conclusion

In this report, we have reformulated the 3D VINS with respect to a moving local frame and developed a lightweight, high-precision, robocentric visual-inertial odometry (R-VIO) algorithm. With this novel reformulation, the resulting VINS does not suffer from the observability mismatch issue encountered in the world-centric systems, thus improving consistency and accuracy. Extensive Monte Carlo simulations and the real-world experiments with different sensor platforms navigating in different environments and using only monocular vision validate our theoretical analysis and show that the proposed R-VIO is versatile and robust to different types of motions and environments. In future, we will focus on improving the proposed approach further, for example, by integrating online calibration and loop closure to deal with sensor parameter variations and bound localization errors.

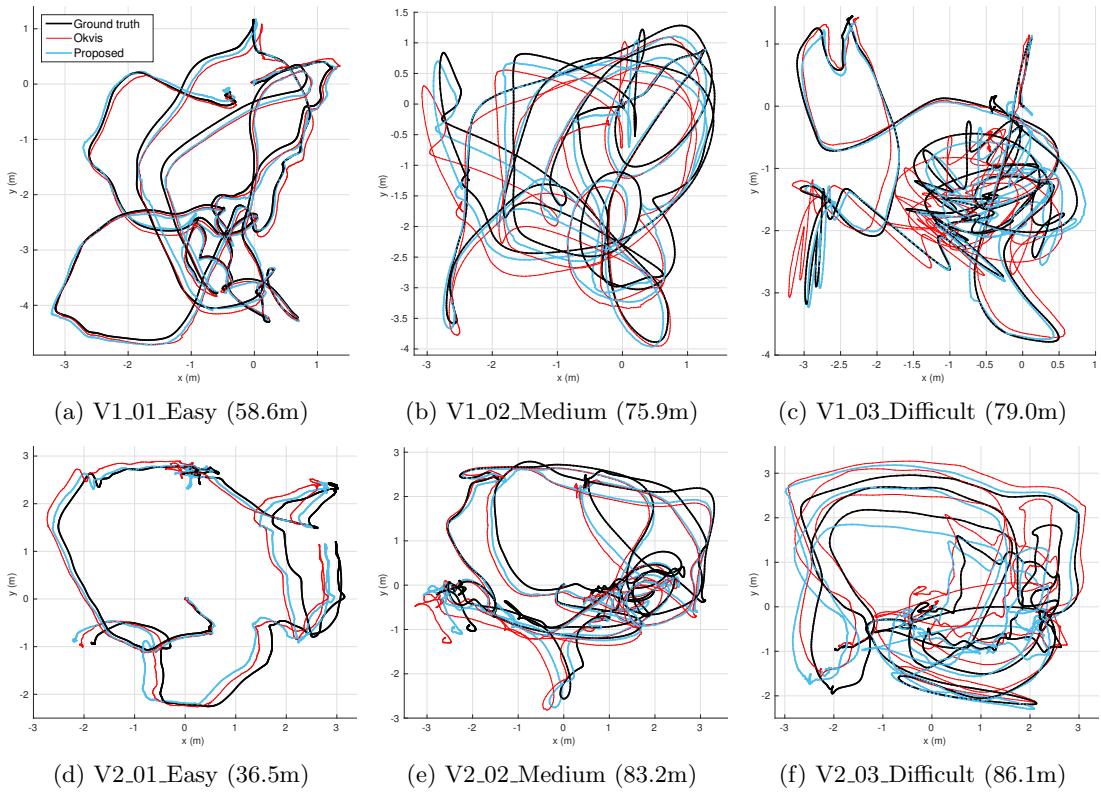


Figure 3: Trajectory estimates vs. ground truth in EuRoC dataset.

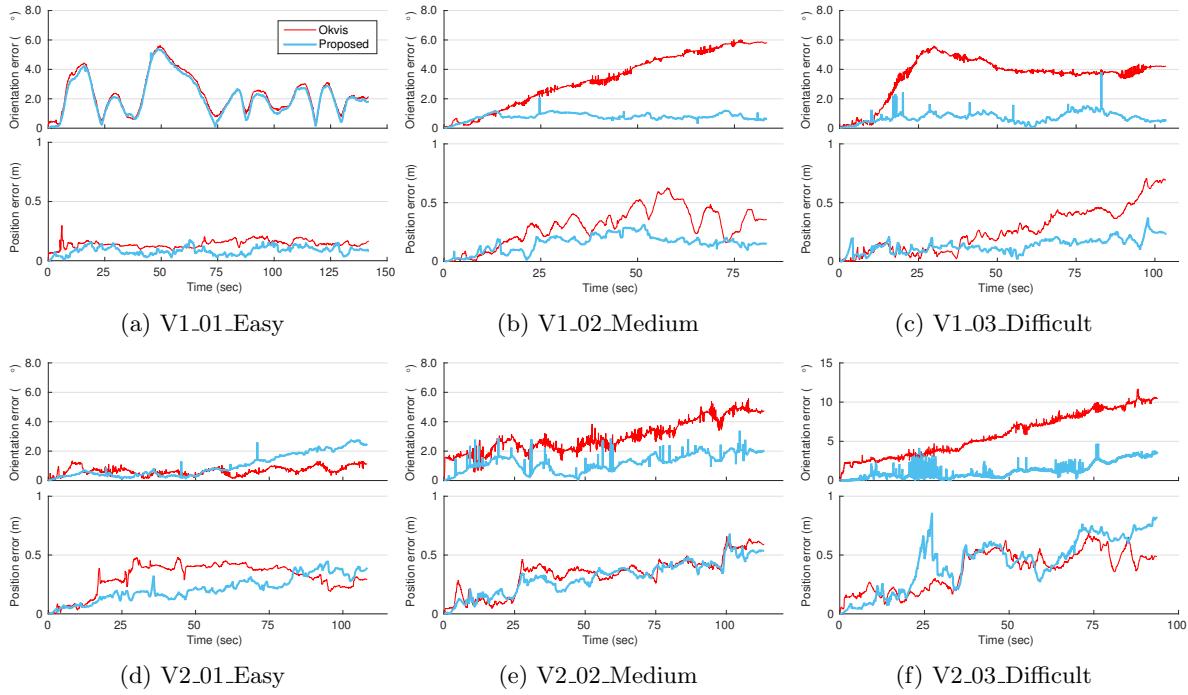


Figure 4: Estimation accuracy (RMSE) in EuRoC dataset.

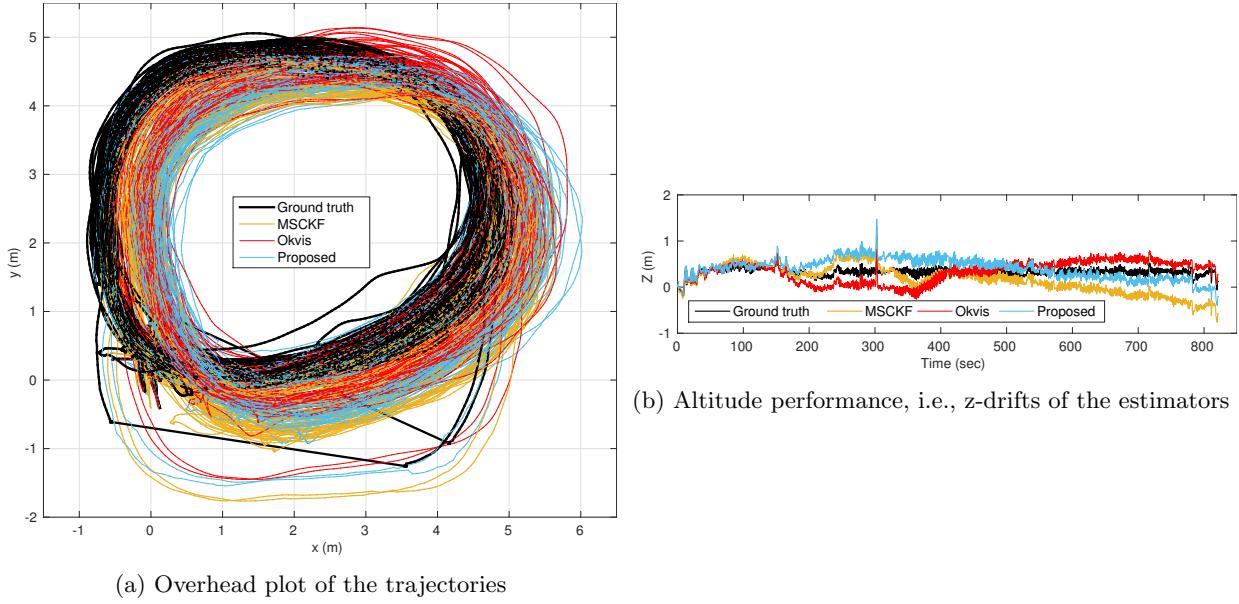


Figure 5: The results vs. ground truth in *Long distance* dataset

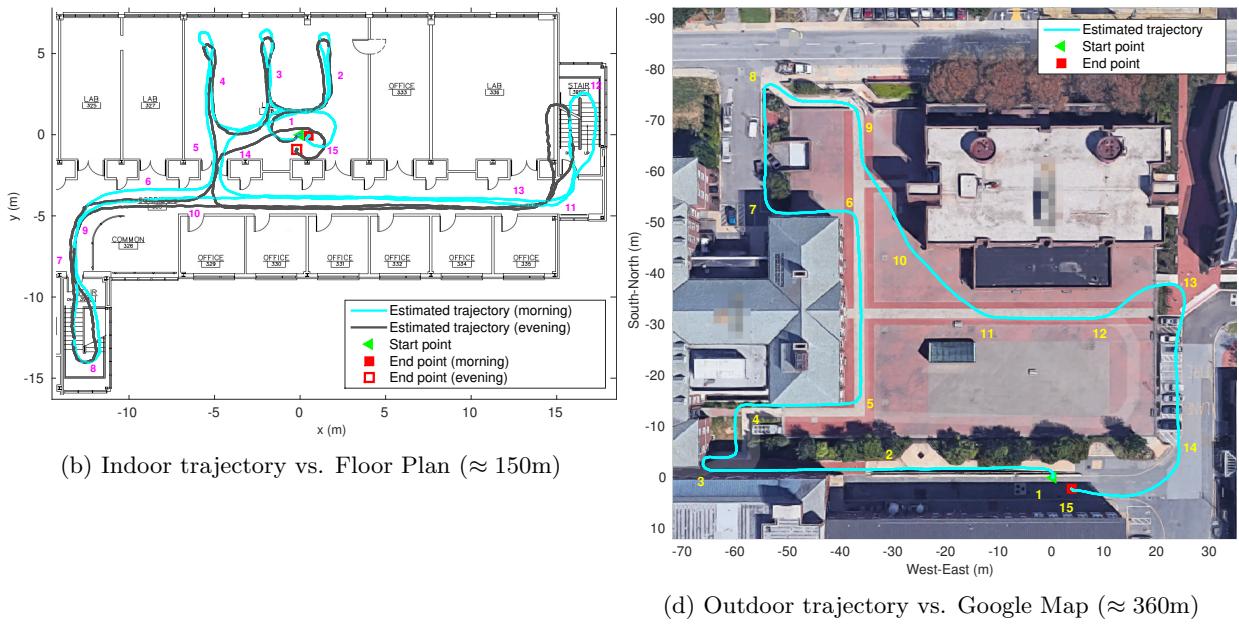


Figure 6: The results in our *Hand-held* experiments

Appendix A: Analytic State Transition Matrix

In this part, we analytically derive the block entries of $\Psi(\ell, k)$ employed by Eq. (164)-(175) in Section 2.5.1:

$$\dot{\Psi}(\ell, k) = \check{\mathbf{F}}\Psi(\ell, k) \quad (226)$$

$$\check{\mathbf{F}} = \begin{bmatrix} \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -[\hat{\omega} \times] & \mathbf{0}_3 & \mathbf{0}_3 & -\mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & -[\ell \mathbf{C}_{\hat{q}}^\top [\hat{\mathbf{v}}_{I_\ell} \times]] & \mathbf{0}_3 & \ell \mathbf{C}_{\hat{q}}^\top & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & -\ell \mathbf{C}_{\hat{q}} & -[\ell \hat{\mathbf{g}} \times] & \mathbf{0}_3 & -[\hat{\omega} \times] & -[\hat{\mathbf{v}}_{I_\ell} \times] & -\mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 \end{bmatrix} \quad (227)$$

$$\Psi(\ell, k) = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \Psi_{44}(\ell, k) & \mathbf{0}_3 & \mathbf{0}_3 & \Psi_{47}(\ell, k) & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \Psi_{53}(\ell, k) & \Psi_{54}(\ell, k) & \mathbf{I}_3 & \Psi_{56}(\ell, k) & \Psi_{57}(\ell, k) & \Psi_{58}(\ell, k) & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \Psi_{63}(\ell, k) & \Psi_{64}(\ell, k) & \mathbf{0}_3 & \Psi_{66}(\ell, k) & \Psi_{67}(\ell, k) & \Psi_{68}(\ell, k) & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \quad (228)$$

Ψ_{44} :

$$\dot{\Psi}_{44}(\ell, k) = -[\hat{\omega} \times] \Psi_{44}(\ell, k) \quad (229)$$

$$\Rightarrow \Psi_{44}(\ell, k) = \Psi_{44}(k, k) \exp \left(\int_{t_k}^{t_\ell} -[\hat{\omega} \times] ds \right) = \exp \left(\int_{t_k}^{t_\ell} -[\hat{\omega} \times] ds \right) = \ell \mathbf{C}_{\hat{q}} \quad (230)$$

Ψ_{47} :

$$\dot{\Psi}_{47}(\ell, k) = -[\hat{\omega} \times] \Psi_{47}(\ell, k) - \mathbf{I}_3 \quad (231)$$

$$\Rightarrow \dot{\Psi}_{47}(\ell, k) + [\hat{\omega} \times] \Psi_{47}(\ell, k) = -\mathbf{I}_3 \quad (232)$$

$$\Rightarrow \exp \left(\int_{t_k}^{t_\ell} [\hat{\omega} \times] ds \right) \dot{\Psi}_{47}(\ell, k) + \exp \left(\int_{t_k}^{t_\ell} [\hat{\omega} \times] ds \right) [\hat{\omega} \times] \Psi_{47}(\ell, k) = -\exp \left(\int_{t_k}^{t_\ell} [\hat{\omega} \times] ds \right) \quad (233)$$

$$\Rightarrow \frac{d}{dt} \left(\exp \left(\int_{t_k}^{t_\ell} [\hat{\omega} \times] ds \right) \Psi_{47}(\ell, k) \right) = -\exp \left(\int_{t_k}^{t_\ell} [\hat{\omega} \times] ds \right) \quad (234)$$

$$\Rightarrow \frac{d}{dt} \left(\ell \mathbf{C}_{\hat{q}}^\top \Psi_{47}(\ell, k) \right) = -\ell \mathbf{C}_{\hat{q}}^\top \quad (235)$$

$$\Rightarrow \ell \mathbf{C}_{\hat{q}}^\top \Psi_{47}(\ell, k) + \Psi_{47}(k, k) = - \int_{t_k}^{t_\ell} \ell \mathbf{C}_{\hat{q}}^\top ds \quad (236)$$

$$\Rightarrow \Psi_{47}(\ell, k) = -\ell \mathbf{C}_{\hat{q}} \int_{t_k}^{t_\ell} \ell \mathbf{C}_{\hat{q}}^\top ds = - \int_{t_k}^{t_\ell} \ell \mathbf{C}_{\hat{q}} ds \quad (237)$$

Ψ_{63} :

$$\dot{\Psi}_{63}(\ell, k) = -\ell \mathbf{C}_{\hat{q}} - [\hat{\omega} \times] \Psi_{63}(\ell, k) \quad (238)$$

$$\Rightarrow \dot{\Psi}_{63}(\ell, k) + [\hat{\omega} \times] \Psi_{63}(\ell, k) = -\ell \mathbf{C}_{\hat{q}} \quad | \text{cf. Eq. (233) -- (235)} \quad (239)$$

$$\Rightarrow \frac{d}{dt} \left(\ell \mathbf{C}_{\hat{q}}^\top \Psi_{63}(\ell, k) \right) = -\ell \mathbf{C}_{\hat{q}}^\top \ell \mathbf{C}_{\hat{q}} = \mathbf{I}_3 \quad (240)$$

$$\Rightarrow {}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top} \Psi_{63}(\ell, k) + \Psi_{63}(k, k) = - \int_{t_k}^{t_{\ell}} \mathbf{I}_3 \, ds \quad (241)$$

$$\Rightarrow \Psi_{63}(\ell, k) = - {}_k^{\ell}\mathbf{C}_{\hat{q}} \Delta t_{k, \ell} \quad (242)$$

Ψ_{64} :

$$\dot{\Psi}_{64}(\ell, k) = -[\hat{\omega} \times] \Psi_{44}(\ell, k) - [\hat{\omega} \times] \Psi_{64}(\ell, k) \quad (243)$$

$$\Rightarrow \dot{\Psi}_{64}(\ell, k) + [\hat{\omega} \times] \Psi_{64}(\ell, k) = -[\hat{\omega} \times] {}_k^{\ell}\mathbf{C}_{\hat{q}} \quad | \text{cf. Eq. (233) - (235)} \quad (244)$$

$$\Rightarrow \frac{d}{dt} \left({}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top} \Psi_{64}(\ell, k) \right) = - {}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top} [\hat{\omega} \times] {}_k^{\ell}\mathbf{C}_{\hat{q}} = -[{}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top}] [\hat{\omega} \times] = -[R_k \hat{\mathbf{g}} \times] \quad (245)$$

$$\Rightarrow {}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top} \Psi_{64}(\ell, k) + \Psi_{64}(k, k) = - \int_{t_k}^{t_{\ell}} [R_k \hat{\mathbf{g}} \times] \, ds \quad (246)$$

$$\Rightarrow \Psi_{64}(\ell, k) = - {}_k^{\ell}\mathbf{C}_{\hat{q}} [R_k \hat{\mathbf{g}} \times] \Delta t_{k, \ell} \quad (247)$$

Ψ_{66} :

$$\dot{\Psi}_{66}(\ell, k) = -[\hat{\omega} \times] \Psi_{66}(\ell, k) \quad | \text{cf. } \Psi_{44} \quad (248)$$

$$\Rightarrow \Psi_{66}(\ell, k) = {}_k^{\ell}\mathbf{C}_{\hat{q}} \quad (249)$$

Ψ_{67} :

$$\dot{\Psi}_{67}(\ell, k) = -[\hat{\omega} \times] \Psi_{47}(\ell, k) - [\hat{\omega} \times] \Psi_{67}(\ell, k) - [\hat{\mathbf{v}}_{I_{\ell}} \times] \quad (250)$$

$$\Rightarrow \dot{\Psi}_{67}(\ell, k) + [\hat{\omega} \times] \Psi_{67}(\ell, k) = [\hat{\omega} \times] {}_k^{\ell}\mathbf{C}_{\hat{q}} \int_{t_k}^{t_{\ell}} {}_k^s \mathbf{C}_{\hat{q}}^{\top} \, ds - [\hat{\mathbf{v}}_{I_{\ell}} \times] \quad | \text{cf. Eq. (233) - (235)} \quad (251)$$

$$\Rightarrow \frac{d}{dt} \left({}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top} \Psi_{67}(\ell, k) \right) = {}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top} [\hat{\omega} \times] {}_k^{\ell}\mathbf{C}_{\hat{q}} \int_{t_k}^{t_{\ell}} {}_k^s \mathbf{C}_{\hat{q}}^{\top} \, ds - {}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top} [\hat{\mathbf{v}}_{I_{\ell}} \times] \quad (252)$$

$$\Rightarrow {}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top} \Psi_{67}(\ell, k) + \Psi_{67}(k, k) = \int_{t_k}^{t_{\ell}} [{}_k^s \mathbf{C}_{\hat{q}}^{\top}] [{}^s \mu \mathbf{C}_{\hat{q}}^{\top}] \, d\mu ds - \int_{t_k}^{t_{\ell}} {}_k^s \mathbf{C}_{\hat{q}}^{\top} [\hat{\mathbf{v}}_{I_s} \times] \, ds \quad (253)$$

$$\Rightarrow \Psi_{67}(\ell, k) = {}_k^{\ell}\mathbf{C}_{\hat{q}} \int_{t_k}^{t_{\ell}} [R_k \hat{\mathbf{g}} \times] \int_{t_k}^s {}_k^{\mu} \mathbf{C}_{\hat{q}}^{\top} \, d\mu ds - {}_k^{\ell}\mathbf{C}_{\hat{q}} \int_{t_k}^{t_{\ell}} {}_k^s \mathbf{C}_{\hat{q}}^{\top} [\hat{\mathbf{v}}_{I_s} \times] \, ds \quad (254)$$

$$\Rightarrow \Psi_{67}(\ell, k) = {}_k^{\ell}\mathbf{C}_{\hat{q}} [R_k \hat{\mathbf{g}} \times] \int_{t_k}^{t_{\ell}} \int_{t_k}^s {}_k^{\mu} \mathbf{C}_{\hat{q}}^{\top} \, d\mu ds - \int_{t_k}^{t_{\ell}} {}_k^s \mathbf{C}_{\hat{q}}^{\top} [\hat{\mathbf{v}}_{I_s} \times] \, ds \quad (255)$$

Ψ_{68} :

$$\dot{\Psi}_{68}(\ell, k) = -[\hat{\omega} \times] \Psi_{68}(\ell, k) - \mathbf{I}_3 \quad | \text{cf. } \Psi_{47} \quad (256)$$

$$\Rightarrow \Psi_{68}(\ell, k) = - {}_k^{\ell}\mathbf{C}_{\hat{q}} \int_{t_k}^{t_{\ell}} {}_k^s \mathbf{C}_{\hat{q}}^{\top} \, ds = - \int_{t_k}^{t_{\ell}} {}_k^s \mathbf{C}_{\hat{q}} \, ds \quad (257)$$

Ψ_{53} :

$$\dot{\Psi}_{53}(\ell, k) = {}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top} \Psi_{63}(\ell, k) \quad (258)$$

$$\Rightarrow \dot{\Psi}_{53}(\ell, k) = - {}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top} {}_k^{\ell}\mathbf{C}_{\hat{q}} \Delta t_{k, \ell} = - \mathbf{I}_3 \Delta t_{k, \ell} \quad (259)$$

$$\Rightarrow \Psi_{53}(\ell, k) + \Psi_{53}(k, k) = - \int_{t_k}^{t_{\ell}} \mathbf{I}_3 \Delta t_{k, s} \, ds \quad (260)$$

$$\Rightarrow \Psi_{53}(\ell, k) = - \frac{1}{2} \mathbf{I}_3 \Delta t_{k, \ell}^2 \quad (261)$$

Ψ_{54} :

$$\dot{\Psi}_{54}(\ell, k) = - {}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top} [\hat{\mathbf{v}}_{I_{\ell}} \times] \Psi_{44}(\ell, k) + {}_k^{\ell}\mathbf{C}_{\hat{q}}^{\top} \Psi_{64}(\ell, k) \quad (262)$$

$$\Rightarrow \dot{\Psi}_{54}(\ell, k) = -\frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top [\hat{\mathbf{v}}_{I_\ell} \times] \frac{\ell}{k} \mathbf{C}_{\hat{q}} - \frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top \frac{\ell}{k} \mathbf{C}_{\hat{q}} [R_k \hat{\mathbf{g}} \times] \Delta t_{k,\ell} \quad (263)$$

$$\Rightarrow \dot{\Psi}_{54}(\ell, k) = -[\frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top \hat{\mathbf{v}}_{I_\ell} \times] - [R_k \hat{\mathbf{g}} \times] \Delta t_{k,\ell} \quad (264)$$

$$\Rightarrow \Psi_{54}(\ell, k) + \Psi_{54}(k, k) = - \int_{t_k}^{t_\ell} [\frac{s}{k} \mathbf{C}_{\hat{q}}^\top \hat{\mathbf{v}}_{I_s} \times] ds - \int_{t_k}^{t_\ell} [R_k \hat{\mathbf{g}} \times] \Delta t_{k,s} ds \quad (265)$$

$$\Rightarrow \Psi_{54}(\ell, k) = - \int_{t_k}^{t_\ell} [R_k \hat{\mathbf{v}}_{I_s} \times] ds - \frac{1}{2} [R_k \hat{\mathbf{g}} \times] \Delta t_{k,\ell}^2 \quad (266)$$

$$\Rightarrow \Psi_{54}(\ell, k) = -[(R_k \hat{\mathbf{p}}_{I_\ell} + \frac{1}{2} R_k \hat{\mathbf{g}} \Delta t_{k,\ell}^2) \times] \quad (267)$$

Ψ_{56} :

$$\dot{\Psi}_{56}(\ell, k) = \frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top \Psi_{66}(\ell, k) \quad (268)$$

$$\Rightarrow \dot{\Psi}_{56}(\ell, k) = \frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top \frac{\ell}{k} \mathbf{C}_{\hat{q}} = \mathbf{I}_3 \quad (269)$$

$$\Rightarrow \Psi_{56}(\ell, k) + \Psi_{56}(k, k) = \int_{t_k}^{t_\ell} \mathbf{I}_3 ds \quad (270)$$

$$\Rightarrow \Psi_{56}(\ell, k) = \mathbf{I}_3 \Delta t_{k,\ell} \quad (271)$$

Ψ_{57} :

$$\dot{\Psi}_{57}(\ell, k) = -\frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top [\hat{\mathbf{v}}_{I_\ell} \times] \Psi_{47}(\ell, k) + \frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top \Psi_{67}(\ell, k) \quad (272)$$

$$\begin{aligned} \Rightarrow \dot{\Psi}_{57}(\ell, k) &= \frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top [\hat{\mathbf{v}}_{I_\ell} \times] \frac{\ell}{k} \mathbf{C}_{\hat{q}} \int_{t_k}^{t_\ell} \frac{s}{k} \mathbf{C}_{\hat{q}}^\top ds + \frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top \frac{\ell}{k} \mathbf{C}_{\hat{q}} [R_k \hat{\mathbf{g}} \times] \int_{t_k}^{t_\ell} \int_{t_k}^s \frac{\mu}{k} \mathbf{C}_{\hat{q}}^\top d\mu ds \\ &\quad - \frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top \frac{\ell}{k} \mathbf{C}_{\hat{q}} \int_{t_k}^{t_\ell} \frac{s}{k} \mathbf{C}_{\hat{q}}^\top [\hat{\mathbf{v}}_{I_s} \times] ds \end{aligned} \quad (273)$$

$$\Rightarrow \dot{\Psi}_{57}(\ell, k) = [\frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top \hat{\mathbf{v}}_{I_\ell} \times] \int_{t_k}^{t_\ell} \frac{s}{k} \mathbf{C}_{\hat{q}}^\top ds + [R_k \hat{\mathbf{g}} \times] \int_{t_k}^{t_\ell} \int_{t_k}^s \frac{\mu}{k} \mathbf{C}_{\hat{q}}^\top d\mu ds - \int_{t_k}^{t_\ell} \frac{s}{k} \mathbf{C}_{\hat{q}}^\top [\hat{\mathbf{v}}_{I_s} \times] ds \quad (274)$$

$$\begin{aligned} \Rightarrow \Psi_{57}(\ell, k) + \Phi_{57}(k, k) &= \int_{t_k}^{t_\ell} [\frac{s}{k} \mathbf{C}_{\hat{q}}^\top \hat{\mathbf{v}}_{I_s} \times] \int_{t_k}^s \frac{\mu}{k} \mathbf{C}_{\hat{q}}^\top d\mu ds + [R_k \hat{\mathbf{g}} \times] \int_{t_k}^{t_\ell} \int_{t_k}^s \int_{t_k}^\mu \lambda \mathbf{C}_{\hat{q}}^\top d\lambda d\mu ds \\ &\quad - \int_{t_k}^{t_\ell} \int_{t_k}^s \frac{\mu}{k} \mathbf{C}_{\hat{q}}^\top [\hat{\mathbf{v}}_{I_\mu} \times] d\mu ds \end{aligned} \quad (275)$$

$$\begin{aligned} \Rightarrow \Psi_{57}(\ell, k) &= \int_{t_k}^{t_\ell} [\frac{s}{k} \mathbf{C}_{\hat{q}}^\top \hat{\mathbf{v}}_{I_s} \times] \int_{t_k}^s \frac{\mu}{k} \mathbf{C}_{\hat{q}}^\top d\mu ds + [R_k \hat{\mathbf{g}} \times] \int_{t_k}^{t_\ell} \int_{t_k}^s \int_{t_k}^\mu \lambda \mathbf{C}_{\hat{q}}^\top d\lambda d\mu ds \\ &\quad - \int_{t_k}^{t_\ell} \int_{t_k}^s \frac{\mu}{k} \mathbf{C}_{\hat{q}}^\top [\hat{\mathbf{v}}_{I_\mu} \times] d\mu ds \end{aligned} \quad (276)$$

Ψ_{58} :

$$\dot{\Psi}_{58}(\ell, k) = \frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top \Psi_{68}(\ell, k) \quad (277)$$

$$\Rightarrow \dot{\Psi}_{58}(\ell, k) = -\frac{\ell}{k} \mathbf{C}_{\hat{q}}^\top \frac{\ell}{k} \mathbf{C}_{\hat{q}} \int_{t_k}^{t_\ell} \frac{s}{k} \mathbf{C}_{\hat{q}}^\top ds = - \int_{t_k}^{t_\ell} \frac{s}{k} \mathbf{C}_{\hat{q}}^\top ds \quad (278)$$

$$\Rightarrow \Psi_{58}(\ell, k) + \Psi_{58}(k, k) = - \int_{t_k}^{t_\ell} \int_{t_k}^s \frac{\mu}{k} \mathbf{C}_{\hat{q}}^\top d\mu ds \quad (279)$$

$$\Rightarrow \Psi_{58}(\ell, k) = - \int_{t_k}^{t_\ell} \int_{t_k}^s \frac{\mu}{k} \mathbf{C}_{\hat{q}}^\top d\mu ds \quad (280)$$

Appendix B: Bundle Adjustment with Euclidean Coordinates

In this section, we show the *nonlinear least-squares* method for computing the position of landmark L_j which has been observed from a set of time-ordered robocentric frames of reference, \mathcal{R}_j . Assume that $\{R_k\}$ is the

first robocentric frame from which the landmark is observed, the landmark position for $i \in \{\mathcal{R}_j\}$:

$${}^{R_k} \mathbf{p}_{L_j} = {}_i^k \mathbf{C}_{\bar{q}} {}^{R_i} \mathbf{p}_{L_j} + {}^{R_k} \mathbf{p}_{R_i} \quad (281)$$

$${}^{R_i} \mathbf{p}_{L_j} = {}_C^I \mathbf{C}_{\bar{q}} {}^{C_i} \mathbf{p}_{L_j} + {}^I \mathbf{p}_C \quad (282)$$

for which we have the estimate of the relative pose, $\{{}_i^k \mathbf{C}_{\bar{q}}, {}^{R_k} \mathbf{p}_{R_i}\}$, and the bearing measurement for ${}^{C_i} \mathbf{p}_{L_j}$:

$$\boldsymbol{\eta}_j^i = \begin{bmatrix} \frac{x_j^i}{z_j^i} \\ \frac{y_j^i}{z_j^i} \\ \frac{v_i}{f} \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{u_i}{f} \\ \frac{v_i}{f} \\ 1 \end{bmatrix}, \quad {}^{C_i} \mathbf{p}_{L_j} = \begin{bmatrix} x_j^i \\ y_j^i \\ z_j^i \end{bmatrix} \quad (283)$$

where (u_i, v_i) is the pixel measurement of L_j centered at the principal point of image, f is the focal length. If we denote the depth of landmark as λ_j^i , then Eq. (281) can be rewritten as:

$${}^{R_k} \mathbf{p}_{L_j} = {}_i^k \mathbf{C}_{\bar{q}} \left({}_C^I \mathbf{C}_{\bar{q}} {}^{C_i} \mathbf{p}_{L_j} + {}^I \mathbf{p}_C \right) + {}^{R_k} \mathbf{p}_{R_i} \quad (284)$$

$$= {}_i^k \mathbf{C}_{\bar{q}} \left({}_C^I \mathbf{C}_{\bar{q}} \lambda_j^i \boldsymbol{\eta}_j^i + {}^I \mathbf{p}_C \right) + {}^{R_k} \mathbf{p}_{R_i} \quad (285)$$

$$= \lambda_j^i {}_i^k \mathbf{C}_{\bar{q}} {}^I \mathbf{C}_{\bar{q}} \boldsymbol{\eta}_j^i + {}_i^k \mathbf{C}_{\bar{q}} {}^I \mathbf{p}_C + {}^{R_k} \mathbf{p}_{R_i} \quad (286)$$

$$= \lambda_j^i \bar{\boldsymbol{\eta}}_j^i + \bar{\mathbf{p}}_i \quad (287)$$

where for brevity we have denoted $\bar{\boldsymbol{\eta}}_j^i = {}_i^k \mathbf{C}_{\bar{q}} {}^I \mathbf{C}_{\bar{q}} \boldsymbol{\eta}_j^i$ and $\bar{\mathbf{p}}_i = {}_i^k \mathbf{C}_{\bar{q}} {}^I \mathbf{p}_C + {}^{R_k} \mathbf{p}_{R_i}$. We can find the perpendicular counterpart of $\bar{\boldsymbol{\eta}}_j^i$ as ${}^\perp \bar{\boldsymbol{\eta}}_j^i$, and premultiply it to the above equation, we have:

$${}^\perp \bar{\boldsymbol{\eta}}_j^i {}^{R_k} \mathbf{p}_{f_j} = {}^\perp \bar{\boldsymbol{\eta}}_j^i {}^\top \bar{\mathbf{p}}_i \quad (288)$$

from which we can form two linearly independent constraints for ${}^{R_k} \mathbf{p}_{L_j}$ (e.g., for $\bar{\boldsymbol{\eta}}_j^i = [\eta_x, \eta_y, \eta_z]^\top$, we have ${}^\perp \bar{\boldsymbol{\eta}}_{j,1}^i = [\eta_y, -\eta_x, 0]^\top$ and ${}^\perp \bar{\boldsymbol{\eta}}_{j,2}^i = [\eta_z, 0, -\eta_x]^\top$):

$$\begin{bmatrix} {}^\perp \bar{\boldsymbol{\eta}}_{j,1}^i {}^\top \\ {}^\perp \bar{\boldsymbol{\eta}}_{j,2}^i {}^\top \end{bmatrix} {}^{R_k} \mathbf{p}_{f_j} = \begin{bmatrix} {}^\perp \bar{\boldsymbol{\eta}}_{j,1}^i {}^\top \bar{\mathbf{p}}_i \\ {}^\perp \bar{\boldsymbol{\eta}}_{j,2}^i {}^\top \bar{\mathbf{p}}_i \end{bmatrix} \quad (289)$$

$$\Rightarrow \mathbf{A}_i {}^{R_k} \mathbf{p}_{L_j} = \mathbf{y}_i \quad (290)$$

If we stack all the resulting equations for $i \in \mathcal{R}_j$, then the following linear least-squares problem is formulated:

$$\mathbf{A}_{2N \times 3} {}^{R_k} \mathbf{p}_{L_j} = \mathbf{y}_{2N \times 1} \quad (291)$$

where the size of \mathcal{R}_j : $N \geq 2$ (including k) for solving the problem, which means the positions of landmarks observed from at least *two* robocentric frames of reference can be solved: ${}^{R_k} \hat{\mathbf{p}}_{L_j} = (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{y}$.

Once obtaining an initial estimate for ${}^{R_k} \mathbf{p}_{L_j}$, we can form a nonlinear least-squares problem using the measurements of L_j to refine the value of ${}^{R_k} \hat{\mathbf{p}}_{L_j}$. If we denote ${}^{R_k} \mathbf{p}_{L_j} = [x, y, z]^\top$, then for $i \in \{\mathcal{R}_j \setminus k\}$:

$${}^{C_i} \mathbf{p}_{L_j} = {}_I^C \mathbf{C}_{\bar{q}} {}^{R_i} \mathbf{p}_{L_j} + {}^C \mathbf{p}_I \quad (292)$$

$$= {}_I^C \mathbf{C}_{\bar{q}} \left({}_i^k \mathbf{C}_{\bar{q}} \left({}^{R_k} \mathbf{p}_{L_j} - {}^{R_k} \mathbf{p}_{R_i} \right) \right) + {}^C \mathbf{p}_I \quad (293)$$

$$:= \begin{bmatrix} h_{i,1}(x, y, z) \\ h_{i,2}(x, y, z) \\ h_{i,3}(x, y, z) \end{bmatrix} \quad (294)$$

However, for seeking the numerical stability of computation, we adopt an equivalent representation of ${}^{R_k} \mathbf{p}_{L_j}$: $[x, y, z]^\top = z[\frac{x}{z}, \frac{y}{z}, 1]^\top$ (assuming depth z is the most variant direction), then:

$${}^{C_i} \mathbf{p}_{L_j} = {}_I^C \mathbf{C}_{\bar{q}} {}^i \mathbf{C}_{\bar{q}} \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - {}^{R_k} \mathbf{p}_{R_i} \right) + {}^C \mathbf{p}_I \quad (295)$$

$$= {}_I^C \mathbf{C}_{\bar{q}k}^i \mathbf{C}_{\bar{q}} \left(z \begin{bmatrix} \frac{x}{z} \\ \frac{y}{z} \\ 1 \end{bmatrix} - {}^{R_k} \mathbf{p}_{R_i} \right) + {}^C \mathbf{p}_I \quad (296)$$

$$= z {}_I^C \mathbf{C}_{\bar{q}k}^i \mathbf{C}_{\bar{q}} \left(\begin{bmatrix} \frac{x}{z} \\ \frac{y}{z} \\ 1 \end{bmatrix} - \frac{1}{z} {}^{R_k} \mathbf{p}_{R_i} \right) + {}^C \mathbf{p}_I \quad (297)$$

$$= z {}_I^C \mathbf{C}_{\bar{q}k}^i \mathbf{C}_{\bar{q}} \left(\begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} - \rho {}^{R_k} \mathbf{p}_{R_i} \right) + {}^C \mathbf{p}_I \quad (298)$$

where $\alpha = \frac{x}{z}$, $\beta = \frac{y}{z}$, and $\rho = \frac{1}{z}$. If we only consider the geometry in the above equation, then it is able to be reformulated with arbitrary scale while still fits for the bearing expression (as the scale will be cancelled in the fraction form), such as multiplying ρ to both sides, which results in the following expression:

$$\rho {}^{C_i} \mathbf{p}_{L_j} = {}_I^C \mathbf{C}_{\bar{q}k}^i \mathbf{C}_{\bar{q}} \left(\begin{bmatrix} \alpha \\ \beta \\ 1 \end{bmatrix} - \rho {}^{R_k} \mathbf{p}_{R_i} \right) + \rho {}^C \mathbf{p}_I \quad (299)$$

$$\Rightarrow {}^{C_i} \bar{\mathbf{p}}_{L_j} = \begin{bmatrix} h_{i,1}(\alpha, \beta, \rho) \\ h_{i,2}(\alpha, \beta, \rho) \\ h_{i,3}(\alpha, \beta, \rho) \end{bmatrix} \quad (300)$$

By substituting Eq. (300) into Eq. (86), we express the measurement model as the function of α , β , and ρ :

$$\mathbf{z}_i = \frac{1}{h_{i,3}(\alpha, \beta, \rho)} \begin{bmatrix} h_{i,1}(\alpha, \beta, \rho) \\ h_{i,2}(\alpha, \beta, \rho) \end{bmatrix} + \mathbf{n}_i \quad (301)$$

$$= \mathbf{h}_i(\alpha, \beta, \rho) + \mathbf{n}_i \quad (302)$$

where $\mathbf{n}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}_i)$ is a zero-mean white Gaussian noise corresponding to the scale of the normalized pixel coordinates (i.e., divided by the focal length f). Given the measurements $\mathbf{z}_i = (\frac{u_i}{f}, \frac{v_i}{f})$, $i \in \{\mathcal{R}_j \setminus k\}$, we can form the following nonlinear least-squares problem for solving $\mathcal{X} = [\alpha, \beta, \rho]^\top$:

$$\mathcal{X}^* = \arg \min_{\mathcal{X}} \sum_{i \in \{\mathcal{R}_j \setminus k\}} \|\mathbf{h}_i(\mathcal{X}) - \mathbf{z}_i\|_{\mathbf{\Lambda}_i} = \arg \min_{\mathcal{X}} \sum_{i \in \{\mathcal{R}_j \setminus k\}} \|\boldsymbol{\varepsilon}_i(\mathcal{X})\|_{\mathbf{\Lambda}_i} \quad (303)$$

where $\|\cdot\|_{\mathbf{\Lambda}}$ denotes the $\mathbf{\Lambda}$ -weighted energy norm, and we define $\boldsymbol{\varepsilon}_i$ as the residual associated to \mathbf{z}_i . This problem can be solved iteratively via Gauss-Newton approximation about the initial estimate of $\hat{\mathcal{X}}$:

$$\delta \mathcal{X}^* = \arg \min_{\delta \mathcal{X}} \sum_{i \in \{\mathcal{R}_j \setminus k\}} \|\boldsymbol{\varepsilon}_i(\hat{\mathcal{X}} + \delta \mathcal{X})\|_{\mathbf{\Lambda}_i} \simeq \arg \min_{\delta \mathcal{X}} \sum_{i \in \{\mathcal{R}_j \setminus k\}} \|\boldsymbol{\varepsilon}_i(\hat{\mathcal{X}}) + \mathbf{H}_i \delta \mathcal{X}\|_{\mathbf{\Lambda}_i} \quad (304)$$

For which the Jacobian of residual $\mathbf{H}_i = \frac{\partial \boldsymbol{\varepsilon}_i(\hat{\mathcal{X}} + \delta \mathcal{X})}{\partial \delta \mathcal{X}}$ evaluated at $\hat{\mathcal{X}}$ can be obtained following the chain rule:

$$\mathbf{H}_i = \frac{\partial \boldsymbol{\varepsilon}_i}{\partial {}^{C_i} \bar{\mathbf{p}}_{L_j}} \frac{\partial {}^{C_i} \bar{\mathbf{p}}_{L_j}}{\partial \mathcal{X}} \frac{\partial \mathcal{X}}{\partial \delta \mathcal{X}} = \frac{\partial \boldsymbol{\varepsilon}_i}{\partial {}^{C_i} \bar{\mathbf{p}}_{L_j}} \frac{\partial {}^{C_i} \bar{\mathbf{p}}_{L_j}}{\partial \mathcal{X}} \quad (305)$$

$$\frac{\partial \boldsymbol{\varepsilon}_i}{\partial {}^{C_i} \bar{\mathbf{p}}_{L_j}} = \frac{1}{\hat{h}_{i,3}} \begin{bmatrix} 1 & 0 & -\frac{\hat{h}_{i,1}}{\hat{h}_{i,3}} \\ 0 & 1 & -\frac{\hat{h}_{i,2}}{\hat{h}_{i,3}} \end{bmatrix}, \quad (306)$$

$$\frac{\partial {}^{C_i} \bar{\mathbf{p}}_{L_j}}{\partial \mathcal{X}} = \begin{bmatrix} \frac{\partial {}^{C_i} \bar{\mathbf{p}}_{L_j}}{\partial [\alpha, \beta]^\top} & \frac{\partial {}^{C_i} \bar{\mathbf{p}}_{L_j}}{\partial \rho} \end{bmatrix} = \begin{bmatrix} {}_I^C \mathbf{C}_{\bar{q}k}^i \mathbf{C}_{\bar{q}} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \\ & {}_I^C \mathbf{C}_{\bar{q}} {}^{R_k} \hat{\mathbf{p}}_{R_k} + {}^C \mathbf{p}_I \end{bmatrix} \quad (307)$$

Once the result of Eq. (304) get converged, we find the optimal parameters: $\mathcal{X}^* = \hat{\mathcal{X}} + \delta \mathcal{X}^*$, thus ${}^{R_k} \mathbf{p}_{L_j}^*$.

Appendix C: Bundle Adjustment with Inverse Depth Parameters

Assuming that we have the same landmark L_j as in Appendix B, the expression for ${}^{R_k}\mathbf{p}_{L_j}$ using our *tailored* inverse depth parametrization (cf. Section 2.3.2) is shown as:

$${}^{R_k}\mathbf{p}_{L_j} = {}_C^I\mathbf{C}_{\bar{q}}{}^{C_k}\mathbf{p}_{L_j} + {}^I\mathbf{p}_C \quad (308)$$

$${}^{C_k}\mathbf{p}_{L_j} = \frac{1}{\rho}\mathbf{e}(\phi, \psi), \quad \mathbf{e} = \begin{bmatrix} \cos \phi \sin \psi \\ \sin \phi \\ \cos \phi \cos \psi \end{bmatrix} \quad (309)$$

where \mathbf{e} is the unit directional vector of the ray from $\{C_k\}$ with ϕ and ψ the elevation and azimuth expressed in $\{C_k\}$, respectively, and ρ denotes the inverse depth of landmark along \mathbf{e} . For $i \in \{\mathcal{R}_j \setminus k\}$, we have:

$${}^{R_i}\mathbf{p}_{L_j} = {}_k^i\mathbf{C}_{\bar{q}}{}^{R_k}\mathbf{p}_{L_j} + {}^{R_i}\mathbf{p}_{R_k} \quad (310)$$

$$= {}_k^i\mathbf{C}_{\bar{q}}\left({}_C^I\mathbf{C}_{\bar{q}}{}^{C_k}\mathbf{p}_{L_j} + {}^I\mathbf{p}_C\right) + {}^{R_i}\mathbf{p}_{R_k} \quad (311)$$

$$= {}_k^i\mathbf{C}_{\bar{q}}\left({}_C^I\mathbf{C}_{\bar{q}}\frac{1}{\rho}\mathbf{e} + {}^I\mathbf{p}_C\right) + {}^{R_i}\mathbf{p}_{R_k} \quad (312)$$

$$= \frac{1}{\rho}{}_k^i\mathbf{C}_{\bar{q}}{}_C^I\mathbf{C}_{\bar{q}}\mathbf{e} + {}_k^i\mathbf{C}_{\bar{q}}{}^I\mathbf{p}_C + {}^{R_i}\mathbf{p}_{R_k} \quad (313)$$

Then, the position of landmark in the corresponding camera frame can be expressed as:

$${}^{C_i}\mathbf{p}_{L_j} = {}_I^C\mathbf{C}_{\bar{q}}{}^{R_i}\mathbf{p}_{L_j} + {}^C\mathbf{p}_I \quad (314)$$

$$= {}_I^C\mathbf{C}_{\bar{q}}\left(\frac{1}{\rho}{}_k^i\mathbf{C}_{\bar{q}}{}_C^I\mathbf{C}_{\bar{q}}\mathbf{e} + {}_k^i\mathbf{C}_{\bar{q}}{}^I\mathbf{p}_C + {}^{R_i}\mathbf{p}_{R_k}\right) + {}^C\mathbf{p}_I \quad (315)$$

$$= \frac{1}{\rho}{}_I^C\mathbf{C}_{\bar{q}}{}_k^i\mathbf{C}_{\bar{q}}{}_C^I\mathbf{C}_{\bar{q}}\mathbf{e} + {}_I^C\mathbf{C}_{\bar{q}}{}_k^i\mathbf{C}_{\bar{q}}{}^I\mathbf{p}_C + {}_I^C\mathbf{C}_{\bar{q}}{}^{R_i}\mathbf{p}_{R_k} + {}^C\mathbf{p}_I \quad (316)$$

$$= \frac{1}{\rho}{}_k^i\bar{\mathbf{C}}_{\bar{q}}\mathbf{e} + {}^i\bar{\mathbf{p}}_k \quad (317)$$

where ${}_k^i\bar{\mathbf{C}}_{\bar{q}} = {}_I^C\mathbf{C}_{\bar{q}}{}_k^i\mathbf{C}_{\bar{q}}{}_C^I\mathbf{C}_{\bar{q}}$, and ${}^i\bar{\mathbf{p}}_k = {}_I^C\mathbf{C}_{\bar{q}}{}_k^i\mathbf{C}_{\bar{q}}{}^I\mathbf{p}_C + {}_I^C\mathbf{C}_{\bar{q}}{}^{R_i}\mathbf{p}_{R_k} + {}^C\mathbf{p}_I$ represent the relative pose between the camera frame $\{C_k\}$ and $\{C_i\}$. If we only consider the geometry in the above equation, then it is able to be reformulated with arbitrary scale, such as being normalized by ρ , which results in the following expression:

$$\rho {}^{C_i}\mathbf{p}_{L_j} = {}_k^i\bar{\mathbf{C}}_{\bar{q}}\mathbf{e} + \rho {}^i\bar{\mathbf{p}}_k \quad (318)$$

$$\Rightarrow {}^i\bar{\mathbf{p}}_{L_j} = \begin{bmatrix} h_{i,1}(\phi, \psi, \rho) \\ h_{i,2}(\phi, \psi, \rho) \\ h_{i,3}(\phi, \psi, \rho) \end{bmatrix} \quad (319)$$

where $h_{i,1}$, $h_{i,2}$, and $h_{i,3}$ are the scalar functions of the entries of ${}^i\bar{\mathbf{p}}_{L_j}$. By substituting Eq. (319) into Eq. (86), the measurement model can be expressed as the function of ϕ , ψ , and ρ :

$$\mathbf{z}_i = \frac{1}{h_{i,3}(\phi, \psi, \rho)} \begin{bmatrix} h_{i,1}(\phi, \psi, \rho) \\ h_{i,2}(\phi, \psi, \rho) \end{bmatrix} + \mathbf{n}_i \quad (320)$$

$$= \mathbf{h}_i(\phi, \psi, \rho) + \mathbf{n}_i \quad (321)$$

where $\mathbf{n}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_i)$ is a zero-mean white Gaussian noise corresponding to the scale of the normalized pixel coordinates (i.e., divided by the focal length f). Given the measurements $\mathbf{z}_i = (\frac{u_i}{f}, \frac{v_i}{f})$, $i \in \{\mathcal{R}_j \setminus k\}$, we can form the following nonlinear least-squares problem for solving $\boldsymbol{\lambda} = [\phi, \psi, \rho]^\top$:

$$\boldsymbol{\lambda}^* = \arg \min_{\boldsymbol{\lambda}} \sum_{i \in \{\mathcal{R}_j \setminus k\}} \|\mathbf{h}_i(\boldsymbol{\lambda}) - \mathbf{z}_i\|_{\boldsymbol{\Lambda}_i} = \arg \min_{\boldsymbol{\lambda}} \sum_{i \in \{\mathcal{R}_j \setminus k\}} \|\boldsymbol{\epsilon}_i(\boldsymbol{\lambda})\|_{\boldsymbol{\Lambda}_i} \quad (322)$$

where $\|\cdot\|_{\Lambda}$ denotes the Λ -weighted energy norm, and we define ϵ_i as the residual associated to \mathbf{z}_i . This problem can be solved iteratively via Gauss-Newton approximation about the initial estimate of $\hat{\lambda}$:

$$\delta\lambda^* = \arg \min_{\delta\lambda} \sum_{i \in \{\mathcal{R}_j \setminus k\}} \|\epsilon_i(\hat{\lambda} + \delta\lambda)\|_{\Lambda_i} \simeq \arg \min_{\delta\lambda} \sum_{i \in \{\mathcal{R}_j \setminus k\}} \|\epsilon_i(\hat{\lambda}) + \mathbf{H}_i \delta\lambda\|_{\Lambda_i} \quad (323)$$

For the initial estimate $\hat{\lambda}$, we can get $[\hat{\phi}, \hat{\psi}]^\top$ using the pixel measurement (u_k, v_k) and the focal length f as:

$$\begin{bmatrix} \hat{\phi} \\ \hat{\psi} \end{bmatrix} = \begin{bmatrix} \arctan(v_k, \sqrt{u_k^2 + f^2}) \\ \arctan(u_k, f) \end{bmatrix} = \begin{bmatrix} \arctan\left(\frac{v_k}{f}, \sqrt{\left(\frac{u_k}{f}\right)^2 + 1}\right) \\ \arctan\left(\frac{u_k}{f}, 1\right) \end{bmatrix} \quad (324)$$

however, the initial value for $\hat{\rho}$ is empirically chosen (e.g., 0.1 [21]), for which we choose 0 to put landmark at infinity first, and let it converge by performing iteration. The Jacobian of residual $\mathbf{H}_i = \frac{\partial \epsilon_i(\hat{\lambda} + \delta\lambda)}{\partial \delta\lambda}$ evaluated at $\hat{\lambda}$ can be obtained following the chain rule:

$$\mathbf{H}_i = \frac{\partial \epsilon_i}{\partial^i \bar{\mathbf{p}}_{L_j}} \frac{\partial^i \bar{\mathbf{p}}_{L_j}}{\partial \lambda} \frac{\partial \lambda}{\partial \delta\lambda} = \frac{\partial \epsilon_i}{\partial^i \bar{\mathbf{p}}_{L_j}} \frac{\partial^i \bar{\mathbf{p}}_{L_j}}{\partial \lambda} \quad (325)$$

$$\frac{\partial \epsilon_i}{\partial^i \bar{\mathbf{p}}_{L_j}} = \frac{1}{\hat{h}_{i,3}} \begin{bmatrix} 1 & 0 & -\frac{\hat{h}_{i,1}}{\hat{h}_{i,3}} \\ 0 & 1 & -\frac{\hat{h}_{i,2}}{\hat{h}_{i,3}} \end{bmatrix}, \quad (326)$$

$$\frac{\partial^i \bar{\mathbf{p}}_{L_j}}{\partial \lambda} = \begin{bmatrix} \frac{\partial^i \bar{\mathbf{p}}_{L_j}}{\partial [\phi, \psi]^\top} & \frac{\partial^i \bar{\mathbf{p}}_{L_j}}{\partial \rho} \end{bmatrix} = \begin{bmatrix} {}_k^i \bar{\mathbf{C}}_{\tilde{q}} & \begin{bmatrix} -\sin \hat{\phi} \sin \hat{\psi} & \cos \hat{\phi} \cos \hat{\psi} \\ \cos \hat{\phi} & 0 \\ -\sin \hat{\phi} \cos \hat{\psi} & -\cos \hat{\phi} \sin \hat{\psi} \end{bmatrix} & {}_i^i \hat{\mathbf{p}}_k \end{bmatrix} \quad (327)$$

Once the result of Eq. (323) get converged, we find the optimal parameters: $\lambda^* = \hat{\lambda} + \delta\lambda^*$, thus ${}^{R_k} \mathbf{p}_{L_j}^*$.

References

- [1] Anastasios I Mourikis and Stergios I Roumeliotis. “A multi-state constraint Kalman filter for vision-aided inertial navigation”. In: *IEEE Intl. Conf. Robot. Autom. (ICRA)*. 2007, pp. 3565–3572.
- [2] Eagle S Jones and Stefano Soatto. “Visual-inertial navigation, mapping and localization: A scalable real-time causal approach”. In: *Intl. J. Robot. Res. (IJRR)* 30.4 (2011), pp. 407–430.
- [3] Jonathan Kelly and Gaurav S Sukhatme. “Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration”. In: *Intl. J. Robot. Res. (IJRR)* 30.1 (2011), pp. 56–79.
- [4] Mingyang Li and Anastasios I Mourikis. “High-precision, consistent EKF-based visual-inertial odometry”. In: *Intl. J. Robot. Res. (IJRR)* 32.6 (2013), pp. 690–711.
- [5] Stefan Leutenegger et al. “Keyframe-based visual–inertial odometry using nonlinear optimization”. In: *Intl. J. Robot. Res. (IJRR)* 34.3 (2015), pp. 314–334.
- [6] Shaojie Shen, Nathan Michael, and Vijay Kumar. “Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft MAVs”. In: *IEEE Intl. Conf. Robot. Autom. (ICRA)*. 2015, pp. 5303–5310.
- [7] Vladyslav Usenko et al. “Direct visual-inertial odometry with stereo cameras”. In: *IEEE Intl. Conf. Robot. Autom. (ICRA)*. 2016, pp. 1885–1892.
- [8] Raúl Mur-Artal and Juan D Tardós. “Visual-inertial monocular SLAM with map reuse”. In: *IEEE Robotics and Automation Letters* 2.2 (2017), pp. 796–803.
- [9] Guoquan P Huang, Anastasios I Mourikis, and Stergios I Roumeliotis. “Observability-based rules for designing consistent EKF SLAM estimators”. In: *Intl. J. of Robot. Res. (IJRR)* 29.5 (2010), pp. 502–528.

- [10] Agostino Martinelli. “State estimation based on the concept of continuous symmetry and observability analysis: The case of calibration”. In: *IEEE Trans. Robot. (TRO)* 27.2 (2011), pp. 239–255.
- [11] Joel A Hesch et al. “Consistency analysis and improvement of vision-aided inertial navigation”. In: *IEEE Trans. Robot. (TRO)* 30.1 (2014), pp. 158–176.
- [12] Guoquan Huang, Michael Kaess, and John J Leonard. “Towards consistent visual-inertial navigation”. In: *IEEE Intl. Conf. Robot. Autom. (ICRA)*. 2014, pp. 4926–4933.
- [13] Teng Zhang et al. “Convergence and consistency analysis for a 3-D Invariant-EKF SLAM”. In: *IEEE Robotics and Automation Letters* 2.2 (2017), pp. 733–740.
- [14] José A Castellanos, José Neira, and Juan D Tardós. “Limits to the consistency of EKF-based SLAM”. In: *IFAC Proceedings Volumes* 37.8 (2004), pp. 716–721.
- [15] Javier Civera et al. “1-point RANSAC for EKF-based structure from motion”. In: *IEEE/RSJ Intl. Conf. Intell. Robot. Syst. (IROS)*. IEEE. 2009, pp. 3498–3504.
- [16] Michael Bloesch et al. “Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback”. In: *Intl. J. of Robot. Res. (IJRR)* 36.10 (2017), pp. 1053–1072.
- [17] W. G. Breckenridge. “Quaternions proposed standard conventions”. In: *NASA Jet Propulsion Laboratory, Tech. Rep.* (1979).
- [18] Nikolas Trawny and Stergios I Roumeliotis. “Indirect Kalman filter for 3D attitude estimation”. In: *UMN Department of Computer Science & Engineering, Tech. Rep.* (2005).
- [19] Kevin Eckenhoff, Patrick Geneva, and Guoquan Huang. “High-Accuracy Preintegration for Visual-Inertial Navigation”. In: *Intl. Workshop on the Algorithmic Foundations of Robotics*. 2016.
- [20] Gene H Golub and Charles F Van Loan. *Matrix Computations*. Vol. 3. JHU Press, 2012.
- [21] Javier Civera, Andrew J Davison, and JM Martinez Montiel. “Inverse depth parametrization for monocular SLAM”. In: *IEEE Trans. Robot. (TRO)* 24.5 (2008), pp. 932–945.
- [22] Zhe Chen, Ke Jiang, and James C Hung. “Local observability matrix and its application to observability analyses”. In: *the 16th Annual Conf. of IEEE Industrial Electronic Society*. 1990, pp. 100–103.
- [23] Jianbo Shi et al. “Good features to track”. In: *IEEE Conf. Comp. Vis. Patt. Reco. (CVPR)*. 1994, pp. 593–600.
- [24] Simon Baker and Iain Matthews. “Lucas-kanade 20 years on: A unifying framework”. In: *Intl. J. Comp. Vis. (IJCV)* 56.3 (2004), pp. 221–255.
- [25] Chiara Troiani et al. “2-point-based outlier rejection for camera-imu systems with applications to micro aerial vehicles”. In: *IEEE Intl. Conf. Robot. Autom. (ICRA)*. 2014, pp. 5530–5536.
- [26] Michael Burri et al. “The EuRoC micro aerial vehicle datasets”. In: *Intl. J. Robot. Res. (IJRR)* 35.10 (2016), pp. 1157–1163.
- [27] Anastasios I Mourikis et al. “Vision-aided inertial navigation for spacecraft entry, descent, and landing”. In: *IEEE Trans. Robot. (TRO)* 25.2 (2009), pp. 264–280.