Hindawi
Journal of Sensors
Volume 2022, Article ID 1311221, 12 pages
https://doi.org/10.1155/2022/1311221

Research Article

A Robust Visual-Aided Inertial Navigation Algorithm for Pedestrians

Langping An, Xianfei Pan, Tingting Li, and Mang Wang

College of Intelligence Science and Technology, National University of Defence Technology, Changsha 410073, China

Correspondence should be addressed to Xianfei Pan; 3312395807@qq.com

Received 10 August 2021; Revised 8 October 2021; Accepted 10 November 2021; Published 6 January 2022

Academic Editor: Qiu-Zhao Zhang

Real-time and robust state estimation for pedestrians is a challenging problem under the satellite denial environment. The zero-velocity-aided foot-mounted inertial navigation system, with the shortcomings of unobservable heading, error accumulation, and poorly adaptable parameters, is a conventional method to estimate the pose relative to a known origin. Visual and inertial fusion is a popular technology for state estimation over the past decades, but it cannot make full use of the movement characteristics of pedestrians. In this paper, we propose a novel visual-aided inertial navigation algorithm for pedestrians, which improves the robustness in the dynamic environment and for multi-motion pedestrians. The algorithm proposed combines the zero-velocity-aided INS with visual odometry to obtain more accurate pose estimation in various environments. And then, the parameters of INS have adjusted adaptively via taking errors between fusion estimation and INS outputs as observers in the factor graphs. We evaluate the performance of our system with real-world experiments. Results are compared with other algorithms to show that the absolute trajectory accuracy in the algorithm proposed has been greatly improved, especially in the dynamic scene and multi-motions trials.

1. Introduction

Pedestrian navigation has been extensively investigated over the last decades, because independent positioning is necessary and challenging under satellite denial environments, such as indoor navigation, mine rescue, and individual combat. Due to the autonomy and continuity of both cameras and inertial measurement units (IMU), visual odometry and inertial navigation system (INS) are the main methods to estimate the pose relative to a known starting point for pedestrians [1]. Besides, visual-inertial odometry (VIO) has become popular in recent years because of the complementary properties of cameras and IMU [2, 3]. Some advanced pedestrian navigation algorithms have already attained satisfactory performance, such as the solution algorithm of the strap-down inertial navigation system (SINS) [1, 4], visual-based methods [5–7], and visual-inertial algorithms [2, 3, 8]. However, several drawbacks, especially poor robustness in the dynamic environment and for multi-motion pedestrians, are limiting the usage of these algorithms in practice.

The error of INS accumulates with time, and it is hard to meet the long-term navigation precision of MIMU for pedestrians. On the one hand, the heading error of pedestrian inertial navigation is not observable, which cannot effectively restrain the heading drift, then the heading error will accumulate over time [4]. On the other hand, the parameters are poor adaptability for different pedestrians under various motion conditions, so that the performance of pedestrian inertial navigation is related to the movement characteristics of pedestrians [9]. Thus, it is the key to adjust parameters adaptively for a robust pedestrian navigation system.

The robustness of the visual navigation in dynamic environments is also challenge for pedestrians. Complex scenes bring unpredictable abnormal observations to the system, which would probably corrupt the quality of the state estimation and even lead to system failure [10]. In addition, the motion characteristics of pedestrians also affects the performance of the system.

The conventional visual-inertial navigation algorithms do not make full use of the human motion characteristics.

Lupton and Sukkah first proposed the theory of inertial integral increment without initial value to solve the problem of inertial vision integrated navigation under high dynamic conditions [11]. The fusion algorithm based on the pre-integration theory only gives scale information through inertial data [3, 8]. We can apply the gait characteristics heading of pedestrians for errors correction.

In this paper, we proposed a robust visual-inertial navigation algorithm to improve the robustness under the condition of limited vision and pedestrian movement, which fuses the cameras with foot-mounted MIMU and adjust parameters of INS adaptively. The conceptual diagram of the algorithm is shown in Figure 1. In the system, the pedestrian state, coming from VO and SINS, is optimized in a batch to obtain a more accurate and robust pose correction. Additionally, we establish the model between zero-velocity interval offset and navigation result error in one step. And taking the optimized result as observation, we estimate the parameters of zero-velocity interval detection to obtain a more accurate pose estimation. In short, our main contributions are as follows:

(i) A novel pose graph optimization algorithm to fusion foot-mounted IMU with visual odometry

(ii) An algorithm to adaptively adjust parameters of the zero-velocity detector, which is driven by the navigation result error

(iii) The general framework to fuse foot-mounted IMU with various sensors, which combines optimization framework and filtering framework to achieves robust localization

(iv) We demonstrate the performance and robustness of our method with extensive experiments. Challenges included dynamic scenarios and multi-motions pedestrians

The remainder of the article is structured as follows. In Sect. II, we discuss relevant literature about ZUPT-aided inertial navigation and the visual-inertial navigation. We give an overview of the algorithm proposed in Sect. III. A pose fusion algorithm between foot-mounted IMU and camera is presented in Sect. IV. Sect. V discusses the result-driven method for adaptive parameter adjustment. The experimental results and their discussion are shown in Sect. VI. Finial, Sect. VII concludes the article.

2. Related Work

2.1. ZUPT-Aided Inertial Navigation. Traditional inertial navigation system integrates IMU measurements to estimate the pedestrian pose relative to a known origin. Typically, pedestrian navigation applies filter approaches [2] or optimized approaches [12–15] to fuse measurements other available sensors with the IMU. Combined with the topic of this paper, we summarize the related research on ZUPT-aided INS, multi-sensors navigation based on optimization framework for pedestrians.

The zero-velocity-aided INS, based on facts that the velocity is zero while the foot touches the ground for pedestrians, fuses the pseudo-measurement of the velocity state with Extended Kalman Filter (EKF) to reduce the accumulated error originating from the integration of noisy IMU measurement. The performance of the navigation system highly relies on the accuracy of the zero-velocity interval detection. Skog *et al.* presented a typical detector named the stance hypothesis optimal detection (SHOE), which achieve good performance for specific pedestrian and movement [16]. However, the parameters of ZUPT are quite different either under various motion states or from person to person. Thus, it is the key to adjust parameters adaptively for a robust pedestrian navigation system [4, 9]. Research work for solving the issue can be classified into two aspects. One is to improve the adaptability of the model-based algorithm by optimizing the parameters in real-time [1, 9, 15, 17–19]. The other is to replace the model-based architecture with a learned algorithm [4]. The adaptive algorithm, driven by the measurement of IMU, analyses the characteristics of parameters. The first is a threshold adjustment method based on speed or movement pattern classification [9, 16, 20]. For example, Brandon *et al.* train two separate support vector machine (SVM) classifiers for adaptive thresholds [4]. One to classify a user's motion type, and another to identify stationary periods given the current motion type. In addition, the model is constructed for zero-velocity detection based on the data characteristics [9]. Seong *et al.* propose a zero-velocity detection algorithm that does not require thresholds, by processing the gyro and accelerometer outputs based on an appropriate algorithm [19]. Different from the above, we present a result-derived algorithm to adjust the zero-velocity detection interval self-adaptively.

Unlike recursive estimation in filter-based frameworks, factor graph optimization estimates the states in a batch to achieve higher accuracy. The theory of inertial pre-integration makes it possible to provide scale information for other pose estimations for the IMU readings.

However, the pre-integration theory cannot be extended to pedestrians. At present, the pedestrian navigation framework based on optimization is based on the PDR algorithm to fuse other information sources. In terms of the use of information sources, posterity has done a lot of research. Researchers fusion GNSS [12, 21], WIFI-fingerprint [22], UWB [20], and other sensor [14, 20] information to optimize the pose of PDR. Compared with the SINS, the multi-sensors navigation based on PDR has drawbacks in accuracy and robustness. We analyse the error characteristics and propose a parameter adaptive adjustment method, which not only is useful for parameter adjustment but also provides support to combine optimization framework and filtering framework in theory.

2.2. Visual-Inertial Navigation. Due to the complementary nature of the IMU and vision, Motion estimation fusing cameras with IMUs has been an extensive research topic for many years. Noticeable approaches include MSCKF [2], VINS-Mono [3], SVO [6], DSO [7], and ORB-SLAM [5]. In this section, we will give a summary of visual-inertial

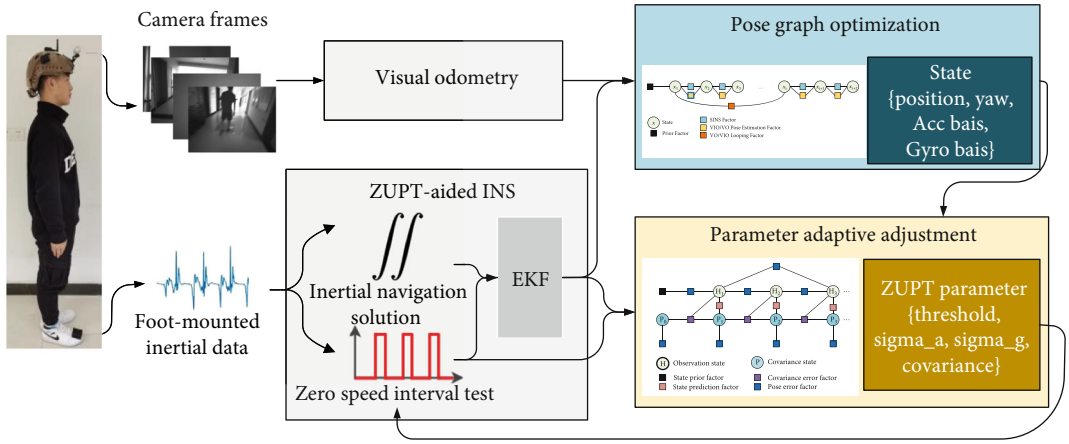

Figure 1: The conceptual diagram of the algorithm illustrating the full pipelines. MIMU and cameras are attached on the pedestrian. Pose graph optimization updates the 4-DOF pose and biases of IMU and parameter adaptive adjustment optimizes the parameters of zero-velocity detection with factor graphs.

odometry methods. We also focus the discussion on the application of visual inertia in pedestrians.

MSCKF [2], a tightly-coupled VIO algorithm based on EKF, uses visual measurements of the same feature across multiple camera views to form a multi-constraint update, in which there is not necessary to include the spatial position of feature points in the observation model. But the problem of inconsistent filter estimation also produces. Different from filtering-based algorithms, the energy minimization-based approaches overall optimize the posture. Lupton presented the theory of pre-integration to realize inertial vision integrated navigation in high dynamic conditions, with which the algorithm based on optimization can be realized. VINS-Mono [3] is a very accurate and robust monocular inertial odometry system, with loop-closing using DBoW2 and 4-DoF pose-graph optimization, and map-merging. Feature tracking is performed with Lucas-Kanade tracker, being slightly more robust than descriptor matching. ORB-SLAM [5] can close loops and reuse the map, which takes advantage of Bag-of-World. A 7-DOF pose graph optimization is followed by loop detection.

As mentioned earlier, visual-inertial fusion is an effective method to improve the accuracy of the navigation system for pedestrians. However, the installation position of the foot-mounted IMU limits the widespread research on visual-inertial fusion for pedestrians. Considering the pedestrian movement patterns and characteristics of sensors, we study a new method to fusion foot-mounted IMU and cameras [23].

2.3. Algorithm Framework. The block diagram of the proposed algorithm for pedestrian navigation is shown in Figure 2. The proposed approach mainly includes four main modules: zero-velocity-aided INS, odometry tracking, pose graph optimization, parameter adaptive adjustment. The first two, the basis of navigation system, provide initial pose estimation, respectively. The latter two are the core of the algorithm proposed. One is for pose fusion odometry with foot-mounted MIMU, and the other is for optimizing zero-velocity intervals derived by navigation results.

The algorithm starts with pose estimation. In the zero-velocity-aided INS module, IMU measurement is integrated to estimate the 6D pose in SINS. And zero-velocity measurements are fused with SINS in EKF to reduce error growth over time. Odometry tracking, based on VINS-Mono, estimates the pedestrian's pose incrementally evolves from the starting point. Unlike a fixed coordinate transformation relation between sensors, the module of pose graph optimization continuously optimizes the coordinate transformation matrix based on both VO/VIO pose estimate and SINS position output. And then parameters are adjusted adaptively according to the fuse position estimates in the parameter optimization module. Specifically, we quantify the influence of inertial navigation parameters on inertial navigation results with the analysis of errors in EFK, which lays a foundation for finding the most accurate zero-velocity intervals.

3. Pose Graph Optimization

3.1. Measurement Pre-Processing. For sensors fusion between cameras and foot-mounted IMU, we assume that there are similar position increments among sensors mounted on pedestrians at the same time. Considering the characteristics of movement for pedestrians, we update the pose according to the frequency of gaits with the moving average filtering of acceleration. As is shown in as in Figure 3, each negative peak is the starting point of each step, and the interval between the two negative peaks is a step.

The input data consist of camera images and IMU measurements. Both are not assumed to be synchronized. The pedestrian is a flexible body but has a similar position estimation in diverse parts of the body. Thus, sensors, different from sensor rigid connection, are attached to the pedestrian and have similar position and yaw at the same time. As is shown in Figure 4, we correct the synchronization by aligning visual keyframes with IMU measurements and the camera keyframes in accordance with the distance. And we think the keyframe has the same pose with the to the closest IMU measurement.

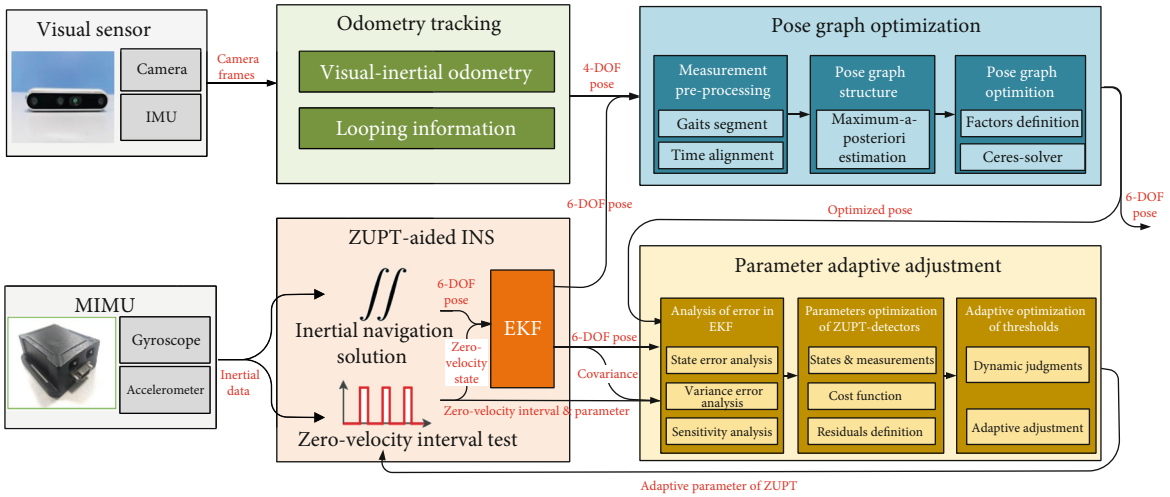

FIGURE 2: Algorithm framework illustrating the full pipelines. Odometry tracking outputs the odometry navigation results and looping information. In ZUPT-aided INS, zero-velocity measurements are fused with SINS in an Extended Kalman Filter to reduce error growth. Pose graph optimization continuously optimizes the pose graph based on both VO/VIO pose estimate and SINS position output. Parameter adaptive adjustment module first analyses the error between truth values and navigation results and construct optimization problem to solve accurate zero-velocity detection thresholds.

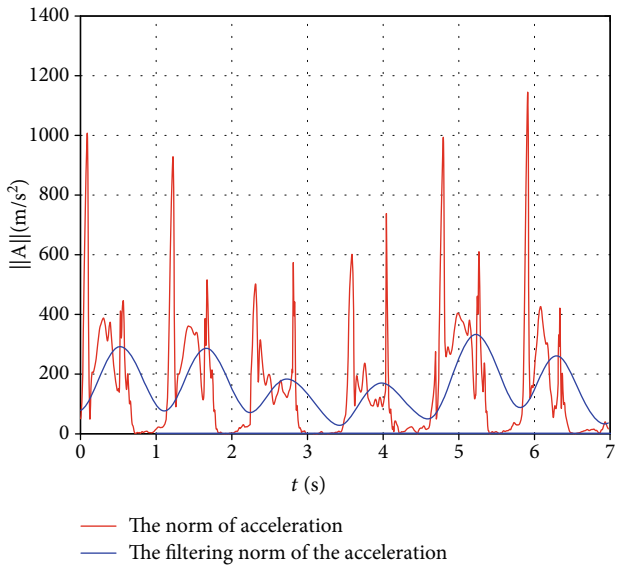

FIGURE 3: The norm and filtering norm of the accelerated velocity vector of IMU during various motions. Each negative peak is the starting of each step, and the interval between the two negative peaks is a step.

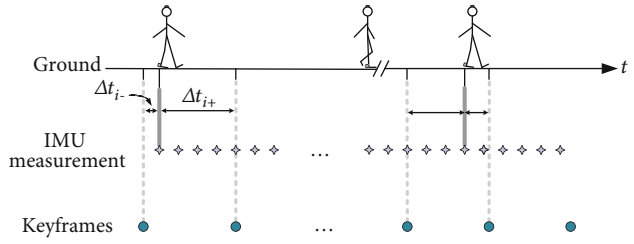

FIGURE 4: Time alignment. We consider camera/IMU time alignment to solve the problems of time mismatch.

3.2. Pose Graph Structure. Assuming that $k_g \in \mathbb{R}$ is the gait indices up to time t_g and there are a number of keyframes $\boldsymbol{C}_j$ between the adjacent foot tribal time $t_i$and t_{i-1} (i is the gait index, and $i \le k_g$), where $j \in \mathbf{m}_i$, $\mathbf{m}_i$ on behalf of the number of keyframes. We then define the objective of the estimation problem as the history of pedestrian states $\mathbf{x}_i$ and keyframes detected up t_g.

$$\chi_f \triangleq \bigcup_{i=0}^{k_g} \left[\{\mathbf{x}_i\}, \bigcup_{\forall j \in \mathbf{m}_i} \{\boldsymbol{C}_j\} \right] \tag{1}$$

The factor graph framework aims to find the most likely posterior state χ_f when given the history of measurements $\mathbf{z}_i$. As is shown in Figure 5, the nature of positioning problem is a Maximum A Posteriori (MAP) problem.

$$\chi_f^* = \arg \max_{\chi_f} \prod_{t=0}^{k_g} \prod_{\tau \in \mathbf{F}} P(\mathbf{z}_t^\tau \mid \chi)$$
$$\mathbf{F} = \{\text{imu, vio, prior state, landmarks}\} \tag{2}$$

$\mathbf{F}$ is the set of measurements, which includes VO/VIO measurements, MIMU measurements, prior state and landmarks, τ is the measurement's type. If the measurements are conditionally independent and corrupted by zero-mean Gaussian noise, the MAP estimate corresponds to the minimum of the negative log-posterior, and (2) is equivalent to a least squares problem of the form.

$$\chi_f^* = \arg \max_{\chi_f} \sum_{i=0}^{k_g} \sum_{\tau \in \mathbf{F}} \mathbf{r}_{\tau_i} \sum_{\tau_i}^{2} \tag{3}$$

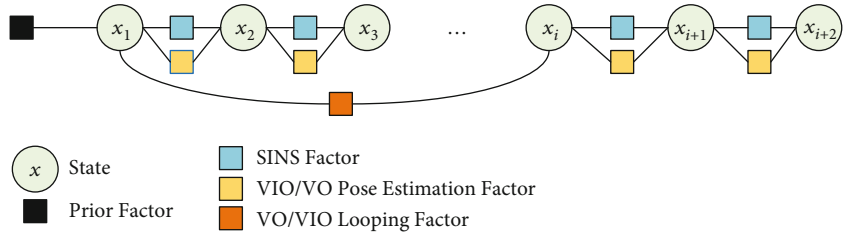

Figure 5: An illustration of the pose fusion graph structure. Every node represents the pedestrian's state in the world frame, which contains position yaw and IMU measurement bias. The edges between two consecutive nodes are constrained by both VO/VIO estimation and the INS pose increment. Others is looping constraint.

where $\mathbf{r}_{\tau_i}$ is the residual of the error between the predicted and measured value of at step index i, the quadratic cost of each residual is weighted by the corresponding covariance $\sum_{\tau_i}$.

The following equation minimizes the energy of the system as a whole:

$$\chi_f^* = \arg \min_{\chi_f} \left\{ \mathrm{r}_{0\Sigma_0}^2 + \sum_{(l_0,j)\in \boldsymbol{C}} \mathbf{r}_{\boldsymbol{C}_{j,0}}{}^2_{\Sigma_{\boldsymbol{C}_{j,0}}} + \sum_{i=0}^{k_g} \mathbf{r}_{I_i}{}^2_{\sum_{\tau_i}} + \sum_{(l,j)\in \boldsymbol{C}} \mathbf{r}_{\boldsymbol{c}_j}{}^2_{\sum_l^{\boldsymbol{C}_j} \#} \right\} \tag{4}$$

Where l is the landmark, l_0 is the prior landmark, and the residuals Σ_0, $\Sigma_{\boldsymbol{C}_{j,0}}$, $\sum\tau_i$, $\sum_l^{\boldsymbol{C}_j \#}$ are from: state prior, landmark prior, IMU factors, odometry factors.

3.3. Factors Definition

(1) State Prior Factors: In the proposed system, prior factors are used to anchor the pose to a fixed reference frame. The residual is defined as the error between the estimated state $\mathbf{x}_0$and the prior $\mathbf{z}_{p0}$

$$\mathbf{r}_0\left(\mathbf{x}_0, \mathbf{z}_{po}\right) = \begin{bmatrix} \mathbf{p}_0 - \mathbf{p}_{p0} \\ y_0 - y_{p0} \\ \mathbf{b}_0^w - \mathbf{b}_{p0}^w \\ \mathbf{b}_0^a - \mathbf{b}_{p0}^a \end{bmatrix} \tag{5}$$

where $\mathbf{p}_i$ and y_i (with $i \in \{0, , ,p, 0\}$) are position and yaw. $\mathbf{b}_i^w$ and $\mathbf{b}_i^a$ express the bias of gyro and accelerometer, respectively, at i. The prior state of the system is determined by IMU initialization.

(2) Landmark Prior Factors: The landmark prior residual $\mathbf{r}_{\boldsymbol{C}_{j,0}}$ is the error between the prior on the landmark location $\boldsymbol{C}_{j,o}$ and the estimated landmark location $\boldsymbol{C}_j$

$$\mathbf{r}_{\boldsymbol{C}_{j,0}}\left(\mathbf{x}_{\boldsymbol{C}_{j,0}}, \mathbf{z}_{\boldsymbol{C}_j}\right) = \mathbf{p}_{\boldsymbol{C}_{j,o}} - \mathbf{p}_{\boldsymbol{C}_j} \tag{6}$$

The landmark prior is generated online through an initial triangulation procedure in visual inertia odometer. The covariance $\Sigma_{\boldsymbol{C}_{j,0}}$ is determined by the triangulation accuracy.

(3) VIO Factors: Since focusing on the relative increment between step g_{t-1} and step g_t, we defifine the residual of VO/VIO factor $\mathbf{r}_{\boldsymbol{C}_j}$ as:

$$\mathbf{r}_{\boldsymbol{C}_j} = \mathbf{z}_t^c - \boldsymbol{h}_t^c(\chi) = \mathbf{z}_t^c - \boldsymbol{h}_t^c(\mathbf{x}_{t-1}, \mathbf{x}_t) = \begin{bmatrix} \mathbf{p}_t^c - \mathbf{p}_t^w \\ y_t^c - y_t^w \end{bmatrix} \tag{7}$$

where $\mathbf{p}_i^r$ and y_i^r (with $r \in \{c, w\}$) is position at time i (with i as t, $t-1$) in the odometer or the global pose estimator. The covariance for VO/VIO measurements, is determined by the estimation accuracy, which is influenced by environmental conditions, pedestrian dynamics, etc. In our case, we adjust if according to the experimental conditions.

(4) SINS Factors: Raw measurements of SINS are position increment and quaternion in the Inertial Coordinate System. In order to fusion the pose with the VIO, we set the positions and yaw as the optimized state, and considering the roll and pitch as the global posture. Generally, knowing the longitude and latitude at the origin point, we can convert them into the Earth Coordinate System. The IMU measurement is obtained according to the SINS. The IMU factor is derived as:

$$\mathbf{r}_{I_i} = \mathbf{z}_i^I - \boldsymbol{h}_i^I(\chi) = \mathbf{z}_i^I - \boldsymbol{h}_i^I(\mathbf{x}_{i-1}, \mathbf{x}_i) = \begin{bmatrix} \mathbf{p}_i^I - \mathbf{p}_i^w \\ y_i^I - y_i^w \\ \mathbf{b}_i^a - \mathbf{b}_{i-1}^a \\ \mathbf{b}_i^w - \mathbf{b}_{i-1}^w \end{bmatrix} \tag{8}$$

where the couple $(\mathbf{q}_i^I, y_i^I)$ is position and orientation at time i in the SINS based on ZUPT. And the couple $(\mathbf{q}_i^w, y_i^w)$ represent the pose of the system. The covariance is determined by performance of IMU devices and accuracy of zero speed detection.

4. Adaptive Parameter Adjustment in INS

ZUPT-aided INS contains two parts: one obtains increment of the pose with inertial integral, the other correct navigation error and measurement bias based on the zero-velocity interval detection and EKF. If correctly identified, zero-velocity updates can significantly improve localization estimates. However, either false-positive or false-negative detections bring observation error to EKF, thus lead to rapid and unbounded error growth. In fact, the error.

between truth values and navigation results is available to analyse the performance of EKF. In practice, we assume navigation errors attribute to inaccurate observations and estimate the length offset of the zero-velocity interval with the navigation error. In addition, variance error analysis is a very important method to measure the estimation of state error. Although there is indeed a consistency deviation between the estimated variance error and the true error. However, we can continuously enhance the consistency because the inconsistency of variance error will be fed back to the state error, and the algorithm is driven by state error. in other words, the difference the estimated variance error and the true error can be measured indirectly and gradually decreases with the correction of the state error.

As is identified following Figure 6, this section presents an optimization algorithm based on error analysis to update parameters of INS. We derive the recurrence formula of state error between two adjacent steps. Then the offset of interval length is translated to the change of observation. Additionally, observations, lengths of the zero-velocity interval, are optimized with a factors graph. Finally, the threshold is updated adaptively in a sliding window.

4.1. Error Analysis in EKF. Under the case of knowing the truth system, the covariance matrix can be propagated based on the system state error in EKF-aid INS, whether the model parameters or the mean square error are inaccurate. For pedestrian navigation, we assume that navigation offset is caused by inexact observation, that is, the zero-velocity interval error $\Delta\mathbf{H}_k$ at the $k-$th gate.

Knowing the truth value $\mathbf{X}_k^r$ and the system output $\widehat{\mathbf{X}}_k$, the state error $\widehat{\mathbf{X}}_k^e$ is defined as follows:

$$\tilde{\mathbf{X}}_k^e \triangleq \mathbf{X}_k^r - \widehat{\mathbf{X}}_k \tag{9}$$

As is our concerned, the offsets of the zero-velocity interval are the only source of a navigation error. Therefore, $\widehat{\mathbf{X}}_k^e$ can be abbreviated as,

$$\begin{aligned}\widehat{\mathbf{X}}_k^e &= (\mathbf{I}-\mathbf{K}_k\mathbf{H}_k)\tilde{\mathbf{X}}_{k|k-1}^p - \mathbf{K}_k\Delta\mathbf{H}_k\mathbf{X}_k^r - \mathbf{K}_k\mathbf{V}_k^r \\ \tilde{\mathbf{X}}_{k|k-1}^p &= \mathbf{\Phi}_{k,k-1}\tilde{\mathbf{X}}_{k-1}^p + \mathbf{\Gamma}_{k-1}^r\mathbf{W}_{k-1}^r\widehat{\mathbf{X}}_{k-1}\end{aligned} \tag{10}$$

where $\tilde{\mathbf{X}}_{k|k-1}^p$ is the prediction mean square error, $\mathbf{K}_k\mathbf{H}_k$ is behalf of filter gain and measurements. $\mathbf{W}_{k-1}^r$represents the noise sequence. Besides, $\mathbf{\Phi}_{g|g-1}$and $\mathbf{\Gamma}_{g|g-1}$ are the state transition matrix in EKF at time t_g. According to the definition of variance, the variance error is formulated as,

$$\begin{aligned}\mathbf{P}_k^P &= (\mathbf{I}-\mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}^P(\mathbf{I}-\mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}} + \mathbf{K}_k\Delta\mathbf{H}_k\mathbf{A}_k\Delta\mathbf{H}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}} \\ &\quad + \mathbf{K}_k\mathbf{R}_k^r\mathbf{K}_k^{\mathrm{T}} - (\mathbf{I}-\mathbf{K}_k\mathbf{H}_k)\mathbf{C}_{k,k-1}^{\mathrm{T}}\Delta\mathbf{H}_k^T\mathbf{K}_k^{\mathrm{T}} \\ &\quad - \mathbf{K}_k\Delta\mathbf{H}_k\mathbf{C}_{k|k-1}(\mathbf{I}-\mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}} \\ \mathbf{P}_{k|k-1}^P &= \mathbf{\Phi}_{k,k-1}\mathbf{P}_{k-1}^P\mathbf{\Phi}_{k,k-1}^{\mathrm{T}} + \mathbf{\Gamma}_{k-1}^r\mathbf{Q}_{k-1}^r\mathbf{\Gamma}_{k-1}^r\end{aligned} \tag{11}$$

Where $\mathbf{R}_g$ is the state matrix in EKF and

$$\begin{aligned}\mathbf{A}_k &= E\left[\mathbf{X}_k^r\mathbf{X}_k^{r\mathrm{T}}\right] \\ \mathbf{C}_k &= E\left[\mathbf{X}_k'\tilde{\mathbf{X}}_k^{P\mathrm{T}}\right] \\ \mathbf{C}_{k|k-1} &= E\left[\mathbf{X}_k^r\tilde{\mathbf{X}}_{k|k-1}^{P\ \mathrm{T}}\right]\end{aligned} \tag{12}$$

In our system, we take the last state as the Initial-values. And initial values of the mean square error are as followed.

$$\begin{aligned}\mathbf{P}_0^P &= E\left[\left(\mathbf{X}_{k-1}^r - \widehat{\mathbf{X}}_{k-1}\right)\left(\mathbf{X}_{k-1}^r - \mathbf{X}\wedge_{k-1}\right)^{\mathrm{T}}\right] \\ \mathbf{A}_0 &= E\left[\mathbf{X}_{k-1}^r\mathbf{X}_{k-1}^{r\ \mathrm{T}}\right] \\ \mathbf{C}_0 &= E\left[\mathbf{X}_{k-1}^r\left(\mathbf{X}_{k-1}^r - \mathbf{X}\wedge_{k-1}\right)^{\mathrm{T}}\right]\end{aligned} \tag{13}$$

We assume that the pedestrian is equipped with a set of multi-rate sensors, with IMU sensors typically producing measurements at high rate and sensors such as monocular or stereo cameras generating measurements at lower rates. Some sensors may become inactive from time to time (*e.g.* GPS), while others may be active only for short periods of time (e.g. signal of opportunity).

4.2. Parameters Optimization of Zero-Velocity Detectors via Factor Graphs. In ZUPT-aided INS, navigation error is suppressed with both state covariance matrix and the zero-velocity observation. We assume that the truth value is known form another available information sources and set the state quantity of the previous time as the initial value of each step approximately. Our goal is to calculate the best INS parameter by fusing all the navigation error.

(1) State Definition: The state the state of the ZUPT-aided system at gate t_g as:

$$\begin{aligned}\mathbf{x}_g^e &= \left[\mathbf{P}_g^e, , , \mathbf{H}_g^e\right] \\ \mathbf{P}_g^e &= \left[P_{g,x}^e, P_{g,y}^e, P_{g,z}^e, P_{g,h}^e\right] \\ \mathbf{H}_g^e &= \left[\Delta v_{g,x}^e, \Delta v_{g,y}^e, \Delta v_{g,z}^e\right]\end{aligned} \tag{14}$$

where the couple $[\mathbf{P}_g^e, , , \mathbf{H}_g^e]$ represents covariance error and measurement offset at the g-th step (with time t_g), respectively.

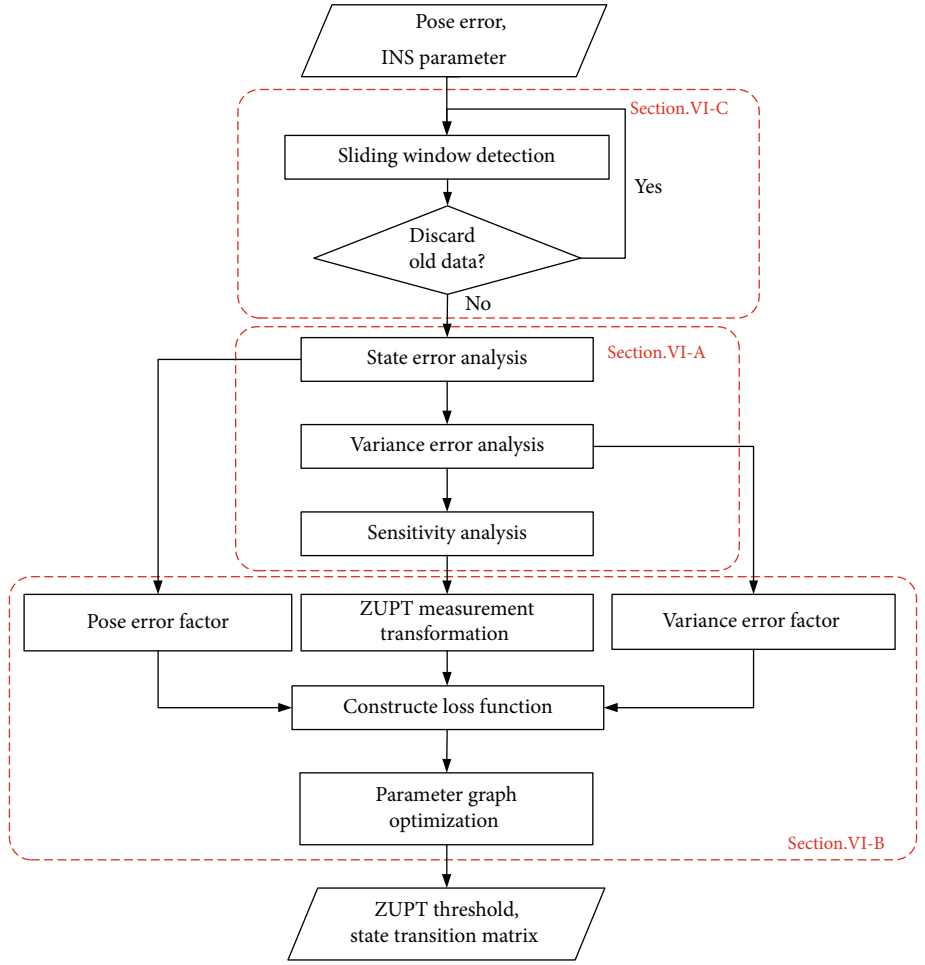

Figure 6: The program flow chart of parameter adaptive optimization.

$P^e_{g,i}$ is the state i ($i \in \{x, , ,y, , ,z, , ,h, e, a, d, i, n, g\}$) variance at the g-th step. Let $\Delta v^e_{g,j}$ be the error of velocity measurement in orientation j.

With the help of we transform the disturbance of zero velocity interval into the change of the observable measurement in one step.

$$\mathbf{H}^e_g = \Delta \boldsymbol{l}_g \boldsymbol{r}_g \mathbf{T}^v_g \tag{15}$$

Where $\mathbf{T}^v_g$ is the Observation transformation matrix and $\Delta \boldsymbol{l}_g$ is the disturbance of zero-velocity interval length in gait g-th, we then define the objective of our estimation problem $\mathbf{X}_k$ as the history of robot states and landmarks detected up to t_g:

$$\mathbf{X}_k \triangleq \bigcup_{t=0}^{t_g} \mathbf{X}^e_g \tag{16}$$

(2) Measurements: The input measurements consist of the pose error $\mathbf{X}^e_k$ and covariance prediction $\mathbf{P}^e_k$, both of which are from truth values of system. k is the time index

$$\begin{aligned} \mathbf{X}^e_k &= \mathbf{X}^e_k - \widehat{\mathbf{X}}_e \\ \mathbf{P}^e_k &= \mathbf{P}^e_k - \widehat{\mathbf{P}}_e \end{aligned} \tag{17}$$

The pose input is the error between navigation results and truth values, and covariance prediction comes from propagation of the pose error in EKF.

(3) Maximum-A-Posteriori Estimation: We assume the uncertainty of measurements is Gaussian distribution with mean and covariance. As is shown in Figure 7, the following equation minimizes the sum of prior and the Mahalanobis norm of all measurement residuals to obtain a maximum posterior estimation in a whole

$$\begin{aligned} \chi^*_p &= \arg \max_{\chi_p} \prod_{t=0}^{K_g} \prod_{k \in \mathbf{S}} P\left(\mathbf{z}^k_t \mid \chi\right) = \arg \min_{\chi_p} \sum_{\tau} \sum_{i \in K_g} \mathbf{r}^2_{ri} \sum\nolimits^{-1}_{\tau_i} \\ &= \arg \min_{\chi} \left\{ \mathbf{r}_p - \mathbf{H}_p \chi^2 + \sum_{k \in \mathcal{P}} \mathbf{r}_{\mathcal{P}} \left(\widehat{\mathbf{z}}^{b_k}_{b_{k+1}}, \chi\right)^2_{\mathbf{P}^{b_k}_{b_{k+1}}} + \sum_{k \in \mathcal{B}} \mathbf{r}_{\mathcal{B}} \left(\widehat{\mathbf{z}}^{b_k}_{b_{k+1}}, \chi\right)^2_{\mathbf{P}^{b_k}_{b_{k+1}}} \right\} \end{aligned} \tag{18}$$

Where $\mathbf{r}_p$, $\mathbf{r}_{\mathcal{B}}$ represent the residual of pose estimation and

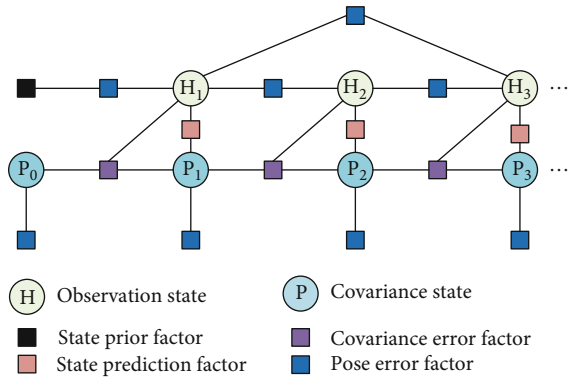

FIGURE 7: An illustration of the parameter graph structure. Every node represents one gait's parameter of pose filtering in ZUPT-aided INS. The node contains the length of zero-velocity interval and states covariances.

state variance, respectively. Detailed definition of the residual terms will be presented later in VI-D. Once the graph is built, optimizing it equals to finding the configuration of nodes that match all edges as much as possible.

We used Ceres Solver to carry out the nonlinear optimization. We run pose graph optimization at the frequency of gait update for a pedestrian. After every optimization, we obtain more accurate estimations fusing the SINS pose with VO/VIO.

(4) State Variance Residual: In contrast to the stepwise recursion of the state covariance matrix, we define the state variance residual with a state covariance error, which is constrained only by two adjacent observations

Consider the EKF state within two consecutive time t_g and t_g, according to analysis of variance error in Sect. V-A, the residual for can be defined as:

$$\mathbf{r}_{\mathscr{V}}\left(\hat{\mathbf{z}}_{t_{g+1}}^{t_g}, \chi_p\right) = \mathbf{P}_g^e \tag{19}$$

(5) Pose Error Residual: We establish a model to describe the influence of zero velocity interval disturbance on pose estimation in a gait. Based on the model, the relative pose constraint is produced

$$\mathbf{r}_{\mathscr{V}}\left(\hat{\mathbf{z}}_{t_{g+1}}^{t_g}, \chi_p\right) = \begin{bmatrix} \delta x_{p_{g+1}}^{p_g} \\ \delta y_{p_{g+1}}^{p_g} \\ \delta z_{P_g+1}^{p_g} \\ \delta \omega_{p_{g+1}}^{p_g} \end{bmatrix} = \mathbf{z}_g^{\mathscr{P}} - \mathbf{h}_g^{\mathscr{P}}\left(\chi_p\right) \tag{20}$$

$$\mathbf{h}_g^{\mathscr{P}}\left(\chi_p\right) = \tilde{\mathbf{X}}_k^e$$

where $\mathbf{h}_g^{\mathscr{P}}(\chi_p)$ extracts the pose offset at the state χ_p.

4.3. Adaptive Optimization of Zero-Velocity Thresholds. We keep a sliding window for graph optimization to get drift-free pose estimation and avoid wrong adjustments when the movement pattern changes. The size of the window is adjusted in real-time with the computation complexity. Additionally, old poses and measurements will be thrown when the motion pattern changed.

We judge whether the movement pattern changes according to positive peaks Within two adjacent steps. The movement pattern is considered to have changed if the ratio of peak values above is beyond the threshold adaptive adjustment rang.

5. Experiment Results

We perform real-world experiments to evaluate the proposed VA-INS system from two aspects in accuracy and robustness. In the indoor environment, which has a dynamic and small viewing field, we test the performance of the algorithm in dynamic environment. We then carry out an outdoor experiment with multi-motions pattern to test the performance of the real-time optimized INS. Additionally, A large-scale experiment is carried out to illustrate the long-time practicability of our system.

The performance of sensors is described in detail in Table 1. The sensor suite contains a foot-mounted MIMU (with gyroscope and accelerometer) operating at 400 Hz, a stereo camera (Intel Realsense D455) with 30 Hz, and the u-blox GNSS modules. As is shown in Figure 8, sensors are connected to the pedestrian but do not have a fixed coordinate relationship.

We get the ground truth with different methods between indoor and outdoor experimental conditions. In indoor experiments, ground truth mark points, either obvious turning points or the end of each step, are used to evaluate the experimental effect. Specifically, the subject pressed a handheld trigger that recorded a timestamp to facilitate temporal alignment with ground truth when he arrived at the mark points. In outdoor experiments, the ground truth, which is calculated by the Differential GPS on the u-blox-NEO-M8N. The position accuracy of Differential GPS is about 0.1 m. Also, google maps are used for error judgment in large-scale experiments. The performance is evaluated by the horizontal position error (HPE) of trajectories. Sensors are connected to the pedestrian but do not have a fixed coordinate relationship.

We get the ground truth with different methods between indoor and outdoor experimental conditions. In indoor experiments, ground truth mark points, either obvious turning points or the end of each step, are used to evaluate the experimental effect. Specifically, the subject pressed a handheld trigger that recorded a timestamp to facilitate temporal alignment with ground truth when he arrived at the mark points. In outdoor experiments, the ground truth, which is calculated by the Differential GPS on the u-blox-NEO-M8N. The position accuracy of Differential GPS is about 0.1 m. Also, google maps are used for error judgment in large-scale experiments. The performance is evaluated by the horizontal position error (HPE) of trajectories.

Table 1: Results in the multi-motion experiment outdoor.

Sensor	Parameter	Index
Gyroscope	Bias stability	<8°/h
	Angle random walk	0.36°/h
	Sampling frequency	400 Hz
Accelerometer	Bias stability	0.03 mg
	Random walk	0.045 m/s/h^0.5
	Sampling frequency	400 Hz
Camera	Image resolution ratio	1280*720
	Baseline	36.2625 mm
	Sampling frequency	30 Hz
U-blox	Navigation sensitivity	−167 dBm
	Sampling frequency	10 Hz

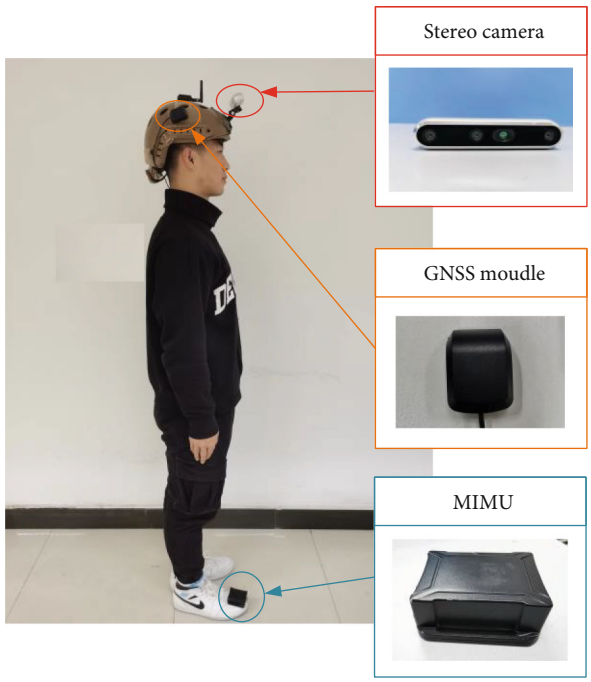

Figure 8: The sensor suite we used for the experiment, which contains ta foot-mounted MIMU and a stereo camera (Intel Realsense D455).

In these experiments, we compare the algorithm proposed with both VINS-Mono and ZUPT-aided INS. VINS-Mono is a robust and versatile monocular visual-inertial state estimator. ZUPT-aided INS uses the SHOE detector to discover the zero-velocity interval and takes zero velocity as the virtual observation of the filtering algorithm to modify the INS results with the Extended Kalman Filter.

6. The Hallway Experiment Indoor

In the hallway experiment indoor, we choose our laboratory environment as the experiment area. The test subject suits sensors and walks at a normal pace in the hallway of the laboratory. During the trial, the subject went along hallways and rooms with the same stride, and returned to the origin along the same path.

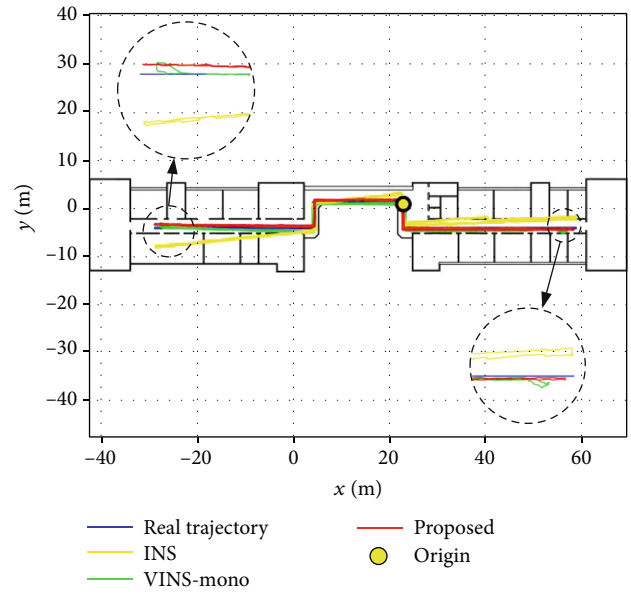

Figure 9: The trajectories of our indoor hallway experiment recovered from real trajectory, ZUPT-aided INS, VINS-Mono and the proposed algorithm, respectively.

Table 2: Results in the mu Hallway Experiment Indoor.

Algorithm	VINS-mono	SHOE	VA-INS
MEAN[m]	0.60	2.01	**0.36**
RMSE[m]	0.69	2.61	**0.39**

From Figure 9, we can see that heading error of INS accumulates over time and the visual odometer has less pose drift than zero-velocity INS in the environment of unrestricted vision. Most notably, VA-INS proposed continuously improves the accuracy of navigation. Table 2 shows the RMSE (Root Mean Square Errors) and MEAN (Mean Errors) of HPE. In the indoor environment with good visual conditions, the performance of a visual-inertial odometer is better than that of pure inertia. Especially, VINS-Mono with loop detection suppresses the drift of course. It is worth noting that VA-INS combines the advantages of the two and achieves the best performance.

7. The Visual-Restricted Experiment

In the mixed experiment of stairs and corridors, the trial walk along the corridors and walk up the stairs from the second floor to the fifth floor, where he encounters pedestrians, low light condition, texture-less area, glass, and reflection. Then, the trial climb down the stairs and walk back to the origin. Key-frames of typical scenes are shown in Figure 10.

As shown in Figure 11, we compare our results with VINS-Mono and fixed-threshold INS. Noticeable VIO drifts

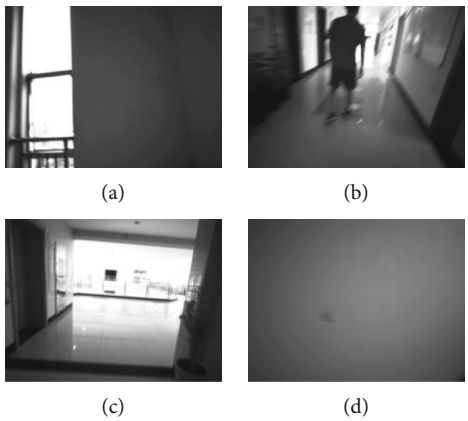

Figure 10: B Experimental key-frames of the camera for pose estimation in VIO, which consist of glass/reflection(a), pedestrian(b), high light condition(c), texture-less area (d).

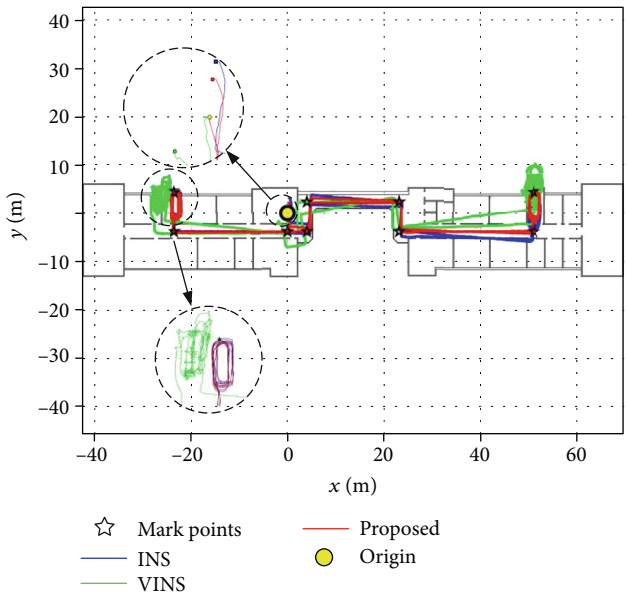

Figure 11: The trajectories of our indoor visual-restricted experiment recovered from real trajectory, ZUPT-aided INS, VINS-Mono and the proposed algorithm, respectively.

Table 3: Results in visual-restricted Experiment indoor.

Algorithm	VINS-mono	SHOE	VA-INS
MEAN[m]	7.62	3.51	**2.38**
RMSE[m]	7.89	3.74	**2.57**

occurred when the experimenter encounters the dynamic objects, and the system is even unavailable when the experimenter climbed up and down the stairs, where texture-less area is in all places. In sharp contrast, although ZUPT-aided INS has the cumulative error with time, it is more robust to changes in experimental scenes. In contrast to them, the fusion foot-mounted IMU with cameras suppresses the error accumulation and improves the robustness by exerting their respective advantages.

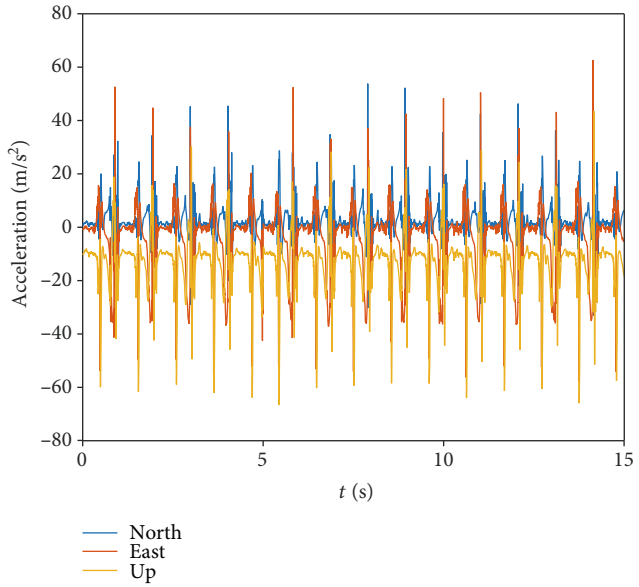

Figure 12: Partial acceleration measurements of foot-mounted MIMU in multi-motion experiment outdoor. For MIMU, there is different characteristics in various motion states.

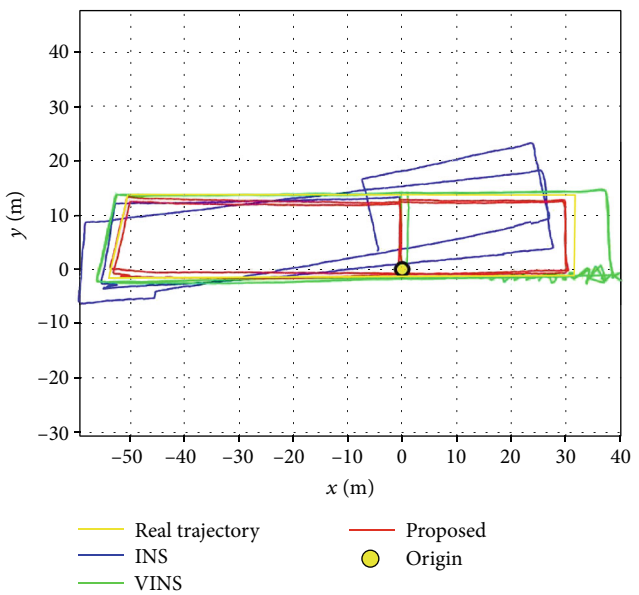

Figure 13: The trajectories of our multi-motion experiment recovered from real trajectory, ZUPT-aided INS, VINS-Mono and the proposed algorithm, respectively.

Table 3 shows the RMSE (Root Mean Square Errors) and MEAN (Mean Errors) of HPE. The robustness of the visual odometer is poor in the indoor environment with limited visual conditions, where the error of INS accumulates with time but is limited. VA-INS uses the pose estimation of VINS-Mono in a good environment to correct the error of INS. Combining the trials above, we can see that the visual odometer is poor robust in the dynamic environment and our algorithm outperforms VINS-Mono and ZUPT-aided INS, which demonstrates the algorithm proposed effectively

Table 4: Results in the multi-motion experiment outdoor.

Algorithm	VINS-mono	SHOE	Proposed
MEAN[m]	5.56	2.54	0.89
RMSE[m]	6.12	3.16	0.91

Table 5: Results in the large-scale experiment.

Algorithm	VINS-mono	SHOE	Proposed
MEAN[m]	13.36	18.75	**5.58**
RMSE[m]	21.73	26.04	**7.53**

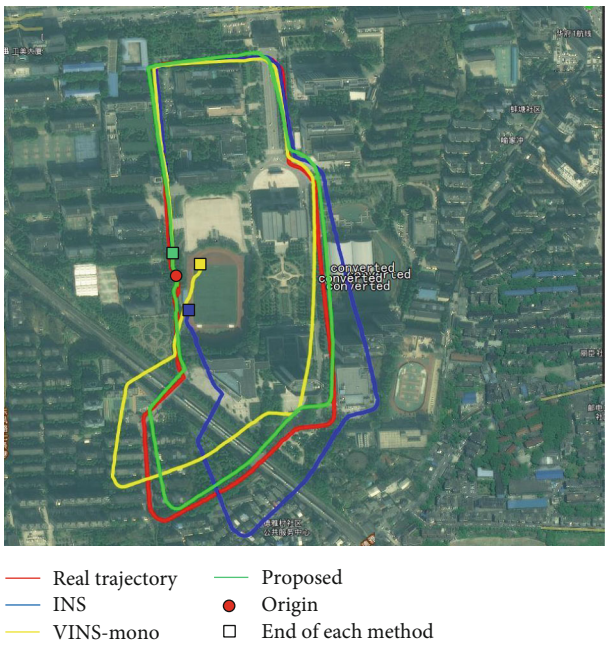

Figure 14: The trajectories of the large-scale experiment outdoor recovered from ZUPT-aided INS, VINS-Mono, and the proposed algorithm, respectively.

reduces the influence of experimental scene on vision and improve navigation accuracy.

7.1. The Multi-Motion Experiment Outdoor. We evaluated our proposed detector over longer trajectories by carrying out a multi-motions trial in an outdoor setting. The trial subject started and ended at the same position on the test course. The experimenter started from the doorway of the building and walked along the building. Then the trial instructed to alternate between jogging, walking, and fast running. Finally, we went back to the building and returned to the origin. The whole trajectory is more than 450 meters and lasts approximately eight minutes. The IMU measurement of multi-motion states can be seen in Figure 12.

Figure 13 shows trajectory comparison is shown. The RMSE (Root Mean Square Errors).

and MEAN (Mean Errors) of HPE is shown in Table 4. Due to the low adaptability of the threshold for various states, we can see obvious translation drift in the estimated trajectory of INS. What's more, VINS-Mono is almost no heading deviation but there are offsets in position estimation when the trial jogs or runs. From the trajectory comparison, we can see that the proposed system improved the accuracy of INS a lot with the real-time adaptive parameters. The algorithm proposed achieved the best performance.

7.2. A Large-Scale Experiment. We have chosen a campus environment as the experimental scene to verify the improvement in the long-range positioning performance. In this large-scale scene test, the trial distance is 2.2 km with about 164028 min.

As is shown in Table 5. The RMSE of HPE in the proposed algorithm is 7.53 m. Under the same test conditions, results by the fixed-threshold INS and VINS-Mono are 26.04 m and 21.73 m. Similarly, the mean error of our algorithm is significantly lower than the others. It is obvious that the algorithm proposed provides accurate positioning even in the large-scale scene.

The estimated trajectory is aligned with Google Map in Figure 14. It can be seen from the figure that the visual loop can correct the navigation error. However, the overall optimization of the trajectory may also affect the overall pose estimation accuracy. Compared with Google Map, we can see our results are almost drift-free in this very long-duration test.

8. Conclusions

In this work, we have proposed a novel visual-aid inertial navigation system for pedestrians with a detailed description of its building blocks and an exhaustive evaluation. The approach could increase the accuracy of pose estimation with flexible-connection sensors fusion and optimize the parameter of INS. We establish the functional relationship between the zero-speed interval disturbance and the navigation results, which is the basis of our work. Using the factor graph, the visual odometer is fused with the foot-mounted MIMU to overcome the error drift in inertial navigation results. Then the fusion-optimized pose is taken as the observation to optimize the zero-velocity interval of each step, which further updates the zero-velocity detection threshold of the pedestrian in the current motion state. We show superior performance by comparing against both the pedestrian inertial navigation algorithm based on fixed threshold and the typical visual-inertial fusion algorithm. Future work will extend the method to fusion other sensors for more accurate pose estimation, such as GPS, WIFI- fingerprint, UWB, and Magnetometer. The goal is to not only further improve the accuracy of pose estimation but also realize the plug-and-play of sensor fusion for pedestrian navigation.

Data Availability

The visual inertia raw data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author(s) declare(s) that they have no conflicts of interest.

Funding

This work is supported in part by National Natural Science Foundation (NNSF) of China under Grant 61773394 and 62073331.

References

[1] Z. Wang, H. Zhao, S. Qiu, and Q. Gao, "Stance-phase detection for ZUPT-aided foot-mounted pedestrian navigation system," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 6, pp. 3170–3181, 2015.

[2] A. I. Mourikis and S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, pp. 3565–3572, Rome, Italy, 2007.

[3] 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.

[4] B. Wagstaff, V. Peretroukhin, and J. Kelly, "Robust data-driven zero-velocity detection for foot-mounted inertial navigation," *IEEE Sensors Journal*, vol. 20, no. 2, pp. 957–967, 2020.

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

[6] C. Forster, Z. Zhang, M. Gassner, M. Werlberger, and D. Scaramuzza, "SVO: Semidirect visual Odometry for monocular and multicamera systems," *IEEE Transactions on Robotics*, vol. 33, no. 2, pp. 249–265, 2017.

[7] J. Engel, V. Koltun, and D. Cremers, "Direct sparse Odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 611–625, 2018.

[8] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[9] R. Zhang, H. Yang, F. Hoflinger, and L. M. Reindl, "Adaptive zero velocity update based on velocity classification for pedestrian tracking," *IEEE Sensors Journal*, vol. 17, no. 7, pp. 2137–2145, 2017.

[10] C. Yu, Z. Liu, X. J. Liu et al., "DS-SLAM: a semantic visual SLAM towards dynamic environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1168–1174, Madrid, Spain, 2018.

[11] T. Lupton and S. Sukkarieh, "Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions," *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 61–76, 2012.

[12] W. Li, X. Cui, and M. Lu, "A robust graph optimization realization of tightly coupled GNSS/INS integrated navigation system for urban vehicles," *Tsinghua Science and Technology*, vol. 23, no. 6, pp. 724–732, 2018.

[13] 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*, pp. 2154–2161, 2012.

[14] J. Tan, X. Fan, S. Wang, and Y. Ren, "Optimization-Based Wi-Fi Radio Map Construction for Indoor Positioning Using Only Smart Phones," *Sensors*, vol. 18, no. 9, p. 3095, 2018.

[15] 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.

[16] I. Skog, J. Nilsson, and P. Händel, "Evaluation of zero-velocity detectors for foot-mounted inertial navigation systems," in *2010 International Conference on Indoor Positioning and Indoor Navigation*, pp. 1–6, Zurich, Switzerland, 2010.

[17] J. Wahlström, I. Skog, F. Gustafsson, A. Markham, and N. Trigoni, "Zero-velocity detection—a Bayesian approach to adaptive thresholding," *IEEE Sensors Letters*, vol. 3, no. 6, pp. 1–4, 2019.

[18] M. Ma, Q. Song, Y. Li, and Z. Zhou, "A zero velocity intervals detection algorithm based on sensor fusion for indoor pedestrian navigation," in *2017 IEEE 2nd information technology, networking, Electronic and Automation Control Conference (ITNEC)*, pp. 418–423, Chengdu, China, 2017.

[19] S. Y. Cho and C. G. Park, "Threshold-less zero-velocity detection algorithm for pedestrian dead reckoning," in *2019 European Navigation Conference (ENC)*, pp. 1–5, Warsaw, Poland, 2019.

[20] F. Liu, J. Wang, J. Zhang, and H. Han, "An Indoor Localization Method for Pedestrians Base on Combined UWB/PDR/Floor Map," *Sensors*, vol. 19, no. 11, p. 2578, 2019.

[21] K. Pan, M. Ren, P. Wang, and Y. Liu, "A federated filtering personal navigation algorithm based on MEMS-INS/GPS integrated," in *2016 Chinese Control and Decision Conference (CCDC)*, pp. 5237–5241, Yinchuan, China, 2016.

[22] M. Nowicki and P. Skrzypczynski, "Indoor Navigation with a Smartphone Fusing Inertial and WiFi Data via Factor Graph Optimization," in *Mobile Computing, Applications, and Services*, S. Sigg, P. Nurmi, and F. Salim, Eds., vol. 162 of Lecture Notes of the Institute for Computer Sciences, Social Informatics, and Telecommunications Engineering, , pp. 280–298, Springer, 2015.

[23] L. An, X. Pan, Z. Chen, M. Wang, Z. Tu, and C. Chu, "A multi-sensor fusion algorithm for pedestrian navigation using factor graphs," in *2021 40th Chinese Control Conference (CCC)*, pp. 3727–3732, Shanghai, China, 2021.