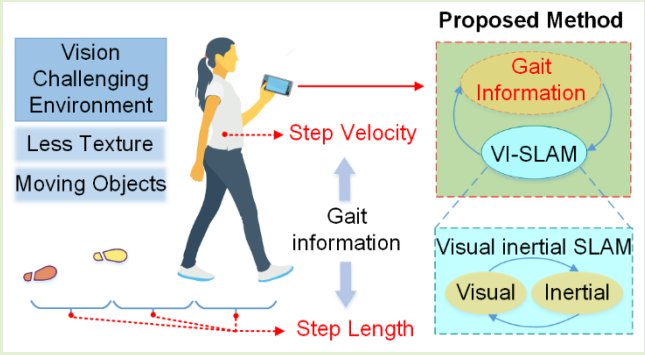
Pedestrian Gait Information Aided Visual Inertial SLAM for Indoor Positioning Using Handheld Smartphones

Yitong Dong, Dayu Yan, Tuan Li, Ming Xia, and Chuang Shi

***Abstract*—Simultaneous localization and mapping (SLAM) is currently a widely used technology for indoor positioning. Many studies have focused on the smartphone-based localization and navigation for its portability and feature-rich sensors. But due to the low-quality sensors and complex environments, current SLAM-based positioning technology using smartphones still poses great challenges, such as inherent cumulative global drift and potential divergence facing texture-less indoor regions. For the regular gait characteristics of pedestrians when naturally walking, the gait motion model can certainly provide effective observation of the pedestrian motion state. We propose a visual-inertial odometry (VIO) assisted by pedestrian gait information for smartphone-based indoor positioning. This work mainly builds two additional state constraints, pedestrian velocity, and step displacement, obtained by the pedestrian dead reckoning (PDR) algorithm for the visual-inertial tracking system. For each step, the corresponding residual term of step length and velocity constraints is constructed and added to the cost function for nonlinear sliding-window optimization. Furthermore, the step displacement is applied again in the four-degree-of-freedom (4-DOF) graph-based optimization to refine the trajectory. VIO system also assists the PDR algorithm in mode switching, to improve the accuracy of gait information by applying the adaptive step length formula. Field experiments were conducted, and the results indicate that with the aiding of the pedestrian gait information, the accuracy and robustness of the visual-inertial pedestrian tracking system using smartphones have been significantly improved. Compared with the state-of-the-art algorithm monocular visual-inertial navigation system (VINS-MONO), our method improves the accuracy by 54.6% on average in our field tests in challenging environments.**



***Index Terms*—Indoor positioning, pedestrian dead reckoning (PDR), smartphones, visual inertial simultaneous localization and mapping (VI-SLAM).**

I. Introduction

OUTDOOR positioning technology has developed rapidly with the fast-growing demand for location-based services (LBS). Positioning and navigation services based on satellite navigation have been widely used in vehicle navigation, pedestrian guidance, and other fields. The deployment of the global navigation satellite system (GNSS) can provide accurate and reliable location services for pedestrians in open-sky environments. However, GNSS positioning capability degrades in harsh environments due to signal attenuation, reflection, and blockage. Therefore, GNSS positioning is not available for indoor positioning and other indoor positioning technology should be explored.

Simultaneous localization and mapping (SLAM) is a very effective system for indoor positioning and has been one of the most active research subjects. SLAM systems have been widely deployed in autonomous vehicles [1], indoor robots [2], and augmented reality (AR) [3]. SLAM can obtain a global and consistent pose estimation of mobile devices, commonly automatic guided vehicles or drones while reconstructing a map of the surrounding environments [4].

Approaches that use only cameras have gained significant interest in the field due to their small size, low

Manuscript received 29 May 2022; revised 14 August 2022; accepted 15 August 2022. Date of publication 7 September 2022; date of current version 14 October 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFC1512003, in part by the Joint Foundation for Ministry of Education of China under Grant 6141A02011907, and in part by the National Natural Science Foundation of China under Grant 61827901. The associate editor coordinating the review of this article and approving it for publication was Prof. Meribout Mahmoud. *(Corresponding author: Tuan Li.)*

Yitong Dong, Dayu Yan, Ming Xia, and Chuang Shi are with the School of Electronic Information Engineering, Beihang University, Beijing 100083, China (e-mail: ytdong@buaa.edu.cn; dyaxb@buaa.edu.cn; xiaming@buaa.edu.cn; shichuang@buaa.edu.cn).

Tuan Li is with the Advanced Research Institute of Multidisciplinary Sciences, Beijing Institute of Technology, Beijing 100081, China (e-mail: tuanli@whu.edu.cn).

Digital Object Identifier 10.1109/JSEN.2022.3203319

cost, and easy hardware setup [5], [6], [7]. However, visual SLAM (V-SLAM) is incapable of recovering the metric scale, therefore, limiting its usage in real-world robotic applications. Visual inertial SLAM (VI-SLAM) assists the vision system with an inertial measurement unit (IMU) to observe the metric scale, as well as roll and pitch angles. The integration of IMU measurements can dramatically improve the motion-tracking performance by bridging the gap between losses of visual tracks due to illumination change, texture-less area, or motion blur.

Although some VI-SLAM systems perform satisfactorily in most environments, their robustness in some scenes is still a challenge. As a mostly-used consumer positioning device for pedestrians, the smartphone with low-quality sensors has worse positioning capability when applying SLAM-based algorithms. However, the regular pedestrian gait information can be combined with the surrounding environment features to improve the accuracy of smartphone-based indoor navigation. Based on the pedestrian gait model and some efficient simplifying assumptions, pedestrian dead reckoning (PDR) technologies have been widely investigated, which provide the heading and step displacement estimations during walking.

Section II conducts state-of-the-art-related works on pose estimation and SLAM indoors. Section III presents the coupling process between PDR and SLAM, i.e., the pose estimation based on step length estimation, and the position estimation using a hand-held model of the equipment. Section IV briefly describes the experiments conducted and the results of the experiments, and the evaluation of the proposed method using smartphones in detail. Finally, we conclude that human motion analysis can be used as additional clues and constraints to extend the monocular visual-inertial odometry (VIO) and improve the accuracy and robustness of smartphone-based VI-SLAM.

The main contributions of this article are listed as follows.

1) A VI-SLAM system assisted by gait information is proposed for pedestrian indoor positioning. We perform the PDR algorithm to obtain the step velocity and step length information. The step velocity information is utilized to construct the residual constraint term and to perform nonlinear optimization. By introducing step velocity information, the accuracy can be improved by 16.3% on average.
2) We apply the pedestrian step length obtained by the PDR algorithm as additional information to construct a novel four-degree-of-freedom (4-DOF) local optimization. Experimental results indicate that the accuracy of position can be improved by 22.3% on average by introducing step length information.
3) We use the heading information obtained from the VIO system to determine the pedestrians' turning events. The recognition of pedestrian turning can assist in the correction of the PDR algorithm to obtain more accurate gait information. By using the step length and velocity information obtained by the adaptive PDR algorithm, more accurate positioning can be achieved. Our approach can improve the accuracy by 54.6% on average.

II. Related Work

VIO has been extensively studied in the past decades and has achieved lots of achievements. VIO can be divided into two categories: 1) extended Kalman filter (EKF)-based VIO and 2) optimization-based VIO. Mourikis proposed the well-known multi-state constraint Kalman filter (MSCKF) method in 2007 [8], [9]. The MSCKF maintained the previous camera poses in the state vector, which used IMU measurements to predict and used visual measurements of the same feature across multiple camera views to form a multiconstraint update. Robust VIO (ROVIO) was a monocular VIO proposed by Bloesch [10]. ROVIO employed an iterated EKF to fuse IMU data and images. None closed-loop and mapping were included, so positioning errors would accumulate unbounded. Schneider *et al.* [11] proposed Maplab, a research-oriented visual-inertial mapping and localization framework processing and manipulating multisession maps. Open Keyframe-based Visual-Inertial SLAM (OKVIS) [12] was a VIO based on keyframe optimization, which optimized the state of vision and IMU together. Keyframes were selected according to spacing rather than considering time-successive poses. The monocular visual-inertial navigation system (VINS-MONO) [13] proposed by Qin was a VI-SLAM based on a tightly-coupled optimization framework that operated with a sliding window. The system had both loop detection and relocalization mechanisms. The VINS-Fusion [14] estimated the state of a robot equipped with an IMU, stereo camera, and global position system (GPS) information which was added compared with VINS-MONO. The satellite positioning information was applied as a constraint to the VI-SLAM to improve the accuracy and robustness of the VINS system. Oriented FAST and rotated BRIEF (ORB)-SLAM was a V-SLAM system, which used ORB features for tracking [15]. In 2020, ORB-SLAM extended to VI-SLAM by adding IMU information, namely ORB-SLAM3 [16], [17]. The system fused visual information and inertial navigation information through a nonlinear optimization method. Basalt VIO [18] provided a globally consistent mapping method using nonlinear optimization. This method had a custom feature tracking front end and smoothed out the mapping information for stereo keyframes.

With the wide use of mobile phones, smartphone-based positioning, and navigation technology have received increasing attention. VINS-Mobile [19] was a VI-SLAM system based on smartphones, which constructed sparse mapping while positioning. Schöps *et al.* [20] transplanted large-scale direct (LSD)-SLAM onto a mobile phone and fit a rough 3-D mesh of a scene to detect physical collisions between the virtual object and the real scene in an AR application. Ondrúška *et al.* [21] proposed MobileFusion, which tried to perform 6 DOF odometry and reconstructed dense surfaces on mobile phones with a monocular camera. However, the positioning performance still suffered from mobile phone hardware devices and the vision-challenging environment. A smartphone's camera requires a longer exposure time than a traditional camera and is thus sensitive to hand tremors. Therefore, the indoor positioning based on smartphones has relatively poor performance, which needs additional constraints to improve robustness and accuracy.

Pedestrian motion has obvious regularity in the process of walking, so the PDR algorithm is proposed to determine the position of pedestrians using inertial sensors. Levi and Judd [22] proposed the PDR algorithm to realize pedestrian positioning, which separated the pedestrian navigation algorithm from the traditional strap-down inertial navigation. The key modules of the PDR algorithm mainly include gait detection, step length estimation, and position update. The position of pedestrians was updated by step events detection, and the pedestrians' step frequency was determined by detecting the periodicity of the acceleration signal. The methods of step events detection include peak detection [23], [24] and zero crossing counting [25]. The location was updated by the last position with the step length obtained by the correlation model and heading obtained from the gyroscope. To decrease the drift of estimation error and get more precise step and heading information from the inertial data, the zero velocity update (ZUPT) model [26], [27], zero angular rate update (ZARU) [28], and heuristic drift elimination (HDE) [29] were proposed. It meant that movement velocity should be zero when the pedestrian's foot touched the ground, and ZUPT could reduce velocity error and improve the accuracy of position estimation. Foxlin [26] and Beauregard [30] proposed resetting the velocity error of the phase when detecting zero velocity for each step. Ojeda and Borenstein [31] applied ZUPT to observation and fed it to the EKF for tracking error correction. In the case of a foot-mounted PDR system, the PDR tracking method avoided the accumulated error caused by the quadratic integration of acceleration. Yan *et al.* [32] recognized the pedestrian walking states, walking straight and turning, by monitoring the angular velocity, and then heuristic heading drift elimination could be achieved. A generalized movement classifier for PDR applications has been proposed and movement segmentation and classification routines have been performed [33]. Martinelli *et al.* [34] introduced a weighted context-based step length estimation algorithm for PDR and six pedestrian contexts are considered. Kunze *et al.* [35] proposed the idea and theoretical analysis that the principal component analysis (PCA) technology could be applied to heading determination, and its effect was also verified by Jin *et al.* [36] through smartphone experiments. Inspired by these works, we propose a visual inertial SLAM assisted by pedestrian gait information obtained from the PDR algorithm for smartphone-based indoor positioning that can improve the accuracy of VI-SLAM in vision-challenging environments.

III. Methodology

The architecture of the proposed system is illustrated in Fig. 1. For the output obtained from the smartphone sensors, a certain preprocessing is first required. In the traditional PDR method, raw data from micro-electro-mechanical system (MEMS) IMU (gyroscope and accelerometer) are used for step detection, and step length estimation. Through the PDR algorithm, the time interval of each step and the corresponding step length and step velocity information can be obtained. For the SLAM system, optimization-based VIO and 4-DOF local optimization are executed after measurement preprocessing, in which the step length and step velocity information can be used to assist the SLAM system, so as to improve the overall positioning accuracy. Meanwhile, the attitude information obtained by the SLAM system can also help PDR system mode switching to obtain more accurate gait information.

Our method consists of three components. The first part is the enhanced nonlinear optimization-based VIO applying step information obtained from the PDR algorithm. The second part is the additional 4-DOF local pose graph optimization assisted by pedestrian step length. The last part introduces the detection of pedestrian turning events by analyzing the attitude outputs of the VIO system, and the PDR mode is switched accordingly.

A. Basic Principles of VI-SLAM Based on Smartphones

The SLAM system can usually be divided into front-end and back-end parts. The front end mainly makes measurement preprocessing after obtaining data collected by sensors. In this period, existing features are tracked by the Kanade-Lucas-Tomasi (KLT) sparse optical flow algorithm for each image [37]. Meanwhile, new corner features are detected [38] to guarantee there are enough features in each image. The IMU preintegration is processed meanwhile. The back-end is mainly a nonlinear-based VIO that infers the global map through the graph optimization algorithm.

To reduce the amount of calculation, VI-SLAM only maintains the keyframe pose and corresponding feature points which can be observed in the sliding window. The state vector in the sliding window can be mainly expressed as follows:

$$\chi = \left[x_n, x_{n+1}, \ldots, x_{n+N}, \lambda_m, \lambda_{m+1}, \ldots, \lambda_{m+M}\right] \tag{1}$$

where x_i represents the IMU state at the time that the corresponding image is captured in the sliding window. It contains the position, velocity, and orientation of the IMU in the world frame, and acceleration bias and gyroscope bias in the IMU body frame. N is the total number of keyframes; M is the total number of features in the sliding window; λ_i represents the state quantity corresponding to the feature point. The IMU state x_i is shown in the following equation:

$$x_i = \left[p_{wb_i}, q_{wb_i}, v_i^w, b_a^{b_i}, b_g^{b_i}\right]^T, \quad i \in [n, n+N] \tag{2}$$

where p_{wb_i} represents the position translation from the body coordinate system corresponding to the ith keyframe to the world coordinate system. In our configuration, the body frame is set to be the IMU frame. The quaternion q_{wb_i} represents the rotation from the body coordinate system corresponding to the ith keyframe to the world coordinate system. The vector v_i^w represents the velocity of the body coordinate system in the world coordinate system. The vector $b_a^{b_i}$ represents the accelerometer bias, and the vector $b_g^{b_i}$ represents the gyroscope bias. In the SLAM backend, by constructing IMU preintegration residuals, visual reprojection residuals, and prior residuals, a cost function is constructed to perform nonlinear optimization solutions. The cost function is as follows:

$$\min_{\chi} \left\{ \rho\left(\left\| r_p - J_p \chi \right\|_{\Sigma_p}^2 + \sum_{i \in B} \rho\left(\left\| r_b\left(z_{b_i b_{i+1}}, \chi\right) \right\|_{\Sigma_{b_i b_{i+1}}}^2 \right) + \sum_{(i,j) \in F} \rho\left(\left\| r_f\left(z_{f_j}^{c_i}, \chi\right) \right\|_{\Sigma_{f_j}^{c_i}}^2 \right) \right) \right\} \tag{3}$$

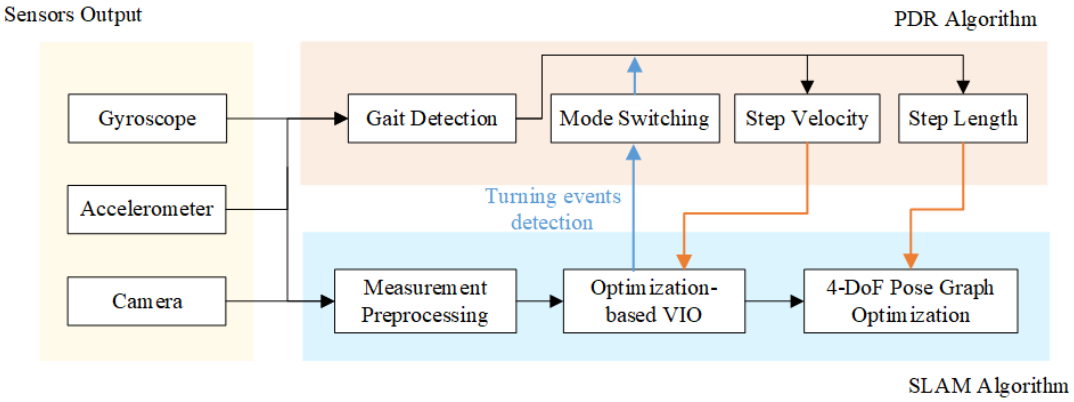


Fig. 1. Architecture of the proposed system.

where $\rho(\|r_p - J_p\chi\|^2_{\Sigma_p})$ is the residual of prior information, and $r_b(z_{b_i b_{i+1}}, \chi)$ and $r_f(z^{c_i}_{f_j}, \chi)$ are residuals for IMU and visual measurements, respectively. B is the set of all IMU measurements, and F is the set of features that have been observed at least twice in the current sliding window. After constructing the cost function, the Levenberg–Marquardt (LM) algorithm is used to iteratively solve the nonlinear optimization which can be achieved by the Ceres solver [39].

B. PDR Technology

During walking, the pedestrians' gait information also has periodicity due to the regularity of the pedestrians' actions. For each step, pedestrians walk along the direction that they are facing. So we can reasonably assume that there is only the speed of the forward direction when pedestrians walk naturally, and that lateral and vertical speed can be negligible during each step. Based on the basic assumption, the PDR technology can be applied in the pedestrian navigation field, as illustrated in Fig. 2.

When a pedestrian is walking while holding the smartphone, the gait cycle can be detected by the accelerometer sensor of the smartphone, and each step of the pedestrian can be identified mainly through peak detection and zero-crossing detection methods. The PDR technology includes three core steps: 1) gait detection; 2) step length estimation; and 3) heading calculation. The PDR algorithm estimates the specific location of the pedestrian by using inertial sensors (gyro, accelerometer, and magnetometer) to estimate the pedestrian's step length, gait, and heading angle.

There are three main models for estimating the length of pedestrian steps: 1) constant model; 2) linear model; and 3) nonlinear model. The constant model divides a measured walking distance by the counted number of steps to get the average step length, that is, the step length is considered to be constant. The linear model collects walking data of pedestrians of different heights, assuming a linear relationship between step length and frequency. The nonlinear model named Weinberg model is as follows [40]:

$$\mathrm{SL} = K \times \sqrt[4]{a_{\max} - a_{\min}} \tag{4}$$

where $a_{\max}$ and $a_{\min}$ are the maximum and minimum values of the synthetic acceleration, and K is the model parameter which represents the scale factor of the step length. We use the Weinberg model to obtain the step length information, because this method is known to perform well even if generalized calibration values are used for different users [41].

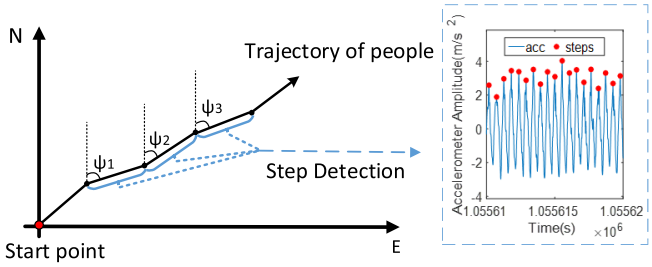


Fig. 2. PDR algorithm architecture. Pedestrian trajectory is determined by the step size and heading of each step (left). Pedestrian walking has obvious periodicity and can the walking period be detected by the acceleration (right).

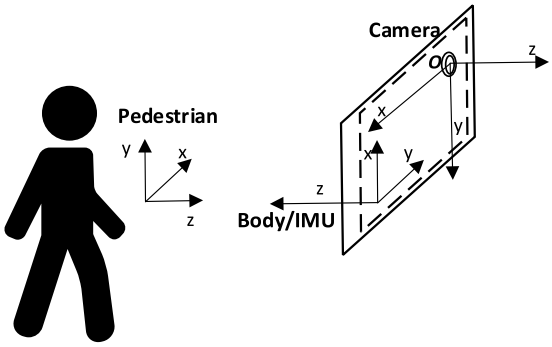


Fig. 3. Pedestrian frame, body frame, camera frame, and their corresponding relationship.

At the same time, the average speed of pedestrians in the pedestrian coordinate system within one step $\tilde{\boldsymbol{v}}^l$ can be expressed as $[\,(\boldsymbol{SL}/\Delta t)\;\;0\;\;0\,]^T$, $\tilde{\boldsymbol{v}}^l$ represents the velocity vector in the l-frame. We assume that pedestrians walk on a level plane in the indoor environment, so the roll and pitch misalignment is equal to roll and pitch [42].

C. PDR-Assisted VI-SLAM Model

Due to the gait information being related to the state vectors of the VIO system, the gait information can be used to assist the VIO system. The gait information obtained by the PDR algorithm is calculated in the pedestrian frame, while the state vector of the VIO system is obtained in the camera and body frame. The relationship between the pedestrian coordinate system, the body coordinate system, and the camera coordinate system is shown in Fig. 3.

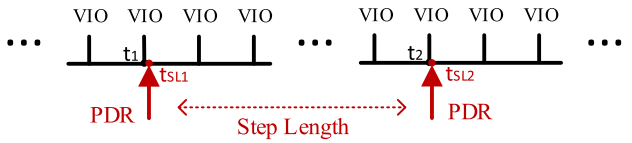


Fig. 4. Selection of keyframes in the sliding window corresponding to step length information. According to the gait information obtained from PDR algorithm, the start time t_{SL1} and end time t_{SL2} of the step length are used to find the nearest keyframe.

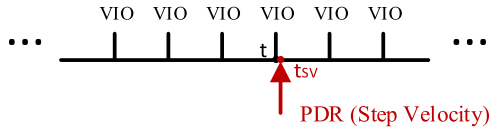


Fig. 5. Selection of keyframes in the sliding window corresponding to step velocity information. According to the gait information obtained from PDR algorithm, the time of the step velocity t_{SV} is used to find the nearest keyframe.

The pedestrian coordinate system defined in this article refers to the hand-held smartphone coordinate system. Since the measurements for calculating the pedestrian step length and step velocity are obtained from smartphones, the pedestrian coordinate system is essentially based on the definition of hand-held smartphones. Therefore, the transformation between the pedestrian coordinate system and the body coordinate system is relatively fixed. Meanwhile, since the state variables (step length and step speed) obtained by the PDR algorithm are decoupled and calculated separately from the heading, it is unnecessary to confirm the initial heading. After the gait information and position information are converted to the same coordinate system, the optimized VIO system assisted by pedestrian gait information can be realized.

For the measurements obtained by the camera and inertial sensors on the smartphones, the SLAM system proceeds with data preprocessing and real-time position estimation, and PDR algorithm obtains pedestrian gait information. For the step length information, when a pedestrian is detected to take a step, the current step length of the pedestrian is obtained through the PDR algorithm, and the start time t_{SL1} and end time t_{SL2} corresponding to the current step are obtained at the same time, as shown in Fig. 4. According to this time, the matched keyframe is found, and the corresponding constraint is constructed for optimizing position. Similarly, for the step velocity information, the matched keyframe is found according to the corresponding time t_{SV} and the error factor is constructed, as shown in Fig. 5. Sometimes, there is a problem that the start time of steps cannot match the timestamp of keyframes exactly. Here, we set a threshold of 0.08 s to find the nearest keyframe with a time difference of less than 0.08 s. At the same time, this error has been considered in the noise error characteristic analysis, so time synchronization is no longer carried out separately.

1) Feasibility Analysis: In order to prove the effectiveness of adding pedestrian gait information, we designed the experiments outdoors and used the Fixposition Vision-real-time kinematic (RTK) as the ground truth acquisition device.

The Fixposition Vision-RTK is a solution that combines computer vision, GNSS, and IMU measurements to achieve high-accuracy positioning, and can achieve centimeter-level positioning. We simulate challenging scenes for indoor positioning and add dynamic objects during the experiment. The Fixposition Vision-RTK is applied to synchronize data collection with smartphones, as shown in Fig. 6. In the process of data collection, the Fixposition Vision-RTK and smartphone remain relatively fixed.

Fig. 6. Front view of the vision-RTK, which can provide a high-accuracy position as the reference in the outdoor environment to verify the feasibility of adding gait information.

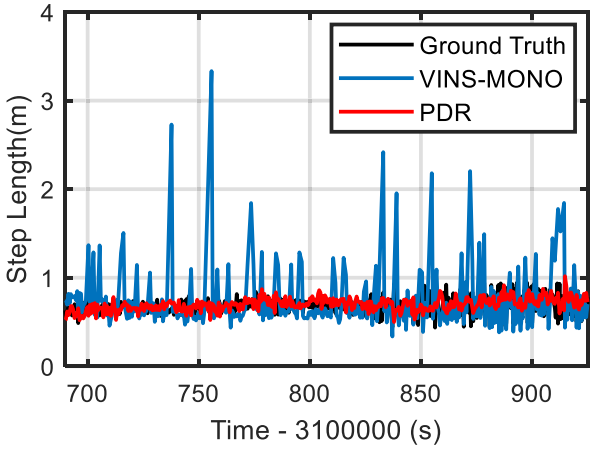


Fig. 7. Comparison of step length between VINS-MONO, PDR algorithm, and the ground truth. It shows that the step length obtained by the PDR algorithm is more consistent with the ground truth.

We compare the step length obtained by the VINS system, the PDR algorithm, and the ground truth respectively, as shown in Fig. 7. The step length information calculated from VINS-MONO contains many data larger than 1 m, which is against the regulation of human movement. In the process of VINS-MONO, the accuracy of positioning will be affected by large noise caused by visual mismatch, and the cumulative error of inertial navigation. Based on this, the problem of large step length error in the process of positioning will appear. The step length obtained by the PDR algorithm is more in line with truth data, which can prove that using PDR constraints can better constrain such outliers. Because the timestamps of the system states were all greater than 31 000 s, the timestamps of the x-axis will be crowded together if the corresponding timestamps were directly displayed which is not very convenient to observe. Therefore, we choose "time-310 000 s" as the scale of the x-axis.

To show the performance of the different algorithms more vividly, the step length information obtained by VINS-MONO and PDR algorithm is compared with the ground truth. The error between step length obtained from an algorithm and the ground truth are calculated. The mean and max step length errors between VINS-MONO, PDR, and ground truth of the first test were shown in Table I.

TABLE I
MEAN AND MAX STEP LENGTH ERROR BETWEEN VINS-MONO, PDR, AND GROUND TRUTH

	Mean Error (m)	Max Error (m)
VINS-MONO	0.1634	2.6769
PDR	0.0986	0.4700

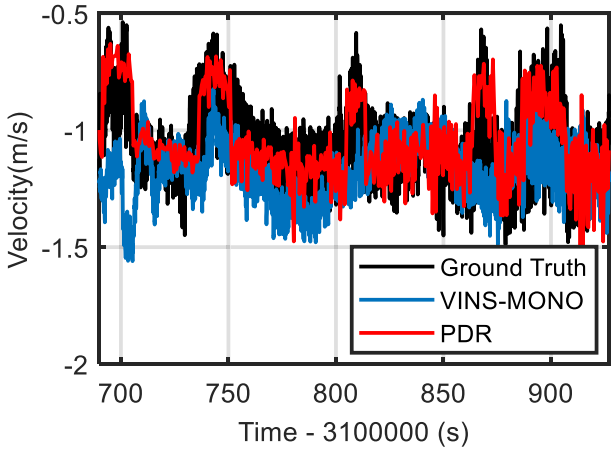


Fig. 8. Comparison of the output speed of *z*-axis between VINS-MONO, PDR algorithm, and the ground truth, which can show that the velocity obtained by PDR algorithm is more consistent with the ground truth.

TABLE II
MEAN AND MAX STEP VELOCITY ERROR BETWEEN VINS-MONO, PDR, AND GROUND TRUTH

	Mean Error (m/s)	Max Error (m/s)
VINS-MONO	0.1677	0.9884
PDR	0.0520	0.4316

While using the Fixposition Vision-RTK to collect the true data of pose, we performed positioning outdoor to obtain the corresponding velocity output by the VINS-MONO system and the velocity obtained by the PDR, as shown in Fig. 8. By comparison, it can be seen that large error between the pedestrian's step velocity obtained by VINS-MONO and the ground truth occurs in the scene with dynamic objects, which is caused by visual observation noise and inertial navigation cumulative error. The velocity obtained by the PDR algorithm is more consistent with the ground truth, which proves that it is feasible to use PDR velocity as aiding information for the SLAM system.

The error between step velocity obtained from an algorithm and ground truth are calculated. The mean and max step velocity error between VINS-MONO, PDR algorithm, and ground truth of the first test were shown in Table II.

2) Noise Error Characteristic Analysis: Generally, the noise of the state can be directly derived from the noise covariance of the IMU (accelerometer and gyroscope) by constructing the state recurrence equation. Since the PDR model is empirical, the noise of state (step length and velocity) obtained from PDR cannot be directly derived from the state equation. It is necessary to analyze the noise error characteristics to determine the noise variance.

We use smartphones to collect the image and IMU data in the outdoor environment for validation. The step length and velocity obtained from the PDR and the ground truth are compared to analyze the gait information error characteristics.

The velocity of the Fixposition Vision-RTK is in an earth-centered earth-fixed (ECEF) coordinate system, and it should be converted to the pedestrian coordinate system before comparison. The corresponding relationship between pedestrian velocity and Fixposition Vision-RTK state is as follows:

$$\hat{v}^l \approx R_{lb} R_{be} \hat{v}^e \tag{5}$$

where R_{lb} is the rotation matrix from the pedestrian coordinate system to the body coordinate system. As referred at the beginning of Section III-C, the pedestrian coordinate system is based on the definition of smartphones. Therefore, the pedestrian coordinate system and the body coordinate system are relatively fixed positions. From the corresponding relationship, we can get as follows:

$$R_{lb} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{6}$$

The corresponding relationship between pedestrian step length and Fixposition Vision-RTK state is as follows:

$$\hat{p}^l_{b_i b_j} \approx R_{lb} R_{be} \hat{p}^e_{b_i b_j}. \tag{7}$$

As shown in Fig. 5, the error of the step length of the PDR algorithm and ground truth is less than 0.5 m. The difference between the ground truth of pedestrian velocity $R_{lb}R_{be}\hat{v}^e$ and the variable $\hat{v}^l$ derived from the PDR algorithm at the current moment is used to analyze the error characteristics of the PDR velocity empirical model. As shown in Fig. 6, the error of step velocity of the PDR algorithm and ground truth is less than 0.5 m/s.

After obtaining the corresponding error characteristics of the PDR model, it can be transformed into the covariance matrix of the VINS-MONO system relative to the PDR velocity observations

$$E\left(\delta v \delta v^T\right) = 0.25 \times I_{3\times 3}. \tag{8}$$

The covariance matrix of step velocity observation noise is the square of velocity error sequence standard obtained by the PDR algorithm.

3) Residual Error Constraints During Optimization: PDR algorithm starts from the start point of the known position and achieves continuous tracking and positioning of pedestrians by measuring the distance and direction of movement. In VI-SLAM, visual information, IMU information, and prior information are used to construct a loss function at the same time. To introduce pedestrian gait constraints, the step length and velocity calculated from the PDR algorithm are used as auxiliary observation values.

The obtained step length and velocity information are combined with the state variables in the current sliding window to construct the corresponding residual constraints, and the corresponding residual factors are added to the loss function for optimization. This residual error constraints method is mainly to optimize the real-time position of the current state, besides the step length information will also be used to construct 4-DOF local optimization to realize the relative correction between two keyframes, which will be described in detail in Section IV. At the same time, because PDR is

essentially an empirical model, the noise covariance matrix needs to use the noise mean square error obtained by error characteristic analysis, which is described in Section II.

Before adding pedestrian steps information, the cost function of the VINS-MONO system mainly consists of the prior residual, the IMU preintegration residual, and the residual of visual reprojection. Adding the residual constructed by step length and velocity information to cost function is the key step of the algorithm.

First, the pedestrian step length information needs to be added to construct the residual constraint and the Jacobian matrix. The residual constraint needs to find the two keyframes corresponding to the start position and end position of the pedestrian step in the sliding window, which is mainly obtained by time matching. The pedestrian step start time t_{SL1} and end time t_{SL2} obtained by the PDR algorithm are compared with keyframe times in the sliding window to find the closest keyframe. And then the corresponding step length and state vectors are used to solve the residual and the Jacobian matrix. The residual is expressed as follows:

$$\begin{aligned}\rho_{\mathrm{PDR}_p} &= \rho\left(\left\|r_{\mathrm{pdr}}\left(\hat{p}^l_{b_ib_j} \mid \chi\right)\right\|^2_{\Sigma_{b_ib_{i+1}}}\right)\\ &= \hat{p}^l_{b_ib_j} - C^l_{b_i}\hat{C}^{b_i}_w\left(\hat{p}^w_{b_j} - \hat{p}^w_{b_i}\right).\end{aligned} \tag{9}$$

The poses at two moments are estimated values and there is noise. To solve the Jacobian matrix, it is expanded as follows:

$$\begin{aligned}\hat{p}^l_{b_ib_j} &\approx C^l_{b_i}\hat{C}^{b_i}_w\left(\hat{p}^w_{b_j} - \hat{p}^w_{b_i}\right)\\ &\approx C^l_{b_i}C^{b_i}_w\left(I + \psi\times\right)\left(p^w_{b_j} + \delta p^w_{b_j} - p^w_{b_i} - \delta p^w_{b_i}\right)\\ &\approx p^l_{b_ib_j} + C^l_{b_i}C^{b_i}_w\delta p^w_{b_j} - C^l_{b_i}C^{b_i}_w\delta p^w_{b_i}\\ &\quad - C^l_{b_i}C^{b_i}_w\left(p^w_{b_j}\times\right)\psi + C^l_{b_i}C^{b_i}_w\left(p^w_{b_i}\times\right)\psi.\end{aligned} \tag{10}$$

Second, the pedestrian steps velocity information obtained from the PDR algorithm is also effective to assist the SLAM system. So pedestrian steps velocity information needs to be added, residual constraint terms need to be constructed and the derivation of the Jacobian matrix needs to be performed. The time of pedestrian velocity t_{SV} calculated by the PDR algorithm is used to compare with the time of keyframes in the sliding window to find the corresponding keyframe. Solve the corresponding residual and Jacobian matrix, the residual is as follows:

$$\rho_{\mathrm{PDR}_v} = \rho\left(\left\|r_{\mathrm{pdr}}\left(\hat{v}^l \mid \chi\right)\right\|^2_{\Sigma_{b_i}}\right) = v^{\ l} - C^l_b\hat{C}^b_w\hat{v}^w \tag{11}$$

where the position and velocity are estimated variables, and they contain the measurement noise. To solve the Jacobian matrix, they are expanded as follows:

$$\hat{v}^l \approx C^l_b\hat{C}^b_w\hat{v}^w \approx v^l + C^l_bC^b_w\delta v^w - C^l_bC^b_w\left(v^w\times\right)\psi. \tag{12}$$

The residual function after adding the pedestrian gait information constraint is as follows:

$$\min_{\chi}\left\{\rho_{\mathrm{prior}} + \rho_{\mathrm{IMU}} + \rho_{\mathrm{image}} + \rho_{\mathrm{PDR}_v} + \rho_{\mathrm{PDR}_l}\right\}. \tag{13}$$

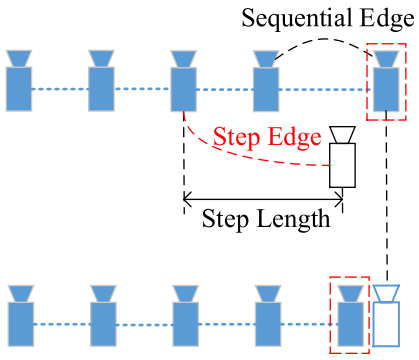


Fig. 9. Pose graph optimization procedure, when each step is detected, the step information is used to construct the 4-DOF constraint, and the path edge and step edge are used for state correction.

Among them, ρ_{prior}, ρ_{IMU}, and ρ_{image} represent the residual of prior information, IMU, and visual measurements, respectively, as shown in (3). $\rho_{\mathrm{PDR}_v} = \sum_{i\in B}\rho(\|r_{\mathrm{pdr}}(\hat{v}^l \mid \chi)\|^2_{\Sigma_{b_i}})$ represents the constrained residual error of the PDR speed on the system, $\rho_{\mathrm{PDR}_l} = \sum_{(i\mid j)\in F}\rho(\|r_{\mathrm{pdr}}(\hat{p}^l_{b_ib_j} \mid \chi)\|^2_{\Sigma_{b_ib_{i+1}}})$ represents the effect of the PDR step increment on the system constrained residuals, and covariance matrix and information matrix are obtained from the analysis of noise error characteristics.

4) 4-DOF Optimization: When the PDR algorithm detects each step, the additional local optimization algorithm is developed to ensure the set of past poses is registered into a locally consistent configuration. For the keyframe of one step, mark it as I_{old} and I_{cur}. The displacement from I_{old} to I_{cur} is the step vector derived from the PDR model.

At the time of the update procedure, the keyframe in the sliding window is aligned with the frame detected by the step size. Since it is assumed that the pedestrian has only the displacement in the forward direction, and the pitch and roll angles do not deviate, the 4-DOF optimization is performed.

Keyframes are added to the pose graph after the VIO process. Every keyframe serves as a vertex in the pose graph, and it connects with other vertexes by two types of edges, as shown in Fig. 9.

1) *Sequential Edge:* A keyframe establishes several sequential edges to its previous keyframes. A sequential edge represents the relative transformation between two keyframes, which is taken directly from VIO. Considering keyframe i and one of its previous keyframes j, the sequential edge only contains relative position $\hat{p}^i_{ij}$ and yaw angle $\hat{\psi}_{ij}$

$$\hat{p}^i_{ij} = \hat{R}^{w-1}_i\left(\hat{p}^w_j - \hat{p}^w_i\right) \tag{14}$$

$$\hat{\psi}_{ij} = \hat{\psi}_j - \hat{\psi}_i. \tag{15}$$

2) *Step Edge:* If the newly marginalized keyframe and the aforementioned keyframe constitute a step size constraint, it will be linked to the keyframe pointing to the previous step in the pose graph through the step edge, and the step edge only contains four degrees of freedom. The value of the step length side is derived from the PDR detection result.

We define the residual of the edge between frames i and j minimally as follows:

$$r_{i,j}\left(p_i^w, \psi_i, p_j^w, \psi_j\right) = \begin{bmatrix} R\left(\hat{\phi}_i, \hat{\theta}_i, \psi_i\right)^{-1}\left(p_j^w - p_i^w\right) - \hat{p}_{ij}^i \\ \psi_j - \psi_i - \hat{\psi}_{ij} \end{bmatrix} \quad (16)$$

where $\hat{\phi}_i$ and $\hat{\theta}_i$ are the fixed estimates of roll and pitch angles, which are obtained from monocular VIO.

The whole graph of sequential edges and step edges is optimized by minimizing the following cost function:

$$\min_{p,\psi}\left\{\sum_{(i,j)\in S}\left\|r_{i,j}\right\|^2 + \sum_{(i,j)\in P}\rho\left(\left\|r_{i,j}\right\|^2\right)\right\} \quad (17)$$

where S is the set of all sequential edges and P is the set of all step edges from the PDR algorithm. The Huber norm $\rho(\cdot)$ is introduced to effectively reduce the impact on the system when there is a large error in the PDR. In contrast, we do not use any robust norms for sequential edges, as these edges are extracted from VIO, which already contains sufficient outlier rejection mechanisms.

5) Turning Events Detection and PDR Mode Switching: Besides the assistance of the PDR algorithm to the VIO system, the turning information obtained from the VIO system can help the PDR algorithm obtain more accurate gait information when the pedestrian is turning.

Pedestrians with different walking states, straight walking, and rotation, have different gait characteristics. If the two states can be successfully distinguished, the PDR mode can switch in real-time to obtain more accurate gait information.

Real-time attitude information can be obtained from the VIO system, so the corresponding pedestrian walking mode can be determined. As shown in Fig. 10, when the pedestrian is in the state of rotation, the attitude data obtained from the VIO system will change significantly.

The moving average method is utilized to determine the pedestrian's turning state. We maintain consecutive several attitude states in a sliding window for turning detection. The average value is determined for the states in the sliding window to get the average attitude. We compare the current sliding window's average value to the previous sliding window's average value

$$\mathrm{dis}_{q_c} = \frac{1}{n} \times \sum_{k\in\mathcal{L}} q_c(k) \quad (18)$$

where $\mathcal{L}$ is the set of all attitude data in the sliding window, n is the total number of vectors in the sliding window, and c is the component of attitude quaternion. The difference between the average value of the quaternion states in the current sliding window and the average value of the quaternion states in the previous sliding window is shown in Fig. 11. As shown in Fig. 11, obvious changes happen when the pedestrian is turning.

When the difference is greater than the threshold, the mode of the PDR algorithm will be switched as a turning event. The K of the Weinberg model is switched to K_1. When the difference is less than the threshold, the mode of the PDR algorithm will be switched to straight walking. The K of the Weinberg model is switched to K_2. By collecting multiple groups of data of pedestrians walking straight and turning, the corresponding parameters K_1 and K_2 can be designed by fitting.

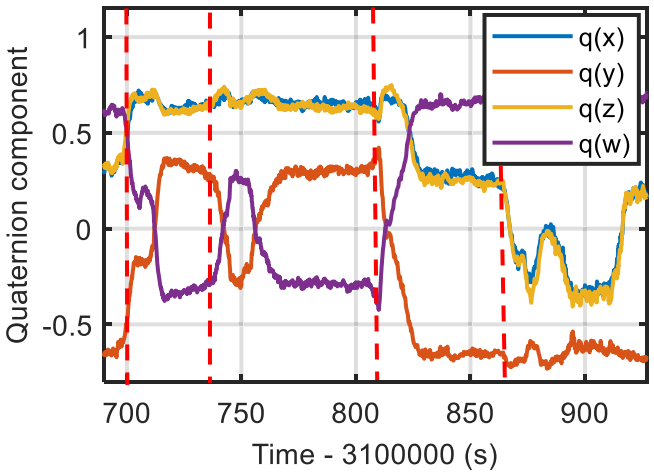


Fig. 10. Output attitude quaternion of the VINS-MONO system. The quaternion component will change when the pedestrian is rotating, and the red dashed line represents the beginning of the pedestrian's turning.

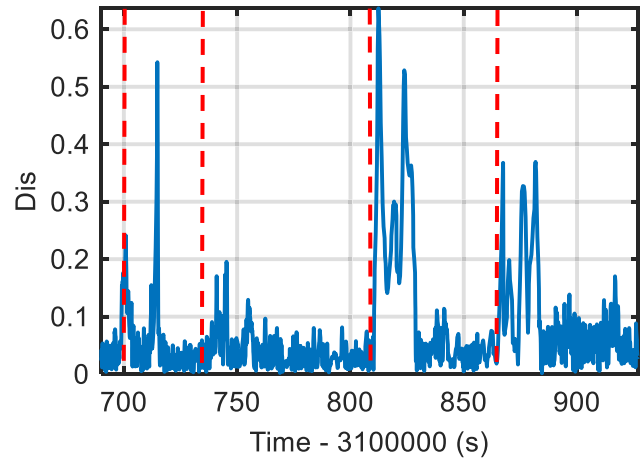


Fig. 11. Difference between the average value of the quaternion states in the current sliding window and the ones in the previous. The red dashed line represents the beginning of a pedestrian's turning.

Since the statistical characteristics of pedestrian gait information are different in different walking states, accurate gait information can be obtained by optimally estimated K_1 and K_2. The attitude information obtained by the VIO system can further modify the PDR model, so as to achieve a more accurate output of gait information and optimize the system.

IV. Experimental Result

In this section, we performed field tests to evaluate the proposed PDR-aided-SLAM system, and the experimental results would be presented to demonstrate the robustness and effectiveness of our proposed method. We test our system in three typical indoor environments with different texture features to evaluate the overall performance. Since the two processes of constructing the pedestrian gait constraints and estimating pedestrian position using VIO can be

Fig. 12. Test environment of path 1, which is a vision-challenging environment with low-texture area.

simultaneously carried out, the proposed framework can still run in real time.

A. Experimental Configuration

All experiments were conducted using a handheld mobile phone, Huawei P30. This device incorporates IMU data samples at 100 Hz and camera frames at 30 Hz. The test data was collected at a normal walking speed by two volunteers, one female and one male who were 160 and 173 cm tall and had different pretrained parameters of the Weinberg model, respectively.

Since there is no effective and reliable method to get the true poses of a handheld smartphone when the pedestrian is walking indoors, we used a foot-mounted strap-down inertial navigation system (SINS) with STIM300 (0.5°/h) to synchronously record the reference position.

During the tests, the volunteers walked naturally holding the smartphone toward the front and followed the predefined path belonging to the specific scene. For a reliable evaluation of the superiority of the proposed method, loop error is introduced besides comparison with the ground truth, thus the closed paths were formed.

B. Evaluation

For a better demonstration of the performance enhancement of indoor positioning, the 2-D localization result of the proposed system together with VINS-Mono would be presented.

The first test was conducted in a texture flaw region, as shown in Fig. 12, where VI-SLAM algorithms behaved poorly and even brought large positioning drift. There existed similar and simple structures, including blank walls and confused floors which may result in poor visual tracks. Without the loss of generality, path 1 was formed with the most simple closed rectangular shape and the total length is about 60 m.

The step length information was added mainly through constructing the residual factor (9) and 4-DOF local optimization (17), and the system assisted by step length information was step length VINS (SL-VINS). The step velocity information was added mainly through constructing the residual factor (11), and the system assisted by step velocity information was SV-VINS. Therefore, the difference between SV-VINS and SL-VINS lies not only in information sources, but also in the ways of information assistance. The step length information additionally was added by the local 4-DOF optimization.

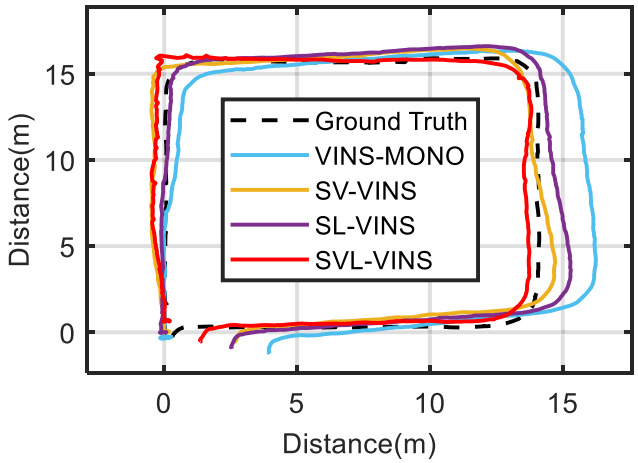


Fig. 13. Comparison of 2-D trajectory between VINS-MONO and proposed methods for path 1.

Fig. 13 shows the 2-D trajectory of our method and VINS-MONO on path 1. As shown in Fig. 11, the trajectory estimated by our method is much closer to the ground truth. In the current indoor scenario, due to the low-quality sensors of the device and fewer texture features, the robustness and accuracy of the VI-SLAM for pedestrians degraded, as large positioning drifts of VINS-Mono were shown in blue. In contrast, adding pedestrian SL-VINS and step velocity information (SV-VINS) separately and adding both at the same time (SVL-VINS) on VINS-Mono achieved better performance.

For the SV-VINS, which applied step velocity in the body frame to aid VINS-MONO, heading drift during pedestrians turning was better suppressed. Since the various state variables were coupled together in the back-end nonlinear optimization, the additional speed observation could not only reduce the velocity error but also be conducted to correct the heading and positioning drift.

The SL-VINS, which applied step length as periodic displacement observation to aid VINS-MONO, provided stable and reliable position increments for the local positioning refinement. With the accumulative local pose graph optimization, a better loop effect and a more accurate track, especially in the straight corridors, were achieved.

Note that a more closed loop and better consistency with the reference path were presented as shown in red. The SVL-VINS, which applied both the pedestrian velocity and step length as additional state measurements, achieved further improvement in positioning performance. Through the joint assistance of enhanced sliding-window-based nonlinear optimization and the novel local 4-DOF pose graph optimization, pedestrian gait information could be fully utilized to achieve a more stable and reliable positioning for pedestrians.

When evaluating the performance of the SLAM system, a common practice is to use absolute trajectory error (ATE). Meanwhile, loop error is an efficient method to evaluate the accuracy of the system which expresses the distance between the start point and the endpoint of a loop form. The loop error and ATE of the first test were shown in Table III. SV-VINS outperformed VINS-MONO in path 1 and

TABLE III
LOOP ERROR AND ATE BETWEEN SVL-VINS AND VINS-MONO, SV-VINS, AND SL-VINS FOR PATH #1

	Loop Error (m)	ATE(m)	Performance Enhancement(%)
VINS-MONO	4.0937	0.6220	-
SV-VINS	2.8084	0.3598	42.15
SL-VINS	2.6926	0.3779	39.24
SVL-VINS	1.8705	0.3743	39.82

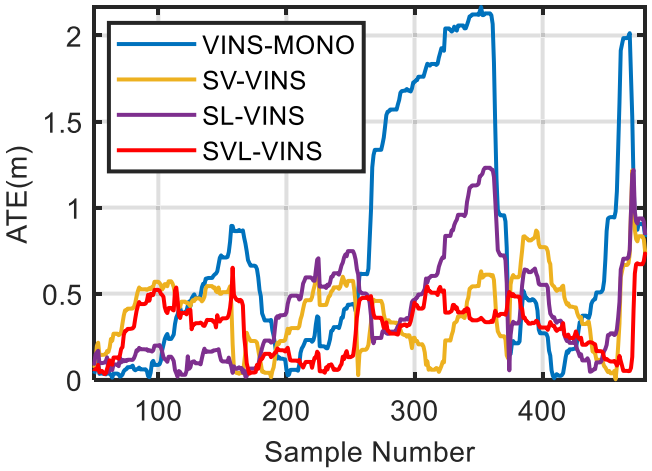


Fig. 14. ATE between VINS-MONO, the proposed methods and ground truth for path 1.

improved the accuracy by 42.15%. SL-VINS outperformed VINS-MONO in path 1 and improved the accuracy by 39.24%. SVL-VINS outperformed VINS-MONO in path 1 and improved the accuracy by 39.82%. Note that a significant improvement in ATE was shown after leveraging the pedestrian gait information. Among them, SV-VINS has the best performance of ATE, which is because the trajectory error in path 1 is mainly caused by heading offset. Therefore, the 4-DOF constraint constructed by step length information does not improve the performance of ATE of the whole system. At the same time, due to the inherent error of the foot-mounted SINS, little further ATE improvement of SVL-VINS than SV-VINS is reasonable. For the loop error index, with the local positioning refinement using step length, the loop error was further mitigated than the original system.

We plot the ATE between several VINS algorithms and the ground truth of path 1 in Fig. 14. When the position becomes divergent due to the pedestrian turning, the addition of step velocity can obviously reduce the error of the VINS algorithm. Fig. 15 shows the positioning cumulative error percentages of VINS-MONO and proposed methods for path 1. The positioning results calculated by the SVL-VINS algorithm have higher accuracy compared with VINS-MONO. The error of the SVL-VINS algorithm can be almost controlled within 0.5 m.

The second scene was located in the hall of the teaching building, as shown in Fig. 16. There were dynamic objects such as pedestrians during the test process, which was a challenging environment for the visual tracking system.

The total length of path 2 is 63 m. Fig. 17 shows the 2-D positioning results of our method and VINS-MONO

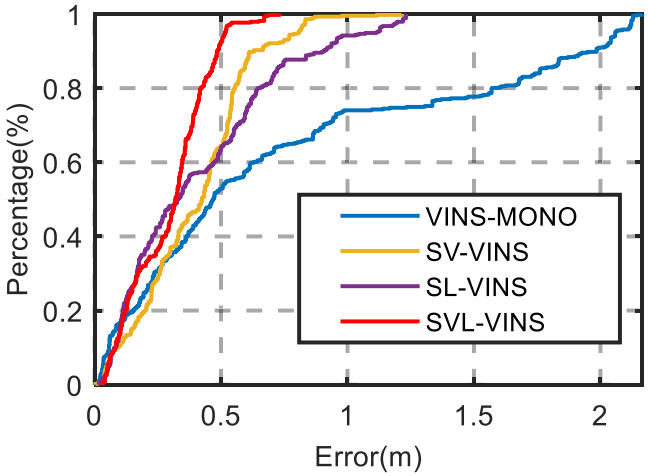


Fig. 15. Position cumulative error percentages of VINS-MONO and proposed methods for path 1.

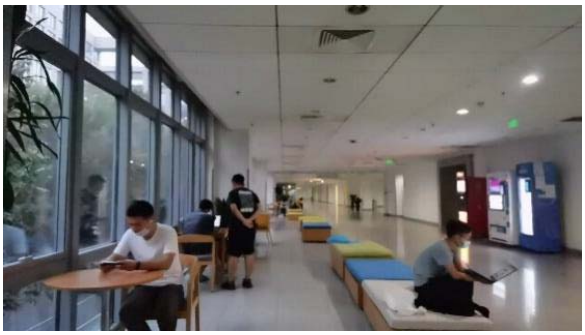


Fig. 16. Test environment of path 2, which is a vision-challenging environment with dynamic objects.

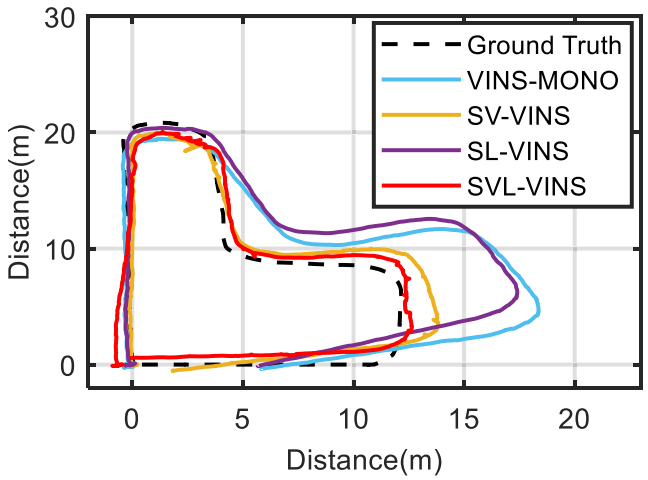


Fig. 17. Comparison of 2-D trajectory between VINS-MONO and proposed methods for path 2.

on the second test. It shows that the trajectory estimated by our method is much closer to the ground truth. Due to the movement of pedestrians during the test, VINS-MONO presented a serious heading deviation in the second corner, after that the large global tracking drift also occurred.

Note that the divergence of motion state estimation was mainly caused by the dynamic objects at the second corner, and also the blank wall lacking texture features. Since the pedestrian gait information would not be influenced by the surrounding environments, it was considered a reliable motion observation.

TABLE IV
LOOP ERROR AND ATE BETWEEN SVL-VINS AND VINS-MONO, SV-VINS, AND SL-VINS FOR PATH #2

	Loop Error (m)	ATE	Performance Enhancement (%)
VINS-MONO	6.0160	1.0772	-
SV-VINS	1.8874	0.5559	48.39
SL-VINS	5.7741	1.1406	-5.89
SVL-VINS	0.8455	0.5313	50.68

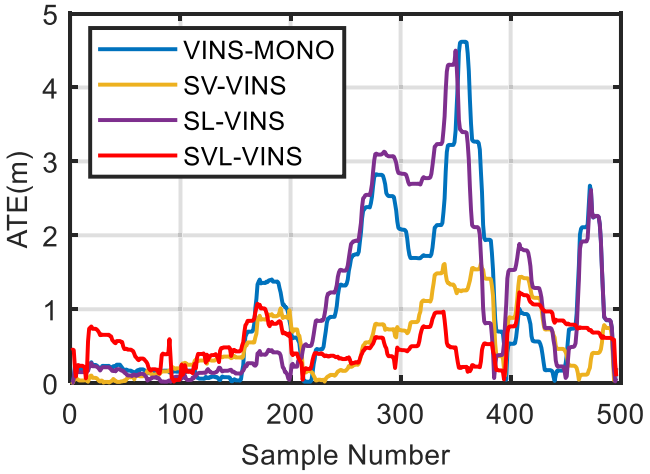


Fig. 18. ATE between VINS-MONO, the proposed methods, and the ground truth for path 2.

As shown in Fig. 17, SV-VINS significantly corrected the heading drift during the pedestrian turning with the help of step velocity. The SL-VINS, which only applied step length in the local pose graph optimization, still suffer the large corner drift although the straight path tracking was better refined. Furthermore, SVL-VINS fully utilized the pedestrian information, thus realizing both closed-loop and little global positioning error, especially the improvement of the heading drift during the corner.

The loop error and ATE of VINS-MONO and the proposed algorithm for path 2 were shown in Table IV. SV-VINS outperformed VINS-MONO in path 2 and improved the accuracy by 48.39%. SVL-VINS outperformed VINS-MONO in path 2 and improved the accuracy by 50.68%. As mentioned before, SL-VINS only refined the local displacement between steps but still suffer the heading drift, thus little improvement was achieved. When adding the pedestrian velocity as an additional observation, SV-VINS and SVL-VINS outperformed other methods.

In addition, we plotted the ATE distribution of proposed algorithms in Fig. 18. When the position estimation became divergent due to the texture-less wall and low-quality sensors, the assistance of step velocity could obviously reduce the error of the VINS system.

Fig. 19 shows the positioning cumulative error percentages of VINS-MONO and proposed methods for path 2. As shown in Fig. 18, the positioning results calculated by the SVL-VINS algorithm had higher accuracy compared with VINS-MONO. The error of the SVL-VINS algorithm could be almost controlled within 1 m.

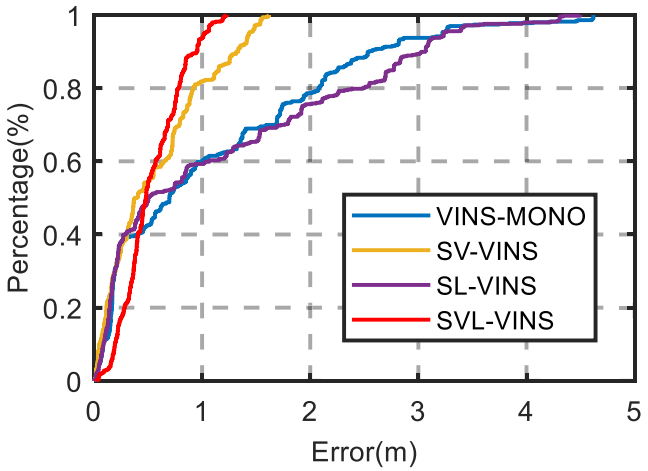


Fig. 19. Position cumulative error percentages of VINS-MONO and proposed methods for path 2.

Fig. 20. Test environment of path 3, which contains multiple sharp turning area.

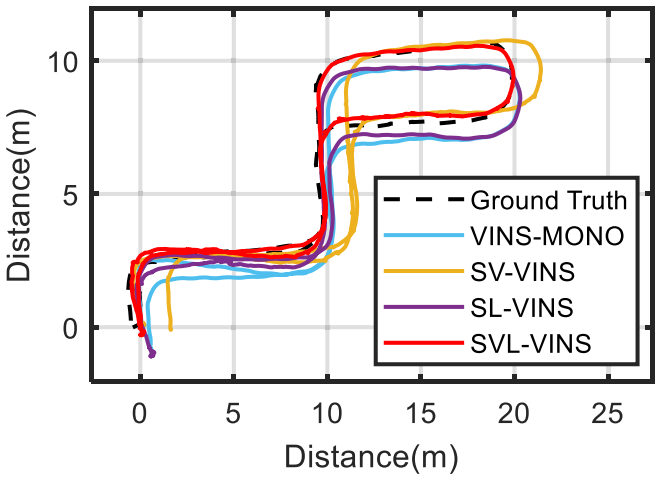


Fig. 21. Comparison of 2-D trajectory between VINS-MONO and proposed methods for path 3.

The third scene was located in an indoor office. There were abundant texture features in the environment, as shown in Fig. 20. The volunteer took multiple sharp turning holding smartphones during the third test. At the same time, the glass door appeared when passing in and out of the room, which may cause visual-tracking's failure due to the texture-less and reflection properties.

The total length of path 3 is 60 m. Fig. 21 shows the 2-D positioning results of our method and VINS-MONO on the third test. It shows that the trajectory estimated by our method is much closer to the ground truth. Due to the multiple

TABLE V
LOOP ERROR AND ATE BETWEEN SVL-VINS AND VINS-MONO, SV-VINS, AND SL-VINS FOR PATH #3

	Loop Error (m)	ATE	Performance Enhancement(%)
VINS-MONO	1.0420	0.3952	-
SV-VINS	1.5480	0.3318	16.04
SL-VINS	1.0953	0.2621	33.68
SVL-VINS	0.2459	0.1046	73.53

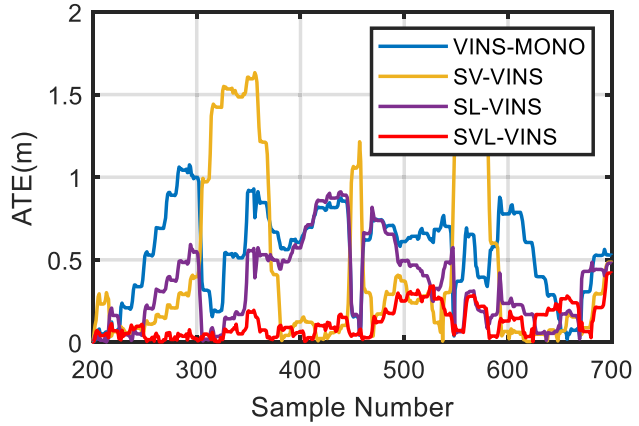


Fig. 22. ATE between VINS-MONO, the proposed methods, and ground truth for path 3.

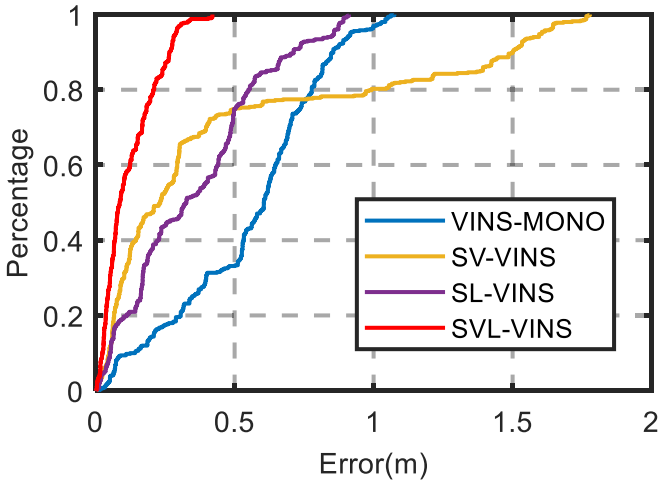


Fig. 23. Position cumulative error percentages of VINS-MONO and proposed methods for path 3.

challenging pedestrians' turning, there are some serious errors in the trajectory. For SV-VINS, the heading drift of the second corner could be corrected significantly. For SL-VINS, which only applied step length in the local pose graph optimization, still suffered the large corner drift though the straight path tracking was better refined. For SVL-VINS, due to the addition of step velocity and length information, the heading and scale divergence had been optimized to a certain extent.

After aligning the truth and the estimated trajectory, we computed the difference between each pair of poses and output ATE. The ATE between several SLAM algorithms and the ground truth is shown in Table V. SV-VINS outperformed VINS-MONO in path 3 and improved the accuracy by 16.04%. SL-VINS outperformed VINS-MONO in path 3 and improved the accuracy by 33.68%. SVL-VINS outperformed VINS-MONO in path 3 and improved the accuracy by 73.53%. SV-VINS and SL-VINS could improve the accuracy of the system partially, and SVL-VINS outperformed others in most cases.

The ATE between the SLAM algorithm and the ground truth of path 3 is shown in Fig. 22. When the position became divergent due to the pedestrian turning, the addition of step velocity and length could obviously reduce the error of the SLAM algorithm.

Fig. 23 shows the positioning cumulative error percentages of VINS-MONO and proposed methods for path 3. As shown in Fig. 23, the positioning results calculated by the SVL-VINS algorithm had higher accuracy compared with VINS-MONO. The error of the SVL-VINS algorithm could be almost controlled within 0.4 m.

V. CONCLUSION

In this article, we proposed an improved smartphone-based SLAM system based on pedestrian gait information for handheld indoor positioning. For the poor positioning caused by the low-performance IMU device of the smartphone and texture-less environment, considering the regularity of the pedestrian walking scene, the PDR algorithm is used to obtain the step length and pace information of the pedestrian at each step. The pace information is used to realize the observation correction in the back-end nonlinear optimization, and the step length information is used to optimize the local 4-DOF pose, to realize the suppression of the heading and positioning divergence. At the same time, the VIO system also assists the PDR algorithm in mode switching, to optimize the PDR algorithm by improving the accuracy of gait information obtained from the PDR algorithm. Through the step-by-step cumulative optimization of the system by pedestrian velocity and pedestrian length, the positioning performance in vision-challenging scenes can be improved. Though the proposed method assumes that the pedestrians walk naturally and the smartphone carrying poses must ensure smooth and unobstructed vision obtained from the camera, it proves that the pedestrian gait information can provide efficient constraints of state estimation in the most common smartphone use cases. In addition, there is no strict hardware requirement for the smartphones as long as they can provide continuous inertial and visual data sequences.

In the future work, we will adopt corresponding pattern recognition based on the different motion patterns of pedestrians, and propose the best solution to achieve more accurate indoor positioning.

REFERENCES

[1] H. Lategahn, A. Geiger, and B. Kitt, "Visual SLAM for autonomous ground vehicles," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 1732–1737.
[2] H. Yu, Q. Fu, Z. Yang, L. Tan, W. Sun, and M. Sun, "Robust robot pose estimation for challenging scenes with an RGB-D camera," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2217–2229, Mar. 2019.
[3] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. 6th IEEE ACM Int. Symp. Mixed Augmented Reality*, Nov. 2007, pp. 225–234.
[4] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: Part I," *IEEE Robot. Autom. Mag.*, vol. 13, no. 2, pp. 99–110, Jun. 2006.

[5] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "MonoSLAM: Real-time single camera SLAM," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1052–1067, Jun. 2007.
[6] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 15–22.
[7] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 834–849.
[8] X. Xie, Y. Yu, X. Lin, and C. Sun, "An EKF SLAM algorithm for mobile robot with sensor bias estimation," in *Proc. 32nd Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, May 2017, pp. 281–285.
[9] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 3565–3572.
[10] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 298–304.
[11] T. Schneider *et al.*, "MAPLAB: An open framework for research in visual-inertial mapping and localization," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1418–1425, Jul. 2018.
[12] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *Int. J. Robot. Res.*, vol. 34, no. 3, pp. 314–334, 2015.
[13] T. Qin, P. Li, and S. Shen, "VINS-Mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Trans. Robot.*, vol. 34, no. 4, pp. 1004–1020, Aug. 2018.
[14] T. Qin, J. Pan, S. Cao, and S. Shen, "A general optimization-based framework for local odometry estimation with multiple sensors," 2019, *arXiv:1901.03638*.
[15] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.
[16] R. Mur-Artal and J. D. Tardós, "Visual-inertial monocular SLAM with map reuse," *IEEE Robot. Automat. Lett.*, vol. 2, no. 2, pp. 796–803, Apr. 2017.
[17] C. Campos, R. Elvira, J. J. G. Rodriguez, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021.
[18] V. Usenko, N. Demmel, D. Schubert, J. Stuckler, and D. Cremers, "Visual-inertial mapping with non-linear factor recovery," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 422–429, Apr. 2020.
[19] P. Li, T. Qin, B. Hu, F. Zhu, and S. Shen, "Monocular visual-inertial state estimation for mobile augmented reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Oct. 2017, pp. 11–21.
[20] T. Schops, J. Engel, and D. Cremers, "Semi-dense visual odometry for AR on a smartphone," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Sep. 2014, pp. 145–150.
[21] P. Ondruska, P. Kohli, and S. Izadi, "MobileFusion: Real-time volumetric surface reconstruction and dense tracking on mobile phones," *IEEE Trans. Vis. Comput. Graphics*, vol. 21, no. 11, pp. 1251–1258, Nov. 2015.
[22] R. W. Levi and T. Judd, "Dead reckoning navigational system using accelerometer to measure foot impacts," U.S. Patent 5 583 776, Dec. 10, 1996.
[23] J. W. Kim, H. J. Jang, D. H. Hwang, and C. Park, "A step, stride and heading determination for the pedestrian navigation system," *J. Global Positioning Syst.*, vol. 3, nos. 1–2, pp. 273–279, Feb. 2015.
[24] J. Qian, J. Ma, R. Ying, P. Liu, and L. Pei, "An improved indoor localization method using smartphone inertial sensors," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat.*, Oct. 2013, pp. 1–7.
[25] P. Goyal, V. J. Ribeiro, H. Saran, and A. Kumar, "Strap-down pedestrian dead-reckoning system," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat.*, Sep. 2011, pp. 1–7.
[26] E. Foxlin, "Pedestrian tracking with shoe-mounted inertial sensors," *IEEE Comput. Graph. Appl.*, vol. 25, no. 6, pp. 38–46, Nov./Dec. 2005.
[27] D. C. Simón, "Step-wise smoothing of ZUPT-aided INS," M.S. thesis, School Elect. Eng., KTH, Stockholm, Sweden, 2012.
[28] S. Rajagopal, "Personal dead reckoning system with shoe mounted inertial sensors," M.S. thesis, School Elect. Eng., KTH, Stockholm, Sweden, 2008.
[29] J. Borenstein and L. Ojeda, "Heuristic drift elimination for personnel tracking systems," *J. Navigat.*, vol. 63, no. 4, pp. 591–606, Oct. 2010.
[30] S. Beauregard, "A helmet-mounted pedestrian dead reckoning system," in *Proc. 3rd Int. Forum Appl. Wearable Comput.*, 2006, pp. 1–11.
[31] L. Ojeda and J. Borenstein, "Personal dead-reckoning system for GPS-denied environments," in *Proc. IEEE Int. Workshop Saf., Secur. Rescue Robot.*, Sep. 2007, pp. 1–6.
[32] D. Yan, C. Shi, T. Li, and Y. Li, "FlexPDR: Fully flexible pedestrian dead reckoning using online multimode recognition and time-series decomposition," *IEEE Internet Things J.*, vol. 9, no. 16, pp. 15240–15254, Aug. 2022.
[33] A. Martinelli, S. Morosi, and E. Del Re, "Daily living movement recognition for pedestrian dead reckoning applications," *Mobile Inf. Syst.*, vol. 2016, pp. 1–13, May 2016.
[34] A. Martinelli, H. Gao, P. D. Groves, and S. Morosi, "Probabilistic context-aware step length estimation for pedestrian dead reckoning," *IEEE Sensors J.*, vol. 18, no. 4, pp. 1600–1611, Feb. 2018.
[35] K. Kunze, P. Lukowicz, K. Partridge, and B. Begole, "Which way am I facing: Inferring horizontal device orientation from an accelerometer signal," in *Proc. Int. Symp. Wearable Comput.*, Sep. 2009, pp. 149–150.
[36] Y. Jin, M. Motani, W.-S. Soh, and J. Zhang, "SparseTrack: Enhancing indoor pedestrian tracking with sparse infrastructure support," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–9.
[37] B. D. Lucas and T. Kanade, *An Iterative Image Registration Technique With an Application to Stereo Vision*. Vancouver, WAS, Canada: British Columbia, 1981.
[38] J. Shi, "Good features to track," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 1994, pp. 593–600.
[39] S. Agarwal and K. Mierle, "Ceres solver: Tutorial & reference," *Google*, vol. 2, no. 72, p. 8, 2012.
[40] H. Weinberg, "Using the ADXL202 in pedometer and personal navigation applications," Analog Devices, Norwood, MA, USA, Appl. Note AN-602, 2002.
[41] J. Jahn, U. Batzer, J. Seitz, L. Patino-Studencka, and J. G. Boronat, "Comparison and evaluation of acceleration based step length estimators for handheld devices," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat.*, Sep. 2010, pp. 1–6.
[42] J. Kuang, X. Niu, and X. Chen, "Robust pedestrian dead reckoning based on MEMS-IMU for smartphones," *Sensors*, vol. 18, no. 5, p. 1391, 2018.

Yitong Dong received the M.S. degree from Beihang University, Beijing, China, in 2022.

Her research interests include the multisensor integrated navigation and visual SLAM.

Dayu Yan is currently pursuing the Ph.D. degree with the School of Electronics and Information Engineering, Beihang University, Beijing, China.

His current research interests include multisource fusion for robots, pedestrian navigation, and indoor positioning, which includes GNSS, inertial, and visual integrated navigation.

Tuan Li received the Ph.D. degree from the GNSS Research Center, Wuhan University, Wuhan, China, in 2019.

He is currently an Assistant Professor at the Beijing Institute of Technology, Beijing, China. His research interests include precise GNSS positioning, GNSS/INS integration, and multisensor fusion for robotics.

Ming Xia is currently a Postdoctoral Researcher at the School of Electronic and Information Engineering, Beihang University, Beijing, China. His research interests include motion recognition, pedestrian inertial positioning, wearable sensor-based positioning, and their applications in location-based service applications.

Chuang Shi is a Professor with the School of Electronics and Information Engineering, Beihang University, Beijing, China. He has been constantly in charge of developing the renowned PANDA software package for multi-GNSS data processing and analysis. He is the primary investigator of several national strategic projects such as "XiHe Project" and "National BeiDou Augmentation Service System." His research interests include high-precision GNSS theories and applications, network adjustment, precise orbit determination of GNSS satellites and LEOs, and real-time precise point positioning (PPP).