2023 IEEE International Conference on Robotics and Automation (ICRA 2023)
May 29 - June 2, 2023. London, UK

PIEKF-VIWO: Visual-Inertial-Wheel Odometry using Partial Invariant Extended Kalman Filter

Tong Hua, Tao Li and Ling Pei*

Abstract— Invariant Extended Kalman Filter (IEKF) has been successfully applied in Visual-inertial Odometry (VIO) as an advanced achievement of Kalman filter, showing great potential in sensor fusion. In this paper, we propose partial IEKF (PIEKF), which only incorporates rotation-velocity state into the Lie group structure and apply it for Visual-Inertial-Wheel Odometry (VIWO) to improve positioning accuracy and consistency. Specifically, we derive the rotation-velocity measurement model, which combines wheel measurements with kinematic constraints. The model circumvents the wheel odometer's 3D integration and covariance propagation, which is essential for filter consistency. And a plane constraint is also introduced to enhance the position accuracy. A dynamic outlier detection method is adopted, leveraging the velocity state output. Through the simulation and real-world test, we validate the effectiveness of our approach, which outperforms the standard Multi-State Constraint Kalman Filter (MSCKF) based VIWO in consistency and accuracy.

Index Terms: Invariant Extended Kalman Filter, sensor fusion, consistency

I. INTRODUCTION

Visual-Inertial Odometry (VIO) has recently been widely studied for its lightweight and high-precision advantage. Camera-IMU pose estimation is available in real commercial applications such as smart drones and mobile phones [1]. However, when visual and inertial sensors are deployed on ground vehicles, degenerative scenes like stationary and constant linear acceleration motions and vehicle vibration noise often lead to significant challenges in pose estimation and system observability. Thus, it is necessary to aid VIO with additional sensors such as wheel encoders [2].

Although the wheel measurements can improve the performance of VIO pose estimation, inconsistency may also negatively affect system state estimation. Experiments in [2] show that the online calibration of Visual-Inertial-Wheel Odometry (VIWO) will lead to high inconsistency with bad initial values. The analysis and improvement of VIO consistency have been mentioned in previous literature. Among them, Invariant Extended Kalman Filter (IEKF) maintains consistency theoretically and can be readily combined with other measurements such as the GPS. In contrast, other methods, such as the submap approach, are usually tailored [3]. Therefore, the introduction of invariance into VIWO is a better choice.

Meanwhile, the propagation matrix is highly related to the velocity and position state, which means IEKF heavily depends on the accuracy of noise parameters [4]. Experiments in [3], [5], [6] are all conducted in the low-speed and small-scale environments. [7] has shown that unpredictable disturbances or substantial outliers may violate the basic assumptions of IEKF and lead to inaccurate approximation to the system model. Unfortunately, state estimation in outdoor scenarios usually suffers from dynamic scene points, unstructured environments with homogeneous and non-textured surfaces [8], which poses a great challenge to the IEKF model.

To address the challenge, a partial IEKF-based VIWO (PIEKF-VIWO) is developed, which integrates the wheel encoder measurements into invariant MSCKF. The model improves the consistency and robustness of VIWO compared with the conventional IEKF model in practice. Our main contributions are as follows:

- A partial IEKF (PIEKF) is derived and applied for MSCKF-based VIWO, whose Lie group structure includes only the rotation and velocity.
- To fit the PIEKF framework, a rotation-velocity measurement model for the wheel odometer is proposed.
- To improve the robustness and accuracy, we introduce a PIEKF-based plane constraint model for the wheel odometer and a velocity-based dynamic outlier detection method for the camera.

II. RELATED WORKS

A. Visual-Inertial-Wheel Odometry

In recent years, a couple of pose estimation methods have incorporated wheel measurements into the visual-inertial navigation system [9], [10], [11], [12], [13]. The scale is proved to be observable by incorporating wheel encoder data, and a plane motion constraint is introduced to achieve higher position accuracy [9]. Based on the observability analysis, [14] proposes a bidirectional trajectory computation method that solves the unobservability before the first tuning. These works usually utilize relative pose constraints integrated by wheel measurements. Although the wheel pre-integration has high time efficiency, it loses velocity information and limits accuracy improvement.

*This work was supported in part by the National Nature Science Foundation of China (NSFC) under Grant 61873163, in part by the Shanghai Science and Technology Committee under Grant 20511103103, and in part by Equipment Preresearch Field Foundation under Grant 80913010303.

All authors are with Shanghai Key Laboratory of Navigation and Location Based Services, Shanghai Jiao Tong University. * Corresponding Author ling.pei@sjtu.edu.cn

[1], [15] extend the wheel encoder application from planar surface to manifolds and achieve high-accuracy 6D pose estimation. Kinematic constraints based on the instantaneous centers of rotation (ICR) model [16] are introduced into the optimization model of visual-inertial localization in [17]. Liu et al. present a tightly-coupled VIWO with an initialization module using wheel encoder readings and online extrinsic calibration [18]. Different strategies for motion tracking are provided in [12] to exploit IMU and wheel encoder cues maximally. However, most of the work is based on the nonlinear optimization method, which is time-consuming and does not focus on consistency. Lee et al. propose a consistent VIWO along with both extrinsic and intrinsic calibration of the wheel encoder and analyze the observability [2]. However, this method requires complex covariance propagation to guarantee system consistency.

B. Filter consistency

For the extended Kalman filter framework, linearization errors can cause filter inconsistencies, leading to overly optimistic covariance estimates, which have been discussed in rich literature (e.g., [19], [20], [3], [21], [22], [23]). There are a couple of solutions for the inconsistency problem. One is modifying the EKF linearization point represented by the First-Estimates Jacobian (FEJ) method [24], [25] which guarantees that the null space of the observability matrix does not degenerate. OC-EKF [26] explicitly enforces the system's observability constraints and maintains the null space. Another solution is robocentric odometry represented by ROVIO [27] and R-VIO [28] which originates from robocentric mapping [29]. This method avoids the alignment with the global gravity vector and the observability mismatch for the world-centric VIO. In recent years, Bonnabel et al. propose IEKF based on a novel Lie group structure [30], [31]. The consistency of this new scheme is demonstrated in [3] and further applied to MSCKF [6], [32] and UKF [33]. In particular, the invariant filter exhibits the potential for long-term inertial navigation when combined with some robust components [34].

III. PRELIMINARY KNOWLEDGE

Before describing our method, the coordinate frames and some notations involved are clarified in Table I. The wheel odometer and body frames are equivalent by default in the following discussion.

In the visual-inertial navigation system, the IMU state is given by:

$$\mathbf{X}_I = ({}^G_I\mathbf{R}, {}^G\mathbf{v}_I, {}^G\mathbf{p}_I, \mathbf{b}_g, \mathbf{b}_a) \tag{1}$$

The IMU continuous kinematic model is given by:

$$\begin{aligned} {}^G_I\dot{\mathbf{R}} &= {}^G_I\mathbf{R}\lfloor {}^I\boldsymbol{\omega} - \mathbf{b}_g - \mathbf{n}_g \rfloor \\ {}^G\dot{\mathbf{v}}_I &= {}^G_I\mathbf{R}({}^I\mathbf{a} - \mathbf{b}_a - \mathbf{n}_a) + {}^G\mathbf{g}, {}^G\dot{\mathbf{p}}_I = {}^G\mathbf{v}_I \\ \dot{\mathbf{b}}_g &= \mathbf{n}_{\omega g}, \dot{\mathbf{b}}_a = \mathbf{n}_{\omega a} \end{aligned} \tag{2}$$

TABLE I
Glossary of notation.

Symbol	Meaning	Symbol	Meaning
G	Global frame	I	IMU Frame
O	Wheel odometer Frame	C	Camera Frame
${}^B_A\mathbf{R}$	Rotation from frame A to frame B	${}^B\mathbf{v}_A$	Frame A's velocity in frame B
${}^B\mathbf{p}_A$	Frame A's position in frame B	$\mathbf{b}_g$	Gyroscope bias
$\mathbf{b}_a$	Acceleromter bias	$\hat{\mathbf{x}}$	Estimated value of $\mathbf{x}$
$\tilde{\mathbf{x}}$	Error value of $\mathbf{x}$	$\mathbf{e}_i$	The ith column of $\mathbf{I}_3$
$\lfloor \mathbf{a} \rfloor$	Skew symmetric matrix of $\mathbf{a} \in \mathbb{R}^3$	$exp(\cdot)$	Mapping from $\mathbb{R}^3$ to the manifold SO(3)
$log(\cdot)$	Mapping from SO(3) to $\mathbb{R}^3$	$\|\|\cdot\|\|$	2-norm of the vector

where ${}^I\boldsymbol{\omega}$ and ${}^I\mathbf{a}$ are the angular velocity and linear acceleration velocity respectively. $\mathbf{n}_g$ and $\mathbf{n}_a$ are the Gaussian noise of the gyroscope and accelerometer measurements respectively, and $\mathbf{n}_{wg}$ and $\mathbf{n}_{wa}$ are the Gaussian noise of gyroscope and accelerometer measurement biases. Base on $SE_2(3)$ proposed in [35], the IMU state on manifold can be constructed [32]:

$$\chi_I = \begin{bmatrix} {}^G_I\mathbf{R} & {}^G\mathbf{v}_I & {}^G\mathbf{p}_I & & & & & \\ & 1 & & & & & & \\ & & 1 & & & & & \\ & & & \mathbf{I}_3 & \mathbf{b}_g & & & \\ & & & & 1 & & & \\ & & & & & \mathbf{I}_3 & \mathbf{b}_a & \\ & & & & & & 1 & \end{bmatrix} \tag{3}$$

We define the right invariant error $\xi_I \in \mathbb{R}^{15}$ by the following retraction:

$$\chi_I = Exp(\xi_I)\hat{\chi}_I \tag{4}$$

where $Exp(\cdot)$ maps the vector $X \in \mathbb{R}^m$ into the manifold, and $\xi_I = \begin{bmatrix} \xi_\theta^T & \xi_v^T & \xi_p^T & \xi_{b_g}^T & \xi_{b_a}^T \end{bmatrix}^T$. And specifically the uncertainty representation of $\mathbf{X}_I$ is shown as follows:

$$\begin{aligned} \mathbf{X}_I = (&exp(\xi_\theta)\hat{\mathbf{R}}, exp(\xi_\theta)\hat{\mathbf{v}} + \mathbf{J}_l(\xi_\theta)\xi_v, exp(\xi_\theta)\hat{\mathbf{p}} + \mathbf{J}_l(\xi_\theta)\xi_p, \\ &\hat{\mathbf{b}}_g + \xi_{b_g}, \hat{\mathbf{b}}_a + \xi_{b_a}) \end{aligned} \tag{5}$$

where $\mathbf{J}_l(\cdot)$ is the left Jacobian for Lie Group. An invariant error state propagation equation is established.

$$\dot{\xi}_I = \mathbf{F}\xi_I + \mathbf{G}\mathbf{n}_I \tag{6}$$

where $\mathbf{n}_I = \begin{bmatrix} \mathbf{n}_g^T & \mathbf{n}_{wg}^T & \mathbf{n}_a^T & \mathbf{n}_{wa}^T \end{bmatrix}^T$. The details of $\mathbf{F}$ and $\mathbf{G}$ can refer to [6].

For MSCKF, the full state vector is augmented by the camera state:

$$\mathbf{X} = (\mathbf{X}_I, \mathbf{X}_{C1}, ..., \mathbf{X}_{Cn}) \tag{7}$$

where n is the number of camera states in the sliding window, and the camera state is represented by:

$$\mathbf{X}_C = ({}^G_C\mathbf{R}, {}^G\mathbf{p}_C) \tag{8}$$

The application of invariant MSCKF on VIO can refer to [6], [32].

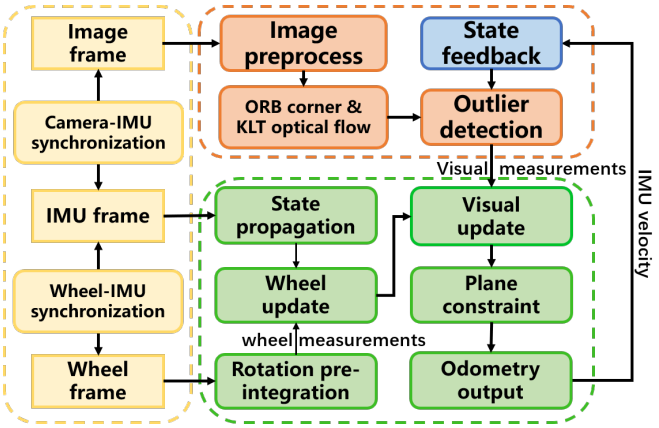


Fig. 1. PIEKF-based VIWO system overview.

IV. PIEKF

The main system components of PIEKF-based VIWO are shown in Fig. 1. In this section, we focus on the wheel measurement model and plane constraints with some details in the appendix. And the IMU-camera and IMU-wheel extrinsic calibration parameters are assumed to be perfectly known.

A. IMU kinematic model

Unlike (5), we only incorporate rotation and velocity into the Lie group structure, which means the filter has partial invariance. By decoupling the position with the rotation, we obtain a new uncertainty representation:

$$\mathbf{X}_I = (exp(\xi_\theta)\hat{\mathbf{R}}, exp(\xi_\theta)\hat{\mathbf{v}} + \mathbf{J}_l(\xi_\theta)\xi_v, \hat{\mathbf{p}} + \xi_p, \hat{\mathbf{b}}_g + \xi_{b_g}, \hat{\mathbf{b}}_a + \xi_{b_a}) \tag{9}$$

And the corresponding propagation matrices $\mathbf{F}$ and $\mathbf{G}$ are given by:

$$\mathbf{F} = \begin{bmatrix} \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -{}^G_I\hat{\mathbf{R}} & \mathbf{0}_{3\times3} \\ \lfloor {}^G\mathbf{g} \rfloor & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -\lfloor {}^G\hat{\mathbf{v}}_I \rfloor {}^G_I\hat{\mathbf{R}} & -{}^G_I\hat{\mathbf{R}} \\ -\lfloor {}^G\hat{\mathbf{v}}_I \rfloor & \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \end{bmatrix} \tag{10}$$

$$\mathbf{G} = \begin{bmatrix} {}^G_I\hat{\mathbf{R}} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \lfloor {}^G\hat{\mathbf{v}}_I \rfloor {}^G_I\hat{\mathbf{R}} & \mathbf{0}_{3\times3} & {}^G_I\hat{\mathbf{R}} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_3 \end{bmatrix} \tag{11}$$

The difference from the standard IEKF model is that the propagation matrix in (10) and (11) does not contains the estimated position ${}^G\hat{\mathbf{p}}_I$ explicitly.

B. Visual measurement model

The camera state is represented as the manifold form χ_C, which yields the state augmentation model:

$$\mathbf{P}_{aug} = \begin{bmatrix} \mathbf{I}_{15+6n} \\ \mathbf{J} \end{bmatrix} \mathbf{P}_{k+1|k} \begin{bmatrix} \mathbf{I}_{15+6n} \\ \mathbf{J} \end{bmatrix}^T \tag{12}$$

where the Jacobian $\mathbf{J}$ is written as:

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times(6+6n)} \\ -\lfloor {}^G_I\mathbf{R}^I\mathbf{p}_C \rfloor & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \mathbf{0}_{3\times(6+6n)} \end{bmatrix} \tag{13}$$

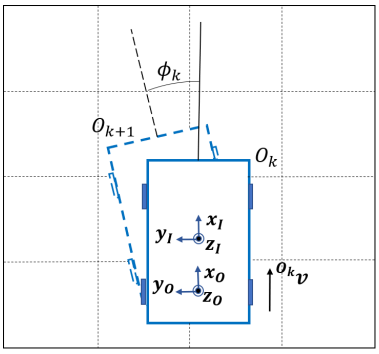


Fig. 2. The vehicle and sensor measurement model.

After initializing the positions of multiple observed features by the triangulation method, the normalized coordinate of the feature point ${}^G\mathbf{p}_f$ is obtained:

$$\mathbf{z} = \pi({}^C\mathbf{p}_f) = \frac{1}{{}^C\mathbf{z}_f} \begin{bmatrix} {}^C\mathbf{x}_f \\ {}^C\mathbf{y}_f \end{bmatrix} = h(\mathbf{X}, {}^G\mathbf{p}_f) \tag{14}$$

where π is the projection function. The corresponding error vector of $\mathbf{X}$ is denoted as $\xi = \begin{bmatrix} \xi_I^T & \xi_{C1}^T & \ldots \xi_{Cn}^T \end{bmatrix}^T \in \mathbb{R}^{15+6n}$. After linearization, the visual measurement model can be obtained as below:

$$\tilde{\mathbf{z}} = \mathbf{H}_X \xi + \mathbf{H}_f {}^G\tilde{\mathbf{p}}_f \tag{15}$$

The Jacobians $\mathbf{H}_x$ and $\mathbf{H}_f$ are given by:

$$\mathbf{H}_x = \mathbf{J}_\pi \begin{bmatrix} \mathbf{0}_{2\times15} & \ldots & \lfloor {}^C\mathbf{p}_f \rfloor {}^G_C\mathbf{R}^T & -{}^G_C\mathbf{R}^T & \ldots \end{bmatrix} \tag{16}$$

$$\mathbf{H}_f = \mathbf{J}_\pi {}^G_C\mathbf{R}^T \tag{17}$$

where $\mathbf{J}_\pi$ represents the Jacobian of π.

C. Wheel measurement model

Ground vehicles usually deploy two differential wheels with the 2D wheel odometer measurements introduced into the filter.

$${}^{O_t}w = \mathbf{e}_3^T {}^{O_t}\boldsymbol{\omega} + n_\omega \tag{18}$$

$${}^{O_t}v = \mathbf{e}_1^T {}^{O_t}\mathbf{v} + n_v \tag{19}$$

where ${}^{O_t}\boldsymbol{\omega}$ and ${}^{O_t}\mathbf{v}$ are the angular velocity and linear velocity [1] at time step t. As shown in Fig. 2, the relative rotation measurement is obtained as follows:

$$\phi_k = \int_{t_k}^{t_{k+1}} {}^{O_\tau}w d\tau \tag{20}$$

Different from relative pose constraints in [2], [9], [18], we adopt a rotation-velocity measurement model:

$${}^{O_k}_{O_{k+1}}\theta = \mathbf{e}_3^T log({}^O_I\mathbf{R}^{I_k}_G\mathbf{R}^G_{I_{k+1}}\mathbf{R}^I_O\mathbf{R}) \tag{21}$$

$${}^{O_k}\mathbf{v} = \lfloor {}^{I_k}\boldsymbol{\omega} \rfloor {}^O_I\mathbf{R}^I\mathbf{p}_O + {}^{O_k}_G\mathbf{R}^G\mathbf{v}_{I_k} \tag{22}$$

Using the right invariant error defined in (4), the residual and measurement Jacobian are given as below:

$$\mathbf{r}_\theta = \mathbf{e}_3^T log({}^{O_k}_{O_{k+1}}\hat{\mathbf{R}}^T {}^{O_k}_{O_{k+1}}\mathbf{R}) \tag{23}$$

$$\mathbf{H}_\theta = \mathbf{e}_3^T \begin{bmatrix} {}^{O_{k+1}}_G\hat{\mathbf{R}} & \mathbf{0}_{3\times12} & \ldots & -{}^{O_{k+1}}_G\hat{\mathbf{R}} & \mathbf{0}_{3\times3} \end{bmatrix} \tag{24}$$

where ${}^{O_k}_{O_{k+1}}\mathbf{R} = R_z(\phi_k)$ is the rotation measurement from the wheel odometer pre-integration. For velocity measurements, the non-holonomic constraints can be applied assuming that frame O is the body frame, i.e., the velocity on the cross-track and vertical direction should be zero. Thus the 3D residual and corresponding Jacobian are obtained.

$$\mathbf{r}_v = \begin{bmatrix} {}^{O_k}v & 0 & 0 \end{bmatrix}^T - {}^O_I\mathbf{R}\lfloor {}^{I_k}\omega \rfloor {}^I\mathbf{p}_O - {}^{O_k}_G\hat{\mathbf{R}}\,{}^G\hat{\mathbf{v}}_{I_k} \tag{25}$$

$$\mathbf{H}_v = \mathbf{e}_3^T \begin{bmatrix} \mathbf{0}_{3\times3} & {}^{O_k}_G\hat{\mathbf{R}} & \mathbf{0}_{3\times3} & {}^O_I\mathbf{R}\lfloor {}^I\mathbf{p}_O \rfloor & \mathbf{0}_{3\times(3+6n)} \end{bmatrix} \tag{26}$$

The Jacobian in (26) is irrelevant to rotation error, which is different from the conventional measurement model [36]. Additionally, velocity measurement update circumvents the 3D integration of wheel measurements and the covariance propagation compared with the relative displacement measurements in [2].

D. Plane constraint

Assuming that the vehicle is running on the plane π which is the initial x-y plane of frame G [9], the rotational and translational constraint is obtained:

$$\mathbf{z}_{rot} = \Lambda {}^\pi_G\mathbf{R}\,{}^G_{O_k}\mathbf{R}\mathbf{e}_3 \tag{27}$$

$$\mathbf{z}_{tran} = \mathbf{e}_3^T {}^\pi_G\mathbf{R}\,{}^G\mathbf{p}_{O_k} \tag{28}$$

where $\Lambda = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix}^T$.The corresponding Jacobians are given as below:

$$\mathbf{H}_{rot} = \Lambda \begin{bmatrix} -{}^\pi_G\mathbf{R}\lfloor {}^G_{O_k}\mathbf{R}\mathbf{e}_3 \rfloor & \mathbf{0}_{3\times(12+6n)} \end{bmatrix} \tag{29}$$

$$\mathbf{H}_{tran} = \mathbf{e}_3^T \begin{bmatrix} -{}^\pi_G\mathbf{R}\lfloor {}^G_{I_k}\mathbf{R}\,{}^I\mathbf{p}_O \rfloor & \mathbf{0}_{3\times3} & {}^\pi_G\mathbf{R} & \mathbf{0}_{3\times(6+6n)} \end{bmatrix} \tag{30}$$

The plane constraint update is performed after removing the redundant camera states.

V. VISUAL OUTLIER DETECTION

In the visual frontend, we preprocess the image using the contrast limited adaptive histogram equalization [37] (CLAHE) method to enhance the contrast of images. And the ORB corners are tracked based on KLT optical flow algorithm [38]. To detect outliers from tracked features, epipolar constraint is a common criterion used in many VIOs [39], [40]. However, this constraint may fail if the translations of the dynamic outliers and the vehicle are close to the same plane. Before describing our outlier detection method, we have three assumptions when the vehicle is approximately running in a straight line:

- The vehicle has a small rotation in two consecutive camera frames, i.e., the angular velocity ${}^I\omega \approx \mathbf{0}$.
- The extrinsic sensor parameters are reliable.
- The outlier has a constant velocity ${}^G\mathbf{v}_f$ between the two frames.

We start from the feature's 3D position in the camera frame:

$${}^C\mathbf{p}_f = \begin{bmatrix} X_f & Y_f & Z_f \end{bmatrix}^T = {}^C_G\mathbf{R}({}^G\mathbf{p}_f - {}^G\mathbf{p}_C) \tag{31}$$

TABLE II
Simulation noise configurations.

Parameter	Value
Gyroscope noise	$0.01 rad/s/\sqrt{Hz}$
Acceleration noise	$0.01 m/s^2/\sqrt{Hz}$
Gyroscope random walk	$0.0001 rad/s^2/\sqrt{Hz}$
Acceleration random walk	$0.0001 m/s^3/\sqrt{Hz}$
Wheel velocity noise	$0.1 m/s$
Wheel angular velocity noise	$0.001 rad/s$
Feature noise	1 pixel

Thus the coordinate in the normalized image plane is $\mathbf{x}_f = \begin{bmatrix} \frac{X}{Z} & \frac{Y}{Z} & 1 \end{bmatrix}^T$. The velocity of feature on the normalized image plane is approximated as:

$$\mathbf{v}_n \approx \frac{\Lambda({}^C v_{f,z}\mathbf{p}_f - {}^C\mathbf{v}_f)}{Z_f} \tag{32}$$

where ${}^C\mathbf{v}_f = {}^C_G\mathbf{R}({}^G\mathbf{v}_f - {}^G\mathbf{v}_C)$, ${}^C v_{f,z}$ is the z-axis component of ${}^C\mathbf{v}_f$. Z_f can be estimated easily by the triangulation method, and supposing each feature is static, i.e., ${}^C\mathbf{v}_{f,est} = -{}^C_G\mathbf{v}_C$, the estimated velocity $\mathbf{v}_{n,est}$ on the normalized plane can be obtained. The measured velocity can be approximated by:

$$\mathbf{v}_m = \frac{\mathbf{x}_f' - \mathbf{x}_f}{\Delta t} \tag{33}$$

where $\mathbf{x}_f$ and $\mathbf{x}_f'$ are the normalized coordinates in two consecutive frames respectively, and Δt is the time step. The velocity error is evaluated as the criterion.

$$\Delta v = ||\mathbf{v}_m - \mathbf{v}_{n,est}|| \tag{34}$$

For the static feature, Δv is close to zero, and for the dynamic outlier, the value depends on ${}^C\mathbf{v}_f$ and Z_f. Therefore we distinguish the outliers by setting the square mean value of all the errors as an adaptive threshold. It is noted that our method does not use the

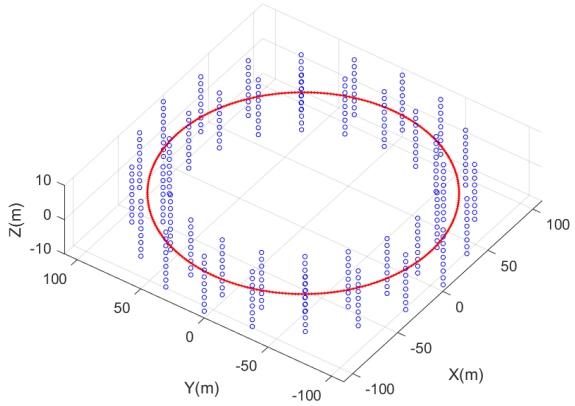
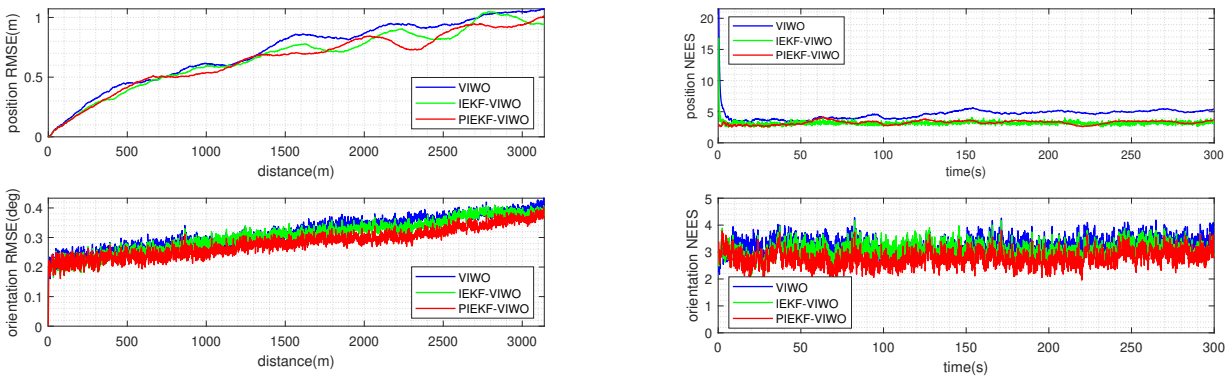


Fig. 3. The error-free landmarks (blue) and trajectory (red).



(a) (b)

Fig. 4. Monte Carlo simulation results. (a): RMSE; (b): ANEES

epipolar constraint, so it still works in the scene under the special motion.

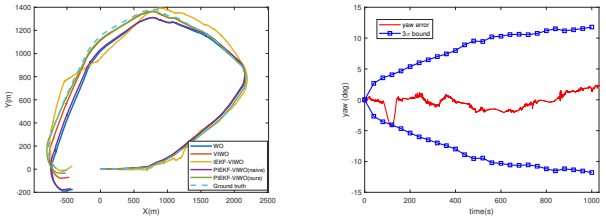
VI. EVALUATION

Both simulations and real-world tests are conducted to validate our proposed method. In the simulation test, we confirm the consistency and position accuracy of our proposed filter and evaluate the effectiveness of outlier detection. In the real-world test, we compare the position accuracy with other filter-based algorithms on Kaist urban dataset [41], and evaluate the plane constraints and time efficiency. Apart from PIEKF-VIWO, we have implemented the MSCKF-based VIWO (denoted as VIWO) and IEKF-based VIWO (denoted as IEKF-VIWO), both of which are the same as PIEKF-VIWO except the filtering framework. Our experiments are conducted on a laptop with Intel(R) Core(TM) i7-10710U CPU@1.10Ghz and 16G RAM.

A. Simulation

1) Consistency and accuracy: We assume that a car equipped with 100Hz IMU, 100Hz wheel odometer and 10Hz camera is moving along a circle on a plane with 360 landmarks scattered around the true trajectory shown in Fig. 3. We conduct the Monte Carlo simulation for 50 runs where sensor measurements are generated with random noise following the noise distribution in Table II. We compare VIWO, IEKF-VIWO, and our proposed PIEKF-VIWO by the accuracy indicator Root Mean Squared Error (RMSE) and the consistency indicator Averaged Normalized Estimation Error Squared (ANEES). The initial pose covariance for three algorithms is set to zero, and the results are shown in Fig. 4. Compared with VIWO and IEKF-VIWO, our proposed method has a slight gain in pose accuracy, whose RMSE is 0.648 m and 0.283 deg, respectively. The ideal NEES for orientation and position is three, and it is shown in Fig. 4(b) that PIEKF-VIWO performs similarly to IEKF-VIWO in consistency, better than the conventional VIWO especially in position NEES.

2) Outlier detection evaluation: We simulate a scenario with 100 static features and 20 dynamic features around the trajectory. The vehicle moves at 54 km/h and each dynamic feature moves at a velocity from a zero mean Gaussian distribution with a standard deviation $\sigma_v = 20$ m/s along each axis per frame. We sample the initial angular velocity from 0 to 0.1 rad/s. The outlier detection result for the first frame is depicted in Fig. 6. In the straight line scenario (Fig. 6(a)), there are a total of



(a) (b)

Fig. 5. Top view of estimated trajectories of algorithms (a) and yaw error with 3σ bound (b) for PIEKF-VIWO in Kaist *urban*32.

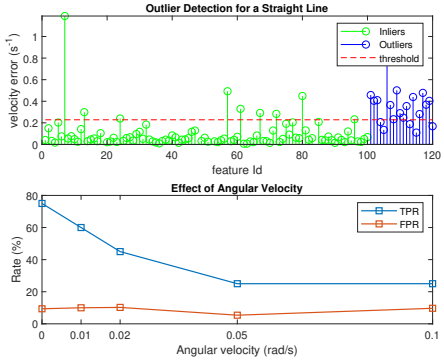


Fig. 6. Outlier detection for simulated features. Features 1-100 are static features, and 101-120 are dynamic features. True Positive Rate (TPR) and False Positive Rate (FPR) are compared for different angular velocities.

TABLE III
Orientation(deg)/Position(m) RMSE on the Kaist urban dataset.

Sequence(urban-)	Algorithm					
	WO	VIO	VIWO	IEKF-VIWO	PIEKF-VIWO(naive)	PIEKF-VIWO(ours)
29(3.2km)	4.13/13.72	4.14/119.07	0.81/2.41	1.27/9.16	1.44/7.59	**0.68**/**1.48**
30(4.7km)	16.64/145.47	63.25/409.59	2.11/**7.57**	16.48/167.30	4.25/20.26	**1.76**/8.39
31(10.7km)	14.47/363.05	22.54/811.06	8.25/177.97	16.01/494.75	24.45/477.40	**5.21**/**91.77**
32(6.4km)	4.00/104.84	×/×	2.14/44.17	12.15/74.88	3.50/92.66	**1.36**/**19.03**
33(7.3km)	10.96/67.34	×/×	4.00/29.52	8.54/79.62	7.11/87.88	**2.94**/**27.36**
34(6.5km)	4.97/85.19	×/×	**3.98**/52.54	7.57/80.57	5.75/58.58	4.72/**51.97**
35(3.2km)	3.55/4.48	13.74/253.85	**1.19**/**1.65**	4.91/59.07	5.02/33.40	2.30/9.07

[1] × means the failure. Bold and underline indicate best and second best in each sequence, respectively.

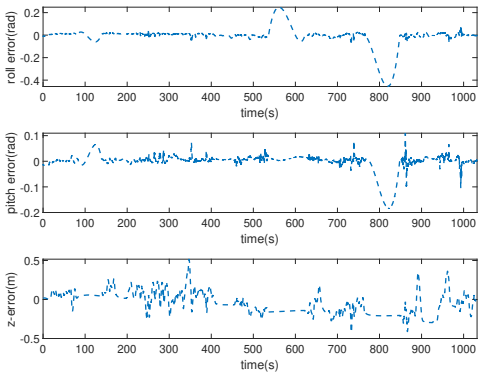


Fig. 7. Roll, pitch and z-axis position error for PIEKF-VIWO on Kaist *urban*32.

116 features tracked, among which 15 dynamic features are detected with a high true positive rate (75%). Fig. 6(b) demonstrates the detection effect is acceptable for this lightweight method when the angular velocity is small enough.

B. Real-world test

To evaluate the performance of our algorithm in the real scenario, KAIST urban 29-35 are chosen for the test, which is collected with the 100Hz IMU, 100Hz wheel encoder, and a 10Hz left camera. The pose computed from the graph SLAM is used as the ground truth. We select the image frame when the vehicle first stops as the starting frame as MSCKF requires a static initialization except for *urban*35. Since *urban*35 does not contain a stationary period, we have to set the initial vehicle velocity to the velocity measured by the wheel encoder. We choose Wheel Odometry (WO), monocular VIO [42] and PIEKF-VIWO without outlier detection and plane constraints (PIEKF-VIWO(naive)) as benchmarks.

1) Accuracy and consistency: Table III lists the RMSE for different algorithms, and the estimated trajectories of *urban*32 are depicted in Fig. 5(a) as an example. It is noted that, unlike the simulation, IEKF-VIWO fails to achieve a good performance, especially in the dataset sequence with long-distance traveling and high speed, such as *urban*31. With the help of the PIEKF framework and additional constraints, PIEKF-VIWO is less vulnerable to noise and performs much better than IEKF and PIEKF-VIWO(naive). It is noted that PIEKF-VIWO is badly-behaved in *urban*35 because of the dynamic initialization. But for most cases, PIEKF-VIWO shows better results regarding both position and orientation accuracy than the conventional VIWO. The estimated yaw result in *urban*32 is also plot in Fig. 5(b), where the error remains in the 3σ boundary, demonstrating the consistency of our algorithm.

2) Plane constraint evaluation: We demonstrate the effect of plane constraints in the partial IEKF algorithm. Real trajectories may have bumpy sections, so we relax the measurement error covariance to 0.01. The estimation error tested on the *urban*32 is depicted in Fig. 7. The estimated roll, pitch, and position along the z-axis are maintained at around zero except in some sections where the constraints are not satisfied.

3) Runtime comparison: Since the algorithm performs wheel measurement updates at a high frequency, we compare the efficiency of our algorithms against VIO by computing the averaged backend processing time per frame for all the tested sequences, which is 2.46 (VIO), 5.84 (IEKF-VIWO) and 5.89 (PIEKF-VIWO) (unit: ms) respectively. IEKF-VIWO and PIEKF-VIWO have approximately the same efficiency in sequence as expected. Though our algorithms consume about twice as much time as VIO, their time consumption remains within 6ms, ensuring real-time in the test.

VII. CONCLUSIONS

This paper proposes a PIEKF-based state estimation algorithm fusing camera, IMU, and wheel odometer. We compare our algorithm with conventional VIWO and demonstrate its better consistency and accuracy through the simulation and real-world experiments. In the future, our work will enhance robustness by introducing online calibration and improving the visual frontend.

References

[1] M. Zhang, X. Zuo, Y. Chen, Y. Liu, and M. Li, "Pose estimation for ground robots: On manifold representation, integration, reparameterization, and optimization," IEEE Transactions on Robotics, 2021.

[2] W. Lee, K. Eckenhoff, Y. Yang, P. Geneva, and G. Huang, "Visual-inertial-wheel odometry with online calibration," in 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2020, pp. 4559–4566.

[3] A. Barrau and S. Bonnabel, "An EKF-SLAM algorithm with consistency properties," arXiv preprint arXiv:1510.06263, 2015.

[4] J. Wang and M. Li, "Covariance regulation based invariant Kalman filtering for attitude estimation on matrix Lie groups," IET Control Theory & Applications, vol. 15, no. 15, pp. 2017–2025, 2021.

[5] R. Hartley, M. Ghaffari, R. M. Eustice, and J. W. Grizzle, "Contact-aided invariant extended Kalman filtering for robot state estimation," The International Journal of Robotics Research, vol. 39, no. 4, pp. 402–430, 2020.

[6] K. Wu, T. Zhang, D. Su, S. Huang, and G. Dissanayake, "An invariant-EKF vins algorithm for improving consistency," in 2017 IEEE/RSJ international conference on intelligent robots and systems, 2017, pp. 1578–1585.

[7] J. Wang, C. Zhang, J. Wu, and M. Liu, "An improved invariant Kalman filter for Lie groups attitude dynamics with heavy-tailed process noise," Machines, vol. 9, no. 9, p. 182, 2021.

[8] M. Buczko and V. Willert, "How to distinguish inliers from outliers in visual odometry for high-speed automotive applications," in 2016 IEEE Intelligent Vehicles Symposium, 2016, pp. 478–483.

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

[10] Z. Dang, T. Wang, and F. Pang, "Tightly-coupled data fusion of vins and odometer based on wheel slip estimation," in 2018 IEEE International Conference on Robotics and Biomimetics, 2018, pp. 1613–1619.

[11] R. Kang, L. Xiong, M. Xu, J. Zhao, and P. Zhang, "Vins-vehicle: A tightly-coupled vehicle dynamics extension to visual-inertial state estimator," in 2019 IEEE Intelligent Transportation Systems Conference, 2019, pp. 3593–3600.

[12] 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. 97 374–97 389, 2019.

[13] J. H. Jung, J. Cha, J. Y. Chung, T. I. Kim, M. H. Seo, S. Y. Park, J. Y. Yeo, and C. G. Park, "Monocular visual-inertial-wheel odometry using low-grade IMU in urban areas," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 2, pp. 925–938, 2020.

[14] J. Liu, W. Gao, and Z. Hu, "Bidirectional trajectory computation for odometer-aided visual-inertial SLAM," IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 1670–1677, 2021.

[15] M. Zhang, Y. Chen, and M. Li, "Vision-aided localization for ground robots," in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2019, pp. 2455–2461.

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

[17] X. Zuo, M. Zhang, Y. Chen, Y. Liu, G. Huang, and M. Li, "Visual-inertial localization for skid-steering robots with kinematic constraints," in The International Symposium of Robotics Research. Springer, 2019, pp. 741–756.

[18] J. Liu, W. Gao, and Z. Hu, "Visual-inertial odometry tightly coupled with wheel encoder adopting robust initialization and online extrinsic calibration," in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2019, pp. 5391–5397.

[19] T. Bailey, J. Nieto, J. Guivant, M. Stevens, and E. Nebot, "Consistency of the EKF-SLAM algorithm," in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006, pp. 3562–3568.

[20] S. Huang and G. Dissanayake, "Convergence and consistency analysis for extended Kalman filter based SLAM," IEEE Transactions on robotics, vol. 23, no. 5, pp. 1036–1049, 2007.

[21] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "Analysis and improvement of the consistency of extended Kalman filter based SLAM," in 2008 IEEE International Conference on Robotics and Automation, 2008, pp. 473–479.

[22] M. Li and A. I. Mourikis, "High-precision, consistent EKF-based visual-inertial odometry," The International Journal of Robotics Research, vol. 32, no. 6, pp. 690–711, 2013.

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

[24] C. Chen, Y. Yang, P. Geneva, and G. Huang, "Fej2: A consistent visual-inertial state estimator design," in Proc. International Conference on Robotics and Automation, Philadelphia, USA, 2022.

[25] G. P. Huang, A. I. Mourikis, and S. I. Roumeliotis, "A first-estimates Jacobian EKF for improving SLAM consistency," in Experimental Robotics. Springer, 2009, pp. 373–382.

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

[27] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," in IEEE/RSJ International Conference on Intelligent Robots & Systems, 2015, pp. 298–304.

[28] Z. Huai and G. Huang, "Robocentric visual-inertial odometry," in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2018, pp. 6319–6326.

[29] J. A. Castellanos, J. Neira, and J. D. Tardós, "Limits to the consistency of EKF-based SLAM," IFAC Proceedings Volumes, vol. 37, no. 8, pp. 716–721, 2004.

[30] S. Bonnabel, "Left-invariant extended Kalman filter and attitude estimation," in 2007 46th IEEE Conference on Decision and Control, 2007, pp. 1027–1032.

[31] S. Bonnable, P. Martin, and E. Salaün, "Invariant extended Kalman filter: theory and application to a velocity-aided attitude estimation problem," in Proceedings of the 48h IEEE Conference on Decision and Control held jointly with 2009 28th Chinese Control Conference, 2009, pp. 1297–1304.

[32] S. Heo and C. G. Park, "Consistent EKF-based visual-inertial odometry on matrix Lie group," IEEE Sensors Journal, vol. 18, no. 9, pp. 3780–3788, 2018.

[33] M. Brossard, S. Bonnabel, and A. Barrau, "Invariant Kalman filtering for visual inertial SLAM," in 2018 21st International Conference on Information Fusion, 2018, pp. 2021–2028.

[34] M. Brossard, A. Barrau, and S. Bonnabel, "RINS-W: Robust inertial navigation system on wheels," in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2019, pp. 2068–2075.

[35] A. Barrau and S. Bonnabel, "The invariant extended Kalman filter as a stable observer," IEEE Transactions on Automatic Control, vol. 62, no. 4, pp. 1797–1812, 2016.

[36] E.-H. Shin and N. El-Sheimy, "Accuracy improvement of low cost ins/gps for land applications," in National Technical Meeting of the Institute of Navigation, 2002, pp. 146–157.

[37] K. Zuiderveld, "Contrast limited adaptive histogram equalization," Graphics gems, pp. 474–485, 1994.

[38] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in 7^{th} International Joint Conference on Artificial Intelligence, vol. 2, 1981, pp. 674–679.

[39] K. Sun, K. Mohta, B. Pfrommer, M. Watterson, S. Liu, Y. Mulgaonkar, C. J. Taylor, and V. Kumar, "Robust stereo visual inertial odometry for fast autonomous flight," IEEE Robotics and Automation Letters, vol. 3, no. 2, pp. 965–972, 2018.

[40] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," IEEE Transactions on Robotics, vol. 34, no. 4, pp. 1004–1020, 2018.

[41] J. Jeong, Y. Cho, Y.-S. Shin, H. Roh, and A. Kim, "Complex urban dataset with multi-level sensors from highly diverse

urban environments," The International Journal of Robotics Research, vol. 38, no. 6, pp. 642–657, 2019.
[42] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in 2007 IEEE International Conference on Robotics and Automation, 2007, pp. 3565–3572.