Kinematics Based Visual Localization for Skid-Steering Robots: Algorithm and Theory

Xingxing Zuo\textsuperscript{1}, Mingming Zhang\textsuperscript{2}, Yiming Chen\textsuperscript{2}, Guoquan Huang\textsuperscript{3}, Yong Liu\textsuperscript{1,*}, and Mingyang Li\textsuperscript{2,*}

Abstract—In recent years, the volume of the commercial mobile robot market has been continuously growing due to the maturity of both hardware and software technology. To build commercial robots, skid-steering mechanical design is of increased popularity due to its manufacturing simplicity and unique hardware properties. However, this causes challenges on software and algorithm design, especially for localization (i.e., determining the robot’s rotation and position). While the general localization algorithms have been extensively studied in research communities, there are still fundamental problems that need to be resolved for localizing skid-steering robots. To tackle this problem, we propose a probabilistic sliding-window estimator dedicated to skid-steering robots, using measurements from a monocular camera, the wheel encoders, and optionally an inertial measurement unit (IMU). Specifically, we explicitly model the kinematics of skid-steering robots by both track instantaneous centers of rotation (ICRs) and correction factors, which are capable of compensating for the complexity of track-to-terrain interaction, the imperfection of mechanical design, terrain conditions and smoothness, and so on. These time and location varying kinematic parameters are estimated online along with other localization states in a tightly-coupled manner. More importantly, we conduct in-depth observability analysis for different sensors and design configurations in this paper, which provides us with theoretical tools in making the correct choice when building real commercial robots. In our experiments, we validate the proposed method by both simulation tests and real-world experiments, which demonstrate that our method outperforms competing methods by wide margins.

I. INTRODUCTION

In recent years, the robotic community has witnessed a growing ‘go-to-market’ trend, by not only building autonomous robots for scientific laboratory usage but also making commercial robots to create new business model and facilitate people’s daily lives. To date, a large amount of commercial outdoor robots, under either daily business usage or active trial operations and tests, are customized skid-steering robots (Newswire, 2019; Vincent, 2019; Rubin, 2019). Instead of having an explicit mechanism of steering control, skid-steering robots rely on adjusting the speed of the left and right tracks to turn around. The simplicity of the mechanical design and the property of being able to turn around with zero-radius make skid-steering robots widely used in both the scientific research community as well as the commercial robotic industry. However, the mechanical simplicity of skid-steering robots has significantly challenged the software and algorithm design in robotic artificial intelligence, especially in autonomous localization.

The localization system provides 6 degrees-of-freedom (DoF) motion estimates (3DoF rotational and 3DoF positional), which is a key component for enabling any autonomous robot. Among a large variety of localization algorithms, monocular visual odometry (VO) is a popular one due to the camera’s low cost, small size, and easy hardware setup (Davison et al., 2007; Forster et al., 2014; Engel et al., 2017). However, VO’s theoretical drawback of unable to uniquely determine the scale in translation significantly limits the estimation accuracy and robustness, making it not suitable for commercial applications. Since mobile robots typically move at ground surfaces, wheel encoders are commonly used in combination with cameras to improve the localization performance (Yap et al., 2011; Zhang et al., 2014; Ganganath and Leung, 2012; Zhang et al., 2019b,c). Alternatively, the low-cost visual-inertial odometry (VIO) which leverages a monocular camera and an inertial measurement unit (IMU) has also been widely used (Li et al., 2014; Leutenegger et al., 2015; Bloesch et al., 2015; Zheng et al., 2017; Qin et al., 2018). With the assist of IMU, the position estimation becomes metric deterministic, and the roll and pitch angles also become observable.

However, all methods mentioned above can not be directly used for localizing skid-steering robots with high precision. On the one hand, readings of wheel encoders from skid-steering robots can not be directly converted into motion estimates (i.e., linear and rotational velocities) of the robot. In skid-steering robots, the track-to-terrain interaction is exceptionally complicated, and the conversion between wheel encoder readings and robot’s motion depends on mechanical design, wheel inflation conditions, load and center of mass, terrain conditions, and so on. On the other hand, although visual-inertial localization has been successfully deployed for a variety of applications including mobile devices and drones, it has been shown that when used for ground robots, the estimation performance will be degraded (Wu et al., 2017; Zhang et al., 2019b). This is due to the incurred extra unobservable states under robots’ motion, additional undesired noise introduced by robots’ vibration, and so on.

To tackle all the problems mentioned above, in this paper, we propose a sliding-window bundle-adjustment based estimator by modeling and online estimating the kinematic parameters of a skid-steering robot. Our method allows explicit conversion between the wheel encoder measurements and the motion estimates (linear and rotational velocities on the local manifold) of the robot, and enables using wheel encoder...
measurements for pose estimation without non-probabilistic assumptions and approximations. In particular, our kinematic parameters include three instantaneous centers of rotation (ICRs) parameters and two correction factors, which are able to compensate for the changes due to slippage, terrain conditions, varying load and center of mass, tire inflation, and so on. By modelling both mechanical and environmental changes, the proposed method is able to operate in complicated environments for long periods of time without performance reduction.

It is also important to point out that the motivation and necessity of our method are different from other recent localization algorithms, which also propose to use kinematic constraints (Scaramuzza, 2011; Nisar et al., 2019). Those algorithms seek to incorporate kinematic constraints to improve the performance of a localization system, which can be individually functioning even without the constraints. However, in our work, as shown in experiments, instead of an optional performance enhancement method, modeling and estimating kinematic constraints for skid-steering robots are mandatory to ensure successful localization, due to the nature of the skid-steering mechanism.

Another key contribution of this work is the detailed estimator observability analysis, to allow correct handling of unobservable parameters for different robots. Since commercial robots might have different sensor configurations, we provide in-depth analysis for the following sensor configurations: i) using a monocular camera and wheel encoders for estimation, ii) incorporating an extra IMU into the sensor system, and iii) performing online sensor extrinsic calibration. We show that, with the correct setup, the skid-steering kinematic parameters are observable under general motion, allowing for consistent and accurate estimator performance. We also note that, the observability properties are not the same under different configurations, which indicates that when building real commercial robots, we must make the correct choices by not entering the un-observable cases. For example, we show that all extrinsic translational parameters between sensors are unobservable, and thus they must be pre-calibrated offline with high precision. Those results are different from the properties of other visual localization systems in which extrinsic parameters of sensors are either full observable (Li and Mourikis, 2014b; Qin and Shen, 2018; Schneider et al., 2019) or only unobservable in one translational axis (Chen et al., 2019). This emphasizes the importance of the observability analysis.

In summary, we focus on accurately localizing steering-skid robots in this paper, by using measurements from a monocular camera, wheel encoders, and optionally an IMU. The key contributions are as follows:

- An efficient tightly-coupled kinematics-constrained visual localization estimator dedicated to skid-steering robots, which jointly estimates the kinematic parameters of the robotic platform and 6DoF poses in a tight-coupled manner.
- Detailed observability analysis under different sensor configurations, and the key results are as follows: (i) by using a monocular camera and wheel encoders, only the three ICR kinematic parameters are observable; (ii) by introducing the additional IMU measurements, both the three ICR kinematic parameters and the two correction factors are observable under general motion; and (iii) the 3D translation and one dimension of the rotation in the extrinsics between the camera and odometer are unobservable with the online estimate of kinematic pa-
rameters, which prevents performing online sensor-to-sensor extrinsic calibration.

- Both simulation tests and real-world experiments were conducted, based on an off-the-shelf Clearpath robotic platform (see Fig. 1). The proposed kinematic-constrained method i) shows high accuracy and robustness under different environmental and mechanical conditions and (ii) outperforms competing methods that do not explicitly model the kinematic constraints.

The rest of the paper is organized as follows. We first review the related literatures in Sec. II. Subsequently, we introduce our kinematic model of skid-steering robots in Sec. III. In Sec. IV, the framework of the tightly-coupled sliding-window estimator is presented in detail. The observability analysis of the proposed method is performed in Sec. V. Finally, to validate the proposed approach, results from both Monte-Carlo simulation tests and real-world experiments are reported in Sec. VI. The paper is concluded in Sec. VII.

II. RELATED WORK

In this section, we group the related work into three main categories: i) localization algorithms for skid-steering robots, ii) camera-based localization, and iii) observability analysis for localization algorithms.

A. Localization Algorithms for Skid-steering Robots

Due to the popularity of scientific and commercial use cases, skid-steering robots have been extensively studied in recent years, and a variety of localization algorithms have been presented (Martínez et al., 2005; Yi et al., 2009; Pentzer et al., 2014; Huskic et al., 2017; Martínez et al., 2017; Sutoh et al., 2018). Early work by Anousaki et al. (Anousaki and Kyriakopoulos, 2004) showed that the standard differential-drive two-wheel vehicle model could not be used to accurately model the motion of a skid-steering robot due to track and wheel slippage. To address this problem, a dead reckoning model was proposed using experimental statistics. By using a similar concept, (Martínez et al., 2005) proposed an approach to calibrate skid-steering kinematic parameters offline experimentally. The performance of localizing skid-steering robots can also be significantly improved by incorporating measurements from complementary sensors, and (Yi et al., 2009) introduced an extended Kalman filter-based dead-reckoning method by fusing the IMU readings and wheel encoder measurements.

A key design choice for formulating an estimator for skid-steering robots is to model the skid-steering kinematic parameters, and we here discuss a couple of commonly used methods in the chronological order. A representative method was to use three ICR parameters for kinematic modeling (Martínez et al., 2005), which was demonstrated to be useful in a variety of environmental conditions. Instead of using three ICR parameters, (Scol and Sembera, 2008) proposed to use a quasi-kinematic dynamic model, (Moosavian and Kalantari, 2008) designed an exponential function to approximate the slippage coefficients, and (Wong, 2008) incorporated the terrain mechanics and vehicle dynamics into a skid-steering model. To have better computational efficiency and allow design simplicity, (Reina and Galati, 2016) used a single-parameter model by computing the skid-steering distance between the left and right tread. To have a more comprehensive error characterization, (Martínez et al., 2017) modelled an additional three sets of parameters, namely sliding, eccentricity and steering efficiency. An alternative algorithm was to model the ratio of the velocities between left and right wheels as an exponential function of the ratio of readings between left and right wheel encoders, which was demonstrated in detail in (Sutoh et al., 2018).

B. Camera based Localization

To date, the majority of camera-based localization algorithms are either generalized ones that can be used for different applications (Hesc et al., 2013; Li and Mourikis, 2013; Engel et al., 2014; Qin et al., 2018), or dedicatedly designed ones to specialized use cases such as i) hand-held or head-mounted mobile devices (Klein and Murray, 2007; Li et al., 2014; Do et al., 2019), ii) large-scale aircrafts and miniature drones (Achtelik et al., 2011; Huang et al., 2017; Nisar et al., 2019), and iii) ground robots and vehicles (Scaramuzza, 2011; Ganganath and Leung, 2012; Zhang et al., 2014, 2019c). The major difference factors between the camera-based localization algorithms used for ground robots and other applications are due to the motion of the camera itself and the availability of other sensors. On the one hand, mobile devices and drones can move freely in 3D, while robots can only navigate on ground surfaces. On the other hand, since most ground robots are equipped with either wheels or tracks, the encoders can be used to provide direct motion measurements.

In localization algorithms for ground robots, the kinematic constraints can be used to improve the localization performance. A variety of algorithms focus on planar environments or rely on high-quality IMUs to obtain roll and pitch of a robot. In those cases, the state space of the robot can be significantly reduced, and the estimation efficiency can be largely improved (Fossel et al., 2015; Zhang et al., 2019a; Zheng and Liu, 2019). To allow accuracy improvement, (Scaramuzza, 2011) proposed a method to use motion constraints in ground vehicles for processing visual features, in which data association and outlier rejection can be better performed. Zhang et al. (2019c) designed a method to jointly estimate the local manifold parameters as well as cameras’ poses, to allow high-precision estimation on non-planar scenes.

On the other hand, due to the low cost and wide availability, the algorithms of combining measurements from cameras and wheel encoders are also extensively studied (Yap et al., 2011; Zhang et al., 2014; Wu et al., 2017; Zhang et al., 2019b,c). (Yap et al., 2011) designed a particle filter-based method for camera-based localization, in which wheel encoder measurements were used for formulating propagation equations, and visual observations were utilized for the update. (Wu et al., 2017) proposed an approach to use wheel encoder measurements in a visual-inertial odometry system by introducing stochastic motion constraints on camera’s roll, pitch, and height. (Censi et al., 2013) designed a pose estimation system with the online wheel odometry parameter (the radius
of left and right wheels as well as the distance between them) calibration for a differential drive robot equipped with two wheels.

C. Observability Analysis for Localization Algorithms

Observability analysis is an essential component in designing localization algorithms, since estimating unobservable parameters will lead to un-predictable estimation performance. To conduct observability analysis, one category of methods is to discretize and linearize a continuous-time nonlinear robotic system, and construct a local observability matrix to investigate its rank and nullspace (Hesch et al., 2013; Li and Mourikis, 2013; Pentzer et al., 2014). Alternatively, observability properties can also be directly analyzed under continuous-time nonlinear format, and a representative family of methods is to compute continuous-time observability matrices via selecting proper equations of Lie derivatives (Guo and Roumeliotis, 2013; Yang and Huang, 2019). Another category of widely used tools for observability analysis is to directly investigate the sensors’ measurement models and robot’s kinematic equations, to identify whether there exist sets of different states that are able to generate identical sensor measurements under the same system inputs, following the definition of observability (Jones and Soatto, 2011; Censi et al., 2013; Li and Mourikis, 2014b; Zuo et al., 2019b).

The observability properties of visual-inertial localization have been widely studied, and researchers have proved that under general motion i) the global translation and rotation about yaw are unobservable while all other motion variables are observable, and ii) the spatial and temporal calibration parameters between the IMU and camera sensors are observable and thus can be calibrated online (Jones and Soatto, 2011; Hesch et al., 2013; Li and Mourikis, 2013, 2014b; Qin et al., 2018). In terms of localization using cameras and wheel encoders, it has been shown that the sensor-to-sensor relative vertical translation can not be identified in all cases since wheel encoders only provide 2D measurements (Guo et al., 2012). The properties of wheel odometry intrinsic parameters were also studied in (Censi et al., 2013), showing that those parameters can be estimated online. Pentzer et al. (2014) investigated the conditions that ICR parameters will be updated in a GPS-aided localization system, by demonstrating that the ICR parameters can be only updated when the robot is turning. However, this is just a glimpse of the observability property. The nature of the GPS measurements and the applicability of that algorithm are fundamentally different from ours.

Our previous work (Zuo et al., 2019b)) performed observability analysis of localizing steering skid robot by using a monocular camera, wheel encoders, and an IMU, and showed that the skid-steering parameters are generally observable. In this work, we significantly extend the analysis in (Zuo et al., 2019b), by explicitly identifying the identifiable and non-identifiable parameters with and without using the IMU. For example, we show that, when an IMU is not included in the sensor system, it is still possible to design an estimator by modeling a reduced skid-steering intrinsic parameter vector and pre-approximating the rest. Additionally, we also conduct detailed analyses of extrinsic parameters between sensors, which are also important factors in sensor fusion system.

III. Preliminaries on ICR-based Kinematics of Skid-Steering Robots

In this section, we present our kinematic model for skid-steering robots. In our derivation, we have assumed that the two wheels of the robots are always in contact with the ground surface. In other words, the case of a robot moving forward with one wheel hanging in the air is not allowed. In addition, the rotational rates of the wheels on each side of the robot are always the same, which is one of the most common mechanical design choices in skid-steering robots.

A. Notations

In this paper, we consider a robotic platform navigating with respect to a global reference frame, \{G\}. The platform is equipped with a camera, an IMU, and wheel odometers, whose frames are denoted by \{C\}, \{I\}, \{O\} respectively. To present transformation, we use \(A_{PB}\) and \(A_{BR}\) to denote position and rotation of frame \{B\} with respect to \{A\}, and \(\hat{A}_{B}q\) is the corresponding unit quaternion of \(A_{BR}\). In addition, \(I\) denotes the identity matrix, and \(0\) denotes the zero matrix. We use \(\tilde{x}\) and \(\delta x\) to represent the current estimated value and error state for variable \(x\). Additionally, we reserve the symbol \(\hat{x}\) to denote the inferred measurement value of \(x\), which is widely used in observability analysis. For the rotation matrix \(\mathbf{R}\), we define the attitude error angle vector \(\delta \theta\) as follows (Trawny and Roumeliotis, 2005):

\[
\mathbf{G}_R = \mathbf{G} \hat{\mathbf{R}} (I + [\delta \theta])
\]

where \([v]\) denotes the skew-symmetric matrix of the vector \(v\):

\[
[v] = \begin{bmatrix}
0 & -v_3 & v_2 \\
v_3 & 0 & -v_1 \\
-v_2 & v_1 & 0
\end{bmatrix}, \quad v = \begin{bmatrix}
v_1 \\
v_2 \\
v_3
\end{bmatrix}
\]

B. ICR-based Kinematics

In order to design a general algorithm to localize skid-steering robots under different conditions, the corresponding kinematic models must be presented in a parametric format. In this work, we employ a model similar to the ones in (Martínez et al., 2005; Pentzer et al., 2014), which contains five kinematic parameters: three ICR parameters and two correction factors, as shown in Fig. 1. To describe the details, we denote \(\text{ICR}_v = (X_v, Y_v)\) the ICR position of the robot frame, and \(\text{ICR}_l = (X_l, Y_l)\) and \(\text{ICR}_r = (X_r, Y_r)\) the ones of the left and right wheels, respectively. The relation between the readings of wheel encoder measurements and the ICR parameters can be derived as follows:

\[
Y_l = -\frac{o_l - o_{v_x}}{o_{w_z}}, \quad Y_r = -\frac{o_r - o_{v_x}}{o_{w_z}} \\
Y_v = \frac{o_{v_x}}{o_{w_z}}, \quad X_v = X_l = X_r = -\frac{o_{v_y}}{o_{w_z}}
\]

where \(o_i\) and \(o_r\) are linear velocities of left and right wheels, \(o_{v_x}\) and \(o_{v_y}\) are robot’s linear velocity along \(x\) and \(y\) axes.
represented in frame $O$ respectively, and $O_{\omega_2}$ denotes the rotational rate about yaw also in frame $O$. Those variables are also visualized in Fig. 1, and we use $\xi_{ICR} = [X_v, Y_l, Y_r]^{\top}$ to represents the set of ICR parameters. Moreover, we have used two scale factors, $\xi_\alpha = [\alpha_l, \alpha_r]^{\top}$, to compensate for factors which might cause changes in scales of the IMU readings. Representative situations include tire inflation, changes of road roughness, varying load of the robot, and so on. With the ICR parameters and correction factors being defined, the skid-steering kinematic model can be written as:

$$
\begin{bmatrix}
O_{v_x} \\
O_{v_y} \\
O_{\omega_2}
\end{bmatrix} = g(\xi, \alpha_l, \alpha_r) = \frac{1}{\Delta Y} \begin{bmatrix}
-Y_l & Y_l \\
X_v & -X_v \\
-1 & 1
\end{bmatrix} \begin{bmatrix}
\alpha_l \\
0 \\
\alpha_r
\end{bmatrix} \begin{bmatrix}
\alpha_l \\
0 \\
\alpha_r
\end{bmatrix}
$$

with

$$
\xi = [\xi_{ICR}^{\top}, \xi_\alpha^{\top}] = [X_v, Y_l, Y_r, \alpha_l, \alpha_r]^{\top}, \Delta Y = Y_l - Y_r
$$

where $\xi$ is the entire set of kinematic parameters. Interestingly, as a special configuration when

$$
b = [0, 0.5b, -0.5b, 1, 1]^{\top}
$$

with $b$ being the distance between left and right wheels, Eq. 4 can be simplified as:

$$
O_{v_x} = \frac{\alpha_l + \alpha_r}{2}, \quad O_{\omega_2} = \frac{\alpha_r - \alpha_l}{b}, \quad O_{v_y} = 0
$$

This is exactly the kinematic model for a wheeled robot moving without slippage (i.e., an ideal differential drive robot), and used by most existing work for localizing wheeled robots (Yap et al., 2011; Wu et al., 2017; Quan et al., 2018). However, in the case of skid-steering robots, if Eq. 7 is employed directly in a localizer, the pose estimation accuracy will be significantly reduced due to the incorrect conversion between wheel encoder readings and robot’s motion estimates (also see experimental results in Sec. VI). We also note that since the kinematic parameter $\xi$ represents the environmental and mechanical conditions of a moving robot, $\xi$ can not be modelled as a constant parameter vector. Instead, to allow high-precision localization, $\xi$ must be estimated online along with other localization states. This is conceptually similar to performing online sensor extrinsic calibration (Censi et al., 2013; Li and Mourikis, 2013, 2014b). However, it is practically possible to calibrate sensors’ extrinsic parameter offline, while infeasible to estimate all skid-steering kinematic parameters before deployment.

IV. KINEMATICS-CONSTRAINED VISUAL-INERTIAL LOCALIZATION

In this paper, we utilize a sliding-window bundle adjustment based estimator for localizing skid-steering robots using a monocular camera, wheel encoders, and optionally an IMU. For presentation simplicity, in this section, we describe our estimator by explicitly considering using the IMU. When the IMU is not included in the sensor system, our presented estimator can be straightforwardly modified by simply deleting the IMU related components.

The architecture of our sliding-window estimator closely follows the design of Eckenhoff et al. (2019); Zhang et al. (2019b), by iteratively optimizing sensor measurement constraints and probabilistically marginalizing old information. We also note that, compared to the articles that focus on estimator architecture novelty, this work is to describe methods to systematically handle skid-steering effects via online calibration. Our goal is to consistently and accurately estimate the motion of a moving robot as well as necessary observability-guided calibration parameters.

A. Estimator Formulation

1) State Vector: To start with, we define the state vector of our estimator as:

$$
x = \left[ x^{\top} O_v^{\top} G_{v_i}^{\top} b_{dk}^{\top} b_{dk}^{\top} m_k^{\top} \xi_k^{\top} \right]^{\top}
$$

where

$$
x = \left[ x^{\top} O_v^{\top} G_{v_i}^{\top} b_{dk}^{\top} b_{dk}^{\top} m_k^{\top} \xi_k^{\top} \right]^{\top}
$$

denotes the sliding-window poses of odometer frame at times $\{k - s, \ldots, k\}$ when keyframe images are captured. $G_{v_i}, b_{dk}$ are the IMU related states, including the IMU velocity in global frame, acceleration bias, and gyroscope bias. If IMU is not available in the system, $G_{v_i}, b_{dk}$ will excluded from the state vector. In addition, $m_k$ denotes the parameters for modeling the local motion manifold of the skid-steering robots across current sliding window. This has been shown in (Zhang, Chen and Li, 2019b) to improve the estimation performance for ground robots, and we also adopt this design in our work. For completeness of the estimator presentation, we also provide the details on $m_k$ in Appendix VI-C. Finally, $\xi_k$, as shown in Eq. 5, represents the skid-steering intrinsic parameter vector, which is explicitly included in the state vector and thus estimated online.

2) Optimization Process: Our optimization process closely follows the design of (Eckenhoff et al., 2019; Zhang et al., 2019b). Specifically, as illustrated in Fig. 2, the sliding-window BA in our estimation algorithm seeks to iteratively minimize a cost function corresponding to a combination of sensor measurement constraints, motion kinematic constraints, and marginalized constraints.

$$
C = C_P + C_V + C_f + C_O + C_M
$$

In what follows, we describe each of the cost terms. Firstly, the marginalized term $C_P$ is critical to consistently keep the algorithm computational complexity bounded, by probabilistically removing the old states in the sliding window. For a constraint $C(x_r, x_m)$ involved with the old states needed to be marginalized $x_m$ and the remaining states $x_r$, we compute the Hessian and gradient matrices with respect to $\left[ x_r^{\top}, x_m^{\top} \right]^{\top}$, which are denoted as:

$$
\begin{bmatrix}
\Lambda_{rr} & \Lambda_{rm} \\
\Lambda_{mr} & \Lambda_{mm}
\end{bmatrix},
\begin{bmatrix}
\mathbf{g}_r \\
\mathbf{g}_m
\end{bmatrix}
$$

The marginalization can be conducted by computing the marginalized Hessian and gradient matrices, i.e., $\Lambda_{marg} =$
$\Lambda_{rr} - \Lambda_{rm}\Lambda_{mm}^{-1}\Lambda_{mr}$ and $g_{marg} = g_r - \Lambda_{rr}\Lambda_{mm}^{-1}g_r$, which represent the uncertainty information for the remaining states $x_r$ in the current sliding window (Eckenhoff et al., 2019). Once marginalization is performed, the prior cost function can be formulated to ensure the remaining states are characterized by the computed uncertainties:

$$C_P(x_r) = \frac{1}{2} \|x_r \mathbin{\Box} \hat{x}_r\|^2_{\Lambda_{marg}} + g_{marg}^T (x_r \mathbin{\Box} \hat{x}_r)$$ (12)

The “boxminus” operator $\mathbin{\Box}$ denotes the generalized minus operation, since we need to perform computations on the manifold. For the marginalization, it should be noted that, as shown in Fig. 2, for limiting the computational complexity, we only leverage the constraints from IMU $C_I$ and odometry $C_O$ between the latest frame $k$ and the second latest frame $k-1$. After $C_I$ and $C_O$ are minimized in the optimization, the information contained in them and the related states will be marginalized into the prior cost term.

The camera term $C_P$, IMU term $C_I$, and motion manifold term $C_M$ used in this work are similar to that of existing literature (Li and Mourikis, 2013; Eckenhoff et al., 2019; Zhang et al., 2019b) but with dedicated design for ground robots. In general, the camera cost term models the geometrical reprojection error of point features in the keyframes, the IMU term computes the error of IMU states between two consecutive keyframes, and the manifold cost term characterizes the motion smoothness across the whole sliding window. To have this article self-contained, the exact cost terms we use are provided in the Appendix VII. Finally, $C_O$ denotes the error induced by wheel odometer measurements. This term is a function of robot pose, measurement input, as well as skid-steering intrinsic parameters, and is discussed in detail in the next section.

It should also be noted that in this work, we assume that the IMU, the wheel odometers, and the camera are synchronized by hardware. Integration of IMU and odometer measurements between the time instants of captured images are required in the constraints $C_I$ and $C_O$. However, since different types of measurements come at varying frequencies, it is unlikely to get IMU/odometer measurements at the exact time instants when capturing the images. Thus, we perform the linear interpolations of IMU and odometer measurements at the image capturing time for performing integration.

**B. ICR-based Kinematic Constraints**

This section provides details on formulating $C_O$. Specifically, by assuming the supporting manifold of the robot is locally planar between $t_k$ and $t_{k+1}$, the local linear and angular velocities, $O(t)v$ and $O(t)\omega$, are a function of the wheel encoders’ measurements of the left and right wheels $o_{lm}(t)$ and $o_{rm}(t)$ as well as the skid-steering kinematic parameters $\xi$ [see (4)]:

$$\begin{bmatrix} O(t)v \\ O(t)\omega \end{bmatrix} = \Pi g(\xi(t), o_l(t), o_r(t))$$

$$= \Pi g(\xi(t), o_{lm}(t) - n_l(t), o_{rm}(t) - n_r(t))$$ (13)

where $\Pi = [e_1 \ e_2 \ 0 \ 0 \ 0 \ e_3]^T$ is the selection matrix with $e_i$ being a $3 \times 1$ unit vector with the $i$th element of 1, $n_l(t)$ and $n_r(t)$ are the odometry noise modeled as zero-mean white Gaussian. With slight abuse of notation, we define $n_o = [n_l \ n_r]^T$.

By using $O(t)v$ and $O(t)\omega$, the wheel odometry based kinematic equations are given by:

$$G_{PO}(t) = G_{O(t)}R \cdot [O(t)v]$$ (14a)

$$G_{rO}(t) = G_{O(t)}R \cdot [O(t)\omega]$$ (14b)

$$\xi(t) = n_\xi(t)$$ (14c)

where we model the noise of the ICR kinematic parameter $\xi$ by using a random walk process, and $n_\xi$ is characterized by zero-mean white Gaussian noise. The motivation of using $n_\xi$ is to capture time-varying characteristics of $\xi$, caused by changes in road conditions, tire pressures, center of mass, and so on. It is important to point out that, unlike sensor extrinsic calibration in which parameters can be modeled as constant parameters, e.g., $C_P = 0$ in (Kelly and Sukhatme, 2011), $\xi$ must be modeled as a time-varying variable.

To propagate pose estimates in a stochastic estimator, we describe the process starting from the estimates $x_{O_{k-1}} = \left[ G_{PO_{k-1}} G_{rO_{k-1}} q_{k-1}^T \right]^T$. Once the instantaneous local velocities of the robot (see Eq. 13) are available, we integrate the differential equations in Eq. 14 over the time interval $t \in (t_{k-1}, t_k)$ by all the intermediate odometer measurements $O_m$, and obtain the predicted robot pose and kinematic parameters at the newest keyframe time $t_k$. This process can be characterized by $\hat{x}_{O_k} = P (x_{O_{k-1}}, O_m)$. Therefore,
the odometer-induced kinematic constraint can be generically written in the following form:

\[ C_O (x_{O_k}, x_{O_{k-1}}) = \|x_{O_k} - f(x_{O_{k-1}}, o_m)\|^2_{\Lambda_O} \]  

(15)

where \( \Lambda_O \) represents the inverse covariance (information) obtained via propagation process. To formulate Eq. 15 in the stochastic estimator, error state characteristics also need to be computed since linearization of the propagation function \( f \) consists of Jacobian matrices with respect to error states. We start with the continuous-time error state model, by linearizing Eq. 14:

\[
\delta^G v \simeq \delta^G \tilde{r} \begin{pmatrix} \hat{O} \dot{v} + \mathbf{J}_{v\xi} \delta \xi + \mathbf{J}_{v\omega} \nu_o \end{pmatrix} - \delta^G \tilde{r} \hat{O} \dot{v} \\
\delta^G \dot{\theta} \simeq -[\hat{O} \omega] \delta \theta + \mathbf{J}_{\omega \xi} \delta \xi + \mathbf{J}_{\omega \omega} \nu_o \\
\delta \xi \simeq n_\xi
\]

(16)

(17)

(18)

We here point out that Eq. 17 can be obtained similar to Eq. 15 of (Trawny and Roumeliotis, 2005). In above equations, \( \mathbf{J}_{v\xi}, \mathbf{J}_{v\omega}, \mathbf{J}_{\omega \xi}, \mathbf{J}_{\omega \omega} \) are the linearized Jacobian matrices, originated from:

\[
\mathbf{O} \dot{v} = \mathbf{O} \dot{v} + \mathbf{J}_{v\xi} \delta \xi + \mathbf{J}_{v\omega} \nu_o \\
\mathbf{O} \omega = \mathbf{O} \omega + \mathbf{J}_{\omega \xi} \delta \xi + \mathbf{J}_{\omega \omega} \nu_o
\]

(19)

(20)

and

\[
\begin{bmatrix}
\hat{\alpha}_1 \dot{O}_1 - \hat{\alpha}_r \omega_r \\
\lambda Y^2 \\
\end{bmatrix} + \begin{bmatrix}
0 & 0 & -\hat{Y}_r & 0 \\
\hat{Y}_r & -\hat{X}_v & 0 & 0 \\
0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

\[
\begin{bmatrix}
0 & 0 & -\hat{Y}_r \omega_l & \hat{Y}_r \omega_l \\
\hat{X}_v \omega_l & \hat{X}_v \omega_l & \hat{X}_v \omega_l & 0 \\
0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

(21)

\[
\begin{bmatrix}
-\hat{\alpha}_1 \hat{Y}_r & \hat{\alpha}_r \hat{Y}_r \\
\hat{X}_v \alpha_l & -\hat{X}_v \alpha_r \\
0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

(22)

\[
\begin{bmatrix}
-\hat{\alpha}_1 \hat{Y}_r & \hat{\alpha}_r \hat{Y}_r \\
\hat{X}_v \alpha_l & -\hat{X}_v \alpha_r \\
0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

(23)

\[
\begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

(24)

Once continuous-time error-state equations are given, the discrete-time Jacobian matrices, e.g., \( \frac{df(t)}{d\xi_{k-1}} \) and \( \frac{df(t)}{d\xi} \) in Eq. 15, can be straightforwardly calculated, similar to the process described in (Trawny and Roumeliotis, 2005). As a result, Eq. 15 encapsulates all information related to the skid-steering effect and enables online estimation of the skid-steering parameters. More details can also be found in our technical report (Zuo, Huang and Li, 2019a). After the constraint in Eq. 15 is minimized, \( \xi_{k-1} \) will be marginalized immediately, ensuring low computational complexity of the system.

C. Initialization of Kinematic Parameters

To allow the estimation of skid-steering kinematic parameters online, an initial estimate of the parameter vector \( \xi \) is required. Generally, in a stochastic estimator, the initial estimate can either be computed purely from sensory data or from prior knowledge. The first type of method is typically used for variables that are independent over trials (e.g., robot initial velocity (Li and Mourikis, 2014a)), and the other type is used for the states that may vary relatively slowly (e.g., sensor extrinsics (Censi, Franchi, Marchionni and Oriolo, 2013)). The skid-steering kinematic parameter belongs to the second case, and thus we propose a prior-based method.

Specifically, we use a simply while effective method by setting:

\[
\xi_{initial} = \begin{bmatrix} 0, 0.5b^t, -0.5b^t, 1, 1 \end{bmatrix}^T
\]

(25)

We emphasize that Eq. 25 is similar but different from Eq. 6. The parameter \( b^t \) represents wheel distance in Eq. 6, which can be correctly used for robot without slippage. However, skid-steering robots are designed to have slippery behaviors, and thus \( b^t \) should not be simply the wheel distance. To compute \( b^t \), we rotate the skid-steering robots and use the fact that rotational velocity reported by the IMU and odometry should be identical, which leads to the follow equation:

\[
b^t = \frac{1}{N} \sum_{i=1}^{N} \frac{||o_{in}(t_i) - o_{rm}(t_i)||}{||\omega_m(t_i)||}
\]

(26)

where \( \omega_m(t_i) \) is gyroscope measurement, and \( N \) is the number of measurements. Although this is not of high precision and the road condition of computing \( b^t \) is different from that of the testing time, this simple initialization method in combination of the proposed online calibration algorithm is able to yield accurate localization results (see our experimental results).

V. Observability Analysis

A critical prerequisite condition for a well-formulated estimator is to only include locally observable (or identifiable\(^1\)) (Bar-Shalom and Fortmann, 1988) sensor and kinematic parameters (i.e., intrinsic and extrinsic parameters) in the online optimization stage. In the skid-steering robot localization system, a subset of estimation parameters inevitably become unobservable under center circumstances, which will be analytically characterized in this section and avoided in a real-world deployment.

Specifically, in this section, we first conduct our analysis by assuming the extrinsic parameters between sensors are perfectly known, and analyze the observability properties of the skid-steering parameters in different sensor system setup. Specifically, we consider three cases: (i) monocular camera and odometer with the 3 ICR parameters and 2 correction factors. Subsequently,

\(^1\)Since the derivative of \( \xi \) is modelled by zero-mean Gaussian, we here use observability and identifiability interchangeably.
we perform the analysis under the case that extrinsic parameters between sensors are unknown. Since estimating extrinsic parameters online is a common estimator design choice in robotics community (Guo, Mirzaei and Roumeliotis, 2012; Heng, Li and Pollefeys, 2013; Geiger, Moosmann, Car and Schuster, 2012), we also investigate the possibility of doing that for skid-steering robots.

A. Methodology Overview

To investigate the observability properties, the analysis can be either conducted in the original nonlinear continuous-time system (Li and Mourikis, 2013) or the corresponding linearized discrete-time system (Guo and Roumeliotis, 2013; Yang and Huang, 2019). As shown in (Li and Mourikis, 2013; Hesch, Kottas, Bowman and Roumeliotis, 2013), the dimension of the nullspace of the observability matrix might subject to changes due to linearization and discretization, and thus we conduct our analysis in the nonlinear continuous-time space in this work.

To conduct the observability analysis, we follow the methodology in (Li and Mourikis, 2014b), to examine the existence of indistinguishable trajectories given the kinematic and sensor measurement models. Specifically, the observability analysis consists of three main steps: firstly, we investigate the information provided by each sensor, and derive inferred ‘abstract’ measurements from the raw measurements; secondly, we use kinematic and measurement constraints to derive equations that indistinguishable trajectories must follow; finally, the observability matrix is constructed by computing the derivatives of the previous derived equations with respect to the states of interests. The observability of the states can be determined by examining the rank and nullspace of the observability matrix (Van Doren et al., 2009).

B. Inferred Measurement Model

We first analyze the information provided by a monocular camera. It is well-known that a monocular camera is able to provide information on rotation and up-to-scale position with respect to the initial camera frame (Hartley and Zisserman, 2003; Li and Mourikis, 2014b) under general motion. Equivalently, the information characterized by a monocular camera can be given by: (i) camera’s angular velocity and (ii) its up-to-scale linear velocity:

\[
\begin{align*}
\bar{\omega}_C(t) &= C_{(t)} \omega + n_\omega(t) \quad (27a) \\
\bar{v}_C(t) &= s^{-1} \bar{C} R \cdot G_{V_C(t)} + n_v(t) \quad (27b)
\end{align*}
\]

where \(n_\omega(t)\) and \(n_v(t)\) are the measurement noises, \(C_{(t)}\omega\) denotes true local angular velocity expressed in camera frame, and \(G_{V_C(t)}\) is the linear velocity of camera with respect to global frame, and finally \(s\) is an unknown scale factor. Additionally, \(\bar{\omega}_C(t)\) and \(\bar{v}_C(t)\) denote the inferred rotational and linear velocity measurements. Moreover, to make our later derivation simpler, we also introduce the rotated inferred measurements as follows:

\[
\bar{\omega}(t) \triangleq \bar{C} R \cdot \bar{\omega}_C(t), \quad \bar{v}(t) \triangleq \bar{C} R \cdot \bar{v}_C(t) \quad (28)
\]

It is important to point out that in the cases when extrinsic parameter calibration between sensors is not considered in the online estimation stage, \(\bar{\omega}(t)\) and \(\bar{v}(t)\) can be uniquely computed from the camera measurement and also treated as the inferred measurement.

C. Observability of \(\xi\) with Monocular Camera and Odometer

We first investigate the case when a system is equipped with a monocular camera and odometers, and their extrinsic parameters are known in advance. To perform observability analysis, we derive system equations that indistinguishable trajectories must satisfy. To start with, we note that the following geometric relationships hold for any camera-odometer system:

\[
O_\omega = \bar{C} R \cdot C_\omega \quad (29)
\]

which allows us to derive the following equations:

\[
\begin{align*}
G_{pO} &= C_{(t)} R \cdot C_{pO} + \bar{G}_{pC} \quad (30a) \\
G_{vO} &= C_{(t)} R \cdot C_{vO} + \bar{G}_{vC} \quad (30b) \\
\bar{G}_R^{-1} \bar{G}_{vO} &= (\bar{G}_{vC})^{-1} \bar{G}_R^{-1} \bar{G}_{vO} R^{-1} \bar{G}_{vC} \quad (30c) \\
o_v &= -[\bar{\omega}] p_o C_o + s \cdot \bar{v} \quad (31)
\end{align*}
\]

Substituting Eq. 28 to Eq. 30d, we obtain the following equation:

\[
\begin{bmatrix}
\bar{\omega}^o_{yC} \\
\bar{\omega}^o_{xC} \\
\bar{\omega}
\end{bmatrix} + s \begin{bmatrix}
\hat{\bar{v}}_x \\
\hat{\bar{v}}_y
\end{bmatrix} = \frac{1}{\Delta Y} \begin{bmatrix}
-\bar{Y}_r & \bar{Y}_l \\
-\bar{X}_v & -\bar{X}_v \\
1 & 0
\end{bmatrix} \begin{bmatrix}
\alpha_l & 0 & 0 \\
0 & \alpha_r & 0 \\
0 & 0 & 0
\end{bmatrix} \begin{bmatrix}
o_l \\
o_r \\
o_o
\end{bmatrix}
\]

where \(\bar{x}_C\) and \(\bar{y}_C\) are the first and second element of \(\bar{r}_C\), and \(\bar{v}_x, \bar{v}_y\) are the first and second element of \(\bar{v}\). For brevity, we use \(\hat{\bar{w}}\) to denote the third element of \(\bar{w}\). By defining \(\beta_r = \Delta Y^{-1} \alpha_r\) and \(\beta_l = \Delta Y^{-1} \alpha_l\), we can write

\[
\begin{bmatrix}
\bar{\omega}^o_{yC} \\
\bar{\omega}^o_{xC} \\
\bar{\omega}
\end{bmatrix} + s \begin{bmatrix}
\hat{\bar{v}}_x \\
\hat{\bar{v}}_y
\end{bmatrix} = \begin{bmatrix}
\bar{\omega}_Y & -\bar{\omega}_X & \beta_l \Delta Y \alpha_l \\
\bar{\omega}_X & \bar{\omega}_Y & 0 \\
-\beta_l \alpha_l + \beta_r \alpha_r & 0 & 0
\end{bmatrix} \begin{bmatrix}
o_l \\
o_r \\
o_o
\end{bmatrix}
\]

Note that, this equation only contains 1) sensor measurements, and 2) a combination of vision scale factor and skid-steering kinematics:

\[
\epsilon = [X_v, Y_l, Y_r, \alpha_l, \alpha_r, s]^T
\]
which allows us to analyze whether indistinguishable sets of \( \epsilon \) exist subject to the provided measurement constraint equations.

The identifiability of \( \epsilon \) can be described as follows:

**Lemma 1.** By using measurements from a monocular camera and wheel odometers, \( \epsilon \) is not locally identifiable.

**Proof.** \( \epsilon \) is locally identifiable if and only if \( \bar{\epsilon} \) is locally identifiable:

\[
\bar{\epsilon} = [Y_t \quad \Delta Y \quad X_v \quad \beta_l \quad \beta_r \quad s]^T
\]

By expanding Eq. 33, we can write the following constraints:

\[
\begin{align*}
  c_x(\epsilon, t) & = \bar{\omega}(t)O^c_y + s\bar{v}_x(t) - \bar{\omega}(t)Y_t - \beta_l \Delta Y o_l(t) = 0 \quad (34a) \\
  c_o(\epsilon, t) & = -\bar{\omega}(t)O^c_x + s\bar{v}_y(t) + \bar{\omega}(t)X_v = 0 \quad (34b) \\
  c_\omega(\epsilon, t) & = \bar{\omega}(t) + \beta_l o_l(t) - \beta_r o_r(t) = 0 \quad (34c)
\end{align*}
\]

A necessary and sufficient condition of \( \bar{\epsilon} \) to be locally identifiable is following observability matrix has full column rank, over a set of time instants \( S = \{t_0, t_1, \ldots, t_s\} \):

\[
O_c = \begin{bmatrix} D(t_0)^T & D(t_1)^T & \ldots & D(t_s)^T \end{bmatrix}^T
\]

where

\[
D(t) = \begin{bmatrix}
\frac{\partial c_x(\epsilon, t)}{\partial \epsilon} & \frac{\partial c_o(\epsilon, t)}{\partial \epsilon} & \frac{\partial c_\omega(\epsilon, t)}{\partial \epsilon}
\end{bmatrix}^T
\]

(35)

Substituting Eq. 36 back into Eq. 35 leads to:

\[
O_c = \begin{bmatrix}
-\bar{\omega}(t_0) & -\beta_l o_l(t_0) & 0 & -\Delta Y o_l(t_0) & 0 & \bar{v}_x(t_0) & 0 & 0 & \bar{\omega}(t_0) & 0 & \bar{v}_y(t_0) & 0 & 0 & 0 & \bar{\omega}(t_0) & 0 & \bar{v}_y(t_0)
\end{bmatrix}
\]

(37)

By defining \( O_c(:, i) \) the \( i \)th block columns of \( O_c \), the following equation holds:

\[
\begin{align*}
  & (-O^c_y + Y_t) \cdot O_c(:, 1) + \Delta Y \cdot O_c(:, 2) \\
  & + (X_v - O^c_x) \cdot O_c(:, 3) + s \cdot O_c(:, 6) = 0
\end{align*}
\]

The above equation demonstrates that \( O_c \) is not of full column rank, indicating that \( \epsilon \) is not identifiable.

To further investigate the indistinguishable states that cause the unobservable situations, we note that for a vector \( \bar{\epsilon}_1 = [Y_t, \Delta Y, X_v, \beta_l, \beta_r, s]^T \) that satisfies Eq. 34, another vector

\[
\bar{\epsilon}_2 = [(1 + \lambda/s)Y_t - (\lambda/s)O^c_y, (1 + \lambda/s)\Delta Y, (1 + \lambda/s)X_v - (\lambda/s)O^c_x, \beta_l, \beta_r, s + \lambda]^T
\]

for any \( \lambda \in \mathbb{R} \) is always valid for the constraints Eq. 34. Thus \( \bar{\epsilon}_1 \) and \( \bar{\epsilon}_2 \) are indistinguishable, and \( \epsilon \) is not locally identifiable. This completes the proof.

**D. Observability of \( \xi_{ICR} \) with Monocular Camera and Odometer**

Since the full kinematic parameters \( \xi = [\xi^T_{ICR}, \xi^T_\omega]^T \) with monocular camera and odometer are not locally identifiable, we look into the case of that only the 3 ICR parameters, i.e., \( \xi_{ICR} \), are estimated without the correction factors. Similar to Eq. 32, the following equation holds:

\[
\begin{bmatrix}
\begin{bmatrix}
\bar{\omega}O^c_y \\
-\bar{\omega}O^c_x
\end{bmatrix} + s \begin{bmatrix}
\bar{v}_x \\
\bar{v}_y
\end{bmatrix} & = \\
\frac{\Delta Y}{\partial Y} (o_l - o_l) + 0
\end{bmatrix} = 0
\]

(38)

The above expression is a function of the odometer and inferred visual measurements \( \bar{\omega}, \bar{v}_x, \bar{v}_y, o_l, o_r \), as well as the kinematic intrinsic parameters \( \xi_{ICR} \) and visual scale factor \( s \):

\[
\gamma = \begin{bmatrix} \xi^T_{ICR} \end{bmatrix}^T = \begin{bmatrix} X_v \quad Y_t \quad Y_r \quad s \end{bmatrix}^T
\]

The local identifiability of \( \gamma \) can be stated as follows:

**Lemma 2.** By using the monocular and odometer measurements, and the 3 ICR parameter vector \( \xi_{ICR} \) to model the kinematics, \( \gamma \) is locally identifiable except for the following degenerate cases: (i) the odometer linear velocity \( o_l(t) \) keeps zero; (ii) the angular velocity \( \bar{\omega}(t) \) keeps zero; (iii) \( o_r(t), o_l(t), \) and \( \bar{\omega}(t) \) are all constants; (iv) the linear velocities of two wheels \( o_l(t), o_r(t) \) keeps identical to each other; (v) the angular velocity \( \bar{\omega}(t) \) is consistently proportional to \( o_l(t) \).

**Proof.** We first note that the local identifiability of \( \gamma \) is equivalent to that of \( \bar{\gamma} \).

\[
\bar{\gamma} = \begin{bmatrix} Y_t \quad \Delta Y \quad X_v \quad s \end{bmatrix}^T
\]

By expanding 38 and considering all the measurements at different time \( t \), we can derive following system constraints:

\[
\begin{align*}
  c_x(\bar{\gamma}_t) & = \bar{\omega}(t)O^c_y + s\bar{v}_x(t) - \bar{\omega}(t)Y_t - \beta_l \Delta Y o_l(t) = 0 \quad (39a) \\
  c_o(\bar{\gamma}_t) & = -\bar{\omega}(t)O^c_x + s\bar{v}_y(t) + \bar{\omega}(t)X_v = 0 \quad (39b) \\
  c_\omega(\bar{\gamma}_t) & = \bar{\omega}(t) + \beta_l o_l(t) - \beta_r o_r(t) = 0 \quad (39c)
\end{align*}
\]

Similar to the case of using full kinematic parameters \( \xi \) in Section. V-C, we derive the following observability matrix for \( \bar{\gamma} \) (using Eq. 35 and. 36):

\[
\bar{O}_c = \begin{bmatrix}
-\bar{\omega}(t_0) & 0 & 0 & \bar{v}_x(t_0) \\
0 & 0 & \bar{\omega}(t_0) & \bar{v}_y(t_0) \\
\vdots & \vdots & \vdots & \vdots \\
0 & \bar{v}_x(t_s) & 0 & \bar{v}_y(t_s) \\
0 & 0 & \bar{\omega}(t_s) & \bar{v}_y(t_s) \\
\end{bmatrix}
\]

(40)

To simplify the structure of the observability matrix, we apply the following linear operations without changing the observability properties:

\[
\begin{align*}
  O_c(:, 2) & \leftarrow (\Delta Y)^2 O_c(:, 2) \\
  O_c(:, 4) & \leftarrow sO_c(:, 4) + (Y_t - O^c_y)O_c(:, 1) + (X_v - O^c_x)O_c(:, 3)
\end{align*}
\]

\[
O_c(:, 2) \leftarrow (\Delta Y)^2 O_c(:, 2)
\]

\[
O_c(:, 4) \leftarrow sO_c(:, 4) + (Y_t - O^c_y)O_c(:, 1) + (X_v - O^c_x)O_c(:, 3)
\]
where $(\cdot) \leftarrow (\cdot)$ represents the operator to replace the left side by the right side. As a result, Eq. 40 can be simplified as:

\[
\mathcal{O}_c = \begin{bmatrix}
-\dot{\omega}(t_0) & 0 & 0 & o_l(t_0) \\
0 & 0 & \dot{\omega}(t_0) & 0 \\
o_r(t_0) - o_l(t_0) & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
-\dot{\omega}(t_s) & 0 & 0 & o_l(t_s) \\
0 & 0 & \dot{\omega}(t_s) & 0 \\
o_r(t_s) - o_l(t_s) & 0 & 0 & 0
\end{bmatrix}
\]

\[ (41) \]

To investigate the observability of the matrix in Eq. 41, we inspect the existence of the non-zero vector $k$ such that $\mathcal{O}_c k = \mathbf{0}$, $k = [k_1 \ k_2 \ k_3 \ k_4]^T \neq \mathbf{0}$. If such a vector $k$ exists, all of the following conditions must be satisfied:

\[-\dot{\omega}(t_1) + o_l(t_1)k_4 = 0, (o_r(t) - o_l(t))k_2 = 0, \dot{\omega}(t_3)k_3 = 0\]

To allow the above equations to be true, one of the following conditions is required:

- $o_l(t)$ keeps constantly zero, $k = [0 \ 0 \ 0 \ \rho]^T$, 
- $\dot{\omega}(t)$ keeps constantly zero, $k = [0 \ 0 \ \rho \ \rho]^T$, 
- $o_r(t)$, $o_l(t)$, and $\dot{\omega}(t)$ are all constants, $k = [0 \ \rho \ 0 \ 0]^T$, 
- $o_l(t)$ keeps identical to $o_r(t)$, $k = [0 \ \rho \ 0 \ 0]^T$, 
- $\dot{\omega}(t)$ keeps proportional to $o_l(t)$, $k = [o_{\beta} / \dot{\omega} \ 0 \ 0 \ \rho]^T$.

where $\rho$ can be any non-zero value that is used to generate valid non-zero vector $k$ such that $\mathcal{O}_c k = \mathbf{0}$. We note that, all above cases are special conditions. When a robot moves under general motion, none of those conditions can be satisfied. Therefore, in a camera and odometers only skid-steering robot localization system, $\xi_{ICR}$ is observable unless entering the specified special conditions listed above.

E. Observability of $\xi$ with a Monocular Camera, an IMU, and Odometer

So far, we have shown that when a robotic system is equipped with a monocular camera and wheel odometer, estimating $\xi$ is not feasible, and the alternative solution is to include $\xi_{ICR}$ in the online stage only. However, it is not an ideal solution to calibrate $\xi_{sa}$ offline and fixed in the online stage since it is subject to changes in road conditions and tire conditions and so on.

To tackle this problem, we investigate the observability of $\xi$ when an IMU is added to the camera-odometer system. Once an IMU is used, similar to the previous analysis, we start by introducing the ‘inferred measurement’. Instead of focusing on visual measurement only, we provide ‘inferred’ measurement by considering the visual-inertial system together. As analyzed in rich existing literature, visual-inertial estimation provides: camera’s local (i) angular velocity and (ii) linear velocity, similar to vision only case (Eq. 28) without having the unknown scale factor (Hesch, Kottas, Bowman and Roumeliotis, 2013; Li and Mourikis, 2014b; Schneider, Li, Cadena, Nieto and Siegwart, 2019). Similarly to Eq. 34a, to simplify the analysis, we prove identifiability of $\hat{\xi}$ instead of $\xi$, since properties of $\xi$ and $\hat{\xi}$ are interchangeable:

\[
\hat{\xi} = [Y_l \ \Delta Y \ X_v \ \beta_l \ \beta_r]^T
\]

**Lemma 3.** By using measurements from a monocular camera, an IMU, and wheel odometer, $\hat{\xi}$ is locally identifiable, except for following degenerate cases: (i) velocity of one of the wheels, $o_r(t)$ or $o_l(t)$, keeps zero; (ii) $\dot{\omega}(t)$ keeps zero; (iii) $o_r(t)$, $o_l(t)$, and $\dot{\omega}(t)$ are all constants; (iv) $o_r(t)$ is always proportional to $o_r(t)$; (v) $\dot{\omega}(t)$ is always proportional to $o_l(t)$.

**Proof.** Similarly to Eq. 34, by removing the scale factor, the corresponding system constraints can be derived as:

\[(42a) \]

\[
\begin{align*}
c_x(\xi, t) &= \dot{\omega}(t)O_C + \ddot{v}_x(t) - \dot{\omega}(t)Y_l - \beta_l \Delta Y o_l(t) = 0 \\
c_y(\hat{\xi}, t) &= -\dot{\omega}(t)O_C + \ddot{v}_y(t) + \dot{\omega}(t)X_v = 0 \\
c_z(\xi, t) &= \dot{\omega}(t) + \beta_l o_l(t) - \beta_r o_r(t) = 0
\end{align*}
\]

Therefore, the observability matrix for $\hat{\xi}$ can be computed by:

\[
\mathcal{O}_c = \begin{bmatrix}
-\dot{\omega}(t_0) & -\beta_l o_l(t_0) & 0 & -\Delta Y o_l(t_0) & 0 \\
0 & 0 & \dot{\omega}(t_0) & 0 & 0 \\
0 & 0 & 0 & o_l(t_0) & -o_r(t_0) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
-\dot{\omega}(t_s) & -\beta_l o_l(t_s) & 0 & -\Delta Y o_l(t_s) & 0 \\
0 & 0 & \dot{\omega}(t_s) & 0 & 0 \\
0 & 0 & 0 & o_l(t_s) & -o_r(t_s)
\end{bmatrix}
\]

\[(43) \]

After the following linear operations:

\[
\mathcal{O}_c(1, 2) \leftarrow -\mathcal{O}_c(1, 2) / \beta_l \\
\mathcal{O}_c(1, 4) \leftarrow \mathcal{O}_c(1, 4) + \Delta Y \mathcal{O}_c(1, 2)
\]

Eq. 43 can be simplified as:

\[
\mathcal{O}_c = \begin{bmatrix}
-\dot{\omega}(t_0) & o_l(t_0) & 0 & 0 & 0 \\
0 & 0 & \dot{\omega}(t_0) & 0 & 0 \\
0 & 0 & 0 & o_l(t_0) & -o_r(t_0) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
-\dot{\omega}(t_s) & o_l(t_s) & 0 & 0 & 0 \\
0 & 0 & \dot{\omega}(t_s) & 0 & 0 \\
0 & 0 & 0 & o_l(t_s) & -o_r(t_s)
\end{bmatrix}
\]

\[(44) \]

Similarly, we investigate the existence of the non-zero vector $k$ such that $\mathcal{O}_c k = \mathbf{0}$. If such a vector $k = [k_1 \ k_2 \ k_3 \ k_4 \ k_5]$ exists, all of the following must be satisfied:

\[-\dot{\omega}(t_1) + o_l(t_1)k_4 = 0, \dot{\omega}(t_3)k_3 = 0, o_r(t_1)k_4 - o_l(t_1)k_5 = 0\]

which requires one of the following conditions to be true:

- $o_l(t)$ is constantly zero, $k = [0 \ \rho_1 \ 0 \ \rho_2 \ 0]^T$, or $o_r(t)$ is constantly zero, $k = [0 \ 0 \ 0 \ 0 \ \rho]^T$.
- $\dot{\omega}(t)$ is constantly zero, $k = [\rho_1 \ 0 \ \rho_2 \ 0 \ 0]^T$.
- $o_r(t)$, $o_l(t)$, and $\dot{\omega}(t)$ are all constants, $k = [0 \ 0 \ \rho \ 0 \ 0]^T$.
- $o_l(t)$ keeps proportional to $o_r(t)$, $k = [0 \ 0 \ 0 \ \rho \ 0]^T$. 

• $\dot{\omega}(t)$ keeps proportional to $\alpha(t)$, $k = [\rho \omega / \omega \ 0 \ 0 \ 0]^T$.

where $\rho, \rho_1, \rho_2$ can be any non-zero value that is used to generate valid non-zero vector $k$ such that $O_c k = 0$. All the above cases are special conditions. Therefore, in a skid-steering robot localization system equipped with a camera, an IMU, and odometers, $\xi$ is observable unless entering the specified special conditions listed above. This completes the proof.

F. Observability of $\xi$ with a Monocular Camera, an IMU, an Odometer and with Online Extrinsic Calibration

It is essential to know the extrinsic transformations between different sensors in a multi-sensor fusion system. Since IMU-camera extrinsic parameters are widely investigated in the existing literature, and practically the IMU-camera system is frequently manufactured as an integrated sensor suite, we here focus on camera-odometer extrinsic parameters.

Since the camera system and wheels of a mobile robot are different hardware components, extrinsic calibration between the corresponding frames is essential for sensor fusion. In this section, we investigate the possibility of performing online extrinsic parameter calibration by including them into the state vector and estimating along with other variables of interests Li and Mourikis (2014b); Censi et al. (2013). To allow this algorithm to function properly, the corresponding extrinsic parameters must be observable, which we investigate as follows.

We first define the parameter state when camera-odometer extrinsics are included:

$$\eta = \begin{bmatrix} X_v & Y_l & Y_r & \alpha_l & \alpha_r & O_{xC} & O_{yC} & O_{zC} & \mathbf{O}_{C} \mathbf{\delta}_T \end{bmatrix}^T$$

where $O_{pC} = [O_{xC} \ O_{yC} \ O_{zC}]$ and $O_{C} \mathbf{\delta} \in \mathbb{R}^3$ are the translational and rotational part of the extrinsic transformation between odometry and camera. $O_{C} \mathbf{\delta}$ is error state (or lie algebra increment) of the 3D rotation matrix $O_{C}$.

Since extrinsic translation and rotation components might be subject to different observability properties, we also define state parameters that contain each of them separately:

$$\eta_p = \begin{bmatrix} X_v & Y_l & Y_r & \alpha_l & \alpha_r & O_{xC} & O_{yC} & O_{zC} \end{bmatrix}^T$$

and

$$\eta_o = \begin{bmatrix} X_v & Y_l & Y_r & \alpha_l & \alpha_r & O_{C} \mathbf{\delta}_T \end{bmatrix}^T$$

To summarize, the objective of this section is to demonstrate the observability properties of $\eta$, $\eta_p$, and $\eta_o$.

**Lemma 4.** By using measurements from a monocular camera, IMU and wheel odometers, $\eta_p$ and $\eta$ are not identifiable. Specifically, the vertical direction of translation in the extrinsics, $O_{zC}$ is always unidentifiable for any type of ground robot, and $O_{xC}$ and $O_{yC}$ become unidentifiable if the skid-steering kinematic parameters are estimated online.

**Lemma 5.** By using measurements from a monocular camera, IMU and wheel odometers, $\eta_o$ is identifiable except the third dimension of the rotation between camera and odometer.

**Proof.** First of all, $\eta$ is locally identifiable if and only if $\eta$ is locally identifiable:

$$\eta = \begin{bmatrix} \dot{Y}_l & \Delta Y & X_v & \beta_l & \beta_r & O_{xC} & O_{yC} & O_{zC} & \mathbf{O}_{C} \mathbf{\delta}_T \end{bmatrix}^T$$

By substituting $\dot{\mathbf{v}}(t) = \mathbf{O}_{C} R \cdot \dot{\mathbf{v}}_C(t)$ in Eq. 28, we are able to derive constraints similar to Eq. 42, as

$$c_{x}(\eta, t) = \dot{\omega}(t) O_{yC} + e_1^T \mathbf{O}_{C} R \dot{\mathbf{v}}_C(t) - \omega Y_l - \beta_l \Delta Y \alpha_l(t) = 0 \quad (45a)$$

$$c_{y}(\eta, t) = -\dot{\omega}(t) O_{xC} + e_2^T \mathbf{O}_{C} R \dot{\mathbf{v}}_C(t) + \omega(t) X_v = 0 \quad (45b)$$

$$c_{z}(\xi, t) = \omega(t) + \beta_l \alpha_l(t) - \beta_r \alpha_r(t) = 0 \quad (45c)$$

Considering the constraints in a set of time instants $S = \{t_0, t_1, \ldots, t_s\}$, we compute the following observability matrices for the systems with calibrating $\eta_p$ and $\eta_o$, given by Eq. 46 and Eq. 47a, respectively. Eq. 47a can be converted to Eq. 47b by linear operations:

$$O_{c}(.; 4) \leftarrow O_{c}(.; 4) - \Delta Y / \beta_l O_{c}(.; 2)$$

$$O_{c}(.; 2) \leftarrow -O_{c}(.; 2) / \beta_l$$

Similar to previous proofs, to look into the properties of $O_{c}$ in Eq. 46, we investigate non-zero vector $k$ such that $O_{c} k = 0$. We can easily find the following non-zero solutions:

$$k_1 = [\rho \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]^T$$

$$k_2 = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]^T$$

$$k_3 = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]^T$$

where $\rho$ can be any non-zero value. We can find that $k_1$, $k_2$ are related with the kinematic parameters, while $k_3$ always holds, which results from no constraints on $O_{zC}$ for ground robots and it has no matter with the kinematic parameters. Through the found null spaces, we can draw the following conclusions: (i) $Y_l$ and $O_{yC}$ are indistinguishable; (ii) $X_v$ and $O_{xC}$ are indistinguishable; (iii) the vertical direction of extrinsic parameters $O_{zC}$ is always unidentifiable for skid-steering robot moving on ground, no matter whether the kinematic parameters are calibrated online.

However, $O_{c}$ in Eq. 47b is under quite different properties. Similarly, we investigate the non-zero $k$ that satisfies $O_{c} k = 0$, which requires the all of the following to be true:

$$-\dot{\omega}(t) k_1 + \alpha_l(t) k_2 - e_1^T \mathbf{O}_{C} R \dot{\mathbf{v}}_C(t_0) \cdot k_0 e_1 = 0, k_0 \cdot 0 = 0$$

$$\dot{\omega}(t) k_3 - e_2^T \mathbf{O}_{C} R \dot{\mathbf{v}}_C(t_0) \cdot k_7 e_2 = 0, \alpha_l(t) k_4 - \alpha_r(t) k_5 = 0$$

However, since $\dot{\omega}(t)$, $\dot{\mathbf{v}}_C(t)$, $\alpha_l(t)$, $\alpha_r(t)$ are time variant under general motion, we can only find such a non-zero vector $k = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]^T$ where $\rho$ can be any non-zero value. We can draw the conclusion: (iv) the first two dimensions of rotation between camera and odometer are identifiable under general motion.

Based on the derived observability properties, it is important to point out the following algorithm design issues: (i) Unlike online calibration algorithms in other literature (Li and Mourikis, 2014b; Qin and Shen, 2018; Schneider et al., 2019), extrinsic parameters between camera and odometer are not
For "localization" sensors, we used a ‘localization’ sensors and ‘ground-truth sensors’ equipped. robots based on the commercially available Clearpath Jackal different settings.

In our experiments, we used two testing skid-steering robots based on the commercially available Clearpath Jackal robot (Clearpath Robotics Inc., 2019) (see Fig. 1), with both ‘localization’ sensors and ‘ground-truth sensors’ equipped. For ‘localization’ sensors, we used a 10Hz monocular global shutter camera at a resolution of 640 × 400, a 200Hz Bosch BMI160 IMU, and 100Hz wheel encoders. The ‘ground truth’ sensor mainly relies on RTK-GPS with centimeter-level precision. All sensors used in our experiment are synchronized by hardware and calibrated offline via (Chen et al., 2019). We note that the offline calibration procedure is an important prerequisite in our experiments since the extrinsic position between the odometer and camera has shown to be constantly unobservable. All the experiments are conducted by first dataset collection and subsequently offline processing using an Intel Core i7-8700 @ 3.20GHz CPU, to allow repeatable comparison between different methods.

### VI. Experimental Results

In this section, we provide experimental results that support our claims in both algorithm design and theoretical analysis. Specifically, we conducted real-world experiments and simulation tests to demonstrate 1) the advantages and necessities of online estimating kinematic parameters in visual (inertial) localization systems and 2) the observability and convergence properties of the skid-steering kinematic parameters under different settings.

In our experiments, we used two testing skid-steering robots based on the commercially available Clearpath Jackal robot (Clearpath Robotics Inc., 2019) (see Fig. 1), with both ‘localization’ sensors and ‘ground-truth sensors’ equipped. For ‘localization’ sensors, we used a 10Hz monocular global shutter camera at a resolution of 640 × 400, a 200Hz Bosch BMI160 IMU, and 100Hz wheel encoders. The ‘ground truth’ sensor mainly relies on RTK-GPS with centimeter-level precision. All sensors used in our experiment are synchronized by hardware and calibrated offline via (Chen et al., 2019). We note that the offline calibration procedure is an important prerequisite in our experiments since the extrinsic position between the odometer and camera has shown to be constantly unobservable. All the experiments are conducted by first dataset collection and subsequently offline processing using an Intel Core i7-8700 @ 3.20GHz CPU, to allow repeatable comparison between different methods.

#### A. Real-world Experiment

In the first set of experiments, we focus on validating the effectiveness of the proposed skid-steering model as well as the localization algorithm. Specifically, we investigated the localization accuracy by estimating skid-steering kinematic parameters \( \xi \) (Eq. 4) online and compared that to the competing methods. To demonstrate the generality of our method, we conducted experiments under various environmental conditions. As shown in Fig 3, the environments involved in our robotic data collection include (a) lawn, (b) cement brick, (c) wooden bridge, (d) muddy road, (e) asphalt road, (f) ceramic tiles, (g) carpet, and (h) wooden floor. Also, Fig. 4 shows one representative trajectory and visual features estimated by the proposed method on a selected sequence, e.g., “SEQ20-CP01”, in which our robot traversed both outdoors and indoors.

---

2 We point out that, we used customized wheel encoder hardware instead of the on-board one on Clearpath robot, to allow accurate hardware synchronization between sensors.
We note that since GPS signal is not always available in all tests (e.g., indoor tests), we use both final drift and root-mean-squared error (RMSE) of absolute translational error (ATE) (Zhang and Scaramuzza, 2018) as our metrics. To make this possible, we started and terminated each experiment in the same position. It is also important to point out that, in the research community, it is preferred to use publicly-available datasets to conduct experiments to facilitate comparison between different researchers. However, most localization datasets publicly available either utilize passenger cars (KITTI (Geiger et al., 2013), Kaist Complex Urban (Jeong et al., 2019), Oxford Robotcar (Maddern et al., 2017)) or lack of one or multiple synchronized low cost sensors (NCLT (Carlevaris-Bianco et al., 2016), and Canadian 3DMap (Tong et al., 2013)). To this end, we also plan to release a comprehensive dataset specifically with low-cost sensors, as our future work.

1) Localization Accuracy: We first conducted an experiment to show the benefits gained by modeling and estimating skid-steering parameters online. In this experiment, three sets of setup are compared, i.e., two provably observable methods and one baseline method. Specifically, those methods are 1) VIO W/ $\xi$: using measurements from a monocular camera, an IMU, and odometer via the proposed estimator by estimating the full 5 skid-steering kinematic parameters $\xi$ online; 2) VO W/ ICR: using monocular camera and odometer measurements (without an IMU), and performing localization by estimating the 3 ICR parameters $\xi_{ICR}$ online; 3) VIO W/O $\xi$: using measurements from monocular camera, an IMU and odometer, and utilizing Eq. 7 for localization without explicitly modeling $\xi$. We note that, in traditional methods when $\xi$ is not modeled, Eq. 7 can be considered as one-parameter (i.e., $b$) approximation of skid-steering kinematics. We also point out that, compared to the third setup, the configuration VO W/O ICR (using a monocular camera, odometer measurements, and using Eq. 7 for localization without modeling $\xi_{ICR}$)
claim that, in order to use odometer measurements of skid-steering robots, the complicated mechanism must be explicitly modelled to avoid accuracy loss. It should be noted all the runs in Table I started from the same initial guess for \( \xi \), to ensure fair comparisons. We also note that, the method of using an IMU and estimating the full 5 kinematic parameters performs best among those methods, by modeling the time-varying scale factors. In fact, the method of estimating only 3 ICR parameters with visual and odometer sensors works well for a portion of the dataset while fails in others (e.g., the datasets under (b,f) categories). This is due to the fact that those datasets involve terrain conditions changes, and the scale factor also changes. If those factors are not modelled, the performance will drop. Moreover, we note that, under those conditions (e.g., (b,f)), the best performing method still works not as good as the performance in other data sequences. This
In this section, we examine the advantages of online estimating the full kinematics parameters $\xi$ in the kinematics-constrained VIO systems, which enables life-long high-precision localization for real-world robotic deployment. In fact, the mechanical parameters of a real robot can not be constants all the time. Some components might be of slow changes (e.g., height and width), and some drift relatively fast (e.g., weights or tire pressures). All those factors might lead to varying kinematic parameters, and we here verify the effectiveness of our method for handling them.

Specifically, we collected datasets under the following practically commonly-seen configurations for our skid-steering robot: (i) normal; (ii) carrying a package with the weight around 3 kg; (iii) under low tire pressure; (iv) carrying a 3-kg package and with low tire pressure. In configurations (i) to (iv), the actual kinematic parameters vary significantly and also deviate from our initial guess. Similar to the previous tests, three algorithms were conducted here by using the measurements from a camera, an IMU and odometer: 1) estimating $\xi_{ICR}$; 2) estimating $\xi$; 3) used fixed $\xi$ with a relatively good initial guess, obtained by the final estimate of running our online estimation algorithm.

We conducted experiments on 8 sequences named ABL-SEQ1 $\sim$ ABL-SEQ8, and each configuration corresponds to two sequences in ascending order (e.g., ABL-SEQ1 and ABL-SEQ2 correspond to the ‘normal’ condition). The evaluation methods used here are as same as the ones used in the

**TABLE II: RMSE of ATE (m) on the Sequences with RTK-GPS Measurements.**

| Segment Length (m) | Sequence Names | VO W/ $\xi$ | VO W/ ICR | VIO W/O $\xi$ |
|-------------------|----------------|-------------|-----------|---------------|
| 15.00             | SEQ21-CP01     | 1.36        | 1.39      | 1.02          |
| 30.00             | SEQ22-CP01     | 1.59        | 1.72      | 1.36          |
| 45.00             | SEQ23-CP01     | 1.84        | 2.00      | 1.59          |
| 60.00             | SEQ24-CP01     | 2.21        | 2.17      | 1.69          |
| 75.00             | SEQ25-CP01     | 2.63        | 2.58      | 1.95          |
| 90.00             | SEQ26-CP01     | 3.00        | 2.86      | 2.18          |
| 105.00            | SEQ27-CP01     | 3.36        | 3.25      | 2.44          |

**TABLE III: Mean of RPE (m) for Different Segment Length on the Sequences with RTK-GPS Measurements.**

| Segment Length (m) | VO W/ $\xi$ | VO W/ ICR | VIO W/O $\xi$ |
|-------------------|-------------|-----------|---------------|
| 15.00             | 0.79        | 1.07      | 0.69          |
| 30.00             | 1.11        | 1.35      | 0.82          |
| 45.00             | 1.42        | 1.64      | 1.00          |
| 60.00             | 1.75        | 2.00      | 1.43          |
| 75.00             | 2.00        | 2.25      | 1.70          |

is due to the fact that we used ‘random walk’ process to model the ‘environmental condition’ changes, which is not the ‘best’ assumption when there are rapid road surface changes. We will also leave the terrain detection as future work.

In some sequences where GPS signals were available across the entire data sequence, we also evaluated the root mean square errors (RMSE) (Bar-Shalom and Fortmann, 1988) of absolute translational error (ATE) (Zhang and Scaramuzza, 2018). To compute that, we interpolated the estimated poses to get the ones corresponding to the timestamp of the GPS measurements. The RMSE errors are shown in Table II, where we highlight the bad results (over 12m) by red color. The results demonstrate that estimating $\xi$ is beneficial for trajectory tracking. In order to provide insight into how the error of each algorithm grows with the trajectory length, we also calculate the calculated relative pose error (RPE) averaged over all the sequences when GPS measurements are available. The RPE results are shown in Table III and Fig. 6, which also support our algorithm claims.

**2) Convergence of Kinematic Parameters:** In this section, we show experimental results to demonstrate the convergence properties of $\xi$ and $\xi_{ICR}$, in systems that we theoretically claim observable. Unlike the experiments in the previous section, which utilized the method described in Sec. IV-C for kinematic parameter initialization, we manually added extra errors to the kinematic parameter for the tests in this section, to better demonstrate the observability properties. Specifically, for the kinematics-constrained VIO system, we added the following extra error terms to initial kinematic parameters

$$\delta X = 0.08, \delta Y = 0.14, \delta R = -0.1, \delta \alpha_l = 0.2, \delta \alpha_r = 0.2$$

For the kinematics-constrained VO system, we only add error terms to $\xi_{ICR}$. To show details in parameter convergence properties, We carried out experiments on representative indoor and outdoor sequences, “SEQ8-CP01, SEQ18-CP01, SEQ19-CP01”. In Fig. 7, the estimates of the full kinematic parameters $\xi$ in VIO are shown, along with the corresponding uncertainty envelopes. The convergence of $\xi_{ICR}$ in VO are also shown in Fig. 8. The results demonstrate that the kinematic parameters $\xi$ in the VIO quickly converge to stable values, and remains slow change rates for the rest of the trajectory. Similar behaviours can also be observed for $\xi_{ICR}$ when only a monocular camera and odometer sensors are used. The results exactly meet our theoretical expectations that $\xi$ in VIO and $\xi_{ICR}$ in VO are both locally identifiable under general motion. We also note that, since it is not feasible for obtaining high-precision ground truth for $\xi$, the correctness of those values cannot be ‘directly’ verified. Instead, they can be evaluated either based on the overall estimation results shown in the previous section or simulation results in Sec. VI-B where ground truth $\xi$ is known.

**3) Ablation Study:** In this section, we examine the advantages of online estimating the full kinematics parameters $\xi$ in the kinematics-constrained VIO systems, which enables life-long high-precision localization for real-world robotic deployment. In fact, the mechanical parameters of a real robot can not be constants all the time. Some components might be of slow changes (e.g., height and width), and some drift relatively fast (e.g., weights or tire pressures). All those factors might lead to varying kinematic parameters, and we here verify the effectiveness of our method for handling them.

Specifically, we collected datasets under the following practically commonly-seen configurations for our skid-steering robot: (i) normal; (ii) carrying a package with the weight around 3 kg; (iii) under low tire pressure; (iv) carrying a 3-kg package and with low tire pressure. In configurations (i) to (iv), the actual kinematic parameters vary significantly and also deviate from our initial guess. Similar to the previous tests, three algorithms were conducted here by using the measurements from a camera, an IMU and odometer: 1) estimating $\xi_{ICR}$; 2) estimating $\xi$; 3) used fixed $\xi$ with a relatively good initial guess, obtained by the final estimate of running our online estimation algorithm.

We conducted experiments on 8 sequences named ABL-SEQ1 $\sim$ ABL-SEQ8, and each configuration corresponds to two sequences in ascending order (e.g., ABL-SEQ1 and ABL-SEQ2 correspond to the ‘normal’ condition). The evaluation methods used here are as same as the ones used in the
previous section, which include both final drift and RMSE. In Table. IV, we show the final drift of three different localization methods. On the other hand, the RMSE of ATE is given in Table. V. Additionally, RPE was shown in Fig. 9 and Table. VI. Those results demonstrate that, when a robot is in normal mechanical condition, and the road condition is without large variance, there are minor differences between estimating the full 5 kinematic parameters \( \xi \) and online estimating only the 3 parameters \( \xi_{ICR} \), when good correction factors \( \alpha_l, \alpha_r \) are given and kept constant. This is due to the fact that the correction factors reflect the transmission efficiency of the robot and are not subject to fast changes in the general case. However, if there are noticeable changes in the robotic mechanical condition, e.g., weight and center of mass change by carrying a large package or tire pressure changes after long-term usage, the correction factors \( \alpha_l, \alpha_r \) will be changed significantly. In such cases, the overall estimation algorithm benefits significantly by online estimating \( \xi \). We also show the estimated trajectories compared with RTK-GPS measurement in Fig. 10, for the representative runs.

**B. Simulation Experiments**

We also perform Monte-Carlo simulations to investigate our proposed method specifically for parameter calibration precision, since this cannot be verified in real-world tests. The synthetic trajectory is generated by simulating a real-world trajectory with a length of 205.4m, using the method introduced in (Li and Mourikis, 2014c). To generate noisy sensory measurements, we have used zero-mean Gaussian vector for all sensors with the following standard deviation (std) values. Pixel std for visual measurements is 0.6 pixels, odometer stds for the left and right wheels are both 0.0245 m/s, gyroscope and accelerometer measurement stds are \( 9 \cdot 10^{-4} \) rad/s and \( 1 \cdot 10^{-2} \) m/s^2, and finally the stds representing the random walk behavior of gyroscope and accelerometer biases are \( 1 \cdot 10^{-2} \) rad/s^2 and \( 1 \cdot 10^{-2} \) m/s^3 respectively. Additionally,
Fig. 10: In the ablation study, the trajectories of RTK-GPS (ground truth) and the estimated trajectory by the localization method: 1) VIO with online estimating the full kinematic parameters $\xi$; 2) VIO with online estimating $\xi_{ICR}$ only; 3) VIO with fixed $\xi$. The skid-steering robot is under four different conditions: (a) normal; (b) carrying a package with the weight around 3 kg; (c) under low tire pressure; (d) carrying a 3-Kg package and with low tire pressure.

TABLE IV: Ablation Experiments Results: Final drift.

| Sequence | Length(m) | Terrain Config. | Norm(m) x(m) y(m) z(m) | Norm(m) x(m) y(m) z(m) | Norm(m) x(m) y(m) z(m) |
|----------|-----------|-----------------|-------------------------|-------------------------|-------------------------|
| ABL-SEQ1 | 167.30    | (e) (i)         | 0.316 0.038 0.167       | 0.304 0.038 0.144       | 0.365 0.029 0.265       |
| ABL-SEQ2 | 147.76    | (e) (i)         | 0.349 -0.109 0.235      | 0.336 -0.110 0.216      | 0.564 -0.190 0.495      |
| ABL-SEQ3 | 152.23    | (e) (ii)        | 0.318 -0.193 -0.195     | 0.311 -0.192 -0.183     | 0.819 -0.240 -0.771     |
| ABL-SEQ4 | 152.80    | (e) (ii)        | 0.406 -0.240 -0.252     | 0.400 -0.240 -0.243     | 0.629 -0.185 -0.578     |
| ABL-SEQ5 | 237.36    | (e) (iii)       | 8.700 0.046 -8.699      | 7.222 -0.069 -7.220     | 9.755 0.037 -9.753      |
| ABL-SEQ6 | 232.43    | (e) (iii)       | 8.102 -0.161 -8.098     | 6.696 -0.242 -6.689     | 8.846 -0.143 -8.842     |
| ABL-SEQ7 | 232.54    | (e) (iv)        | 9.509 0.189 -9.505      | 7.771 0.053 -7.769      | 10.502 0.161 -10.498    |
| ABL-SEQ8 | 233.07    | (e) (iv)        | 10.055 0.256 -10.050    | 8.204 0.080 -8.202      | 10.601 0.173 -10.598    |
| Mean     |           |                 | 4.719 -0.022 -4.550     | 3.906 -0.085 -3.743     | 5.319 -0.045 -5.291     |

TABLE V: Ablation Experiments Results: RMSE of ATE (m).

|                | VIO W/ ICR | VIO W/ $\xi$ | VIO W/ Fixed $\xi$ |
|----------------|------------|--------------|--------------------|
| ABL-SEQ1       | 0.24       | 0.23         | 0.28               |
| ABL-SEQ2       | 0.17       | 0.17         | 0.24               |
| ABL-SEQ3       | 0.14       | 0.14         | 0.27               |
| ABL-SEQ4       | 0.16       | 0.16         | 0.23               |
| ABL-SEQ5       | 2.20       | 1.85         | 2.49               |
| ABL-SEQ6       | 2.20       | 1.87         | 2.40               |
| ABL-SEQ7       | 2.34       | 2.13         | 2.82               |
| ABL-SEQ8       | 2.54       | 2.09         | 2.75               |
| Mean           | 1.27       | 1.08         | 1.44               |

TABLE VI: Ablation Experiments Results: Mean of RPE (m) for Different Segment Length.

| Segment Length (m) | VIO W/ ICR | VIO W/ $\xi$ | VIO W/ Fixed $\xi$ |
|--------------------|------------|--------------|--------------------|
| 9.00m              | 0.53       | 0.48         | 0.54               |
| 18.00m             | 1.00       | 1.33         | 1.49               |
| 27.00m             | 1.47       | 1.75         | 1.95               |
| 36.00m             | 1.92       | 2.14         | 2.39               |

since skid-steering kinematic parameters cannot be known in advance, we initialize $\xi$ in our simulation tests by adding an error vector to the ground truth values. The noise vector is sampled from zero-mean Gaussian distribution with std $8\cdot10^{-2}$ for all elements in $\xi$.

To collect algorithm statistics, we conducted 15 Monte-Carlo tests and compute parameter estimation results for $\xi$. Specifically, we computed the mean and std of calibration errors for all elements in $\xi$, averaged from the Monte-Carlo tests. The results for each element in $\xi$ are: $-0.0211 \pm 0.0095$, $0.0102 \pm 0.0030$, $-0.0081 \pm 0.0026$, $0.0212 \pm 0.0109$, $0.0216 \pm 0.0108$. Those results indicate that, the skid-steering parameters can be accurately calibrated by significantly reducing uncertainty values. It is also interesting to look into a representative run, in which the initial estimate of $\xi$ is subject to the following error vector $\delta\xi = [0.15 0.15 -0.15 0.1 0.1]$. In this case, the calibration errors averaged over the second half of the trajectory are: $-0.0276 \pm 0.0067$, $0.0199 \pm 0.0118$, $0.0054 \pm 0.0026$, $0.0192 \pm 0.0157$, $0.0189 \pm 0.0157$.

Since simulation tests provide absolute ground truth, it is also interesting to investigate the accuracy gain by estimating $\xi$ online. Fig. 11 demonstrates the estimated trajectory when $\xi$ is estimated online, or $\xi$ is fixed during estimation as well as the ground truth. This clearly demonstrates that, by the online estimation process, the localization accuracy can be significantly improved. The averaged RMSE of rotation and translation for those two competing methods in this
Monte-Carls tests are $0.042 \pm 0.023 rad$, $2.051 \pm 0.830 m$ and $0.154 \pm 0.0635 rad$, $4.617 \pm 2.563 m$, respectively.

Fig. 11: In simulation experiments, estimated Trajectories aligned with the ground truth trajectory.

VII. CONCLUSIONS

In this paper, we propose a novel kinematics-constrained visual localization method specialized for skid-steering robots, where multi-modal measurements are fused in a tightly-coupled sliding-window BA. In particular, in order to compensate for the complicated track-to-terrain interactions, the imperfectness of mechanical design, and terrain conditions, we explicitly model the kinematics of skid-steering robots by using both track ICRs and correction factors, which are online estimated along with the other states of interests. To ensure reliable localization performance, we conduct detailed observability analysis for the proposed algorithm under different setup conditions. Specifically, we show that the kinematic parameter vector $\xi$ is observable under general motion when measurements from an IMU are added and odometer-to-camera extrinsic parameters are calibrated offline. In other situations, degenerate cases might be entered and reduced precision might be incurred. Extensive real-world experiments and simulation tests are also provided, which demonstrate that the proposed method is able to compute skid-steering parameters online and yield accurate localization results. Experimental results also validate our observability analysis, showing that under theoretically observable conditions the corresponding parameters can converge quickly.

APPENDIX

To make this article self-contained, we also provide detailed formulation for each term in our cost function.

A. Camera Cost Function

In the sliding-window BA, only the keyframes are optimized for computational saving. We use a simple heuristic for keyframe selection: the odometer prediction has a translation or rotation over a certain threshold (in all the experiments, 0.2 meter and 3 degrees). Since the movement form of the ground robot is simple, and it can be well predicted by the odometer in a short period of time. Unlike existing methods (Qin et al., 2018; Leutenegger et al., 2015), which extract features and analyze the distribution of the features for keyframe selection, the non-keyframe will be dropped immediately without any extra operations in our framework. Among keyframes selected into the sliding-window, corner feature points are extracted in a fast way (Rosten and Drummond, 2006) and tackled with FREAK (Alahi et al., 2012) descriptors.

The successfully tracked features across multiple keyframes will be initialized in the 3D space by triangulation. By denoting $z_{i,j}$ the visual measurement of a 3D feature $G\cdot p_{f_j}$ observed by the $C_i$th camera keyframe, the visual reprojection error (Hartley and Zisserman, 2003) in normalized image coordinate is given by:

$$C_v(G, R, G\cdot p_{O}) = || z_{i,j} - \pi(G, R, G\cdot p_{C_i}, p_{f_j}) ||^2_{A_V},$$

$$G\cdot R = G_{O_i} R G_{C}, \quad G\cdot p_{C_i} = G\cdot p_{O_i} + G_{O_i} R G_{C} p_C$$

In the above expressions, $\pi(\cdot)$ denotes the perspective function of an intrinsically calibrated camera, and $A_V$ represents the inverse of noise covariance in the observation $z_{i,j}, G\cdot R$ and $G\cdot p_{C}$ are the extrinsic transformation between camera and odometer, which are offline calibrated and remain constant during the online estimation process. We also note that since the camera to odometer extrinsic parameters are shown to be unobservable in Sec. V-F, they can not be included in the online process. Furthermore, we chose not to incorporate visual feature into the state vector due to the limited computational resources. Thus $G\cdot p_{f_j}$, and the pose of oldest keyframe over the sliding window will be marginalized immediately after the iteratively minimization.

B. IMU Constraints

The IMU provides readings of both accelerometer and gyroscope as follows:

$$\omega_m = \omega + b_\omega + n_\omega$$  \hspace{1cm} (49a)

$$a_m = a - \frac{1}{G} R G_{c} g + b_a + n_a$$  \hspace{1cm} (49b)

where $G\cdot g$ is the known global gravity vector, $b_\omega$ and $b_a$ the time-varying gyro and accelerometer bias vectors, and $n_\omega$ and $n_a$ denote white Gaussian measurement noise. The IMU integration process is characterized by:

$$x_{1k} = \begin{bmatrix} G\cdot p_{O_{1k}}, G\cdot q_{1k}^T, G\cdot q_{P_{1k}}, \hat{b}_{\omega_{1k}}, \hat{\omega}_{1k} \\ \end{bmatrix}^T = f(x_{1k-1}, W_m, A_m)$$  \hspace{1cm} (50)

where $W_m, A_m$ are the gyroscope and accelerometer measurements during the time interval $t \in (t_{k-1}, t_k)$, and $f(\cdot)$ is the IMU integration function. Since the IMU integration is widely investigated in research communities (Mourikis and Roumeliotis, 2007; Li and Mourikis, 2013; Eckenhoff et al., 2019), and we here ignore the details on $f(\cdot)$. The associated uncertainty matrix (i.e., linearized noise information matrix) of the prediction process $A_f$ can also be obtained by linearizing
the function \( f(\cdot) \). As a result, the IMU cost term can be summarized by:

\[
C_I(x_{k}, x_{k-1}) = \left\| x_{k} \oplus f(x_{k-1}, \mathcal{W}_m, \mathcal{A}_m) \right\|_\Lambda^2
\]

(51)

which provides pose constraints between consecutive keyframes. We also note that, the IMU cost function requires odometer to IMU extrinsic parameters to transform states in odometer frame to IMU frame, which are also calibrated offline. After minimizing Eq. 51, the states \( \hat{\mathbf{v}}^T_{t-1}, \hat{\mathbf{b}}^T_{a_{t-1}}, \hat{\mathbf{b}}^T_{w_{t-1}} \) will be marginalized, and the contained information will be incorporated into the prior cost term.

C. Motion Mainfold constraints

Finally, since the skid-steer robot navigates on ground surfaces, its trajectories can also be constrained by the prior knowledge about the shape of surface manifold. Specifically, we utilize our method presented in (Zhang, Chen and Li, 2019b) to approximate ground surfaces using quadratic polynomials, the following holds:

\[
m_p(G_{PO}) = \frac{1}{2} G_{PO}^T \begin{bmatrix} a_1 & a_2 & a_3 \\ a_2 & b_1 & b_2 \\ a_3 & b_2 & c \end{bmatrix} G_{PO} + \begin{bmatrix} b_1 \\ b_2 \\ c \end{bmatrix}^T \quad (52)
\]

where \( m = [a_1, a_2, a_3, b_1, b_2, c]^T \) to denote the manifold parameters. By utilizing the quadratic surface approximation, we are able to define the following cost function for both rotation and position terms (Zhang, Chen and Li, 2019b):

\[
m_r(G_{OQ}, G_{PO}) = \left( G_{OQ} e_3 \right)^T_{12} \frac{\partial m_p}{\partial G_{PO}} = 0, \quad \text{and} \quad m_p(G_{PO}) = 0
\]

where \( |v|_{12} \) denotes the first and second rows of a symmetric matrix of the 3D vector \( v \). The above constraints reflect the fact that, the motion manifold \( m \) has explicitly defined roll and pitch of a ground robot, which should be in consistent with the rotation \( G_{OQ} \). Therefore, the motion manifold cost term for keyframes in the sliding window can be written as:

\[
C_M(G_{Q}, G_{PO}, m_k, m_{k-1}) = \left\| \begin{bmatrix} m_k - m_{k-1} \\ m_p(G_{PO}) \\ m_r(G_{OQ}, G_{PO}) \end{bmatrix} \right\|_{\Lambda_m}
\]

(53)

for all \( i \in [k-s+1, k] \). \( m_{k-1} \) and \( m_k \) denotes the manifold parameters characterize the motion parameters across the last and the current sliding window, respectively. Moreover, \( \Lambda_m \) is the information matrix describing the uncertainties in both localization states and the surface manifold approximation itself, which is described in detail in (Zhang, Chen and Li, 2019b).

**REFERENCES**

Achtelik MW, Achtelik M, Weiss S and Siegwart R (2011) Onboard IMU and monocular vision based control for MAVs in unknown in-and outdoor environments. In: IEEE International Conference on Robotics and Automation.

Alahi A, Ortiz R and Vandergeheynst P (2012) Freak: Fast retina keypoint. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 510–517.

Anousaki G and Kyriakopoulos KJ (2004) A dead-reckoning scheme for skid-steered vehicles in outdoor environments. In: IEEE International Conference on Robotics and Automation, volume 1. pp. 580–585.

Bar-Shalom Y and Fortmann TE (1988) Tracking and Data Association. New York: Academic Press.

Bloesch M, Omari S, Hutter M and Siegwart R (2015) Robust visual inertial odometry using a direct ekf-based approach. In: IEEE/RSJ international conference on intelligent robots and systems (IROS). pp. 298–304.

Carlevaris-Bianco N, Ushani AK and Eustice RM (2016) University of michigan north campus long-term vision and lidar dataset. The International Journal of Robotics Research 35(9): 1023–1035.

Censi A, Franchi A, Marchionni L and Oriolo G (2013) Simultaneous calibration of odometry and sensor parameters for mobile robots. IEEE Transactions on Robotics 29(2): 475–492.

Chen Y, Zhang M, Hong D, Deng C and Li M (2019) Perception system design for low-cost commercial ground robots: Sensor configurations, calibration, localization, and mapping. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE.

Clearpath Robotics Inc (2019) Clearpath ground vehicle. Available: https://www.clearpathrobotics.com/jackal-small-unmanned-ground-vehicle/.

Davison AJ, Reid ID, Molton ND and Stasse O (2007) Monoslam: Real-time single camera slam. IEEE Transactions on Pattern Analysis & Machine Intelligence (6): 1052–1067.

Do T, Neira L, Yang Y and Roumeliotis S (2019) Attitude tracking from a camera and an accelerometer on gyroless devices. In: The International Symposium on Robotics Research (ISRR).

Eckenhoff K, Geneva P and Huang G (2019) Closed-form preintegration methods for graph-based visual-inertial navigation. International Journal of Robotics Research 38(5): 563–586.

Engel J, Koltun V and Cremers D (2017) Direct sparse odometry. IEEE transactions on pattern analysis and machine intelligence 40(3): 611–625.

Engel J, Schöps T and Cremers D (2014) LSD-SLAM: Large-scale direct monocular SLAM. In: European Conference on Computer Vision. pp. 834–849.

Forster C, Pizzoli M and Scaramuzza D (2014) Svo: Fast semi-direct monocular visual odometry. In: IEEE international conference on robotics and automation (ICRA). pp. 15–22.

Fossel JD, Tuyls K and Sturm J (2015) 2D-SDF-SLAM: A signed distance function based SLAM frontend for laser
scanners. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1949–1955.

Ganganath N and Leung H (2012) Mobile robot localization using odometry and kinect sensor. In: *IEEE International Conference on Emerging Signal Processing Applications*. pp. 91–94.

Geiger A, Lenz P, Stiller C and Urtasun R (2013) Vision meets robotics: The kitti dataset. *International Journal of Robotics Research* .

Geiger A, Moosmann F, Car Ő and Schuster B (2012) Automatic camera and range sensor calibration using a single shot. In: *IEEE International Conference on Robotics and Automation*. pp. 3936–3943.

Guo CX, Mirzaei FM and Roumeliotis SI (2012) An analytical least-squares solution to the odometer-camera extrinsic calibration problem. In: *IEEE International Conference on Robotics and Automation*. pp. 3962–3968.

Guo CX and Roumeliotis SI (2013) Imu-rrgb camera 3d pose estimation and extrinsic calibration: Observability analysis and consistency improvement. In: *IEEE International Conference on Robotics and Automation*. pp. 2935–2942.

Hartley R and Zisserman A (2003) *Multiple view geometry in computer vision*. Cambridge university press.

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

Hesch JA, Kottas DG, Bowman SL and Roumeliotis SI (2013) Towards consistent vision-aided inertial navigation. In: *Algorithmic Foundations of Robotics X*. pp. 559–574.

Huang AS, Bachrach A, Henry P, Krainin M, Maturana D, Fox D and Roy N (2017) Visual odometry and mapping for autonomous flight using an rgb-d camera. In: *The International Journal on Robotics Research*. pp. 235–252.

Husnik G, Buck S and Zell A (2017) Path following control of skid-steered wheeled mobile robots at higher speeds on different terrain types. In: *IEEE International Conference on Robotics and Automation (ICRA)*. pp. 3734–3739.

Jeong J, Cho Y, Shin YS, Roh H and Kim A (2019) Complex urban dataset with multi-level sensors from highly diverse urban environments. *The International Journal of Robotics Research* 38(6): 642–657.

Jones ES and Soatto S (2011) Visual-inertial navigation, mapping and localization: A scalable real-time causal approach. *The International Journal of Robotics Research* 30(4): 407–430.

Kelly J and Sukhatme GS (2011) Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration. *The International Journal of Robotics Research* 30(1): 56–79.

Klein G and Murray D (2007) Parallel tracking and mapping for small ar workspaces. In: *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality*. pp. 1–10.

Leutenegger S, Lynen S, Bosse M, Siegwart R and Furgale P (2015) Keyframe-based visual–inertial odometry using non-linear optimization. *The International Journal of Robotics Research* 34(3): 314–334.

Li M and Mourikis AI (2013) High-precision, consistent EKF-based visual-inertial odometry. *The International Journal of Robotics Research* 32(6): 690–711.

Li M and Mourikis AI (2014a) A convex formulation for motion estimation using visual and inertial sensors .

Li M and Mourikis AI (2014b) Online temporal calibration for camera–imu systems: Theory and algorithms. *The International Journal of Robotics Research* .

Li M and Mourikis AI (2014c) Vision-aided inertial navigation with rolling-shutter cameras. *The International Journal of Robotics Research* 33(11): 1490–1507.

Li M, Yu H, Zheng X and Mourikis AI (2014) High-fidelity sensor modeling and self-calibration in vision-aided inertial navigation. In: *IEEE International Conference on Robotics and Automation*. pp. 409–416.

Maddern W, Pascoe G, Linegar C and Newman P (2017) 1 year, 1000km: The oxford robotcar dataset. *The International Journal of Robotics Research* 36(1): 3–15.

Martínez JL, Mandow A, Morales J, Pedraza S and García-Cerezo A (2005) Approximating kinematics for tracked mobile robots. *The International Journal of Robotics Research* 24(10): 867–878.

Martínez JL, Morales J, Mandow A, Pedraza S and García-Cerezo A (2017) Inertia-based icr kinematic model for tracked skid-steer robots. In: *IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*. pp. 166–171.

Moosavian SAA and Kalantari A (2008) Experimental slip estimation for exact kinematics modeling and control of a tracked mobile robot. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp. 95–100.

Mourikis AI and Roumeliotis SI (2007) A multi-state constraint kalman filter for vision-aided inertial navigation. In: *IEEE International Conference on Robotics and Automation*. pp. 3565–3572.

Newswire P (2019) Skid steer loader market. Available: https://finance.yahoo.com/news/skid-steer-loader-market-anticipated-102009939.html.

Nisar B, Foehn P, Falanga D and Scaramuzza D (2019) Vimo: Simultaneous visual inertial model-based odometry and force estimation. *IEEE Robotics and Automation Letters* .

Pentzer J, Brennan S and Reichard K (2014) Model-based prediction of skid-steer robot kinematics using online estimation of track instantaneous centers of rotation. *Journal of Field Robotics* 31(3): 455–476.

Qin T, Li P and Shen S (2018) VINS-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics* 34(4): 1004–1020.

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

Quan M, Piao S, Tan M and Huang SS (2018) Tightly-coupled monocular visual-odometric SLAM using wheels and a mems gyroscope. arXiv:1804.04854 .

Reina G and Galati R (2016) Slip-based terrain estimation
with a skid-steer vehicle. *Vehicle System Dynamics* 54(10): 1384–1404.

Rosten E and Drummond T (2006) Machine learning for high-speed corner detection. In: *European conference on computer vision*. Springer, pp. 430–443.

Rubin BF (2019) Amazon’s scout robots. https://www.cnet.com/news/amazons-scout-robots-thats-no-cooler-thats-your-prime-delivery/.

Scaramuzza D (2011) 1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints. *International journal of computer vision* 95(1): 74–85.

Schneider T, Li M, Cadena C, Nieto J and Siegwart R (2019) Observability-aware self-calibration of visual and inertial sensors for ego-motion estimation. *IEEE Sensors Journal* 19(10): 3846–3860.

Solc F and Sembera J (2008) Kinetic model of a skid steered robot. In: *Proceedings of the International Conference on Signal Processing, Robotics and Automation*. pp. 61–65.

Sutoh M, Iijima Y, Sakakieda Y and Wakabayashi S (2018) Motion modeling and localization of skid-steering wheeled rover on loose terrain. *IEEE Robotics and Automation Letters* 3(4): 4031–4037.

Tong CH, Gingras D, Larose K, Barfoot TD and Dupuis E (2013) The canadian planetary emulation terrain 3d mapping dataset. *The International Journal of Robotics Research* 32(4): 389–395.

Trawny N and Roumeliotis SI (2005) Indirect Kalman filter for 3D pose estimation. *University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep.* 2.

Van Doren JF, Douma SG, Van den Hof PM, Jansen JD and Bosgra OH (2009) Identiﬁability: from qualitative analysis to model structure approximation. *IFAC Proceedings Volumes* 42(10): 664–669.

Vincent J (2019) Fedex unveils autonomous delivery robot. Available: https://www.theverge.com/2019/2/27/18242834/delivery-robot-fedex-sameday-bot-autonomous-trials.

Wong JY (2008) *Theory of ground vehicles*. John Wiley & Sons.

Wu KJ, Guo CX, Georgiou G and Roumeliotis SI (2017) VINS on wheels. In: *IEEE International Conference on Robotics and Automation*. pp. 5155–5162.

Yang Y and Huang G (2019) Observability analysis of aided ins with heterogeneous features of points, lines, and planes. *IEEE Transactions on Robotics*.

Yap T, Li M, Mourikis AI and Shelton CR (2011) A particle filter for monocular vision-aided odometry. In: *IEEE International Conference on Robotics and Automation*. pp. 5663–5669.

Yi J, Wang H, Zhang J, Song D, Jayasuriya S and Liu J (2009) Kinematic modeling and analysis of skid-steered mobile robots with applications to low-cost inertial-measurement-unit-based motion estimation. *IEEE transactions on robotics* 25(5).

Zhang J, Singh S and Kantor G (2014) Robust monocular visual odometry for a ground vehicle in undulating terrain. In: *Field and Service Robotics*. pp. 311–326.

Zhang M, Chen Y and Li M (2019a) SDF-Loc: Signed distance field based 2D relocalization and map update in dynamic environments. In: *American Control Conference (ACC)*. pp. 1997–2004.

Zhang M, Chen Y and Li M (2019b) Vision-aided localization for ground robots. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Zhang M, Zuo X, Chen Y, Liu Y and Li M (2019c) Localization for ground robots: On manifold representation, integration, re-parameterization, and optimization. *arXiv preprint*.

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

Zheng F and Liu YH (2019) Visual-odometric localization and mapping for ground vehicles using se (2)-xyz constraints. In: *International Conference on Robotics and Automation (ICRA)*, pp. 3556–3562.

Zheng X, Moratto Z, Li M and Mourikis AI (2017) Photometric patch-based visual-inertial odometry. In: *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3264–3271.

Zuo X, Huang G and Li M (2019a) Visual localization for skid-steering robots. Technical report. Available: http://udel.edu/~ghuang/papers/tr_icr.pdf.

Zuo X, Zhang M, Chen Y, Liu Y, Huang G and Li M (2019b) Visual-inertial localization for skid-steering robots with kinematic constraints. In: *The International Symposium on Robotics Research (ISRR)*.