Visual-Inertial-Wheel Odometry with Online Calibration

Woosik Lee, Kevin Eckenhoff, Yulin Yang, Patrick Geneva, and Guoquan Huang

***Abstract*— In this paper, we introduce a novel visual-inertial-wheel odometry (VIWO) system for ground vehicles, which efficiently fuses multi-modal visual, inertial and 2D wheel odometry measurements in a sliding-window filtering fashion. As multi-sensor fusion requires both intrinsic and extrinsic (spatiotemproal) calibration parameters which may vary over time during terrain navigation, we propose to perform VIWO along with online sensor calibration of wheel encoders' intrinsic and extrinsic parameters. To this end, we analytically derive the 2D wheel odometry measurement model from the raw wheel encoders' readings and optimally fuse this 2D relative motion information with 3D visual-inertial measurements. Additionally, an observability analysis is performed for the linearized VIWO system, which identifies five commonly-seen degenerate motions for wheel calibration parameters. The proposed VIWO system has been validated extensively in both Monte-Carlo simulations and real-world experiments in large-scale urban driving scenarios.**

I. INTRODUCTION

Autonomous ground vehicles have found their ways into many applications from autonomous driving and warehousing to military and agriculture robots navigating off-road rough terrains. High-precision consistent 3D localization with low-cost multi-modal sensors (given that engineered robotic systems have limited cost budgets) is essential to enable such autonomy. Among all possible sensor suites, cameras, IMUs and wheel encoders are appealing because of their ubiquity and low costs while providing sufficient information for 3D motion estimation. While visual-inertial odometry (VIO) has witnessed great success in recent years and has shown that even a single camera and IMU can provide accurate 3D motion tracking [1], VIO can suffer from unobservability if it undergoes planar motions such as constant acceleration which is a common case for ground vehicles. Thus, it is necessary to further aid VIO with additional sensors such as wheel encoders (which are typically come with wheeled ground vehicles) [2], [3]. As such, in this work we aim to develop efficient visual-inertial-wheel odometry (VIWO).

It is well understood that accurate calibration is a prerequisite for multi-sensor fusion, which is often obtained offline and assumed to be constant during operation (e.g., [2]). However, calibration parameters may change over time, for example, due to mechanical vibration and environmental effects, which, if not addressed properly, can hurt the estimation performance (e.g., see Fig. 1). We thus advocate to perform online sensor calibration of both intrinsic and extrinsic parameters in the proposed VIWO.

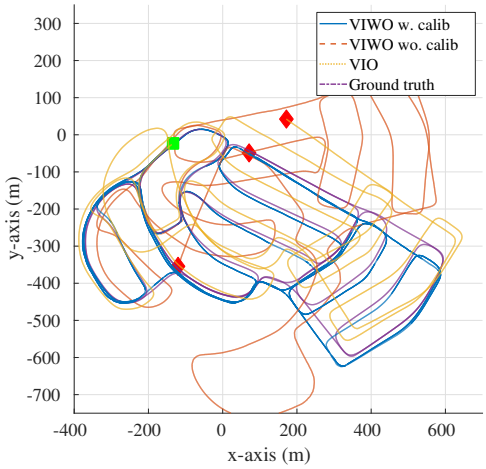

Fig. 1: Simulation results for VIWO with online calibration (blue), VIWO without calibration (red), and VIO (yellow). The green square and red diamond correspond to the start and end of the 8.9km trajectory, respectively.

In particular, an efficient, consistent, and tightly-coupled VIWO estimator is developed, which also performs online sensor calibration of the wheel encoder's intrinsics and the spationtemporal extrinsics between the wheel odometer and IMU/camera. Note that IMU/camera calibration has been well studied in our prior work [4] and is assumed known in this work without a loss of generality. Note also that we employ a differential-drive kinematic model [5], instead of the instantaneous center of rotation (ICR)-based model for skid-steering robots as in [3], arguably because it better models the actual kinematics of two-wheel driven (not tracked) vehicles and its intrinsic calibration provides robustness to non-systematic errors. Specifically, the main contributions of this work include:

- We develop an efficient MSCKF-based VIWO estimator to optimally fuse IMU, camera, and preintegrated wheel measurements, which particularly models the wheel-IMU time offset and performs online calibration of both intrinsic and extrinsic spatiotemproal parameters.
- We derive a state transition matrix for the MSCKF-based linearized system, which implicitly incorporates the stochastic cloning and perform the observability analysis based on it to identify degenerate motions that cause calibration parameters to become unobservable.
- We evaluate the proposed VIWO extensively in simulations, showing the calibration convergence, the robustness to poor initial calibration values, and the estimation consistency. We demonstrate the unobservability of the calibration parameters when the vehicle undergoes degenerate motions. The proposed VIWO is also validated in a large-scale urban driving experiment.

This work was partially supported by the University of Delaware (UD) College of Engineering, the NSF (IIS-1924897), the ARL (W911NF-19-2-0226, JWS 10-051-003), and Google ARCore. Yang was partially supported by the University Doctoral Fellowship and Geneva by the Delaware Space Grant College and Fellowship Program (NASA Grant NNX15AI19H).

The authors are with the Robot Perception and Navigation Group (RPNG), University of Delaware, Newark, DE 19716, USA. {woosik, keck, yuyang, pgeneva, ghuang}@udel.edu

II. RELATED WORK

Rich literature exists on differential drive kinematic models and their offline intrinsic calibration [6]–[9], for example, Kummerle et al. [10] performed online calibration within a graph optimization framework and [11], [12] calibrated sensors within the filtering framework. These methods focus on processing odometry measurements at the sensor rate and thus can suffer computational burdens as the frequency increases. In contrast, preintegration of odometry measurements over time create an inferred measurement about the change of the robot state, and allows update at a lower rate, thus reducing computational burdens. For example, wheel odometry measurements are preintegrated for 2D localization with online intrinsic calibration in [13], [14]. Note that in [3], sliding-window optimization for 3D localization of an ICR-based skid-steering robot [15], along with online wheel intrinsic calibration was performed. Although the ICR model encompasses differential drive models as a special case, as the model incorporates the wheel slippage into the intrinsics, the ICR intrinsic parameters were modeled as a random walk, which requires more parameters to tune and may introduce more degenerate cases.

Offline extrinsic calibration of the wheel odometry and other sensors has also been well studied. For example, Antonelli et al. [16] performed relative 3D orientation and 2D position calibration between the camera and the odometry frame with known landmarks, while Heng et al. [17] calibrated the wheel and multi-camera without the need of environmental targets. Online calibration approaches also exist, for instance, in [18]–[20] graph optimization was used for 6DOF spatial calibration. However, all these methods assumed the sensors are time-synchronized. Note that the time offset is different from sensor asynchronicity and describes the disagreement between sensor clocks. As shown in our previous work [21], even a small time offset error between the sensors can greatly affect estimation errors. To the best of our knowledge, no work has yet investigated this time offset calibration between wheel odometry and IMU/camera, although significant research efforts in VIO has studied both online calibration of IMU/camera time offset [22]–[26] and offline calibration [27]–[30].

Observability analysis [31], [32] is important due to its ability to indicate the minimum states needed to initialize an estimator, identify degenerate motions which can affect estimation accuracy [2], [33], and design consistent estimators [34]. Understanding observability properties and degenerate motions is especially essential for ground vehicles due to their constrained motions that prevent the full excitation of the sensor suite. For the differential drive model, Martinelli et al. [35] showed that with camera and wheel sensors, the wheel odometry intrinsic and 2D extrinsic calibration parameters are observable up to the scale under general motion. The author also found a degenerate motion where the parameters are not observable when each wheel keeps the same speed ratio. Censi et al. [36] performed intrinsic and 2D extrinsic calibration between wheel and laser sensors and showed three motion profiles for which the parameters are observable, but did not provide specific degenerate motions. Other then differential drive model, Zuo et al. [3] analyzed the observability of intrinsics of skid-steering model with degenerate motion studies, Yang et al. [4] did comprehensive observability studies on extrinsic and time offset parameters between a camera and an IMU and provided the degenerate motion analysis.

III. VIO WITH MSCKF

In this section, we briefly review the VIO within the standard MSCKF framework [37], which serve as the baseline for the proposed VIWO system. Specifically, at time t_k, the state vector $\mathbf{x}_k$ consists of the current inertial state $\mathbf{x}_{I_k}$ and n historical IMU pose clones $\mathbf{x}_{C_k}$ represented in the global frame $\{G\}$:

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

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

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

where ${}^{I_k}_{G}\bar{q}$ is the JPL unit quaternion [38] corresponding to the rotation ${}^{I_k}_{G}\mathbf{R}$ from $\{G\}$ to IMU frame $\{I\}$, ${}^{G}\mathbf{p}_{I_k}$ and ${}^{G}\mathbf{v}_{I_k}$ are the position and velocity of $\{I\}$ in $\{G\}$, and $\mathbf{b}_g$ and $\mathbf{b}_a$ are the biases of the gyroscope and accelerometer. We define $\mathbf{x} = \hat{\mathbf{x}} \boxplus \tilde{\mathbf{x}}$, where $\mathbf{x}$ is the true state, $\hat{\mathbf{x}}$ is its estimate, $\tilde{\mathbf{x}}$ is the error state, and the operation $\boxplus$ which maps the error state vector to its corresponding manifold [39].

A. IMU Kinematic Model

The state is propagated forward in time using the IMU linear acceleration $\mathbf{a}_m$ and angular velocity $\boldsymbol{\omega}_m$ measurements:

$$\mathbf{a}_m = \mathbf{a} + {}^{I}_{G}\mathbf{R}\mathbf{g} + \mathbf{b}_a + \mathbf{n}_a \ , \quad \boldsymbol{\omega}_m = \boldsymbol{\omega} + \mathbf{b}_g + \mathbf{n}_g \tag{4}$$

where $\mathbf{a}$ and $\boldsymbol{\omega}$ are the true local acceleration and angular velocity, $\mathbf{g} \approx [0\ 0\ 9.81]^\top$ is the global gravity, and $\mathbf{n}_a$ and $\mathbf{n}_g$ are zero mean Gaussian noises. These measurements are used to propagate the state estimate and covariance from time t_k to t_{k+1} based on the standard inertial kinematic model $\mathbf{f}(\cdot)$ [38]:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{f}(\hat{\mathbf{x}}_{k|k}, \mathbf{a}_m, \boldsymbol{\omega}_m, \mathbf{0}, \mathbf{0}) \tag{5}$$

$$\mathbf{P}_{k+1|k} = \boldsymbol{\Phi}(t_{k+1}, t_k)\mathbf{P}_{k|k}\boldsymbol{\Phi}(t_{k+1}, t_k)^\top + \mathbf{Q}_k \tag{6}$$

where $\hat{\mathbf{x}}_{a|b}$ denotes the estimate at time t_a formed by processing the measurements up to time t_b, and $\boldsymbol{\Phi}$ and $\mathbf{Q}$ are the state transition matrix and discrete noise covariance.

B. Camera Measurement Model

Sparse corner features are detected and tracked over a window of images associated with the cloned frames $\mathbf{x}_{C_k}$. The resulting bearing measurements, $\mathbf{z}_k$, are given by:

$$\mathbf{z}_k = \boldsymbol{\Pi}({}^{C_k}\mathbf{p}_f) + \mathbf{n}_k \tag{7}$$

$${}^{C_k}\mathbf{p}_f = {}^{C}_{I}\mathbf{R}\,{}^{I_k}_{G}\mathbf{R}({}^{G}\mathbf{p}_f - {}^{G}\mathbf{p}_{I_k}) + {}^{C}\mathbf{p}_I \tag{8}$$

where $\boldsymbol{\Pi}\left([x\ y\ z]^\top\right) = \begin{bmatrix}\frac{x}{z} & \frac{y}{z}\end{bmatrix}^\top$ is the perspective projection, ${}^{G}\mathbf{p}_f$ is the 3D point feature position, and ${}^{C}_{I}\mathbf{R}$ and ${}^{C}\mathbf{p}_I$ are the camera to IMU extrinsic transformation.[1] Stacking

[1]Note that in this paper we assume that the camera-IMU extrinsics along with the temporal calibration (i.e., the time offset between the two sensors), are known – which however can be calibrated either offline or online [40] – so that we can focus on the online intrinsic and extrinsic calibrations of the odometer and IMU in the proposed VIWO system.

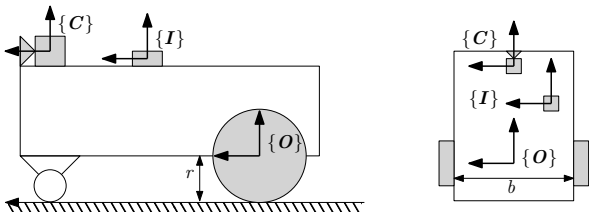

Fig. 2: Sensor frames in this paper: IMU frame {I}, camera sensor frame {C}, and odometry frame {O} located at the center of the baselink

all measurements corresponding to a single feature and performing linear marginalization of the feature's position (via a null space operation) results in a residual that depends only on the state [37]:

$$\tilde{\mathbf{z}}_{c_k} = \mathbf{H}_{x_k}\tilde{\mathbf{x}}_k + \mathbf{n}_{f_k} \tag{9}$$

This then can be directly used in EKF update without storing features in the state, leading to substantial computational savings as the problem size remains bounded over time.

IV. WHEEL-ENCODER MEASUREMENT MODEL

Building upon the preceding VIO models, we now generalize our 3D motion tracking system to optimally incorporate 2D wheel-encoder measurements that are commonplace in ground vehicles. In particular, a ground vehicle is often driven by two differential (left and right) wheels mounted on a common axis (baselink), each equipped with an encoder providing local angular rate readings [5]:

$$\omega_{ml} = \omega_l + n_{\omega_l}, \qquad \omega_{mr} = \omega_r + n_{\omega_r} \tag{10}$$

where ω_l and ω_r are the true angular velocities of each wheel, and n_{ω_l} and n_{ω_r} are the corresponding zero-mean white Gaussian noises. These encoder readings can be combined to provide 2D linear and angular velocities about the vehicle body or odometer frame $\{O\}$ at the center of the baselink:

$${}^{O}\omega = (\omega_r r_r - \omega_l r_l)/b\ , \qquad {}^{O}v = (\omega_r r_r + \omega_l r_l)/2 \tag{11}$$

where $\mathbf{x}_{WI} := [r_l\ r_r\ b]^\top$ are the left and right wheel radii and the baselink length, respectively.

A. Wheel Odometry Preintegration

As the wheel encoders typically provide measurements of higher rate (e.g., 100-500 Hz) than the camera, it would be too expensive to perform EKF update at their rate. On the other hand, as a sliding window of states corresponding to the imaging times are stochastically cloned in the state vector [see (1)], we naturally preintegrate the wheel odometry measurements (11) between the two latest camera poses and then use this integrated 2D motion measurement for the MSCKF update together with the visual feature measurements [see (9)]. As a result, the state vector of our VIWO remains the same (up to online calibration) as that of the VIO, incurring only a small extra computational overhead.

Consider preintegrating wheel odometry measurements between two clone times t_k and t_{k+1}. The continuous-time 2D kinematic model for $t_\tau \in [t_k, t_{k+1}]$ is given by:

$$\dot{\theta}_{O_k}^{O_\tau} = {}^{O_\tau}\omega \tag{12}$$

$${}^{O_k}\dot{x}_{O_\tau} = {}^{O_\tau}v\ \cos(\theta_{O_k}^{O_\tau}) \tag{13}$$

$${}^{O_k}\dot{y}_{O_\tau} = {}^{O_\tau}v\ \sin(\theta_{O_k}^{O_\tau}) \tag{14}$$

Algorithm 1 VIWO Odometry Measurement Update

procedure WHEEL_UPDATE($\mathbf{x}_{k+1|k}$, $\{\omega_{ml}, \omega_{mr}\}_{k:k+1}$)
  // Preintegrate measurements and Jacobian
  $\mathbf{z} = \mathbf{0}_{3\times 1}$
  $\frac{\partial \mathbf{g}}{\partial \tilde{\mathbf{x}}_{WI}},\ \mathbf{P}_m = \mathbf{0}_3$
  for $\omega_{ml(\tau)}, \omega_{mr(\tau)} \in \{\omega_{ml(k:k+1)}, \omega_{mr(k:k+1)}\}$ **do**
    $\mathbf{z} = \mathbf{z} + \Delta\mathbf{z}$
    $\mathbf{P}_m = \mathbf{\Phi}_{tr,\tau}\mathbf{P}_m\mathbf{\Phi}_{tr,\tau}^\top + \mathbf{\Phi}_{n,\tau}\mathbf{Q}_\tau\mathbf{\Phi}_{n,\tau}^\top$
    $\frac{\partial \mathbf{g}}{\partial \tilde{\mathbf{x}}_{WI}} = \mathbf{\Phi}_{tr,\tau}\frac{\partial \mathbf{g}}{\partial \tilde{\mathbf{x}}_{WI}} + \mathbf{\Phi}_{WI,\tau}$
  end for
  // Compute residual and Jacobian
  $\tilde{\boldsymbol{z}} = \mathbf{z} - \mathbf{h}(\hat{\mathbf{x}}_I, \hat{\mathbf{x}}_C, \hat{\mathbf{x}}_{WE}, {}^{O}\hat{t}_I, \hat{\mathbf{x}}_{WI})$
  $\mathbf{H} = \begin{bmatrix} \frac{\partial \mathbf{h}}{\partial \tilde{\mathbf{x}}_I} & \frac{\partial \mathbf{h}}{\partial \tilde{\mathbf{x}}_C} & \frac{\partial \mathbf{h}}{\partial \tilde{\mathbf{x}}_{WE}} & -\frac{\partial \mathbf{g}}{\partial \tilde{\mathbf{x}}_{WI}} & \frac{\partial \mathbf{h}}{\partial {}^{O}\tilde{t}_I} \end{bmatrix}$
  // Perform χ^2 test & update
  if $\chi^2(\tilde{\boldsymbol{z}}, \mathbf{H}, \mathbf{P}_m) == Pass$ **then**
    $\hat{\mathbf{x}}_{k+1|k+1} = EKF\_Update(\hat{\mathbf{x}}_{k+1|k}, \tilde{\boldsymbol{z}}, \mathbf{H}, \mathbf{P}_m)$
  end if
end procedure

where $\theta_{O_k}^{O_\tau}$ is the local yaw angle, ${}^{O_k}x_{O_\tau}$ and ${}^{O_k}y_{O_\tau}$ are the 2D position of $\{O_\tau\}$ in the starting integration frame $\{O_k\}$. Note that this model reveals the fact that the 2D orientation evolves over the integration period. We then integrate these differential equations from t_k to t_{k+1} and obtain the 2D relative pose measurement as follows:

$$\mathbf{z}_{k+1} = \begin{bmatrix} \theta_{O_k}^{O_{k+1}} \\ {}^{O_k}\mathbf{d}_{O_{k+1}} \end{bmatrix} = \begin{bmatrix} \int_{t_k}^{t_{k+1}} {}^{O_t}\omega dt \\ \int_{t_k}^{t_{k+1}} {}^{O_t}v\ \cos(\theta_{O_k}^{O_t}) dt \\ \int_{t_k}^{t_{k+1}} {}^{O_t}v\ \sin(\theta_{O_k}^{O_t}) dt \end{bmatrix} \tag{15}$$

$$=: \mathbf{g}(\omega_{l(k:k+1)}, \omega_{r(k:k+1)}, \mathbf{x}_{WI}) \tag{16}$$

where $\omega_{(k:k+1)}$ denote all the wheel measurements integrated t_k to t_{k+1}. If both extrinsic and time offset (spatiotemporal) calibration parameters between the odometer and IMU/camera are perfectly known, the above integrated odometry measurements can be readily used in the MSCKF update as in [2]. However, in practice, this often is not the case, for example, due to inaccurate prior calibration or mechanical vibration. To cope with possible time-varying calibration parameters during terrain navigation, the proposed VIWO performs online calibration of the wheel-encoders' intrinsics $\mathbf{x}_{WI}$, and the extrinsics $\mathbf{x}_{WE} = [{}^{O}_{I}\bar{q}^\top\ {}^{O}\mathbf{p}_I^\top]^\top$ and time offset ${}^{O}t_I$ between the odometer and IMU. Note again that the IMU and camera are assumed to be calibrated and synchronized for presentation brevity. To this end, we augment the state vector (1) with these parameters:

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{x}_{I_k}^\top & \mathbf{x}_{C_k}^\top & \mathbf{x}_{WE}^\top & {}^{O}t_I & \mathbf{x}_{WI}^\top \end{bmatrix}^\top \tag{17}$$

In what follows, we will derive in detail the relation between the preintegrated wheel odometry measurements (16) and the augmented state (17) by properly taking into account the intrinsic/extrinsic calibration parameters:

$$\mathbf{z}_{k+1} = \mathbf{h}(\mathbf{x}_{I_{k+1}}, \mathbf{x}_{C_{k+1}}, \mathbf{x}_{WE}, {}^{O}t_I, \mathbf{x}_{WI}) \tag{18}$$

B. Odometry Measurement wrt. Intrinsics

As evident from (11) and (16), the wheel-odometry integration entangles the intrinsics $\mathbf{x}_{WI}$, and ideally we should

re-integrate these measurements whenever a new estimate of intrinsics is available, which however negates the computational efficiency of preintegration. To address this issue, we linearize the preintegrated odometry measurements about the current estimate of the intrinsics while properly taking into account the measurement uncertainty due to the linearization errors of the intrinsics and the noise [see (16)]:

$$\mathbf{z}_{k+1} \simeq \mathbf{g}(\omega_{ml(k:k+1)}, \omega_{mr(k:k+1)}, \hat{\mathbf{x}}_{WI}) + \frac{\partial \mathbf{g}}{\partial \tilde{\mathbf{x}}_{WI}} \tilde{\mathbf{x}}_{WI} + \frac{\partial \mathbf{g}}{\partial \mathbf{n}_w} \mathbf{n}_w \tag{19}$$

where $\mathbf{n}_\omega$ is the stacked noise vector whose τ-th block is corresponding to the encoder measurement noise at $t_\tau \in [t_k, t_{k+1}]$ (i.e., $[n_{\omega_{l,\tau}} \; n_{\omega_{r,\tau}}]^\top$) [see (10)].

Clearly, performing EKF update with this measurement requires the Jacobians with respect to both the intrinsics and the noise in (19). It is important to note that as the preintegration of $\mathbf{g}(\cdot)$ is computed incrementally using the encoders' measurements in the interval $[t_k, t_{k+1}]$, we accordingly calculate the measurement Jacobians incrementally one step at a time. Note also that since the noise Jacobian and $\mathbf{n}_\omega$ are often of high dimensions and may be computationally expensive when computing the stacked noise covariance during the update, we instead compute the noise covariance by performing small matrix operations at each step.

Specifically, linearization of (15) at time t_τ yields the following recursive equations:

$$\tilde{\theta}_{O_k}^{O_{\tau+1}} = \tilde{\theta}_{O_k}^{O_\tau} + \mathbf{H}_{1,\tau} \tilde{\mathbf{x}}_{WI} + \mathbf{H}_{2,\tau} \mathbf{n}_{\omega,\tau} \tag{20}$$

$${}^{O_k}\tilde{x}_{O_{\tau+1}} = {}^{O_k}\tilde{x}_{O_\tau} + \mathbf{H}_{3,\tau} \tilde{\theta}_{O_k}^{O_\tau} + \mathbf{H}_{4,\tau} \tilde{\mathbf{x}}_{WI} + \mathbf{H}_{5,\tau} \mathbf{n}_{\omega,\tau} \tag{21}$$

$${}^{O_k}\tilde{y}_{O_{\tau+1}} = {}^{O_k}\tilde{y}_{O_\tau} + \mathbf{H}_{6,\tau} \tilde{\theta}_{O_k}^{O_\tau} + \mathbf{H}_{7,\tau} \tilde{\mathbf{x}}_{WI} + \mathbf{H}_{8,\tau} \mathbf{n}_{\omega,\tau} \tag{22}$$

where the intermediate Jacobians, $\mathbf{H}_{l,\tau}, l = 1 \ldots 8$, can be found in our companion technical report [41]. With the above equations, we can recursively compute the noise covariance $\mathbf{P}_{m,\tau+1}$ and the Jacobian $\frac{\partial \mathbf{g}_{\tau+1}}{\partial \tilde{\mathbf{x}}_{WI}}$ as follows:

$$\mathbf{\Phi}_{tr,\tau} = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{H}_{3,\tau} & 1 & 0 \\ \mathbf{H}_{6,\tau} & 0 & 1 \end{bmatrix}, \mathbf{\Phi}_{WI,\tau} = \begin{bmatrix} \mathbf{H}_{1,\tau} \\ \mathbf{H}_{4,\tau} \\ \mathbf{H}_{7,\tau} \end{bmatrix}, \mathbf{\Phi}_{n,\tau} = \begin{bmatrix} \mathbf{H}_{2,\tau} \\ \mathbf{H}_{5,\tau} \\ \mathbf{H}_{8,\tau} \end{bmatrix} \tag{23}$$

$$\mathbf{P}_{m,\tau+1} = \mathbf{\Phi}_{tr,\tau} \mathbf{P}_{m,\tau} \mathbf{\Phi}_{tr,\tau}^\top + \mathbf{\Phi}_{n,\tau} \mathbf{Q}_\tau \mathbf{\Phi}_{n,\tau}^\top \tag{24}$$

$$\frac{\partial \mathbf{g}_{\tau+1}}{\partial \tilde{\mathbf{x}}_{WI}} = \mathbf{\Phi}_{tr,\tau} \frac{\partial \mathbf{g}_\tau}{\partial \tilde{\mathbf{x}}_{WI}} + \mathbf{\Phi}_{WI,\tau} \tag{25}$$

where $\mathbf{Q}_\tau$ is the noise covariance of wheel encoder measurement at t_τ. These equations show how the Jacobian and the noise covariance evolve during the preintegration interval. We thus can recursively compute measurement noise covariance $\mathbf{P}_m$ and the Jacobian matrix $\frac{\partial \mathbf{g}}{\partial \tilde{\mathbf{x}}_{WI}}$ at the end of preintegration t_{k+1}, based on the zero initial condition (i.e., $\mathbf{P}_{m,0}$, $\frac{\partial \mathbf{g}_0}{\partial \tilde{\mathbf{x}}_{WI}} = \mathbf{0}_3$).

C. Odometry Measurement wrt. Extrinsics

1) Spatial calibration: Note that the preintegrated wheel measurement (16) provides *only* the 2D relative motion on the odometer's plane, while the VIWO state vector (17) contains the 3D IMU/camera poses. In order to establish the connection of the preintegrated odometry with the state, clearly the relative transformation (extrinsic calibration) between the IMU and the odometer is required [see (15)]:

$$\begin{bmatrix} \tilde{\theta}_{O_k}^{O_{k+1}} \\ {}^{O_k}\tilde{\mathbf{d}}_{O_{k+1}} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_3^\top \mathrm{Log}({}_I^O\mathbf{R} \, {}_G^{I_{k+1}}\mathbf{R} \, {}_G^{I_k}\mathbf{R}^\top \, {}_I^O\mathbf{R}^\top) \\ \Lambda_I^O \mathbf{R} \, {}_G^{I_k}\mathbf{R} ({}^G\mathbf{p}_{I_{k+1}} + {}_{I_{k+1}}^G\mathbf{R}^I \mathbf{p}_O - {}^G\mathbf{p}_{I_k}) + {}^O\mathbf{p}_I \end{bmatrix} \tag{26}$$

where $\Lambda = [\mathbf{e}_1 \; \mathbf{e}_2]^\top$, $\mathbf{e}_i$ is the i-th standard unit basis vector, and $\mathrm{Log}(\cdot)$ is the $SO(3)$ matrix logarithm function [42]. As this measurement depends on the two consecutive poses as well as the odometer/IMU extrinsics, when updating with it in the MSCKF, the corresponding measurement Jacobians are needed and computed as follows [see (26) and (18)]:

$$\frac{\partial \mathbf{h}}{\partial \tilde{\mathbf{x}}_{I_{k+1}}} = \begin{bmatrix} \mathbf{e}_3^\top \, {}_I^O\hat{\mathbf{R}} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 9} \\ -\Lambda_I^O \hat{\mathbf{R}} \, {}_{I_{k+1}}^{I_k}\hat{\mathbf{R}} \lfloor {}^I\hat{\mathbf{p}}_O \rfloor & \Lambda_I^O \hat{\mathbf{R}} \, {}_G^{I_k}\hat{\mathbf{R}} & \mathbf{0}_{2\times 9} \end{bmatrix} \tag{27}$$

$$\frac{\partial \mathbf{h}}{\partial \tilde{\mathbf{x}}_{C_{k+1}}} = \begin{bmatrix} -\mathbf{e}_3^\top \, {}_I^O\hat{\mathbf{R}} \, {}_{I_k}^{I_{k+1}}\hat{\mathbf{R}} \, \hat{\mathbf{R}} & \mathbf{0}_{1\times 3} \\ \Lambda_I^O \hat{\mathbf{R}} \lfloor {}_G^{I_k}\hat{\mathbf{R}} ({}^G\hat{\mathbf{p}}_{O_{k+1}} - {}^G\hat{\mathbf{p}}_{I_k}) \rfloor & -\Lambda_I^O \hat{\mathbf{R}} \, {}_G^{I_k}\hat{\mathbf{R}} \, \hat{\mathbf{R}} \end{bmatrix} \tag{28}$$

$$\frac{\partial \mathbf{h}}{\partial \tilde{\mathbf{x}}_{WE}} = \begin{bmatrix} \mathbf{e}_3^\top (\mathbf{I} - {}_{O_k}^{O_{k+1}}\hat{\mathbf{R}}) & \mathbf{0}_{1\times 3} \\ \Lambda(\lfloor {}_I^O\hat{\mathbf{R}} \, {}^{I_k}\hat{\mathbf{p}}_{O_{k+1}} \rfloor + {}_{O_{k+1}}^{O_k}\hat{\mathbf{R}} \lfloor {}^O\hat{\mathbf{p}}_I \rfloor) & \Lambda(\mathbf{I} - {}_{O_{k+1}}^{O_k}\hat{\mathbf{R}}) \end{bmatrix} \tag{29}$$

where $\lfloor \cdot \rfloor$ is the skew symmetric matrix.

2) Temporal calibration: To account for the difference between sensor clocks and measurement delay, we model an unknown constant time offset between the IMU clock and the odometer clock:[2] ${}^I t_k = {}^O t_k + {}^O t_I$, where ${}^I t_k$ and ${}^O t_k$ are the times when measurement $\mathbf{z}_k$ was collected in the IMU and odometer's clocks, and ${}^O t_I$ is the time offset between the two time references. Consider that we want to derive preintegrated odometry constraints between two cloned states at the true IMU times ${}^I t_k$ and ${}^I t_{k+1}$. Using the current best estimate of the time offset ${}^O\hat{t}_I$, we can integrate our wheel encoder measurements between the odometer times ${}^O t_k = {}^I t_k - {}^O\hat{t}_I$ and ${}^O t_{k+1} = {}^I t_{k+1} - {}^O\hat{t}_I$, whose corresponding times in the IMU clock are:

$${}^I t'_k := {}^I t_k - {}^O\hat{t}_I + {}^O t_I \quad = {}^I t_k + {}^O\tilde{t}_I \tag{30}$$

$${}^I t'_{k+1} := {}^I t_{k+1} - {}^O\hat{t}_I + {}^O t_I = {}^I t_{k+1} + {}^O\tilde{t}_I \tag{31}$$

After preintegration we have the 2D relative pose measurement between the times ${}^I t'_k$ and ${}^I t'_{k+1}$ while the corresponding states are at the times ${}^I t_k$ and ${}^I t_{k+1}$. To update with this measurement, we employ the following first-order approximation by accounting the time-offset estimation error:

$${}_G^{I({}^I t'_k)}\mathbf{R} \approx (\mathbf{I} - \lfloor {}^{I(t_k)}\boldsymbol{\omega} \, {}^O\tilde{t}_I \rfloor) \, {}_G^{I(t_k)}\mathbf{R} \tag{32}$$

$${}^G\mathbf{p}_{I({}^I t'_k)} \approx {}^G\mathbf{p}_{I(t_k)} + {}^G\mathbf{v}_{I(t_k)} \, {}^O\tilde{t}_I \tag{33}$$

Using this relationship, we can compute the measurement Jacobian of the time offset as follows [see (26)]:

$$\frac{\partial \mathbf{h}}{\partial {}^O\tilde{t}_I} = \begin{bmatrix} \mathbf{e}_3^\top \, {}_I^O\hat{\mathbf{R}} ({}^{I_{k+1}}\boldsymbol{\omega} - {}_{I_k}^{I_{k+1}}\hat{\mathbf{R}} \, {}^{I_k}\boldsymbol{\omega}) \\ \Lambda_I^O \hat{\mathbf{R}} (\lfloor {}^{I_k}\hat{\mathbf{p}}_{O_{k+1}} \rfloor \, {}^{I_k}\boldsymbol{\omega} - {}_{I_{k+1}}^{I_k}\hat{\mathbf{R}} \lfloor {}^I\hat{\mathbf{p}}_O \rfloor \, {}^{I_{k+1}}\boldsymbol{\omega} + {}^{I_k}\hat{\mathbf{v}}_{I_{k+1}}) \end{bmatrix} \tag{34}$$

Note that in our experiments we use the IMU angular rate measurement and the current estimate of velocity in computing the above Jacobian.

D. Odometry Measurement Update

At this point, we have obtained the preintegrated wheel odometry measurements along with their corresponding Jacobians which are readily used for the MSCKF update:

$$\begin{aligned} \tilde{\mathbf{z}}_{k+1} &:= \mathbf{g}(\omega_{ml(k:k+1)}, \omega_{mr(k:k+1)}, \hat{\mathbf{x}}_{WI}) - \mathbf{h}(\hat{\mathbf{x}}_I, \hat{\mathbf{x}}_C, \hat{\mathbf{x}}_{WE}, {}^O\hat{t}_I) \\ &\approx \underbrace{\begin{bmatrix} \frac{\partial \mathbf{h}}{\partial \tilde{\mathbf{x}}_I} & \frac{\partial \mathbf{h}}{\partial \tilde{\mathbf{x}}_C} & \frac{\partial \mathbf{h}}{\partial \tilde{\mathbf{x}}_{WE}} & -\frac{\partial \mathbf{g}}{\partial \tilde{\mathbf{x}}_{WI}} & \frac{\partial \mathbf{h}}{\partial {}^O\tilde{t}_I} \end{bmatrix}}_{\mathbf{H}_{k+1}} \tilde{\mathbf{x}}_{k+1} - \frac{\partial \mathbf{g}}{\partial \mathbf{n}_\omega} \mathbf{n}_\omega \end{aligned} \tag{35}$$

[2] We assume that the two wheel encoders are hardware synchronized and thus their readings have the same timestamps.

Note that similar to how we treat visual features, we also employ the Mahalanobis distance test to reject bad preintegrated odometry measurements (which can be due to some unmodelled errors such as slippage) and only those passing the χ^2 test will be used for EKF update. To summarize, the main steps of the odometry measurement update are outlined in Algorithm 1.

V. Observability Analysis

As system observability plays an important role for state estimation [31], [32], we perform the observability analysis to gain insights about the state/parameter identifiability for the proposed VIWO. For brevity, in the following we only present the key results of our analysis while the detailed analysis can be found in the companion technical report [41].

For the analysis purpose, in analogy to [43], we consider the following state vector which includes a single cloned pose of $\mathbf{x}_{I_{k-1}}$ and a single 3D point feature ${}^G\mathbf{p}_f$:

$$\mathbf{x}_k = \begin{bmatrix}\mathbf{x}_{I_k}^\top & \mathbf{x}_{C_k}^\top & \mathbf{x}_{etc}^\top\end{bmatrix}^\top,\ \mathbf{x}_{etc} = \begin{bmatrix}\mathbf{x}_{WE}^\top & {}^O t_I & \mathbf{x}_{WI}^\top & {}^G\mathbf{p}_f^\top\end{bmatrix}^\top \quad (36)$$

The observability matrix for the linearized system is:

$$\mathbf{M} = \begin{bmatrix}\mathbf{H}_0^\top & (\mathbf{H}_1\mathbf{\Phi}_{(1,0)})^\top & \dots & (\mathbf{H}_k\mathbf{\Phi}_{(k,0)})^\top\end{bmatrix}^\top \quad (37)$$

where $\mathbf{\Phi}_{(k,0)}$ is the state transition matrix which is not obvious when including the clone in the state vector and will be derived below, and $\mathbf{H}_k$ is the measurement Jacobian at time step k (e.g., see (35)). If we can find matrix $\mathbf{N}$ that satisfies $\mathbf{MN} = \mathbf{0}$, the basis of $\mathbf{N}$ indicate the unobservable directions of the linearized system.

A. State Transition Matrix

In the MSCKF-based linearized system, the state transition matrix corresponding to the cloned state essentially reveals the stochastic cloning process. To see this, first recall how the cloned states are processed [44]: (i) augment the state with the current IMU pose when a new image is available, (ii) propagate the cloned pose with zero dynamics, (iii) marginalize the oldest clone after update if reaching the maximum size of the sliding window. In the case of one clone, this cloning corresponds to the following operation (while marginalization in the covariance form is trivial):

$$\hat{\mathbf{x}}_{k|k} \leftarrow \begin{bmatrix}\mathbf{I}_6 & \mathbf{0}_{6\times9} & \mathbf{0}_6 & \mathbf{0}_{6\times13}\\ \mathbf{0}_{9\times6} & \mathbf{I}_9 & \mathbf{0}_{9\times6} & \mathbf{0}_{9\times13}\\ \mathbf{I}_6 & \mathbf{0}_{6\times9} & \mathbf{0}_6 & \mathbf{0}_{6\times13}\\ \mathbf{0}_{13\times6} & \mathbf{0}_{13\times9} & \mathbf{0}_{13\times6} & \mathbf{I}_{13}\end{bmatrix}\hat{\mathbf{x}}_{k|k} \quad (38)$$

One may model this state transition by including all the clones (with zero dynamics) in the state vector. This may complicate the analysis as it needs to explicitly include all the clone constraints, while we here construct the state transition matrix with implicit clone constraints.

We discover that cloning and propagating the current error state $\tilde{\mathbf{x}}_k$ can be unified by the following linear mapping:

$$\tilde{\mathbf{x}}_{k+1} = \underbrace{\begin{bmatrix}\mathbf{\Phi}_{I_{11}(k+1,k)} & \mathbf{\Phi}_{I_{12}(k+1,k)} & \mathbf{0}_6 & \mathbf{0}_{6\times13}\\ \mathbf{\Phi}_{I_{21}(k+1,k)} & \mathbf{\Phi}_{I_{22}(k+1,k)} & \mathbf{0}_{9\times6} & \mathbf{0}_{9\times13}\\ \mathbf{I}_6 & \mathbf{0}_{6\times9} & \mathbf{0}_6 & \mathbf{0}_{6\times13}\\ \mathbf{0}_{13\times6} & \mathbf{0}_{13\times9} & \mathbf{0}_{13\times6} & \mathbf{I}_{13}\end{bmatrix}}_{\mathbf{\Xi}_{(k+1,k)}}\tilde{\mathbf{x}}_k \quad (39)$$

where $\mathbf{\Phi}_I = \begin{bmatrix}\mathbf{\Phi}_{I_{11}} & \mathbf{\Phi}_{I_{12}}\\ \mathbf{\Phi}_{I_{21}} & \mathbf{\Phi}_{I_{22}}\end{bmatrix}$ is the IMU state transition matrix (see [43]). Note that $\mathbf{I}_6$ in the third block row copies the IMU pose of $\tilde{\mathbf{x}}_k$ into $\tilde{\mathbf{x}}_{k+1}$ as a cloned pose without changing its value, while the cloned state in $\tilde{\mathbf{x}}_k$ has been discarded (marginalized). The above operation clearly reveals the MSCKF cloning process and thus, we will leverage this linear system (39) for the ensuing analysis.

Specifically, during the time interval $[t_0, t_{k+1}]$, we have the following linear dynamic system:

$$\tilde{\mathbf{x}}_{k+1} = \underbrace{\mathbf{\Xi}(t_{k+1},t_k)\mathbf{\Xi}(t_k,t_{k-1})\cdots\mathbf{\Xi}(t_1,t_0)}_{\mathbf{\Xi}(t_{k+1},t_0)}\tilde{\mathbf{x}}_0 \quad (40)$$

$$\mathbf{\Xi}_{(k+1,0)} = \begin{bmatrix}\mathbf{\Phi}_{I_{11}(k+1,0)} & \mathbf{\Phi}_{I_{12}(k+1,0)} & \mathbf{0}_6 & \mathbf{0}_{6\times13}\\ \mathbf{\Phi}_{I_{21}(k+1,0)} & \mathbf{\Phi}_{I_{22}(k+1,0)} & \mathbf{0}_{9\times6} & \mathbf{0}_{9\times13}\\ \mathbf{\Psi}_{(k+1,0)} & \mathbf{0}_{6\times9} & \mathbf{0}_6 & \mathbf{0}_{6\times13}\\ \mathbf{0}_{13\times6} & \mathbf{0}_{13\times9} & \mathbf{0}_{13\times6} & \mathbf{I}_{13}\end{bmatrix} \quad (41)$$

We also enforce the constraint that the initial IMU pose and the clone state at time t_0 are identical, whose error states have the following constraint:

$$\begin{bmatrix}\mathbf{I}_6 & \mathbf{0}_{6\times9}\end{bmatrix}\tilde{\mathbf{x}}_{I_0} = \tilde{\mathbf{x}}_{C_0} \quad (42)$$

With (40) and (42), we finally have the following state transition matrix $\mathbf{\Phi}_{(k+1,0)}$ for our observability analysis:

$$\begin{bmatrix}\tilde{\mathbf{x}}_{I_{k+1}}\\ \tilde{\mathbf{x}}_{C_{k+1}}\\ \tilde{\mathbf{x}}_{etc}\end{bmatrix} = \underbrace{\begin{bmatrix}\mathbf{\Phi}_{I(k+1,0)} & \mathbf{0}_{15\times6} & \mathbf{0}_{15\times13}\\ \mathbf{0}_{6\times9} & \mathbf{\Psi}_{(k+1,0)} & \mathbf{0}_{6\times13}\\ \mathbf{0}_{13\times15} & \mathbf{0}_{13\times6} & \mathbf{I}_{13}\end{bmatrix}}_{\mathbf{\Phi}_{(k+1,0)}}\begin{bmatrix}\tilde{\mathbf{x}}_{I_0}\\ \tilde{\mathbf{x}}_{C_0}\\ \tilde{\mathbf{x}}_{etc}\end{bmatrix} \quad (43)$$

B. Observability Properties

Based on the measurement Jacobians and state transition matrix [see (9), (35) and (43)], we are able to construct the observability matrix (37) of the MSCKF-based linearized system under consideration. Careful inspection of this matrix reveals the following results (see [41] for details):

Lemma 5.1: The proposed VIWO has the following observability properties:

- With general motions, there are four unobservable directions corresponding to the global position and the yaw angle as in VINS [43].
- As matrix blocks of the observability matrix that are related to the wheel-encoder's extrinsic and intrinsic calibration parameters highly depend on dynamic states and encoder's readings, these blocks can be full-rank given sufficient motions, implying that these parameters are observable, which has been validated in simulations in terms of convergence (see Fig. 3).
- We identify the following five degenerate motions that cause the odometer's calibration parameters to become unobservable, which might be commonly seen for ground vehicles:

Motion	**Unobservable**
Pure translation	${}^O\mathbf{p}_I$, b
1-axis rotation	${}^O\mathbf{p}_I$
Constant angular and linear velocity	${}^O t_I$
No left/right wheel velocity	r_l / r_r
No motion	${}^O_I\mathbf{R}$, ${}^O\mathbf{p}_I$, r_l, r_r, b

VI. Simulation Validations

The proposed VIWO was implemented within the *Open VINS* [45] framework which provides both simulation and evaluation utilities. We expanded the visual-inertial simulator

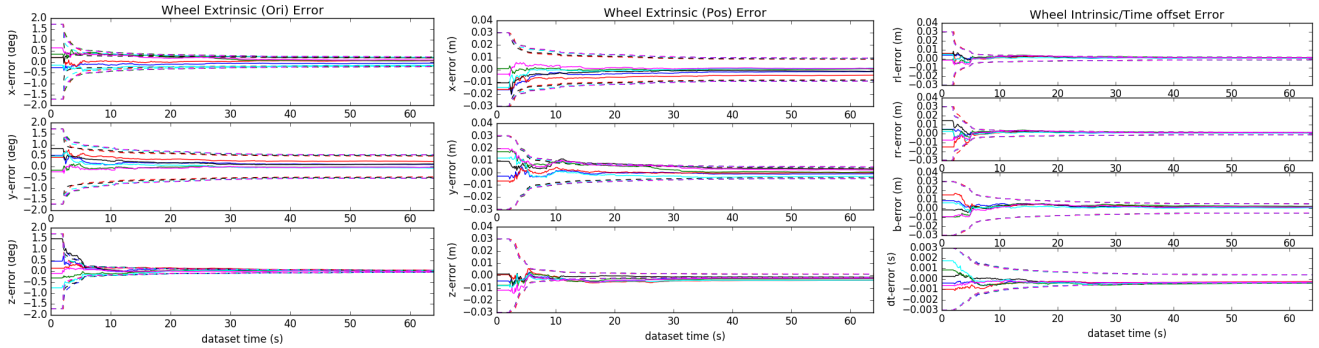

Fig. 3: Calibration errors of each parameter (solid) and 3σ bound (dotted) for six different runs under random motion. Each colors denote runs with different realization of the measurement noise and the initial values. Both the estimation errors and 3σ bounds can converge in about 10 seconds.

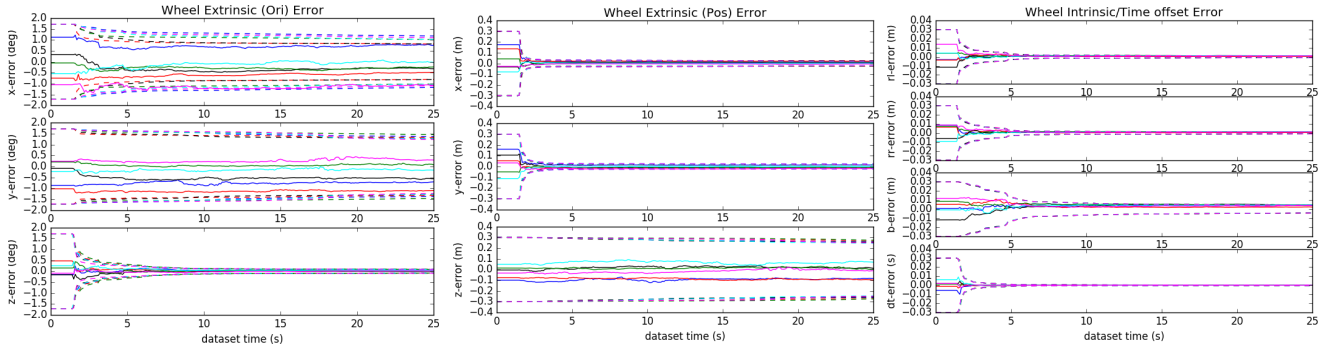

Fig. 4: Calibration error results of each parameter under planar motion. Each colors denote runs with different realization of the measurement noise and the initial values. Both the estimation errors (solid) and 3σ (dotted) bounds are reported.

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

Parameter	Value	Parameter	Value
Cam Freq. (hz)	10	IMU Freq. (hz)	200
Wheel Freq. (hz)	50	Num. Clones	15
Max. Feats	200	Feat. Rep.	GLOBAL
Pixel Proj. (px)	1	Wheel. White Noise	1.0e-03
Gyro. White Noise	1.0e-04	Gyro. Rand. Walk	1.0e-04
Accel. White Noise	1.0e-04	Accel. Rand. Walk	1.0e-04
Wheel Ext (Ori). Ptrb.	1.0e-02	Wheel Ext (Pos). Ptrb.	1.0e-01
Wheel Int. Ptrb.	1.0e-02	Wheel Toff. Ptrb.	1.0e-02

TABLE II: Relaive pose error (RPE) of each algorithm (degree/meter).

	50m	100m	200m	NEES
VIO	0.362 / 1.252	0.494 / 2.245	0.657 / 3.930	3.921 / 3.895
true w. cal	0.277 / 0.550	0.365 / 0.908	0.479 / 1.573	1.952 / 2.020
true wo. cal	0.259 / 0.384	0.340 / 0.622	0.443 / 1.125	1.698 / 1.473
bad w. cal	0.276 / 0.543	0.365 / 0.888	0.486 / 1.526	1.943 / 1.826
bad wo. cal	0.572 / 0.510	1.104 / 1.142	2.239 / 3.367	59.678 / 183.538

to additionally simulate a differential drive robot which can only have velocity in the local x-direction (non-holonomic constraint [5]) and have listed the key simulation parameters in Table I. In order to validate the proposed calibration method, we tested VIWO on a large-scale neighborhood simulation dataset with four different combinations: with "bad" or "true" initial calibration parameters and with or without online calibration. The bad initial values were drawn from the prior distributions specified in Table I. 50m, 100m, and 200m relative pose error (RPE, [46]) and the average normalized estimation error squared (NEES) results of the four configurations and standard VIO for comparison are shown in Table II. VIWO with "bad" initial values with calibration showed similar RPE results compare to those had "true" initial values, while VIWO with "bad" initial

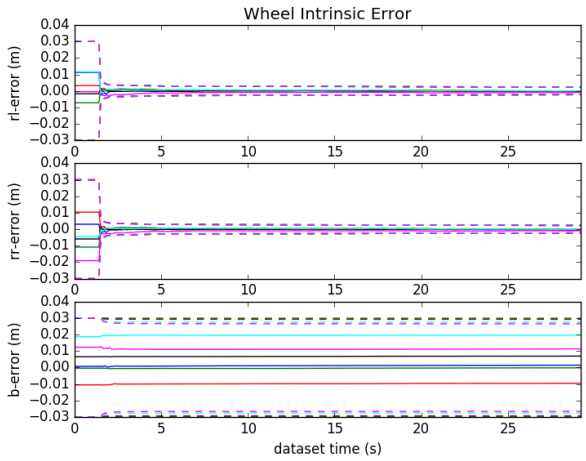

Fig. 5: Intrinsic calibration error under straight line motion

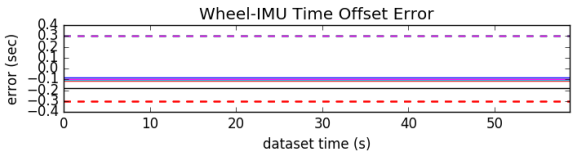

Fig. 6: Time offset calibration error under straight line motion

values without calibration has a very poor estimation performance and a large NEES due to treating the incorrect calibration as being true. Fig. 1 shows the trajectory of the algorithms, and clearly shows the failure of estimation when not performing calibration. We found that the system was particularly sensitive to the wheel intrinsics and even with small perturbations the system without online calibration provided unusable estimation results.

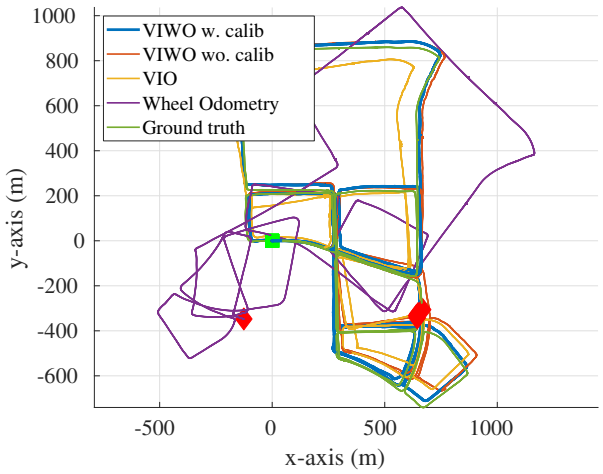

Fig. 7: $urban39$ results of VIWO with calibration (blue), VIWO without calibration (red), VIO (yellow) and wheel odomtry (purple). The green square and red diamond correspond to the start and end of the 11.06 km trajectory, respectively.

A. Extrinsic, Intrinsic, and Time offset Calibration

To validate that all calibration parameters are able to converge for random motion, we simulated a trajectory which excited all-axes motions without invalidating non-holonomic constraint. Shown in Fig. 3, we performed online extrinsic, intrinsic, and time offset calibration six times with different initial perturbations. We additionally plot the 3σ bounds, which should bound the error in the case that our estimator is consistent. It is clear that all calibration parameters are able to quickly converge to near their true values and remain within 3σ bounds.

B. Degenerate Motion Validation

One of the degenerate motions that are of particular interest to grounded vehicles is planar motion. Fig. 4 shows the calibration results for the starting segment of the planar trajectory shown in Fig. 1. The extrinsic z-axis does not converge since the trajectory is in the global x-y plane and thus the robot only rotates around its z-axis causing this extrinsic to become unobservable. Note that the roll and pitch portion of the orientation (left top and left middle of Fig. 4) has a slower rate of convergence compared to other parameters because planar motion does not provide good geometrical constraints.

A straight line motion with constant linear velocity and no rotation, which is a special case of both pure translation and constant angular/linear velocities, is also simulated. To prevent the unobservable parameters from affecting other calibration, we only estimate either the wheel intrinsics or time offset and show their estimation results. Shown in Fig. 5, it is clear that the baselink is unable to converge, while the wheel radius can still converge due to having scale from the IMU. Fig. 6 shows the time offset is unable to be calibrate in this straight line case.

VII. Real World Experiments

We further evaluate the proposed VIWO in a real-world urban dataset $urban39$ [9] which is collected in urban area with 11.06 km long trajectory, and used the stereo camera, IMU, and wheel encoder measurements for the estimation. We compared VIWO with calibration (VIWO w. cal), VIWO without calibration (VIWO wo. cal), and VIO to evaluate the performance. Additionally, we also included the pure wheel odometry trajectory, to deliver the sense of wheel measurements errors. We used 25 clones and 200 features for the real-time estimation along with online IMU-camera intrinsic and spatiotemporal extrinsic calibration, and the resulting trajectories of each algorithm are shown in Fig. 7. The dataset provides the ground truth which is computed from graph SLAM. The root mean squared error (RMSE) of orientation and position of each algorithm were: VIWO w. cal 1.713 / 42.748 , VIWO wo. cal 2.874 / 52.657, VIO 5.276 / 66.316, and Wheel Odometry 50.439 / 481.093 (degrees / meters). The initial values of parameters and the final calibrated values are listed in Table. III. We found the time offset calibration quickly converged to a nontrivial value of -0.0273 seconds which given the 6.0 m/s average vehicle velocity equates to 0.163 meters error if not properly calibrated, and thus validates the need for online estimation of this parameter. The rest of the calibration parameters also showed convergence to values that are different from initially given by small amount, but such change can have large effect in long-term estimation as shown in Fig. 7.

TABLE III: Values of calibration parameters before/after calibration

Parameter	Before	After
left wheel radius	0.311740	0.312262
right wheel radius	0.311403	0.311843
base length	1.52439	1.53201
Ext. Pos	[-0.070, 0.000, 1.400]	[-0.062, 0.003, 1.384]
Ext. Ori	[0.000, 0.000, 0.000]	[0.000, 0.001, -0.002]
Time offset	0.00000	-0.02723

VIII. Conclusions and Future Work

In this paper, we have developed an efficient and consistent MSCKF-based visual-inertial-wheel odometry (VIWO) system that fuses multi-modal wheel-encoder, IMU and camera measurements. In particular, to compensate for possible time-varying calibration and for improving estimation, the proposed VIWO performs online sensor calibration of the spatiotempral extrinsics of odometer-IMU/camera as well as the wheel encoder's intrinsics. To better understand the VIWO estimatability, we have conducted in-depth observability anaylsis for the MSCKF-based linearized system by leveraging a new state transition matrix that models the stochastic cloning in the MSCKF, and have identified five degenerate motions that may cause online calibration unobservable. In the future, we will extend the proposed VIWO to characterize the sensor uncertainty online instead of using the prior.

References

[1] G. Huang, "Visual-inertial navigation: A concise review," in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.

[2] K. J. Wu, C. X. Guo, G. Georgiou, and S. I. Roumeliotis, "Vins on wheels," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 5155–5162.

[3] X. Zuo, M. Zhang, Y. Chen, Y. Liu, G. Huang, and M. Li, "Visual-inertial localization for skid-steering robots with kinematic constraints," *arXiv preprint arXiv:1911.05787*, 2019.

[4] Y. Yang, P. Geneva, K. Eckenhoff, and G. Huang, "Degenerate motion analysis for aided INS with online spatial and temporal calibration," *IEEE Robotics and Automation Letters (RA-L)*, vol. 4, no. 2, pp. 2070–2077, 2019.

[5] R. Siegwart, I. R. Nourbakhsh, and D. Scaramuzza, *Introduction to autonomous mobile robots*. MIT press, 2011.

[6] W. Lee, H. Cho, S. Hyeong, and W. Chung, "Practical modeling of gnss for autonomous vehicles in urban environments," *Sensors*, vol. 19, no. 19, p. 4236, 2019.

[7] J. Borenstein and L. Feng, "Measurement and correction of systematic odometry errors in mobile robots," *IEEE Transactions on robotics and automation*, vol. 12, no. 6, pp. 869–880, 1996.

[8] Y. Maddahi, N. Sepehri, A. Maddahi, and M. Abdolmohammadi, "Calibration of wheeled mobile robots with differential drive mechanisms: An experimental approach," *Robotica*, vol. 30, no. 6, pp. 1029–1039, 2012.

[9] J. Jeong, Y. Cho, Y.-S. Shin, H. Roh, and A. Kim, "Complex urban dataset with multi-level sensors from highly diverse urban environments," *The International Journal of Robotics Research*, vol. 38, no. 6, pp. 642–657, 2019.

[10] R. Kümmerle, G. Grisetti, and W. Burgard, "Simultaneous parameter calibration, localization, and mapping," *Advanced Robotics*, vol. 26, no. 17, pp. 2021–2041, 2012.

[11] A. Martinelli, "Local decomposition and observability properties for automatic calibration in mobile robotics," in *2009 IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 4182–4188.

[12] N. Seegmiller, F. Rogers-Marcovitz, G. Miller, and A. Kelly, "Vehicle model identification by integrated prediction error minimization," *The International Journal of Robotics Research*, vol. 32, no. 8, pp. 912–931, 2013.

[13] A. Kelly, "Fast and easy systematic and stochastic odometry calibration," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, vol. 4. IEEE, 2004, pp. 3188–3194.

[14] J. Deray, J. Solà, and J. Andrade-Cetto, "Joint on-manifold self-calibration of odometry model and sensor extrinsics using pre-integration," in *2019 European Conference on Mobile Robots (ECMR)*. IEEE, 2019, pp. 1–6.

[15] J. L. Martinez, A. Mandow, J. Morales, S. Pedraza, and A. Garcia-Cerezo, "Approximating kinematics for tracked mobile robots," *The International Journal of Robotics Research*, vol. 24, no. 10, pp. 867–878, 2005.

[16] G. Antonelli, F. Caccavale, F. Grossi, and A. Marino, "Simultaneous calibration of odometry and camera for a differential drive mobile robot," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 5417–5422.

[17] L. Heng, B. Li, and M. Pollefeys, "Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 1793–1800.

[18] D. A. Cucci and M. Matteucci, "Position tracking and sensors self-calibration in autonomous mobile robots by gauss-newton optimization," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 1269–1275.

[19] D. Li, K. Eckenhoff, K. Wu, Y. Wang, R. Xiong, and G. Huang, "Gyro-aided camera-odometer online calibration and localization," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 3579–3586.

[20] X. Wang, H. Chen, Y. Li, and H. Huang, "Online extrinsic parameter calibration for robotic camera–encoder system," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 8, pp. 4646–4655, 2019.

[21] W. Lee, K. Eckenhoff, P. Geneva, and G. Huang, "Intermittent gps-aided vio: Online initialization and calibration," in *Proc. of the IEEE International Conference on Robotics and Automation*, Paris, France, 2020.

[22] M. Li and A. I. Mourikis, "Online temporal calibration for camera–imu systems: Theory and algorithms," *The International Journal of Robotics Research*, vol. 33, no. 7, pp. 947–964, 2014.

[23] T. Qin and S. Shen, "Online temporal calibration for monocular visual-inertial systems," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 3662–3669.

[24] K. Eckenhoff, P. Geneva, and G. Huang, "Sensor-failure-resilient multi-imu visual-inertial navigation," in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.

[25] K. Eckenhoff, P. Geneva, J. Bloecker, and G. Huang, "Multi-camera visual-inertial navigation with online intrinsic and extrinsic calibration," in *Proc. International Conference on Robotics and Automation*, Montreal, Canada, May 2019.

[26] Y. Yang, B. P. W. Babu, C. Chen, G. Huang, and L. Ren, "Analytic combined imu integrator for visual-inertial navigation," in *Proc. of the IEEE International Conference on Robotics and Automation*, Paris, France, 2020.

[27] F. Tungadi and L. Kleeman, "Time synchronisation and calibration of odometry and range sensors for high-speed mobile robot mapping," in *Proc. Australasian Conference on Robotics and Automation*, 2008.

[28] R. Voges and B. Wagner, "Timestamp offset calibration for an imu-camera system under interval uncertainty," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 377–384.

[29] J. Rehder, R. Siegwart, and P. Furgale, "A general approach to spatiotemporal calibration in multisensor systems," *IEEE Transactions on Robotics*, vol. 32, no. 2, pp. 383–398, 2016.

[30] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 1280–1286.

[31] G. Huang, "Improving the consistency of nonlinear estimators: Analysis, algorithms, and applications," Ph.D. dissertation, Department of Computer Science and Engineering, University of Minnesota, 2012.

[32] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Observability-constrained vision-aided inertial navigation," *University of Minnesota, Dept. of Comp. Sci. & Eng., MARS Lab, Tech. Rep*, vol. 1, p. 6, 2012.

[33] Y. Yang and G. Huang, "Observability analysis of aided ins with heterogeneous features of points, lines and planes," *IEEE Transactions on Robotics*, vol. 35, no. 6, pp. 1399–1418, 2019.

[34] G. Huang, A. I. Mourikis, and S. I. Roumeliotis, "Observability-based rules for designing consistent EKF SLAM estimators," *International Journal of Robotics Research*, vol. 29, no. 5, pp. 502–528, Apr. 2010.

[35] A. Martinelli, "State estimation based on the concept of continuous symmetry and observability analysis: The case of calibration," *IEEE Transactions on Robotics*, vol. 27, no. 2, pp. 239–255, 2011.

[36] A. Censi, A. Franchi, L. Marchionni, and G. Oriolo, "Simultaneous calibration of odometry and sensor parameters for mobile robots," *IEEE Transactions on Robotics*, vol. 29, no. 2, pp. 475–492, 2013.

[37] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 3565–3572.

[38] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep., Mar. 2005.

[39] C. Hertzberg, R. Wagner, U. Frese, and L. Schröder, "Integrating generic sensor fusion algorithms with sound state representations through encapsulation of manifolds," *Information Fusion*, vol. 14, no. 1, pp. 57–77, 2013.

[40] M. Li, "Visual-inertial odometry on resource-constrained systems," Ph.D. dissertation, UC Riverside, 2014.

[41] W. Lee and G. Huang, "Visual inertial wheel odometry: Online calibration technical report," Robot Perception and Navigation Group (RPNG), University of Delaware, Tech. Rep., 2020. [Online]. Available: http://udel.edu/~ghuang/papers/tr_wheel-vio.pdf

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

[43] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Consistency analysis and improvement of vision-aided inertial navigation," *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 158–176, Feb 2014.

[44] Y. Yang and G. Huang, "Acoustic-inertial underwater navigation," in *Proc. of the IEEE International Conference on Robotics and Automation*, Singapore, May 29–Jun.3, 2017, pp. 4927–4933.

[45] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "Openvins: A research platform for visual-inertial estimation," in *Proc. of the IEEE International Conference on Robotics and Automation*, Paris, France, 2020. [Online]. Available: https://github.com/rpng/open_vins

[46] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 7244–7251.