

Factor Graph Optimization-Based Smartphone IMU-Only Indoor SLAM With Multihypothesis Turning Behavior Loop Closures

SHIYU BAI ¹, Member, IEEE

WEISONG WEN ¹, Member, IEEE

LI-TA HSU ¹, Senior Member, IEEE

PEIWEN YANG ¹

Hong Kong Polytechnic University, Hong Kong, SAR, China

Pedestrian dead reckoning (PDR) using smartphones is a popular method for indoor localization. However, it encounters challenges due to the drift of position errors. While external resources like Wi-Fi, Bluetooth, and indoor maps can correct the drift, they require the preinstallation of facilities or information, limiting their application. Inspired by the recent advancement of the simultaneous localization and mapping (SLAM), such as the visual SLAM, this article proposes a factor graph optimization (FGO)-based smartphone inertial measurement unit (IMU)-only indoor SLAM with multihypothesis turning behavior loop closures. In this article, the turning behavior that can be observed repetitively is regarded as landmarks, which are like the visual landmarks in visual SLAM. FGO is employed, wherein the

motion deduced from PDR is used to establish relative constraints between consecutive variable nodes. Meanwhile, the robust loop closure constraint inferred from the multihypothesis behavior data association is formed to correct the drift of the PDR. Experimental tests in different scenarios are done to evaluate the localization performance, and the time consumption of the proposed method is given.

NOMENCLATURE

$\tilde{\mathbf{a}}_k^b$	Measured acceleration in the body frame at the time instant t_k .
$\tilde{\mathbf{w}}_k^b$	Measured angular velocity in the body frame at the time instant t_k .
$\tilde{\mathbf{m}}_k^b$	Measured magnetic-field vector in the body frame at the time instant t_k .
$\hat{\mathbf{a}}_k^b$	Smoothed acceleration in the body frame at the time instant t_k .
$\tilde{\mathbf{w}}_{k-n:n}^b$	Measured angular velocity sequence in the body frame from the time instant t_{k-n} to t_k .
$\hat{\mathbf{w}}_{k-n:n}^b$	Smoothed angular velocity sequence in the body frame from the time instant t_{k-n} to t_k .
$\hat{\mathbf{w}}_{k-n:n}^w$	Smoothed angular velocity sequence in the world frame from the time instant t_{k-n} to t_k .
$\hat{a}_{v,\max}^w$	Smoothed vertical acceleration peak in the world frame.
$\hat{a}_{v,\min}^w$	Smoothed vertical acceleration valley in the world frame.
$\hat{\mathbf{R}}_k$	Estimated rotation matrix at the time instant t_k .
\hat{L}_k	Estimated step length at the time instant t_k .
$\hat{\theta}_k$	Estimated heading at the time instant t_k .
$\hat{\mathbf{R}}_{k-n:k}$	Estimated rotation matrix sequence from the time instant t_{k-n} to t_k .
\mathbf{r}^p	PDR residual.
\mathbf{r}^c	Turning behavior loop closure residual.

I. INTRODUCTION

Accurate indoor localization solution for pedestrians [1], [2] is critical in many applications, such as periodic inspection, emergency response, and incident management [3], [4], [5]. With the rapid development of microelectromechanical systems (MEMS)-based inertial measurement units (IMUs), the IMUs with higher performance and smaller size can be installed in the smartphone or put on the body to provide a lightweight localization solution when the global navigation satellite system (GNSS) is unavailable. In daily life, people almost carry their smartphones at all times. In addition, IMUs are compactly integrated into smartphones, which facilitates the use of smartphone IMUs for localization purposes [6]. Pedestrian dead reckoning (PDR) based on IMU measurements is a popular method to achieve indoor localization. Although PDR is a self-contained localization tool, its position error accumulates with time and distance, which cannot be effectively applied to long-distance situations. Some external [7], [8], [9] or prior information can be used to restrain PDR's error accumulation, but they put higher demands on additional equipment and preliminary calibration, leading to increased cost and workload.

Manuscript received 19 February 2024; revised 15 May 2024; accepted 13 July 2024. Date of publication 18 July 2024; date of current version 6 December 2024.

DOI: No. 10.1109/TAES.2024.3430256

Refereeing of this contribution was handled by J. Seo.

This work was supported in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515110771 and in part by the University Grants Committee of Hong Kong through the scheme Research Impact Fund under Grant R5009-21. An earlier version of this paper was presented in part at the 2023 International Conference on Indoor Positioning and Indoor Navigation (IPIN) [DOI: 10.1109/IPIN57070.2023.10332510].

Authors' address: Shiyu Bai, Weisong Wen, Li-Ta Hsu, and Peiwen Yang are with Hong Kong Polytechnic University, Hong Kong, SAR, China, E-mail: (shiyu.bai@polyu.edu.hk; welson.wen@polyu.edu.hk; lt.hsu@polyu.edu.hk; peiwen1.yang@connect.polyu.hk). (Corresponding author: Li-Ta Hsu.)

The video (YouTube) or (bilibili) is also shared to display our research.

0018-9251 © 2024 IEEE

To solve these problems, a factor graph optimization (FGO)-based smartphone IMU-only indoor simultaneous localization and mapping (SLAM) with multihypothesis turning behavior loop closures is proposed. This work is the extension of our conference paper [10]. In the conference paper, we only presented a preliminary methodology and initial results. However, in this work, we have introduced substantial new content. The main contributions of this article are as follows.

- 1) *Factor Graph Optimization-Based IMU-Only Environment-Induced Behavior SLAM (Hereafter Referred to as Behavior SLAM)*: This article introduces a behavior SLAM framework based on graph optimization. The turning behavior is established as the landmark. Rather than using a particle filter (PF), FGO is utilized in the creation of behavior SLAM, leading to better estimation for localization and trajectory than filter-based methods compared in this article. This article introduces a more sophisticated methodology compared to the one in our conference version. It provides a detailed technical scheme for PDR. Furthermore, it presents the turning behavior sequence and the formation of the confidence ellipse for improved data association.
- 2) *Multihypothesis Turning Behavior Loop Closures*: This article builds a multihypothesis turning behavior loop closure to restrain the cumulative error of SLAM. Multiple factor graphs are maintained to ensure the availability and reliability of data association. The turning behavior sequence is introduced to identify different types of turns, thereby excluding improper loop closures. In addition, the turning behavior sequence is used to determine data associations among key points, avoiding the loss of loop closures.
- 3) Experiments in different indoor scenes are conducted for the evaluation. Moreover, the performance of the proposed method is assessed and analyzed from various perspectives, including accuracy across different paths, reliability, uncertainty settings, and time consumption. These aspects have not been addressed in existing papers dealing with behavior SLAM.

The rest of this article is organized as follows. Section II delves into relevant literature, while Section III provides an overview of the methodology of the proposed method. Sections IV and V introduce the construction of the factor graph and the multihypothesis turning behavior data association. Section VI gives experiment results and discussions. Finally, Section VII concludes this article.

II. RELATED WORK

PDR has been widely used in indoor localization, which tracks pedestrians using inertial measurements [11], [12], [13]. The main problem in PDR is error cumulation [14]. Thereby, some supporting methods have been

thoroughly investigated to improve the indoor localization performance.

One feasible tool is to utilize the external infrastructure to provide the position measurement. GNSS is the most frequently used method to achieve localization. Although GNSS signal is unavailable in indoor environments, it can be amplified and retransmitted to indoor receivers using repeaters. Li [15] proposed an indoor localization way assisted by GNSS repeaters, and the Kalman filter was used to fuse position results of different methods with transferred GNSS signals, leading to improved localization accuracy. However, interference between signals is unavoidable, which easily causes accuracy reduction. Radio-frequency (RF) signal-based techniques, such as Wi-Fi [16], Bluetooth [17], and ultra-wideband (UWB) [18], can provide relatively stable localization results in indoor scenarios. Moreover, the integration of PDR and RF has also been discussed. Poulose et al. [19] presented a sensor fusion framework based on PDR and Wi-Fi received signal strength indication (RSSI) measurements. Wi-Fi fingerprinting and trilateration were combined to calibrate PDR errors and achieved better indoor localization accuracy than traditional Wi-Fi localization. Li et al. [20] proposed a Wi-Fi/PDR integrated navigation system based on a constrained Kalman filter, which can mitigate gross errors to improve the reliability of indoor localization. Wang et al. [13] incorporated PDR and Bluetooth in tightly coupled integration using graph optimization, leading to better estimation accuracy than the extended Kalman filter (EKF) and unscented Kalman filter (UKF). Tian et al. [21] proposed a tightly coupled PDR and UWB integration based on PF, and an adaptive ranging uncertainty model was built to mitigate the ranging errors. However, one key challenge of RF signal-based indoor localization is the dependence on external facilities. Moreover, the localization accuracy declines severely when non-line-of-sight (NLOS) errors occur in RF ranging measurements [22].

Enhanced dead reckoning-based indoor localization methods have been developed to lower the strong dependence on external infrastructure. Poulose et al. [23] proposed a position-estimation algorithm that combined the accelerometer, magnetometer, and gyroscope to estimate the attitude, and a pitch-based step detection algorithm was presented, which can outperform traditional PDR methods. Moreover, the data-driven inertial navigation method has also shown its advantage in position error suppression. Yan et al. [24] introduced a robust IMU double integration (RIDI) for inertial navigation, in which a velocity vector can be regressed from a sequence of IMU data to correct the accelerometer bias. A more accurate trajectory can be calculated by integrating corrected linear accelerations. Chen et al. [25] formulated the physical model based on Newtonian mechanics and utilized a recurrent neural network (RNN) to mitigate the drift of inertial navigation. Herath and Yan [26] introduced a robust neural inertial navigation (RoNIN) that can be achieved by ResNet [27], long short-term memory network (LSTM) [28], and temporal convolutional network (TCN) [29]. Liu et al. [30]

presented a tight learned inertial odometry (TLIO). In this method, 3-D displacements and their uncertainties were initially regressed, and a stochastic cloning EKF was utilized for state estimation. However, these methods continue to experience drift.

Employing the map is one effective method to constrain dead reckoning-based pedestrian localization. Human behaviors can be associated with particular areas in indoor environments [31]. According to the behaviors, the pedestrian location can be speculated, assisted by a prior map. Munoz et al. [32] investigated a landmark-based drift compensation in which the corner and stair were used as landmarks. Data association was based on the Mahalanobis distance and a predefined threshold, which can easily cause the wrong matching as only a single behavior is used for determination. Utilizing a sequence of behaviors can improve reliability. An indoor localization method based on activity sequences was proposed in [33]. IMU, barometer, and digital compass first detected the behaviors. Then, detected behaviors were matched to the indoor road network based on hidden Markov model [34] to obtain the position estimation. Gu et al. [35] introduced a context-based map matching approach for indoor scenarios, where the combination of IMU and barometer was used to identify behaviors, such as turning, shopping, and more. Kang and Shin [36] introduced a smartphone-based indoor navigation method using IMU and map. The turn points were detected and matched with the map to rectify position errors. However, the prior map is necessary in most of these methods. In addition, indoor structures may change due to artificial adaptations and sudden incidents, leading to the failure of prior information. Meanwhile, frequently updating the indoor map is still an open challenge.

SLAM [37], [38] can be used to improve indoor localization performance in unknown scenes. Mainstream SLAM methods using smartphones are mainly based on camera. Cortés et al. [39] developed an authentic dataset for visual-inertial odometry, the real-world benchmark set for visual-inertial SLAM using smartphones, providing a wide range of raw data from Google Pixel Android phones and iPhones with a high-quality ground truth track. Poulou and Han [40] proposed a hybrid indoor SLAM based on smartphone IMU and camera. The errors of IMU-based and camera-based localization results can be mutually compensated, and better positioning results than the individual localization method can be achieved. Although visual SLAM methods can achieve satisfactory localization solutions in indoor environments, they call for some requirements in the phone and consume vast amounts of computing resources, which is unsuitable for people in daily life. Behavior SLAM based on IMU is another type of method to suppress accumulative errors. Human behaviors can be built as landmarks in a continuously updated map. When the pedestrian passes the same landmark, the current position can be updated to improve the localization accuracy. Hardegger et al. [41] developed the ActionSLAM, in which location-related behaviors were regarded as landmarks, and these landmarks were used to compensate for the position error drift based

on PF. Hardegger et al. [42] extended his work to unknown multifloor environments, in which PDR from foot-mounted IMU and the recognition of behaviors were combined to update the pedestrian position. Angermann and Robertson [43] investigated FootSLAM to achieve a long-term, stable, infrastructure-less pedestrian positioning method. The human motion was represented in a 2-D hexagonal grid, and the Rao–Blackwellized particle filtering method was utilized for the state estimation. Abdennasser et al. [44] formulated an unsupervised indoor localization scheme, SemanticSLAM, using the environment signatures of inertial sensors as landmarks to calibrate PDR errors. Shokry et al. [45] introduced DynamicSLAM, in which building-induced behaviors and human-based anchors were utilized as landmarks to achieve indoor localization using a PF. Osman et al. [46] proposed an indoor SLAM based on motion and magnetic maps, which was also achieved using a Rao–Blackwellized PF to correct the pedestrian pose.

PF is the most common method to achieve state estimation. However, filter-based methods suffer from the marginalization of past states and measurements. The proper relinearization cannot be done since past states are not involved in the filter [47], [48]. Recently, graph-based sensor fusion has been widely used for state estimation. Indelman et al. [49] introduced a factor graph-based information fusion in inertial navigation systems. This method can offer an efficient and universal framework for state estimation. It can achieve greater precision than filter-based methods by mitigating nonlinear errors. Graph-based fusion has been applied to various communities [50], [51], [52]. Furthermore, it is gradually applied in indoor environments. Li et al. [53] provided an incremental smoothing method based on the Tukey kernel function to fuse UWB and PDR, whose performance was evaluated compared to EKF, and the proposed method has stronger robustness. Wang et al. [54] introduced a factor graph model to integrate Wi-Fi and PDR. An adaptive robust adjustment was added to suppress the influence of the gross errors. Yang et al. [55] proposed a unified framework based on the factor graph for ubiquitous indoor localization, which can integrate fingerprinting-based and ranging-based information to achieve submeter-level localization accuracy. Wang et al. [5] studied a graph optimization-based indoor localization method to fuse the learning-based inertial odometry and magnetic measurement, which was robust to magnetic field perturbation. Zhao et al. [56] proposed an uncertainty-aware Gaussian mixture model (GMM) for UWB time-difference-of-arrival (TDOA) localization in indoor environments. In this method, a bilevel optimization was designed for joint localization and GMM noise model learning, which can lead to enhanced localization performance.

Graph-based methods have also shown their potential advantage in SLAM fields, and building reliable data associations has always been a core problem to be solved. Thrun and Montemerlo [57] proposed GraphSLAM, in which a sparse graph represented the state estimation for the robot

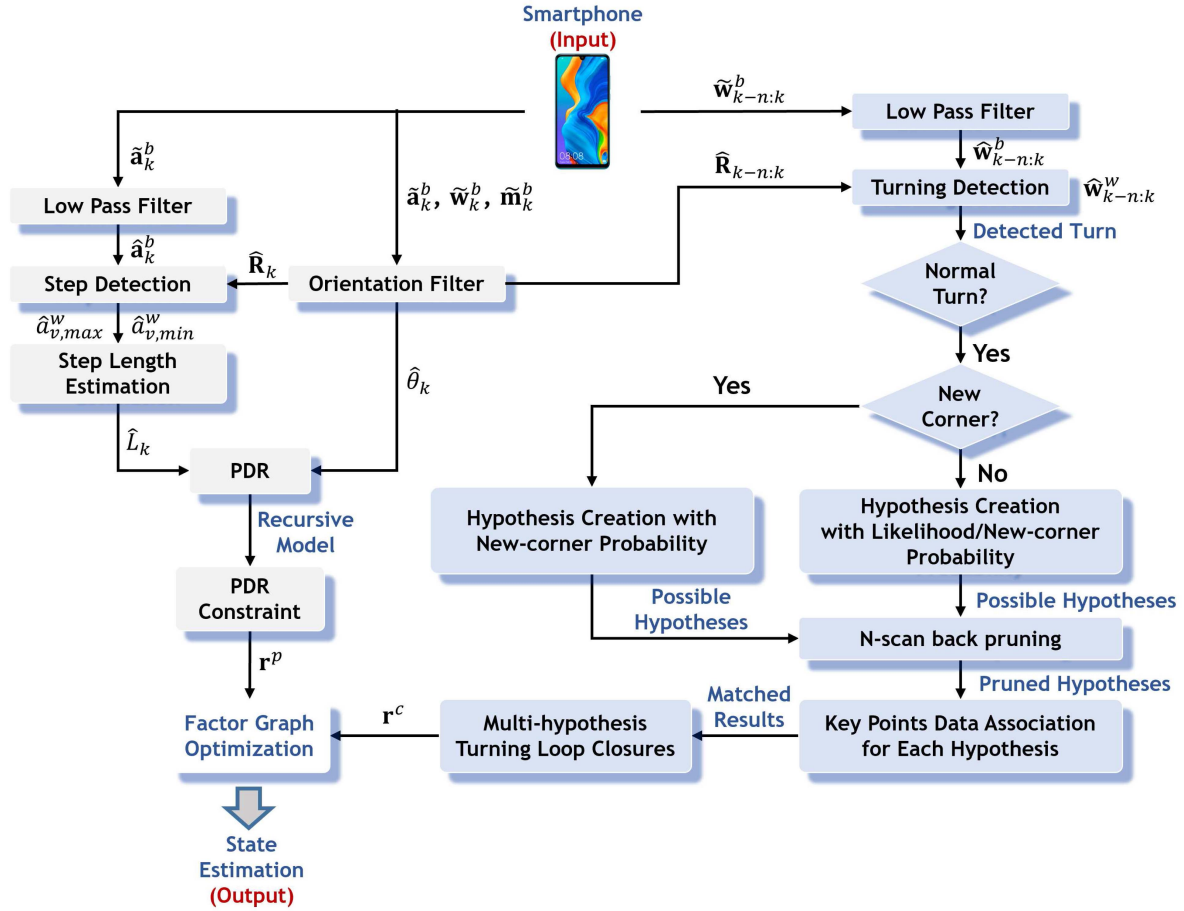


Fig. 1. Framework of the proposed method, in which the input is the nine-axis IMU measurements (acceleration, angular velocity, and magnetic-field vectors) from a smartphone. Multiple hypotheses are generated, with each one assigned a probability to assess its likelihood of occurrence. The output is the state estimation corresponding to the hypothesis with the highest probability.

path and map. Unknown data associations were solved based on the correspondence test. The correspondence probability was calculated, and a threshold was set to determine the data association. Grisetti et al. [58] provided a tutorial on the graph-based SLAM and reviewed data association methods, including spectral clustering, joint compatibility branch and bound, and backtracking. Wang et al. [59] investigated a factor graph-aided multihypothesis tracking, which can reduce the time of track swap and increase the correct data association rate. Hsiao and Kaess [60] developed the multihypothesis incremental smoothing and mapping using Bayes tree and Hypo-tree (MH-iSAM2) to handle ambiguous data association in SLAM. A multihypothesis factor graph can be formed to enhance the robustness of the state estimation. Although many graph-based SLAM methods use multihypothesis tracking, existing methods are based on vision or LiDAR. There is still no corresponding algorithm for smartphone IMU-based SLAM, whose models and data association schemes differ from current ones.

To the best of our knowledge, we are the first to develop an FGO-based IMU-only SLAM for smartphones, and the multihypothesis turning behavior data association is built

to form the loop closure, which achieves better reliability than single-hypothesis FGO methods. Unlike graph-based SLAM methods using vision and LiDAR, models and data association schemes for FGO are first studied when only smartphone IMU is available. Furthermore, this article comprehensively discusses the performance of the proposed method from different perspectives, including the accuracy in different paths, reliability, uncertainty setting, and time consumption, which are essential to behavior SLAM but cannot be found in existing papers dealing with behavior SLAM.

III. METHODOLOGY OVERVIEW

Fig. 1 depicts the framework of the proposed method. The input of the proposed method is acceleration, angular velocity, and magnetic-field vector from the nine-axis IMU embedded in the smartphone. The FGO is based on the Georgia Tech Smoothing and Mapping library (GTSAM) [61], a library of C++ classes that implements sensor fusion based on factor graphs. GTSAM offers an incremental inference algorithm based on the Bayes tree, which can ensure the estimation accuracy and computational efficiency of solving the optimization problem. The proposed

method consists of two parts. The first part is to utilize PDR to obtain continuous position estimations. Meanwhile, the PDR factor, the residual based on pose and heading constraints between two consecutive variable nodes derived from the PDR model, is formed.

Another part is the formulation of multihypothesis turning loop closures. A sequence of angular velocity observations is used to detect turning behaviors. The type of turning will be judged. If it is a normal turn, defined as the turning at a corner rather than a U-turn, it has to be determined whether this behavior occurs at a new corner. Therefore, multiple hypotheses are formed, each associated with a probability, to evaluate their likelihood of occurrence. In the proposed method, the concept of multiple-hypothesis turnings suggests that a turning could potentially occur at any previous landmark or even at a new corner. This results in the generation of multiple hypotheses. Considering that the rise of hypotheses can cause a vast computational load in factor graphs, back pruning [62] needs to be done to reduce the number of candidate hypotheses. According to pruned hypotheses, key points within the turning behavior sequence are interrelated to construct loop closures. Multiple factor graphs are simultaneously built to achieve optimization. The most likely hypothesis, which has the maximum probability, is used as the instant output.

The typical symbols that are used throughout the article are also presented in this part, which is shown in the Nomenclature. The superscripts b and w denote the body and world frame, respectively. The body frame is aligned with three axes of IMU. The world frame has its origin in the body frame at the initial instant. Its three axes point to east, north, and up, respectively.

IV. FACTOR GRAPH OPTIMIZATION-BASED SLAM AIDED BY TURNING BEHAVIORS

The factor graph is a bipartite graph $\mathbf{F} = (\mathbf{G}, \mathbf{X}, \mathbf{E})$ that is composed of two types of nodes: 1) factor node $\mathbf{g}_i \in \mathbf{G}$ and 2) variable node $\mathbf{x}_j \in \mathbf{X}$. \mathbf{G} is the set of factor nodes and \mathbf{X} is the set of variable nodes (state vector). An edge denoted as $\mathbf{e}_{ij} \in \mathbf{E}$ is present between the factor node and the variable node when the variable is involved in the factor [63]. \mathbf{E} is the set of edges. Given the assumption that the noise follows a Gaussian distribution, we can denote a generic factor node as

$$\mathbf{g}_i(\mathbf{X}) = \|\mathbf{z}_i - \mathbf{h}_i(\mathbf{X})\|_{\Sigma}^2 \quad (1)$$

where \mathbf{X} represents the state vector that needs to be estimated, which can be expressed as

$$\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n] \quad (2)$$

where

$$\mathbf{x}_n = [p_{x,n}, p_{y,n}, \theta_n] \quad (3)$$

where \mathbf{x}_n denotes the state (variable node) at the time instant t_n , which includes the horizontal position and heading. \mathbf{z}_i and $\mathbf{h}_i(\mathbf{X})$ are the i th actual measurement and predicted measurement, respectively. Σ is the covariance matrix of (1). $\|\mathbf{d}\|_{\Sigma}^2$ denotes the Mahalanobis distance that can be

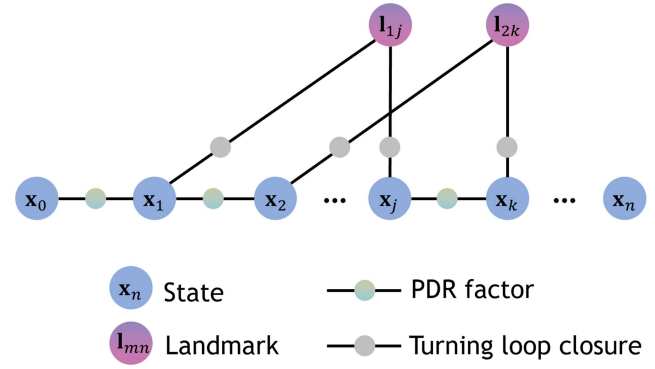


Fig. 2. Factor graph of the proposed method, in which the PDR factor connects two consecutive states, and turning loop closure connects two states when the pedestrian revisits the same corner.

expressed as $\mathbf{d}^T \Sigma^{-1} \mathbf{d}$, where \mathbf{d} is a generic symbol to express the residual.

In nonlinear optimization, the solution that satisfies all constraints is found by minimizing the global error, expressed as

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left(\sum_i \mathbf{g}_i(\mathbf{X}) \right) \quad (4)$$

where $\hat{\mathbf{X}}$ represents the optimal state vector. (4) can be solved by repeated linearization via the Gauss–Newton-style nonlinear optimizer.

Fig. 2 illustrates the factor graph in our method, where the state is represented by a blue circle. The purple circles labeled with \mathbf{l}_{1j} and \mathbf{l}_{2k} denote the landmarks. Their subscripts represent that the states at time instants t_1 and t_j , as well as the states at time instants t_2 and t_k , are associated by different landmarks. The PDR factor and the turning behavior loop closure are represented by the green circle and the grey circle, respectively.

Given that the landmarks are indeed part of the trajectory, the position of the landmarks is not included in the state.

The cost function in this article is formulated as

$$\mathbf{X}^* = \arg \min \left(\sum_{k=1}^n \|\mathbf{r}^p(\hat{\mathbf{z}}_k^p, \mathbf{X})\|_{\Sigma^p}^2 + \sum_{k=1}^m \|\mathbf{r}^c(\hat{\mathbf{z}}_k^c, \mathbf{X})\|_{\Sigma^c}^2 \right) \quad (5)$$

where $\mathbf{r}^p(\hat{\mathbf{z}}_k^p, \mathbf{X})$ and $\mathbf{r}^c(\hat{\mathbf{z}}_k^c, \mathbf{X})$ represent the PDR residual and turning loop closure residual, respectively. Σ^p and Σ^c denote the respective covariance matrix. The covariance matrix of PDR residual can be obtained by carrying out a statistical analysis for estimated PDR errors with the aid of LiDAR odometry, whose drift is smaller than $0.05\% \times \text{Distance}$ [64]. The covariance matrix of turning loop closure residual is set smaller than the PDR to strengthen the correction. n and m signify the number of PDR residual and turning loop closure residual. In this article, the optimization is carried out using iSAM2 [65]. The residual models in (5) are given in the following (6) and (7).

The PDR residual is used to constrain consecutive states and can be formulated as

$$\mathbf{r}^p(\hat{\mathbf{z}}_k^p, \mathbf{X}) = \begin{bmatrix} \hat{L}_k \cos \hat{\theta}_k - p_{x,k} + p_{x,k-1} \\ \hat{L}_k \sin \hat{\theta}_k - p_{y,k} + p_{y,k-1} \\ \Delta \hat{\theta}_{k:k-1} - \theta_k + \theta_{k-1} \end{bmatrix} \quad (6)$$

where \hat{L}_k and $\hat{\theta}_k$ are estimated step length and heading at the time instant t_k . $\Delta \hat{\theta}_{k:k-1}$ expresses the heading change measurement between two consecutive states.

The turning loop closure residual is expressed as

$$\mathbf{r}^c(\hat{\mathbf{z}}_k^c, \mathbf{X}) = \begin{bmatrix} -p_{x,i} + p_{x,j} \\ -p_{y,i} + p_{y,j} \\ \Delta \hat{\theta}_{ij} - \theta_i + \theta_j \end{bmatrix} \quad (7)$$

where $[p_{x,i}, p_{y,i}, \theta_i]^T$ and $[p_{x,j}, p_{y,j}, \theta_j]^T$ denote the pose and heading at the time instants t_i and t_j , respectively. The subscript in $\hat{\mathbf{z}}_k^c$ denotes the k th loop closure. Unlike visual or LiDAR SLAM, the feature in behavior SLAM using IMU is notably sparse. It can only pinpoint where the corner is. When the loop closure is activated, we can only presume that the user passes through the same point. Consequently, the difference between their current and historical position is zero (the turning behavior occurs on the trajectory). However, the heading can be computed with the IMU and magnetometer, which can form a relative heading transformation in the loop closure using absolute headings at different time instants.

V. MULTIHYPOTHESIS TURNING BEHAVIOR DATA ASSOCIATION

It is important to note that the proposed method involves multiple factor graphs, and Section IV only refers to the factor formulation in one such graph. In this section, we will introduce the process of constructing multihypothesis data associations. Rather than relying on PF, this article integrates multiple hypotheses with the factor graph to realize the IMU-only behavior SLAM.

A. Turning Detection and Multihypothesis Model Formulation

In this article, the smoothed angular velocity sequence is employed to detect the turning behavior based on the peak detection algorithm [66]. Given that $\hat{\mathbf{w}}_{k-n:n}^b$ is expressed in the body frame, it needs to be transformed into the world frame using the rotation matrix sequence $\hat{\mathbf{R}}_{k-n:k}$, which is estimated by an orientation filter based on acceleration, angular velocity, and magnetic-field vector. The vertical angular rate is used to detect the turning and threshold for the turning detection is set to 30 deg/s based on empirical data when a person takes a turn. The first angular rate less than 2 deg/s on both sides of the peak is the valley, which is shown in Fig. 3. The time instant at which the peak occurs is denoted as t_o . The turning behavior sequence $\hat{\mathbf{w}}_{v,o-a:o+b}^w$, a sequence of vertical angular rates expressed in the world frame, can be determined by two valleys at time instants t_{o-a} and t_{o+b} , where a and b are integers. Throughout the

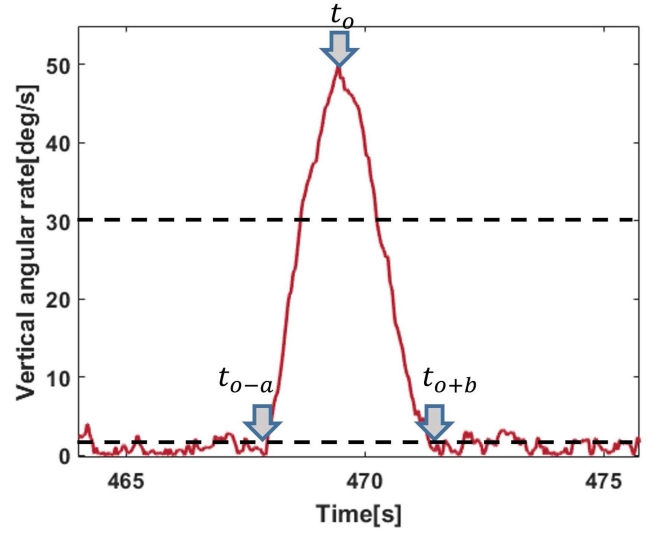


Fig. 3. Turning behavior sequence detection, in which the sequence can be determined by two valleys of vertical angular rates, $\hat{w}_{v,o-a}^w$ and $\hat{w}_{v,o+b}^w$, expressed in the world frame, as shown with the time instant t_{o-a} and t_{o+b} .

article, the subscript v denotes the vertical component of the corresponding vector.

The turning behavior can be divided into two categories: 1) normal turn, and 2) U-turn. In this article, the U-turn will not be built as a landmark as it can randomly happen at any point in the indoor space, which cannot be used to update the state. The normal turn can be recognized using the change of heading, which can be expressed as

$$\Delta \hat{\theta} = \text{abs}(\text{mean}(\hat{\theta}_{o-a:o}) - \text{mean}(\hat{\theta}_{o:o+b})) \quad (8)$$

where $\hat{\theta}_{o-a:o}$ and $\hat{\theta}_{o:o+b}$ denote the estimated heading sequence from the time instant t_{o-a} to t_o and the time instant t_o to t_{o+b} , respectively. The Madgwick filter [67] is used to estimate orientation, in which the measured angular velocity $\hat{\mathbf{w}}_k^b$ is used to predict the orientation, and measured acceleration $\hat{\mathbf{a}}_k^b$ and magnetic-field vector $\hat{\mathbf{m}}_k^b$ are used to correct the orientation. The normal turn and U-turn can be determined by

$$\begin{cases} \text{Normal turn} & \Delta \hat{\theta} < \theta_T \\ U - \text{turn} & \Delta \hat{\theta} \geq \theta_T \end{cases} \quad (9)$$

where θ_T is the threshold that is set as 90 deg in this article, which is set according to our test results.

In the proposed method, loop closure is the primary tool to restrain accumulative errors. To guarantee the reliability of data association, multihypothesis tracking [68], [69] is used to formulate the data association. The pedestrian's current position and its uncertainty are combined to detect possible landmarks.

When there is no turning behavior, PDR is employed to calculate the current position. The step length, denoted as L_k , is estimated using the Weinberg model [70]. This can

be expressed by

$$\hat{L}_k = K\sqrt{\hat{\sigma}_{v,\max}^w - \hat{\sigma}_{v,\min}^w} \quad (10)$$

where K denotes the constant, which is set according to our test results. The update of position in the horizontal direction is expressed as

$$\begin{aligned} \hat{p}_{x,k} &= \hat{p}_{x,k-1} + \hat{L}_k \cos \hat{\theta}_k \\ \hat{p}_{y,k} &= \hat{p}_{x,k-1} + \hat{L}_k \sin \hat{\theta}_k \end{aligned} \quad (11)$$

where $\hat{p}_{x,k}$ and $\hat{p}_{y,k}$ represent the x -axis and y -axis positions at the time instant t_k , $\hat{p}_{x,k-1}$ and $\hat{p}_{y,k-1}$ are the x -axis and y -axis positions at the time instant t_{k-1} . During the turning, the landmark position $[\hat{p}_{x,o}^m, \hat{p}_{y,o}^m]^T$ is obtained. The sequence number of the given landmark is represented as m .

The uncertainty of current position is expressed by the covariance and its propagation model at the time instant t_o can be expressed as [71]

$$\hat{\Sigma}_o = \hat{\mathbf{F}}_o \hat{\Sigma}_{o-1} \hat{\mathbf{F}}_o^T + \hat{\mathbf{G}}_o \mathbf{Q} \hat{\mathbf{G}}_o^T \quad (12)$$

where $\hat{\mathbf{F}}_o$ is the state transition matrix of PDR in (11). \mathbf{Q} is the noise covariance matrix of the step length and heading. $\hat{\mathbf{G}}_o$ is the coefficient matrix of noise covariance matrix \mathbf{Q} , which can be represented as

$$\hat{\mathbf{G}}_o = \begin{bmatrix} \cos \hat{\theta}_o & -\hat{L}_o \sin \hat{\theta}_o \\ \sin \hat{\theta}_o & -\hat{L}_o \cos \hat{\theta}_o \end{bmatrix}. \quad (13)$$

Equation (12) indicates the prediction process of the uncertainty of pedestrian position. An update process should also be included to avoid an unlimited uncertainty increase. After each optimization, the posterior distribution of states $p(\mathbf{X}|\mathbf{Z})$ can be obtained [61], in which \mathbf{X} and \mathbf{Z} represent the state vector and measurement vector. $p(\mathbf{X}|\mathbf{Z})$ is assumed to follow a Normal distribution with an estimated mean of $\hat{\mu}$ and covariance $\hat{\Sigma}$. $\hat{\Sigma}_o$ will then be updated by $\hat{\Sigma}$.

The uncertainty of pedestrian's position $\hat{\Sigma}_o$ can be used to obtain the confidence ellipse, which is shown in Fig. 4. Assuming that $\hat{\Sigma}_o$ can be expressed as follows:

$$\hat{\Sigma}_o = \begin{bmatrix} \hat{P}_{11} & \hat{P}_{12} \\ \hat{P}_{12} & \hat{P}_{22} \end{bmatrix}. \quad (14)$$

The radius of the ellipse can be obtained by [72]

$$\begin{aligned} \hat{\lambda}_1 &= \sqrt{\frac{\hat{P}_{11} + \hat{P}_{22}}{2} + \sqrt{\left(\frac{\hat{P}_{11} - \hat{P}_{22}}{2}\right)^2 + \hat{P}_{12}^2}} \\ \hat{\lambda}_2 &= \sqrt{\frac{\hat{P}_{11} + \hat{P}_{22}}{2} - \sqrt{\left(\frac{\hat{P}_{11} - \hat{P}_{22}}{2}\right)^2 + \hat{P}_{12}^2}} \end{aligned} \quad (15)$$

where $\hat{\lambda}_1$ and $\hat{\lambda}_2$ denote the radius of the major and minor axis, the angle $\hat{\alpha}$ from the positive x -axis to the ellipse's major axis can be expressed as

$$\hat{\alpha} = \begin{cases} 0 & \hat{P}_{12} = 0 \text{ and } \hat{P}_{11} \geq \hat{P}_{22} \\ \frac{\pi}{2} & \hat{P}_{12} = 0 \text{ and } \hat{P}_{11} < \hat{P}_{22} \\ \text{atan2}(\hat{\lambda}_1^2 - \hat{P}_{11}, \hat{P}_{12}) & \text{otherwise} \end{cases} \quad (16)$$

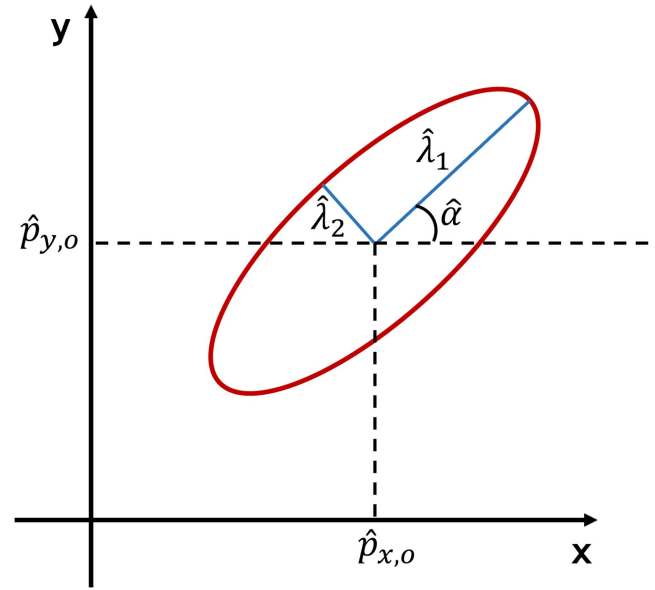


Fig. 4. Illustration of the confidence ellipse. $\hat{\lambda}_1$ and $\hat{\lambda}_2$ denote the radius of the major and minor axis, $\hat{\alpha}$ expresses the angle from the positive x -axis to the ellipse's major axis, $\hat{p}_{x,o}$ and $\hat{p}_{y,o}$ signify ellipse's center coordinates.

According to (15) and (16), the parametric equation of confidence ellipse can be expressed as [73]

$$\begin{aligned} \hat{x}(t) &= \sqrt{cs}\hat{\lambda}_1 \cos \hat{\alpha} \cos t - \sqrt{cs}\hat{\lambda}_2 \sin \hat{\alpha} \sin t + \hat{p}_{x,o} \\ \hat{y}(t) &= \sqrt{cs}\hat{\lambda}_1 \sin \hat{\alpha} \cos t + \sqrt{cs}\hat{\lambda}_2 \cos \hat{\alpha} \sin t + \hat{p}_{y,o} \end{aligned} \quad (17)$$

where $t \in [0, 2\pi]$ and cs stands for the chi-square value. In this article, potential loop closures are selected from landmarks within the ellipse. A confidence interval must be selected, which represents the range where the estimated position is expected to lie. Essentially, it defines the size of the confidence ellipse. Therefore, choosing a suitable confidence interval is crucial to prevent both the exclusion of landmarks and an increase in computational burden. In this article, we adopt a two-sigma confidence interval, corresponding to 0.95. Meanwhile, the freedom of the horizontal position is 2. Therefore, the chi-square value is 5.991 based on the chi-square distribution table.

The relationship among the moving trajectory, confidence ellipse, and known landmarks is given in Fig. 5. The known landmark represents the previous turning behavior. If there is a known landmark within the range of confidence ellipse, it means that the pedestrian passes one known corner. Otherwise, it is built as a new landmark.

There can be multiple known landmarks within the ellipse. In addition, even though there are known landmarks inside the ellipse, the turning action represented by the black dotted line in Fig. 5 may happen at a new corner. Thus, different hypotheses need to be formulated to ensure the reliability of data association. According to the multihypothesis tracking theory, multiple hypotheses are formed, each corresponding to a unique state estimate. Each hypothesis is assigned a probability to indicate its likelihood of occurrence. The hypothesis with the highest probability

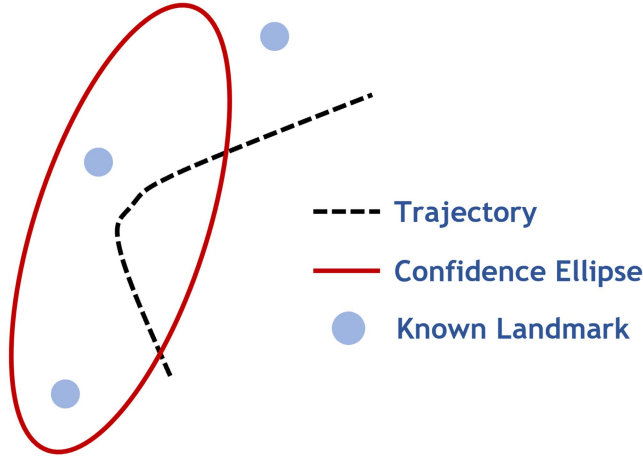


Fig. 5. Relationship among the moving trajectory, confidence ellipse, and known landmarks. Known landmarks within the ellipse are considered as candidates for loop closures.

at a given moment will be selected, and its associated state estimate is considered as the current estimation. The probability of one hypothesis at the time instant t_o is represented as

$$\text{prob}(\Omega_o^j) = \eta(f(\mathbf{z}_o) \text{prob}_o^{\text{det}})^\delta \circ \lambda_{\text{new}}^{1-\delta} \circ \text{prob}(\Omega_{o-1}^j) \quad (18)$$

where Ω_o^j is the j th hypothesis at the time instant t_o . η denotes the normalizer, which is calculated by dividing the sum of all probabilities by the corresponding probability. Its purpose is to adjust each computed probability so that their cumulative sum equals 1. $f(\mathbf{z}_o)$ is the probability density of the actual measurement, and $\text{prob}_o^{\text{det}}$ represents the reciprocal of the landmark number in the ellipse. For example, if there are N landmarks in the ellipse, $\text{prob}_o^{\text{det}}$ is $1/N$. λ_{new} represents the occurrence rate of a new landmark, which varies depending on the specific environment. The denser the landmarks, the higher this parameter is set. Conversely, in areas where landmarks are relatively sparse, such as corners, this parameter is set lower. An effective method for evaluating the order of magnitude of λ_{new} involves dividing the area of the site by the number of landmarks. Based on the order of magnitude, we have chosen a suitable value of 0.001 for λ_{new} , as determined by our test results. When a landmark is recognized, the value of δ is set to 1. If not, it is set to 0. In (18), the probability density of actual measurement, given that the mean and covariance matrix of the multivariate Normal distribution are $\hat{\mathbf{z}}_o$ and $\hat{\mathbf{U}}_o$, is represented as

$$f(\mathbf{z}_o) = \mathcal{N}(\mathbf{z}_o; \hat{\mathbf{z}}_o, \hat{\mathbf{U}}_o) \quad (19)$$

where \mathbf{z}_o and $\hat{\mathbf{z}}_o$ are the actual and predicted measurement, respectively. $\hat{\mathbf{U}}_o$ expresses the covariance matrix of $\hat{\mathbf{z}}_o$. Given that the measurement is the horizontal position difference between the pedestrian and landmark and the pedestrian is assumed to revisit the same location, the actual measurement can be expressed as

$$\mathbf{z}_o = [0, 0]^T. \quad (20)$$

$\hat{\mathbf{z}}_o$ signifies the predicted position difference between the landmark and pedestrian, given by

$$\begin{aligned} \hat{z}_{x,o} &= \hat{p}_{x,o}^m - \hat{p}_{x,o} \\ \hat{z}_{y,o} &= \hat{p}_{y,o}^m - \hat{p}_{y,o}. \end{aligned} \quad (21)$$

$\hat{\mathbf{U}}_o$ is represented as

$$\hat{\mathbf{U}}_o = \hat{\mathbf{H}}_o \hat{\Sigma}_o \hat{\mathbf{H}}_o^T + \hat{\mathbf{H}}_o^m \hat{\Sigma}_o^m \hat{\mathbf{H}}_o^{mT} + \Sigma_o^z \quad (22)$$

where $\hat{\mathbf{H}}_o$ and $\hat{\mathbf{H}}_o^m$ represent the Jacobian matrices of (21), corresponding to the positions of the pedestrian and the landmark, respectively. $\hat{\Sigma}_o^m$ is the noise covariance matrix of the position of the m th landmark. Σ_o^z is the noise covariance matrix of the predicted measurement $\hat{\mathbf{z}}_o$. $\hat{\mathbf{U}}_o$ is given as follows based on the Cholesky decomposition:

$$\hat{\mathbf{U}}_o = \hat{\mathbf{L}} \hat{\mathbf{L}}^T. \quad (23)$$

Therefore, (19) can be further indicated as

$$f(\mathbf{z}_o) = \frac{1}{2\pi |\hat{\mathbf{L}}|} \exp\left(-\frac{(\mathbf{z}_o - \hat{\mathbf{z}}_o)^T \hat{\mathbf{U}}_o^{-1} (\mathbf{z}_o - \hat{\mathbf{z}}_o)}{2}\right). \quad (24)$$

According to (18), the likelihood of each hypothesis is calculated. As multiple hypotheses are maintained, the number of hypotheses continues to increase, potentially leading to a higher computational load. To manage this, the N-scan back pruning method is used to trim hypotheses, thereby maintaining a manageable number of hypotheses.

B. Key Points Data Association

While the landmark represents a single point in the turn, several potential points during the turn are also utilized for data association. In the corner, three key points are defined: 1) the entry point, 2) the turning point, and 3) the exit point. These correspond to the time instants t_{o-a} , t_o , and t_{o+b} , respectively, as shown in Fig. 3. Time interval $[t_{o-a}, t_{o+b}]$ can be represented by

$$\mathbf{T}_o = [t_{o-a}, t_{o+b}]. \quad (25)$$

In this article, six typical turning behaviors of the pedestrian that can be used for the formulation of loop closures are discussed, which are shown in Fig. 6. The red and blue lines represent the trajectory, and arrows with corresponding color stand for the moving direction. The blue circle denotes the data association point. In Fig. 6(a) and 6(b), three key points are all used to perform the data association given that lines are similar, but their direction is the reverse. In Fig. 6(c)–6(f), only one key point is utilized to be associated as their trajectories are different.

The pedestrian heading difference is used to determine the specific situation. Assuming that two turnings happen in \mathbf{T}_m and \mathbf{T}_n , four heading differences of the pedestrian are calculated as

$$\begin{aligned} \Delta\hat{\theta}_1 &= \text{abs}(\text{mean}(\hat{\theta}_{m-a:n}) - \text{mean}(\hat{\theta}_{n-a:n})) \\ \Delta\hat{\theta}_2 &= \text{abs}(\text{mean}(\hat{\theta}_{m:m+b}) - \text{mean}(\hat{\theta}_{n:n+b})) \\ \Delta\hat{\theta}_3 &= \text{abs}(\text{mean}(\hat{\theta}_{m-a:n}) - \text{mean}(\hat{\theta}_{n:n+b})) \\ \Delta\hat{\theta}_4 &= \text{abs}(\text{mean}(\hat{\theta}_{m:m+b}) - \text{mean}(\hat{\theta}_{n-a:n})) \end{aligned} \quad (26)$$

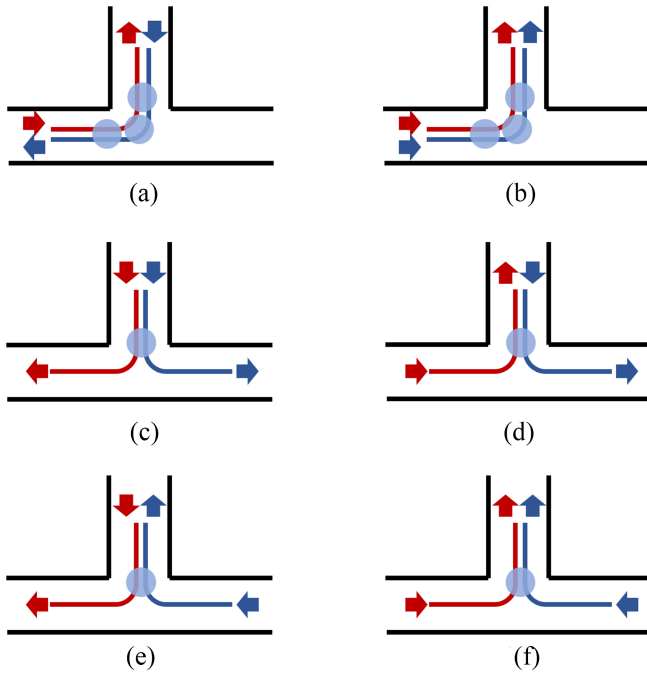


Fig. 6. Local trajectory in different situations. In (a) and (b), red and blue trajectories nearly overlap, and entry, turning, and exit points are associated to build loop closures. In (c)–(f), red and blue trajectories are not similar, and only one key point can be chosen to build the loop closure.

where a in \mathbf{T}_m is not equal to the counterpart in \mathbf{T}_n . This holds also for b . $\hat{\theta}_{o-a:o}$ is the estimated heading sequence from the time instant t_{o-a} to t_o , and $\hat{\theta}_{o:o+b}$ stands for the estimated heading sequence from the time instant t_o to t_{o+b} . Assuming that red and blue lines represent turns in \mathbf{T}_m and \mathbf{T}_n , the corresponding relation between the situation in Fig. 6 and the heading difference is given as follows:

$$\begin{cases} (a) & \Delta\hat{\theta}_3 > \theta_T \text{ and } \Delta\hat{\theta}_4 > \theta_T \\ (b) & \Delta\hat{\theta}_1 < \theta_T \text{ and } \Delta\hat{\theta}_2 < \theta_T \\ (c) & \Delta\hat{\theta}_1 < \theta_T \text{ and } \Delta\hat{\theta}_2 > \theta_T \\ (d) & \Delta\hat{\theta}_3 < \theta_T \text{ and } \Delta\hat{\theta}_4 > \theta_T \\ (e) & \Delta\hat{\theta}_3 > \theta_T \text{ and } \Delta\hat{\theta}_4 < \theta_T \\ (f) & \Delta\hat{\theta}_1 > \theta_T \text{ and } \Delta\hat{\theta}_2 < \theta_T. \end{cases} \quad (27)$$

Corresponding key points are matched based on (27). For instance, in Fig. 6(a) and 6(b), the turning points on the red and blue trajectories are paired. However, the entry point on the red trajectory is linked to the exit point on the blue trajectory in Fig. 6(a). In Fig. 6(c)–6(f), only one point is decided to do the data association.

Based on the explanations of the proposed method, we will outline the process of constructing multiple factor graphs. Let us assume that at this moment, there is only one factor graph (Root), which includes n states and three landmarks, denoted as A, B, and C. The construction of multiple factor graphs can be summarized as follows.

- 1) When a turning behavior is detected at the next state ($n + 1$), and the confidence ellipse is calculated. Let us assume that A and B are within the confidence

ellipse, but C is outside of it. We can hypothesize that either A or B is the loop closure, but we do not know which one. Therefore, two hypotheses are derived: one suggests that A is the loop closure point, while the other suggests that B is the loop closure point.

- 2) Moreover, there is a third hypothesis that neither A nor B are the loop closure points, which implies that the user has taken a turn at a new corner. Finally, the Root derives three new factor graphs based on these hypotheses. These three factor graphs inherit all factors of the Root up to the n th state.
- 3) The differences lie in the following: In the first factor graph, there is an edge between the $n + 1$ th state and the state where the A landmark was first built. In the second factor graph, there is an edge between the $n + 1$ th state and the state where the B landmark was first built. The edge is the loop closure. In the third factor graph, there is no edge connecting the $n + 1$ th state with the previous states, but the time instant corresponding to the $n + 1$ th state is regarded as the moment when a new landmark, marked as D, is first built.
- 4) We will calculate the likelihood of occurrence for these three hypotheses, which will be assigned to each factor graph and propagated over time. These three factor graphs will be used as the new Roots. When the next turning behavior is detected, they will derive new hypotheses (factor graphs) based on the above schemes.

VI. EXPERIMENT RESULTS AND DISCUSSION

Performance validation of the proposed method is carried out through experimental tests, the scheme of which is shown in Fig. 7. Test scenarios include the office building and underground car park. In all scenarios, the smartphone was held in hand, in a texting mode. We conducted a series of experiments in various environments, each with a distinct purpose as follows.

- 1) *Section C and D*: This part evaluates the positioning performance of different methods when the user is circling in the office building and underground car park.
- 2) *Section E*: Given that loop closure is key to restraining the positioning error in the proposed method, the method can be very effective in the environments of Section C and D. Unlike Section C and D, Section E describes a situation where the user departs for a destination and then returns. It aims to evaluate the proposed method when the loop closure cannot be triggered frequently.
- 3) *Section F*: This section aims to evaluate the robustness of the proposed method compared to the traditional factor graph.
- 4) *Section G*: The proposed method assumes that the user passes through the same point when the loop

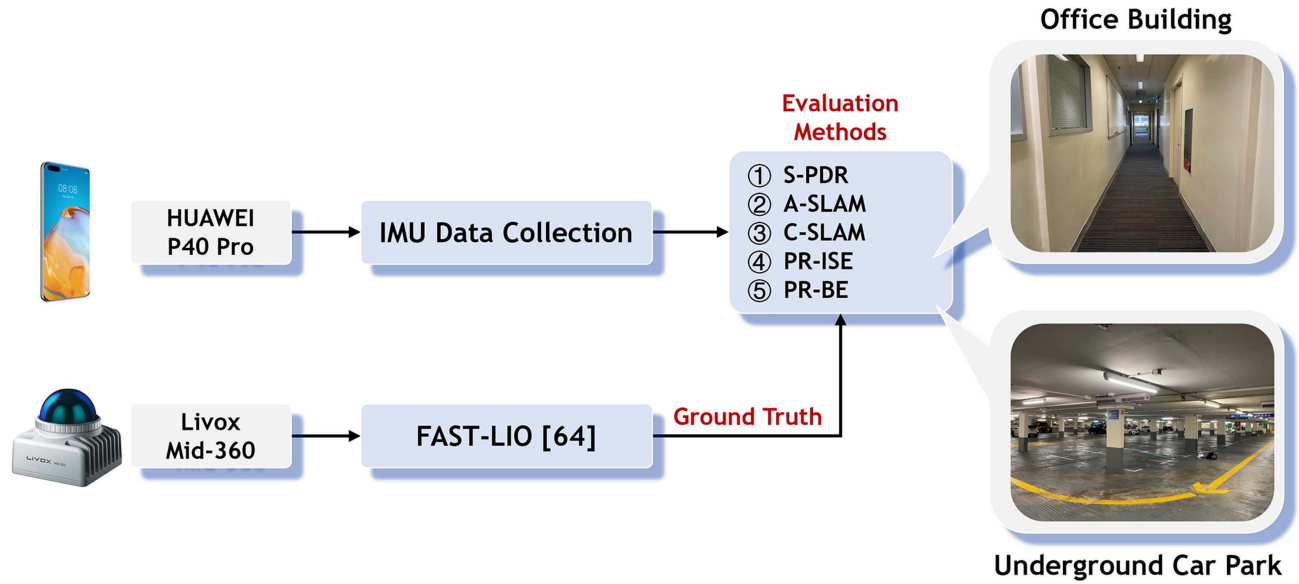


Fig. 7. Experimental test scheme. A LiDAR-inertial system was used to provide the ground truth. Test scenarios include the office building and underground car park.

closure is triggered. This part aims to test the proposed method when the user does not follow the same route to activate the loop closure.

- 5) *Section H*: The confidence ellipse is key to determining potential loop closures, which is affected by the uncertainty setting of PDR errors. This part aims to evaluate the proposed method under different uncertainty settings.
- 6) *Section I*: This part aims to compare the time consumption of different methods.
- 7) *Section J*: This part aims to evaluate the proposed method in comparison to the version presented at our conference.

A. Experiment Setup

During the experiment, IMU data was collected using an Android-based Huawei P40 Pro smartphone at a rate of 100 Hz. A 3-D LiDAR-inertial integrated system Mid-360 produced by Livox Technology Company was utilized to collect the point cloud and IMU data to perform fast, robust LiDAR-inertial odometry (FAST-LIO) [64], which was utilized as the ground truth, denoted as GT. Its weight and dimensions are 265 g and $65 \times 65 \times 60$ mm, so it is effortless for the pedestrian to hold it during the test. Meanwhile, its horizontal and vertical field of view (FOV) are 360 deg and 59 deg, and the maximum detection range is 70 m, which applies to indoor scenes. In addition, its drift is smaller than $0.05\% \times \text{Distance}$ [64]. In our tests, the longest trajectory is about 981 m, which increases with the distance covered by a person. In our tests, the longest trajectory is approximately 981 m, and its error at the endpoint can be considered to be less than 0.5 m. Furthermore, given that only the IMU is used, the LiDAR-inertial odometry is sufficiently accurate to serve as a reference. A desktop computer that is equipped with 13th Gen Intel Core i9-13900 K \times 32 processor, 32 GB

RAM and Ubuntu 20.04 operating system (OS) was used to execute different methods.

B. Evaluation Methods and Metrics

In this article, we utilize three conventional methods for comparative analysis.

- 1) *S-PDR*: This method tracks the position of pedestrians using smartphone-based dead reckoning [12], [13], [70].
- 2) *A-SLAM*: This method tracks the position of pedestrians by leveraging PDR and behavior landmarks within a PF framework [41], [45], [46].
- 3) *C-SLAM*: This method tracks the position of pedestrians by leveraging PDR and behavior landmarks within a factor graph framework. A single-hypothesis graph is constructed, and the iterative closest point (ICP) technique is employed for matching [74].

In addition, the proposed method is used in two forms for the comparison, denoted as the proposed instantaneous estimation (PR-ISE) and the proposed batch estimation (PR-BE). PR-ISE and PR-BE essentially use the optimization algorithm, but their applications differ. PR-ISE serves as an online positioning solution, while PR-BE is used for offline trajectory estimation. Specifically, PR-ISE obtains the position estimation at a given time instant t using data up to the time instant t . PR-BE uses all available information from start to finish to generate an enhanced trajectory after all data has been collected.

To quantify the localization error of different methods, absolute trajectory error (ATE) and relative trajectory error (RTE) are used. The detailed calculation method for ATE and RTE can be found in [75].

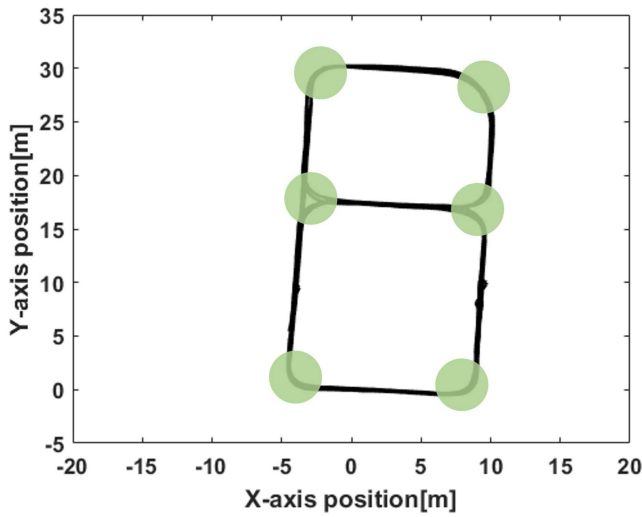


Fig. 8. Trajectory and landmarks in the office building.

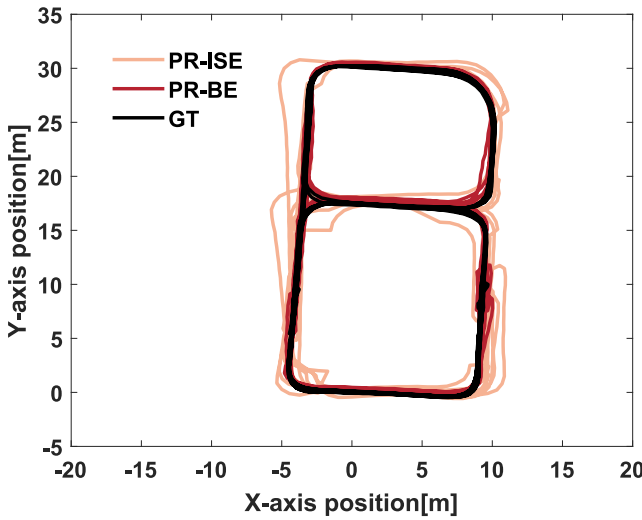


Fig. 9. Comparison of trajectory among PR-ISE, PR-BE, and GT in the office building.

C. Experimental Test in Office Building

The office building was first selected as the field to do the evaluation. The trajectory and landmarks in the office building are indicated in Fig. 8, where six green circles are used to denote the landmarks that achieve loop closure. The length of the moving trajectory is about 763 m.

The comparison of trajectory among PR-ISE, PR-BE, and GT in the office building is given in Fig. 9. The comparison of horizontal position error among all methods is indicated in Fig. 10. In all the box plots presented in our article, the red line represents the median of the position error. The red cross signifies the outliers of the position error. The black lines above and below denote the maximum and minimum of the position error, respectively. Lastly, the blue box illustrates the interquartile range of the position error.

It can be observed that the S-PDR method presents more significant errors than other methods, as it is based

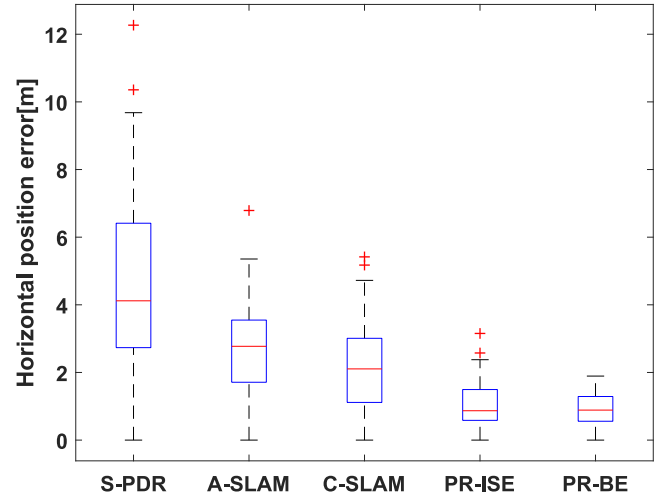


Fig. 10. Comparison of horizontal position error among all methods in the office building.

purely on dead reckoning. In the A-SLAM method, turning behaviors are treated as landmarks. This allows for the updating of pedestrian positions whenever a known landmark is identified. However, only the current state is maintained while previous states are discarded in A-SLAM. Known landmarks can be updated only if they are revisited, which can introduce extra errors to the state estimation. In C-SLAM, a predefined threshold for the match must be set, which heavily influences localization accuracy. On the one hand, the position uncertainty grows with the movement when no loop closure occurs. It is challenging to recognize loop closures with a fixed predefined threshold. On the other hand, if the turning behavior resembles (c)–(f) in Fig. 6, the inability of ICP to determine loop closure results in decreased localization accuracy.

The proposed methods can achieve better localization results than the current methods. Estimation results of PR-ISE come from the real-time hypothesis that has the highest likelihood of occurrence. It shows that PR-ISE presents a divergence in some sections since only past and current information is used. It can only rely on PDR to estimate the position before the subsequent loop closure. However, the position error can be corrected once the loop closure is effective. PR-ISE outperforms A-SLAM as past states are retained and can be iteratively optimized using the following information. Specifically, the historical landmark can be corrected if other data association is triggered, improving the estimation performance when the loop closure belonging to this landmark is built. In addition, position uncertainty is used to detect the potential known landmarks, multiple hypotheses are maintained, preventing PR-ISE from losing data associations. Meanwhile, PR-ISE can form point-to-point loop closures; thus, it can be superior to C-SLAM. PR-BE can achieve the best accuracy as all information is used, leading to more accurate trajectory estimation compared with other methods.

The horizontal position error CDF is indicated in Fig. 11. The horizontal position ATE and RTE is indicated in Table I.

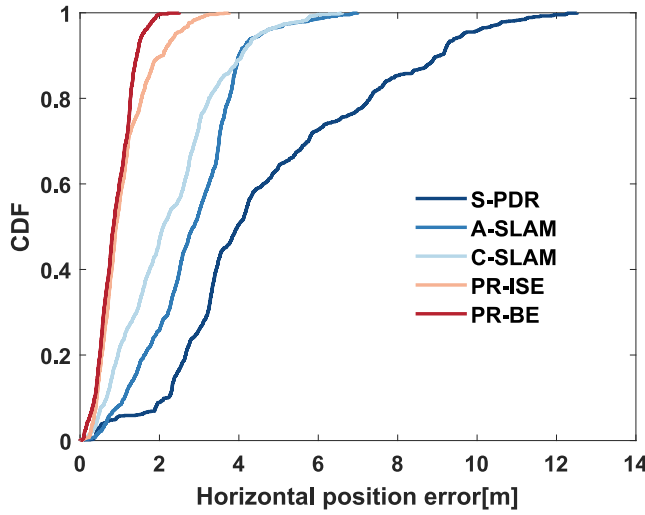


Fig. 11. Comparison of horizontal position error CDF among all methods in the office building.

TABLE I
Comparison of Horizontal Position ATE and RTE Among All Methods in the Office Building

Method	ATE (m)	RTE (m)	95% Error (m)
S-PDR	3.84	2.33	9.66
A-SLAM	2.13	1.76	4.56
C-SLAM	1.79	2.11	4.54
PR-ISE	0.89	1.28	2.41
PR-BE	0.72	0.88	1.58

The localization error of PR-BE is 1.58 m (95%) with an ATE of 0.72 m and RTE of 0.88 m. The localization error of PR-ISE is 2.41 m (95%) with an ATE of 0.89 m and RTE of 1.28 m. The localization errors of S-PDR, A-SLAM, and C-SLAM are 9.66 m (95%), 4.56 m (95%), and 4.54 m (95%), respectively. The ATE of PR-ISE is reduced by 76.82%, 58.22%, and 50.28% compared to S-PDR, A-SLAM, and C-SLAM, respectively. The ATE of PR-BE is decreased by 19.10% compared with PR-ISE.

D. Experimental Test in Underground Car Park

The underground car park was then used to do further evaluation. The trajectory and turning behavior landmarks in the underground car park are indicated in Fig. 12, where seven green circles are used to denote the landmark. The length of the moving trajectory is about 932 m.

The comparison of trajectory among PR-ISE, PR-BE, and GT in the underground car park is shown in Fig. 13. The comparison of horizontal position error among all methods is indicated in Fig. 14. It shows that the trajectory estimated by PR-BE is closer to the GT than PR-ISE. PR-ISE presents the divergence as subsequent loop closures cannot be utilized to correct the real-time localization results. As for traditional methods, S-PDR indicates the most significant errors as the pedestrian position is derived using the accumulation of displacements in two horizontal directions. Unlike the factor graph, A-SLAM uses only the measurement and state prediction at the current time

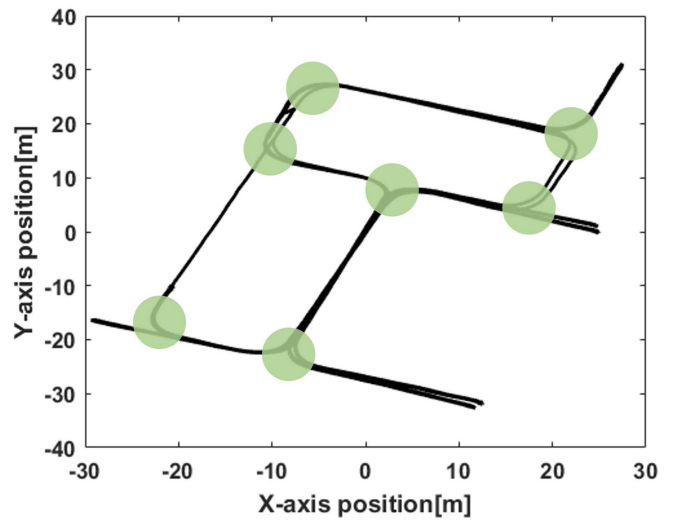


Fig. 12. Trajectory and landmarks in the underground car park.

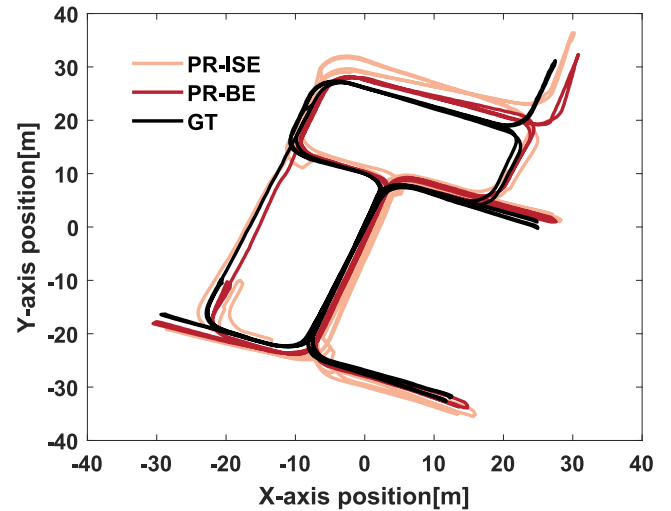


Fig. 13. Comparison of trajectory among PR-ISE, PR-BE, and GT in the underground car park.

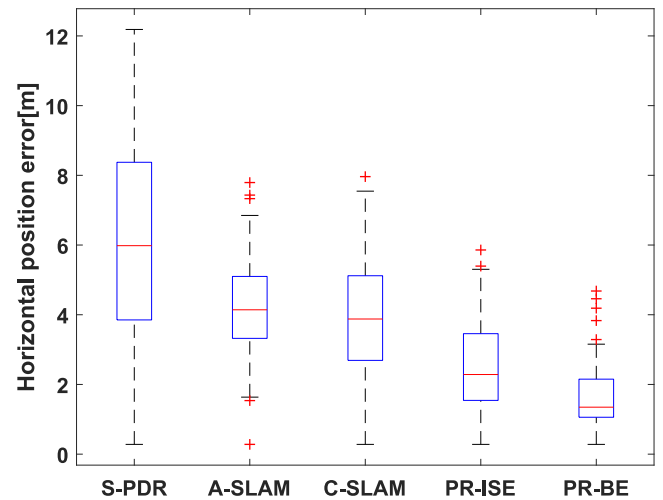


Fig. 14. Comparison of horizontal position error among all methods in the underground car park.

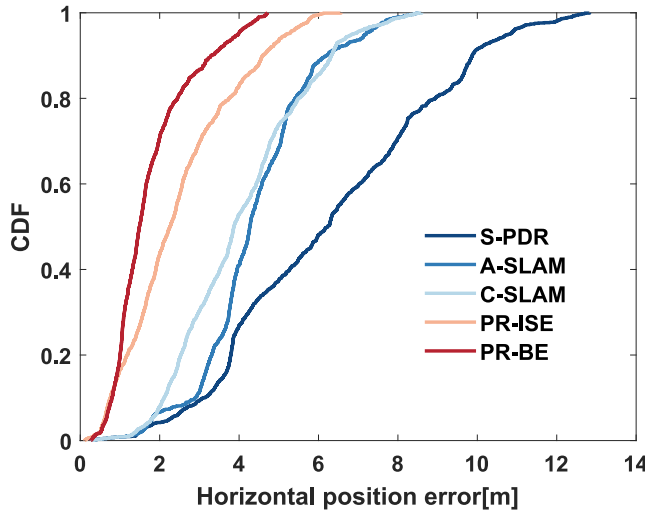


Fig. 15. Comparison of horizontal position error CDF among all methods in the underground car park.

TABLE II
Comparison of Horizontal Position ATE and RTE Among All Methods in the Underground Car Park

Method	ATE (m)	RTE (m)	95% Error (m)
S-PDR	4.81	2.30	10.83
A-SLAM	3.21	2.11	7.20
C-SLAM	3.04	2.27	6.90
PR-ISE	1.97	1.93	5.19
PR-BE	1.38	1.30	3.97

instant for state estimation. The use of a limited amount of data means that it cannot optimize historical states, which could affect the accuracy of the current state estimation. In C-SLAM, the fixed threshold for matching gives rise to the loss of loop closures, which can also lead to reduced localization performance.

Fig. 15 shows the comparison of horizontal position error CDF. Table II gives the comparison of ATE and RTE. The localization error of S-PDR is 10.83 m (95%) with an ATE of 4.81 m and RTE of 2.30 m. The localization error of A-SLAM is 7.20 m (95%) with an ATE of 3.22 m and RTE of 2.11 m. The localization error of C-SLAM is 6.90 m (95%) with an ATE of 3.04 m and RTE of 2.27 m. The localization error of PR-ISE is 5.19 m (95%) with an ATE of 1.97 m and RTE of 1.93 m. It shows that PR-ISE can obtain the best real-time localization accuracy with a decrease of 59.04%, 38.82%, and 35.20% in ATE compared with S-PDR, A-SLAM, and C-SLAM. The localization error of PR-BE is 3.97 m (95%) with an ATE of 1.38 m and RTE of 1.30 m. The ATE of PR-BE is reduced by 29.95% compared with PR-ISE.

E. Experimental Test in Return Journey Scenarios

In Section VI-C and VI-D, given that there is no destination in either test, the pedestrian moves in a survey mode. This implies that the pedestrian is circling around this area to obtain a more accurate localization and behavior map for future navigation purposes. In this section, we present a

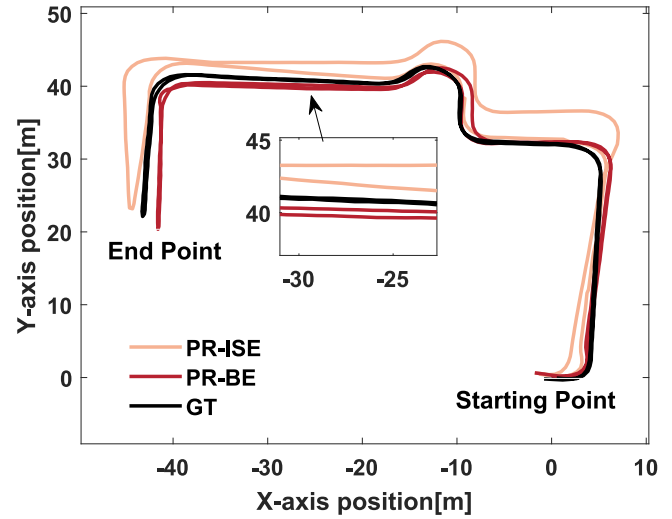


Fig. 16. Comparison of trajectory among PR-ISE, PR-BE, and GT in the first return journey scenario.

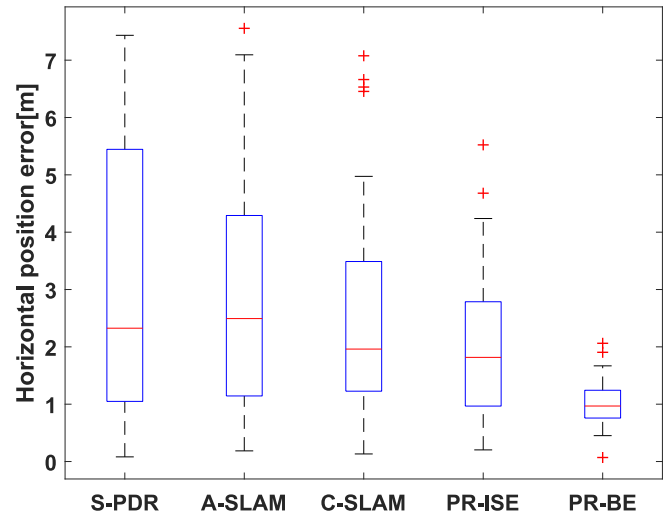


Fig. 17. Comparison of horizontal position error among all methods in the first return journey scenario.

situation that is more typical of our daily lives to evaluate the performance of the proposed method. This situation is referred to as the return journey. In this scenario, the pedestrian departs from a starting point $[0, 0]^T$, travels to a selected endpoint, and then returns to the starting point. Two other sets of data in the office building were used to evaluate, and the lengths of the two moving trajectories are about 220 and 201 m, respectively.

The comparisons of trajectory are given in Figs. 16 and 19, respectively. Given that the user follows nearly the same route when departing for a selected point and then returning, the trajectory appears as a single line. An enlarged view is incorporated into each figure to clearly illustrate the departure and return process. It has been observed that PR-BE can achieve superior trajectory estimation. PDR can maintain the accuracy of short-term positioning, and there is no noticeable position drift when the first landmark is established. By utilizing all the data collected during the

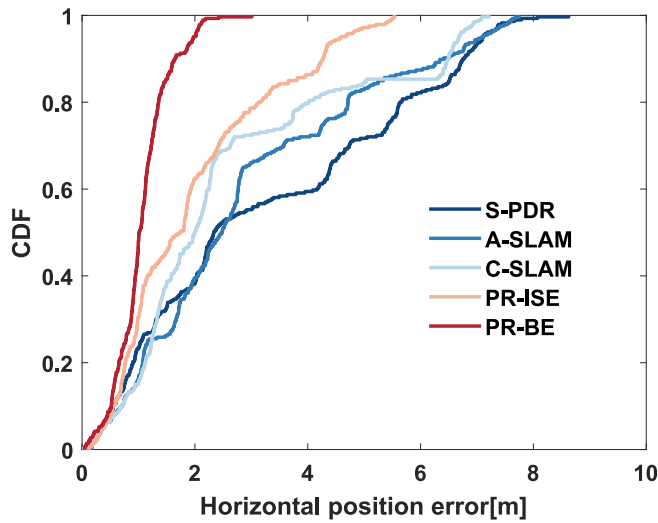


Fig. 18. Comparison of horizontal position error CDF among all methods in the first return journey scenario.

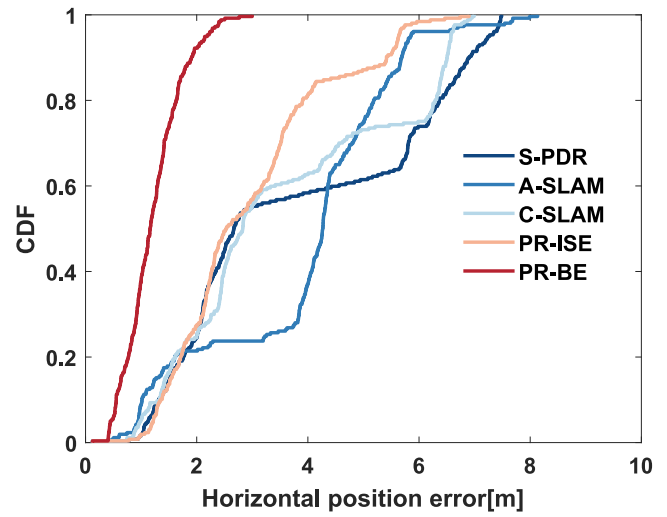


Fig. 21. Comparison of horizontal position error CDF among all methods in the second return journey scenario.

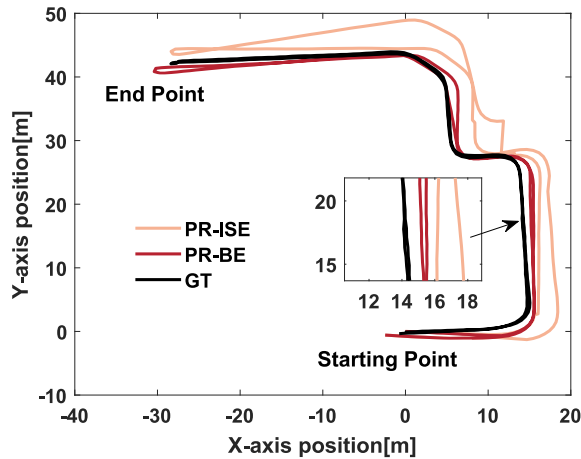


Fig. 19. Comparison of trajectory among PR-ISE, PR-BE, and GT in the second return journey scenario.

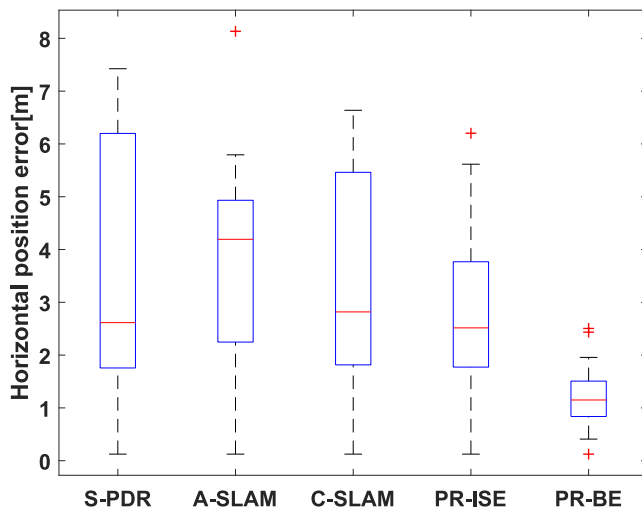


Fig. 20. Comparison of horizontal position error among all methods in the second return journey scenario.

TABLE III
Comparison of Horizontal Position ATE and RTE Among All Methods in the First Return Journey Scenario

Method	ATE (m)	RTE (m)
S-PDR	2.80	13.15
A-SLAM	2.50	4.39
C-SLAM	2.28	4.62
PR-ISE	1.74	1.63
PR-BE	0.78	0.88

test, the entire trajectory can be corrected. It shows that PR-ISE continues to drift even if the pedestrian revisits the same corner after they return from the end point. This is primarily because PR-ISE selects the result from multiple hypotheses based on probabilities calculated by (18). The position estimation corresponding to the hypothesis with the highest probability is then used as the output. Although the proposed method produces one hypothesis in which the loop closure is triggered in the same corner when the pedestrian returns from the end point, its likelihood of occurrence is less than the counterpart in which this corner is regarded as a new landmark due to the PDR drift. PR-ISE converges only when the pedestrian passes several corners, thereby increasing the likelihood of occurrence for the hypothesis associated with turning loop closures.

The horizontal position errors are given in Figs. 17 and 20. Figs. 18 and 21 indicate the horizontal position error CDF in both scenarios. Tables III and IV show the comparison of horizontal position ATE and RTE. It is noticed that PR-BE can still achieve better localization results than traditional methods. However, the performance of PR-ISE degrades in return journey scenarios. When the pedestrian travels in a single direction, the corner cannot be revisited for an extended duration, leading to the easy accumulation of PDR errors. When the pedestrian returns and revisits one corner, the real-time positioning results cannot be instantly

TABLE IV
Comparison of Horizontal Position ATE
and RTE Among All Methods in the
Second Return Journey Scenario

Method	ATE (m)	RTE (m)
S-PDR	3.06	2.57
A-SLAM	3.00	2.17
C-SLAM	2.81	4.68
PR-ISE	2.28	1.52
PR-BE	0.93	1.51

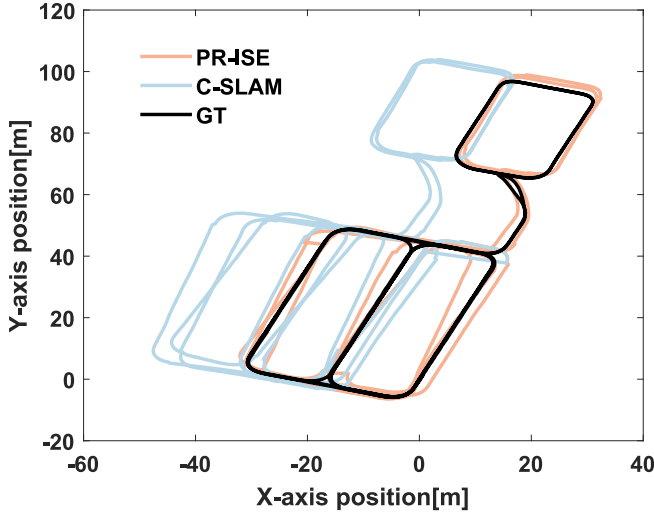


Fig. 22. Comparison of trajectory among PR-ISE, C-SLAM, and GT in the underground car park.

corrected. After several corners are revisited, the accumulated errors can be calibrated.

F. Discussion: Reliability of Multihypothesis Factor Graph Optimization

To guarantee the reliability of turning loop closures, the multihypothesis data association is utilized in the proposed method to formulate multiple factor graphs. This part shows the enhanced reliability with multihypothesis.

C-SLAM is based on a single hypothesis. Considering that there is only one possibility in this method, C-SLAM must choose one landmark for data association when confronted with several possible candidates, which easily leads to wrong loop closures when ambiguous data associations exist. The deterioration of state estimation is quite common in single-hypothesis factor graph for behavior SLAM. Another test in the underground car park was conducted, and the length of the moving trajectory is about 981 m. The comparisons of trajectory with PR-ISE and PR-BE are given in Figs. 22 and 23, respectively.

In Figs. 22 and 23, it can be observed that C-SLAM presents a wrong trajectory due to the incorrect data association. In C-SLAM, the single hypothesis ignores the potential possibilities of other data associations, which easily gives rise to wrong localization results. If one position is wrong, the subsequent state estimations are all affected. The proposed method maintains multiple factor graphs

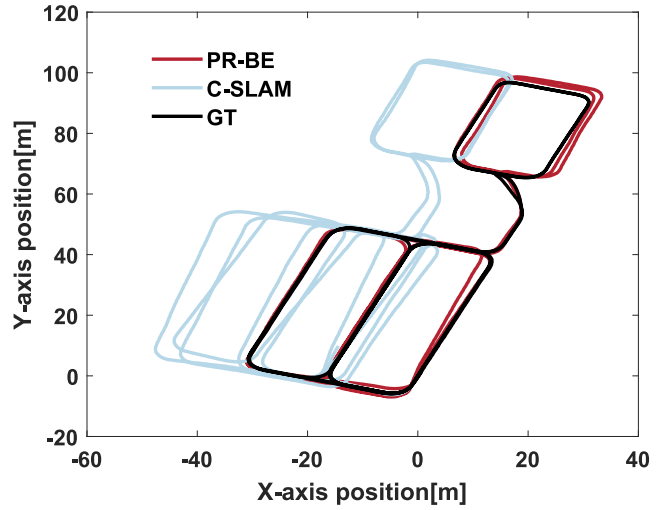


Fig. 23. Comparison of trajectory among PR-BE, C-SLAM, and GT in the underground car park.

with different possibilities simultaneously. The likelihood of occurrence for the hypothesis with correct associations is getting more prominent with time. Therefore, the multi-hypothesis FGO can improve the performance of IMU-only behavior SLAM.

G. Discussion: Point-Based Location Loop Closure With Different Paths

In visual or LiDAR SLAM approaches, the loop closure is often a transformation between poses at different time instants. On the contrary, the transformation is unavailable using only IMU due to different measuring schemes. Therefore, the user is assumed to walk along the same path when they revisit the same corner. In the residual, a solid constraint is formed to adjust associated points to be the same. However, it is improper at a wide corridor as pedestrians may walk along different paths.

Supplementary experimentations were done in the underground car park to evaluate the performance of the proposed approach in this situation. The length of the moving trajectory is about 548 m. The comparison of trajectory among different methods when the pedestrian walks along different paths is shown in Fig. 24, and the vertical angular rate in this test is given in Fig. 25. It can be noticed that S-PDR presents an evident drift, which makes it hard to obtain the transformation or define an area to build the loop closure. In the proposed approach, the points in the corner are associated to form loop closures. As shown in Fig. 25, the turnings in T_m and T_n happen at the same corner. Given that the walking directions are roughly the same at these two durations, the time instant t_m is associated with the time instant t_n , the time instant t_{m-a} is associated with the time instant t_{n-a} , and the time instant t_{m+b} is associated with the time instant t_{n+b} . The locations at the associated instants are regarded the same. Therefore, the trajectories at corners in the proposed method overlap. However, from the ground truth, the pedestrian does not walk along the same path. From Fig. 24, it can be observed that the differences

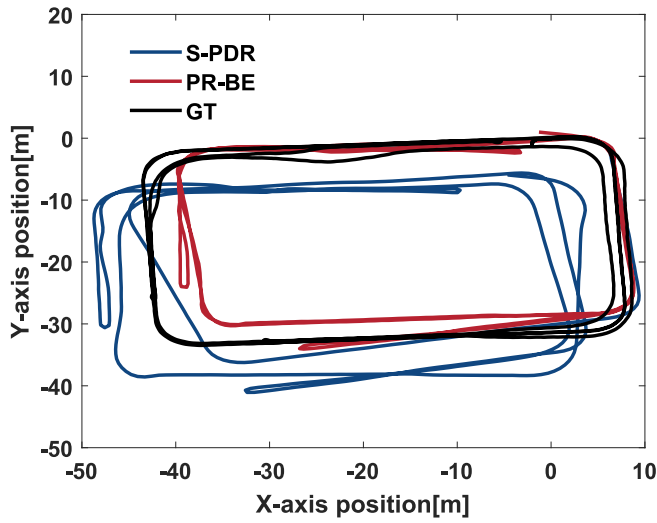


Fig. 24. Comparison of trajectory among S-PDR, PR-BE, and GT when the pedestrian walks along different paths.

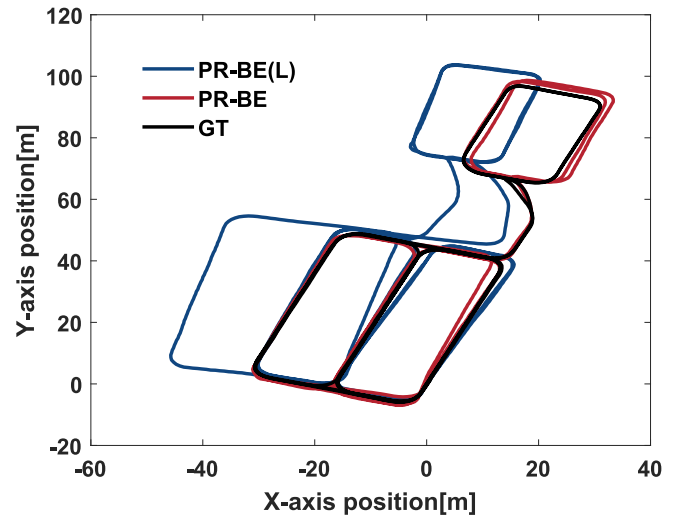


Fig. 26. Comparison of trajectory based on offline position estimation with different noise covariance matrix settings, in which L means the noise covariance matrix is set smaller.

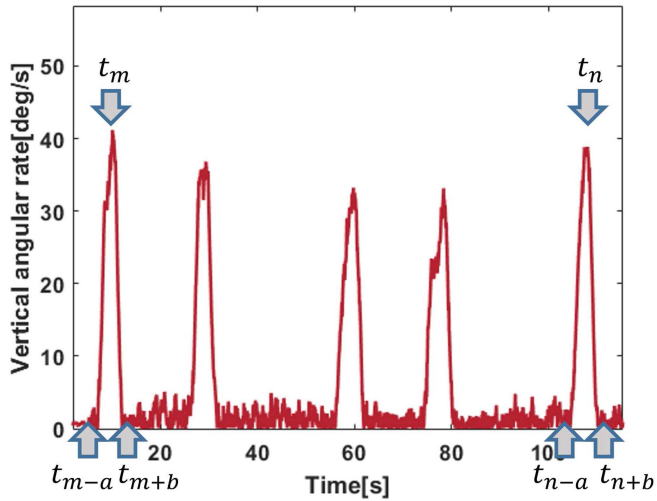


Fig. 25. Vertical angular rate in this test. Turnings in T_m and T_n happen at the same corner, and key points in T_m and T_n are associated to build loop closures.

in the paths are prominent in four areas: 1) the top left corner, 2) the top right corner, 3) the bottom right corner, and 4) the right trajectories. The solid constraints align all path differences closely, which can result in a slightly scaled-down trajectory.

H. Discussion: Performance Comparison of the Proposed Method With Different Uncertainty Setting

In the proposed method, the confidence ellipse obtained from the uncertainty of the current position is used to detect the potential known landmarks. When there is no loop closure, the uncertainty of human position will be propagated based on the prior setting of noise covariance matrix Q in the (12). Recognizing potential inaccuracies in the statistical analysis and avoiding improper settings that distort the result, Q in real-world applications needs to be set slightly larger than the result from the statistical analysis.

In PR-ISE and PR-BE, Q is set as $\text{diag}(0.40^2, 0.02^2)$, where the standard deviations of step length and heading noise are 0.40 m and 0.02 rad, respectively. This ensures a broader coverage range of the confidence ellipse, thereby preventing the loss of information. This part discusses the estimation results under different noise covariance matrix settings. The comparison of trajectory based on PR-BE with different noise covariance matrix settings is given in Fig. 26. (L) means that the noise covariance matrix is set smaller than statistical results.

It can be observed that positioning results from PR-BE(L) diverge. The small noise covariance matrix in PR-BE(L) can cause a narrow range of confidence ellipse, leading to the loss of loop closure constraints. We tend to set the noise covariance matrix larger than statistical results to guarantee the inclusion of the true loop closure.

I. Discussion: Time Consumption of Multihypothesis Factor Graph Optimization

The time consumption of the proposed method is also given in this article. When the number of hypotheses exceeds six, the N-scan back pruning is implemented, based on our test results. The hypotheses number during the operation is shown in Fig. 27. It also shows that the number of hypotheses is sometimes more than six from Fig. 27, because the newly added hypotheses may be more than the pruned ones. However, it can maintain the number of hypotheses at a stable range. The relationship between the processing time and data duration among A-SLAM, C-SLAM, and PR-ISE is given in Fig. 28. It indicates the total computing time for a duration of time. Thus, the value in the vertical coordinate grows with the horizontal coordinate. In the test, A-SLAM incorporates 100 particles. Both C-SLAM and PR-ISE operate on the iSAM2 framework. It can be noticed that the C-SLAM possesses the fastest processing speed as there is one factor graph in this method. Even though

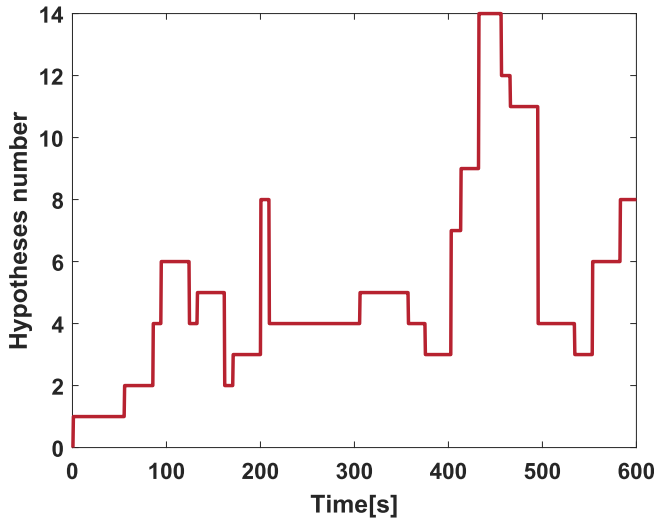


Fig. 27. Number of hypotheses in the proposed method during the operation.

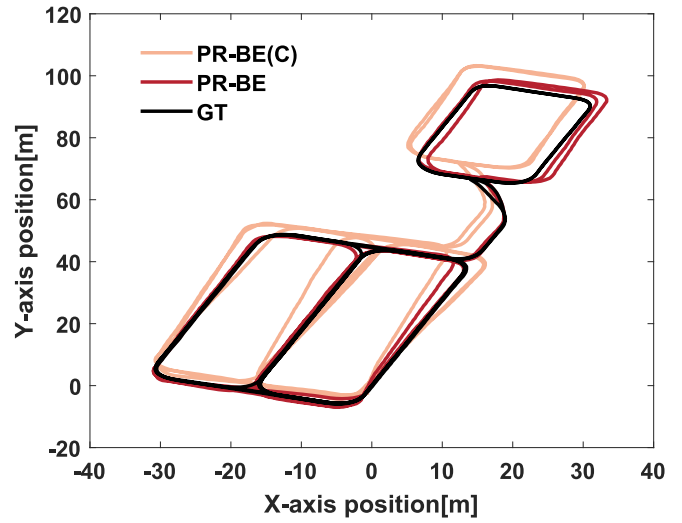


Fig. 29. Comparison of trajectory among PR-BE(C) and PR-BE in the underground car park.

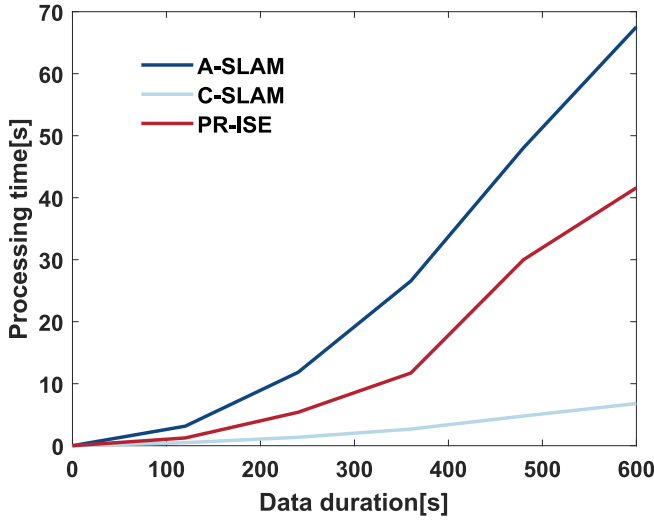


Fig. 28. Relationship between the processing time and data duration among A-SLAM, C-SLAM, and PR-ISE.

the filter only requires estimating the state at the current time instant, processing 100 particles can add additional computational burden. The results indicate that PR-ISE can outpace A-SLAM in processing speed if the number of hypotheses is constrained. From 360 to 480 s, there is an increase in the slope of PR-ISE's processing time as the number of hypotheses increases during this interval.

It is crucial to note that Fig. 28 only displays real-time positioning methods. Although PR-ISE and PR-BE both employ optimization, their operational methods differ. Each time PR-ISE processes new data, it optimizes both the current state and all previous states, subsequently providing the current position. In contrast, PR-BE uses data from the start to the end, delivering the trajectory only after all the data has been collected. PR-BE is better suited for offline trajectory estimation applications, which implies that its operational scheme differs from those of A-SLAM,

C-SLAM, and PR-ISE. Therefore, PR-BE is not included in this discussion.

J. Discussion: Performance Comparison of the Proposed Method With the Conference Paper

Compared with our conference paper [10], this article introduces the turning behavior sequence and the formation of confidence ellipse for data association. In [10], we only adopted a simple method to determine if known landmarks can be associated for loop closures. Specifically, it considers the minimum value among diagonal elements of the user's position covariance matrix. The square root of this minimum value is then calculated to get the range. If the distance between the user's position and a landmark's position is less than this computed range, the landmark is deemed a candidate. The simplified method is not always effective. In certain instances, it fails to include a sufficient number of candidates, leading to missed loop closures. Moreover, without the turning behavior sequence, using a single point for heading difference calculation can also lead to the loss of loop closures. Although it indicates that the results from the proposed method in Section VI-D align with those in our conference paper, this is given that the data used in the test is relatively ideal. This suggests that all the correct landmarks are incidentally incorporated in this test. In this section, we have further compared the performance of the proposed method with that presented in our conference paper using another set of data collected in the underground car park.

The comparisons of the trajectory and horizontal position error in the underground car park are presented in Figs. 29 and 30, respectively. The comparison of the ATE and RTE for the horizontal position is shown in Table V. Methods ending with (C) denote the corresponding methods used in the conference paper. From Fig. 29, it is evident that there is a deviation between PR-BE(C) and the ground truth, which is due to the loss of loop closure points. Meanwhile, Fig. 30 and Table V demonstrate that both

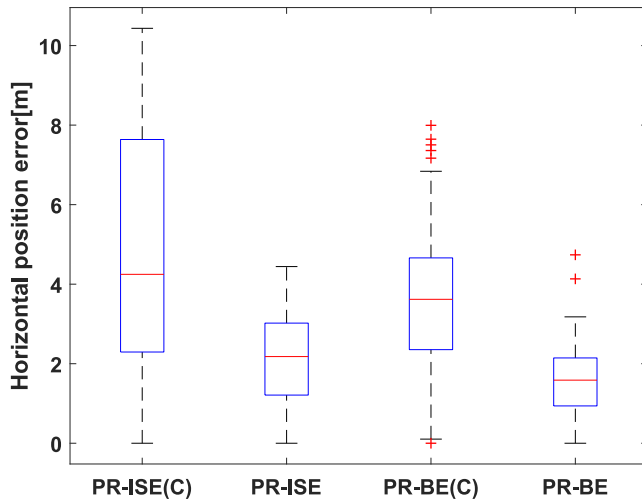


Fig. 30. Comparison of horizontal position error among PR-ISE(C), PR-ISE, PR-BE(C), and PR-BE in the underground car park.

TABLE V
Comparison of Horizontal Position ATE and RTE Among PR-ISE(C), PR-ISE, PR-BE(C), and PR-BE in the Underground Car Park

Method	ATE (m)	RTE (m)
PR-ISE(C)	3.95	3.58
PR-ISE	1.71	2.42
PR-BE(C)	2.90	2.68
PR-BE	1.32	1.97

PR-ISE and PR-BE outperform PR-ISE(C) and PR-BE(C). This validates the effectiveness of the introduced improved data association method.

K. Discussion: Challenges to the Proposed Method

Some challenges that limit the proposed method still exist. First, although the multihypothesis is utilized to handle the data association, the right trajectory cannot be recovered due to the very close turns. Furthermore, current turning behavior detection is barely based on the vertical angular rate, easily leading to missed or false recognition when the actual turns cannot be matched with predefined standards. Finally, the proposed method may fail in an environment with sparse features. For example, if there are fewer rapid turns over a long time, the system will be operated in pure PDR mode, leading to rapid error accumulation. Meanwhile, false loop closures can occur if the pedestrian has a completely random walk in an empty room, generating too many random turns.

VII. CONCLUSION

A factor graph-based smartphone IMU-only SLAM with multihypothesis turning behavior loop closures is proposed in this article. Multiple factor graphs, each associated with different probabilities (values used to assess the likelihood of occurrence), are constructed to ensure the reliability of data association. Key locations during

the turning behavior are associated to build loop closures. The experiments are done in the office building and underground car park. The results show that the proposed method outperforms the considered filter-based methods in instant localization estimation performance, as it utilizes more information, and historical states can be iteratively optimized to improve the estimation accuracy for current states further. Meanwhile, the proposed method can also obtain a smoother and more accurate trajectory. This article also discusses the performance of the proposed method in return journey scenarios and advantages of multihypothesis FGO compared to single-hypothesis FGO. Moreover, the performance of the proposed method under different paths and uncertainty settings is given. Finally, we present the processing time of the proposed method in comparison with other methods, and highlight its advantages over the approach outlined in our conference paper. It is worth noting that even though only one type of smartphone was used in our tests, the proposed method can be adapted to other types of smartphones. This is because the performance differences among IMUs in various mobile devices are not significant, considering that they are all low cost, which do not result in a significant difference in PDR and turning detection.

In addition, using the proposed method puts forward some requirements for the context. First, the scenario needs to possess some structured features, such as corners and corridors to constrain the movement of the pedestrian. Then, given that PDR suffers from severe drift, the pedestrians need to revisit previously visited places quite often. It will help when exploring small areas in return journey mode but fails when navigating to a remote goal. However, it is important to note that the method can be extended to a map-matching mode when the pedestrian travels to a distant location. This implies that the proposed SLAM method is utilized to obtain a behavior map, which can then be used to provide absolute positioning correction using map matching. Finally, although the online navigation error can be corrected once the loop closure is triggered and the proposed online method (PR-ISE) outperforms existing methods, the postprocessing mode (PR-BE) can perform better. Thus, PR-BE is more applicable to trajectory recovery and other offline analysis.

In the future, we will add more behaviors, such as taking the elevator and escalator and going up and down stairs, to this framework to overcome the issue in environments with sparse features. Intermittent sensing data, such as Wi-Fi, Bluetooth, or UWB, can also be added into the current framework to reduce the PDR error accumulation when the behavior landmark has not been encountered, which can help improve the SLAM data association. Moreover, the same type of landmarks close to each other can be combined to formulate a composite landmark to do the loop closure, and AI can be exploited to improve the recognition rate of behaviors, solving the problem of disabled detection. Ultimately, we plan to integrate our proposed method into mobile devices.

- [1] Y. Wang, A. Chernyshoff, and A. M. Shkel, "Study on estimation errors in ZUPT-aided pedestrian inertial navigation due to IMU noises," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 3, pp. 2280–2291, Jun. 2020.
- [2] C.-S. Jao et al., "PINDOC: Pedestrian indoor navigation system integrating deterministic, opportunistic, and cooperative functionalities," *IEEE Sensors J.*, vol. 22, no. 14, pp. 14424–14435, Jul. 2022.
- [3] S. Monica and G. Ferrari, "UWB-based localization in large indoor scenarios: Optimized placement of anchor nodes," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 2, pp. 987–999, Apr. 2015.
- [4] F. Gu, S. Valaee, K. Khoshelham, J. Shang, and R. Zhang, "Landmark graph-based indoor localization," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8343–8355, Sep. 2020.
- [5] Y. Wang, J. Kuang, Y. Li, and X. Niu, "Magnetic field-enhanced learning-based inertial odometry for indoor pedestrian," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022.
- [6] Q. Wang et al., "Recent advances in pedestrian inertial navigation based on smartphone: A review," *IEEE Sensors J.*, vol. 22, no. 23, pp. 22319–22343, Dec. 2022.
- [7] J. Xiao, Z. Zhou, Y. Yi, and L. M. Ni, "A survey on wireless indoor localization from the device perspective," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 1–31, 2016.
- [8] F. Liu et al., "Survey on WiFi-based indoor positioning techniques," *IET Commun.*, vol. 14, no. 9, pp. 1372–1383, 2020.
- [9] H. Lee, A. A. Abdallah, J. Park, J. Seo, and Z. M. Kassas, "Neural network-based ranging with LTE channel impulse response for localization in indoor environments," in *Proc. 20th Int. Conf. Control, Automat. Syst.*, 2020, pp. 939–944.
- [10] S. Bai, W. Wen, L.-T. Hsu, and Y. Yu, "Factor graph optimization-based indoor pedestrian SLAM with probabilistic exact activity loop closures using smartphone," in *Proc. 13th Int. Conf. Indoor Positioning Indoor Navigation*, 2023, pp. 1–8.
- [11] W. Kang and Y. Han, "SmartPDR: Smartphone-based pedestrian dead reckoning for indoor localization," *IEEE Sensors J.*, vol. 15, no. 5, pp. 2906–2916, May 2015.
- [12] T.-M. T. Dinh, N.-S. Duong, and K. Sandrasegaran, "Smartphone-based indoor positioning using BLE iBeacon and reliable lightweight fingerprint map," *IEEE Sensors J.*, vol. 20, no. 17, pp. 10283–10294, Sep. 2020.
- [13] X. Wang et al., "Tightly coupled integration of pedestrian dead reckoning and bluetooth based on filter and optimizer," *IEEE Internet Things J.*, vol. 10, no. 8, pp. 7327–7342, Apr. 2023.
- [14] M. Zhu, Y. Wu, and S. Luo, "f²IMU-R: Pedestrian navigation by low-cost foot-mounted dual IMUs and interfoot ranging," *IEEE Trans. Control Syst. Technol.*, vol. 30, no. 1, pp. 247–260, Jan. 2022.
- [15] X. Li, "Cellular base station assisted indoor positioning," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 2, pp. 592–606, Apr. 2019.
- [16] C. Yang and H.-R. Shao, "WiFi-based indoor positioning," *IEEE Commun. Mag.*, vol. 53, no. 3, pp. 150–157, Mar. 2015.
- [17] R. Faragher and R. Harle, "An analysis of the accuracy of bluetooth low energy for indoor positioning applications," in *Proc. 27th Int. Tech. Meeting Satell. Division Inst. Navigation*, 2014, pp. 201–210.
- [18] A. Poullose and D. S. Han, "UWB indoor localization using deep learning LSTM networks," *Appl. Sci.*, vol. 10, no. 18, 2020, Art. no. 6290.
- [19] A. Poullose, J. Kim, and D. S. Han, "A sensor fusion framework for indoor localization using smartphone sensors and Wi-Fi RSSI measurements," *Appl. Sci.*, vol. 9, no. 20, 2019, Art. no. 4379.
- [20] Z. Li, L. Zhao, C. Qin, and Y. Wang, "WiFi/PDR integrated navigation with robustly constrained Kalman filter," *Meas. Sci. Technol.*, vol. 31, no. 8, 2020, Art. no. 084002.
- [21] Q. Tian, K. I.-K. Wang, and Z. Salcic, "An INS and UWB fusion approach with adaptive ranging error mitigation for pedestrian tracking," *IEEE Sensors J.*, vol. 20, no. 8, pp. 4372–4381, Apr. 2020.
- [22] S. Li, M. Hedley, I. B. Collings, and D. Humphrey, "Joint trajectory and ranging offset estimation for accurate tracking in NLOS environments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 1, pp. 3–14, Feb. 2020.
- [23] A. Poullose, O. S. Eyobu, and D. S. Han, "An indoor position-estimation algorithm using smartphone IMU sensor data," *IEEE Access*, vol. 7, pp. 11165–11177, 2019.
- [24] H. Yan, Q. Shan, and Y. Furukawa, "RIDI: Robust IMU double integration," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 621–636.
- [25] C. Chen, X. Lu, A. Markham, and N. Trigoni, "IONet: Learning to cure the curse of drift in inertial odometry," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 6468–6476.
- [26] S. Herath, H. Yan, and Y. Furukawa, "RoNIN: Robust neural inertial navigation in the wild: Benchmark, evaluations, & new methods," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 3146–3152.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [29] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.
- [30] W. Liu et al., "TLIO: Tight learned inertial odometry," *IEEE Robot. Automat. Lett.*, vol. 5, no. 4, pp. 5653–5660, Oct. 2020.
- [31] S. Khalifa and M. Hassan, "Evaluating mismatch probability of activity-based map matching in indoor positioning," in *Proc. Int. Conf. Indoor Positioning Indoor Navigation*, 2012, pp. 1–9.
- [32] E. Munoz Diaz, M. Caamano, and F. J. Fuentes Sánchez, "Landmark-based drift compensation algorithm for inertial pedestrian navigation," *Sensors*, vol. 17, no. 7, 2017, Art. no. 1555.
- [33] B. Zhou, Q. Li, Q. Mao, W. Tu, and X. Zhang, "Activity sequence-based indoor pedestrian localization using smartphones," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 5, pp. 562–574, Oct. 2015.
- [34] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, vol. 3, no. 1, pp. 4–16, Jan. 1986.
- [35] Y. Gu, D. Li, Y. Kamiya, and S. Kamijo, "Integration of positioning and activity context information for lifelog in urban city area," *NAVIGATION: J. Inst. Navigation*, vol. 67, no. 1, pp. 163–179, 2020.
- [36] T. Kang and Y. Shin, "Indoor navigation algorithm based on a smartphone inertial measurement unit and map matching," in *Proc. Int. Conf. Inf. Commun. Technol. Convergence*, 2021, pp. 1421–1424.
- [37] C. Cadena et al., "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, Dec. 2016.
- [38] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardes, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.
- [39] S. Cortés, A. Solin, E. Rahtu, and J. Kannala, "ADVIO: An authentic dataset for visual-inertial odometry," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 419–434.
- [40] A. Poullose and D. S. Han, "Hybrid indoor localization using IMU sensors and smartphone camera," *Sensors*, vol. 19, no. 23, 2019, Art. no. 5084.
- [41] M. Hardegger, S. Mazilu, D. Caraci, F. Hess, D. Roggen, and G. Tröster, "ActionSLAM on a smartphone: At-home tracking with a fully wearable system," in *Proc. Int. Conf. Indoor Positioning Indoor Navigation*, 2013, pp. 1–8.
- [42] M. Hardegger, D. Roggen, and G. Tröster, "3D ActionSLAM: Wearable person tracking in multi-floor environments," *Pers. Ubiquitous Comput.*, vol. 19, pp. 123–141, 2015.
- [43] M. Angermann and P. Robertson, "FootSLAM: Pedestrian simultaneous localization and mapping without exteroceptive sensors-hitchhiking on human perception and cognition," *Proc. IEEE*, vol. 100, no. Special Centennial Issue, pp. 1840–1848, May 2012.
- [44] H. Abdelnasser et al., "SemanticSLAM: Using environment landmarks for unsupervised indoor localization," *IEEE Trans. Mobile Comput.*, vol. 15, no. 7, pp. 1770–1782, Jul. 2016.
- [45] A. Shokry, M. Elhamshary, and M. Youssef, "DynamicSLAM: Leveraging human anchors for ubiquitous low-overhead indoor localization," *IEEE Trans. Mobile Comput.*, vol. 20, no. 8, pp. 2563–2575, Aug. 2021.

- [46] M. Osman, F. Viset, and M. Kok, "Indoor SLAM using a foot-mounted IMU and the local magnetic field," in *Proc. 25th Int. Conf. Inf. Fusion*, 2022, pp. 1–7.
- [47] V. Indelman, S. Williams, M. Kaess, and F. Dellaert, "Information fusion in navigation systems via factor graph based incremental smoothing," *Robot. Auton. Syst.*, vol. 61, no. 8, pp. 721–738, 2013.
- [48] W. Wen, T. Pfeifer, X. Bai, and L.-T. Hsu, "Factor graph optimization for GNSS/INS integration: A comparison with the extended Kalman filter," *NAVIGATION: J. Inst. Navigation*, vol. 68, no. 2, pp. 315–331, 2021.
- [49] V. Indelman, S. Williams, M. Kaess, and F. Dellaert, "Factor graph based incremental smoothing in inertial navigation systems," in *Proc. 15th Int. Conf. Inf. Fusion*, 2012, pp. 2154–2161.
- [50] D. Wisth, M. Camurri, and M. Fallon, "Robust legged robot state estimation using factor graph optimization," *IEEE Robot. Automat. Lett.*, vol. 4, no. 4, pp. 4507–4514, Oct. 2019.
- [51] W. Wen, X. Bai, Y. C. Kan, and L.-T. Hsu, "Tightly coupled GNSS/INS integration via factor graph and aided by fish-eye camera," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10651–10662, Nov. 2019.
- [52] S. Bai, J. Lai, P. Lyu, B. Ji, B. Wang, and X. Sun, "A novel plug-and-play factor graph method for asynchronous absolute/relative measurements fusion in multisensor positioning," *IEEE Trans. Ind. Electron.*, vol. 70, no. 1, pp. 940–950, Jan. 2023.
- [53] X. Li, Y. Wang, and K. Khoshelham, "Comparative analysis of robust extended Kalman filter and incremental smoothing for UWB/PDR fusion positioning in NLOS environments," *Acta Geodaetica et Geophysica*, vol. 54, pp. 157–179, 2019.
- [54] Y. Wang, Z. Li, J. Gao, and L. Zhao, "Deep neural network-based Wi-Fi/pedestrian dead reckoning indoor positioning system using adaptive robust factor graph model," *IET Radar, Sonar Navigation*, vol. 14, no. 1, pp. 36–47, 2020.
- [55] L. Yang, N. Wu, B. Li, W. Yuan, and L. Hanzo, "Indoor localization based on factor graphs: A unified framework," *IEEE Internet Things J.*, vol. 10, no. 5, pp. 4353–4366, Mar. 2023.
- [56] W. Zhao, A. Goudar, M. Tang, X. Qiao, and A. P. Schoellig, "Uncertainty-aware Gaussian mixture model for UWB time difference of arrival localization in cluttered environments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2023, pp. 5266–5273.
- [57] S. Thrun and M. Montemerlo, "The graph SLAM algorithm with applications to large-scale mapping of urban structures," *Int. J. Robot. Res.*, vol. 25, no. 5/6, pp. 403–429, 2006.
- [58] G. Grisetti, R. Kümmerle, C. Stachniss, and W. Burgard, "A tutorial on graph-based SLAM," *IEEE Intell. Transp. Syst. Mag.*, vol. 2, no. 4, pp. 31–43, winter 2010.
- [59] H. Wang, J. Sun, S. Lu, and S. Wei, "Factor graph aided multiple hypothesis tracking," *Sci. China Inf. Sci.*, vol. 56, pp. 1–6, 2013.
- [60] M. Hsiao and M. Kaess, "MH-iSAM2: Multi-hypothesis iSAM using Bayes tree and hypo-tree," in *Proc. Int. Conf. Robot. Automat.*, 2019, pp. 1274–1280.
- [61] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," Georgia Inst. Technol., Atlanta, GA, USA, Tech. Rep. GT-RIM-CP&R-2012-002, 2012.
- [62] I. Cox and S. Hingorani, "An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 2, pp. 138–150, Feb. 1996.
- [63] F. Dellaert et al., "Factor graphs for robot perception," *Foundations Trends Robot.*, vol. 6, no. 1/2, pp. 1–139, 2017.
- [64] W. Xu and F. Zhang, "FAST-LIO: A fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 3317–3324, Apr. 2021.
- [65] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *Int. J. Robot. Res.*, vol. 31, no. 2, pp. 216–235, 2012.
- [66] M. Mladenov and M. Mock, "A step counter service for Java-enabled devices using a built-in accelerometer," in *Proc. 1st Int. Workshop Context-Aware Middleware Serv.: Affiliated 4th Int. Conf. Commun. Syst. Softw. Middleware*, 2009, pp. 1–5.
- [67] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, 2011, pp. 1–7.
- [68] D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. Autom. Control*, vol. 24, no. 6, pp. 843–854, Dec. 1979.
- [69] J. Xiong, Z. Xiong, Y. Ding, J. W. Cheong, and A. G. Dempster, "Multihypothesis Gaussian belief propagation for radio ranging-based localization and mapping," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022.
- [70] H. Weinberg, "Using the ADXL202 in pedometer and personal navigation applications," in *Analog Devices, One Technology Way*, Norwood, MA, USA, AN-602 Appl. Note, 2002.
- [71] F. Bourgault, A. A. Makarenko, S. B. Williams, B. Grocholsky, and H. F. Durrant-Whyte, "Information based adaptive robotic exploration," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2002, pp. 540–545.
- [72] G. Cowan, *Statistical Data Analysis*. London, U.K.: Oxford University Press, 1998.
- [73] "How to draw a covariance error ellipse?," 2020. Accessed: Jul. 23, 2020. [Online]. Available: <https://users.cs.utah.edu/tch/CS6640F2020/resources/How%20to%20draw%20a%20covariance%20error%20ellipse.pdf>
- [74] R. Liu et al., "Collaborative SLAM based on WiFi fingerprint similarity and motion information," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 1826–1840, Mar. 2020.
- [75] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 7244–7251.



Shiyu Bai (Member, IEEE) was born in Xuzhou, Jiangsu, China. He received the Ph.D. degree in navigation, guidance, and control from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2022.

He is currently a Postdoctoral Fellow with the Department of Aeronautical and Aviation Engineering, Hong Kong Polytechnic University, Hong Kong. His research interests include inertial navigation, multi-sensor fusion, indoor positioning, and vehicular positioning.



Weisong Wen (Member, IEEE) received the B.Eng. degree in mechanical engineering from Beijing Information Science and Technology University (BISTU), Beijing, China, in 2015, and the M.Eng. degree in mechanical engineering from China Agricultural University, Beijing, China, in 2017, and the Ph.D. degree in mechanical engineering from the Hong Kong Polytechnic University (PolyU), in 2020.

He was also a visiting Ph.D. student with the Faculty of Engineering, University of California, Berkeley, CA, USA, in 2018. Since 2021, he has been a Research Assistant Professor with Department of Aeronautical and Aviation Engineering, PolyU, where he joined as an Assistant Professor in 2023. He has authored or coauthored 30 SCI papers and 40 conference papers in the field of GNSS (ION GNSS+) and navigation for robotic systems (IEEE International Conference on Robotics and Automation, IEEE Intelligent Transportation Systems Conference), such as autonomous driving vehicles.

Dr. Wen was the recipient of the innovation award from TechConnect 2021, the Best Presentation Award from the Institute of Navigation (ION) in 2020, and the First Prize in Hong Kong Section in Qianhai-Guangdong-Macao Youth Innovation and Entrepreneurship Competition in 2019 based on his research achievements in 3-D LiDAR aided GNSS positioning for robotics navigation in urban canyons. The developed 3-D LiDAR-aided GNSS positioning method has been reported by top magazines, such as Inside GNSS and has attracted industry recognition with remarkable knowledge transfer.



Li-Ta Hsu (Senior Member, IEEE) received the B.S. and Ph.D. degrees in aeronautics and astronautics from National Cheng Kung University, Tainan City, Taiwan, in 2007 and 2013, respectively.

In 2012, he was a Visiting Scholar with University College London, London, U.K. He served as a Postdoctoral Researcher with the Institute of Industrial Science, University of Tokyo, Tokyo, Japan. He is currently an Assistant Professor with the Department of Aeronau-

tical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong. His research interests include GNSS positioning in challenging environments and localization for pedestrian, autonomous driving vehicle, and unmanned aerial vehicle.



Peiwen Yang received the M.S. degree in electronic and communication engineering from the School of Information and Electronics, Beijing Institute of Technology, Beijing, China, in 2019.

His research interests include aerial vehicle control, computer vision, and robotics.