Invariant Extended Kalman Filtering for Pedestrian Deep-Inertial Odometry

Shiyu Bai[1], Weisong Wen[1], Yue Yu[2], Li-Ta Hsu[1]

[1] Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong - shiyu.bai@polyu.edu.hk, welson.wen@polyu.edu.hk, lt.hsu@polyu.edu.hk

[2] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong - michael-yue.yu@polyu.edu.hk

Keywords: Invariant Extended Kalman Filtering, Deep learning, Inertial odometry, State estimation, Pedestrian localization, Low-cost.

Abstract

Indoor localization for pedestrians, which relies solely on inertial odometry, has been a topic of great interest. Its significance lies in its ability to provide positioning solutions independently, without the need for external data. Although traditional strap-down inertial navigation shows rapid drift, the introduction of pedestrian dead reckoning (PDR), and artificial intelligence (AI) has enhanced the applicability of inertial odometry for indoor localization. However, inertial odometry continues to be affected by drift, inherent to the nature of dead reckoning. This implies that even a slight error at a given moment can lead to a significant decrease in accuracy after continuous integration operations. In this paper, we propose a novel approach aimed at enhancing the positioning accuracy of inertial odometry. Firstly, we derive a learning-based forward speed using inertial measurements from a smartphone. Unlike mainstream methods where the learned speed is directly used to determine the position, we use the forward speed combined with non-holonomic constraint (NHC) as a measurement to update the state predicted within a strap-down inertial navigation framework. Secondly, we employ an invariant extended Kalman filter (IEKF)-based state estimation to facilitate fusion to cope with the non-linearity arising from the system and measurement model. Experimental tests are carried out in different scenarios using an iPhone 12, and traditional methods, including PDR, robust neural inertial navigation (RONIN), and the EKF-based method, are compared. The results suggest that the method we propose surpasses these traditional methods in performance.

1. Introduction

Indoor pedestrian localization is crucial for a range of emerging applications, including virtual/augmented reality (VR/AR) (Yi et al., 2021), car finder services (Ren et al., 2024), and emergency response (Tseng et al., 2022). At present, the primary solutions for indoor localization rely on radio frequency (RF) signals, such as Wi-Fi, Bluetooth low energy (BLE), and ultra-wideband (UWB) (Liu et al., 2007). While these tools can provide satisfactory localization performance, they are dependent on existing infrastructure and necessitate prior calibration. This dependency can increase costs and manpower requirements, thus restricting their application. Moreover, their performance can be easily affected by external interference and non-line-of-sight (NLOS) conditions (Yang et al., 2021). Perception-based localization methods, such as those using cameras (Qin et al., 2018) and light detection and ranging (LiDAR) (Zhang and Singh, 2014), do not rely on infrastructures and have demonstrated superior positioning accuracy in indoor environments. However, most related algorithms for perception are challenging to implement on mobile devices for pedestrian use due to the requirements for corresponding hardware and computational capabilities.

Inertial measurement units (IMUs), as low-cost sensors, are extensively integrated into existing mobile devices such as smartphones and smartwatches. They are typically utilized for applications like games and screen orientation changes. In addition, when other information is not available, IMUs can be utilized to establish inertial odometry, thereby providing an alternative solution for indoor pedestrian localization. In current solutions, inertial odometry is typically utilized as the backbone for achieving SLAM or sensor fusion (Bai et al., 2023), as it facilitates completely egocentric motion tracking (Chen and Pan, 2024). Inertial odometry can be divided into three main categories: strap-down inertial navigation, pedestrian dead reckoning (PDR), and data-driven inertial navigation (Wang et al., 2022). Strap-down inertial navigation, a classic form of dead reckoning, utilizes data from an inertial measurement unit (IMU) to estimate an object's position, velocity, and attitude. However, the IMUs typically used for indoor pedestrian localization are low-cost and can cause rapid drift when strap-down inertial navigation is applied. To mitigate this drift, researchers utilize a zero-velocity update (ZUPT) (Wang et al., 2015). Despite the improved accuracy, it is worth noting that ZUPT is typically employed with foot-mounted IMUs, which are not frequently used by pedestrians in their daily lives. On the other hand, PDR is an effective method that can achieve less drift than strap-down inertial navigation and is compatible with mobile devices like smartphones. PDR operates based on step detection. The lengths and directions of steps are subsequently estimated to determine the position. However, the performance of PDR can easily degrade when the pedestrian's motion state changes. Additionally, inaccuracies in heading also significantly impact the performance of PDR.

Recently, with the rapid advancement of artificial intelligence (AI), data-driven inertial navigation methods have been shown to effectively suppress error accumulation. Chen introduced IONet (Chen et al., 2018), a method that utilizes a window length of IMU data to calculate changes in distance and heading. Given an initial location and heading, this method can determine a pedestrian's location. Notably, this approach outperforms PDR and strap-down inertial navigation. Herath proposed robust neural inertial navigation (RONIN) (Herath et al., 2020), in which horizontal positions and headings can

be derived from a sequence of IMU data. RONIN encompasses three variants, each of which can be operated using ResNet, long short-term memory (LSTM), and temporal convolutional network (TCN). Wang presented a loss formulation for smartphone-based inertial odometry (Wang et al., 2021). The velocity was represented as the average velocity magnitude and moving direction, which can improve the trajectory estimation. In addition to directly predicting a pedestrian's position using learned displacement or velocity, the learned measurement can be integrated with strap-down inertial navigation to estimate the position. Cortés used a convolutional neural network (CNN)-based deep learning model to regress the instantaneous speed, which was then used to update the inertial navigation system to obtain the pose (Cortés et al., 2018). Liu developed TLIO (Liu et al., 2020). The 3D displacement in a local gravity-aligned frame was used as the measurement, and a stochastic cloning extended Kalman filter (EKF) was employed to determine the state.

Currently, the state estimation methods for the integration can be primarily divided into filtering and optimization. The comparison between them remains a key topic in the navigation and positioning community (Wen et al., 2019). While the factor graph can surpass the standard EKF in terms of estimation accuracy, it can impose a higher computational burden. This becomes a significant concern when it is implemented on widely used mobile devices such as smartphones. Recently, the invariant extended Kalman filtering (IEKF), an extension of the EKF, has shown successful results in state estimation and navigation (Barrau and Bonnabel, 2016). The IEKF has been mathematically proven to exhibit strong convergence and consistency properties. Potokar introduced IEKF-based underwater localization using IMU and doppler velocity logs (DVLs) (Potokar et al., 2021). A Monte Carlo simulation was conducted, and the results show that the proposed method outperforms a quaternion-based EKF in terms of accuracy and convergence speed. Wang proposed an IEKF-based pedestrian motion tracking (Wang et al., 2020). The average speed, which was derived from the estimated position, was used as the measurement. Additionally, a measurement noise parameter adapter was designed to enhance the accuracy of the estimation.

In this paper, we investigate a novel inertial odometry for indoor localization on smartphones. To the best of our knowledge, this is the first paper to develop a deep-inertial odometry (DIO) within an IEKF framework. The main contributions of this paper are as follows:

1) We propose a deep-inertial odometry based on velocity updates. In this method, a neural network is utilized to regress forward speed from IMU data. This speed is then combined with non-holonomic constraint (NHC) to assist strap-down inertial navigation. This method can achieve superior performance by avoiding the influence of heading uncertainty.

2) We formulate an integration model based on the IEKF, which uses learned speed and strap-down inertial navigation to enhance the accuracy of state estimation, offering enhanced performance compared to the EKF-based method. Additionally, we provide the derivation of the model used in deep-inertial odometry.

3) For evaluation purposes, we carry out experiments across different scenarios. We use conventional PDR, RONIN, and EKF-based methods for comparative analysis.

The structure of this paper is as follows: Section 2 presents the learning-based forward speed model, while Section 3 outlines the formulation of the IEKF-based DIO. The results of experimental tests are discussed in Section 4. Finally, Section 5 provides the conclusion of the paper.

2. Learning-based Forward Speed Model

Regarding the training data, we utilized an iPhone 12 to collect both IMU data and ground truth. With the assistance of ARKit, we can obtain a visual-inertial odometry (VIO) solution, which served as the ground truth. Despite the drift experienced by VIO in the absence of loop closure, we can use the velocity derived from VIO as the training target, considering it a drift-free measurement. A total of 73 sequences were collected from various parts of the Hong Kong Polytechnic University campus, amounting to approximately 5 hours of data.

The proposed method utilizes IMU data to estimate the forward speed of the pedestrian. This estimation is then combined with the NHC, forming the measurement to update the prediction in the filter. This method effectively mitigates the impact of heading uncertainty in real-world scenarios. To optimize the regression performance of the forward speed, the collected data must be transformed into corresponding frames. In this paper, we define three distinct frames to describe the transformation relationships among them. These frames are referred to as the world frame, the smartphone frame, and the body frame, symbolized by w, s, and b, respectively. The world frame serves as the reference for navigation. The smartphone and body frames are depicted in Figure 1. The x_s and y_s axes of the smartphone frame are defined along the longitudinal and lateral axes, respectively. The z_s axis is oriented perpendicularly to the smartphone screen. The x_b and y_b axes of the body frame are oriented towards the front and left of the human body, respectively. The plane formed by the x_b and y_b axes is assumed to be parallel to the ground. The z_b axis is perpendicular to this plane.

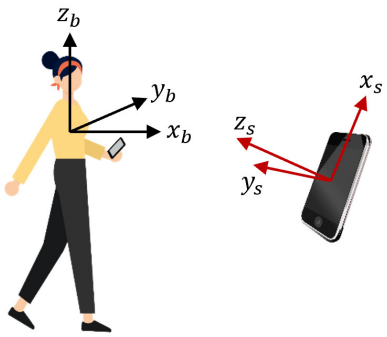

Figure 1. The illustration of smartphone and body frames.

Based on the position of VIO within the world frame, we can determine its velocity in the same frame by applying a differential operation, denoted as v^{wx} and v^{wy}, respectively. By computing the norm, we can derive the forward speed from the ground truth. This can be articulated as follows:

$$v^b = \sqrt{(v^{wx})^2 + (v^{wy})^2} \tag{1}$$

To mitigate the influence of different holding attitudes for the smartphone, the IMU data need to be transformed to the body

frame, which can be expressed as:

$$\begin{aligned}\hat{\mathbf{a}}_t^b &= \hat{\mathbf{R}}_s^b \tilde{\mathbf{a}}_t \\ \hat{\mathbf{w}}_t^b &= \hat{\mathbf{R}}_s^b \tilde{\mathbf{w}}_t\end{aligned} \tag{2}$$

where $\hat{\mathbf{R}}_s^b$ is the rotation matrix from the smartphone to the body frame. The form of $\hat{\mathbf{d}}$ denotes the estimated variable in this paper, where $\mathbf{d}$ is a general symbol. $\tilde{\mathbf{a}}_t$ and $\tilde{\mathbf{w}}_t$ are the IMU data at the time instant t, and $\hat{\mathbf{a}}_t^b$ and $\hat{\mathbf{w}}_t^b$ are the transformed IMU data. In $\hat{\mathbf{R}}_s^b$, the angles along the x and y axes can be determined by the roll and pitch of the smartphone. Given that the smartphone is held in front of the body by the pedestrian, the angle along the z axis is set to zero in this paper. In addition, by removing the gravity vector from $\hat{\mathbf{a}}_t^b$, we can obtain the components of linear acceleration, denoted as $\hat{\mathbf{l}}_t^b$.

This paper leverages a 1D version of ResNet-18 architecture, adding a fully connected layer at the end to regress the forward speed of the pedestrian. The input dimension of the network is $n \times 6$, which includes n IMU data in the pedestrian's body frame, The regression of forward speed can be expressed as follows:

$$\hat{v}^b = f_{ResNet}\left(\hat{\mathbf{l}}_{1:n}^b, \hat{\mathbf{w}}_{1:n}^b\right) \tag{3}$$

The loss function is defined as the mean square error (MSE) form, which is expressed as:

$$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{v}_i^b - v_i^b\right)^2 \tag{4}$$

where N denotes the amount of data in the training dataset.

In the model training, the model is implemented using PyTorch and the training process is done through NVIDIA GeForce RTX 4090. The Adam optimizer (Kingma and Ba, 2014) is used with an initial learning rate of 0.0001, zero weight decay, and dropouts with a probability of 0.5 for the FC layers. The models are trained for 1000 epochs.

3. Deep-Inertial Odometry using Invariant Extended Kalman Filtering

The framework of the proposed method, as depicted in Figure 2, operates as follows: The specific force and angular velocity from a built-in smartphone IMU serve as the input. Strap-down inertial navigation is implemented to predict the motion state, which is then used for error propagation. A neural network regresses the forward speed, which is used to update the state in an IEKF-based integration. Finally, the estimated state is output, which is then used for prediction and frame transformation.

3.1 State Prediction

The state variables in this paper are expressed as:

$$\mathbf{x}_k = (\mathbf{R}_k,\ \mathbf{v}_k, \mathbf{p}_k, \mathbf{b}_k^w, \mathbf{b}_k^a) \tag{5}$$

where $\mathbf{x}_k$ is the state at the time instant t_k, which includes the rotation, velocity, position, gyroscope bias, and accelerometer bias. The state propagation based on IMU data can be expressed as:

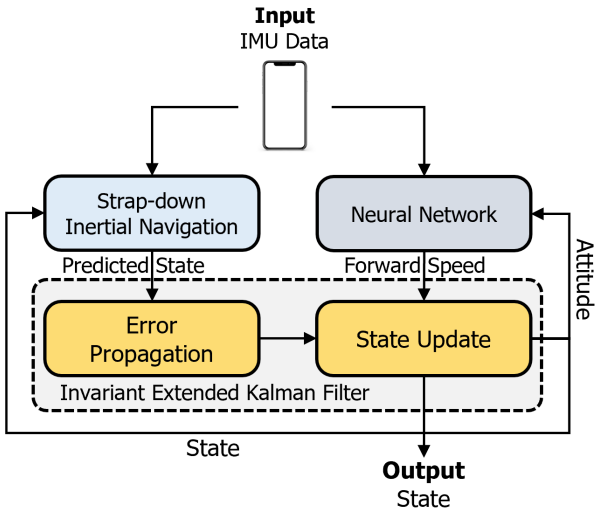

Figure 2. The structure of the proposed method. Smartphone IMU data is utilized for both strap-down inertial navigation and the neural network. The speed, regressed by the network, is then used to update the state in an IEKF-based integration.

$$\begin{aligned}\dot{\mathbf{R}}_t &= \mathbf{R}_t\left[\tilde{\mathbf{w}}_t - \mathbf{b}_t^w - \mathbf{n}_t^w\right]_\times \\ \dot{\mathbf{v}}_t &= \mathbf{R}_t\left(\tilde{\mathbf{a}}_t - \mathbf{b}_t^a - \mathbf{n}_t^a\right) + \mathbf{g} \\ \dot{\mathbf{p}}_t &= \mathbf{v}_t \\ \dot{\mathbf{b}}_t^w &= \mathbf{n}_t^{b_w} \\ \dot{\mathbf{b}}_t^a &= \mathbf{n}_t^{b_a}\end{aligned} \tag{6}$$

where $\tilde{\mathbf{w}}_t$ and $\tilde{\mathbf{a}}_t$ represent the measured angular velocity and specific force, respectively. $\mathbf{n}_t^w$ and $\mathbf{n}_t^a$ represent the noise from the gyroscope and accelerometer. IMU biases are modelled as a random walk, with their derivatives represented by $\mathbf{n}_t^{b_w}$ and $\mathbf{n}_t^{b_a}$.

The state variables of the rotation, velocity, and position can form the matrix Lie Group. Therefore, the state can be rewritten as:

$$\begin{aligned}\boldsymbol{\chi}_t &= \begin{pmatrix}\mathbf{R}_t & \mathbf{v}_t & \mathbf{p}_t \\ \mathbf{0}_{1\times3} & 1 & 0 \\ \mathbf{0}_{1\times3} & 0 & 1\end{pmatrix} \\ \boldsymbol{\Theta}_t &= \begin{pmatrix}\mathbf{b}_t^w \\ \mathbf{b}_t^a\end{pmatrix}\end{aligned} \tag{7}$$

This paper adopts the right invariant error. The error of $\boldsymbol{\chi}_t$ can be given by:

$$\boldsymbol{\eta}_t = \hat{\boldsymbol{\chi}}_t \boldsymbol{\chi}_t^{-1} \tag{8}$$

where

$$\boldsymbol{\chi}_t^{-1} = \begin{pmatrix}\mathbf{R}_t^{\mathrm{T}} & -\mathbf{R}_t^{\mathrm{T}}\mathbf{v}_t & -\mathbf{R}_t^{\mathrm{T}}\mathbf{p}_t \\ \mathbf{0}_{1\times3} & 1 & 0 \\ \mathbf{0}_{1\times3} & 0 & 1\end{pmatrix} \tag{9}$$

Then, $\boldsymbol{\eta}_t$ can be expressed as:

$$\begin{aligned}\boldsymbol{\eta}_t &= \begin{pmatrix}\hat{\mathbf{R}}_t\mathbf{R}_t^{\mathrm{T}} & \hat{\mathbf{v}}_t - \hat{\mathbf{R}}_t\mathbf{R}_t^{\mathrm{T}}\mathbf{v}_t & \hat{\mathbf{p}}_t - \hat{\mathbf{R}}_t\mathbf{R}_t^{\mathrm{T}}\mathbf{p}_t \\ \mathbf{0}_{1\times3} & 1 & 0 \\ \mathbf{0}_{1\times3} & 0 & 1\end{pmatrix} \\ &= \begin{pmatrix}\boldsymbol{\eta}_{R_t} & \boldsymbol{\xi}_{v_t} & \boldsymbol{\xi}_{p_t} \\ \mathbf{0}_{1\times3} & 1 & 0 \\ \mathbf{0}_{1\times3} & 0 & 1\end{pmatrix}\end{aligned} \tag{10}$$

Given that the right invariant error is used, the error of IMU bias can be expressed as:

$$\zeta_t = \begin{pmatrix} \hat{\mathbf{b}}_t^w - \mathbf{b}_t^w \\ \hat{\mathbf{b}}_t^a - \mathbf{b}_t^a \end{pmatrix} = \begin{pmatrix} \zeta_t^w \\ \zeta_t^a \end{pmatrix} \tag{11}$$

Let $\boldsymbol{\xi}_{R_t}$ be the Lie algebra of $\boldsymbol{\eta}_{R_t}$, the system equation of the state error can be expressed as:

$$\begin{pmatrix} \dot{\boldsymbol{\xi}}_{R_t} \\ \dot{\boldsymbol{\xi}}_{v_t} \\ \dot{\boldsymbol{\xi}}_{p_t} \\ \dot{\zeta}_t^w \\ \dot{\zeta}_t^a \end{pmatrix} = \hat{\mathbf{F}}_t \begin{pmatrix} \boldsymbol{\xi}_{R_t} \\ \boldsymbol{\xi}_{v_t} \\ \boldsymbol{\xi}_{p_t} \\ \zeta_t^w \\ \zeta_t^a \end{pmatrix} + \hat{\mathbf{G}}_t \begin{pmatrix} \mathbf{n}_t^w \\ \mathbf{n}_t^a \\ \mathbf{0}_{3\times1} \\ \mathbf{n}_t^{b_w} \\ \mathbf{n}_t^{b_a} \end{pmatrix} \tag{12}$$

where $\hat{\mathbf{F}}_t$ is the state transition matrix, given as:

$$\hat{\mathbf{F}}_t = \begin{pmatrix} \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -\hat{\mathbf{R}}_t & \mathbf{0}_{3\times3} \\ [\mathbf{g}]_\times & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -[\hat{\mathbf{v}}_t]_\times \hat{\mathbf{R}}_t & -\hat{\mathbf{R}}_t \\ \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & -[\hat{\mathbf{p}}_t]_\times \hat{\mathbf{R}}_t & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \end{pmatrix} \tag{13}$$

where $\hat{\mathbf{G}}_t$ is the noise matrix, which is expressed as:

$$\hat{\mathbf{G}}_t = \begin{pmatrix} \hat{\mathbf{R}}_t & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ [\hat{\mathbf{v}}_t]_\times \hat{\mathbf{R}}_t & \hat{\mathbf{R}}_t & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ [\hat{\mathbf{p}}_t]_\times \hat{\mathbf{R}}_t & \mathbf{0}_{3\times3} & \hat{\mathbf{R}}_t & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -\mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -\mathbf{I}_{3\times3} \end{pmatrix} \tag{14}$$

The, the propagation of the covariance matrix of the state can be expressed as:

$$\hat{\boldsymbol{\Sigma}}_t = \hat{\boldsymbol{\Phi}}_t \hat{\boldsymbol{\Sigma}}_{t-1} \hat{\boldsymbol{\Phi}}_t^{\mathrm{T}} + \hat{\boldsymbol{\Phi}}_t \hat{\mathbf{G}}_t \mathbf{Q} \hat{\mathbf{G}}_t^{\mathrm{T}} \hat{\boldsymbol{\Phi}}_t^{\mathrm{T}} \Delta t \tag{15}$$

where $\mathbf{Q}$ is the noise covariance matrix formed by the noise term in equation (12). $\hat{\boldsymbol{\Phi}}_t$ can be given as:

$$\hat{\boldsymbol{\Phi}}_t = \exp\left(\hat{\mathbf{F}}_\mathbf{t} \Delta \mathrm{t}\right) \tag{16}$$

3.2 Measurement Update

The regressed speed, expressed in the pedestrian's body frame, needs be transformed to the smartphone frame s. Additionally, the nonholonomic constraint (NHC) is applied. Consequently, the measurement can be expressed as follows:

$$\tilde{\mathbf{z}}_t = \hat{\mathbf{R}}_b^s \begin{bmatrix} \hat{v}^b & 0 & 0 \end{bmatrix}^{\mathrm{T}} \tag{17}$$

where $\hat{\mathbf{R}}_b^s$ is the rotation matrix from the body to the smartphone frame, which is the inverse matrix of $\hat{\mathbf{R}}_s^b$.

The measurement model can be expressed as:

$$\tilde{\mathbf{z}}_t = \mathbf{R}_t^{\mathrm{T}} \mathbf{v}_t + \mathbf{s}_t \tag{18}$$

where $\mathbf{s}_t$ represents the measurement noise. The equation (18) is further expressed as:

$$\begin{aligned} \tilde{\mathbf{Z}}_t &= \boldsymbol{\chi}_t^{-1} \mathbf{M} + \mathbf{S}_t \\ \begin{pmatrix} \tilde{\mathbf{z}}_t \\ -1 \\ 0 \end{pmatrix} &= \begin{pmatrix} \mathbf{R}_t^{\mathrm{T}} & -\mathbf{R}_t^{\mathrm{T}} \mathbf{v}_t & -\mathbf{R}_t^{\mathrm{T}} \mathbf{p}_t \\ \mathbf{0}_{1\times3} & 1 & 0 \\ \mathbf{0}_{1\times3} & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{0}_{3\times1} \\ -1 \\ 0 \end{pmatrix} + \begin{pmatrix} \mathbf{s}_t \\ 0 \\ 0 \end{pmatrix} \end{aligned} \tag{19}$$

The measurement matrix can be expressed as:

$$\mathbf{H}_t = \begin{pmatrix} \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \end{pmatrix} \tag{20}$$

Then the state update can be represented as:

$$\begin{aligned} \hat{\boldsymbol{\chi}}_t &= \exp\left(\mathbf{K}^{\xi} \boldsymbol{\Pi} \hat{\boldsymbol{\chi}}_\mathrm{t} \tilde{\mathbf{Z}}_\mathrm{t}\right) \hat{\boldsymbol{\chi}}_\mathrm{t} \\ \hat{\boldsymbol{\Theta}}_t &= \hat{\boldsymbol{\Theta}}_t + \mathbf{K}^{\zeta} \boldsymbol{\Pi} \hat{\boldsymbol{\chi}}_t \tilde{\mathbf{Z}}_t \end{aligned} \tag{21}$$

where $\mathbf{K}^{\xi}$ and $\mathbf{K}^{\zeta}$ denote the filtering gains. $\hat{\boldsymbol{\chi}}_t \tilde{\mathbf{Z}}_t$ represents the innovation. $\boldsymbol{\Pi}$ is an auxiliary matrix to remove the zero terms of innovation.

4. Experimental Tests

Experimental tests were conducted in different scenarios to validate the performance of the proposed method. The test sites selected were an underground car park and a campus ground. During these tests, IMU data was collected using an iPhone 12 at a rate of 100 Hz. A LiDAR-inertial system, Mid-360, produced by Livox Technology Company, was utilized to run FAST-LIO (Xu and Zhang, 2021), providing the ground truth. Traditional methods, including PDR, RONIN, and EKF, were used for comparison. In the EKF method, the system also utilizes the learned forward speed to update the state.

4.1 Underground Car Park

Firstly, the test results in the underground car park are presented. Figure 3 and Figure 4 show the comparisons of trajectory and positioning errors among different methods. It is evident that both PDR and RONIN suffer from noticeable drift. In the test scenario, only a 6-axis IMU is used, which means that the heading can be determined solely via the integration of gyroscopes. However, compensating for the unknown bias in the gyroscope is challenging in PDR, leading to rapid drift in heading estimation and affecting the overall PDR results. Similarly, in RONIN, the IMU data must first be transformed into the world frame before being fed into the neural network for displacement regression. The inaccuracies in heading estimation can compromise the quality of RONIN's positioning results.

In both the EKF and the proposed method, the bias of the IMU can be estimated online with the aid of velocity measurements. This leads to better positioning results than those achieved by PDR and RONIN. Although the EKF can achieve performance similar to the proposed method in the first half of the data, its performance gradually degrades over time. The primary source of error stems from the linearization procedure, where the predicted states are typically used in the Taylor expansion. This introduces additional errors for state estimation if the predicted states have significant errors. However, as can be observed from equation (13), the coefficients related to pose and velocity are independent of the current state estimates in the invariant extended Kalman filter. This independence can mitigate the influence of nonlinear errors and enhance the accuracy of odometry.

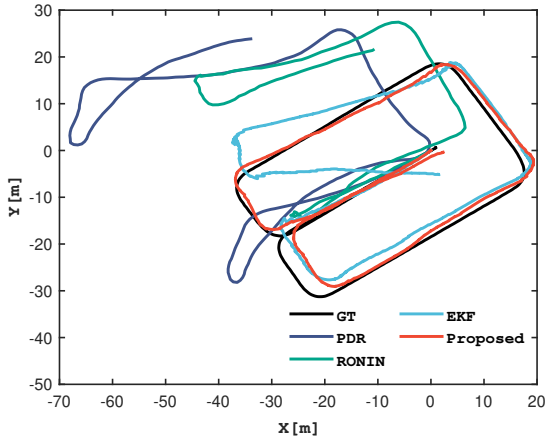

Figure 3. The trajectory comparison in the underground car park.

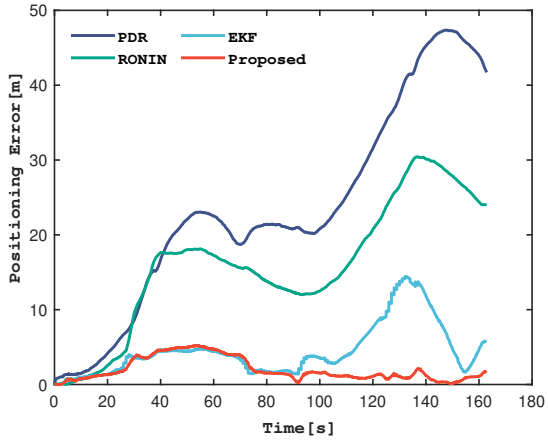

Figure 4. The positioning error comparison in the underground car park.

The comparison of root mean squared error (RMSE) among different methods in the underground car park is presented in Table 1. It is evident that the proposed method outperforms the others, achieving the best positioning performance.

Table 1. The RMSE comparison in the underground car park.

Method	PDR	RONIN	EKF	Proposed
RMSE (m)	27.14	18.07	5.68	2.53

4.2 Campus Ground

Another test was conducted on the campus grounds to further validate the performance of the proposed method. The comparison of trajectories is depicted in Figure 5. Both PDR and RONIN exhibit noticeable drift due to the impact of heading inaccuracies. By employing a filter-based integration, both EKF and the proposed method can outperform PDR and RONIN in terms of positioning performance. However, due to the presence of nonlinear errors, the position estimation provided by EKF gradually deteriorates.

The comparison of positioning error and RMSE are depicted in Figure 6 and Table 2, respectively. It can be observed that PDR and RONIN exhibit significant drift right from the start. While EKF can somewhat mitigate the accumulation of errors, its positioning outcomes still fall short when compared to the proposed method.

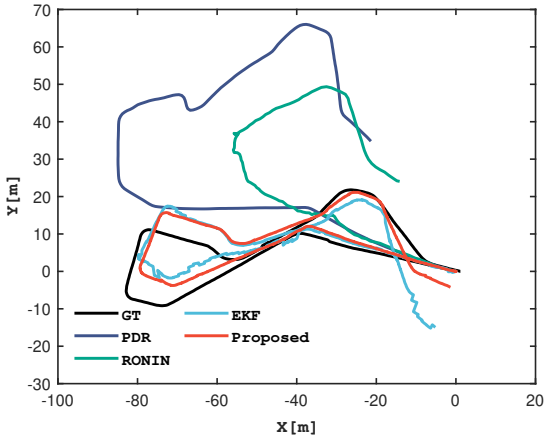

Figure 5. The trajectory comparison in the campus ground.

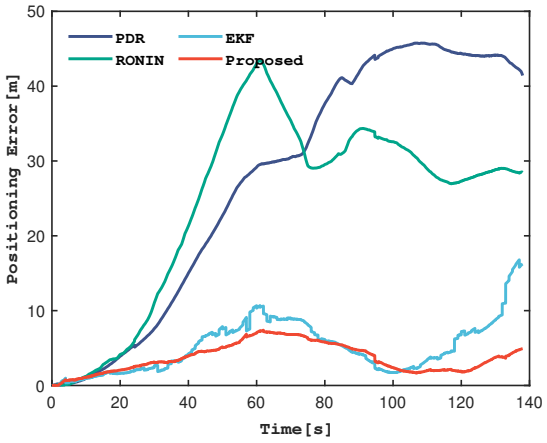

Figure 6. The positioning error comparison in the campus ground.

Table 2. The RMSE comparison in the campus ground.

Method	PDR	RONIN	EKF	Proposed
RMSE (m)	32.18	27.52	6.24	4.16

5. Conclusion

This paper introduces an invariant extended Kalman filter for pedestrian deep-inertial odometry. In the proposed method, IMU data is utilized to estimate the pedestrian’s forward speed using a neural network. This estimation, when combined with NHC, allows for the calculation of the pedestrian’s position via an invariant extended Kalman filter based on velocity measurement updates. This approach effectively mitigates the impact of heading inaccuracies and nonlinear errors found in traditional methods. The proposed method is validated using real-world data collected by an iPhone 12 in an underground car park and on a campus ground. For comparison, PDR, RONIN, and an EKF-based method are employed. The results demonstrate that the proposed method can effectively limit error accumulation in the odometry, thereby enhancing indoor localization performance using only an IMU.

The proposed DIO is fundamentally based on dead reckoning, it continues to experience drift over extended periods. In the future, we plan to use the proposed DIO as the foundation and incorporate external information such as UWB, Wi-Fi RTT, or 5G for absolute position correction. Simultaneously, we aim to integrate the proposed method into a smartphone to establish a real-time indoor localization system.

Acknowledgements

This work received support from several sources. It was funded by the Guangdong Basic and Applied Basic Research Foundation (2021A1515110771) and the University Grants Committee of Hong Kong under the Research Impact Fund (R5009-21). Partial support was also provided by the Research Centre for Data Sciences & Artificial Intelligence at The Hong Kong Polytechnic University, through the project titled “Data-driven-assisted GNSS RTK/INS Navigation for Autonomous Systems in Urban Canyons.” Additionally, the Research Center of Deep Space Exploration at The Hong Kong Polytechnic University contributed to this work, through the project “Multi-robot Collaborative Operations in Lunar Areas for Regolith Processing.”

References

Bai, S., Wen, W., Hsu, L.-T., Yu, Y., 2023. Factor graph optimization-based indoor pedestrian SLAM with probabilistic exact activity loop closures using smartphone. *2023 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 1–8.

Barrau, A., Bonnabel, S., 2016. The invariant extended Kalman filter as a stable observer. *IEEE Transactions on Automatic Control*, 62(4), 1797–1812.

Chen, C., Lu, X., Markham, A., Trigoni, N., 2018. Ionet: Learning to cure the curse of drift in inertial odometry. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32number 1.

Chen, C., Pan, X., 2024. Deep learning for inertial positioning: A survey. *IEEE Transactions on Intelligent Transportation Systems*.

Cortés, S., Solin, A., Kannala, J., 2018. Deep learning based speed estimation for constraining strapdown inertial navigation on smartphones. *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, IEEE, 1–6.

Herath, S., Yan, H., Furukawa, Y., 2020. Ronin: Robust neural inertial navigation in the wild: Benchmark, evaluations, & new methods. *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 3146–3152.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Liu, H., Darabi, H., Banerjee, P., Liu, J., 2007. Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(6), 1067–1080.

Liu, W., Caruso, D., Ilg, E., Dong, J., Mourikis, A. I., Daniilidis, K., Kumar, V., Engel, J., 2020. Tlio: Tight learned inertial odometry. *IEEE Robotics and Automation Letters*, 5(4), 5653–5660.

Potokar, E. R., Norman, K., Mangelson, J. G., 2021. Invariant extended Kalman filtering for underwater navigation. *IEEE Robotics and Automation Letters*, 6(3), 5792–5799.

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

Ren, X., Zhu, S., Liu, F., Li, H., Li, H., Xiao, X., Gao, R., Zhang, Z., 2024. Help you locate the car: A smartphone-based car-finding system in underground parking lot. *IEEE Sensors Journal*, 24(5), 7107-7118.

Tseng, P.-Y., Lin, J. J., Chan, Y.-C., Chen, A. Y., 2022. Real-time indoor localization with visual SLAM for in-building emergency response. *Automation in Construction*, 140, 104319.

Wang, Q., Fu, M., Wang, J., Luo, H., Sun, L., Ma, Z., Li, W., Zhang, C., Huang, R., Li, X. et al., 2022. Recent advances in pedestrian inertial navigation based on smartphone: A review. *IEEE Sensors Journal*, 22(23), 22319–22343.

Wang, Y., Cheng, H., Meng, M. Q.-H., 2020. Pedestrian motion tracking by using inertial sensors on the smartphone. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 4426–4431.

Wang, Y., Cheng, H., Wang, C., Meng, M. Q.-H., 2021. Pose-invariant inertial odometry for pedestrian localization. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–12.

Wang, Z., Zhao, H., Qiu, S., Gao, Q., 2015. Stance-phase detection for ZUPT-aided foot-mounted pedestrian navigation system. *IEEE/ASME Transactions on Mechatronics*, 20(6), 3170–3181.

Wen, W., Bai, X., Kan, Y. C., Hsu, L.-T., 2019. Tightly coupled GNSS/INS integration via factor graph and aided by fish-eye camera. *IEEE Transactions on Vehicular Technology*, 68(11), 10651–10662.

Xu, W., Zhang, F., 2021. Fast-lio: A fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter. *IEEE Robotics and Automation Letters*, 6(2), 3317–3324.

Yang, X., Wang, J., Song, D., Feng, B., Ye, H., 2021. A novel NLOS error compensation method based IMU for UWB indoor positioning system. *IEEE Sensors Journal*, 21(9), 11203–11212.

Yi, X., Zhou, Y., Xu, F., 2021. Transpose: Real-time 3D human translation and pose estimation with six inertial sensors. *ACM Transactions on Graphics (TOG)*, 40(4), 1–13.

Zhang, J., Singh, S., 2014. LOAM: LiDAR odometry and mapping in real-time. *Robotics: Science and Systems*, 2number 9, Berkeley, CA, 1–9.