Visual-Based Kinematics and Pose Estimation for Skid-Steering Robots

Xingxing Zuo⬤, Mingming Zhang, Mengmeng Wang⬤, Yiming Chen, Guoquan Huang⬤, *Senior Member, IEEE*, Yong Liu⬤, and Mingyang Li⬤, *Member, IEEE*

Abstract— To build commercial robots, skid-steering mechanical design is of increased popularity due to its manufacturing simplicity and unique mechanism. However, these also cause significant challenges on software and algorithm design, especially for the pose estimation (i.e., determining the robot's rotation and position) of skid-steering robots, since they change their orientation with an inevitable skid. 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 imperfectness of mechanical design, terrain conditions and smoothness, etc. To prevent performance reduction in robots' long-term missions, the time- and location- varying kinematic parameters are estimated online along with pose estimation 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.

Note to Practitioners—This paper was motivated by the problem of long-term pose estimation of the commonly commercial-used skid-steering robots with only low-cost sensors. Skid-steering robots change their orientation with a skid, which poses a significant challenge for pose estimation when using the wheel encoders. We propose to online estimate the robot's kinematics, which succeeds in compensating for the complexity of track-to-terrain interaction, due to the slippage, the imperfectness of mechanical design, terrain conditions and smoothness. It is

critical to estimate the kinematics and poses jointly to prevent performance reduction in robots' long-term missions. We further theoretically analyze whether the kinematics parameters can be estimated under different sensor configurations, and find out the special degrade motions that make the parameters unobservable.

Index Terms— Kinematics, pose estimation, visual odometry, skid-steering robots, observability.

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 [3], [4], [5]. 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 [6], [7], [8], [9], [10], [11], [12], [13].

The localization system provides motion estimates, which is a key component for enabling any autonomous robot. To localize skid-steering robots, there is a large body of relevant literature [6], [7], [8], [9], [10], [11], [12], [13]. Early work by Anousaki and Kyriakopoulos [6] 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, Martínez et al. [7] proposed an approach to approximate the kinematics of skid-steering robots based on instantaneous centers of rotation (ICRs). Although some other kinematic models of skid-steering robots are also proposed [11], [12], [14], [15], ICR based kinematics is still popular due to its simplicity and feasibility [9], [10], [13], especially for real-time robotic applications. In [9], IMU readings and wheel encoder measurements are fused in an EKF-based motion-estimation system for skid-steered robots. The ICR-based kinematics is utilized to compute virtual velocity measurements for robot motion estimation. However, the easily changed ICR parameters are not estimated in the estimator, which may lead to performance

reduction. In [10], ICR parameters and navigation states are estimated online in an EKF-based estimator for 3-DoF motion estimation of skid-steering robots. Wheel odometer measurements and GPS measurements are fused in the system. It is also revealed that ICR-based kinematics is only valid in low dynamics, and will have a degraded performance when the vehicle is operated at a high speed. The work [13] leveraged ICR-based kinematics and fused the readings from wheel encoders and a GPS-compass integrated sensor to estimate the ICR parameters and 3-DoF poses of the robot.

In contrast to the above existing works, we estimate the kinematics (formulated by ICRs and correction factors), and *full 6-DoF* poses (3-DoF rotations and 3-DoF translations) of the skid-steering robots jointly in a sliding-window bundle adjustment (BA) based estimator. Extracted visual features from camera and wheel encoder readings, optional IMU readings, are fused to optimize the estimated states in a tightly-coupled way and track the states of the skid-steering robot during its long-term mission. 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, slippage, etc. In the long-term mission of the skid-steering robots, such as delivery, the kinematic parameters can be inevitably changed. Thus we estimate the kinematic parameters online to guarantee accurate pose estimation of the robots in complicated environments without performance reduction.

For a complicated estimator, it is critical to conduct observability analysis [1], [10], [16], [17], [18], [19] to study the identifiability of estimated states. Pentzer et al. [10] 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 our visual-based systems. This paper is *evolved* from our previous conference paper [1], which 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 extensively extend [1] and fully explore the observability properties of the visual-based kinematics and pose estimation system, by explicitly identifying the identifiable and non-identifiable parameters with and without using the IMU. Besides, we further investigate the suitability of kinematics estimation with joint online sensor extrinsic calibration, which is commonly required in state estimation systems with sensor fusion. The results are with significant differences from the properties of other visual localization systems [19], [20], [21]. This emphasizes the importance of the observability analysis in this paper.

In summary, we focus on pose and kinematics estimation of steering-skid robots to enable their long-term mission in complicated environment, by using measurements from a monocular camera, wheel encoders, and optionally an IMU. The main contributions are as follows:

- A visual-based estimator dedicated to skid-steering robots, which jointly estimates the ICR-based kinematic parameters of the robotic platform and 6-DoF poses in a tight-coupled manner. The formulation, error state propagation, and initialization of the kinematic parameters are presented in detail.

- Detailed observability analysis of the estimator 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 3-DoF extrinsic translation between the camera and odometer are unobservable with the online estimate of kinematic parameters, which prevents performing online sensor-to-sensor extrinsic calibration.

- Extensive experiments including both simulation tests and real-world experiments were conducted for evaluations. Ablation study is also investigated to standout the feasibility of the proposed method. In general, the proposed method i) shows high accuracy and great robustness under different environmental and mechanical conditions to enable the long-term mission of the robots and (ii) outperforms the competing methods that do and do not estimate the kinematics.

The rest of the paper is organized as follows. We introduce the kinematics model of skid-steering robots in Sec. II. Subsequently, the framework of the tightly-coupled sliding-window estimator is introduced in Sec. III, and we illustrate the kinematics estimation in detail. The observability analysis of the estimator under different configurations is performed in Sec. IV. Experimental results are presented in Sec. V. Finally, the paper is concluded in Sec. VI.

II. ICR-BASED KINEMATICS OF 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 ${}^A\mathbf{p}_B$ and ${}^A_B\mathbf{R}$ to denote position and rotation of frame $\{B\}$ with respect to $\{A\}$, and ${}^A_B\mathbf{q}$ is the corresponding unit quaternion of ${}^A_B\mathbf{R}$. In addition, $\mathbf{I}$ denotes the identity matrix, and $\mathbf{0}$ denotes the zero matrix. We use $\hat{\mathbf{x}}$ and $\delta\mathbf{x}$ to represent the current estimated value and error state for variable $\mathbf{x}$. Additionally, we reserve the symbol $\check{z}$ to denote the inferred measurement value of z, which is widely used in observability analysis. For the rotation matrix ${}^G_O\mathbf{R}$, we define the attitude error angle vector $\delta\boldsymbol{\theta}$ as follows [23]:

$$\begin{matrix}{}^G_O\mathbf{R} = {}^G_O\hat{\mathbf{R}}(\mathbf{I} + \lfloor\delta\boldsymbol{\theta}\rfloor)\end{matrix} \tag{1}$$

We use $\lfloor\mathbf{v}\rfloor$ to denote the skew-symmetric matrix of a 3d vector $\mathbf{v}$:

$$\lfloor\mathbf{v}\rfloor = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}, \tag{2}$$
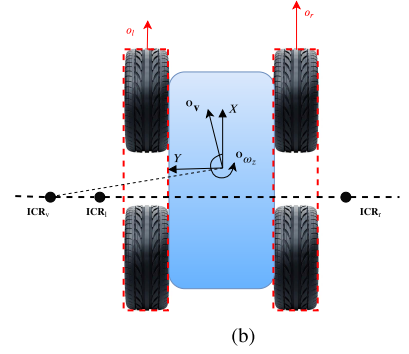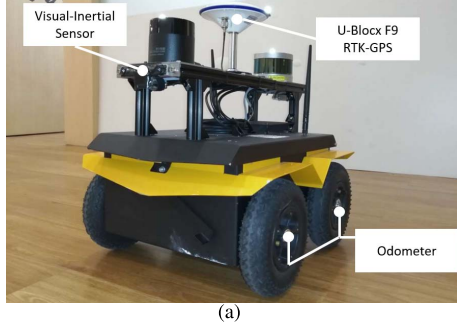
(a)   (b)

Fig. 1.   Skid-steering robotic platform used in our tests, as well as the corresponding kinematic model [1]. (a) Our testing robot, built on the Clearpath Jackal Platform [2]. The equipped low-cost sensors (i.e., a monocular camera, an imu, and wheel encoders) are leveraged in the proposed system, while the others (i.e., LiDAR and RTK-GPS) are not required in our system. (b) The odometer measurements and the instantaneous center of rotation (ICR, denoted by $[\mathbf{ICR}_v, \mathbf{ICR}_l, \mathbf{ICR}_r]$) of a skid-steering robot. $^{\mathbf{O}}\mathbf{v}$ represents the robot velocity in odometer frame, and $^{\mathbf{O}}\omega_z$ is the angular velocity along the yaw direction.

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 [7] and [10], which contains five kinematic parameters: three ICR parameters and two correction factors, as shown in Fig. 1. To describe the details, we denote $\mathbf{ICR}_v = (X_v, Y_v)$ the ICR position of the robot frame, and $\mathbf{ICR}_l = (X_l, Y_l)$ and $\mathbf{ICR}_r = (X_r, Y_r)$ 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:

$$Y_l = -\frac{o_l - {}^{\mathbf{O}}v_x}{{}^{\mathbf{O}}\omega_z}, \quad Y_r = -\frac{o_r - {}^{\mathbf{O}}v_x}{{}^{\mathbf{O}}\omega_z}$$

$$Y_v = \frac{{}^{\mathbf{O}}v_x}{{}^{\mathbf{O}}\omega_z}, \quad X_v = X_l = X_r = -\frac{{}^{\mathbf{O}}v_y}{{}^{\mathbf{O}}\omega_z} \quad (3)$$

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

$$\begin{bmatrix} {}^{\mathbf{O}}v_x \\ {}^{\mathbf{O}}v_y \\ {}^{\mathbf{O}}\omega_z \end{bmatrix} = g(\boldsymbol{\xi}, o_l, o_r) = \frac{1}{\Delta Y} \begin{bmatrix} -Y_r & Y_l \\ X_v & -X_v \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_l & 0 \\ 0 & \alpha_r \end{bmatrix} \begin{bmatrix} o_l \\ o_r \end{bmatrix} \quad (4)$$

with

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}_{ICR}^\top & \boldsymbol{\xi}_\alpha^\top \end{bmatrix} = \begin{bmatrix} X_v & Y_l & Y_r & \alpha_l & \alpha_r \end{bmatrix}^\top, \quad \Delta Y = Y_l - Y_r \quad (5)$$

where $\boldsymbol{\xi}$ is the entire set of kinematic parameters.

Interestingly, as a special configuration when

$$\boldsymbol{\xi} = \begin{bmatrix} 0, & 0.5\, b, & -0.5\, b, & 1, & 1 \end{bmatrix}^T \quad (6)$$

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

$$^{\mathbf{O}}v_x = \frac{o_l + o_r}{2}, \quad {}^{\mathbf{O}}\omega_z = \frac{o_r - o_l}{b}, \quad {}^{\mathbf{O}}v_y = 0 \quad (7)$$

This is exactly the kinematic model for a wheeled robot moving without slippage (i.e., an ideal differential drive robot kinematics), and used by most existing work for localizing wheeled robots [24], [25], [26]. 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. V).

III. VISUAL-INERTIAL KINEMATICS AND POSE ESTIMATION

In this paper, we utilize a sliding-window bundle adjustment (BA) based estimator for the kinematics and pose estimation of 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 follows the pose estimation method for ground robots [27], where visual constraints, IMU constraints, motion manifold (the profile of ground surface) constraints, ideal differential drive model induced odometer constraints are formulated and iteratively optimized. Notably, kinematics estimation is not touched in [27], while it is the focus of this work. We also note that, compared to [27] and the other papers that focus on estimator architecture novelty, this work introduces methods to systematically handle skid-steering effects via online estimation. Our goal is to consistently and accurately estimate the motion of a moving robot as well as necessary observability-guided kinematic parameters of the robots.

A. Estimator Formulation

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

$$\mathbf{x} = \begin{bmatrix} \boldsymbol{\chi}_{\mathbf{O}}^\top, & {}^{\mathbf{G}}\mathbf{v}_{\mathbf{I}_k}^\top, & \mathbf{b}_{a_k}^\top, & \mathbf{b}_{\omega_k}^\top, & \mathbf{m}_k^\top, & \boldsymbol{\xi}_k^\top \end{bmatrix}^\top \quad (8)$$

where

$$\boldsymbol{\chi}_{\mathbf{O}} = \left[{}^{\mathbf{G}}_{\mathbf{O}_{k-s}}\mathbf{q}^\top, {}^{\mathbf{G}}\mathbf{p}_{\mathbf{O}_{k-s}}^\top, \ldots, {}^{\mathbf{G}}_{\mathbf{O}_{k-1}}\mathbf{q}^\top, {}^{\mathbf{G}}\mathbf{p}_{\mathbf{O}_{k-1}}^\top, {}^{\mathbf{G}}_{\mathbf{O}_k}\mathbf{q}^\top, {}^{\mathbf{G}}\mathbf{p}_{\mathbf{O}_k}^\top \right]^\top \quad (9)$$

denotes the sliding-window poses of odometer frame at times $\{k - s, \ldots, k\}$ when keyframe images are captured. ${}^{\mathbf{G}}\mathbf{v}_{\mathbf{I}_k}, \mathbf{b}_{a_k}, \mathbf{b}_{\omega_k}$ are the IMU related states, including the IMU velocity in global frame, accelerometer bias, and gyroscope bias. If IMU is not available in the system, ${}^{\mathbf{G}}\mathbf{v}_{\mathbf{I}_k}, \mathbf{b}_{a_k}, \mathbf{b}_{\omega_k}$ will excluded from the state vector. In addition, $\mathbf{m}_k$ denotes the parameters for modeling the local motion manifold of the skid-steering robots across current sliding window. The motion manifold is parameterized by a quadratic polynomial, and $\mathbf{m}_k$ is the 6-dimensional vectors to formulate the polynomial. This has been shown in [28] and [27] to improve the estimation performance for ground robots, and we also adopt this design in our work. Finally, $\boldsymbol{\xi}_k$, 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) Bundle Adjustment Optimization: Our optimization process follows the design of [27]. Specifically, the sliding-window bundle adjustment in our estimation algorithm seeks to iteratively minimize a cost function corresponding to a combination of sensor measurement constraints, kinematics constraints, and marginalized constraints.

$$\mathcal{C} = \mathcal{C}_P + \mathcal{C}_V + \mathcal{C}_I + \mathcal{C}_O + \mathcal{C}_M \quad (10)$$

In what follows, we describe each of the cost terms. Firstly, the marginalized term $\mathcal{C}_P$ is critical to consistently keep the algorithm computational complexity bounded, by probabilistically removing the old states in the sliding window [29]. The camera term $\mathcal{C}_V$, IMU term $\mathcal{C}_I$, and motion manifold term $\mathcal{C}_M$ used in this work are similar to that of existing literature [18], [27], [30], respectively, but with dedicated design for ground robots. In general, the camera cost term models the geometrical reprojection error of point features in the keyframes (see Sec. III-B), 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. Finally, $\mathcal{C}_O$ denotes the kinematic constraints induced by wheel odometer measurements. This term is a function of robot pose, measurement input, as well as skid-steering intrinsic kinematic parameters, and is discussed in detail in Sec. III-C. Due to limited space, we omit the details of $\mathcal{C}_P$, $\mathcal{C}_I$, and $\mathcal{C}_M$ in this paper, and readers are encouraged to refer to our supplementary material.

B. Visual Constraints

In the sliding-window BA with visual constraints, the poses of only visual 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 [31], [32], 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 [33] and tacked with FREAK [34] descriptors.

The successfully tracked features across multiple keyframes will be initialized in the 3D space by triangulation. By denoting $\mathbf{z}_{i,j}$ the visual measurement of a 3D feature ${}^{\mathbf{G}}\mathbf{p}_{f_j}$ observed by the $\mathbf{C}_i$th camera keyframe, the visual reprojection error [35] in normalized image coordinate is given by:

$$\mathcal{C}_V({}^{\mathbf{G}}_{\mathbf{O}_i}\mathbf{R}, {}^{\mathbf{G}}\mathbf{p}_{\mathbf{O}_i}) = \left\| \mathbf{z}_{i,j} - \pi({}^{\mathbf{G}}_{\mathbf{C}_i}\mathbf{R}, {}^{\mathbf{G}}\mathbf{p}_{\mathbf{C}_i}, {}^{\mathbf{G}}\mathbf{p}_{f_j}) \right\|_{\mathbf{\Lambda}_V}^2,$$
$${}^{\mathbf{G}}_{\mathbf{C}_i}\mathbf{R} = {}^{\mathbf{G}}_{\mathbf{O}_i}\mathbf{R}{}^{\mathbf{O}}_{\mathbf{C}}\mathbf{R}, \quad {}^{\mathbf{G}}\mathbf{p}_{\mathbf{C}_i} = {}^{\mathbf{G}}\mathbf{p}_{\mathbf{O}_i} + {}^{\mathbf{G}}_{\mathbf{O}_i}\mathbf{R}{}^{\mathbf{O}}\mathbf{p}_{\mathbf{C}} \quad (11)$$

In the above expressions, $\pi(\cdot)$ denotes the perspective function of an intrinsically calibrated camera, and $\mathbf{\Lambda}_V$ represents the inverse of noise covariance in the observation $\mathbf{z}_{i,j}$. ${}^{\mathbf{O}}_{\mathbf{C}}\mathbf{R}$ and ${}^{\mathbf{O}}\mathbf{p}_{\mathbf{C}}$ are the extrinsic transformation between camera and odometer. Furthermore, we choose not to incorporate visual features into the state vector due to the limited computational resources. Thus ${}^{\mathbf{G}}\mathbf{p}_{f_j}$ and the pose of oldest keyframe over the sliding window will be marginalized immediately after the iterative minimization.

C. ICR-Based Kinematic Constraints

This section provides details on formulating ICR-based kinematic constraints $\mathcal{C}_O$. 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, ${}^{\mathbf{O}(t)}\mathbf{v}$ and ${}^{\mathbf{O}(t)}\boldsymbol{\omega}$, 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 $\boldsymbol{\xi}$ [see eq. (4)]:

$$\begin{bmatrix} {}^{\mathbf{O}(t)}\mathbf{v} \\ {}^{\mathbf{O}(t)}\boldsymbol{\omega} \end{bmatrix} = \mathbf{\Pi} \, g(\boldsymbol{\xi}(t), o_l(t), o_r(t))$$
$$= \mathbf{\Pi} \, g(\boldsymbol{\xi}(t), o_{lm}(t) - n_l(t), o_{rm}(t) - n_r(t)) \quad (12)$$

where $\mathbf{\Pi} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{e}_3 \end{bmatrix}^T$ is the selection matrix with $\mathbf{e}_i$ being a 3×1 unit vector with the ith 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 $\mathbf{n}_o = \begin{bmatrix} n_l & n_r \end{bmatrix}^\top$.

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

$${}^{\mathbf{G}}\dot{\mathbf{p}}_{\mathbf{O}(t)} = {}^{\mathbf{G}}_{\mathbf{O}(t)}\mathbf{R} \cdot {}^{\mathbf{O}(t)}\mathbf{v} \quad (13a)$$
$${}^{\mathbf{G}}_{\mathbf{O}(t)}\dot{\mathbf{R}} = {}^{\mathbf{G}}_{\mathbf{O}(t)}\mathbf{R} \cdot \lfloor {}^{\mathbf{O}(t)}\boldsymbol{\omega} \rfloor \quad (13b)$$
$$\dot{\boldsymbol{\xi}}(t) = \mathbf{n}_\xi(t) \quad (13c)$$

where we model the noise of the ICR kinematic parameter $\boldsymbol{\xi}$ by using a random walk process, and $\mathbf{n}_\xi$ is characterized by zero-mean white Gaussian noise. The motivation of using $\mathbf{n}_\xi$ is to capture time-varying characteristics of $\boldsymbol{\xi}$, caused by changes in road conditions, tire pressures, center of mass. It is important to point out that, unlike sensor extrinsic calibration in which parameters can be modeled as constant parameters, e.g., ${}^C\dot{\mathbf{p}}_I = \mathbf{0}$ in [36], $\boldsymbol{\xi}$ must be modeled as a time-varying variable.

To propagate pose estimates in a stochastic estimator, we describe the process starting from the estimates $\hat{\mathbf{x}}_{O_{k-1}} = \begin{bmatrix} {}^{\mathbf{G}}\hat{\mathbf{p}}_{\mathbf{O}_{k-1}}^T, {}^{\mathbf{G}}_{\mathbf{O}_{k-1}}\hat{\mathbf{q}}^T, \hat{\boldsymbol{\xi}}_{k-1}^T \end{bmatrix}^T$. Once the instantaneous local velocities of the robot (see Eq. (12)) are available, we integrate the

differential equations in Eq. (13) over the time interval $t \in (t_{k-1}, t_k)$ by all the intermediate odometer measurements $\mathcal{O}_m$, and obtain the predicted robot pose and kinematic parameters at the newest keyframe time t_k. This process can be characterized by $\hat{\mathbf{x}}_{O_k} = f(\hat{\mathbf{x}}_{O_{k-1}}, \mathcal{O}_m)$. Therefore, the odometer-induced kinematic constraint can be generically written in the following form:

$$\mathcal{C}_O(\mathbf{x}_{O_k}, \mathbf{x}_{O_{k-1}}) = \left\| \mathbf{x}_{O_k} \boxminus f(\mathbf{x}_{O_{k-1}}, \mathcal{O}_m) \right\|_{\mathbf{\Lambda}_O}^2 \qquad (14)$$

where $\mathbf{\Lambda}_O$ represents the inverse covariance (information) obtained via propagation process, and "$\boxminus$" denotes the minus operation on manifold [37]. To formulate Eq. (14) 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. (13):

$$\delta^{\mathbf{G}} \dot{\mathbf{p}}_{\mathbf{O}} \simeq {}^{\mathbf{G}}_{\mathbf{O}}\hat{\mathbf{R}}(\mathbf{I} + \lfloor \delta\boldsymbol{\theta} \rfloor)({}^{\mathbf{O}}\hat{\mathbf{v}} + \mathbf{J}_{v\xi}\delta\boldsymbol{\xi} + \mathbf{J}_{vo}\mathbf{n}_o) - {}^{\mathbf{G}}_{\mathbf{O}}\hat{\mathbf{R}}{}^{\mathbf{O}}\hat{\mathbf{v}}$$

$$\simeq -{}^{\mathbf{G}}_{\mathbf{O}}\hat{\mathbf{R}}\lfloor {}^{\mathbf{O}}\hat{\mathbf{v}} \rfloor \delta\boldsymbol{\theta} + {}^{\mathbf{G}}_{\mathbf{O}}\hat{\mathbf{R}}\mathbf{J}_{v\xi}\delta\boldsymbol{\xi} + {}^{\mathbf{G}}_{\mathbf{O}}\hat{\mathbf{R}}\mathbf{J}_{vo}\mathbf{n}_o \qquad (15)$$

$$\dot{\delta\boldsymbol{\theta}} \simeq -\lfloor {}^{\mathbf{O}}\hat{\boldsymbol{\omega}} \rfloor \delta\boldsymbol{\theta} + \mathbf{J}_{\omega\xi}\delta\boldsymbol{\xi} + \mathbf{J}_{\omega o}\mathbf{n}_o \qquad (16)$$

$$\dot{\delta\boldsymbol{\xi}} \simeq \mathbf{n}_\xi \qquad (17)$$

We here point out that Eq. (16) can be obtained similar to Eq. (156) of [23]. In above equations, $\mathbf{J}_{v\xi}, \mathbf{J}_{\omega\xi}, \mathbf{J}_{vo}, \mathbf{J}_{\omega o}$ are the linearized Jacobian matrices, originated from:

$$^{\mathbf{O}}\mathbf{v} = {}^{\mathbf{O}}\hat{\mathbf{v}} + \mathbf{J}_{v\xi}\delta\boldsymbol{\xi} + \mathbf{J}_{vo}\mathbf{n}_o \qquad (18)$$

$$^{\mathbf{O}}\boldsymbol{\omega} = {}^{\mathbf{O}}\hat{\boldsymbol{\omega}} + \mathbf{J}_{\omega\xi}\delta\boldsymbol{\xi} + \mathbf{J}_{\omega o}\mathbf{n}_o \qquad (19)$$

and

$$\mathbf{J}_{v\xi} = \frac{\hat{\alpha}_l o_l - \hat{\alpha}_r o_r}{\Delta \hat{Y}^2} \begin{bmatrix} 0 & \hat{Y}_r & -\hat{Y}_l & 0 & 0 \\ \Delta\hat{Y} & -\hat{X}_v & \hat{X}_v & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$+ \frac{1}{\Delta\hat{Y}} \begin{bmatrix} 0 & 0 & 0 & -\hat{Y}_r o_l & \hat{Y}_l o_r \\ 0 & 0 & 0 & \hat{X}_v o_l & -\hat{X}_v o_r \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad (20)$$

$$\mathbf{J}_{vo} = -\frac{1}{\Delta\hat{Y}} \begin{bmatrix} -\hat{\alpha}_l \hat{Y}_r & \hat{\alpha}_r \hat{Y}_l \\ \hat{X}_v \hat{\alpha}_l & -\hat{X}_v \hat{\alpha}_r \\ 0 & 0 \end{bmatrix} \qquad (21)$$

$$\mathbf{J}_{\omega\xi} = \frac{1}{\Delta\hat{Y}^2} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -\hat{\alpha}_r o_r + \hat{\alpha}_l o_l & \hat{\alpha}_r o_r - \hat{\alpha}_l o_l & -\Delta\hat{Y} o_l & \Delta\hat{Y} o_r \end{bmatrix} \qquad (22)$$

$$\mathbf{J}_{\omega o} = -\frac{1}{\Delta\hat{Y}} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -\hat{\alpha}_l & \hat{\alpha}_r \end{bmatrix} \qquad (23)$$

Once continuous-time error-state equations are given, the discrete-time state transition matrices needed when minimizing Eq. (14), can be straightforwardly calculated by numerical integration. In general, Eq. (14) encapsulates all information related to the skid-steering effect and enables online estimation of the skid-steering parameters. After the constraint in Eq. (14) is minimized, $\boldsymbol{\xi}_{k-1}$ will be marginalized immediately, ensuring low computational complexity of the system.

D. Initialization of Kinematic Parameters

To allow the estimation of skid-steering kinematic parameters online, an initial estimate of the parameter vector $\boldsymbol{\xi}$ is required. Specifically, we use a simply while effective method by setting:

$$\boldsymbol{\xi}_{initial} = \left[0, 0.5\, b^\dagger, -0.5\, b^\dagger, 1, 1\right]^T \qquad (24)$$

We emphasize that Eq. (24) is similar but different from Eq. (6). The parameter b 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 $b^\dagger$ should not be simply the wheel distance. To compute $b^\dagger$, 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:

$$b^\dagger = \frac{1}{N} \sum_{i=1}^{N} \frac{||o_{lm}(t_i) - o_{rm}(t_i)||}{||\boldsymbol{\omega}_m(t_i)||} \qquad (25)$$

where $\boldsymbol{\omega}_m(t_i)$ is gyroscope measurement, and N is the number of measurements. Although this is not of high precision and the road condition of computing $b^\dagger$ 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).

IV. OBSERVABILITY ANALYSIS

A critical prerequisite condition for a well-formulated estimator is to *only* include locally observable (or identifiable[1]) [38] states 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; (ii) monocular camera and odometer with 3 ICR parameter only; (iii) monocular camera, odometer and an IMU with 3 ICR parameters and 2 correction factors. Subsequently, 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 [39], [40], [41], 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 [18] or the corresponding linearized discrete-time system [42], [43]. As shown in [18] and [17], the dimension

---

[1]Since the derivative of $\boldsymbol{\xi}$ is modelled by zero-mean Gaussian, we here use observability and identifiability interchangeably.

of the nullspace of the observability matrix might subject to changes due to linearization and dicrestization, and thus we conduct our analysis in the nonlinear continuous-time space in this work.

To conduct the observability analysis, we are inspired by [19], in which information provided by each sensor is investigated and subsequently combined together for deriving the final results. By doing this, 'abstract' measurements instead of the 'raw' measurements are used for analysis, which significantly simplifies our derivation and is helpful for intuitive understanding. 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 [44].

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 [19], [35] 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:

$$\breve{\boldsymbol{\omega}}_C(t) = {}^{C_{(t)}}\boldsymbol{\omega} + \mathbf{n}_\omega(t) \tag{26a}$$

$$\breve{\mathbf{v}}_C(t) = s^{-1} \cdot {}^{C}_{G}\mathbf{R} \cdot {}^{G}\mathbf{v}_{C_{(t)}} + \mathbf{n}_v(t) \tag{26b}$$

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

$$\breve{\boldsymbol{\omega}}(t) \triangleq {}^{O}_{C}\mathbf{R} \cdot \breve{\boldsymbol{\omega}}_C(t), \quad \breve{\mathbf{v}}(t) \triangleq {}^{O}_{C}\mathbf{R} \cdot \breve{\mathbf{v}}_C(t) \tag{27}$$

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

C. Observability of ξ 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:

$$^{O}\boldsymbol{\omega} = {}^{O}_{C}\mathbf{R} \cdot {}^{C}\boldsymbol{\omega} \tag{28}$$

which allows us to derive the following equations:

$$^{G}\mathbf{p_O} = {}^{G}_{C}\mathbf{R} \cdot {}^{C}\mathbf{p_O} + {}^{G}\mathbf{p_C} \tag{29a}$$

$$^{G}\mathbf{v_O} = {}^{G}\dot{\mathbf{p}}_\mathbf{O} = {}^{G}_{C}\mathbf{R}\lfloor{}^{C}\boldsymbol{\omega}\rfloor{}^{C}\mathbf{p_O} + {}^{G}\mathbf{v_C} \tag{29b}$$

$$^{O}_{G}\mathbf{R}{}^{G}\mathbf{v_O} = {}^{O}_{C}\mathbf{R}\lfloor{}^{C}\boldsymbol{\omega}\rfloor{}^{C}\mathbf{p_O} + {}^{O}_{C}\mathbf{R}{}^{C}_{G}\mathbf{R}{}^{G}\mathbf{v_C} \tag{29c}$$

$$^{O}\mathbf{v} = -\lfloor{}^{O}\boldsymbol{\omega}\rfloor{}^{O}\mathbf{p_C} + {}^{O}_{C}\mathbf{R}{}^{C}_{G}\mathbf{R}{}^{G}\mathbf{v_C} \tag{29d}$$

Substituting Eq. (27) to Eq. (29d), we obtain the following equation:

$$^{O}\mathbf{v} = -\lfloor\breve{\boldsymbol{\omega}}\rfloor{}^{O}\mathbf{p_C} + s \cdot \breve{\mathbf{v}} \tag{30}$$

where ${}^{O}\mathbf{p_C}$ is known and ${}^{O}\mathbf{v}$ is velocity expressed in the odometer frame. We also note that, during the observability analysis, the noise terms are ignored, following the standard procedure of performing the observability analysis.

On the other hand, as mentioned in Sec. II, odometer provides observations for the speed of left and right wheels, i.e., o_l and o_r respectively. By linking o_l, o_r, $\breve{\boldsymbol{\omega}}(t)$, $\breve{\mathbf{v}}(t)$, and kinematic parameter vector $\boldsymbol{\xi}$ together, specifically substituting Eq. (30) into Eq. (4), we obtain:

$$
\begin{bmatrix} \begin{bmatrix} \breve{\omega}{}^{O}y_\mathbf{C} \\ -\breve{\omega}{}^{O}x_\mathbf{C} \end{bmatrix} + s\begin{bmatrix} \breve{v}_x \\ \breve{v}_y \end{bmatrix} \\ \breve{\omega} \end{bmatrix} = \frac{1}{\Delta Y}\begin{bmatrix} -Y_r & Y_l \\ X_v & -X_v \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \alpha_l & 0 \\ 0 & \alpha_r \end{bmatrix}\begin{bmatrix} o_l \\ o_r \end{bmatrix}
$$
$$
= \begin{bmatrix} \breve{\omega}Y_l \\ -\breve{\omega}X_v \\ \frac{1}{\Delta Y}\begin{bmatrix} -1 & 1 \end{bmatrix}\begin{bmatrix} \alpha_l & 0 \\ 0 & \alpha_r \end{bmatrix}\begin{bmatrix} o_l \\ o_r \end{bmatrix} \end{bmatrix} + \begin{bmatrix} \alpha_l o_l \\ 0 \\ 0 \end{bmatrix} \tag{31}
$$

where ${}^{O}x_\mathbf{C}$, ${}^{O}y_\mathbf{C}$ are the first and second element of ${}^{O}\mathbf{p_C}$, and $\breve{v}_x$, $\breve{v}_y$ are the first and second element of $\breve{\mathbf{v}}$. For brevity, we use $\breve{\omega}$ to denote the third element of $\breve{\boldsymbol{\omega}}$. By defining $\beta_r = \Delta Y^{-1}\alpha_r$, and $\beta_l = \Delta Y^{-1}\alpha_l$, we can write

$$
\begin{bmatrix} \begin{bmatrix} \breve{\omega}{}^{O}y_\mathbf{C} \\ -\breve{\omega}{}^{O}x_\mathbf{C} \end{bmatrix} + s\begin{bmatrix} \breve{v}_x \\ \breve{v}_y \end{bmatrix} \\ \breve{\omega} \end{bmatrix} = \begin{bmatrix} \breve{\omega}Y_l \\ -\breve{\omega}X_v \\ -\beta_l o_l + \beta_r o_r \end{bmatrix} + \begin{bmatrix} \beta_l \Delta Y o_l \\ 0 \\ 0 \end{bmatrix} \tag{32}
$$

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

$$\boldsymbol{\epsilon} = \begin{bmatrix} X_v & Y_l & Y_r & \alpha_l & \alpha_r & s \end{bmatrix}^\top$$

which allows us to analyze whether indistinguishable sets of $\boldsymbol{\epsilon}$ exist subject to the provided measurement constraint equations.

The identifiability of $\boldsymbol{\epsilon}$ can be described as follows:

Lemma 1: By using measurements from a monocular camera and wheel odometers, $\boldsymbol{\epsilon}$ is not locally identifiable.

Proof: $\boldsymbol{\epsilon}$ is locally identifiable if and only if $\bar{\boldsymbol{\epsilon}}$ is locally identifiable:

$$\bar{\boldsymbol{\epsilon}} = \begin{bmatrix} Y_l & \Delta Y & X_v & \beta_l & \beta_r & s \end{bmatrix}^\top$$

By expanding Eq. (32), we can write the following constraints:

$$c_x(\bar{\boldsymbol{\epsilon}}, t) = \breve{\omega}(t){}^{O}y_\mathbf{C} + s\breve{v}_x(t) - \breve{\omega}(t)Y_l - \beta_l\Delta Y o_l(t) = 0 \tag{33a}$$

$$c_y(\bar{\boldsymbol{\epsilon}}, t) = -\breve{\omega}(t){}^{O}x_\mathbf{C} + s\breve{v}_y(t) + \breve{\omega}(t)X_v = 0 \tag{33b}$$

$$c_\omega(\bar{\boldsymbol{\epsilon}}, t) = \breve{\omega}(t) + \beta_l o_l(t) - \beta_r o_r(t) = 0 \tag{33c}$$

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 $\mathcal{S} = \{t_0, t_1, \ldots, t_s\}$,:

$$\mathbf{M} = \begin{bmatrix} \mathbf{D}(t_0)^\top & \mathbf{D}(t_1)^\top & \ldots & \mathbf{D}(t_s)^\top \end{bmatrix}^\top \quad (34)$$

where

$$\mathbf{D}(t) = \begin{bmatrix} \frac{\partial c_x(\bar{\epsilon}, t)}{\partial \bar{\epsilon}} & \frac{\partial c_y(\bar{\epsilon}, t)}{\partial \bar{\epsilon}} & \frac{\partial c_\omega(\bar{\epsilon}, t)}{\partial \bar{\epsilon}} \end{bmatrix}^\top$$
$$= \begin{bmatrix} -\breve{\omega}(t) & -\beta_l o_l(t) & 0 & -\Delta Y o_l(t) & 0 & \breve{v}_x(t) \\ 0 & 0 & \breve{\omega}(t) & 0 & 0 & \breve{v}_y(t) \\ 0 & 0 & 0 & o_l(t) & -o_r(t) & 0 \end{bmatrix} \quad (35)$$

Substituting Eq. (35) back into Eq. (34) leads to:

$$\mathbf{M} = \begin{bmatrix} -\breve{\omega}(t_0) & -\beta_l o_l(t_0) & 0 & -\Delta Y o_l(t_0) & 0 & \breve{v}_x(t_0) \\ 0 & 0 & \breve{\omega}(t_0) & 0 & 0 & \breve{v}_y(t_0) \\ 0 & 0 & 0 & o_l(t_0) & -o_r(t_0) & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -\breve{\omega}(t_s) & -\beta_l o_l(t_s) & 0 & -\Delta Y o_l(t_s) & 0 & \breve{v}_x(t_s) \\ 0 & 0 & \breve{\omega}(t_s) & 0 & 0 & \breve{v}_y(t_s) \\ 0 & 0 & 0 & o_l(t_s) & -o_r(t_s) & 0 \end{bmatrix} \quad (36)$$

By defining $\mathbf{M}(:, i)$ the ith block columns of $\mathbf{M}$, the following equation holds:

$$(-{}^{\mathbf{O}}y_{\mathbf{C}} + Y_l) \cdot \mathbf{M}(:, 1) + \Delta Y \cdot \mathbf{M}(:, 2)$$
$$+ (X_v - {}^{\mathbf{O}}x_{\mathbf{C}}) \cdot \mathbf{M}(:, 3) + s \cdot \mathbf{M}(:, 6) = \mathbf{0}$$

The above equation demonstrates that $\mathbf{M}$ is *not* of full column rank, indicating that ϵ is not identifiable.

To further investigate the indistinguishable states that cause the unobservable situations, we note that for a vector $\bar{\epsilon}_1 = [Y_l, \Delta Y, X, \beta_l, \beta_r, s]^\top$ that satisfies Eq. (33a), another vector

$$\bar{\epsilon}_2 = [(1 + \lambda/s)Y_l - (\lambda/s){}^{\mathbf{O}}y_{\mathbf{C}}, (1 + \lambda/s)\Delta Y,$$
$$(1 + \lambda/s)X - (\lambda/s){}^{\mathbf{O}}x_{\mathbf{C}}, \beta_l, \beta_r, s + \lambda]^\top$$

for any $\lambda \in \mathbb{R}$ is always valid for the constraints Eq. (33a). Thus $\bar{\epsilon}_1$ and $\bar{\epsilon}_2$ are indistinguishable, and $\bar{\epsilon}$ is not locally identifiable. This completes the proof. □

D. Observability of ξ_{ICR} With Monocular Camera and Odometer

Since the full kinematic parameters $\xi = [\xi_{ICR}^T, \xi_\alpha^T]^T$ with monocular camera and odometer are not locally identifiable, we look into the case of that only the 3 ICR parameters, i.e., ξ_{ICR}, are estimated without the correction factors. Similar to Eq. (31), the following equation holds:

$$\left[\begin{bmatrix} \breve{\omega}{}^{\mathbf{O}}y_{\mathbf{C}} \\ -\breve{\omega}{}^{\mathbf{O}}x_{\mathbf{C}} \end{bmatrix} + s \begin{bmatrix} \breve{v}_x \\ \breve{v}_y \end{bmatrix} \right] = \begin{bmatrix} \breve{\omega}Y_l \\ -\breve{\omega}X_v \\ \frac{1}{\Delta Y}(o_r - o_l) \end{bmatrix} + \begin{bmatrix} o_l \\ 0 \\ 0 \end{bmatrix} \quad (37)$$

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

$$\gamma = [\xi_{ICR}^T \ s]^T = [X_v \ Y_l \ Y_r \ s]^\top$$

The local identifiability of γ can be stated as follows:

Lemma 2: By using the monocular and odometer measurements, and the 3 ICR parameter vector ξ_{ICR} to model the kinematics, γ is locally identifiable except for the following degenerate cases: (i) the odometer linear velocity $o_l(t)$ keeps zero; (ii) the angular velocity $\breve{\omega}(t)$ keeps zero; (iii) $o_r(t)$,

$o_l(t)$, and $\breve{\omega}(t)$ 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 $\breve{\omega}(t)$ is consistently proportional to $o_l(t)$.

Proof: We first note that the local identifiability of γ is equivalent to that of $\bar{\gamma}$,

$$\bar{\gamma} = \begin{bmatrix} Y_l & \Delta Y & X_v & s \end{bmatrix}^T$$

By expanding 37 and considering all the measurements at different time t, we can derive following system constraints:

$$c_x(\bar{\gamma}, t) = \breve{\omega}(t){}^{\mathbf{O}}y_{\mathbf{C}} + s\breve{v}_x(t) - \breve{\omega}Y_l - o_l(t) = 0 \quad (38a)$$
$$c_y(\bar{\gamma}, t) = -\breve{\omega}(t){}^{\mathbf{O}}x_{\mathbf{C}} + s\breve{v}_y(t) + \breve{\omega}(t)X_v = 0 \quad (38b)$$
$$c_\omega(\bar{\gamma}, t) = \breve{\omega}(t) + \frac{o_l(t) - o_r(t)}{\Delta Y} = 0 \quad (38c)$$

Similar to the case of using full kinematic parameters ξ in Section. IV-C, we derive the following observability matrix for $\bar{\gamma}$ (using Eq. (34) and. 35):

$$\mathbf{M} = \begin{bmatrix} -\breve{\omega}(t_0) & 0 & 0 & \breve{v}_x(t_0) \\ 0 & 0 & \breve{\omega}(t_0) & \breve{v}_y(t_0) \\ 0 & \frac{o_r(t_0) - o_l(t_0)}{(\Delta Y)^2} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ -\breve{\omega}(t_s) & 0 & 0 & \breve{v}_x(t_s) \\ 0 & 0 & \breve{\omega}(t_s) & \breve{v}_y(t_s) \\ 0 & \frac{o_r(t_s) - o_l(t_s)}{(\Delta Y)^2} & 0 & 0 \end{bmatrix} \quad (39)$$

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

$$\mathbf{M}(:, 2) \leftarrow (\Delta Y)^2 \mathbf{M}(:, 2)$$
$$\mathbf{M}(:, 4) \leftarrow s\mathbf{M}(:, 4) + (Y_l - {}^{\mathbf{O}}y_{\mathbf{C}})\mathbf{M}(:, 1) + (X_v - {}^{\mathbf{O}}x_{\mathbf{C}})\mathbf{M}(:, 3)$$

where $(\cdot) \leftarrow (\cdot)$ represents the operator to replace the left side by the right side. As a result, Eq. (39) can be simplified as:

$$\mathbf{M} = \begin{bmatrix} -\breve{\omega}(t_0) & 0 & 0 & o_l(t_0) \\ 0 & 0 & \breve{\omega}(t_0) & 0 \\ 0 & o_r(t_0) - o_l(t_0) & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ -\breve{\omega}(t_s) & 0 & 0 & o_l(t_s) \\ 0 & 0 & \breve{\omega}(t_s) & 0 \\ 0 & o_r(t_s) - o_l(t_s) & 0 & 0 \end{bmatrix} \quad (40)$$

To investigate the observability of the matrix in Eq. (40), we inspect the existence of the non-zero vector $\mathbf{k}$ such that $\mathbf{M}\mathbf{k} = \mathbf{0}, \mathbf{k} = \begin{bmatrix} k_1 & k_2 & k_3 & k_4 \end{bmatrix}^\top \neq \mathbf{0}$. If such a vector $\mathbf{k}$ exists, all of the following conditions must be satisfied:

$$-\breve{\omega}(t)k_1 + o_l(t)k_4 = 0, \ (o_r(t) - o_l(t))k_2 = 0, \ \breve{\omega}(t)k_3 = 0$$

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

- $o_l(t)$ keeps constantly zero, $\mathbf{k} = \begin{bmatrix} 0 & 0 & 0 & \rho \end{bmatrix}^\top$,
- $\breve{\omega}(t)$ keeps constantly zero, $\mathbf{k} = \begin{bmatrix} 0 & 0 & \rho & 0 \end{bmatrix}^\top$,
- $o_r(t)$, $o_l(t)$, and $\breve{\omega}(t)$ are all constants, $\mathbf{k} = \begin{bmatrix} 0 & \rho & 0 & 0 \end{bmatrix}^\top$,
- $o_l(t)$ keeps identical to $o_r(t)$, $\mathbf{k} = \begin{bmatrix} 0 & \rho & 0 & 0 \end{bmatrix}^\top$,
- $\breve{\omega}(t)$ keeps proportional to $o_l(t)$, $\mathbf{k} = \begin{bmatrix} \rho o_l/\breve{\omega} & 0 & 0 & \rho \end{bmatrix}^\top$.

where ρ can be any non-zero value that is used to generate valid non-zero vector $\mathbf{k}$ such that $\mathbf{M}\mathbf{k} = \mathbf{0}$. 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, $\boldsymbol{\xi}_{ICR}$ is observable unless entering the specified special conditions listed above.   □

E. Observability of $\boldsymbol{\xi}$ 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 $\boldsymbol{\xi}$ is not feasible, and the alternative solution is to include $\boldsymbol{\xi}_{ICR}$ in the online stage only. However, it is not an ideal solution to calibrate $\boldsymbol{\xi}_\alpha$ offline and fixed in the online stage since it is subject to the changes in road conditions and tire conditions, etc.

To tackle this problem, we investigate the observability of $\boldsymbol{\xi}$ when an IMU and camera are fused with odometer measurements. Once IMU and camera are 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. (27)) without having the unknown scale factor [17], [19], [21]. Similarly to Eq. (33), to simplify the analysis, we prove identifiability of $\bar{\boldsymbol{\xi}}$ instead of $\boldsymbol{\xi}$, since properties of $\bar{\boldsymbol{\xi}}$ and $\boldsymbol{\xi}$ are interchangeable:

$$\bar{\boldsymbol{\xi}} = \begin{bmatrix} Y_l & \Delta Y & X_v & \beta_l & \beta_r \end{bmatrix}^\top$$

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

Proof: Similarly to Eq. (33a), by removing the scale factor, the corresponding system constraints can be derived as:

$$c_x(\bar{\boldsymbol{\xi}}, t) = \breve{\omega}(t)^{\mathbf{O}}y_{\mathbf{C}} + \breve{v}_x(t) - \breve{\omega}(t)Y_l - \beta_l \Delta Y o_l(t) = 0 \tag{41a}$$

$$c_y(\bar{\boldsymbol{\xi}}, t) = -\breve{\omega}(t)^{\mathbf{O}}x_{\mathbf{C}} + \breve{v}_y(t) + \breve{\omega}(t)X_v = 0 \tag{41b}$$

$$c_\omega(\bar{\boldsymbol{\xi}}, t) = \breve{\omega}(t) + \beta_l o_l(t) - \beta_r o_r(t) = 0 \tag{41c}$$

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

$$\mathbf{M} = \begin{bmatrix} -\breve{\omega}(t_0) & -\beta_l o_l(t_0) & 0 & -\Delta Y o_l(t_0) & 0 \\ 0 & 0 & \breve{\omega}(t_0) & 0 & 0 \\ 0 & 0 & 0 & o_l(t_0) & -o_r(t_0) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\breve{\omega}(t_s) & -\beta_l o_l(t_s) & 0 & -\Delta Y o_l(t_s) & 0 \\ 0 & 0 & \breve{\omega}(t_s) & 0 & 0 \\ 0 & 0 & 0 & o_l(t_s) & -o_r(t_s) \end{bmatrix} \tag{42}$$

After the following linear operations:

$$\mathbf{M}(:, 2) \leftarrow -\mathbf{M}(:, 2)/\beta_l$$

$$\mathbf{M}(:, 4) \leftarrow \mathbf{M}(:, 4) + \Delta Y \mathbf{M}(:, 2)$$

Eq. (42) can be simplified as:

$$\mathbf{M} = \begin{bmatrix} -\breve{\omega}(t_0) & o_l(t_0) & 0 & 0 & 0 \\ 0 & 0 & \breve{\omega}(t_0) & 0 & 0 \\ 0 & 0 & 0 & o_l(t_0) & -o_r(t_0) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\breve{\omega}(t_s) & o_l(t_s) & 0 & 0 & 0 \\ 0 & 0 & \breve{\omega}(t_s) & 0 & 0 \\ 0 & 0 & 0 & o_l(t_s) & -o_r(t_s) \end{bmatrix} \tag{43}$$

Similarly, we investigate the existence of the non-zero vector $\mathbf{k}$ such that $\mathbf{M}\mathbf{k} = 0$. If such a vector $\mathbf{k} = \begin{bmatrix} k_1 & k_2 & k_3 & k_4 & k_5 \end{bmatrix}$ exists, all of the following must be satisfied:

$$-\breve{\omega}(t)k_1 + o_l(t)k_2 = 0, \quad \breve{\omega}(t)k_3 = 0, \quad o_r(t)k_4 - o_l(t)k_5 = 0$$

which requires one of the following conditions to be true:

- $o_l(t)$ is constantly zero, $\mathbf{k} = \begin{bmatrix} 0 & \rho_1 & 0 & \rho_2 & 0 \end{bmatrix}^\top$, or $o_r(t)$ is constantly zero, $\mathbf{k} = \begin{bmatrix} 0 & 0 & 0 & 0 & \rho \end{bmatrix}^\top$
- $\breve{\omega}(t)$ is constantly zero, $\mathbf{k} = \begin{bmatrix} \rho_1 & 0 & \rho_2 & 0 & 0 \end{bmatrix}^\top$,
- $o_r(t)$, $o_l(t)$, and $\breve{\omega}(t)$ are all constants, $\mathbf{k} = \begin{bmatrix} 0 & 0 & \rho & 0 & 0 \end{bmatrix}^\top$,
- $o_l(t)$ keeps proportional to $o_r(t)$, $\mathbf{k} = \begin{bmatrix} 0 & 0 & 0 & \rho & 0 \end{bmatrix}^\top$,
- $\breve{\omega}(t)$ keeps proportional to $o_l(t)$, $\mathbf{k} = \begin{bmatrix} \rho o_l/\breve{\omega} & \rho & 0 & 0 & 0 \end{bmatrix}^\top$.

where ρ, ρ_1, ρ_2 can be any non-zero value that is used to generate valid non-zero vector $\mathbf{k}$ such that $\mathbf{M}\mathbf{k} = 0$. All the above cases are *special* conditions. Therefore, in a skid-steering robot localization system equipped with a camera, an IMU, and odometers, $\boldsymbol{\xi}$ is observable unless entering the specified special conditions listed above. This completes the proof.   □

F. Observability of $\boldsymbol{\xi}$ With a Monocular Camera, an IMU, an Odometer and With Online Extrinsics 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. We first define the parameter state when camera-odometer extrinsics are included:

$$\boldsymbol{\eta} = \begin{bmatrix} X_v & Y_l & Y_r & \alpha_l & \alpha_r & {}^{\mathbf{O}}x_{\mathbf{C}} & {}^{\mathbf{O}}y_{\mathbf{C}} & {}^{\mathbf{O}}z_{\mathbf{C}} & {}^{\mathbf{O}}_{\mathbf{C}}\delta\boldsymbol{\theta}^\top \end{bmatrix}^\top$$

where ${}^{\mathbf{O}}\mathbf{p}_{\mathbf{C}} = \begin{bmatrix} {}^{\mathbf{O}}x_{\mathbf{C}} & {}^{\mathbf{O}}y_{\mathbf{C}} & {}^{\mathbf{O}}z_{\mathbf{C}} \end{bmatrix}^\top$ and ${}^{\mathbf{O}}_{\mathbf{C}}\delta\boldsymbol{\theta} \in \mathbb{R}^3$ are the translational and rotational part of the extrinsic transformation between odometer and camera. ${}^{\mathbf{O}}_{\mathbf{C}}\delta\boldsymbol{\theta}$ is error state (or Lie algebra increment) of the 3D rotation matrix ${}^{\mathbf{O}}_{\mathbf{C}}\mathbf{R}$.

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

$$\boldsymbol{\eta}_p = \begin{bmatrix} X_v & Y_l & Y_r & \alpha_l & \alpha_r & {}^{\mathbf{O}}x_{\mathbf{C}} & {}^{\mathbf{O}}y_{\mathbf{C}} & {}^{\mathbf{O}}z_{\mathbf{C}} \end{bmatrix}^\top$$

and

$$\boldsymbol{\eta}_\theta = \begin{bmatrix} X_v & Y_l & Y_r & \alpha_l & \alpha_r & {}^{\mathbf{O}}_{\mathbf{C}}\delta\boldsymbol{\theta}^\top \end{bmatrix}^\top$$

To summarize, the objective of this section is to demonstrate the observability properties of $\boldsymbol{\eta}$, $\boldsymbol{\eta}_p$, and $\boldsymbol{\eta}_\theta$.

Lemma 4: By using measurements from a monocular camera, IMU and wheel odometers, $\boldsymbol{\eta}_p$ and $\boldsymbol{\eta}$ are not identifiable. Specifically, the vertical direction of translation in the extrinsics, ${}^{\mathbf{O}}z_{\mathbf{C}}$ is always unidentifiable for any type of ground robot, and ${}^{\mathbf{O}}x_{\mathbf{C}}$ and ${}^{\mathbf{O}}y_{\mathbf{C}}$ become unidentifiable if the skid-steering kinematic parameters are estimated online.

Lemma 5: By using measurements from a monocular camera, IMU and wheel odometers, $\boldsymbol{\eta}_\theta$ is identifiable in general motion, which allows an estimation algorithm to perform online calibration on extrinsic rotational parameters.

Proof: First of all, $\boldsymbol{\eta}$ is locally identifiable if and only if $\bar{\boldsymbol{\eta}}$ is locally identifiable:

$$\bar{\boldsymbol{\eta}} = \begin{bmatrix} Y_l & \Delta Y & X_v & \beta_l & \beta_r & {}^{\mathbf{O}}x_{\mathbf{C}} & {}^{\mathbf{O}}y_{\mathbf{C}} & {}^{\mathbf{O}}z_{\mathbf{C}} & {}^{\mathbf{O}}_{\mathbf{C}}\delta\boldsymbol{\theta}^\top \end{bmatrix}^\top$$

By substituting $\check{\mathbf{v}}(t) \triangleq {}^{\mathbf{O}}_{\mathbf{C}}\mathbf{R} \cdot \check{\mathbf{v}}_C(t)$ in Eq. (27), we are able to derive constraints similar to Eq. (41), as

$$c_x(\bar{\boldsymbol{\eta}}, t) = \check{\omega}(t){}^{\mathbf{O}}y_{\mathbf{C}} + \mathbf{e}_1^\top {}^{\mathbf{O}}_{\mathbf{C}}\mathbf{R}\check{\mathbf{v}}_C(t) - \check{\omega}Y_l - \beta_l \Delta Y o_l(t) = 0 \tag{44a}$$

$$c_y(\bar{\boldsymbol{\eta}}, t) = -\check{\omega}(t){}^{\mathbf{O}}x_{\mathbf{C}} + \mathbf{e}_2^\top {}^{\mathbf{O}}_{\mathbf{C}}\mathbf{R}\check{\mathbf{v}}_C(t) + \check{\omega}(t)X_v = 0 \tag{44b}$$

$$c_\omega(\bar{\boldsymbol{\xi}}, t) = \check{\omega}(t) + \beta_l o_l(t) - \beta_r o_r(t) = 0 \tag{44c}$$

Considering the constraints in a set of time instants $\mathcal{S} = \{t_0, t_1, \ldots, t_s\}$, we compute the following observability matrices for the systems with calibrating $\bar{\boldsymbol{\eta}}_p$ and $\bar{\boldsymbol{\eta}}_\theta$, given by Eq. (45), as shown at the bottom of the next page, and Eq. (46a), as shown at the bottom of the next page, respectively. Eq. (46a) can be converted to Eq. (46b), as shown at the bottom of the next page, by linear operations:

$$\mathbf{M}(:, 4) \leftarrow \mathbf{M}(:, 4) - \Delta Y/\beta_l \mathbf{M}(:, 2)$$
$$\mathbf{M}(:, 2) \leftarrow -\mathbf{M}(:, 2)/\beta_l$$

Similar to previous proofs, to look into the properties of $\mathbf{M}$ in Eq. (45), we investigate non-zero vector $\mathbf{k}$ such that $\mathbf{M}\mathbf{k} = \mathbf{0}$. We can easily find the following non-zero solutions:

$$\mathbf{k}_1 = \begin{bmatrix} \rho & 0 & 0 & 0 & 0 & \rho & 0 \end{bmatrix}^\top$$
$$\mathbf{k}_2 = \begin{bmatrix} 0 & 0 & \rho & 0 & 0 & \rho & 0 & 0 \end{bmatrix}^\top$$
$$\mathbf{k}_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & \rho \end{bmatrix}^\top$$

where ρ can be any non-zero value. We can find that $\mathbf{k}_1$, $\mathbf{k}_2$ are related with the kinematic parameters, while $\mathbf{k}_3$ always holds, which results from no constraints on ${}^{\mathbf{O}}z_{\mathbf{C}}$ 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 ${}^{\mathbf{O}}y_{\mathbf{C}}$ are indistinguishable; (ii) X_v and ${}^{\mathbf{O}}x_{\mathbf{C}}$ are indistinguishable; (iii) the vertical direction of extrinsic parameters ${}^{\mathbf{O}}z_{\mathbf{C}}$ is always unidentifiable for skid-steering robot moving on ground, no matter whether the kinematic parameters are calibrated online.

However, $\mathbf{M}$ in Eq. (46b) is under quite different properties. Similarly, we investigate the non-zero $\mathbf{k}$ that satisfies $\mathbf{M}\mathbf{k} = \mathbf{0}$, which requires the all of the following to be true:

$$-\check{\omega}(t)k_1 + o_l(t)k_2 - \mathbf{e}_1^\top {}^{\mathbf{O}}_{\mathbf{C}}\mathbf{R}\lfloor \check{\mathbf{v}}_C(t_0)\rfloor \cdot \begin{bmatrix} k_6 & k_7 & k_8 \end{bmatrix}^\top = 0$$
$$\check{\omega}(t)k_3 - \mathbf{e}_2^\top {}^{\mathbf{O}}_{\mathbf{C}}\mathbf{R}\lfloor \check{\mathbf{v}}_C(t_0)\rfloor \cdot \begin{bmatrix} k_6 & k_7 & k_8 \end{bmatrix}^\top = 0$$
$$o_r(t)k_4 - o_l(t)k_5 = 0$$

However, since $\check{\omega}(t), \check{\mathbf{v}}_C(t), o_l(t), o_r(t)$ are time variant under general motion, we can not find such a non-zero vector $\mathbf{k}$. We can draw the conclusion: (iv) the 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 [19], [20], [21], extrinsic parameters between camera and odometer are not observable and cannot be calibrated online. (ii) The extrinsic rotation between sensors can be observable in general motion and thus can be included in the state vector of a state estimation algorithm for online calibration. Since extrinsic rotation usually does not undergo severe changes, and its calibration also suffers from degenerate motions, we do not recommend calibrating it online.

V. 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: (i) the advantages and necessities of online estimating kinematic parameters in visual (inertial) localization systems for skid-steering robots, (ii) the observability and convergence properties of the skid-steering kinematic parameters under different settings, (iii) the robustness of the proposed kinematics and pose estimation method against the changes of mass center, tire inflation condition, terrains, etc., to enable the long-term mission completeness of the robots without performance reduction. In our experiments, we used the adapted skid-steering robots based on the commercially available Clearpath Jackal robot [2] (see Fig. 1), with both 'kinematics and pose estimation' sensors and 'ground-truth sensors' equipped. For 'kinematics and pose estimation' sensors, we used a 10Hz monocular global shutter camera at a resolution of 640×400, a 200Hz Bosch BMI160 IMU, and 100Hz wheel encoders .[2] 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 [22]. We note that the offline calibration procedure is an important prerequisite in our experiments since the extrinsic translation between the odometer and camera has shown to be constantly unobservable. All the experiments are conducted on an Intel Core i7-8700 @ 3.20GHz CPU for comparisons. To fully examine the proposed method, we only evaluate the poses from the odometry system without loop closures and prior maps to standout the performance of our

---

[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.
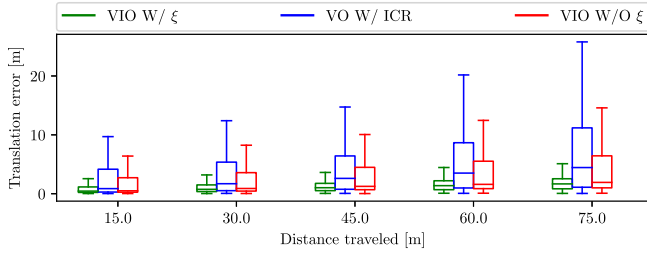
Fig. 2. Boxplot of the relative trajectory error (RPE) statistics over all the sequences where RTK-GPS measurements are available. This plot best seen in color.

proposed kinematic and pose estimation method. The sliding-window size of the estimator is 8 in all the evaluations presented in this paper.

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 $\boldsymbol{\xi}$ (Eq. (4)) online and compared that to the competing methods. To demonstrate the generality of our method, we conducted experiments under various environmental conditions. 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. The representative figures in 8 types of terrains are also shown in the experiments of [1].

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) [45] as our metrics. 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 [46], Kaist Complex Urban [47], Oxford Robotcar [48]) or lack of one or multiple synchronized low cost sensors (NCLT [49], and Canadian 3DMap [50]). To this end, we also plan to release a comprehensive dataset specifically with low-cost sensors, as our future work.

1) Pose Estimation Accuracy: We first conducted an experiment to show the benefits gained by modeling and estimating skid-steering parameters online. In this experiment, three setups are compared, i.e., two provably observable methods and one baseline method. Specifically, those methods are 1) **VIO (visual-inertial odometry) W/ $\boldsymbol{\xi}$:** using measurements from a monocular camera, an IMU, and odometer via the proposed estimator by estimating the full 5 skid-steering kinematic parameters $\boldsymbol{\xi}$ online; 2) **VO (visual odometry) W/ ICR:** using monocular camera and odometer measurements (without an IMU), and performing localization by estimating the 3 ICR parameters $\boldsymbol{\xi}_{ICR}$ online; 3) **VIO W/O $\boldsymbol{\xi}$:** using measurements from monocular camera, an IMU and odometer, and utilizing

$$\mathbf{M}=\begin{bmatrix} -\breve{\omega}(t_0) & -\beta_l o_l(t_0) & 0 & -\Delta Y o_l(t_0) & 0 & 0 & \breve{\omega}(t_0) & 0 \\ 0 & 0 & \breve{\omega}(t_0) & 0 & 0 & -\breve{\omega}(t_0) & 0 & 0 \\ 0 & 0 & 0 & o_l(t_0) & -o_r(t_0) & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -\breve{\omega}(t_s) & -\beta_l o_l(t_s) & 0 & -\Delta Y o_l(t_s) & 0 & 0 & \breve{\omega}(t_s) & 0 \\ 0 & 0 & \breve{\omega}(t_s) & 0 & 0 & -\breve{\omega}(t_s) & 0 & 0 \\ 0 & 0 & 0 & o_l(t_s) & -o_r(t_s) & 0 & 0 & 0 \end{bmatrix} \quad (45)$$

$$\mathbf{M}=\begin{bmatrix} -\breve{\omega}(t_0) & -\beta_l o_l(t_0) & 0 & -\Delta Y o_l(t_0) & 0 & -\mathbf{e}_1^\top {}^O_C\mathbf{R}\lfloor\breve{\mathbf{v}}_C(t_0)\rfloor \\ 0 & 0 & \breve{\omega}(t_0) & 0 & 0 & -\mathbf{e}_2^\top {}^O_C\mathbf{R}\lfloor\breve{\mathbf{v}}_C(t_0)\rfloor \\ 0 & 0 & 0 & o_l(t_0) & -o_r(t_0) & \mathbf{0}_{3\times1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -\breve{\omega}(t_s) & -\beta_l o_l(t_s) & 0 & -\Delta Y o_l(t_s) & 0 & -\mathbf{e}_1^\top {}^O_C\mathbf{R}\lfloor\breve{\mathbf{v}}_C(t_s)\rfloor \\ 0 & 0 & \breve{\omega}(t_s) & 0 & 0 & -\mathbf{e}_2^\top {}^O_C\mathbf{R}\lfloor\breve{\mathbf{v}}_C(t_s)\rfloor \\ 0 & 0 & 0 & o_l(t_s) & -o_r(t_s) & \mathbf{0}_{3\times1} \end{bmatrix} \quad (46a)$$

$$\mathbf{M}=\begin{bmatrix} -\breve{\omega}(t_0) & o_l(t_0) & 0 & 0 & 0 & -\mathbf{e}_1^\top {}^O_C\mathbf{R}\lfloor\breve{\mathbf{v}}_C(t_0)\rfloor \\ 0 & 0 & \breve{\omega}(t_0) & 0 & 0 & -\mathbf{e}_2^\top {}^O_C\mathbf{R}\lfloor\breve{\mathbf{v}}_C(t_0)\rfloor \\ 0 & 0 & 0 & o_l(t_0) & -o_r(t_0) & \mathbf{0}_{3\times1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -\breve{\omega}(t_s) & -o_l(t_s) & 0 & 0 & 0 & -\mathbf{e}_1^\top {}^O_C\mathbf{R}\lfloor\breve{\mathbf{v}}_C(t_s)\rfloor \\ 0 & 0 & \breve{\omega}(t_s) & 0 & 0 & -\mathbf{e}_2^\top {}^O_C\mathbf{R}\lfloor\breve{\mathbf{v}}_C(t_s)\rfloor \\ 0 & 0 & 0 & o_l(t_s) & -o_r(t_s) & \mathbf{0}_{3\times1} \end{bmatrix} \quad (46b)$$

TABLE I
RMSE OF ATE (M) ON THE SEQUENCES WITH RTK-GPS MEASUREMENTS

|	SEQ21-CP01	SEQ22-CP01	SEQ23-CP01	SEQ3-CP01	SEQ4-CP01	SEQ5-CP01	SEQ8-CP01	SEQ12-CP01	SEQ14-CP01	SEQ17-CP01	Mean
Length (m)	629.16	633.53	651.94	632.64	629.96	626.83	204.81	436.19	372.15	110.55	
Terrain	(b,f)	(b,f)	(b)	(b,f)	(b,f)	(b,f)	(e)	(e)	(b)	(b)	
VIO W/ ξ	**1.36**	**2.35**	1.50	**2.81**	**1.17**	**2.70**	**0.26**	**0.80**	1.65	0.32	**1.492**
VO W/ ICR	43.12	38.49	**0.61**	48.04	39.62	40.60	0.44	3.12	**0.74**	**0.20**	21.498
VIO W/O ξ [27]	5.88	6.02	1.62	9.82	10.43	8.61	0.50	1.86	4.97	0.45	5.016
VINS-on-Wheels [25]	6.52	8.37	3.81	8.97	13.39	8.93	0.52	1.78	5.69	0.56	5.854
VINS-Mono [31]	10.36	14.84	3.19	12.53	18.29	11.37	0.62	3.04	5.04	0.52	7.980
Wheel Odo. W/ ξ	36.38	48.26	4.07	63.33	43.91	54.94	1.85	4.62	7.71	0.85	26.592
Inertial Aided Wheel Odo.	39.35	45.38	4.29	58.21	40.05	48.92	1.63	3.63	5.93	0.69	24.808
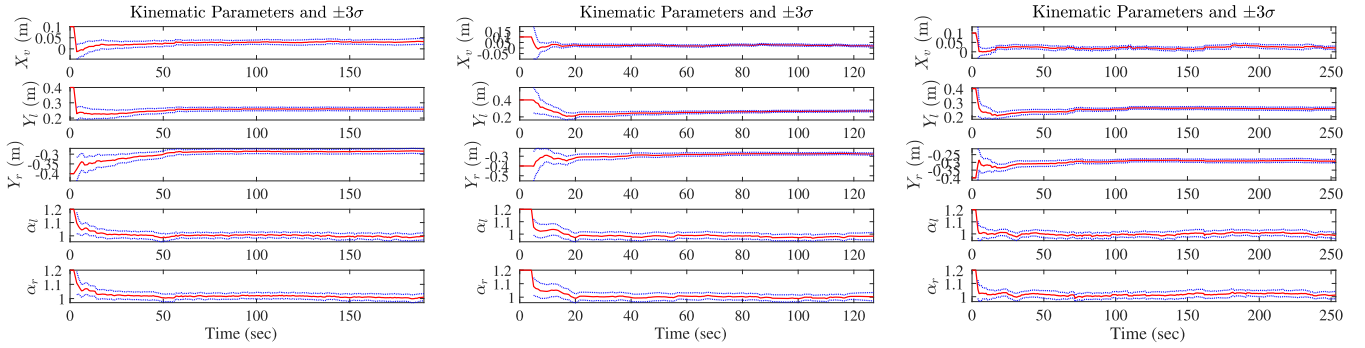
Fig. 3. While given bad initial values, the kinematic parameters ξ are able to converge to the reasonable values in the visual-inertial navigation system. The online estimated kinematic parameters (red) and the associated $\pm 3\sigma$ envelopes (blue) are shown on sequences " SEQ8-CP01, SEQ18-CP01, SEQ19-CP01" from left to right.
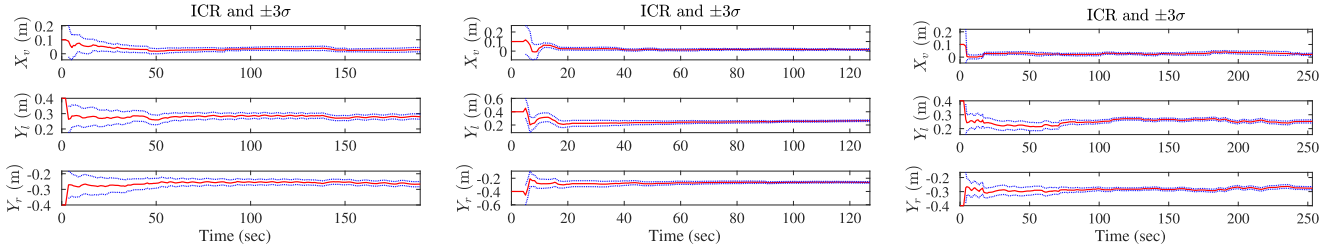

Fig. 4. While given bad initial values, the ICR parameters ξ_{ICR} are able to converge to the reasonable values in the kinematics-constrained VO system. The online estimated ξ_{ICR} parameters (red) and the associated $\pm 3\sigma$ envelopes (blue) are shown on sequences " SEQ8-CP01, SEQ18-CP01, SEQ19-CP01" from left to right.

differential drive kinematics in Eq. (7) for localization [27] without explicitly modeling ξ. Notably, the configuration **VIO W/O ξ** is exactly our baseline method [27] for ground robots. In [27], similar to the optimized cost function in Eq. (10), the constraints from prior term $\mathcal{C}_P$, camera term $\mathcal{C}_V$, IMU term $\mathcal{C}_I$, and motion manifold term $\mathcal{C}_M$ are taken into account, while the odometer constraint $\mathcal{C}_O$ is simply induced from ideal differential drive model (see Eq. (7)) instead of leveraging the ICR-based kinematics constraint (see Section. III-C) proposed in this work.

We show the final drift errors on 23 representative sequences in the Table.II of the supplementary material, which cover all eight types of terrains (a)-(g). Notably, some sequences also cover multiple types of terrains. In the sequences where GPS signals were available across the entire data sequence, we also evaluated the root mean square errors (RMSE) [38] of absolute

translational error (ATE) [45]. To compute that, we interpolated the estimated poses to get the ones corresponding to the timestamp of the GPS measurements. In addition to the aforementioned configurations, **VIO W/ ξ**, **VO W/ ICR**, and **VIO W/O ξ**, we also compare to the state-of-the-art wheel odometer aided VIO methtod **VINS-on-Wheels** [25], the state-of-the-art VIO method **VINS-Mono** [31], and also the vision-free methods, including **Wheel Odo. W/ ξ** and **Inertial Aided Wheel Odo.**. In the method, **Wheel Odo. W/ ξ**, only the wheel odometer measurements are used by propagating poses forward based on the ICR kinematics with the full 5 kinematic parameters ξ (see Sec. III-C). The kinematic parameters are kept fixed at the given initial values, and will not be updated due to the lack of constraints from the other sensor modalities. The fixed values can not reflect the real-time robot kinematic status, and can lead to significant errors in

TABLE II
MEAN OF RPE (M) FOR DIFFERENT SEGMENT LENGTH ON THE
SEQUENCES WITH RTK-GPS MEASUREMENTS

Segment Length (m)	VIO W/ ξ	VO W/ ICR	VIO W/O ξ
15.00	0.78	10.27	2.37
30.00	1.11	11.33	2.82
45.00	1.42	13.14	3.40
60.00	1.75	15.61	4.07
75.00	2.00	17.76	4.66

pose estimation. As for the method, **Inertial Aided Wheel Odo.**, both the measurements from IMU and wheel odometers are fused. Besides the forward propagation based on the full ICR-based kinetic model with wheel odometers measurements, IMU measurements are also propagated forward to formulate the relative pose constraints between two virtual keyframes. Here we selected virtual keyframes once the wheel odometer pose prediction has a translation over 0.2 meters or rotation over 3 degrees. The IMU velocity and biases, as well as the kinematic parameters, are optimized in this method.

The RMSE errors of compared methods are shown in Table. I, where we highlight the best results in bold, while the bad results (RMSE of ATE is over 12m) by underlines. The results clearly demonstrate that when skid-steering kinematic parameter ξ is estimated online, the localization accuracy can be significantly improved. This validates our claim that, in order to use odometer measurements of skid-steering robots, the complicated mechanism must be explicitly modelled to avoid accuracy loss. We also note that, the method of using an IMU and estimating the full 5 kinematic parameters, **VIO W/ ξ**, 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 model, 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 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. The pure VIO method without wheel odometers, **VINS-Mono** [31] has a poor performance because of the degeneration motion of ground robots. The state-of-the-art odometer aided VIO method, **VINS-on-Wheels** [25], shows relatively good pose estimation results, while it is inferior to the proposed method, due to the planar ground assumption and the ideal differential drive model assumption, which does not address the slippage issue. The vision-free methods, **Wheel Odo. W/ ξ** and **Inertial Aided Wheel Odo.**, shows large pose estimation error on most of the sequences, which stands out the substantial help from visual measurements. Representative Trajectory estimates on representative sequences are also shown in the supplementary material. 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. II and Fig. 2, which also support our algorithm claims.

To further examine the advantages of online estimating the full kinematic parameters ξ in the kinematics-constrained VIO systems, we conduct ablation study and compare the three variants of the proposed method: 1) estimating 3 parameters ξ_{ICR} only; 2) estimating the full ξ with 5 parameters; 3) used fixed ξ with a relatively good initial guess. Besides, in the ablation study, the skid-steering robot is tested under the following practically commonly-seen mechanism configurations during its life-long service: (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. The experimental results can be found in our supplementary material, which demonstrates the advantage of estimating the full kinematic parameters ξ and the effectiveness of the proposed methods under the aforementioned mechanism configurations.

2) Convergence of Kinematic Parameters: In this section, we show experimental results to demonstrate the convergence properties of ξ and ξ_{ICR}, in systems that we theoretically claim observable. Unlike the experiments in the previous section, which utilized the method described in Sec. III-D 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

$$\delta X_v = 0.08, \ \delta Y_l = 0.14, \ \delta Y_r = -0.1, \ \delta \alpha_l = 0.2, \ \delta \alpha_r = 0.2$$

For the kinematics-constrained VO system, we only add error terms to ξ_{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. 3, the estimates of the full kinematic parameters ξ in VIO are shown, along with the corresponding $\pm 3\sigma$ uncertainty envelopes (where σ is the square root of the corresponding diagonal component of the estimated covariance matrix). The convergence of ξ_{ICR} in VO are also shown in Fig. 4. The results demonstrate that the kinematic parameters ξ 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 ξ_{ICR} when only a monocular camera and odometer sensors are used. The results exactly meet our theoretical expectations that ξ in VIO and ξ_{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 ξ, 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. V-B where ground truth ξ is known.

3) Time Efficiency: The proposed ICR-kinematic based method is computationally effective to run in real time, thus is applicable in various robotic applications. In this section, we showcase the running time of the proposed method on a typical sequence "SEQ21-CP01" with a traversed

TABLE III

MEAN AND (STANDARD DEVIATION) OF THE RUNNING
TIME FOR THE MAIN PROCESSING STAGES

Stages	Running Time (ms)		
	VO W/ ICR	VIO W/ ξ	VIO W/O ξ
Feature Detection	2.09(0.67)	2.06(0.66)	2.09(0.67)
Feature Tracking	4.35(1.04)	4.29(1.00)	4.40(1.06)
BA Optimization	4.09(1.33)	4.77(2.11)	4.35(1.54)
Marginalization	2.22(0.83)	2.71(0.90)	2.63(0.91)

Fig. 5. In simulation experiments, estimated Trajectories aligned with the ground truth trajectory.

distance of 629.16 meters, and there are about 2400 cycles of sliding-window optimization occurring in our estimators. As noted previously, the run time is also evaluated on a desktop with Intel Core i7-8700 @ 3.20GHz CPU. The implementation of our method is in C++ and adequately optimized for high efficiency. We disclose the average running time and its standard deviation of four main processing stages: feature detection, feature tracking, BA optimization, and marginalization. The ICR-based kinematic constraints are formulated and leveraged in BA optimization, as well as the other constraints. The results are shown in Table III, including the three configurations **VIO W/ ξ**, **VO W/ ICR**, and **VIO W/O ξ**. Comparing the runtime of **VIO W/ ξ** to the one of **VIO W/O ξ**, it is obvious to find that introducing the ICR-based kinematic constraints into VIO has a negligible burden on the computation. Considering the significant improvement in the pose estimation accuracy by incorporating kinematic constraints (see experiments in Sec. V-A1), it is quite worthwhile to estimate the kinematic parameters and poses jointly by the proposed method.

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 205.4 m, using the method introduced in [51]. 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, since skid-steering kinematic parameters can not be known in advance, we initialize ξ 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 $8 \cdot 10^{-2}$ for all elements in ξ.

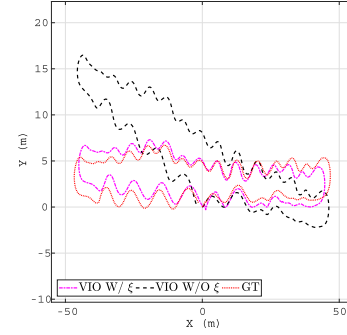To collect algorithm statistics, we conducted 15 Monte-Carlo tests and compute parameter estimation results for ξ. Specifically, we computed the mean and std of calibration errors for all elements in ξ, averaged from the Monte-Carlo tests. The estimation error for each element in ξ are: -0.0211 ± 0.0095, 0.0102 ± 0.0030, -0.0081 ± 0.0026, 0.0212 ± 0.0109, 0.0216 ± 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 ξ is subject to the following error vector $\delta\xi = \begin{bmatrix} 0.15 & 0.15 & -0.15 & 0.1 & 0.1 \end{bmatrix}^\top$. In this case, the calibration errors averaged over the second half of the trajectory are: -0.0276 ± 0.0067, 0.0199 ± 0.0118, 0.0054 ± 0.0026, 0.0192 ± 0.0157, 0.0189 ± 0.0157.

Since simulation tests provide absolute ground truth, it is also interesting to investigate the accuracy gain by estimating ξ online. Fig. 5 demonstrates the estimated trajectory when ξ is estimated online, or ξ 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.

VI. CONCLUSION AND FUTURE WORK

In this paper, we propose a novel kinematics and pose estimation 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, mass center changes, tire inflation changes, and terrain conditions, we explicitly model the kinematics of skid-steering robots by using both track ICRs and correction factors, which are online estimated to prevent performance reduction in the long-term mission of skid-steering robots. To guide the estimator design, we conduct detailed observability analysis for the proposed algorithm under different setup conditions. Specifically, we show that the kinematic parameter vector ξ 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 including ablation study and

simulation tests are also provided, which demonstrate that the proposed method is able to compute skid-steering kinematic parameters online and yield accurate pose estimation results.

There are also limitations to the proposed method. Notably, ICR-based kinematic model is only valid when a vehicle is operated in low dynamics. Although low dynamics are common for skid-steering robots, the feasibility at high speed is also essential. Moreover, abrupt changes of the traversed terrain or robot mechanism status will cause significant changes in the kinematic parameters. In that case, estimators will have some latency to converge. In the future, it is worthwhile investigating a more general kinematic model for mobile robots through data-driven deep-learning methods. A general kinematic model is supposed to function well regardless of the moving speed, interacted terrains, and robot mechanism status. It is also interesting to actively detect the abrupt changes in terrains and robot status. State estimation can converge quickly by enlarging the uncertainty of kinematic parameters when abrupt changes are detected.

REFERENCES

[1] X. Zuo, M. Zhang, Y. Chen, Y. Liu, G. Huang, and M. Li, "Visual-inertial localization for skid-steering robots with kinematic constraints," in *Proc. Int. Symp. Robot. Res.* Springer, 2019, pp. 741–756.

[2] Clearpath Robotics. (2019). *Clearpath Ground Vehicle*. [Online]. Available: https://www.clearpathrobotics.com/jackal-small-unmanned-ground-vehicle/

[3] P. Newswire. (2019). *Skid Steer Loader Market*. [Online]. Available: https://finance.yahoo.com/news/skid-steer-loader-market-anticipated-102000939.html

[4] J. Vincent. (2019). *Fedex Unveils Autonomous Delivery Robot*. [Online]. Available: https://www.theverge.com/2019/2/27/18242834/delivery-robot-fedex-sameday-bot-autonomous-trials

[5] B. F. Rubin. (2019). *Amazon's Scout Robots*. [Online]. Available: https://www.cnet.com/news/amazons-scout-robots-thats-no-cooler-thats-your-prime-delivery/

[6] G. Anousaki and K. J. Kyriakopoulos, "A dead-reckoning scheme for skid-steered vehicles in outdoor environments," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Dec. 2004, pp. 580–585.

[7] J. L. Martínez, A. Mandow, J. Morales, S. Pedraza, and A. García-Cerezo, "Approximating kinematics for tracked mobile robots," *Int. J. Robot. Res.*, vol. 24, pp. 867–878, Oct. 2005.

[8] A. Mandow, J. L. Martinez, J. Morales, J. L. Blanco, A. Garcia-Cerezo, and J. Gonzalez, "Experimental kinematics for wheeled skid-steer mobile robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2007, pp. 1222–1227.

[9] J. Yi, H. Wang, J. Zhang, D. Song, S. Jayasuriya, and J. Liu, "Kinematic modeling and analysis of skid-steered mobile robots with applications to low-cost inertial-measurement-unit-based motion estimation," *IEEE Trans. Robot.*, vol. 25, no. 5, pp. 1087–1097, Oct. 2009.

[10] J. Pentzer, S. Brenman, and K. Reichard, "Model-based prediction of skid-steer robot kinematics using online estimation of track instantaneous centers of rotation," *J. Field Robot.*, vol. 31, no. 3, pp. 455–476, May 2014.

[11] J. L. Martinez, J. Morales, A. Mandow, S. Pedraza, and A. Garcia-Cerezo, "Inertia-based ICR kinematic model for tracked skid-steer robots," in *Proc. IEEE Int. Symp. Saf., Secur. Rescue Robot. (SSRR)*, Oct. 2017, pp. 166–171.

[12] M. Sutoh, Y. Iijima, Y. Sakakieda, and S. Wakabayashi, "Motion modeling and localization of skid-steering wheeled rover on loose terrain," *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 4031–4037, Oct. 2018.

[13] C. Wang, W. Lv, X. Li, and M. Mei, "Terrain adaptive estimation of instantaneous centres of rotation for tracked robots," *Complexity*, vol. 2018, pp. 1–10, Nov. 2018.

[14] F. Solc and J. Sembera, "Kinetic model of a skid steered robot," in *Proc. Int. Conf. Signal Process., Robot. Autom.*, 2008, pp. 61–65.

[15] J. Y. Wong, *Theory of Ground Vehicles*. Hoboken, NJ, USA: Wiley, 2008.

[16] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "Observability-based rules for designing consistent EKF SLAM estimators," *Int. J. Robot. Res.*, vol. 29, no. 5, pp. 502–528, 2010.

[17] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Towards consistent vision-aided inertial navigation," in *Algorithmic Foundations of Robotics X*. Berlin, Germany: Springer, 2013, pp. 559–574.

[18] M. Li and A. I. Mourikis, "High-precision, consistent EKF-based visual-inertial odometry," *Int. J. Robot. Res.*, vol. 32, no. 6, pp. 690–711, 2013.

[19] M. Li and A. I. Mourikis, "Online temporal calibration for camera-IMU systems: Theory and algorithms," *Int. J. Robot. Res.*, vol. 33, no. 7, pp. 947–964, 2014.

[20] T. Qin and S. Shen, "Online temporal calibration for monocular visual-inertial systems," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 3662–3669.

[21] T. Schneider, M. Li, C. Cadena, J. Nieto, and R. Siegwart, "Observability-aware self-calibration of visual and inertial sensors for ego-motion estimation," *IEEE Sensors J.*, vol. 19, no. 10, pp. 3846–3860, Jan. 2019.

[22] Y. Chen, M. Zhang, D. Hong, C. Deng, and M. Li, "Perception system design for low-cost commercial ground robots: Sensor configurations, calibration, localization and mapping," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 6663–6670.

[23] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D pose estimation," Dept. Comp. Sci. Eng., Univ. Minnesota, Minneapolis, MN, USA, Tech. Rep. 2, 2005.

[24] T. Yap, M. Li, A. I. Mourikis, and C. R. Shelton, "A particle filter for monocular vision-aided odometry," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 5663–5669.

[25] K. J. Wu, C. X. Guo, G. Georgiou, and S. I. Roumeliotis, "VINS on wheels," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 5155–5162.

[26] M. Quan, S. Piao, M. Tan, and S.-S. Huang, "Tightly-coupled monocular visual-odometric SLAM using wheels and a MEMS gyroscope," *IEEE Access*, vol. 7, pp. 97374–97389, 2019.

[27] M. Zhang, X. Zuo, Y. Chen, Y. Liu, and M. Li, "Pose estimation for ground robots: On manifold representation, integration, reparameterization, and optimization," *IEEE Trans. Robot.*, vol. 37, no. 4, pp. 1081–1099, Aug. 2021.

[28] M. Zhang, Y. Chen, and M. Li, "Vision-aided localization for ground robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 2455–2461.

[29] K. Eckenhoff, P. Geneva, and G. Huang, "Closed-form preintegration methods for graph-based visual–inertial navigation," *Int. J. Robot. Res.*, vol. 38, no. 5, pp. 563–586, Apr. 2019.

[30] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 3565–3572.

[31] T. Qin, P. Li, and S. Shen, "VINS-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.

[32] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *Int. J. Robot. Res.*, vol. 34, no. 3, pp. 314–334, 2015.

[33] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2006, pp. 430–443.

[34] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 510–517.

[35] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[36] J. Kelly and G. S. Sukhatme, "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration," *Int. J. Robot. Res.*, vol. 30, no. 1, pp. 56–79, 2011.

[37] T. D. Barfoot, *State Estimation for Robotics*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[38] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*. New York, NY, USA: Academic, 1988.

[39] C. X. Guo, F. M. Mirzaei, and S. I. Roumeliotis, "An analytical least-squares solution to the odometer-camera extrinsic calibration problem," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2012, pp. 3962–3968.

[40] L. Heng, B. Li, and M. Pollefeys, "CamOdoCal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 1793–1800.

[41] A. Geiger, F. Moosmann, O. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2012, pp. 3936–3943.

[42] C. X. Guo and S. I. Roumeliotis, "IMU-RGBD camera 3D pose estimation and extrinsic calibration: Observability analysis and consistency improvement," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2013, pp. 2935–2942.

[43] Y. Yang and G. Huang, "Observability analysis of aided INS with heterogeneous features of points, lines, and planes," *IEEE Trans. Robot.*, vol. 35, no. 6, pp. 1399–1418, Dec. 2019.

[44] J. F. M. Van Doren, S. G. Douma, P. M. J. Van den Hof, J. D. Jansen, and O. H. Bosgra, "Identifiability: From qualitative analysis to model structure approximation," *IFAC Proc. Volumes*, vol. 42, no. 10, pp. 664–669, 2009.

[45] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual(-Inertial) odometry," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 7244–7251.

[46] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.

[47] J. Jeong, Y. Cho, Y.-S. Shin, H. Roh, and A. Kim, "Complex urban dataset with multi-level sensors from highly diverse urban environments," *Int. J. Robot. Res.*, vol. 38, no. 6, pp. 642–657, May 2019.

[48] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000km: The Oxford robotcar dataset," *Int. J. Robot. Res.*, vol. 36, no. 1, pp. 3–15, 2017.

[49] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, "University of Michigan north campus long-term vision and lidar dataset," *Int. J. Robot. Res.*, vol. 35, no. 9, pp. 1023–1035, Aug. 2016.

[50] C. H. Tong, D. Gingras, K. Larose, T. D. Barfoot, and É. Dupuis, "The Canadian planetary emulation terrain 3D mapping dataset," *Int. J. Robot. Res.*, vol. 32, no. 4, pp. 389–395, Apr. 2013.

[51] M. Li and A. I. Mourikis, "Vision-aided inertial navigation with rolling-shutter cameras," *Int. J. Robot. Res.*, vol. 33, no. 11, pp. 1490–1507, Sep. 2014.

Yiming Chen received the B.Eng. degree in automation from Tianjin University, Tianjin, China, in 2009, the dual M.S. degrees in electrical engineering and mathematics, and the Ph.D. degree in electrical engineering from the University of California, Riverside, USA, in 2011, 2013, and 2014, respectively.

He is currently a Research Engineer/the Manager with Alibaba Group. His current research interests include sensor fusion, state estimation, computer vision, and in particular, their applications in robotics.

Guoquan Huang (Senior Member, IEEE) received the B.S. degree in automation (electrical engineering) from the University of Science and Technology Beijing, China, in 2002, and the M.S. and Ph.D. degrees in computer science from the University of Minnesota–Twin Cities, in 2009 and 2012, respectively. He is currently an Associate Professor of mechanical engineering (ME), computer and information sciences (CIS), and electrical and computer engineering (ECE), at the University of Delaware (UD), where he is leading the Robot Perception and Navigation Group (RPNG). From 2012 to 2014, he was a Post-Doctoral Associate with the MIT CSAIL (Marine Robotics). His research interests focus on nonlinear estimation and spatial perception for autonomous vehicles and mobile devices, including probabilistic sensing, calibration, localization, mapping, tracking, perception, and navigation. He has received the 2015 UD Research Award (UDRF), the 2015 NASA DE Space Research Seed Award, the 2016 NSF CRII Award, the 2018 SATEC Robotics Delegation (one of ten U.S. experts invited by ASME), the 2018 Google Daydream Faculty Research Award, the 2019 Google AR/VR Faculty Research Award, and the Winner for the IROS 2019 FPV Drone Racing VIO Competition. He was also a recipient of the ICRA 2022 Outstanding Navigation Paper Award and the Finalists of the RSS 2009 Best Paper Award and ICRA 2021 Best Paper Award in Robot Vision. He serves as an Associate Editor for the IEEE TRANSACTIONS ON ROBOTICS (T-RO), IEEE ROBOTICS AND AUTOMATION LETTERS (RA-L), and *IET Cyber-Systems and Robotics* (CSR).

Xingxing Zuo received the B.Eng. degree in mechanical design, manufacturing and automation from the University of Electronic Science and Technology of China (UESTC) in 2016, and the Ph.D. degree in control science and engineering from Zhejiang University in 2021. He has been a Post-Doctoral Researcher with the Department of Informatics, Technical University of Munich, since 2021. His research interests include robot perception, 3-D computer vision, state estimation, sensor fusion, and localization and mapping for autonomous robots in complex environments. He was the finalist for the "Best Paper Award in Robot Vision" at the 2021 IEEE International Conference on Robotics and Automation (ICRA).

Mingming Zhang received the B.Eng. and M.S. degrees in mechatronics engineering from the State Key Laboratory of Robotics and System, Harbin Institute of Technology (HIT), Harbin, China, in 2014 and 2016, respectively.

He is currently a Staff Research Engineer/the Manager with Alibaba Group. His research interests include sensor fusion, large scale mapping, visual perception, and visual navigation, especially, their applications in robotics and mobile platforms.
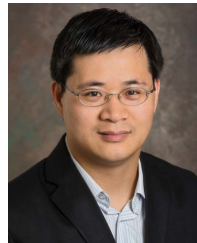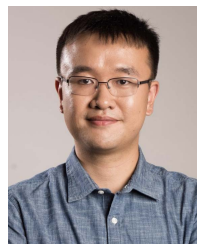
Yong Liu received the B.S. degree in computer science and engineering and the Ph.D. degree in computer science from Zhejiang University, Hangzhou, China, in 2001 and 2007, respectively.

He is currently a Professor with the Department of Control Science and Engineering, Institute of Cyber Systems and Control, Zhejiang University. His research interests include machine learning, robotics vision, information processing, and granular computing.

Mengmeng Wang received the B.S. and M.S. degrees in control science and engineering from Zhejiang University, Zhejiang, China, in 2015 and 2018, respectively, where she is currently pursuing the Ph.D. degree with the Laboratory of Advanced Perception on Robotics and Intelligent Learning, College of Control Science and Engineering. Her research interests include visual tracking, action recognition, computer vision, and deep learning.

Mingyang Li (Member, IEEE) received the B.Eng. degree in automation engineering from the University of Electronic Science and Technology of China (UESTC) in 2009, and the Ph.D. degree in electrical engineering from the University of California, Riverside, in 2014.

He is currently a Research Director at the Damo Academy, Alibaba Group. His research intrests include computer vision, sensor fusion, and machine learning in robotic and autonomous driving applications.