Towards Scale-Aware, Robust, and Generalizable Unsupervised Monocular Depth Estimation by Integrating IMU Motion Dynamics (Supplementary Material)

Sen Zhang, Jing Zhang, and Dacheng Tao

The University of Sydney, Sydney, Australia
szha2609@uni.sydney.edu.au
{jing.zhang1,dacheng.tao}@sydney.edu.au

Abstract. This document provides supplementary information on (1) the derivation of camera-centric IMU preintegration, (2) the derivation of camera-centric EKF propagation, (3) the derivation of camera-centric EKF update, and (4) more qualitative generalization results on Make3D.

S1 Derivation of Camera-Centric IMU Preintegration

Let $\{\boldsymbol{p}_{wb_t}, \boldsymbol{q}_{wb_t}\}$ and $\boldsymbol{v}_t^w$ denote the translation and rotation from the body frame to the world frame, and the velocity expressed in the world frame at time t, where $\boldsymbol{q}_{wb_t}$ is the corresponding quaternion of $\boldsymbol{R}_{wb_t}$. The first-order derivatives of $\{\boldsymbol{p}, \boldsymbol{v}, \boldsymbol{q}\}$ read: $\dot{\boldsymbol{p}_{wb_t}} = \boldsymbol{v}_t^w$, $\dot{\boldsymbol{v}_t^w} = \boldsymbol{a}_t^w$, and $\dot{\boldsymbol{q}_{wb_t}} = \boldsymbol{q}_{wb_t} \otimes [0, \frac{1}{2}\boldsymbol{w}^{b_t}]^T$. Then the continuous IMU motion dynamics from time i to j is given by:

$$\boldsymbol{p}_{wb_j} = \boldsymbol{p}_{wb_i} + \boldsymbol{v}_i^w \Delta t + \int\int_{t\in[i,j]} (\boldsymbol{R}_{wb_t}\boldsymbol{a}^{b_t} - \boldsymbol{g}^w)\mathrm{d}t^2, \tag{1}$$

$$\boldsymbol{v}_j^w = \boldsymbol{v}_i^w + \int_{t\in[i,j]} (\boldsymbol{R}_{wb_t}\boldsymbol{a}^{b_t} - \boldsymbol{g}^w)\mathrm{d}t, \tag{2}$$

$$\boldsymbol{q}_{wb_j} = \int_{t\in[i,j]} \boldsymbol{q}_{wb_t} \otimes [0, \frac{1}{2}\boldsymbol{w}^{b_t}]^T \mathrm{d}t, \tag{3}$$

where Δt is the time gap between i and j, and $\otimes$ denotes quaternion multiplication. By leveraging the multiplicative property of rotation, i.e., $\boldsymbol{q}_{wb_t} = \boldsymbol{q}_{wb_i} \otimes \boldsymbol{q}_{b_ib_t}$, we have:

$$\boldsymbol{p}_{wb_j} = \boldsymbol{p}_{wb_i} + \boldsymbol{v}_i^w \Delta t - \frac{1}{2}\boldsymbol{g}^w \Delta t^2 + \boldsymbol{R}_{wb_i}\boldsymbol{\alpha}_{b_ib_j}, \tag{4}$$

$$\boldsymbol{v}_j^w = \boldsymbol{v}_i^w - \boldsymbol{g}^w \Delta t + \boldsymbol{R}_{wb_i}\boldsymbol{\beta}_{b_ib_j}, \tag{5}$$

$$\boldsymbol{q}_{wb_j} = \boldsymbol{q}_{wb_i} \otimes \boldsymbol{q}_{b_ib_j}, \tag{6}$$

where the three integration terms that can be pre-computed read:

$$\boldsymbol{\alpha}_{b_ib_j} = \int\int_{t\in[i,j]} (\boldsymbol{R}_{b_ib_t}\boldsymbol{a}^{b_t})\mathrm{d}t^2, \tag{7}$$

$$\boldsymbol{\beta}_{b_ib_j} = \int_{t\in[i,j]} (\boldsymbol{R}_{b_ib_t}\boldsymbol{a}^{b_t})\mathrm{d}t, \tag{8}$$

$$\boldsymbol{q}_{b_ib_j} = \int_{t\in[i,j]} \boldsymbol{q}_{b_ib_t} \otimes [0, \frac{1}{2}\boldsymbol{w}^{b_t}]^T\mathrm{d}t, \tag{9}$$

Given the extrinsics $\{\boldsymbol{R}_{cb}, \boldsymbol{p}_{cb}\}$ and $\{\boldsymbol{R}_{bc}, \boldsymbol{p}_{bc}\}$ between the IMU and the camera frames, based on Eq. (9), we can first derive the camera-centric IMU preintegrated rotation $\check{R}_{c_kc_{k+1}}$ as:

$$\boldsymbol{R}_{c_k\check{c}_{k+1}} = \boldsymbol{R}_{cb}\mathcal{F}^{-1}(\boldsymbol{q}_{b_kb_{k+1}})\boldsymbol{R}_{bc}, \tag{10}$$

where $\mathcal{F}$ denotes the transformation from rotation matrix to quaternion. Then by rearranging Eq. (4), we have:

$$\begin{aligned}
\boldsymbol{\alpha}_{b_kb_{k+1}} &= \boldsymbol{R}_{b_kw}(\boldsymbol{p}_{wb_{k+1}} - \boldsymbol{p}_{wb_k}) - \boldsymbol{R}_{b_kw}\boldsymbol{v}_i^w\Delta t + \frac{1}{2}\boldsymbol{R}_{b_kw}\boldsymbol{g}^w\Delta t^2 && (11)\\
&= \boldsymbol{p}_{b_kb_{k+1}} - \boldsymbol{v}_i^{b_k}\Delta t + \frac{1}{2}\boldsymbol{g}^{b_k}\Delta t^2 && (12)\\
&= \boldsymbol{R}_{b_kc_k}(\boldsymbol{p}_{c_kb_{k+1}} - \boldsymbol{p}_{c_kb_k}) - \boldsymbol{v}_i^{b_k}\Delta t + \frac{1}{2}\boldsymbol{g}^{b_k}\Delta t^2. && (13)
\end{aligned}$$

By left-multiplying $\boldsymbol{R}_{cb}$ to both sides of Eq. (13), we have:

$$\boldsymbol{R}_{cb}\boldsymbol{\alpha}_{b_kb_{k+1}} = \boldsymbol{p}_{c_kb_{k+1}} - \boldsymbol{p}_{cb} - \boldsymbol{v}_i^{c_k}\Delta t + \frac{1}{2}\boldsymbol{g}^{c_k}\Delta t^2. \tag{14}$$

Then we consider the following two equations w.r.t. translation:

$$\begin{aligned}
\boldsymbol{p}_{cb} &= -\boldsymbol{R}_{cb}\boldsymbol{p}_{bc}, && (15)\\
\boldsymbol{p}_{c_kb_{k+1}} &= \boldsymbol{p}_{c_kc_{k+1}} - \boldsymbol{R}_{c_kb_{k+1}}\boldsymbol{p}_{b_{k+1}c_{k+1}} && (16)\\
&= \boldsymbol{p}_{c_kc_{k+1}} - \boldsymbol{R}_{c_kc_{k+1}}\boldsymbol{R}_{c_{k+1}b_{k+1}}\boldsymbol{p}_{b_{k+1}c_{k+1}} && (17)\\
&= \boldsymbol{p}_{c_kc_{k+1}} - \boldsymbol{R}_{c_kc_{k+1}}\boldsymbol{R}_{cb}\boldsymbol{p}_{bc}. && (18)
\end{aligned}$$

By inserting Eq. (15-18) into Eq. (14) and rearranging the resulting formula, we obtain the camera-centric IMU preintegrated translation:

$$\boldsymbol{p}_{c_k\check{c}_{k+1}} = \boldsymbol{R}_{cb}\boldsymbol{\alpha}_{b_kb_{k+1}} + \boldsymbol{R}_{c_kc_{k+1}}\boldsymbol{R}_{cb}\boldsymbol{p}_{bc} - \boldsymbol{R}_{cb}\boldsymbol{p}_{bc} + \boldsymbol{v}^{c_k}\Delta t - \frac{1}{2}\boldsymbol{g}^{c_k}\Delta t^2. \tag{19}$$

S2 Derivation of Camera-Centric EKF Propagation

Let c_k denote the camera frame at time t_k, and $\{b_t\}$ denote the IMU frames between t_k and time t_{k+1} when we receive the next visual measurement. We then propagate the IMU information according to the state transition model: $\boldsymbol{x}_t = f(\boldsymbol{x}_{t-1}, \boldsymbol{u}_t) + \boldsymbol{w}_t$, where $\boldsymbol{u}_t$ is the IMU record at time t, $\boldsymbol{w}_t$ is the noise term, and $\boldsymbol{x}_t = [\boldsymbol{\phi}_{c_k b_t}^T, \boldsymbol{p}_{c_k b_t T}, \boldsymbol{v}^{c_k T}, \boldsymbol{g}^{c_k T}, \boldsymbol{b}_w^{b_t T}, \boldsymbol{b}_a^{b_t T}]^T$ is the state vector expressed in the camera frame c_k except for $\{\boldsymbol{b}_w, \boldsymbol{b}_a\}$. $\boldsymbol{\phi}_{c_k b_t}$ denotes the so(3) Lie algebra of the rotation matrix $\boldsymbol{R}_{c_k b_t}$ s.t. $\boldsymbol{R}_{c_k b_t} = exp([\boldsymbol{\phi}_{c_k b_t}]^\wedge)$, where $[\cdot]^\wedge$ denotes the operation from a so(3) vector to the corresponding skew symmetric matrix. To facilitate the derivation of the propagation process, we further separate the state into the nominal states denoted by $\bar{(\cdot)}$, and the error states $\delta \boldsymbol{x}_{b_t} = [\delta \boldsymbol{\phi}_{c_k b_t}^T, \delta \boldsymbol{p}_{c_k b_t}^T, \delta \boldsymbol{v}^{c_k T}, \delta \boldsymbol{g}^{c_k T}, \delta \boldsymbol{b}_w^{b_t T}, \delta \boldsymbol{b}_a^{b_t T}]^T$, such that:

$$\boldsymbol{R}_{c_k b_t} = \bar{\boldsymbol{R}}_{c_k b_t} exp([\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge), \quad \boldsymbol{p}_{c_k b_t} = \bar{\boldsymbol{p}}_{c_k b_t} + \delta \boldsymbol{p}_{c_k b_t}, \tag{20}$$

$$\boldsymbol{v}^{c_k} = \bar{\boldsymbol{v}}^{c_k} + \delta \boldsymbol{v}^{c_k}, \quad \boldsymbol{g}^{c_k} = \bar{\boldsymbol{g}}^{c_k} + \delta \boldsymbol{g}^{c_k}, \tag{21}$$

$$\boldsymbol{b}_w^{b_t} = \bar{\boldsymbol{b}_w}^{b_t} + \delta \boldsymbol{b}_w^{b_t}, \quad \boldsymbol{b}_a^{b_t} = \bar{\boldsymbol{b}_a}^{b_t} + \delta \boldsymbol{b}_a^{b_t}. \tag{22}$$

The nominal states can be computed using the preintegration terms, while the error states are used for propagating the covariances. It is noteworthy that the state transition model of $\delta \boldsymbol{x}_{b_t}$ is non-linear, which prevents a naive use of the Kalman filter. EKF addresses this problem and performs propagation by linearizing the state transition model at each time step using the first-order Taylor approximation. Therefore, let $\dot{(\cdot)}$ denote the derivative w.r.t. time t, we derive the continuous-time propagation model for the error states as:

$$\delta \dot{\boldsymbol{x}}_{b_t} = \boldsymbol{F} \delta \boldsymbol{x}_{b_t} + \boldsymbol{G} \boldsymbol{n}, \tag{23}$$

where $\boldsymbol{n} = \{\boldsymbol{n}_w^T, \boldsymbol{n}_{bw}^T, \boldsymbol{n}_a^T, \boldsymbol{n}_{ba}^T\}$. $\boldsymbol{n}_w$ and $\boldsymbol{n}_a$ denote the white Gaussian noise in the commonly-used IMU noise model, and $\boldsymbol{n}_{bw}$ and $\boldsymbol{n}_{ba}$ denote the Gaussian steps for the white Gaussian random walks $\boldsymbol{b}_w^{b_t}$ and $\boldsymbol{b}_a^{b_t}$, respectively. The derivations of $\boldsymbol{F}$ and $\boldsymbol{G}$ are given as following.

We first consider $\delta \dot{\boldsymbol{g}}^{c_k}$. Since $\delta \boldsymbol{g}^{c_k}$ is a constant w.r.t. time t, we have:

$$\delta \dot{\boldsymbol{g}}^{c_k} = 0. \tag{24}$$

And by the definition of the Gaussian random walks $\{\boldsymbol{b}_w^{b_t}, \boldsymbol{b}_w^{b_t}\}$, we have:

$$\delta \dot{\boldsymbol{b}}_w^{b_t} = \boldsymbol{n}_{bw}, \tag{25}$$

$$\delta \dot{\boldsymbol{b}}_a^{b_t} = \boldsymbol{n}_{ba}, \tag{26}$$

We then come to $\delta \dot{\phi}_{c_k b_t}$. Since $\delta \boldsymbol{\phi}_{c_k b_t}$ presents a small amount increment, by using Eq. (20) and first-order Taylor expansion, we have:

$$\begin{aligned} \boldsymbol{R}_{c_k b_t} &= \bar{\boldsymbol{R}}_{c_k b_t} exp([\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge) && (27) \\ &\approx \bar{\boldsymbol{R}}_{c_k b_t} (\boldsymbol{I} + [\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge). && (28) \end{aligned}$$

Then by using the derivative of $\boldsymbol{R}_{c_k b_t}$ w.r.t. time t, i.e., $\dot{\boldsymbol{R}}_{c_k b_t} = \boldsymbol{R}_{c_k b_t} [\boldsymbol{w}^{b_t}]^\wedge$, we can take the derivative of both sides of Eq. (28), leading to:

$$\boldsymbol{R}_{c_k b_t} [\boldsymbol{w}^{b_t}]^\wedge \approx \bar{\boldsymbol{R}}_{c_k b_t} [\bar{\boldsymbol{w}}^{b_t}]^\wedge (\boldsymbol{I} + [\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge) + \bar{\boldsymbol{R}}_{c_k b_t} \delta \dot{\boldsymbol{\phi}}_{c_k b_t}, \tag{29}$$

where $\bar{\boldsymbol{w}}^{b_t}$ denotes the nominal angular velocity expressed in the IMU body frame at time t. By inserting Eq. (28) into Eq. (29), we have:

$$\bar{\boldsymbol{R}}_{c_k b_t}(\boldsymbol{I} + [\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge)[\boldsymbol{w}^{b_t}]^\wedge \approx \bar{\boldsymbol{R}}_{c_k b_t}[\bar{\boldsymbol{w}}^{b_t}]^\wedge(\boldsymbol{I} + [\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge) + \bar{\boldsymbol{R}}_{c_k b_t}[\delta\dot{\boldsymbol{\phi}}_{c_k b_t}]^\wedge. \tag{30}$$

By cancelling $\bar{\boldsymbol{R}}_{c_k b_t}$ in Eq. (30) and rearranging the formula, we have:

$$\begin{aligned}
[\delta\dot{\boldsymbol{\phi}}_{c_k b_t}]^\wedge &\approx (\boldsymbol{I} + [\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge)[\boldsymbol{w}^{b_t}]^\wedge - [\bar{\boldsymbol{w}}^{b_t}]^\wedge(\boldsymbol{I} + [\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge) && (31)\\
&= (\boldsymbol{I} + [\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge)[\bar{\boldsymbol{w}}^{b_t} + \delta\boldsymbol{w}^{b_t}]^\wedge - [\bar{\boldsymbol{w}}^{b_t}]^\wedge(\boldsymbol{I} + [\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge) && (32)\\
&= (\boldsymbol{I} + [\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge)([\bar{\boldsymbol{w}}^{b_t}]^\wedge + [\delta\boldsymbol{w}^{b_t}]^\wedge) - [\bar{\boldsymbol{w}}^{b_t}]^\wedge(\boldsymbol{I} + [\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge). && (33)
\end{aligned}$$

By rearranging Eq. (33) and using the equation $[\boldsymbol{u}^\wedge\boldsymbol{v}]^\wedge = \boldsymbol{u}^\wedge\boldsymbol{v}^\wedge - \boldsymbol{v}^\wedge\boldsymbol{u}^\wedge$:

$$\begin{aligned}
[\delta\dot{\boldsymbol{\phi}}_{c_k b_t}]^\wedge &\approx [\delta\boldsymbol{w}^{b_t}]^\wedge + [\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge[\delta\boldsymbol{w}^{b_t}]^\wedge + [\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge[\bar{\boldsymbol{w}}^{b_t}]^\wedge - [\bar{\boldsymbol{w}}^{b_t}]^\wedge[\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge && (34)\\
&\approx [\delta\boldsymbol{w}^{b_t}]^\wedge + [\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge[\delta\boldsymbol{w}^{b_t}]^\wedge + [[\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge\bar{\boldsymbol{w}}^{b_t}]^\wedge. && (35)
\end{aligned}$$

By neglecting the high-order small term $[\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge[\delta\boldsymbol{w}^{b_t}]^\wedge$, and using the equation $\boldsymbol{u}^\wedge\boldsymbol{v} = -\boldsymbol{v}^\wedge\boldsymbol{u}$, we have:

$$\begin{aligned}
[\delta\dot{\boldsymbol{\phi}}_{c_k b_t}]^\wedge &\approx [\delta\boldsymbol{w}^{b_t}]^\wedge + [[\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge\bar{\boldsymbol{w}}^{b_t}]^\wedge && (36)\\
&= [\delta\boldsymbol{w}^{b_t} + [\delta\boldsymbol{\phi}_{c_k b_t}]^\wedge\bar{\boldsymbol{w}}^{b_t}]^\wedge. && (37)\\
\delta\dot{\boldsymbol{\phi}}_{c_k b_t} &\approx \delta\boldsymbol{\phi}_{c_k b_t}]^\wedge\bar{\boldsymbol{w}}^{b_t} && (38)\\
&= -[\bar{\boldsymbol{w}}^{b_t}]^\wedge\delta\boldsymbol{\phi}_{c_k b_t} + \delta\boldsymbol{w}^{b_t} && (39)
\end{aligned}$$

We then derive $\bar{\boldsymbol{w}}^{b_t}$ and $\delta\boldsymbol{w}^{b_t}$ to complete Eq. (39) for $\delta\dot{\boldsymbol{\phi}}_{c_k b_t}$. Recall that we have the following noise model for the gyroscope measurement:

$$\boldsymbol{w}_m^{b_t} = \boldsymbol{w}^{b_t} + \boldsymbol{b}_w^{b_t} + \boldsymbol{n}_w, \quad \boldsymbol{n}_w \sim N(0, \sigma_w^2\boldsymbol{I}). \tag{40}$$

By inserting Eq. (22) in to Eq. (40) and rearranging the formula:

$$\boldsymbol{w}^{b_t} = \boldsymbol{w}_m^{b_t} - \bar{\boldsymbol{b}}_w^{b_t} - \delta\boldsymbol{b}_w^{b_t} - \boldsymbol{n}_w. \tag{41}$$

By separating the nominal and stochastic terms in Eq. (41), we have:

$$\begin{aligned}
\bar{\boldsymbol{w}}^{b_t} &= \boldsymbol{w}_m^{b_t} - \bar{\boldsymbol{b}}_w^{b_t}, && (42)\\
\delta\boldsymbol{w}^{b_t} &= -\delta\boldsymbol{b}_w^{b_t} - \boldsymbol{n}_w., && (43)
\end{aligned}$$

which complete the derivation of $\delta\dot{\boldsymbol{\phi}}_{c_k b_t}$ in Eq. (39) w.r.t. $\delta\boldsymbol{x}_{b_t}$ and $\boldsymbol{n}$.

We next derive $\delta\dot{\boldsymbol{p}}_{c_k b_t}$. Taking the derivative w.r.t. both sides of Eq. (20), i.e., $\boldsymbol{p}_{c_k b_t} = \bar{\boldsymbol{p}}_{c_k b_t} + \delta\boldsymbol{p}_{c_k b_t}$, and rearranging the resulting equation leads to:

$$\begin{aligned}
\delta\dot{\boldsymbol{p}}_{c_k b_t} &= \dot{\boldsymbol{p}}_{c_k b_t} - \dot{\bar{\boldsymbol{p}}}_{c_k b_t} && (44)\\
&= \boldsymbol{v}_t^{c_k} - \boldsymbol{v}_t^{\bar{c}_k}. && (45)
\end{aligned}$$

By approximating $\boldsymbol{v}_t^{c_k}$ and $\boldsymbol{v}_t^{\bar{c}_k}$ by $\boldsymbol{v}^{c_k}$ and $\boldsymbol{v}^{\bar{c}_k}$, and inserting Eq. (21) into the approximated Eq. (45), we have:

$$\begin{aligned} \delta \dot{\boldsymbol{p}}_{c_k b_t} &\approx \boldsymbol{v}_t^{\bar{c}_k} + \delta \boldsymbol{v}^{c_k} - \boldsymbol{v}_t^{\bar{c}_k} && (46) \\ &= \delta \boldsymbol{v}^{c_k}. && (47) \end{aligned}$$

Finally, we give the derivation of $\delta \dot{\boldsymbol{v}}^{c_k}$ as following. We first take the derivative to both sides of Eq. (21) and rearrange the formula, leading to:

$$\begin{aligned} \delta \dot{\boldsymbol{v}}^{c_k} &= \dot{\boldsymbol{v}}^{c_k} - \dot{\boldsymbol{v}}^{\bar{c}_k} && (48) \\ &= \boldsymbol{a}^{c_k} - \bar{\boldsymbol{a}}^{c_k}. && (49) \end{aligned}$$

$\boldsymbol{a}^{c_k}$ and $\bar{\boldsymbol{a}}^{c_k}$ can be derived as:

$$\begin{aligned} \boldsymbol{a}^{c_k} &= \boldsymbol{R}_{c_k b_t} \boldsymbol{a}^{b_t} && (50) \\ &= \bar{\boldsymbol{R}}_{c_k b_t} exp([\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge)(\bar{\boldsymbol{a}}^{b_t} + \delta \boldsymbol{a}^{b_t}) && (51) \\ &\approx \bar{\boldsymbol{R}}_{c_k b_t} (\boldsymbol{I} + [\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge)(\bar{\boldsymbol{a}}^{b_t} + \delta \boldsymbol{a}^{b_t}), && (52) \\ \bar{\boldsymbol{a}}^{c_k} &= \bar{\boldsymbol{R}}_{c_k b_t} \bar{\boldsymbol{a}}^{b_t}. && (53) \end{aligned}$$

By inserting Eq. (52-53) to Eq. (49), we have:

$$\begin{aligned} \delta \dot{\boldsymbol{v}}^{c_k} \approx\ & \bar{\boldsymbol{R}}_{c_k b_t} \bar{\boldsymbol{a}}^{b_t} + \bar{\boldsymbol{R}}_{c_k b_t} \delta \boldsymbol{a}^{b_t} + \bar{\boldsymbol{R}}_{c_k b_t} [\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge \bar{\boldsymbol{a}}^{b_t} \\ & + \bar{\boldsymbol{R}}_{c_k b_t} [\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge \delta \boldsymbol{a}^{b_t} - \bar{\boldsymbol{R}}_{c_k b_t} \bar{\boldsymbol{a}}^{b_t} && (54) \\ =\ & \bar{\boldsymbol{R}}_{c_k b_t} \delta \boldsymbol{a}^{b_t} + \bar{\boldsymbol{R}}_{c_k b_t} [\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge \bar{\boldsymbol{a}}^{b_t} + \bar{\boldsymbol{R}}_{c_k b_t} [\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge \delta \boldsymbol{a}^{b_t}. && (55) \end{aligned}$$

By neglecting the high-order small term $\bar{\boldsymbol{R}}_{c_k b_t} [\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge \delta \boldsymbol{a}^{b_t}$ in Eq. (55) and using the equation $\boldsymbol{u}^\wedge \boldsymbol{v} = -\boldsymbol{v}^\wedge \boldsymbol{u}$, we have:

$$\delta \dot{\boldsymbol{v}}^{c_k} \approx \bar{\boldsymbol{R}}_{c_k b_t} \delta \boldsymbol{a}^{b_t} - \bar{\boldsymbol{R}}_{c_k b_t} [\bar{\boldsymbol{a}}^{b_t}]^\wedge \delta \boldsymbol{\phi}_{c_k b_t}. \quad (56)$$

We then derive $\bar{\boldsymbol{a}}^{b_t}$ and $\delta \boldsymbol{a}^{b_t}$ to complete Eq. (56). Recall that we have the following noise model for the accelerometer measurement:

$$\boldsymbol{a}_m^{b_t} = \boldsymbol{a}^{b_t} + \boldsymbol{R}_{b_t c_k} \boldsymbol{g}^{c_k} + \boldsymbol{b}_a^{b_t} + \boldsymbol{n}_a, \quad \boldsymbol{n}_a \sim N(0, \sigma_w^2 \boldsymbol{I}). \quad (57)$$

By inserting Eq. (20-22) to Eq. (57) and using $\boldsymbol{R}^T = \boldsymbol{R}^{-1}$, we have:

$$\begin{aligned} \boldsymbol{a}_m^{b_t} =\ & \boldsymbol{a}^{b_t} + [\bar{\boldsymbol{R}}_{c_k b_t} exp([\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge)]^T (\bar{\boldsymbol{g}}^{c_k} + \delta \boldsymbol{g}^{c_k}) \\ & + \bar{\boldsymbol{b}}_a^{b_t} + \delta \boldsymbol{b}_a^{b_t} + \boldsymbol{n}_a. && (58) \end{aligned}$$

We rearrange the second term in Eq. (58) as below:

$$\begin{aligned} & [\bar{\boldsymbol{R}}_{c_k b_t} exp([\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge)]^T (\bar{\boldsymbol{g}}^{c_k} + \delta \boldsymbol{g}^{c_k}) && (59) \\ \approx & [\bar{\boldsymbol{R}}_{c_k b_t} (\boldsymbol{I} + [\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge)]^T (\bar{\boldsymbol{g}}^{c_k} + \delta \boldsymbol{g}^{c_k}) && (60) \\ = & [\bar{\boldsymbol{R}}_{c_k b_t} + \bar{\boldsymbol{R}}_{c_k b_t} [\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge]^T (\bar{\boldsymbol{g}}^{c_k} + \delta \boldsymbol{g}^{c_k}) && (61) \\ = & (\bar{\boldsymbol{R}}_{c_k b_t}^T + [[\delta \boldsymbol{\phi}_{c_k b_t}]^\wedge]^T \bar{\boldsymbol{R}}_{c_k b_t}^T)(\bar{\boldsymbol{g}}^{c_k} + \delta \boldsymbol{g}^{c_k}). && (62) \end{aligned}$$

Since $[\delta\boldsymbol{\phi}_{c_kb_t}]^\wedge$ is a skew symmetric matrix, Eq. (62) can be rewritten as:

$$\begin{aligned}
&[\bar{\boldsymbol{R}}_{c_kb_t}exp([\delta\boldsymbol{\phi}_{c_kb_t}]^\wedge)]^T(\bar{\boldsymbol{g}}^{c_k}+\delta\boldsymbol{g}^{c_k}) && (63)\\
\approx&(\bar{\boldsymbol{R}}_{c_kb_t}^T+[[\delta\boldsymbol{\phi}_{c_kb_t}]^\wedge]^T\bar{\boldsymbol{R}}_{c_kb_t}^T)(\bar{\boldsymbol{g}}^{c_k}+\delta\boldsymbol{g}^{c_k}) && (64)\\
=&(\bar{\boldsymbol{R}}_{c_kb_t}^T-[\delta\boldsymbol{\phi}_{c_kb_t}]^\wedge\bar{\boldsymbol{R}}_{c_kb_t}^T)(\bar{\boldsymbol{g}}^{c_k}+\delta\boldsymbol{g}^{c_k}). && (65)
\end{aligned}$$

By inserting Eq. (65) into Eq. (58) and rearranging the resulting formula:

$$\begin{aligned}
\boldsymbol{a}^{b_t}=\;&\boldsymbol{a}_m^{b_t}-\bar{\boldsymbol{b}}_a^{b_t}-\delta\boldsymbol{b}_a^{b_t}-\boldsymbol{n}_a-\bar{\boldsymbol{R}}_{c_kb_t}^T\bar{\boldsymbol{g}}^{c_k}-\bar{\boldsymbol{R}}_{c_kb_t}^T\delta\boldsymbol{g}^{c_k}\\
&+[\delta\boldsymbol{\phi}_{c_kb_t}]^\wedge\bar{\boldsymbol{R}}_{c_kb_t}^T\bar{\boldsymbol{g}}^{c_k}+[\delta\boldsymbol{\phi}_{c_kb_t}]^\wedge\bar{\boldsymbol{R}}_{c_kb_t}^T\delta\boldsymbol{g}^{c_k}. && (66)
\end{aligned}$$

By separating the nominal and stochastic terms in Eq. (66), we have:

$$\begin{aligned}
\bar{\boldsymbol{a}}^{b_t}=\;&\boldsymbol{a}_m^{b_t}-\bar{\boldsymbol{R}}_{c_kb_t}^T\bar{\boldsymbol{g}}^{c_k}-\bar{\boldsymbol{b}}_a^{b_t}, && (67)\\
\delta\boldsymbol{a}^{b_t}=\;&-\delta\boldsymbol{b}_a^{b_t}-\boldsymbol{n}_a-\bar{\boldsymbol{R}}_{c_kb_t}^T\delta\boldsymbol{g}^{c_k}\\
&+[\delta\boldsymbol{\phi}_{c_kb_t}]^\wedge\bar{\boldsymbol{R}}_{c_kb_t}^T\bar{\boldsymbol{g}}^{c_k}+[\delta\boldsymbol{\phi}_{c_kb_t}]^\wedge\bar{\boldsymbol{R}}_{c_kb_t}^T\delta\boldsymbol{g}^{c_k} && (68)\\
\approx\;&-\delta\boldsymbol{b}_a^{b_t}-\boldsymbol{n}_a-\bar{\boldsymbol{R}}_{c_kb_t}^T\delta\boldsymbol{g}^{c_k}+[\delta\boldsymbol{\phi}_{c_kb_t}]^\wedge\bar{\boldsymbol{R}}_{c_kb_t}^T\bar{\boldsymbol{g}}^{c_k}, && (69)
\end{aligned}$$

where the high-order small term $[\delta\boldsymbol{\phi}_{c_kb_t}]^\wedge\bar{\boldsymbol{R}}_{c_kb_t}^T\delta\boldsymbol{g}^{c_k}$ in Eq. (68) is neglected. By inserting Eq. (69) into Eq. (56), we have:

$$\begin{aligned}
\delta\dot{\boldsymbol{v}}^{c_k}\approx\;&-\bar{\boldsymbol{R}}_{c_kb_t}\delta\boldsymbol{b}_a^{b_t}-\bar{\boldsymbol{R}}_{c_kb_t}\boldsymbol{n}_a-\bar{\boldsymbol{R}}_{c_kb_t}\bar{\boldsymbol{R}}_{c_kb_t}^T\delta\boldsymbol{g}^{c_k}\\
&+\bar{\boldsymbol{R}}_{c_kb_t}[\delta\boldsymbol{\phi}_{c_kb_t}]^\wedge\bar{\boldsymbol{R}}_{c_kb_t}^T\bar{\boldsymbol{g}}^{c_k}-\bar{\boldsymbol{R}}_{c_kb_t}[\bar{\boldsymbol{a}}^{b_t}]^\wedge\delta\boldsymbol{\phi}_{c_kb_t} && (70)\\
=\;&-\bar{\boldsymbol{R}}_{c_kb_t}\delta\boldsymbol{b}_a^{b_t}-\bar{\boldsymbol{R}}_{c_kb_t}\boldsymbol{n}_a-\delta\boldsymbol{g}^{c_k}\\
&-\bar{\boldsymbol{R}}_{c_kb_t}[\bar{\boldsymbol{R}}_{c_kb_t}^T\bar{\boldsymbol{g}}^{c_k}]^\wedge\delta\boldsymbol{\phi}_{c_kb_t}-\bar{\boldsymbol{R}}_{c_kb_t}[\bar{\boldsymbol{a}}^{b_t}]^\wedge\delta\boldsymbol{\phi}_{c_kb_t} && (71)\\
=\;&-\bar{\boldsymbol{R}}_{c_kb_t}\delta\boldsymbol{b}_a^{b_t}-\bar{\boldsymbol{R}}_{c_kb_t}\boldsymbol{n}_a-\delta\boldsymbol{g}^{c_k}\\
&-\bar{\boldsymbol{R}}_{c_kb_t}([\bar{\boldsymbol{R}}_{c_kb_t}^T\bar{\boldsymbol{g}}^{c_k}]^\wedge+[\bar{\boldsymbol{a}}^{b_t}]^\wedge)\delta\boldsymbol{\phi}_{c_kb_t} && (72)\\
=\;&-\bar{\boldsymbol{R}}_{c_kb_t}\delta\boldsymbol{b}_a^{b_t}-\bar{\boldsymbol{R}}_{c_kb_t}\boldsymbol{n}_a-\delta\boldsymbol{g}^{c_k}\\
&-\bar{\boldsymbol{R}}_{c_kb_t}[\bar{\boldsymbol{R}}_{c_kb_t}^T\bar{\boldsymbol{g}}^{c_k}+\bar{\boldsymbol{a}}^{b_t}]^\wedge\delta\boldsymbol{\phi}_{c_kb_t}. && (73)
\end{aligned}$$

Based on Eq. (24,25,26,39,47,73) and the continuous-time error propagation model Eq. (23), $\boldsymbol{F}$ and $\boldsymbol{G}$ can be written as:

$$\boldsymbol{F}=\begin{bmatrix}
-[\bar{\boldsymbol{w}}^{b_t}]^\wedge & 0 & 0 & 0 & -\boldsymbol{I}_3 & 0\\
0 & 0 & \boldsymbol{I}_3 & 0 & 0 & 0\\
-\bar{\boldsymbol{R}}_{c_kb_t}[\bar{\boldsymbol{R}}_{c_kb_t}^T\bar{\boldsymbol{g}}^{c_k}+\bar{\boldsymbol{a}}^{b_t}]^\wedge & 0 & 0 & -\boldsymbol{I}_3 & 0 & -\bar{\boldsymbol{R}}_{c_kb_t}\\
0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix},\boldsymbol{G}=\begin{bmatrix}
-\boldsymbol{I}_3 & 0 & 0 & 0\\
0 & 0 & 0 & 0\\
0 & 0 & -\bar{\boldsymbol{R}}_{c_kb_t} & 0\\
0 & 0 & 0 & 0\\
0 & \boldsymbol{I}_3 & 0 & 0\\
0 & 0 & 0 & \boldsymbol{I}_3
\end{bmatrix}. \quad (74)$$

$\bar{\boldsymbol{w}}^{b_t}$ and $\bar{\boldsymbol{a}}^{b_t}$ are given in Eq. (42) and Eq. (67), respectively.

Given the continuous error propagation model and the initial condition $\boldsymbol{\Phi}_{t_\tau,t_\tau}=\boldsymbol{I}_{18}$, the discrete state-transition matrix $\boldsymbol{\Phi}_{(t_{\tau+1},t_\tau)}$ can be found by solving $\dot{\boldsymbol{\Phi}}_{(t_{\tau+1},t_\tau)} = \boldsymbol{F}_{t_{\tau+1}}\boldsymbol{\Phi}_{(t_{\tau+1},t_\tau)}$ [3]:

$$\boldsymbol{\Phi}_{t_{\tau+1},t_\tau} = exp(\int_{t_\tau}^{t_{\tau+1}} \boldsymbol{F}(s)\mathrm{d}s) \approx \boldsymbol{I}_{18} + \boldsymbol{F}\delta t + \frac{1}{2}\boldsymbol{F}^2\delta t^2, \quad \delta t = t_{\tau+1} - t_\tau. \tag{75}$$

Let $\check{\boldsymbol{P}}$ and $\hat{\boldsymbol{P}}$ denote the prior and posterior covariance estimates during propagation and after an update given new observations. Then we have [1,3]:

$$\check{\boldsymbol{P}}_{t_{\tau+1}} = \boldsymbol{\Phi}_{t_{\tau+1},t_\tau}\check{\boldsymbol{P}}_{t_\tau}\boldsymbol{\Phi}^T_{t_{\tau+1},t_\tau} + \boldsymbol{Q}_{t_\tau}, \tag{76}$$

$$\boldsymbol{Q}_{t_\tau} = \int_{t_\tau}^{t_{\tau+1}} \boldsymbol{\Phi}_{s,t_\tau}\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^T\boldsymbol{\Phi}^T_{s,t_\tau}\mathrm{d}s \approx \boldsymbol{\Phi}_{t_{\tau+1},t_\tau}\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^T\boldsymbol{\Phi}^T_{t_{\tau+1},t_\tau}\delta t, \tag{77}$$

where $\boldsymbol{Q} = \mathcal{D}([\sigma_w^2\boldsymbol{I}_3, \sigma_{b_w}^2\boldsymbol{I}_3, \sigma_a^2\boldsymbol{I}_3, \sigma_{b_a}^2\boldsymbol{I}_3])$. $\mathcal{D}$ is the diagonalization function.

S3 Derivation of Camera-Centric EKF Update

In general, given an observation measurement $\boldsymbol{\xi}_{k+1}$ and its corresponding covariance $\boldsymbol{\Gamma}_{k+1}$ from the camera sensor at time t_{k+1}, we assume the following observation model: $\boldsymbol{\xi}_{k+1} = h(\boldsymbol{x}_{k+1}) + \boldsymbol{n}_r,\ \boldsymbol{n}_r \sim N(0, \boldsymbol{\Gamma}_{k+1})$.

Let $\boldsymbol{H}_{k+1} = \frac{\partial h(\boldsymbol{x}_{k+1})}{\partial \delta \boldsymbol{x}_{k+1}}$. Then the EKF update applies as following:

$$\boldsymbol{K}_{k+1} = \check{\boldsymbol{P}}_{k+1}\boldsymbol{H}^T_{k+1}(\boldsymbol{H}_{k+1}\check{\boldsymbol{P}}_{k+1}\boldsymbol{H}^T_{k+1} + \boldsymbol{\Gamma}_{k+1})^{-1}, \tag{78}$$

$$\hat{\boldsymbol{P}}_{k+1} = (\boldsymbol{I}_{18} - \boldsymbol{K}_{k+1}\boldsymbol{H}_{k+1})\check{\boldsymbol{P}}_{k+1}, \tag{79}$$

$$\delta\hat{\boldsymbol{x}}_{k+1} = \boldsymbol{K}_{k+1}(\boldsymbol{\xi}_{k+1} - h(\check{\boldsymbol{x}}_{k+1})). \tag{80}$$

In DynaDepth, the observation measurement is defined as the ego-motion predicted by $\mathcal{M}_p$, i.e., $\boldsymbol{\xi}_{k+1} = [\tilde{\boldsymbol{\phi}}^T_{c_kc_{k+1}}, \tilde{\boldsymbol{p}}^T_{c_kc_{k+1}}]^T$. Accordingly, we define $h(\boldsymbol{x}_{k+1})$ as $h(\boldsymbol{x}_{k+1}) = [h^T_{\boldsymbol{\phi}}(\boldsymbol{x}_{k+1}), h^T_{\boldsymbol{p}}(\boldsymbol{x}_{k+1})]^T$. We first consider the observation model $h_\phi(\boldsymbol{x}_{k+1})$ for rotation. Assuming $[\cdot]^\vee$ as the inverse function of $[\cdot]^\wedge$, then we have:

$$h_{\boldsymbol{\phi}}(\boldsymbol{x}_{k+1}) = \phi_{c_kc_{k+1}} = ln([\boldsymbol{R}_{c_kb_{k+1}}\boldsymbol{R}_{bc}]^\vee). \tag{81}$$

$\{\boldsymbol{R}_{bc}, \boldsymbol{p}_{bc}\}$ and $\{\boldsymbol{R}_{cb}, \boldsymbol{p}_{cb}\}$ denote the extrinsics between camera and IMU. By inserting Eq. (20) into Eq. (81), we have:

$$\begin{aligned} h_{\boldsymbol{\phi}}(\boldsymbol{x}_{k+1}) &= ln([\boldsymbol{R}_{c_kb_{k+1}}\boldsymbol{R}_{bc}]^\vee) && (82)\\ &= ln([\bar{\boldsymbol{R}}_{c_kb_{k+1}}exp([\delta\boldsymbol{\phi}_{c_kb_{k+1}}]^\wedge)\boldsymbol{R}_{bc}]^\vee) && (83)\\ &= ln([\bar{\boldsymbol{R}}_{c_kb_{k+1}}\boldsymbol{R}_{bc}\boldsymbol{R}_{cb}exp([\delta\boldsymbol{\phi}_{c_kb_{k+1}}]^\wedge)\boldsymbol{R}_{bc}]^\vee). && (84) \end{aligned}$$

We then separate the expression in $[\cdot]^\vee$ in Eq. (84) into the following two parts:

$$\begin{aligned} \bar{\boldsymbol{R}}_{c_kb_{k+1}}\boldsymbol{R}_{bc} &= \bar{\boldsymbol{R}}_{c_kc_{k+1}} = exp([\bar{\boldsymbol{\phi}}_{c_kc_{k+1}}]^\wedge), && (85)\\ \boldsymbol{R}_{cb}exp([\delta\boldsymbol{\phi}_{c_kb_{k+1}}]^\wedge)\boldsymbol{R}_{bc} &\approx \boldsymbol{R}_{cb}(\boldsymbol{I} + [\delta\boldsymbol{\phi}_{c_kb_{k+1}}]^\wedge)\boldsymbol{R}_{bc} && (86)\\ &= \boldsymbol{I} + \boldsymbol{R}_{cb}[\delta\boldsymbol{\phi}_{c_kb_{k+1}}]^\wedge\boldsymbol{R}_{bc}. && (87) \end{aligned}$$

By using the equation $[\boldsymbol{R}\delta\boldsymbol{\phi}]^\wedge = \boldsymbol{R}[\delta\boldsymbol{\phi}]^\wedge\boldsymbol{R}^T$, Eq. (87) can be rewritten as:

$$\begin{aligned}\boldsymbol{R}_{cb}exp([\delta\boldsymbol{\phi}_{c_k b_{k+1}}]^\wedge)\boldsymbol{R}_{bc} &\approx \boldsymbol{I} + [\boldsymbol{R}_{cb}\delta\boldsymbol{\phi}_{c_k b_{k+1}}]^\wedge \quad (88)\\ &\approx exp([\boldsymbol{R}_{cb}\delta\boldsymbol{\phi}_{c_k b_{k+1}}]^\wedge). \quad (89)\end{aligned}$$

By inserting Eq. (85) and Eq. (89) into Eq. (84), and approximating the resulting exponential function using the Baker–Campbell–Hausdorff (BCH) approximation formula [1], we have:

$$\begin{aligned}h_{\boldsymbol{\phi}}(\boldsymbol{x}_{k+1}) &\approx ln([exp([\bar{\boldsymbol{\phi}}_{c_k c_{k+1}}]^\wedge)exp([\boldsymbol{R}_{cb}\delta\boldsymbol{\phi}_{c_k b_{k+1}}]^\wedge)]^\vee) \quad (90)\\ &\approx \bar{\boldsymbol{\phi}}_{c_k c_{k+1}} + J_l^{-1}(-\bar{\boldsymbol{\phi}}_{c_k c_{k+1}})\boldsymbol{R}_{cb}\delta\boldsymbol{\phi}_{c_k b_{k+1}}. \quad (91)\end{aligned}$$

The definition of the inversed SO(3) left Jacobian $J_l^{-1}(\cdot)$ is given by [1]:

$$J_l^{-1}(\boldsymbol{\phi}) = \frac{\phi}{2}cot\frac{\phi}{2}\mathbf{1} + (1 - \frac{\phi}{2}cot\frac{\phi}{2})\boldsymbol{\alpha}\boldsymbol{\alpha}^T - \frac{\phi}{2}\boldsymbol{\alpha}^\wedge, \quad (92)$$

where $\phi = |\boldsymbol{\phi}|$ and $\boldsymbol{\alpha} = \boldsymbol{\phi}/\phi$. Based on Eq. (91), we can compute the nominal prior and the derivative w.r.t. $\delta\boldsymbol{x}_{k+1}$ for the rotation as:

$$\begin{aligned}h_{\boldsymbol{\phi}}(\check{\boldsymbol{x}}_{k+1}) &= \bar{\boldsymbol{\phi}}_{c_k c_{k+1}}, \quad (93)\\ \frac{\partial h_{\boldsymbol{\phi}}(\boldsymbol{x}_{k+1})}{\partial \delta\boldsymbol{x}_{k+1}} &= \left[J_l(-\bar{\boldsymbol{\phi}}_{c_k c_{k+1}})^{-1}\boldsymbol{R}_{cb}\ 0\ 0\ 0\ 0\ 0\right]. \quad (94)\end{aligned}$$

We then derive the observation model $h_{\boldsymbol{p}}(\boldsymbol{x}_{k+1})$ for the translation as below:

$$h_{\boldsymbol{p}}(\boldsymbol{x}_{k+1}) = \boldsymbol{p}_{c_k c_{k+1}} = \boldsymbol{R}_{c+kb_{k+1}}\boldsymbol{p}_{bc} + \boldsymbol{p}_{c_k b_{k+1}}. \quad (95)$$

By inserting Eq. (20) into Eq. (95) and using the equation $\boldsymbol{u}^\wedge\boldsymbol{v} = -\boldsymbol{v}^\wedge\boldsymbol{u}$:

$$\begin{aligned}h_{\boldsymbol{p}}(\boldsymbol{x}_{k+1}) &= \boldsymbol{R}_{c+kb_{k+1}}\boldsymbol{p}_{bc} + \boldsymbol{p}_{c_k b_{k+1}} \quad (96)\\ &= \bar{\boldsymbol{R}}_{c_k b_{k+1}}exp([\delta\boldsymbol{\phi}_{c_k b_{k+1}}]^\wedge)\boldsymbol{p}_{bc} + \bar{\boldsymbol{p}}_{c_k b_{k+1}} + \delta\boldsymbol{p}_{c_k b_{k+1}} \quad (97)\\ &\approx \bar{\boldsymbol{R}}_{c_k b_{k+1}}(\boldsymbol{I} + [\delta\boldsymbol{\phi}_{c_k b_{k+1}}]^\wedge)\boldsymbol{p}_{bc} + \bar{\boldsymbol{p}}_{c_k b_{k+1}} + \delta\boldsymbol{p}_{c_k b_{k+1}} \quad (98)\\ &= \bar{\boldsymbol{R}}_{c_k b_{k+1}}\boldsymbol{p}_{bc} + \bar{\boldsymbol{R}}_{c_k b_{k+1}}[\delta\boldsymbol{\phi}_{c_k b_{k+1}}]^\wedge\boldsymbol{p}_{bc} + \bar{\boldsymbol{p}}_{c_k b_{k+1}} + \delta\boldsymbol{p}_{c_k b_{k+1}} \quad (99)\\ &= \bar{\boldsymbol{R}}_{c_k b_{k+1}}\boldsymbol{p}_{bc} + \bar{\boldsymbol{p}}_{c_k b_{k+1}} - \bar{\boldsymbol{R}}_{c_k b_{k+1}}[\boldsymbol{p}_{bc}]^\wedge\delta\boldsymbol{\phi}_{c_k b_{k+1}} + \delta\boldsymbol{p}_{c_k b_{k+1}}. \quad (100)\end{aligned}$$

Based on Eq. (100), we can then compute the nominal prior and the derivative w.r.t. $\delta\boldsymbol{x}_{k+1}$ for the translation as:

$$\begin{aligned}h_{\boldsymbol{p}}(\check{\boldsymbol{x}}_{k+1}) &= \bar{\boldsymbol{R}}_{c_k b_{k+1}}\boldsymbol{p}_{bc} + \bar{\boldsymbol{p}}_{c_k b_{k+1}}, \quad (101)\\ \frac{\partial h_{\boldsymbol{p}}(\boldsymbol{x}_{k+1})}{\partial \delta\boldsymbol{x}_{k+1}} &= \left[-\bar{\boldsymbol{R}}_{c_k b_{k+1}}[\boldsymbol{p}_{bc}]^\wedge\ \boldsymbol{I}_3\ 0\ 0\ 0\ 0\right]. \quad (102)\end{aligned}$$

To finish the camera-centric EKF update step, we combine the derivation results in Eq. (93, 94, 101, 102), and write $h(\check{\boldsymbol{x}}_{k+1})$ and $\boldsymbol{H}_{k+1}$ as:

$$h(\check{\boldsymbol{x}}_{k+1}) = \begin{bmatrix}\bar{\boldsymbol{\phi}}_{c_k c_{k+1}}\\ \bar{\boldsymbol{R}}_{c_k b_{k+1}}\boldsymbol{p}_{bc} + \bar{\boldsymbol{p}}_{c_k b_{k+1}}\end{bmatrix}, \quad (103)$$

$$\boldsymbol{H}_{k+1} = \begin{bmatrix}J_l(-\bar{\boldsymbol{\phi}}_{c_k c_{k+1}})^{-1}\boldsymbol{R}_{cb} & 0 & 0\ 0\ 0\ 0\\ -\bar{\boldsymbol{R}}_{c_k b_{k+1}}[\boldsymbol{p}_{bc}]^\wedge & \boldsymbol{I}_3 & 0\ 0\ 0\ 0\end{bmatrix}. \quad (104)$$

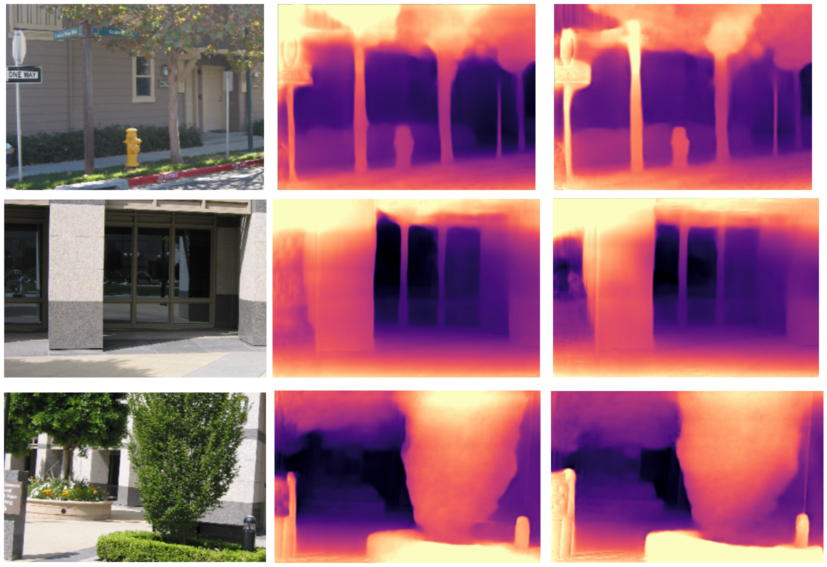

Fig. 1: More generalization results on Make3D using models trained on KITTI with (w/) and without (w.o/) IMU.

Finally, by inserting Eq. (103-104) into Eq. (78-80), we can perform the camera-centric EKF update step to get the updated posterior error states $\delta\hat{\boldsymbol{x}}_{k+1}$ and calculate the EKF updated camera ego-motion, based on $\delta\hat{\boldsymbol{x}}_{k+1}$ and the propagated nominal states which can be obtained from the camera-centric IMU preintegration results, i.e., Eq. (10) and Eq.(19).

S4 More Generalization Results on Make3D

We present more generalization results on Make3D [4] using models trained on KITTI [2] with (w/) and without (w.o/) IMU in Fig. 1-3. By using IMU, it can be seen that the model generalizes better in unseen datasets, especially in the glass and shadow areas, where the underlying assumption of visual photometric consistency can be easily violated. In addition, the model using IMU recovers more delicate texture details, which further justifies the benefit of using the IMU motion dynamics that is independent with the visual information during training.

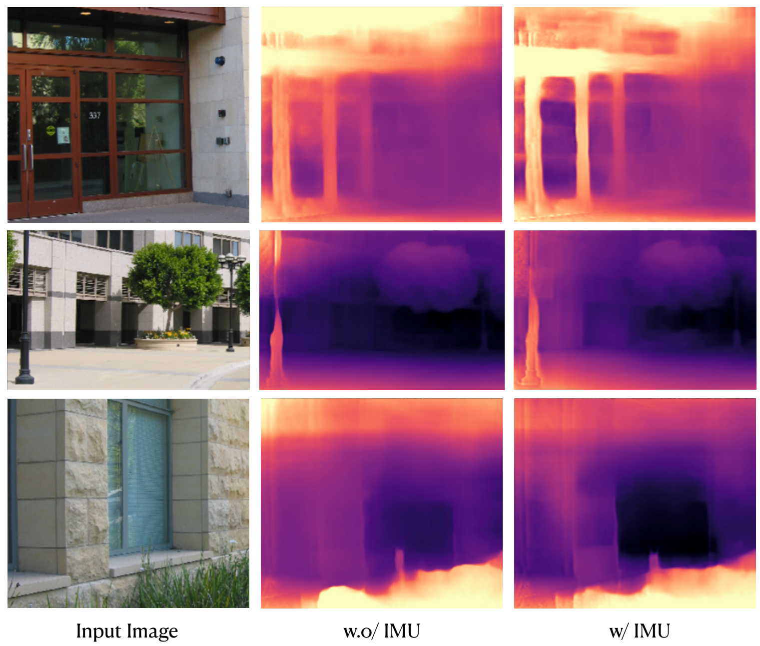

Fig. 2: More generalization results on Make3D using models trained on KITTI with (w/) and without (w.o/) IMU.

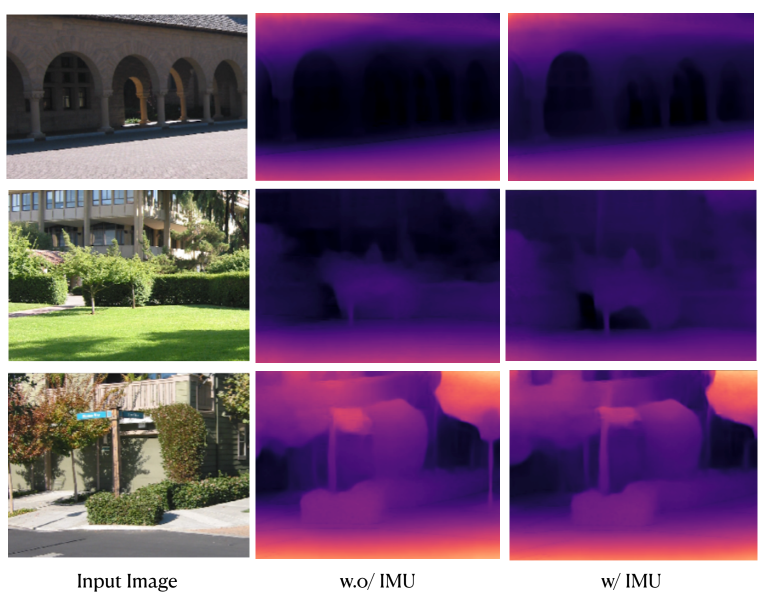

Fig. 3: More generalization results on Make3D using models trained on KITTI with (w/) and without (w.o/) IMU.

References

1. Barfoot, T.D.: State estimation for robotics. Cambridge University Press (2017)
2. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. The International Journal of Robotics Research **32**(11), 1231–1237 (2013)
3. Huang, G.: Visual-inertial navigation: A concise review. In: 2019 IEEE International Conference on Robotics and Automation (ICRA). pp. 9572–9582. IEEE (2019)
4. Saxena, A., Sun, M., Ng, A.Y.: Make3d: Learning 3d scene structure from a single still image. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(5), 824–840 (2008)