Free-walking: Pedestrian inertial navigation based on dual foot-mounted IMU

Qu Wang [a, b, 1], Meixia Fu [a], Jianquan Wang [a, *], Lei Sun [a], Rong Huang [c], Xianda Li [c], Zhuqing Jiang [d], Yan Huang [e], Changhui Jiang [f]

[a] School of Automation Science and Electrical Engineering, University of Science and Technology Beijing, Beijing, 100083, China
[b] Shunde Innovation School, University of Science and Technology Beijing, Fo Shan, 528399, China
[c] Research Institute of China Unicom, Beijing, 100190, China
[d] School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, 100876, China
[e] Department of Automation Tsinghua University, Beijing, 100084, China
[f] GEOLOC Laboratory, Université Gustave Eiffel, Paris, 77454, France

ARTICLE INFO

ABSTRACT

The inertial navigation system (INS), which is frequently used in emergency rescue operations and other situations, has the benefits of not relying on infrastructure, high positioning frequency, and strong real-time performance. However, the intricate and unpredictable pedestrian motion patterns lead the INS localization error to significantly diverge with time. This paper aims to enhance the accuracy of zero-velocity interval (ZVI) detection and reduce the heading and altitude drift of foot-mounted INS via deep learning and equation constraint of dual feet. Aiming at the observational noise problem of low-cost inertial sensors, we utilize a denoising autoencoder to automatically eliminate the inherent noise. Aiming at the problem that inaccurate detection of the ZVI detection results in obvious displacement error, we propose a sample-level ZVI detection algorithm based on the U-Net neural network, which effectively solves the problem of mislabeling caused by sliding windows. Aiming at the problem that Zero-Velocity Update (ZUPT) cannot suppress heading and altitude error, we propose a bipedal INS method based on the equation constraint and ellipsoid constraint, which uses foot-to-foot distance as a new observation to correct heading and altitude error. We conduct extensive and well-designed experiments to evaluate the performance of the proposed method. The experimental results indicate that the position error of our proposed method did not exceed 0.83% of the total traveled distance.

1. Introduction

With the increasing need for indoor location-based service (LBS) in various scenarios and sectors, high-precision and highly robust LBS solutions have become a research hotspot [1]. Location awareness is emerging as an Internet of Things (IoTs)-enabled technology and an essential part of the ambient assisted living framework [2]. Global Navigation Satellite Systems (GNSS) have provided millimeter to centimeter level positioning in outdoor environment [3]. Due to the obstruction of satellite signals, it is difficult for global navigation satellite systems to obtain accurate positioning results in indoor environments [4]. Therefore, indoor positioning systems built on wireless signals such as cellular [5,6], Radio Frequency Identification (RFID) [7], ZigBee, LiDAR [8], Ultra-WideBand (UWB) [9], Bluetooth [10], and Wi-Fi [11] have been studied and implemented extensively. The unreliable variations in Wi-Fi signals, such as the complicated signal multipath problem, limit its application. Vision-based indoor localization methods achieve high-precision localization results, but privacy and security issues limit their application [12]. In addition, issues such as high computational overhead limit its application to embedded devices [13]. Visible light-based localization methods achieve high localization accuracy in a small range, but sunlight affects the robustness of the localization system [14]. In summary, the existing indoor positioning technologies have their strengths. However, they all

have problems such as high deployment costs, small coverage, poor robustness, and low positioning accuracy, which cannot fully meet the high-precision LBS requirements in various application.

With the rapid development of microelectromechanical system (MEMS) technology, MEMS-based inertial navigation systems (INS) play a vital role in providing seamless indoor and outdoor LBS for arbitrary unprepared and unfamiliar environments. Compared with other navigation and positioning methods, pedestrian inertial navigation technology based on MEMS sensors does not need to deploy base stations in advance, and only relies on the sensors they carry, such as accelerometers, gyroscopes and magnetometers, to complete pedestrians' navigation and positioning [15]. Inertial navigation has the characteristics of strong autonomy, high positioning frequency, low power consumption and real-time performance. Therefore, it is more suitable for applications such as individual combat, anti-terrorism action, emergency rescue, and fire rescue where the environment is unknown and infrastructure cannot be deployed in advance, as shown in Fig. 1. Especially in military applications, inertial navigation can not only improve training efficiency and duty efficiency, but also provide accurate and continuous location services for areas such as jungles, indoor environments, and urban canyons, enabling multi-soldier collaborative operations. However, the work of such personnel is rather special. In addition to normal walking, more situations are running, turning, going up and downstairs, and even jumping and crawling, which brings greater challenges to navigation and positioning algorithm.

Although there has been much research work on strapdown INS, foot-mounted INS still has the following difficulties and challenges:

(1) Low-cost MEMS sensors with poor performance and high noise; (2) The complex and changeable pedestrian motion modes make it difficult to accurately detect the zero-velocity interval (ZVI); (3) Heading and altitude error accumulate over time in Zero-Velocity Update (ZUPT)-assisted pedestrian navigation systems. Thus, the positioning error of ZUPT-assisted INS is not well suppressed and diverges with time and movement distance [16].

This paper aims to improve the ZVI detection accuracy and reduce the heading and altitude drift of foot-mounted INS via deep learning and the equation constraint of dual feet. The key contributions of our study are as follows:

(1) We propose a novel dense sequence labelling mechanism to solve the multiclass window problem. This mechanism can accurately label the data for each timestamp, and then the dense prediction of temporal data is achieved by encoding and decoding data with accurate labelling.
(2) Aiming at the problem that traditional ZVI detectors based on sliding windows cannot provide sample-level detection results, we propose a novel sample-level ZVI detection model based on a dense sequence labelling mechanism and a U-Net network, which outputs binary zero-velocity classification results for each sensor measurement.
(3) We propose a bipedal INS method based on the equality constraint and ellipsoid constraint. This method fuses two monopod systems through foot-to-foot ranging, and uses foot-to-foot distance as a new observation of the Kalman filter to correct heading and altitude error. The method can

Fig. 1. Applications of foot-mounted pedestrian inertial navigation.

effectively suppress heading and altitude error, and achieve long-term accurate and stable pedestrian positioning.

(4) We conduct extensive and well-designed experiments to verify the performance of the proposed ZVI detection method and bipedal INS method, and obtain a satisfactory performance.

The remainder of this paper is organized as follows: Section 2 reviews the related literature. Section 3 details the proposed sample-level ZVI and bipedal INS method. Section 4 conducts an extensive experiment to verify the proposed method. Finally, Section 5 draws a conclusion.

2. Related work

The methods for pedestrian inertial navigation can be classified into three categories: (1) Pedestrian dead reckoning (PDR) consists of step detection [17], step length estimation [18] and heading estimation [19], which obtain high-precision positioning results in a short period through device built-in inertial sensors [20]; (2) Deep learning-based inertial odometry [21] recognizes pedestrian's moving modes and regresses pedestrian's position without any constraints on the usage mode of the device [21]; (3) The strapdown inertial navigation system (SINS) integrates the angular velocity with time to obtain the attitude change, and double integrates the acceleration with time in the navigation coordinate system to obtain the position change. The SINS sensors can be worn on the waist, calf, head, or upper arm, and more are tied to the foot or built into the insole. In this paper, we focus on foot-mounted inertial navigation.

However, the localization error diverges over time due to inertial sensor noise (i.e., biases, cross-coupling errors, scale factors and random noise) and non-strapdown relationship between user and device [15]. To overcome the divergence of accumulative inertial navigation error, the zero-velocity correction is the most practical constraint method, which consists of two parts: ZVI detection and zero-velocity update. ZUPT technology suppresses error accumulation by the fact that the velocity of pedestrian is zero when the foot contacts the ground [22]. ZVI detection is the basis of zero-velocity updating, which directly affects the effect of zero-velocity correction [23]. The ZVI detection is often modeled as a hypothesis-testing problem, and statistical detection theory is used to analyze recognize zero-velocity interval [24]. Extensive ZVI detection methods based on inertial sensors, including the generalized likelihood ratio test (GLRT) [25], acceleration magnitude (MAG) detector, acceleration moving mean detector, acceleration moving variance (MV) detector, angular rate moving mean detector, angular rate moving variance (ARMV) detector and angular rate energy (ARE) detector, have been proposed. These methods are all based on predefined thresholds, and their performance depends on pedestrian movement patterns and threshold settings.

Selecting the threshold value of each ZVI detection method, which must be selected according to the judgement principle of the zero-velocity state and a large number of experiments, is complicated. In addition, human movement is relatively complex and cannot be in a single movement mode, such as walking all the time. To solve the problem of the fixed threshold method, Ren et al. [26] leveraged a hidden Markov model (HMM) to adaptively adjust the ZVI detection threshold according to pedestrian movement modes. Yang et al. [27] utilized a Mahony filtering algorithm to merge and de-noise the original foot acceleration and angular velocity information to provide high-precision pitch angle information. Then, they utilized the pitch angle sliding variance for zero-velocity detection. Wahlstrom et al. [28] proposed a threshold-adaptive ZVI detection algorithm, which selects the optimal threshold

online according to the user's speed, gait frequency, or movement pattern. However, this algorithm requires collecting large quantity datasets at different walking velocities and different movement modes.

With the rapid development of deep learning technology, some researchers try to use a data-driven approach to perform ZVI detection, which can be applied to more general motion models without any manual feature extraction [29]. The ZVI detection methods based on deep learning or machine learning import the original inertial measurement units (IMU) data and corresponding labels into the neural network for training, which can effectively solve the dependence of traditional methods on the threshold setting and greatly improve the ZVI detection accuracy [30]. Wagstaff et al. [31] used a Long Short-Term Memory (LSTM) neural network to build a ZVI detection model, and improved model performance through rotation, scaling, and dithering. Zhu et al. [32] combined Convolutional Neural Networks (CNN) and LSTM, and used one-dimensional CNN to process raw data to achieve motion classification and improve the ZVI detection accuracy. Chen et al. [23] trained a contrastive neural network by comparing with the known static inertial data to recognize ZVI. Yang et al. [33] proposed Symmetrical-Net that leverages deep Recurrent Convolutional Neural Networks to adaptively detect ZVI.

Researchers have also explored ZVI detection methods based on images, foot pressure, radar signals, and echo feedback, which require additional equipment and are computationally expensive. In addition, the introduction of additional facilities can provide the ZVI detection accuracy, but the positioning system loses its autonomy and cannot deal with emergencies or cannot collect data in advance.

Compared with the ZVI detection accuracy, the heading and altitude error in inertial navigation has a greater impact on the positioning results. In the IEZ framework, the accumulation of the heading error cannot be eliminated. Therefore, limiting the heading error has become a research hotspot. In 2005, Foxlin et al. [34] corrected the navigation error according to the gait characteristics of walking to limit the navigation error, and combined ZVI detection with the extended Kalman filter to form an INS + EKF + ZUPT (IEZ) foot-mounted INS framework. The algorithm not only corrects velocity error, but also helps limit the increase in position and attitude error. Based on the IEZ framework, Zhang et al. [35] proposed the Zero Angular Rate Update (ZARU) method based on the assumption that the heading change is zero when the device is static or quasi-static, which limits the accumulation of gyroscope error and improves the heading stability. Since then, navigation system using the IEZ framework has been widely used. To enhance localization accuracy, scholars have improved various key parts of the IEZ framework. Sun et al. [36] improved the data quality of inertial sensor by calibration to improve the navigation accuracy.

Some studies have attempted to wear multiple inertial sensors on pedestrians, which can use multiple sensors to constrain each other to improve the navigation and positioning results. Zhu et al. [37] and Wu et al. [38] fixed two IMUs on pedestrian feet, using a maximum distance constraint between the feet to improve heading stability. Skog et al. [39] and Chen et al. [40] integrated multiple small and low-cost IMU on a PCB board to form an IMU array, and then used unscented Kalman filtering to fuse the navigation information of multiple IMUs to obtain a more robust and precise heading estimation. Pedestrians will be constrained by the surrounding environment during walking. Therefore, some researchers correct pedestrian inertial navigation error based on map assistance or map matching [41]. Khairi et al. [42] used building structures to assist pedestrian navigation, effectively constraining the heading divergence. Sahoo et al. [43] proposed a ramp calibration algorithm that detects ramps on the map and uses the ramp

position to correct pedestrian positions to improve localization accuracy.

In addition, some researchers combine foot-mounted INS with Global Navigation Satellite System (GNSS) [3], pseudolite [44], Bluetooth [45], cellular [6], vision [13], LiDAR, ultra-wideband (UWB) [9], Wi-Fi and other wireless positioning technologies to provide absolute position information for foot-mounted INS to enhance positioning accuracy and robustness. However, the method suppresses the heading error by fusing the magnetometer and GNSS can only be used outdoors, and the magnetometer and GNSS will fail in the indoor environment. In addition, the integrating sensors such as cameras and UWB will make the system too complicated and bloated, which is not conducive to the widespread use of pedestrian inertial navigation.

3. Materials and methods

3.1. System architecture

Fig. 2 illustrates the overall structure of our proposed pedestrian inertial navigation based on dual foot-mounted IMU. We sanitize inertial sensor noise with a denoising autoencoder. Subsequently, we propose a sample-level ZVI detection method that effectively solves the problem of mislabeling caused by sliding windows. Furthermore, we leverage the equality constraint and ellipsoid constraint to correct heading and altitude error.

3.2. Sensor noise sanitization based on denoising autoencoder

Mining the intrinsic features of inertial sensor measurements can effectively enhance the performance of ZVI detection and INS mechanization. In this study, the denoising autoencoder is utilized to learn the intrinsic characteristics of sensor data before training the ZVI detection model. The autoencoder is an unsupervised learning method with the same size as the input and output layers. Autoencoder consists of encoding and decoding procedures. For input x, the hidden encoding layers predict a target, and the decoding procedure reconstructs the input with z through decoding mirror layers. The objective of autoencoder is to minimize the difference between the raw measurements $\boldsymbol{x}$ of inertial sensors and reconstructed input.

$$\min_{\varnothing,\varphi}\|\boldsymbol{x} - \varphi(\varnothing(\boldsymbol{x}))\|^2 \tag{1}$$

The denoising autoencoder [46] is a stochastic variant of the autoencoder. The denoising autoencoder is an extremely effective tool for learning the essential features, which can be used for noise reduction. Compared with autoencoder, denoising autoencoders require an additional stochastic corruption step in the training phase. As shown in Fig. 3, each denoising autoencoder consists of three parts: an encoder layer, a noise-added layer, and a decoder layer. The noise-added layer obtains corrupted inertial data $\tilde{x}$ by adding the masking noise to the original sensor measurements $\boldsymbol{x}$. The encoder layer maps the corrupted data $\tilde{\boldsymbol{x}}$ into hidden representation $\boldsymbol{y}$.

$$\boldsymbol{y} = f(\boldsymbol{w}\tilde{\boldsymbol{x}} + \boldsymbol{b}) \tag{2}$$

where $\boldsymbol{w}$ is an $N \times M$ encoding matrix, and $\boldsymbol{b}$ is a bias vector. M represents the number of input units and N represents the number of hidden units. $f(\cdot)$ denotes the activation function. A similar process is performed by the decoder layer to map the hidden representation back to the raw sensor measurements.

$$\boldsymbol{z} = g(\boldsymbol{w}'\boldsymbol{y} + \boldsymbol{b}') \tag{3}$$

where $\boldsymbol{w}'$ is an $M \times N$ decoding matrix, and $\boldsymbol{b}'$ is a bias vector. $\boldsymbol{z}$ is the reconstructed vector. $g(\cdot)$ denotes the activation function. In this paper, we choose the sigmoid function as the activation function.

The goal of denoising is to minimize the reconstruction error, which is done by minimizing the square error loss function $L(x_i, z_i)$, where

$$L(\boldsymbol{x}_i, \boldsymbol{z}_i) = \frac{1}{2} \sum_{j=1}^{M} (\boldsymbol{x}_j - \boldsymbol{z}_j)^2 \tag{4}$$

A sparsity term applies to the objective function of the denoising autoencoder to make it work even if there are more hidden units than input units. Therefore, the cost function J_{DAE} is rewritten as
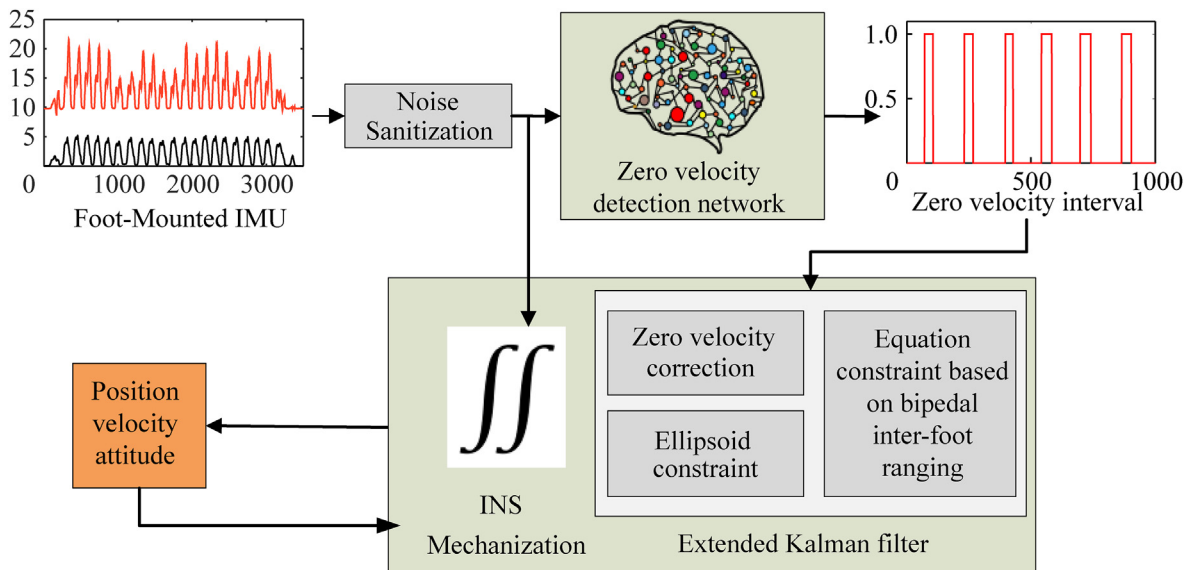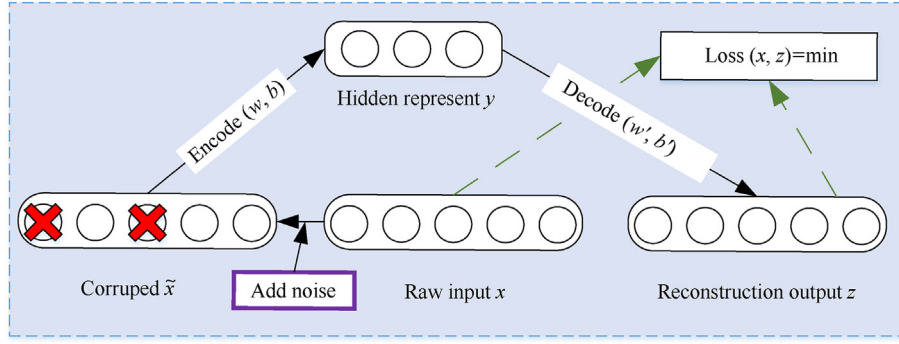

Fig. 2. System architecture.

Q. Wang, M. Fu, J. Wang et al.

Fig. 3. Structure diagram of a denoising autoencoder.

$$J_{\mathrm{DAE}} = L(\boldsymbol{x}_i, \boldsymbol{z}_i) + \beta \sum_{j=1}^{N} KL(\rho \| \widehat{\rho}_j) = \frac{1}{2} \sum_{j=1}^{M} (\boldsymbol{x}_j - \boldsymbol{z}_j)^2$$

$$+ \beta \sum_{j=1}^{N} \left[\rho \lg(\rho / \widehat{\rho}_j) + (1-\rho) \lg\left(\frac{1-\rho}{1-\widehat{\rho}_j} \right) \right] \quad (5)$$

The first term is the cost function of autoencoder, and the second term is the sparse penalty term, where N is the number of hidden units; ρ and $\widehat{\rho}_j$ are the sparsity parameter and the average activation of hidden unit j, respectively; and β is the sparsity penalty.

3.3. Zero-velocity interval detection based on U-Net neural network

ZVI detection is the most critical technology in pedestrian inertial navigation systems based on zero-velocity assistance. Enhancing the ZVI detection accuracy can effectively improve positioning accuracy [16]. However, due to the high-frequency foot vibration during pedestrian movement, it is challenging to accurately recognize the zero-velocity state and the non-zero-velocity state from the gyroscope and accelerometer measurements. Especially for mixed motion modes such as running and jumping, further research is needed to accurately detect the ZVI. The traditional ZVI detection methods are based on sliding window, which suffers from a multiclass window problem, i.e., the samples within each window cannot always share the same label. To solve the multiclass window problem, we analyze the pedestrian gait cycle, present a dense labelling mechanism, and propose a novel sample-level ZVI detection method based on U-Net neural network.

3.3.1. Gait analysis

A entire gait cycle of pedestrian is defined as the process from the heel of one foot on the ground to the heel of the opposite foot on the ground again when they walk [49]. Fig. 4 shows the different stages in the pedestrian gait cycle. The time for the foot to fully contact the ground during walking is very short, usually approximately 0.1−0.3 s, and it is even shorter when running [47,48]. Fig. 5 shows the acceleration change during the walking process. As shown in Fig. 5, the periodic amplitude change in the inertial data corresponds to the alternating occurrence of ZVI and non-ZVI, and has a relatively apparent front-to-back dependency in the time dimension.

3.3.2. Traditional zero-velocity interval detection methods

During movement such as walking and running, the feet will alternately leave and touch ground. When their feet touch the ground, the pedestrian's speed is zero. Many research works are based on this phenomenon to correct the positioning error. The traditional ZVI detection methods based on IMU are classified into the following three categories: those based on acceleration, those based on angular velocity, or those based on a combination of acceleration and angular velocity [50]. A sliding window is usually used to enhance detection robustness. Detection method based on acceleration amplitude threshold (MAGD) [42],

$$A_{\mathrm{a}} = \begin{cases} \frac{1}{N} \sum_{k=i}^{i+N-1} (\| \boldsymbol{a}_k \| - g)^2 < t_{\mathrm{a}} & 1 \\ \\ \text{others} & 0 \end{cases} \quad (6)$$

Detection method based on acceleration variance threshold (MVD) [51],

$$A_{\sigma^2} = \begin{cases} \frac{1}{N-1} \sum_{k=i}^{i+N-1} (\boldsymbol{a}_k - \overline{\boldsymbol{a}}_N)^2 < t_{\sigma} & 1 \\ \\ \text{others} & 0 \end{cases} \quad (7)$$

Detection method based on angular velocity energy threshold (ARED) [52],

$$A_{\omega} = \begin{cases} \frac{1}{N} \sum_{k=i}^{i+N-1} (\| \omega_k \|)^2 < t_{\omega} & 1 \\ \\ \text{others} & 0 \end{cases} \quad (8)$$

Combined detection method (generalized likelihood ratio test, GLRT),

$$A_c = \begin{cases} \frac{1}{N} \sum_{k=i}^{i+N-1} \frac{1}{\sigma_a^2} \| \boldsymbol{a}_k - g \frac{\overline{\boldsymbol{a}}_k}{\| \overline{\boldsymbol{a}}_k \|} \|^2 + \frac{1}{\sigma_{\omega}^2} \| \omega_k \| < t_{\mathrm{c}} & 1 \\ \\ \text{others} & 0 \end{cases} \quad (9)$$

$$\overline{\boldsymbol{a}}_k = \frac{1}{N} \sum_{k=i}^{k+N-1} \boldsymbol{a}_k \quad (10)$$

where $\boldsymbol{a}_k = \begin{bmatrix} a_{xk} & a_{yk} & a_{zk} \end{bmatrix}^{\mathrm{T}}$ is the measurement of the accelerometer at time k; $\omega_k = \begin{bmatrix} \omega_{xk} & \omega_{yk} & \omega_{zk} \end{bmatrix}^{\mathrm{T}}$ is the measurement of the gyroscope at time k; t_{a}, t_{ω}, t_{σ} and t_{c} are the zero-velocity detection thresholds of acceleration, angular velocity, acceleration variance and combined observation, respectively; and σ_a^2 and σ_{ω}^2 are the random noise variances in acceleration and angular velocity, respectively. The value of the window size N depends on the sampling frequency of the sensor and the walking velocity of the pedestrian, and is usually determined empirically.
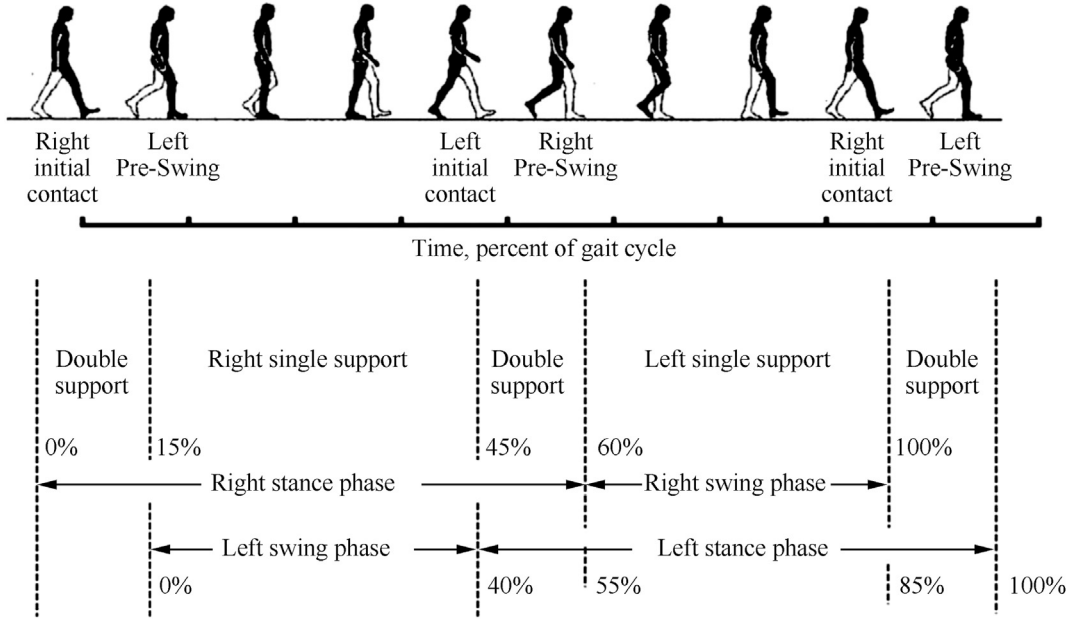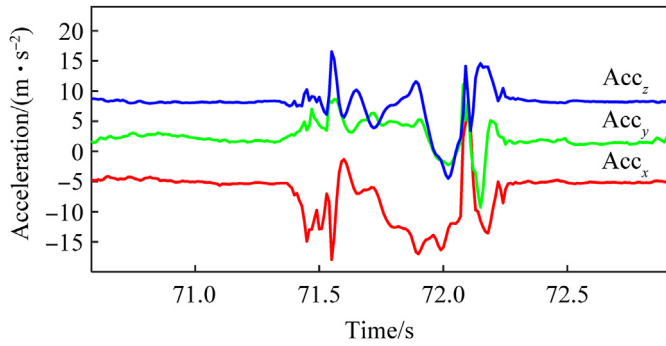
Fig. 4. An entire gait cycle [47,48].

Fig. 5. Acceleration variation within one gait cycle.

The above fixed threshold-based methods are excellent for ZVI detection in uniform motion, but when the pedestrian frequently switches motion modes and intensities, the detection accuracy drops sharply [33].

3.3.3. Dense labelling

ZVI detection may be considered as an issue of sequence labelling in which the zero-velocity states at each timestamp in long-term time-series data are anticipated. The traditional ZVI detection method typically uses explicit windowing to segment the data. The explicit windowing method divides data into sub-windows of equal length and assigns an identical label to each sample in each subwindow. As illustrated in Fig. 6, a multiclass windows occur when samples within a subwindow do not share the same label. In addition, there is no clear consensus on the definite window size to utilize because it is challenging to define the boundaries of state change.

Fig. 6 provides a visual comparison of sliding window labeling and dense sequence labeling. Red indicates the swinging phase (S1), and blue represents the stance phase (S2). For instance, when using the labeling strategy with the most occurring classes, the label of this subwindow is assigned as S1. Conversely, S2. However, this subwindow essentially involves class S1 and class S2
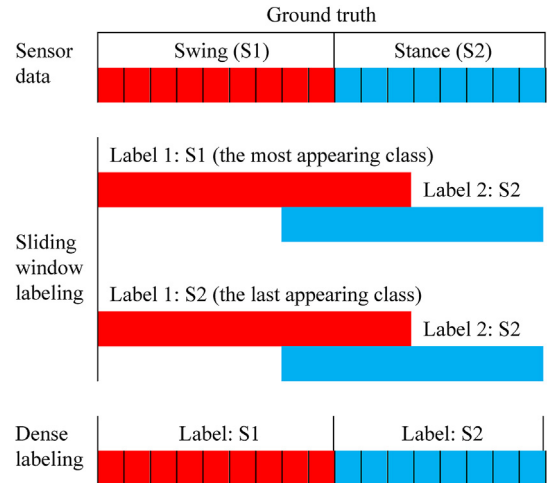

Fig. 6. Illustration of sliding window labeling and dense sequence labeling. The top of the figure shows the sensor data in different states, with different colors indicating different state (swing or stance). The middle of the figure shows two different approaches in sliding window labeling, which both lead to loss of label information. The bottom of the figure shows dense sequence labeling, which avoids the multiclass window problem by assigning labels to each sample.

information. Therefore, it leads to losing the accurate information of the other class. Using incorrect sequence semantic information leads to an inaccurate recognition result. The bottom of Fig. 6 shows that dense sequence labeling can train and predict every sample's label rather than every window.

As technical jargon, the term "dense sequence labeling" is proposed by the semantic segmentation field in computer vision. We contributed the dense sequence labeling method to circumvent the multiclass window problem, giving a single label for each sample in every timestamp rather than assigning the same label for each subwindow. The dense sequence labeling result is denoted as $Y_L = \{y_1, y_2, \cdots y_i, \cdots y_L\}$, where y_i represents the state of the ith sample. The output of dense labeling is the same length as the time-series data.

Q. Wang, M. Fu, J. Wang et al.

3.3.4. Zero-velocity interval detection

Based on dense labels, we propose a novel sample-level ZVI detection method based on U-Net neural network. U-Net neural network is an improved end-to-end image semantic segmentation network based on a fully convolutional neural network [53], which contains a shrinking path for feature extraction and a dilation path for expanding the size of feature maps. The shrinking and dilation paths are a pair of symmetric structures (encoder-decoder structures). In U-Net, the input image first performs feature extraction on the image through the convolution-pooling operation, then uses deconvolution in the expansion path to enlarge the size of the feature map to the raw image size, and finally predicts category of each pixel to realize semantic segmentation of images.

The model architecture of the proposed sample-level ZVI detection method is shown in Fig. 7. We denote the sanitized two-dimensional sensor data as a single pixel column and multichannel image with a size of (1, N, C), where N is the length of the input subsequence, and C is the number of the sensor channels. Our method outputs the time of heel strike and toe-off (zero-velocity interval).

3.4. INS mechanization

As shown in Fig. 8, INS mechanization obtains the position, velocity, and attitude of an object by integrating acceleration twice and angular rate once provided by accelerometers and gyroscope. Compared with the noise of low-cost MEMS sensors, the navigation performance improvement brought by optimizing parameters such as Earth rotation is negligible. Therefore, the INS mechanization equations can be simplified as follows [54]:

$$\begin{bmatrix} \boldsymbol{p}_k^n \\ \boldsymbol{v}_k^n \\ \boldsymbol{C}_{b,k}^n \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_{k-1}^n + \boldsymbol{v}_k^n \Delta t \\ \boldsymbol{v}_{k-1}^n + \left[\boldsymbol{C}_{b,k}^n \left(\tilde{\boldsymbol{f}}_k^b - \boldsymbol{b}_a \right) - \boldsymbol{g}^n \right] \Delta t \\ \boldsymbol{C}_{b,k-1}^n + \boldsymbol{C}_{b,k-1}^n \Omega \left[\left(\tilde{\boldsymbol{w}}_k^b - \boldsymbol{b}_g \right) \Delta t \right] \end{bmatrix} \quad (11)$$

where k is a time index, $\Delta t = t_k - t_{k-1}$ is the time interval between two consecutive observations, $\boldsymbol{p}^n$ and $\boldsymbol{v}^n$ indicate the position vectors and velocity vectors in the navigation system, respectively, $\boldsymbol{C}_b^n$ represents the rotation matrix from the carrier system to the navigation system, $\boldsymbol{g}^n = [0,0,-g]^T$ represents the Earth's gravity vectors in the navigation system, $\tilde{\boldsymbol{f}}^b$ and $\tilde{\boldsymbol{w}}^b$ represent the acceleration and angular rate observation vectors in the carrier system, respectively, $\boldsymbol{b}_a$ and $\boldsymbol{b}_g$ are the bias vectors of the accelerometer and gyroscope, respectively, and $\Omega[\bullet]$ is the cross-product form.

3.5. Heading error correction based on equation constraints of bipedal INS and inter-foot ranging

Traditional heading correction methods use magnetometer observations mounted on the IMU to correct heading estimation error. This method can effectively provide position estimation accuracy, but the magnetometer is easily contaminated by building materials, electronic equipment and other external magnetic sources. Other heading correction methods that rely on sources or prior knowledge (WLAN fingerprinting, map matching) require infrastructure to be installed in the area to be located or environmental data to be collected in advance. These methods lose the substantial autonomy of inertial navigation, and cannot be applied to emergencies such as emergency rescue.

To calibrate the heading error while maintaining the autonomy of inertial navigation, we calibrate the heading error by means of bipedal INS and inter-foot ranging. We design a bipedal pedestrian navigation system that consists of two IMU modules and an ultrasonic module. The system obtains the distance between the feet through the ultrasonic module and synchronizes it with the inertial data. The inter-foot distance between is calculated as follows:

$$d = \| \boldsymbol{p}_L^e + \boldsymbol{C}_{b,L}^e \boldsymbol{I}_L^b - \boldsymbol{p}_R^e - \boldsymbol{C}_{b,R}^e \boldsymbol{I}_R^b \| + n_d \triangleq \| \Delta \boldsymbol{I} \| + n_d \quad (12)$$

where L and R represent the left and right feet, respectively. $\boldsymbol{I}^b$ represents the lever arm of IMU module and ultrasound module. $\boldsymbol{p}^n$ indicates the position vectors. $\boldsymbol{C}_b^n$ represents the rotation matrix. n_d represents the inter-foot distance observation noise with a covariance of σ_d^2. $\| \bullet \|$ represents the vector magnitude. In this paper, the bipedal distance is incorporated into the EKF as a new observation.

We define the error state $\delta \boldsymbol{x}$ as the state estimate $\hat{\boldsymbol{x}}$ minus the true state $\boldsymbol{x}$.
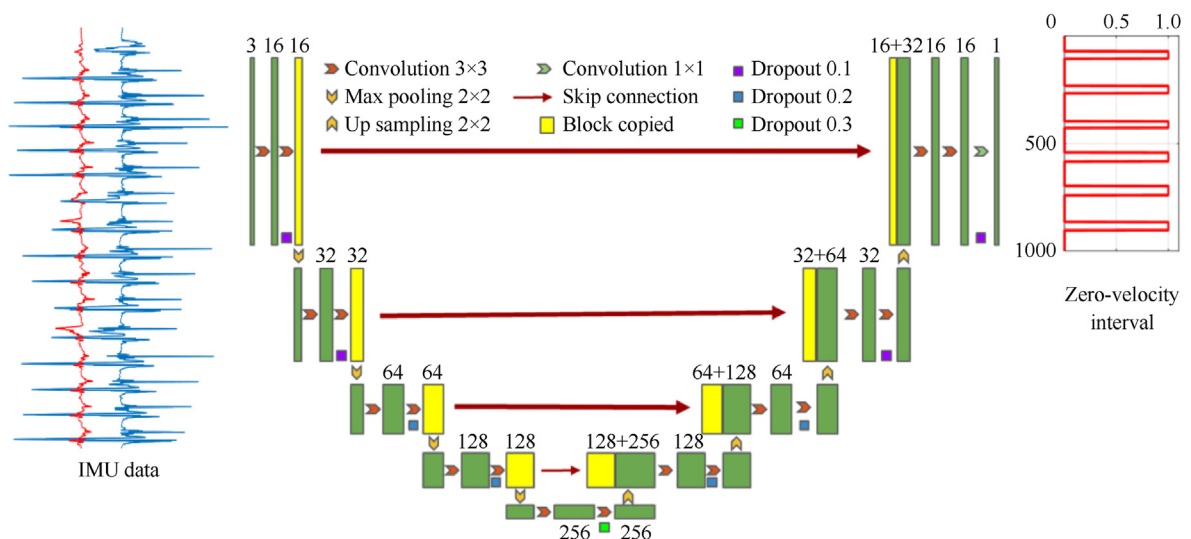

Fig. 7. System architecture of the proposed robust zero-velocity interval detection method based on U-Net network. Input refers to IMU raw measurements surrounding the under determining point. Output infers the time of heel strike and toe-off.
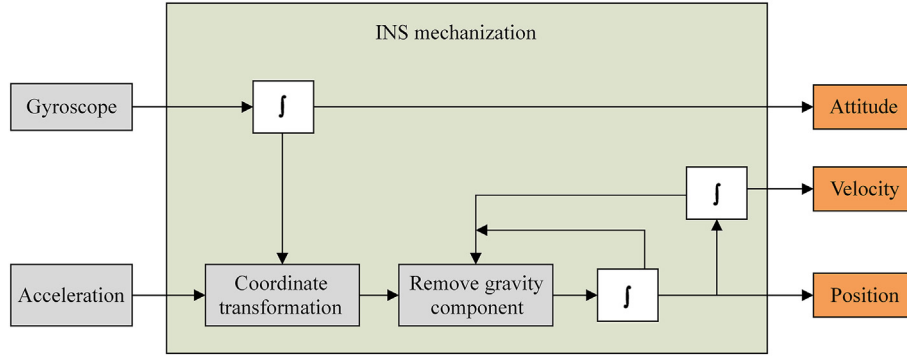
ARTICLE IN PRESS

Q. Wang, M. Fu, J. Wang et al.   Defence Technology xxx (xxxx) xxx

Fig. 8. INS mechanization.

$$\delta \boldsymbol{x} = \widehat{\boldsymbol{x}} - \boldsymbol{x} \tag{13}$$

According to the monopod model, the left and right feet are each subsystem. We combine the two subsystems into a bipedal system by introducing a bipedal intra-foot distance constraint. Correspondingly, the system state vectors $\delta \boldsymbol{X}$ of the biped are expanded to

$$\delta \boldsymbol{X} = \begin{bmatrix} \delta \boldsymbol{x}_{\mathrm{L}} \\ \delta \boldsymbol{x}_{\mathrm{R}} \end{bmatrix} \tag{14}$$

where $\delta \boldsymbol{x}_L$ and $\delta \boldsymbol{x}_R$ represent the error state space equations of the left and right foot subsystems, respectively.

Its state space equation $\delta \boldsymbol{X}$ and observation equation $\delta \boldsymbol{Y}$ are expressed as

$$\delta \boldsymbol{X}_k = \boldsymbol{F} \delta \boldsymbol{X}_{k-1} + \boldsymbol{D} \boldsymbol{W} \tag{15}$$

$$\delta \boldsymbol{Y} \equiv \begin{bmatrix} \delta \boldsymbol{y}_{v,\mathrm{L}} \\ \delta \boldsymbol{y}_{v,\mathrm{R}} \\ \delta \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} H_v & 0_{3\times 15} \\ 0_{3\times 15} & H_v \\ H_{d,\mathrm{L}} & H_{d,\mathrm{R}} \end{bmatrix} \delta \boldsymbol{X} + \begin{bmatrix} n_{v,\mathrm{L}} \\ n_{v,\mathrm{R}} \\ n_d \end{bmatrix} \tag{16}$$

$$\delta \boldsymbol{d} = \boldsymbol{d}^n - \boldsymbol{d} \tag{17}$$

where $\boldsymbol{F}$ and $\boldsymbol{D}$ represent the state transition matrix and noise gain matrix of the inertial navigation system, respectively. $\boldsymbol{H}$ and $\boldsymbol{n}$ represent the observation matrix and observation noise, respectively. $\boldsymbol{d}^n$ is the ultrasonic measurement and the process noise $\boldsymbol{W}$ is

$$\boldsymbol{W} = \begin{bmatrix} \mathscr{w}_{\mathrm{L}}^{\mathrm{T}} \\ \mathscr{w}_{\mathrm{R}}^{\mathrm{T}} \end{bmatrix} \tag{18}$$

where

$$H_{d,\mathrm{L}} = \left[\Delta \boldsymbol{I}^{\mathrm{T}} \left(\boldsymbol{C}_{b,\mathrm{L}}^e \boldsymbol{I}_{\mathrm{L}}^b \right) \times \boldsymbol{0}_{1\times 3} \Delta \boldsymbol{I}^{\mathrm{T}} \boldsymbol{0}_{1\times 6} \right] \Big/ \parallel \Delta \boldsymbol{I} \parallel \tag{19}$$

$$H_{d,\mathrm{R}} = \left[-\Delta \boldsymbol{I}^{\mathrm{T}} \left(\boldsymbol{C}_{b,\mathrm{R}}^e \boldsymbol{I}_{\mathrm{R}}^b \right) \times \boldsymbol{0}_{1\times 3} - \Delta \boldsymbol{I}^{\mathrm{T}} \boldsymbol{0}_{1\times 6} \right] \Big/ \parallel \Delta \boldsymbol{I} \parallel \tag{20}$$

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{f}_{\mathrm{L}} & \boldsymbol{0}_{15} \\ \boldsymbol{0}_{15} & \boldsymbol{f}_{\mathrm{R}} \end{bmatrix} \tag{21}$$

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{d}_{\mathrm{L}} & \boldsymbol{0}_{15\times 12} \\ \boldsymbol{0}_{15\times 12} & \boldsymbol{d}_{\mathrm{R}} \end{bmatrix} \tag{22}$$

where $\boldsymbol{I}$ represents identity matrix. $\boldsymbol{0}$ represents zero matrix. $\boldsymbol{C}$ represents the rotation matrix. $\boldsymbol{f}$ represent the acceleration observation vectors.

The filtering process is as follows.

3.5.1. Initialization phase

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{P}^{\mathrm{L}} & \boldsymbol{0}_{15\times 15} \\ \boldsymbol{0}_{15\times 15} & \boldsymbol{P}^{\mathrm{R}} \end{bmatrix} \tag{23}$$

$\boldsymbol{P}^{\mathrm{L}}$ and $\boldsymbol{P}^{\mathrm{R}}$ are the covariance matrices of the left and right feet, respectively.

$$\boldsymbol{P}^{\mathrm{L}} = \begin{bmatrix} \boldsymbol{p}_{11}^{\mathrm{L}} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} \\ \boldsymbol{0}_{3\times 3} & \boldsymbol{p}_{22}^{\mathrm{L}} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} \\ \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{p}_{33}^{\mathrm{L}} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} \\ \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{p}_{44}^{\mathrm{L}} & \boldsymbol{0}_{3\times 3} \\ \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{p}_{55}^{\mathrm{L}} \end{bmatrix} \tag{24}$$

$$\boldsymbol{P}^{\mathrm{R}} = \begin{bmatrix} \boldsymbol{p}_{11}^{\mathrm{R}} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} \\ \boldsymbol{0}_{3\times 3} & \boldsymbol{p}_{22}^{\mathrm{R}} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} \\ \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{p}_{33}^{\mathrm{R}} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} \\ \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{p}_{44}^{\mathrm{R}} & \boldsymbol{0}_{3\times 3} \\ \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{0}_{3\times 3} & \boldsymbol{p}_{55}^{\mathrm{R}} \end{bmatrix} \tag{25}$$

where $\boldsymbol{p}_{ii}$ represents the empirical value of the navigation parameter.

3.5.2. Prediction stage

$$\delta \boldsymbol{X}_k = \boldsymbol{F}_k \delta \boldsymbol{X}_{k-1} + \boldsymbol{D}_k \boldsymbol{W}_k \tag{26}$$

where $\delta \boldsymbol{X}_{k-1}$ is the error estimate from the previous step. $\boldsymbol{F}$ and $\boldsymbol{D}$ represent the state transition matrix and noise gain matrix of the inertial navigation system. Since the noises of the two subsystems are independent of each other, the covariance matrix $\boldsymbol{Q}$ of the process noise $\boldsymbol{W}_k$ is calculated as follows:

Q. Wang, M. Fu, J. Wang et al.

$$Q = \begin{bmatrix} Q^L & 0_{6\times6} \\ 0_{6\times6} & Q^R \end{bmatrix} \tag{27}$$

where Q^L and Q^R are the observation noise covariance matrix of the left and right feet, respectively.

$$Q^L = \begin{bmatrix} q^L_{11} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & q^L_{22} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & q^L_{33} \end{bmatrix} \tag{28}$$

$$Q^R = \begin{bmatrix} q^R_{11} & 0_{3\times3} & 0_{3\times3} \\ 0_{3\times3} & q^R_{22} & 0_{3\times3} \\ 0_{3\times3} & 0_{3\times3} & q^R_{33} \end{bmatrix} \tag{29}$$

where q_{11}, q_{22}, and q_{33} are the empirical covariance values of the accelerometer, gyroscope and ultrasound in the initialization stage, respectively.

3.5.3. Observation update phase

The motion of each foot of the pedestrian alternates between swinging and stance phase, the movements of the feet alternate with each other. Therefore, the observation equations δy are also different in different phases. When the left foot is in stance and the right foot is swinging.

$$\delta y = \begin{bmatrix} \delta y_{v,L} \\ \delta d \end{bmatrix} = \begin{bmatrix} H_v & 0_{3\times15} \\ H_{d,L} & H_{d,R} \end{bmatrix} \delta X + \begin{bmatrix} n_{v,L} \\ n_d \end{bmatrix} \tag{30}$$

When the right foot is in stance and the left foot is swinging.

$$y = \begin{bmatrix} \delta y_{v,R} \\ \delta d \end{bmatrix} = \begin{bmatrix} H_v & 0_{3\times15} \\ H_{d,L} & H_{d,R} \end{bmatrix} \delta X + \begin{bmatrix} n_{v,R} \\ n_d \end{bmatrix} \tag{31}$$

When both left and right feet are in stance at the same time.

$$\delta y = \begin{bmatrix} \delta y_{v,L} \\ \delta y_{v,R} \\ \delta d \end{bmatrix} = \begin{bmatrix} H_v & 0_{3\times15} \\ 0_{3\times15} & H_v \\ H_{d,L} & H_{d,R} \end{bmatrix} \delta X + \begin{bmatrix} n_{v,L} \\ n_{v,R} \\ n_d \end{bmatrix} \tag{32}$$

where H and n represent the observation matrix and observation noise, respectively. 0 represents zero matrix.

The observation is input into the extended Kalman filter, and the filter gain is calculated. Then, the innovation is calculated, and the innovation is substituted into the state estimation equation to obtain the optimal estimation of the error. Finally, the optimal estimation of the error is compensated for the state obtained by the solution.

3.6. Altitude error correction based on ellipsoid constraint

Although the height error of each step is only centimeters. However, altitude errors accumulate over time and step count, eventually resulting in very significant altitude errors, especially with low-cost sensors. Therefore, altitude error correction is very meaningful and necessary for low-cost inertial sensors. The altitude drift of each step of pedestrians is usually less than 3 cm, which is significantly smaller than the height of stair (decimeters). When the altitude change is less than the threshold, it is considered to be drift and needs to be set to zero; when it is greater than the threshold, it is considered to be a real altitude change and is not set
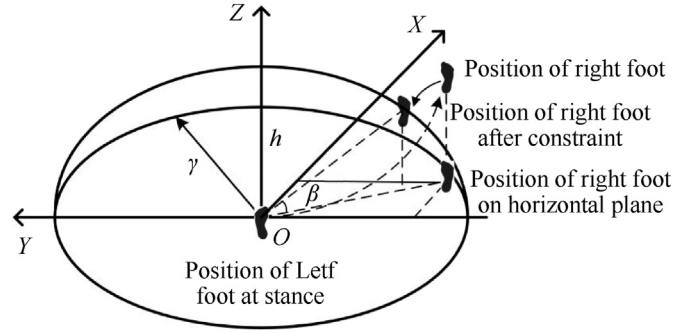

Fig. 9. Ellipsoid constraint correction diagram.

to zero. As shown in Fig. 9, the foot lies on a common ellipsoid surface during the neighboring stance phases, the ellipsoid constraint can be used to suppress the drift of altitude. The ellipsoid constraint means that the height change in the same foot (position of foot-mounted IMU) between two adjacent supports is less than a threshold ε.

$$|h(t_{k-m}) - h(t_k)| < \varepsilon \tag{33}$$

where h is height. t_{k-m} and t_k are previous and current moment, respectively. In two adjacent support phases, the feet lie on the same ellipsoid. This constraint satisfies

$$1 = \frac{(x^e)^2 + (y^e)^2}{(R_E + h)^2} + \frac{(z^e)^2}{\left[R_E\left(1 - f^2\right) + h\right]^2} + n_{ec} \tag{34}$$

where f is the eccentricity of the Earth, and R_E is the lateral radius of curvature. R_E and h are the radius of curvature and height at time t_{k-m}. n_{ec} is ellipsoid bound noise.

We convert Eq. (34) to the observation matrix form by first-order Taylor expansion as follows:

$$H_{ec} = \left[0_{1\times6}, \ \frac{2x^e}{(R_E + h)^2}, \ \frac{2y^e}{(R_E + h)^2}, \ \frac{2z^e}{\left(R_E\left(1 - f^2\right) + h\right)^2}, \ 0_{1\times6} \right] \tag{35}$$

Each foot obeys the ellipsoid constraints. Therefore, their observation equations δy_{ec} are as follows:

$$\delta y_{ec} = \begin{bmatrix} \delta y_{ec,L} \\ \delta y_{ec,R} \end{bmatrix} = \begin{bmatrix} H_{ec,L} & 0_{3\times15} \\ 0_{1\times15} & H_{ec,R} \end{bmatrix} \delta X + \begin{bmatrix} n_{ec,L} \\ n_{ec,R} \end{bmatrix} \tag{36}$$

where H_{ec} and n_{ec} represent the height observation matrix and observation noise. δX represents the system state vectors of the biped.

In extreme cases, this approach will fail and lead to height errors when the user is walking on a very gentle slope or ramp. This problem can be solved by ramp detection [55].

4. Experimentation and evaluation

We conduct well-designed and extensive experiments on different scenarios to verify the performance of the proposed method. The experimental setup is first introduced. Section B validates the ZVI recognition accuracy. Section C validates the positioning accuracy of the proposed method in typical scenarios. Section D compares the proposed method with other methods in a complex indoor and outdoor environment.

Q. Wang, M. Fu, J. Wang et al.

4.1. Experimental setup

To alleviate the heading error, we install an IMU on each foot of pedestrian, and obtain the foot distance measurement value through the ultrasonic ranging module. The two monopod INS subsystems are fused by equality constraints through the inter-foot distance. As illustrated in Fig. 10, the proposed system consists of two foot-mounted IMU modules (ICM-42605 [56]), a pair of ultrasound sensors and an Arduino development board for collecting sensor data. The key performance parameters of the IMU module and the ultrasound module are shown in Table 1 and Table 2, respectively.

Inertial navigation is relative positioning. Its positioning error depends not only on the sensor performance and positioning algorithm, but also on the positioning time and travel distance. Therefore, we adopt absolute trajectory error (ATE), distance-normalized relative trajectory error (D-RTE), time-normalized relative trajectory error (T-RTE) and position drift error (PDE) to fully evaluate the performance of the proposed method. To explore the localization error distribution, we introduce the positioning error cumulative distribution function (CDF).

ATE is the major metric used to evaluate the indoor positioning accuracy. ATE is the RMSE of the estimation error and reflects the global consistency between the estimated trajectory and the true trajectory, and the error grows with movement time and distance.

$$\text{ATE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N} \| x_t - \tilde{x}_t \|} \tag{37}$$

T-RTE is the RMSE of average error over a fixed time-interval window span t_i ($t_i = 10$ s in our case). It reflects the local consistency between the estimated trajectory and the true trajectory.

$$\text{T} - \text{RTE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N} \| (x_{t+t_i} - x_t) - (\tilde{x}_{t+t_i} - \tilde{x}_t) \|} \tag{38}$$

D-RTE is the RMSE of average error over a fixed travel distance d ($d = 10$ m in our case). It also reflects the local consistency between the estimated trajectory and the true trajectory.

$$\text{D} - \text{RTE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N} \| (x_{t+t_d} - x_t) - (\hat{x}_{t+t_d} - \hat{x}_t) \|} \tag{39}$$

PDE reflects the final position (at timestamp m) drift over the total traveled distance (traj. len).

Table 1
Performance parameters of IMUs module.

Name	Parameter
Operating voltage/V	3.3−5
Output frequency/Hz	0.1−200
Range (A/g), ($G/(\degree \cdot s^{-1})$)	A: ±2/4/8/16;
	G: ±250/500/1000/2000
Acceleration stability/(g·s^{-1})	0.01
Angular velocity stability/(($\degree$)·s^{-1})	0.05

Table 2
Performance parameters of the ultrasonic module.

Name	Parameter
Sound frequency/kHz	40
Receiving distance/cm	3−500
Receiving angle/($\degree$)	<90
Data fluctuation range/mm	>2
Data sending frequency/Hz	50
Ranging accuracy/cm	±2

$$\text{PDE} = \| x_m - \tilde{x}_m \| / \text{traj. len} \tag{40}$$

where x_t and $\tilde{x}_t$ represent the position in the ground truth trajectory and in the predicted path at timestamp t, respectively. t_d is the time required to travel a distance of d.

4.2. Accuracy of zero-velocity interval recognition

To verify the performance of the ZVI detection method, we compare the proposed method with traditional fixed threshold-based ZVDs methods (acceleration-moving variance detector (MVD) [51], acceleration-magnitude detector (MAGD) [42], angular rate energy detector (ARED) [52], and stance hypothesis optimal detector (SHOE) [57]) and data-driven methods (CNN, LSTM, CNN-LSTM). Table 3 compare these methods in typical movement modes (forward walking, backward walking, running, lateral walking). As shown in Table 3, for normal movement (forward walking), the fixed threshold method and the data-driven method are similar, and both achieve more than 95% recognition accuracy. For abnormal movements (backward walking, running, lateral walking), it is difficult for the fixed threshold methods to maintain a recognition accuracy of more than 90%, while the data-driven methods still achieve a recognition accuracy of more than 90%. For abnormal movements, the recognition accuracy of the data-
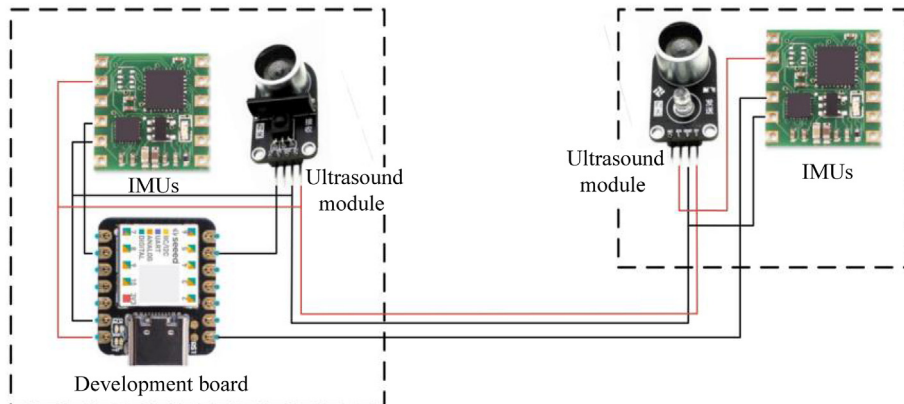

Fig. 10. Data acquisition equipment.

Table 3
Comparison of zero-velocity recognition accuracy in different motion modes.

Methods	Forward walking/%	Backward walking/%	Running/%	Lateral walking/%	Average/%
MVD	95.8	86.6	78.5	82.3	85.80
ARED	97.4	88.9	81.2	84.7	88.05
MAGD	96.3	87.7	85.3	89.2	89.63
SHOE	98.2	89.3	84.8	88.6	90.23
CNN	96.1	91.6	92.8	93.2	93.43
LSTM	98.3	95.1	93.6	96.3	95.83
CNN-LSTM	99.2	96.3	95.1	97.5	97.03
Proposed	**99.5**	**98.1**	**97.9**	**98.7**	**98.55**

driven method is significantly better than that of the fixed threshold method. As shown in Table 3, the average accuracy of each ZVI detector in different motion modes are 85.80%, 88.05%, 89.63%, 90.23%, 93.43%, 95.83%, 97.03% and 98.55%, respectively. The proposed method outperforms other compared methods on four common motion patterns. All compared methods are based on the sliding window for zero-velocity interval detection, only knowing whether there is a zero-velocity point in the window, and cannot specifically detect whether each observation is zero-

velocity. The proposed method provides more. The proposed method provides more accurate and finer ZVI detection for each observation.

In this work, we collect acceleration, angular velocity data, and trained the zero-velocity interval detection model on the PC with python language and PyTorch deep learning platform. U-Net is a neural network with high efficiency and low computational overhead. The zero-velocity interval detection model based on U-Net can efficiently detect the zero-velocity interval in real time.

4.3. Positioning accuracy in typical scenarios

To more fully and systematically evaluate the proposed positioning method, we invited 7 volunteers to conduct experiments in four typical indoor and outdoor positioning scenarios, including offices, shopping malls, streets, and parking lots. Table 4 details 7 volunteers' information. Volunteers randomly generated multiple trajectories of 50–300 m in each scenario. The cumulative walking distance of each scenario is approximately 5 km. Fig. 11 shows the comparison of localization accuracy for the four scenarios in the
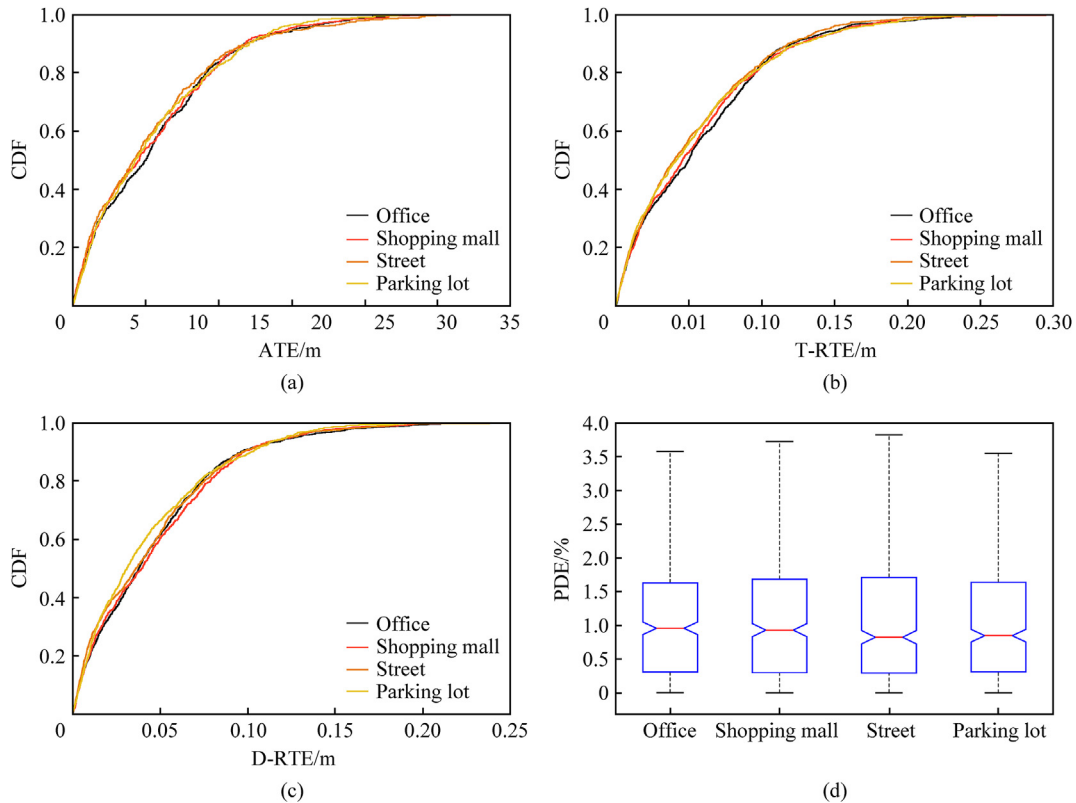
Table 4
Description of volunteers.

Subject	Gender	Age	Height/cm	Weight/kg
S1	M	27	171	68
S2	M	26	183	82
S3	F	23	155	45
S4	M	35	192	85
S5	F	25	161	54
S6	F	27	158	57
S7	M	31	175	70

Fig. 11. Accuracy comparison in typical scenarios: (a) ATE; (b) T-RTE; (c) D-RTE; (d) PDE.

Table 5
Accuracy comparison in typical scenarios.

Scenarios	ATE/m	T-RTE/m	D-RTE/m	PDE/%
Offices	0.98	0.050	0.039	0.97
Shopping malls	0.85	0.046	0.040	0.93
Streets	0.81	0.044	0.034	0.83
Parking lots	0.82	0.041	0.032	0.85
Average	**0.86**	**0.045**	**0.036**	**0.89**

form of CDF and boxplot. As shown in Fig. 11(a), the 80% ATE of offices, shopping malls, streets and parking lots are 1.90 m, 1.87 m, 1.83 m and 1.84 m, respectively. As shown in Fig. 11(b), the 80% T-RTE of offices, shopping malls, streets and parking lots are 0.092 m, 0.099 m, 0.091 m and 0.096 m, respectively. As shown in Fig. 11(c), the 80% D-RTE of offices, shopping malls, streets and parking lots are 0.074 m, 0.079 m, 0.075 m, and 0.072 m, respectively. As shown in Fig. 11(d), the average PDE of offices, shopping malls, streets and parking lots are 0.97%, 0.93%, 0.83% and 0.85%, respectively. The localization performance of the four scenarios is very similar, which proves that the proposed method has satisfactory universality and robustness. Table 5 shows that the average ATE, T-RTE, D-RTE and PDE of the proposed method in typical scenarios are 0.86 m, 0.045 m, 0.036 m, and 0.89%, respectively.

4.4. Comparison with other methods

To justify the superiority of our proposed bipedal INS method based on equality constraint and ellipsoid constraint, we compared our proposed method with the following foot-mounted INS.

(1) The IEZ (INS + EKF + ZUPT) framework was proposed by Foxlin et al. [34], which uses ZUPT method to estimate and correct the drift error of the sensors, and then uses an Extended Kalman Filter (EKF) to correct the solution results of the INS.
(2) Wagstaff et al. [31] trained an Support Vector Machine (SVM) to classify pedestrian's movement modes, adaptively updated the ZVI detector threshold and trained an LSTM-based learning strategy to directly classify pedestrians' state (stationary or moving).

Fig. 12 compares our proposed method and abovementioned methods in a comprehensive environment, including multiple indoor and outdoor scenes. The preplanned comprehensive path is up to 3.3 km, covering typical scenarios, such as cover football fields, indoor passages, round flower beds, streets, outdoor stairs and parking lots. In the comparative experiment, we invited 5 volunteers, and each volunteer walked along the planned path 3 times.

The experimental results are shown in Fig. 13. The yellow line of the proposed method is steeper than the other plots, which indicates that the overall error of our proposed method are significantly lower than those of the compared methods. As shown in Fig. 13(a), 80% ATE of our proposed method is 3.51 m, while those of Wagstaff's method and IEZ are 5.47 m and 8.89 m, respectively. As shown in Fig. 13(b), 80% T-RTE of our proposed method is 0.10 m, while those of Wagstaff's method and IEZ are 0.15 m and 0.22 m, respectively. As shown in Fig. 13(c), 80% D-RTE of our proposed method is 0.06 m, while those of Wagstaff's method and IEZ are 0.10 m and 0.16 m, respectively. Table 6 demonstrates that our proposed method achieves the best results in terms of ATE, T-RTE,
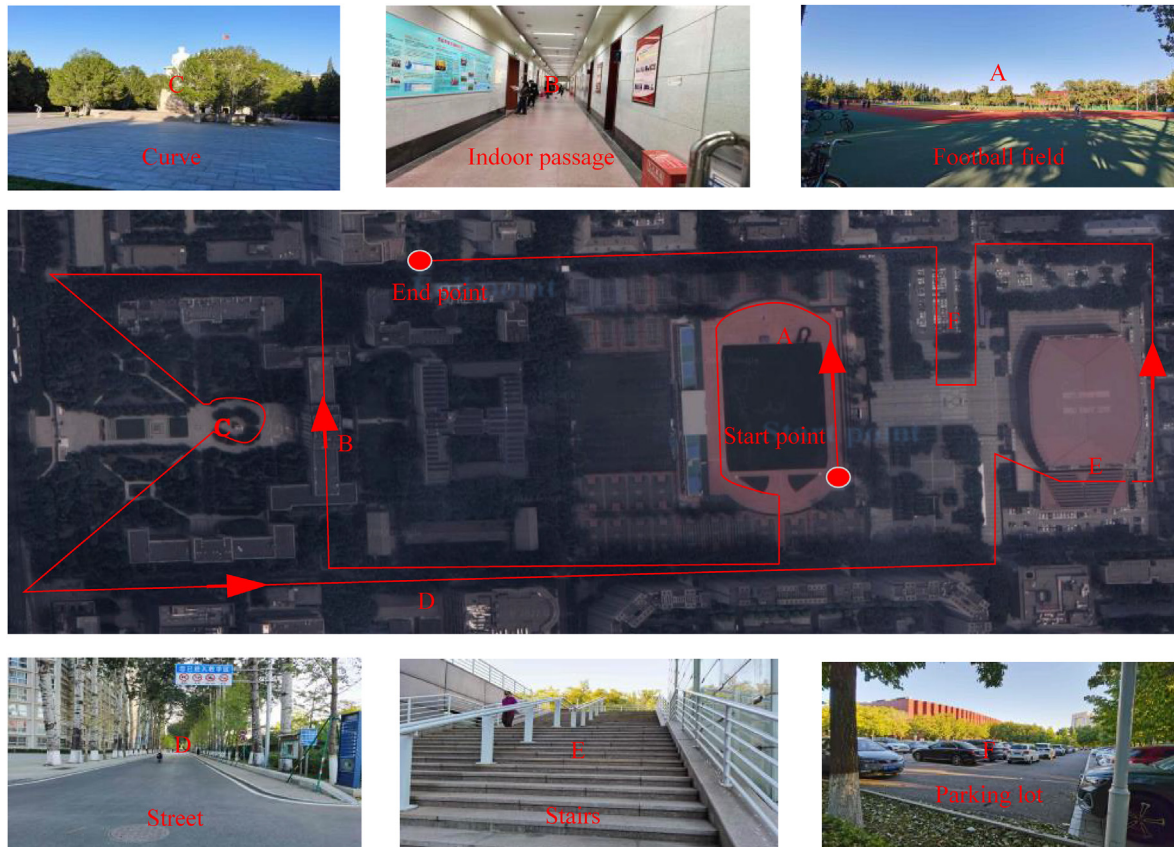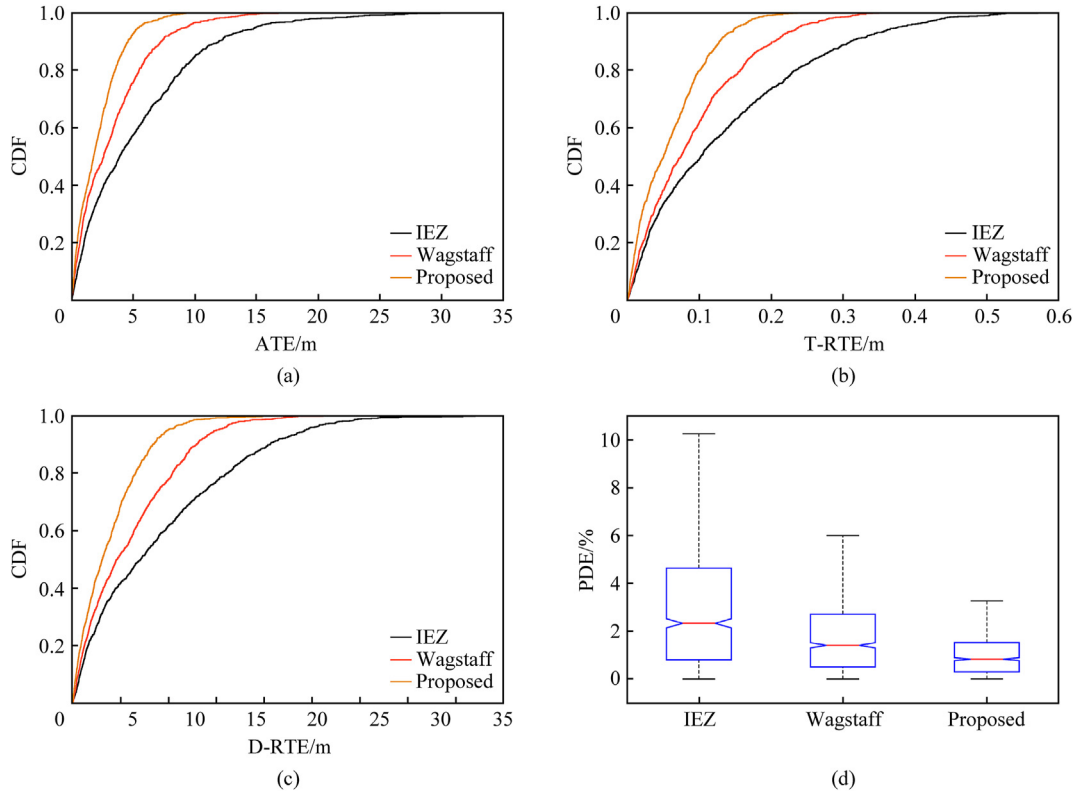

Fig. 12. Preplanned comprehensive paths.

Fig. 13. Accuracy comparison of different methods: (a) ATE; (b) T-RTE; (c) D-RTE. (d) PDE.

Table 6
Accuracy comparison of different methods.

Scenarios	ATE/m	T-RTE/m	D-RTE/m	PDE/%
IEZ	3.88	0.11	0.07	2.33
Wagstaff et al.	2.61	0.07	0.04	1.41
Proposed	**1.77**	**0.05**	**0.03**	**0.83**

D-RTE and PDE metrics. As shown in Table 6, the proposed method improves the average ATE by 32.18% and 54.38% over Wagstaff`s method and IEZ, respectively. Wagstaff`s method adaptively adjust the zero-velocity detection threshold according to the motion mode to obtain more accurate zero-velocity detection and positioning accuracy. Wagstaff`s method and IEZ are based on the sliding window for zero-velocity interval detection, only knowing whether a certain window is zero-velocity, and cannot specifically detect whether each observation is zero-velocity. The proposed method provides zero-velocity judgment for each observation, while using equality constraints and ellipsoid constraints to effectively suppress positioning errors.

5. Conclusions

The main issues of inertial navigation are that the observation contains noise, the pedestrian's complex and changeable movement patterns lead to inaccurate ZVI detection, and the heading and elevation error are difficult to suppress. To conquer these issues, we propose a pedestrian inertial navigation method named free-walking that consists of a sensor denoising module, a sample-level ZVI detection module, and a dual-foot INS module based on intra-foot ranging. The proposed method not only improves the robustness of ZVI detection in complex motion modes, but also provides sample-level ZVI detection. We proposed a bipedal INS combined with inter-foot ranging. Based on bipedal INS, we establish equation constraints and ellipsoid constraint models, and integrate the bipedal distance as an observation into the bipedal system through EKF. The method can effectively suppress heading and altitude error, and achieve long-term accurate and stable pedestrian positioning. The proposed method does not rely on any additional infrastructure and historical training data to achieve indoor localization. Therefore, the proposed method is not only easy to deploy in arbitrary unfamiliar and unprepared environments, but also easy to integrate with other localization algorithms or IoTs systems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Ashraf I, Hur S, Park Y. Recent advancements in indoor positioning and localization. Electronics 2022;11:2047. https://doi.org/10.3390/electronics11132047.

[2] Guo X, Ansari N, Hu F, Shao Y, Elikplim NR, Li L. A survey on fusion-based indoor positioning. IEEE Commun Surv Tutorials 2020;22:566–94. https://doi.org/10.1109/COMST.2019.2951036.

[3] Dai H, Bian H, Wang R, Ma H. An INS/GNSS integrated navigation in GNSS denied environment using recurrent neural network. Def Technol 2020;16: 334–40. https://doi.org/10.1016/j.dt.2019.08.011.

[4] Zhang M, Dai H, Hu B, Chen Q. Robust adaptive UKF based on SVR for inertial based integrated navigation. Def Technol 2020;16:846–55. https://doi.org/10.1016/j.dt.2019.10.012.

[5] Chen R, Yuan S, Ma C, Zhao H, Feng Z-Y. Tailored hidden Markov model: a tailored hidden Markov model optimized for cellular-based map matching. IEEE Trans Ind Electron 2022;69:13818–27. https://doi.org/10.1109/TIE.2021.3135645.

[6] Abdallah AA, Jao C, Kassas ZM, Shkel AM. A pedestrian indoor navigation system using deep-learning-aided cellular signals and ZUPT-aided foot-mounted IMUs. IEEE Sensor J 2022;22:5188–98. https://doi.org/10.1109/JSEN.2021.3118695.

[7] Lai J, Luo C, Wu J, Li J, Wang J, Chen J, et al. TagSort : accurate relative localization exploring RFID phase spectrum matching for Internet of Things. IEEE Int Things J 2020;7:389–99. https://doi.org/10.1109/JIOT.2019.2950174.

[8] Sağlam A, Baykan N. A new color distance measure formulated from the cooperation of the euclidean and the vector angular differences for lidar point cloud segmentation. Int J Eng Geosci 2020;6:117–24. https://doi.org/10.26833/ijeg.709212.

[9] Meng Q, Song Y, Li S, Zhuang Y. Resilient tightly coupled INS/UWB integration method for indoor UAV navigation under challenging scenarios. Def Technol 2022. https://doi.org/10.1016/j.dt.2022.12.013.

[10] Wen Z, Yang G, Cai Q, Chen T. A novel bluetooth-odometer-aided smartphone-based vehicular navigation in satellite-denied environments. IEEE Trans Ind Electron 2023;70:3136–46. https://doi.org/10.1109/TIE.2022.3169714.

[11] Hao L, Huang B, Jia B, Mao G. DHCLoc: a device-heterogeneity-tolerant and channel-adaptive passive WiFi localization method based on DNN. IEEE Internet Things J 2022;9:4863–74. https://doi.org/10.1109/JIOT.2021.3106902.

[12] Ai Y, Rui T, Yang X, He J, Fu L, Li J, et al. Visual SLAM in dynamic environments based on object detection. Def Technol 2021;17:1712–21. https://doi.org/10.1016/j.dt.2020.09.012.

[13] Cui J, Zhang F, Feng D, Li C, Li F, Tian Q. An improved SLAM based on RK-VIF: vision and inertial information fusion via Runge-Kutta method. Def Technol 2021. https://doi.org/10.1016/j.dt.2021.10.009.

[14] Tran HQ, Ha C. Machine learning in indoor visible light positioning systems: a review. Neurocomputing 2022;491:117–31. https://doi.org/10.1016/j.neucom.2021.10.123.

[15] Li Y, Chen R, Niu X, Zhuang Y, Gao Z, Hu X, et al. Inertial sensing meets machine learning: opportunity or challenge? IEEE Trans Intell Transport Syst 2022;23:9995–10011. https://doi.org/10.1109/TITS.2021.3097385.

[16] Wahlstrom J, Skog I. Fifteen years of progress at zero velocity: a review. IEEE Sensor J 2021;21:1139–51. https://doi.org/10.1109/JSEN.2020.3018880.

[17] Abiad N Al, Kone Y, Renaudin V, Robert T. Smartstep: a robust STEP detection method based on SMARTphone inertial signals driven by gait learning. IEEE Sensor J 2022;22:12288–97. https://doi.org/10.1109/JSEN.2022.3169621.

[18] Soni R, Trapasiya S. A survey of step length estimation models based on inertial sensors for indoor navigation systems. Int J Commun Syst 2022;35: 1–22. https://doi.org/10.1002/dac.5053.

[19] Manos A, Hazan T, Klein I. Walking direction estimation using smartphone sensors: a deep network-based framework. IEEE Trans Instrum Meas 2022;71: 1–12. https://doi.org/10.1109/TIM.2022.3144225.

[20] Wang Q, Luo H, Xiong H, Men A, Zhao F, Xia M, et al. Pedestrian dead reckoning based on walking pattern recognition and online magnetic fingerprint trajectory calibration. IEEE Internet Things J 2021;8:2011. https://doi.org/10.1109/JIOT.2020.3016146. −26.

[21] Wang Y, Cheng H, Meng MQ-H. Inertial odometry using hybrid neural network with temporal attention for pedestrian localization. IEEE Trans Instrum Meas 2022;71:1–10. https://doi.org/10.1109/TIM.2022.3186704.

[22] Jao C, Wang Y, Shkel AM. Pedestrian inertial navigation system augmented by vision-based foot-to-foot relative position measurements. IEEE/ION Position, Locat. Navig. Symp., IEEE; 2020. https://doi.org/10.1109/PLANS46316.2020.9109993. 2020, pp. 900–7.

[23] Chen Z, Pan X, Chen C, Wu M. Contrastive learning of zero-velocity detection for pedestrian inertial navigation. IEEE Sensor J 2022;22:4962–9. https://doi.org/10.1109/JSEN.2021.3072160.

[24] Wahlstrom J, Markham A, Trigoni N. FootSLAM meets adaptive thresholding. IEEE Sensor J 2020;20:9351–8. https://doi.org/10.1109/JSEN.2020.2987813.

[25] Liu X, Wang S, Zhang T, Huang R, Wang Q. A zero-velocity detection method with transformation on generalized likelihood ratio statistical curve. Measurement 2018;127:463–71. https://doi.org/10.1016/j.measurement.2018.05.113.

[26] Ren M, Pan K, Liu Y, Guo H, Zhang X, Wang P. A novel pedestrian navigation

[27] Yang Y, Zhao D, Liu W, Cao H, Wang C, Tang J, et al. Pitch angle sliding variance test method based on Mahony filter for zero-velocity detection. Measurement 2022;205:112145. https://doi.org/10.1016/j.measurement.2022.112145.

[28] Wahlstrom J, Skog I, Gustafsson F, Markham A, Trigoni N. Zero-velocity detection—a bayesian approach to adaptive thresholding. IEEE Sensors Lett 2019;3:1–4. https://doi.org/10.1109/LSENS.2019.2917055.

[29] Chen C, Zhao P, Lu CX, Wang W, Markham A, Trigoni N. Deep-learning-based pedestrian inertial navigation: methods, data set, and on-device inference. IEEE Internet Things J 2020;7:4431–41. https://doi.org/10.1109/JIOT.2020.2966773.

[30] Zhu T, Pan X, Zhang S. A zero velocity detection method for soldier navigation based on deep learning. J Phys Conf Ser 2019;1345:032018. https://doi.org/10.1088/1742-6596/1345/3/032018.

[31] Wagstaff B, Peretroukhin V, Kelly J, Member S. Robust data-driven zero-velocity detection for foot-mounted inertial navigation. IEEE Sensor J 2020;20: 957–67. https://doi.org/10.1109/JSEN.2019.2944412.

[32] Deng Z, Wang P, Yan D, Shang K. Foot-mounted pedestrian navigation method based on gait classification for three-dimensional positioning. IEEE Sensor J 2020;20. https://doi.org/10.1109/JSEN.2019. 2949060. 2045–55.

[33] Yang M, Zhu R, Xiao Z, Yan B. Symmetrical-net: adaptive zero velocity detection for ZUPT-aided pedestrian navigation system. IEEE Sensor J 2022;22:5075–85. https://doi.org/10.1109/JSEN.2021.3094301.

[34] Foxlin E, Intersense A, Foxlin E. Pedestrian tracking with shoe-mounted inertial sensors. IEEE Comput Graph Appl 2005;25:38–46. https://doi.org/10.1109/MCG.2005.140.

[35] Zhang L, Su Z, Li Q. Pedestrian navigation using iZES framework for bounding heading drift. Commun. Comput. Inf. Sci. 2016:235–44. https://doi.org/10.1007/978-981-10-2672-0_24.

[36] Sang S, Zhai R, Zhang W, Sun Q, Zhou Z. A self-developed indoor three-dimensional pedestrian localization platform based on MEMS sensors. Sens Rev 2015;35:157–67. https://doi.org/10.1108/SR-07-2014-682.

[37] Zhu M, Wu Y, Luo S. f²IMU-R: pedestrian navigation by low-cost foot-mounted dual IMUs and Interfoot ranging. IEEE Trans Control Syst Technol 2022;30:247–60. https://doi.org/10.1109/TCST.2021.3063533.

[38] Wu C-J, Kuo C-H, Lin Y-H, Liu W-Y. A feasible model training for LSTM-based dual foot-mounted pedestrian INS. IEEE Sensor J 2021;21:13616–27. https://doi.org/10.1109/JSEN.2021.3070534.

[39] Skog I, Nilsson J-O, Handel P. an open-source multi inertial measurement unit (MIMU) platform. 2014 Int. Symp. Inert. Sensors Syst.. IEEE; 2014. p. 1–4. https://doi.org/10.1109/ISISS.2014.6782523.

[40] Chen S, Lu Z, Xu X, Liu J, Bi Z. Foot-mounted dual-sensor single-board pedestrian inertial navigation system based on position and velocity constraints. Sensor Mater 2022;34:2075. https://doi.org/10.18494/SAM3853.

[41] Borenstein J, Ojeda L. Heuristic drift elimination for personnel tracking systems. J Navig 2010;63:591–606. https://doi.org/10.1017/S0373463310000184.

[42] Abdulrahim K, Hide C, Moore T, Hill C. Integrating low cost IMU with building heading in indoor pedestrian navigation. J Glob Position Syst 2011;10:30–8. https://doi.org/10.5081/jgps.10.1.30.

[43] Sahoo S, Saboo M, Pratihar DK, Mukhopadhyay S. Real-time detection of actual and early gait events during level-ground and ramp walking. IEEE Sensor J 2020;20:8128–36. https://doi.org/10.1109/JSEN.2020.2980863.

[44] Zhu R, Wang Y, Cao H, Yu B, Gan X, Huang L, et al. RTK/Pseudolite/LAHDE/IMU-PDR integrated pedestrian navigation system for urban and indoor environments. Sensors 2020;20:1791. https://doi.org/10.3390/s20061791.

[45] Chen J, Zhou B, Bao S, Liu X, Gu Z, Li L, et al. A data-driven inertial navigation/bluetooth fusion algorithm for indoor localization. IEEE Sensor J 2022;22: 5288–301. https://doi.org/10.1109/JSEN.2021.3089516.

[46] Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders. Proc. 25th Int. Conf. Mach. Learn. - '08. New York, New York, USA: ACM Press; 2008. p. 1096–103. https://doi.org/10.1145/1390156.1390294.

[47] Khandelwal S. Gait event detection in the real world. Halmstad University; 2018.

[48] Zhu C, Tomizawa Y, Luo X, Kawamura A. Biped walking with variable ZMP, frictional constraint, and inverted pendulum model. Proc - 2004 IEEE Int Conf Robot Biomimetics, IEEE ROBIO 2004. https://doi.org/10.1109/robio.2004.1521816. 425–30.

[49] Ruppelt J, Kronenwett N, Trommer GF. A novel finite state machine based step detection technique for pedestrian navigation systems. 2015 Int. Conf. Indoor Position. Indoor Navig.. IEEE; 2015. p. 1–7. https://doi.org/10.1109/IPIN.2015.7346771.

[50] Skog I, Händel P, Nilsson JO, Rantakokko J. Zero-velocity detection-An algorithm evaluation. IEEE Trans Biomed Eng 2010;57:2657–66. https://doi.org/10.1109/TBME.2010.2060723.

[51] Godha S, Lachapelle G. Foot mounted inertial system for pedestrian navigation. Meas Sci Technol 2008;19:075202. https://doi.org/10.1088/0957-0233/19/7/075202.

[52] Ojeda L, Borenstein J. Non-GPS navigation for security personnel and first Responders. J Navig 2007;60:391. https://doi.org/10.1017/S0373463307004286.

[53] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for

biomedical image segmentation. Lect Notes Comput Sci 2015:234—41. https://doi.org/10.1007/978-3-319-24574-4_28.

[54] Kuang J, Niu X, Chen X. Robust pedestrian dead reckoning based on MEMS-IMU for smartphones. Sensors 2018;18:1391. https://doi.org/10.3390/s18051391.

[55] Jiménez AR, Seco F, Zampella F, Prieto JC, Guevara J. PDR with a foot-mounted IMU and ramp detection. Sensors 2011;11:9393—410. https://doi.org/10.3390/s111009393.

[56] ICM-42605 n.d. https://invensense.tdk.com/products/motion-tracking/6-axis/icm-42605/. [Accessed 20 September 2022]. accessed

[57] Skog I, Handel P, Nilsson JO, Rantakokko J. Zero-velocity detection—an algorithm evaluation. IEEE Trans Biomed Eng 2010;57:2657—66. https://doi.org/10.1109/TBME.2010.2060723.