Factor Graph Optimization-based Indoor Pedestrian SLAM with Probabilistic Exact Activity Loop Closures using Smartphone

Shiyu Bai
Department of Aeronautical and Aviation Engineering
The Hong Kong Polytechnic University
Hong Kong, China
shiyu.bai@polyu.edu.hk

Weisong Wen
Department of Aeronautical and Aviation Engineering
The Hong Kong Polytechnic University
Hong Kong, China
welson.wen@polyu.edu.hk

Li-Ta Hsu
Department of Aeronautical and Aviation Engineering
The Hong Kong Polytechnic University
Hong Kong, China
lt.hsu@polyu.edu.hk

Yue Yu
Department of Land Surveying and Geo-Informatics
The Hong Kong Polytechnic University
Hong Kong, China
yue806.yu@connect.polyu.hk

2023 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN) | 979-8-3503-2011-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/IPIN57070.2023.10332510

***Abstract*—Indoor localization by smartphones has indicated its promising application prospect in daily life. Smartphone-based pedestrian dead reckoning (PDR) is a common method to obtain the locations. However, PDR suffers from position error accumulation. Although radio frequency (RF) and indoor map can be utilized to restrain the error drift, it requires the prior deployment of facilities or information, which is unsuitable for unknown environments. This paper proposes a factor graph optimization (FGO)-based indoor pedestrian simultaneous localization and mapping (SLAM) with probabilistic exact activity loop closures using a smartphone. In this paper, the smartphone built-in inertial measurement unit (IMU) is solely used to achieve SLAM, in which the human turning activity is regarded as the landmark. Repeatedly observed activities are then used to form loop closures to restrain the drift. FGO is first utilized to formulate pedestrian IMU-only SLAM, which achieves better estimation accuracy than the filter-based method. Moreover, multi-hypothesis tracking is employed to deal with ambiguous data association. During the turning, key points are defined and mutually matched to form exact loop closures to improve estimation accuracy. Simulations and experimental tests are both done to evaluate the performance of the proposed method.**

***Index Terms*—indoor localization, simultaneous localization and mapping (SLAM), factor graph optimization (FGO), smartphone, human activity**

I. Introduction

Indoor pedestrian localization has recently received great attention [1], [2]. Accurate localization solutions can considerably promote efficiency and increase the success rate of many applications, such as emergency management, search and rescue, and safety check. Indoor pedestrian localization is mainly based on wearable devices and smartphones. In the former, professional instruments are needed, and it requires that pedestrians place these devices on their bodies and shoes. On the contrary, people are carrying their smartphones almost at any time. Moreover, smartphones are generally integrated with multiple sensors in compact form, which offers a prominent advantage in indoor localization [3].

In indoor scenarios, pedestrian dead reckoning (PDR) [4] using a smartphone built-in inertial measurement unit (IMU) has been widely adopted to achieve localization. The step length can be calculated based on accelerometer data, and the position of the pedestrian can be updated by combining the previous position and displacements in different directions derived from step length. However, PDR is a kind of recursive algorithm via integration, which cannot be used in long-distance applications. Although some prebuilt infrastructures and floorplans can be employed to obtain localization solutions to correct PDR, it heavily relies on prior deployment and information, making the localization function lack scalability. Pedestrian simultaneous localization and mapping (SLAM) is an alternative manner to improve localization accuracy. When a human passes the same position, the loop closure can be formed to restrain the cumulative error. Actually, loop closures in an unknown environment can be easily triggered by humans in some applications, such as search and rescue.

In pedestrian SLAM, human activities (such as turning corners and taking elevators or escalators) could be regarded as landmarks related to specific positions in indoor scenarios. For instance, turning mainly happens in the corner, and the corner's position will be simultaneously calculated. Once the human goes through the same corners, its position could be updated by known landmarks to restrain cumulative errors. A particle filter (PF) is commonly used in pedestrian SLAM to estimate the state. However, filter-based methods only employ the state

at the last instant and current information. Past states cannot be further adjusted, and the inherent error will be introduced into the following estimation, leading to compromised localization results. Moreover, the data association cannot be elaborately done in filter-based methods due to the loss of historical information.

Recently, factor graph optimization (FGO) has become a popular way to achieve state estimation [5], [6]. In FGO, historical information can be reserved and utilized to perform the joint estimation. On the one hand, FGO can restrain non-linear errors in time-series information by iterative optimization. On the other hand, historical states can be re-estimated, which can indirectly improve the estimation of the current state. Meanwhile, the reserve of historical information in state estimation can be fully exploited to construct exact loop closures to improve the estimation accuracy in pedestrian SLAM.

In this paper, FGO-based indoor pedestrian SLAM with probabilistic exact activity loop closures is proposed. FGO is firstly employed to formulate pedestrian SLAM using activity landmarks with only smartphone IMU. The main contributions of this paper are as follows:

1) This paper proposes a graph-based indoor pedestrian SLAM framework that only uses smartphone IMU to achieve localization. Instead of PF, FGO is first used to achieve SLAM that regards human activity (turning corners in this paper) as landmarks, which can achieve better accuracy than the filter-based method.
2) This paper builds a probabilistic exact activity loop closure to restrain the accumulative error. Multi-hypothesis data association is utilized to maintain the trajectories by loop closures with different probabilities. Key location points during the turning are defined and selected to form exact loop closures.
3) Numerical simulation and experimental tests are performed to validate the performance of the proposed method. The results of position and trajectory estimation of the proposed method are both discussed. Meanwhile, the processing time of different methods is also compared.

The rest of the paper is structured as follows: In Section II, the relevant literature is discussed. In Section III, the proposed method is described in detail. In Section IV, simulation and test results are given. Finally, Section V concludes the paper.

II. Related Work

To restrain the position error accumulations in PDR, deep learning-based methods have been used in indoor pedestrian localization. Herath proposed data-driven inertial navigation [7]; more accurate horizontal positions and heading directions are regressed from a sequence of IMU sensor measurements. Liu introduced a network to regress 3D displacements and their uncertainty. Then a stochastic cloning Extended Kalman Filter (EKF) was used to estimate the pedestrian pose [8]. Although learning-based methods regress more accurate displacements than traditional PDR, the error cumulation still exists as there is no absolute constraint.

One way to provide the constraint is based on the prebuilt infrastructures, such as Wi-Fi [9], Ultra-Wideband (UWB) [10], and Bluetooth [11]. However, device deployment consumes huge amounts of labor and financial resource. Another method is to utilize an indoor map to achieve the correction. In most cases, the activities that a person performs can be related to specific locations [12]. For example, the activity of turning a corner or taking an escalator mainly happens when the pedestrian is at the corner or escalator. When these location-related activities are recognized, corresponding location information can be used to eliminate the accumulated errors. Zhou proposed an activity sequence-based indoor localization, the users activities are firstly detected by smartphone sensors, and then the activity sequence can be mapped into the indoor road network to obtain location information [13]. Gu proposed a context-based map matching for an indoor scenario in which IMU and barometer are combined to recognize the activity of turning, shopping, taking an escalator, and so on [14]. More location-related activities can be detected to restrain error accumulation. However, this type of method requires a prior map to perform matching, which cannot be applied in an environment without prior information.

Recently, SLAM has been widely used in the robotics community [15]–[17]. SLAM is commonly applied to locate the object in previously unknown scenarios as the features of surrounding environments are estimated simultaneously. Some researchers have developed pedestrian SLAM in which the activity is regarded as a landmark, and a local landmark map of the users environment can be built. Hardegger proposed ActionSLAM, in which the PDR and observation of activities are combined in a particle filter framework to perform location tracking [18]. Hardegger presented 3D ActionSLAM, in which an improved motion model is introduced, and systematic evaluations of the systems boundary operating conditions are provided [19].

At present, traditional pedestrian SLAM using smartphone are mainly formed via particle filter. Filter-based methods can obtain rapid estimation by marginalizing out past states and only estimating current state. However, this manner cannot perform a proper re-linearization on past states as they are not included in the state vector [20], leading to a compromised accuracy. More importantly, the inherent position errors of known landmarks will be introduced to the following estimation as the historical states cannot be iteratively optimized, which can also reduce the localization performance.

III. Factor Graph Optimization-based SLAM with Probabilistic Exact Activity Loop Closures

The overall structure of the proposed method is shown in Fig. 1. Smartphone IMU is employed to achieve the pedestrian SLAM with probabilistic exact activity loop closures. First of all, IMU data is utilized to perform PDR. The step length can be derived from acceleration data. According to the Weinberg model [21], the step length L_k can be estimated by

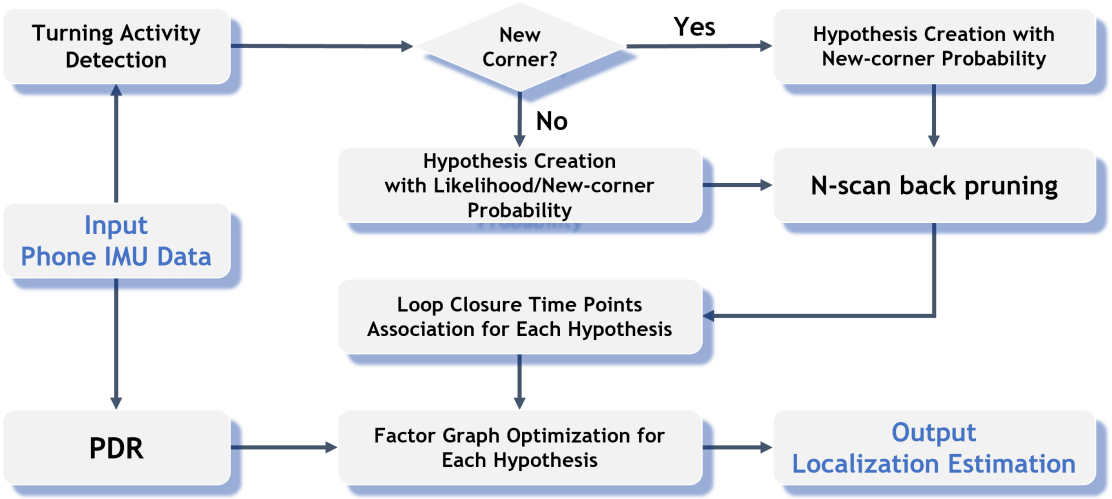

Fig. 1. The structure of the proposed method

$$L_k = K \sqrt[4]{a_{v,max} - a_{v,min}} \tag{1}$$

where $a_{v,max}$ and $a_{v,min}$ denote the maximum and minimum values of vertical acceleration in each stride, and K is the constant. The horizontal positions are updated by

$$\begin{aligned} p_{x,k} &= p_{x,k-1} + L_k \cos\theta_k \\ p_{y,k} &= p_{y,k-1} + L_k \sin\theta_k \end{aligned} \tag{2}$$

where $p_{x,i}$ and $p_{y,i}$ denote the horizontal positions in different directions at instant t_i. L_k stands for the step length, and θ_k is the heading provided by the orientation from the integration of 9-axis IMU measurements.

In the proposed method, turning activities are regarded as landmarks. The turning activity is detected through gyroscope measurements, and the location point with the maximal turning rate at instant t_o during the change of heading is labeled as a landmark $\left[p^m_{x,o}, p^m_{y,o}\right]^{\mathrm{T}}$, and m denotes the serial number of this landmark.

Whether the detected landmark is located in a new corner or not needs to be determined to construct the loop closure. In the proposed method, the covariance matrix of pedestrian position is utilized. According to the covariance matrix of position, the uncertainty ellipsoid can be obtained, and landmarks inside the ellipsoid are regarded as candidate-known activities to achieve loop closures. Otherwise, it is an unknown feature.

The propagation equation of covariance matrix is shown as

$$\mathbf{\Sigma}_o = \mathbf{F}_o \mathbf{\Sigma}_{o-1} \mathbf{F}_o^{\mathrm{T}} + \mathbf{G}_o \mathbf{Q} \mathbf{G}_o^{\mathrm{T}} \tag{3}$$

where $\mathbf{F}_o$ is the state transition matrix that can be expressed by a two-dimensional identity matrix. $\mathbf{Q}$ stands for the covariance matrix of noise and $\mathbf{G}_o$ denotes the noise matrix that is expressed as

$$\mathbf{G}_o = \begin{bmatrix} \cos\theta_o & -L_o \sin\theta_o \\ \sin\theta_o & -L_o \cos\theta_o \end{bmatrix} \tag{4}$$

It can be noticed from equation (3) that the uncertainty of the position is growing with time. Thus, the covariance matrix will be updated after each optimization.

To improve the accuracy of SLAM, multiple hypotheses with different loop closures are formulated. The probabilities of different hypotheses need to be obtained. Based on Multi-hypothesis tracking theory [22], [23], the probability of the j-th hypothesis Ω^j_o at instant t_o is expressed as

$$p\left(\Omega^j_o\right) = \eta \left(\mathcal{N}\left(\mathbf{z}_o\right) p^{det}_o\right)^{\delta} \cdot \lambda^{1-\delta}_{new} \cdot p\left(\Omega^j_{o-1}\right) \tag{5}$$

where η denotes the normalizer, $\mathcal{N}\left(\mathbf{z}_o\right)$ and p^{det}_o signify the measurement likelihood and the inverse of the number of alternative activities. λ_{new} expresses the average rate of turning activity. When a known landmark is found, δ is 1. Otherwise, it is 0. In equation (5), the measurement likelihood can be expressed as

$$\mathcal{N}\left(\mathbf{z}_o\right) = \mathcal{N}\left(\mathbf{z}_o, \hat{\mathbf{z}}_o, \mathbf{U}_o\right) \tag{6}$$

where $\mathbf{z}_o = \left[z_{x,o}, z_{y,o}\right]^{\mathrm{T}}$ is the position difference measurement. In view of landmarks in the proposed method are actually in the pedestrian trajectory, $z_{x,o}$ and $z_{y,o}$ are both zero. $\hat{\mathbf{z}}_o$ is the predicted position difference, which can be expressed as

$$\begin{aligned} \hat{z}_{x,o} &= p^m_{x,o} - p_{x,o} \\ \hat{z}_{y,o} &= p^m_{y,o} - p_{y,o} \end{aligned} \tag{7}$$

In equation (6), $\mathbf{U}_o$ is the covariance matrix of $\hat{\mathbf{z}}_o$, which can be calculated as

$$\mathbf{U}_o = \mathbf{H}_o \mathbf{\Sigma}_o \mathbf{H}_o^{\mathrm{T}} + \mathbf{H}^m_o \mathbf{\Sigma}^m_o \mathbf{H}^{m\mathrm{T}}_o + \mathbf{\Sigma}_{z,o} \tag{8}$$

where $\mathbf{H}_o$ stands for the Jacobian matrix of equation (7) with respect to $p_{x,o}$ and $p_{y,o}$. $\mathbf{H}^m_o$ denotes the Jacobian matrix with respect to $p^m_{x,o}$ and $p^m_{y,o}$. $\mathbf{\Sigma}^m_o$ is the covariance matrix of m-th landmark that is set based on the covariance matrix of position at the m-th landmark. $\mathbf{\Sigma}_{z,o}$ denotes the covariance matrix of measurement $\mathbf{z}_o$. According to equation (5), the probability of

each hypothesis with different loop closures could be obtained. It shows that the number of hypotheses grows exponentially with time. Each time a new hypothesis is created, a new factor graph is formulated, leading to a huge computational burden. Therefore, the N-scan back pruning is performed, and N is set as 4 in this paper.

Although the landmark in every corner is single point, key location points around the landmark are all used to build loop closures to add constraints. In the proposed method, there are three key location points in the turning activity, enter point t_{en}, middle point t_{mi}, and exit point t_{ex}. The vertical angular rate during the turning is shown in Fig. 2. Therefore, the instants

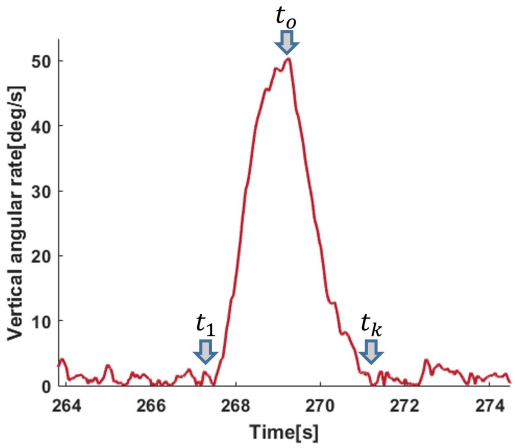

Fig. 2. The vertical angular rate during the turning

of key location points can be expressed as

$$t_{en} = t_1, t_{mi} = t_o, t_{ex} = t_k \tag{9}$$

The local trajectory in different situations is shown in Fig. 3. The red and blue lines stand for the motion trajectories when the pedestrian passes the same corner, which can be divided into two categories. In the first one, the red and blue lines nearly overlap each other, as shown in (a) and (b). In the second one, the red and blue lines do not overlap, as shown in (c) to (f). In the first category, enter point, middle point, and exit point are all utilized to construct the loop closures. According to the heading measurements, corresponding points are selected and matched. For example, the middle points in the red and blue lines are matched in both (a) and (b). In (a), however, the enter point in the red line is matched with the exit point in the blue line. In the second category, loop closures can also be constructed, but only a single point is used to achieve the loop closure due to the mismatching trajectories.

The graph structure of the proposed method is indicated in Fig. 4. The blue circle with $\mathbf{X}_n$ stands for the variable node that includes the state of the pedestrian. The green circle stands for the PDR factor that constrains two consecutive variable nodes in the factor graph. The purple circle represents the activity loop closure to associate two variable nodes at different instants with restraining the accumulative position

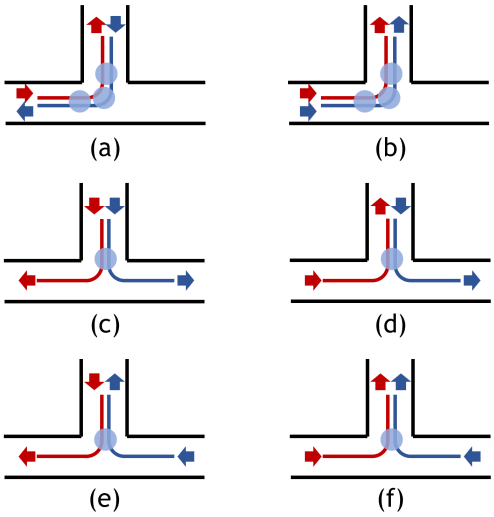

Fig. 3. The local trajectory in different situations

errors. The triangle denotes the key points during the turning-corner activities.

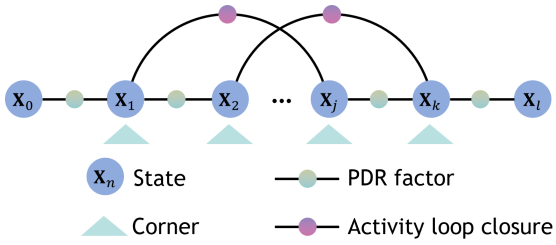

Fig. 4. Graph structure of the proposed method

The state vector in the proposed method is given by

$$\begin{aligned} \mathbf{X}_n &= [\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_n] \\ \mathbf{x}_k &= [p_{x,k}, p_{y,k}, \theta_k] \end{aligned} \tag{10}$$

where $\mathbf{X}_n$ stands for the states set, and $\mathbf{x}_k$ denotes the states at instant t_k that includes horizontal positions in two directions and heading. $n+1$ stands for the number of states in the set. Considering that landmarks are in pedestrian trajectory, the landmark positions are the specific points in the pedestrian position; thus, they are not included in the state vector.

The cost fuction of the proposed method can be described as

$$\mathbf{X}^* = argmin \left(\sum_{i=1}^{n} \left\| \mathbf{r}_p \left(\hat{\mathbf{z}}_p, \mathbf{X} \right) \right\|^2_{\mathbf{\Sigma}_p} + \sum_{i=1}^{m} \left\| \mathbf{r}_c \left(\hat{\mathbf{z}}_c, \mathbf{X} \right) \right\|^2_{\mathbf{\Sigma}_c} \right) \tag{11}$$

where $\mathbf{X}^*$ denotes the optimal state set, $\|\mathbf{r}\|^2_{\mathbf{\Sigma}}$ is the Mahalanobis norm that equals $\mathbf{r}^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{r}$, $\mathbf{r}$ denotes the residual, and $\mathbf{\Sigma}$ is the covariance matrix. m represents the number of activity loop closures. $\mathbf{r}_p\left(\hat{\mathbf{z}}_p, \mathbf{X}\right)$ and $\mathbf{r}_c\left(\hat{\mathbf{z}}_c, \mathbf{X}\right)$ stand for the residual

of PDR and loop closures. In this paper, iSAM2 [24] and GTSAM [25] are used to solve the optimization problem to obtain the real-time performance.

The residual of PDR can be formulated based on equation (2) as

$$\mathbf{r}_p\left(\hat{\mathbf{z}}_p, \mathbf{X}\right) = \begin{bmatrix} L_k \cos\theta_k - p_{x,k} + p_{x,k-1} \\ L_k \sin\theta_k - p_{y,k} + p_{y,k-1} \\ \Delta\theta_{kk-1} - \theta_k + \theta_{k-1} \end{bmatrix} \tag{12}$$

The residual of activity loop closure can be formulated as

$$\mathbf{r}_c\left(\hat{\mathbf{z}}_c, \mathbf{X}\right) = \begin{bmatrix} -p_{x,i} + p_{x,j} \\ -p_{y,i} + p_{y,j} \\ \Delta\theta_{ij} - \theta_i + \theta_j \end{bmatrix} \tag{13}$$

where $p_{x,i}$, $p_{y,i}$ and θ_i stand for the pose at instant t_i, $p_{x,j}$, $p_{y,j}$ and θ_j represent the pose at instant t_j. These two instants are associated by the loop closure.

IV. Simulations and Experimental Tests

Simulations and experimental tests are both carried out to validate the performance of the proposed method. During the evaluation, the following three pipelines are employed for the comparison.

1) **S-PDR** [4]: which tracks pedestrians via a typical dead reckoning using data from inertial sensors embedded in smartphones. This is to express the localization accuracy of the dead reckoning using a smartphone.
2) **A-SLAM** [18]: which tracks pedestrians by utilizing PDR with the observations of activities to build and update a local landmark map of the environments. This is to show the performance of traditional SLAM based on particle filters.
3) **C-SLAM** [26]: which tracks pedestrians by utilizing PDR, radio frequency (RF), and turning activities. RF signal is excluded in the comparison. Iterative closest point (ICP) is used to perform the turning matching. This is to express the performance of traditional FGO-based activity loop closures.

In addition, the proposed methods based on position and trajectory estimation are conducted respectively for the comparison, denoted as Proposed-ISE and Proposed-BE. An optimized trajectory using all information is given in the latter, while instant position estimation is presented in the former.

A. Simulations

Numerical simulation is first carried out to compare the performance of different methods. The displacements are simulated as step length with a noise of 0.3 m, and a noise of heading is set as 1 deg. The pedestrian goes through nine loop closure corners as shown with green circle in Fig. 5.

The pedestrian trajectory comparison of different methods is indicated in Fig. 6. GT is the ground truth of the pedestrian trajectory. It can be noticed from Fig. 6 that Proposed-ISE and Proposed-BE are closer to the ground truth. The horizontal position error comparison in the simulation, the absolute position error (APE), and the relative position error (RPE) [27]

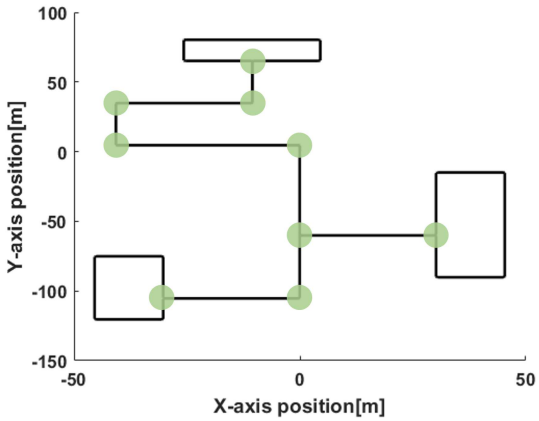

Fig. 5. The pedestrian trajectory in simulation

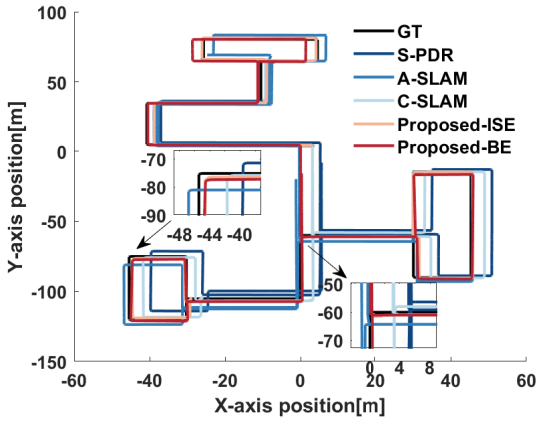

Fig. 6. The pedestrian trajectory comparison in the simulation

comparison are shown in Fig. 7 and TABLE I, respectively. It indicates that S-PDR presents a fast divergence compared to other methods as the recursive algorithm is used to derive the pedestrian position, and there is no absolute position constraint. In A-SLAM, turning activities are regarded as landmarks, and the pedestrian positions can be updated every time a known landmark is recognized. However, A-SLAM marginalizes past states and only estimates states at the current instant, which can decrease the estimation accuracy for the current state. Furthermore, the turning activity is a continuous process, and the marginalization of past states makes it difficult to determine the appropriate time point for the update, which can also lead to a compromised estimation accuracy. In C-SLAM, only one hypothesis is determined every time a turning behavior is detected, which can easily lead to the wrong data association. Therefore, the loop closure will not be built if there are ambiguous data associations. Meanwhile, if the local trajectory is just like (c) to (f) in Fig. 3, the data association cannot be determined by ICP, leading to reduced localization accuracy.

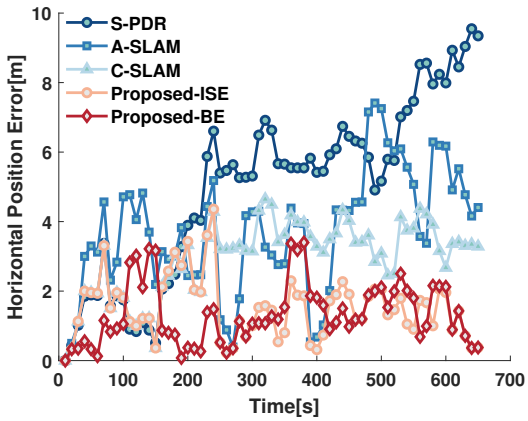

Fig. 7. The horizontal position error comparison in the simulation

TABLE I
THE HORIZONTAL POSITION APE AND RPE COMPARISON IN THE SIMULATION

	APE (m)	RPE (m)
S-PDR	4.00	2.11
A-SLAM	2.93	2.65
C-SLAM	2.22	1.41
Proposed-ISE	1.24	1.39
Proposed-BE	1.13	1.25

B. Experimental Tests

Experimental tests are carried out further to evaluate the localization performance of the proposed method. HUAWEI P40 Pro is used to collect IMU data that includes angular rate, acceleration, and orientation. A handheld LiDAR-inertial navigation system based on Fast-lio [28] is used as the ground truth. The test is carried out in the underground garage of The Hong Kong Polytechnic University. In the test, the smartphone was hold by hand in front of the body. The pedestrian goes through seven loop closure corners as shown with green circle in Fig. 8.

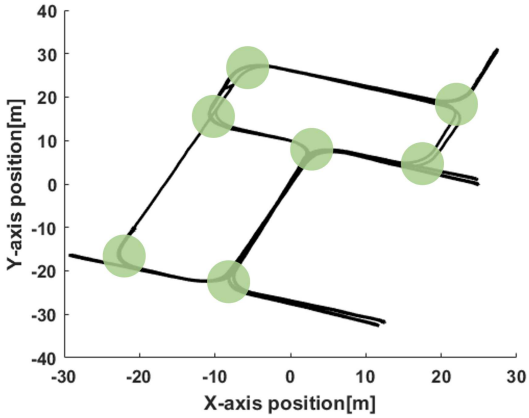

Fig. 8. The pedestrian trajectory comparison in the underground garage

The pedestrian trajectory comparison, the horizontal position error comparison, the horizontal position APE, and RPE comparison in the experimental test are indicated in Fig. 9, Fig. 10, and TABLE II. It could be noticed that the localization results of other methods all present divergent trends. Proposed-ISE and Proposed-BE can achieve better localization accuracy compared with current methods. Proposed-BE can outperform Proposed-ISE as more information can be utilized in nonlinear optimization, leading to better trajectory estimation.

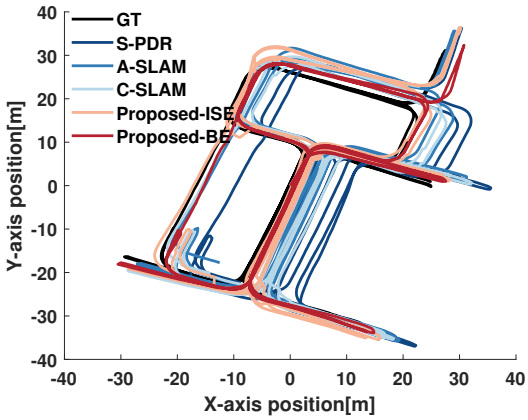

Fig. 9. The pedestrian trajectory comparison in the underground garage

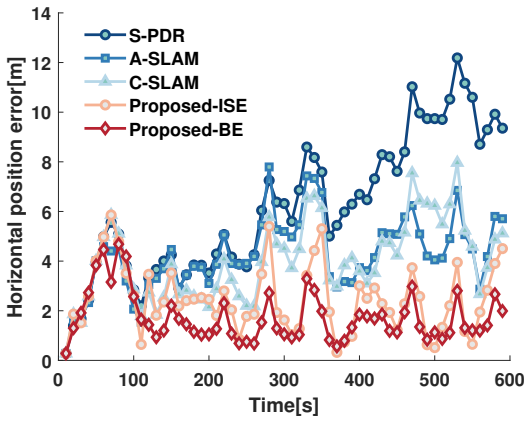

Fig. 10. The horizontal position error comparison in the underground garage

TABLE II
THE HORIZONTAL POSITION APE AND RPE COMPARISON IN THE UNDERGROUND GARAGE

	APE (m)	RPE (m)
S-PDR	4.81	2.30
A-SLAM	3.21	2.11
C-SLAM	3.04	2.27
Proposed-ISE	1.97	1.93
Proposed-BE	1.38	1.30

Considering that there are multiple factor graphs in the proposed method, its processing time needs to be discussed.

The N-scan back pruning is utilized to avoid the unlimited growth of the hypotheses. The processing time of A-SLAM and Proposed-ISE is given in Fig. 11. A-SLAM is based on particle filter that involves 100 particles in the test. Proposed-ISE is based on the iSAM2 framework and all past information respect to any instant is used. Although only the state at the current instant needs to be estimated in filter-based methods, it can introduce extra computational load when processing 100 particles. It shows that Proposed-ISE can achieve a faster processing speed than A-SLAM as long as the hypotheses number is maintained. From 360 s to 480 s, the slope of processing time from Proposed-ISE increases as the hypotheses number rises during this period.

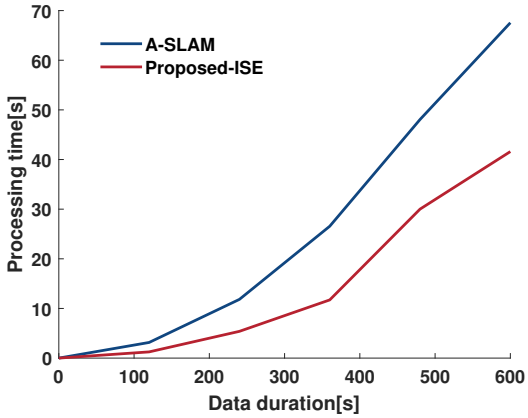

Fig. 11. The processing time of different methods

V. CONCLUSION

This paper proposes an FGO-based indoor pedestrian SLAM with probabilistic exact activity loop closures. Instead of PF, FGO is firstly used to form pedestrian SLAM using activity landmarks with only smartphone IMU. First, Proposed-ISE can achieve better estimation performance for instant localization as more information is used than filter-based methods. Second, a more accurate trajectory can be obtained by Proposed-BE. In this paper, the key points during the turning are defined and selected to form exact loop closures. Furthermore, multiple hypotheses based on different loop closures are formed to improve the performance of SLAM. Simulations and experimental tests are performed to evaluate the localization accuracy of the proposed method. Compared to traditional methods, the proposed method can effectively restrain the position error accumulation of smartphone-based PDR with only IMU data. Furthermore, proposed method has a faster processing speed than the particle filter, even though there are multiple factor graphs in the proposed method.

ACKNOWLEDGMENT

This work was supported by the Guangdong Basic and Applied Basic Research Foundation (2021A1515110771) and University Grants Committee of Hong Kong under the scheme Research Impact Fund (R5009-21). This research was also supported by the Faculty of Engineering, The Hong Kong Polytechnic University under the project “Perception-based GNSS PPP-RTK/LVINS integrated navigation system for unmanned autonomous systems operating in urban canyons.”.

REFERENCES

[1] K. Nguyen-Huu, K. Lee, and S.-W. Lee, “An indoor positioning system using pedestrian dead reckoning with WiFi and map-matching aided,” in *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2017, pp. 1–8.

[2] Y. Wang, J. Kuang, Y. Li, and X. Niu, “Magnetic field-enhanced learning-based inertial odometry for indoor pedestrian,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022.

[3] Q. Wang, M. Fu, J. Wang, H. Luo, L. Sun, Z. Ma, W. Li, C. Zhang, R. Huang, X. Li *et al.*, “Recent advances in pedestrian inertial navigation based on smartphone: A review,” *IEEE Sensors Journal*, 2022.

[4] W. Kang and Y. Han, “SmartPDR: Smartphone-based pedestrian dead reckoning for indoor localization,” *IEEE Sensors Journal*, vol. 15, no. 5, pp. 2906–2916, 2014.

[5] V. Indelman, S. Williams, M. Kaess, and F. Dellaert, “Factor graph based incremental smoothing in inertial navigation systems,” in *2012 15th International Conference on Information Fusion*. IEEE, 2012, pp. 2154–2161.

[6] 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 Transactions on Industrial Electronics*, vol. 70, no. 1, pp. 940–950, 2022.

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

[8] W. Liu, D. Caruso, E. Ilg, J. Dong, A. I. Mourikis, K. Daniilidis, V. Kumar, and J. Engel, “Tlio: Tight learned inertial odometry,” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5653–5660, 2020.

[9] R. C. Luo and T.-J. Hsiao, “Indoor localization system based on hybrid Wi-Fi/BLE and hierarchical topological fingerprinting approach,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10 791–10 806, 2019.

[10] G. T. Lee, S. B. Seo, and W. S. Jeon, “Indoor localization by Kalman filter based combining of UWB-positioning and PDR,” in *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2021, pp. 1–6.

[11] J. Chen, B. Zhou, S. Bao, X. Liu, Z. Gu, L. Li, Y. Zhao, J. Zhu, and Q. Li, “A data-driven inertial navigation/Bluetooth fusion algorithm for indoor localization,” *IEEE Sensors Journal*, vol. 22, no. 6, pp. 5288–5301, 2021.

[12] S. Khalifa and M. Hassan, “Evaluating mismatch probability of activity-based map matching in indoor positioning,” in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2012, pp. 1–9.

[13] B. Zhou, Q. Li, Q. Mao, W. Tu, and X. Zhang, “Activity sequence-based indoor pedestrian localization using smartphones,” *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 5, pp. 562–574, 2014.

[14] Y. Gu, D. Li, Y. Kamiya, and S. Kamijo, “Integration of positioning and activity context information for lifelog in urban city area,” *NAVIGATION: Journal of the Institute of Navigation*, vol. 67, no. 1, pp. 163–179, 2020.

[15] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, “Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age,” *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[16] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, “ORB-SLAM: a versatile and accurate monocular SLAM system,” *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

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

[18] 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 *International Conference on Indoor Positioning and Indoor Navigation*. IEEE, 2013, pp. 1–8.

[19] M. Hardegger, D. Roggen, and G. Tröster, "3D ActionSLAM: wearable person tracking in multi-floor environments," *Personal and Ubiquitous Computing*, vol. 19, pp. 123–141, 2015.
[20] V. Indelman, S. Williams, M. Kaess, and F. Dellaert, "Information fusion in navigation systems via factor graph based incremental smoothing," *Robotics and Autonomous Systems*, vol. 61, no. 8, pp. 721–738, 2013.
[21] H. Weinberg, "Using the ADXL202 in pedometer and personal navigation applications," *Analog Devices AN-602 application note*, vol. 2, no. 2, pp. 1–6, 2002.
[22] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979.
[23] S. Grzonka, A. Karwath, F. Dijoux, and W. Burgard, "Activity-based estimation of human trajectories," *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 234–245, 2011.
[24] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the bayes tree," *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012.
[25] F. Dellaert, "Factor graphs and GTSAM: A hands-on introduction," Georgia Institute of Technology, Tech. Rep., 2012.
[26] R. Liu, S. H. Marakkalage, M. Padmal, T. Shaganan, C. Yuen, Y. L. Guan, and U.-X. Tan, "Collaborative SLAM based on WiFi fingerprint similarity and motion information," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1826–1840, 2019.
[27] Z. Zhang and D. Scaramuzza, "A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 7244–7251.
[28] W. Xu and F. Zhang, "Fast-lio: A fast, robust LiDAR-inertial odometry package by tightly-coupled iterated Kalman filter," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3317–3324, 2021.