



Article

Autonomous Multi-Floor Localization Based on Smartphone-Integrated Sensors and Pedestrian Indoor Network

Chaoyang Shi ¹, Wenxin Teng ², Yi Zhang ³, Yue Yu ^{4,*}, Liang Chen ², Ruizhi Chen ² and Qingquan Li ⁵

¹ School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan 430079, China

² State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan 430079, China

³ Department of Building and Real Estate, The Hong Kong Polytechnic University, Hong Kong 999077, China

⁴ Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong 999077, China

⁵ Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China

* Correspondence: yue806.yu@connect.polyu.hk; Tel.: +852-6218-2854

Abstract: Autonomous localization without local wireless facilities is proven as an efficient way for realizing location-based services in complex urban environments. The precision of the current map-matching algorithms is subject to the poor ability of integrated sensor-based trajectory estimation and the efficient combination of pedestrian motion information and the pedestrian indoor network. This paper proposes an autonomous multi-floor localization framework based on smartphone-integrated sensors and pedestrian network matching (ML-ISNM). A robust data and model dual-driven pedestrian trajectory estimator is proposed for accurate integrated sensor-based positioning under different handheld modes and disturbed environments. A bi-directional long short-term memory (Bi-LSTM) network is further applied for floor identification using extracted environmental features and pedestrian motion features, and further combined with the indoor network matching algorithm for acquiring accurate location and floor observations. In the multi-source fusion procedure, an error ellipse-enhanced unscented Kalman filter is developed for the intelligent combination of a trajectory estimator, human motion constraints, and the extracted pedestrian network. Comprehensive experiments indicate that the presented ML-ISNM achieves autonomous and accurate multi-floor positioning performance in complex and large-scale urban buildings. The final evaluated average localization error was lower than 1.13 m without the assistance of wireless facilities or a navigation database.

Keywords: autonomous localization; data and model dual-driven; trajectory estimator; floor identification; error ellipse; unscented Kalman filter



Citation: Shi, C.; Teng, W.; Zhang, Y.; Yu, Y.; Chen, L.; Chen, R.; Li, Q. Autonomous Multi-Floor Localization Based on Smartphone-Integrated Sensors and Pedestrian Indoor Network. *Remote Sens.* **2023**, *15*, 2933. <https://doi.org/10.3390/rs15112933>

Academic Editors: Jay Hyoun Kwon, Chang-Ki Hong, Tae-Suk Bae and Hung-Kyu Lee

Received: 9 May 2023

Revised: 23 May 2023

Accepted: 3 June 2023

Published: 4 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Accurate localization within indoor environments is considered a crucial component of intelligent urban infrastructure. Due to the intricate and multifaceted nature of indoor spaces, it is imperative to develop autonomous and cost-effective indoor positioning services. A variety of existing indoor positioning systems (IPs), including Wireless Fidelity (Wi-Fi) [1], Bluetooth Low Energy (BLE) [2], ultra-wideband (UWB) [3], sound source [4], and integrated sensors [5], have been developed to provide indoor positioning with varying degrees of precision. However, further advancements are necessary to ensure optimal performance in complex indoor environments, which are typical of modern smart cities. Among IPs, at this stage, the existing positioning systems mostly rely on local facilities or wireless signals, and meter-level localization precision can be achieved with a combination

of external signals and smartphone-integrated sensors, which is labor-consuming and high cost. A comparison between existing IPSs is shown in Table 1.

Table 1. Comparison of existing positioning technologies.

Location Sources	Accuracy	Robustness	Complexity	Scalability	Cost
Wi-Fi [1]	Fingerprinting: 2–5 m; RTT ranging: 1–3 m	Affected by environmental factors and human bodies	Time-consuming for database generation	Easy	No additional cost
Bluetooth/BLE [2]	Fingerprinting: 2–5 m AOA array: <1 m	Limited by the changeable environments	Time-consuming to construct the database	Easy	High cost of antenna array
UWB [3]	Centimeter level	Good	Medium	Medium	High
Sound Source [4]	Meter level	Affected by NLOS factor	Medium	Good	Medium
Integrated Sensors [5]	Cumulative error exists	Good	Medium	Good	Low
NFC [6]	Centimeter level, short effective distance	Good	Low	Easy	A large number of NFC tags are required
Cellular Network [7]	Ten meters to tens of meters	Affected by environments	Medium	Good	High
RFID [8]	1–5 m	Affected by environments	Medium	Medium	Medium
Infrared Ray [9]	Meter level	LOS required	Medium	Low	Medium, additional transceivers are required
Visible Light [10]	1–5 m	Medium	Medium	Good	Low
Ultrasound [11]	Centimeter level	Good	Low	Low	Medium, additional transceivers are required
Magnetic Field [12]	2–5 m	Affected by environments	High	Good	Low
Computer Vision [13]	Camera rendezvous: centimeter level; others: meter level	Medium, affected by the ambient light and quality of the image	Very high	Good	Medium

The autonomous localization system (ALS) is advanced by its autonomy characteristics because no additional facilities or pre-collected fingerprinting database are required in the positioning phase, and the efficacy of the ALS is hindered by the imprecise nature of the trajectory data obtained from smartphones. This inaccuracy may be attributed to the varied ways in which users hold their devices [14], and the accumulation of errors within the sensors integrated into mobile terminals [15]. As a result, the accuracy of the ALS could be significantly improved through the development of more precise trajectory data collection methods and optimization of the sensor technology in mobile devices. The existing problems include complex human motion modes [16] and a lack of an efficient combination of built-in sensor-based location sources and an existing indoor map or pedestrian network information [17], which are the main factors that affect the accuracy of ALS.

Numerous research studies have been conducted to address the aforementioned issues. For instance, Yan et al. [18] proposed a robust smartphone-based RIDI system that estimates walking speed and localization information accurately, even with varying handheld positions. Their approach yielded more efficient and similar accuracy than the traditional visual inertial odometry (VIO) method. Additionally, they developed the RoNIN network to enhance the precision and stability of inertial odometry using a new dataset comprising more than 40 h of data from smartphone-integrated sensors under complex pedestrian motion modes [19]. Guo et al. [20] proposed a machine learning-based classification approach to estimate walking speed under complex pedestrian motion modes by incorporating a handheld mode awareness strategy. Various mobile data were collected, and model training processes were conducted to evaluate the accuracy of the proposed speed estimator. Similarly, Zhang et al. [21] developed the SmartMTra framework, which leverages feature extraction and motion detection techniques, along with handheld mode classification, to achieve robust dead reckoning performance. Furthermore, Poulouse et al. [22] comprehensively evaluated the heading calculation precision using different multi-sensor fusion algorithms—namely the Kalman filter (KF), extended Kalman filter (EKF), unscented Kalman filter (UKF), particle filter (PF), and complementary filters (CF). Their evaluation results demonstrated that the UKF realized much better heading estimation accuracy, while PF showed the lowest accuracy, which is not suitable for heading calculation. These studies have highlighted the importance of developing robust and accurate solutions for indoor positioning and tracking in complex urban environments.

Indoor mapping information is essential for generating and updating crowdsourced databases with high accuracy. To improve map-matching precision while reducing calculation complexity, Wu et al. [23] proposed the HTrack system, which takes into account pedestrian heading and geospatial data. Xia et al. [24] combined the pedestrian dead reckoning model with BLE-based distance measurement and map constraints. The particle filter was applied to enhance localization robustness and precision, resulting in an RMSE of 1.48 m. Li et al. [25] introduced a fingerprinting precision-level predictor that can autonomously evaluate the performance of Wi-Fi- and magnetic field-provided locations based on wireless indexes, map information, and database-related characteristics. This approach significantly increases the precision of integrated positioning based on the crowdsensing trajectories. The importance of accurate and efficient indoor mapping approaches has been demonstrated by these studies, as they are critical for achieving reliable indoor tracking and positioning.

Detecting the correct floor level is crucial for enhancing the effectiveness and accuracy of classifying trajectory data collected through crowd-sourcing, as well as for generating comprehensive databases. In a study by Zhao et al. [26], they proposed an HYFI system that estimates the initial floor level based on the distribution of local wireless access points (APs). To minimize the effects of environmental factors, the system also integrated pressure information into its approach. The authors reported an accuracy rate exceeding 96.1% when compared to using only a single source. Similarly, Shao et al. [27] introduced an adaptive algorithm for detecting floors using wireless technology, which is especially suitable for large-scale indoor areas with multiple floors. Their method involves analyzing the Wi-Fi radio signal strength indicator (RSSI) and spatial similarity features and segmenting local environments using block models, and achieved an average accuracy rate of 97.24%. These studies have demonstrated the importance of developing robust and accurate floor detection approaches to ensure reliable indoor tracking and positioning in complex indoor environments. Thus, it is crucial to continue the research on improving floor detection algorithms to enhance the overall performance of indoor positioning systems.

The classical smartphone-based indoor mapping and trajectory optimization structures include Walkie-Markie [28], PiLoc [29], and MPiLoc [30], among others. In Walkie-Markie, the indoor pathway is generated using the basis of the detection of Wi-Fi AP-based landmarks and trajectory matching. The limitation is that the collected RSSI value is subjected to changeable indoor environments and the absolute location of the generated pathway cannot be acquired. PiLoc classifies similar crowdsourced trajectories by their shapes and the similarity of collected Wi-Fi RSSI information and merges the similar trajectories using point-to-point fusion. MPiLoc further extends the floor plan from 2D to 3D and uses the sparsely acquired GNSS-reported locations as the absolute points. The disadvantage is that both PiLoc and MPiLoc rely on the accurate estimation of heading information, while the precise absolute heading may not be available all the time. Li et al. [31] presented the IndoorWaze system, using the crowdsourced Wi-Fi fingerprinting data and POI information collected by shopping mall employees to generate a robust and marked floor plan. The final experiments showed that the designed IndoorWaze framework can accurately mark the pathways and location of the store for indoor navigation purposes.

Integrated navigation technology has become more and more popular due to its improved robustness and precision compared to single-location sources in complex indoor environments. Several fusion methods, such as the Kalman filter (KF) [32], extended Kalman filter (EKF) [33], unscented Kalman filter (UKF) [34], and particle filter (PF) [35], are applied as the typical integration methods for multi-source fusion-based indoor localization. Huang et al. [36] developed a cost-effective and user-convenient indoor localization technique that utilizes existing Li-Fi lighting and Wi-Fi infrastructure to achieve significantly improved positioning accuracy. Their technique involves Li-Fi-assisted coefficient calibration and was experimentally verified. Chen et al. [37] developed an indoor dynamic positioning method that utilizes the symmetrical characteristics of human motion to quickly define the basis of the human motion process and address existing issues. They introduced

an ultra-wideband (UWB) method and applied an unscented Kalman filter to fuse inertial sensors and UWB data. Inertial positioning compensated for UWB signal obstacles, while UWB positioning overcame the error accumulation of the inertial positioning. Chen et al. [38] proposed an INS and Wi-Fi integration model that utilizes multi-dimensional dynamic time warping (MDTW) to calculate the distance between the measured signals and fingerprints in a database. They also introduced an MDTW-based weighted least squares (WLS) method for fusing multiple fingerprint localization results, resulting in improved positioning accuracy and robustness.

To further enhance the performance of smartphone-integrated sensor-based pedestrian trajectory estimation and optimization, and indoor network extraction and trajectory matching, this paper proposes an ML-ISNM framework, which realizes accurate multi-floor positioning performance without the assistance of local facilities. The main contributions of this work are summarized as follows:

- (1) This paper proposes a robust data and model dual-driven pedestrian trajectory estimator for accurate integrated sensor-based positioning in complex motion modes and disturbed environments. The proposed approach considers factors such as handheld modes, lateral errors, and step-length constraints while updating the location based on a period of observations rather than solely relying on the last moment;
- (2) A new floor detection algorithm based on Bi-LSTM is implemented to offer floor index references for estimated trajectories. This involves extracting hybrid features from wireless signals, human motion, and map-related data to improve the recognition precision, leading to a more accurate initial location and floor information provided to users;
- (3) This work models an extracted pedestrian indoor network, formulates it as the combination of a matrix, and develops a grid search algorithm for network matching and further walking route calibration with the reference of the initial location and floor detection results. The matched network information is further applied as the observation under the fusion phase;
- (4) Using the outcomes of trajectory estimation, floor recognition, and indoor network matching, an error ellipse-supported unscented Kalman filter (EE-UKF) is suggested to robustly combine data from integrated sensors, pedestrian motion, and indoor network information. This approach can achieve meter-level positioning accuracy without requiring additional local facilities for assistance.

The following content is outlined in the sections of this article. Section 2 provides an overview of the trajectory estimator based on a data and model dual-driven approach. Section 3 describes the developed Bi-LSTM-based floor recognition, trajectory matching, calibration, and error ellipse-supported UKF-based intelligent integration. Section 4 discusses the evaluation results of the developed ML-ISNM. Finally, Section 5 presents a conclusion of this work.

2. Data and Model Dual-Driven Trajectory Estimator

This research introduces an ML-ISNM framework, which combines trajectory estimation, multi-level observations, and indoor network information to enable the autonomous detection of the user's location and floors without requiring local facilities, while also facilitating accurate multi-source fusion. The proposed trajectory estimator integrates sensor data from tri-gyroscopes, tri-accelerometers, tri-magnetometers, and barometers to obtain the raw 3D location, speed, and attitude vectors. Next, Bi-LSTM-based floor detection and pedestrian indoor network matching are applied to provide an accurate initial reference location and floor indexes for the proposed trajectory estimator. In addition, non-holonomic constraints, a quasi-static magnetic field, and an indoor network reference are generated as the multi-level observations and combined with the trajectory estimator by an enhanced EE-UKF, and the forward trajectory is further calibrated by the backward EE-UKF after obtaining the network's extracted landmark points. The overall structure of the developed ML-ISNM framework is described in Figure 1. This part presents an

accurate data and model dual-driven trajectory estimator under complex handheld modes and disturbed indoor environments, and the estimated trajectory information is further integrated with other observations in order to realize much more accurate and robust positioning performance.

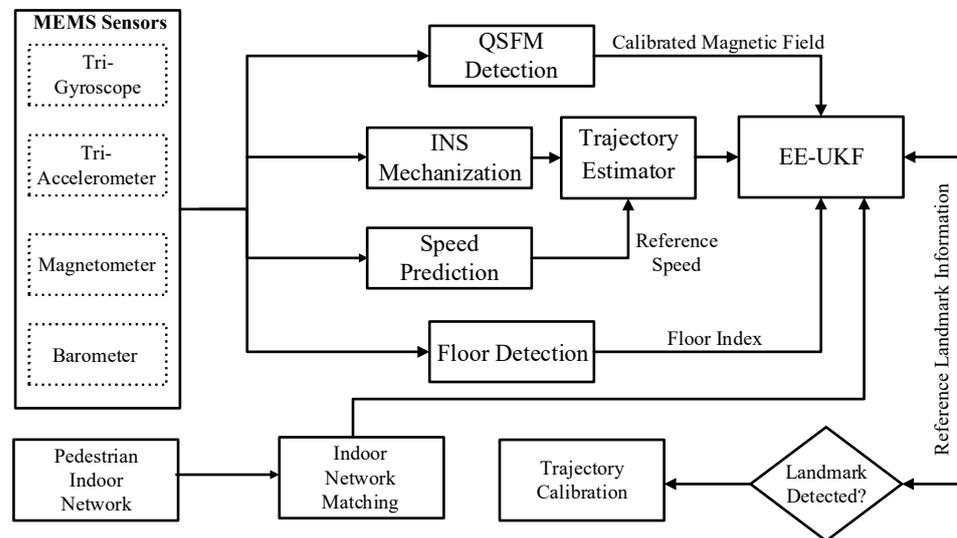


Figure 1. Structure of developed ML-ISNM.

2.1. Hybrid Deep-Learning Model Enhanced Walking Speed Prediction

When it comes to achieving inertial odometry for mobile devices that depend on indoor pedestrian localization, INS and PDR mechanisms are viewed as viable solutions. However, conventional approaches using INS/PDR have limited precision because of the varied handheld modes of these devices and accumulated errors in their inertial sensors. Furthermore, relying solely on the previous moment for location updates through these methods may result in missing crucial motion information during the chosen walking period.

To address these limitations, this work proposes a novel data and model dual-driven trajectory estimation approach that leverages advanced machine learning techniques to enhance the accuracy and robustness of inertial odometry. This approach considers factors such as handheld positioning, lateral error, and step-length constraints while updating the location based on a period of observations rather than solely relying on the last moment. By incorporating these techniques into INS/PDR-based systems, the resulting inertial odometry is significantly more accurate and reliable, making them ideal for deployment in complex indoor environments.

This research addresses the limitations of the existing dead reckoning (DR) methods by proposing an enhanced deep learning-based walking velocity estimator. The aim was to realize precise speed observation for PDR mechanization, taking into account the varied handheld modes of mobile terminals and overcoming cumulative errors that can limit accuracy. To detect the different handheld modes of existing smartphones, this approach leverages MLP models that consider period-specific acceleration vectors, the angular speed vector, and related modeled characteristics. Additionally, the Bi-LSTM structure is used to predict pedestrians' continuous walking speed using similar inputs. Figure 2 depicts the whole framework of the developed deep-learning-based approach for walking velocity prediction. By utilizing these advanced machine learning techniques, the resulting PDR mechanism achieves significantly higher accuracy and robustness in indoor tracking and positioning, which makes it ideal for deployment in challenging indoor environments where signal interference and other environmental factors can pose significant challenges.

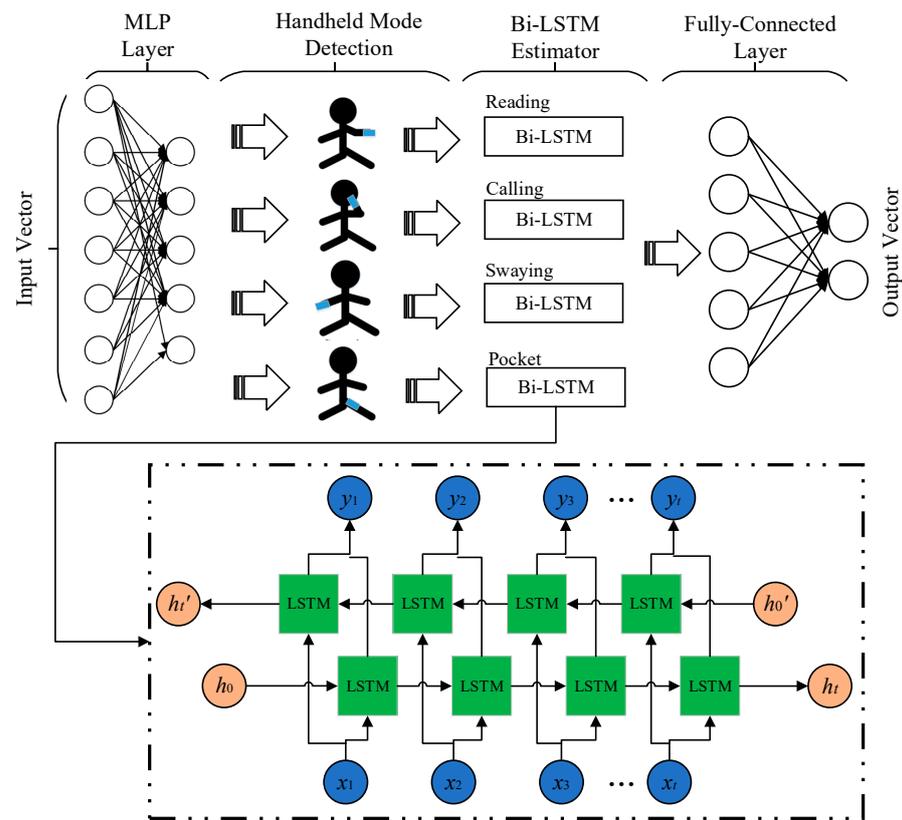


Figure 2. Deep-learning model for walking velocity estimation.

The above diagram depicts the structure of the proposed walking speed estimator, which relies on deep learning. The model comprises several components, with the MLP serving as the initial element to detect handheld modes. The detected modes and generated features are then input into the Bi-LSTM model, which provides precise estimates of pedestrian walking speed-originated features. Lastly, a fully-connected layer is utilized for final-result modeling to deliver real-time estimations of pedestrian moving velocity information. By integrating these advanced machine learning techniques into the design of the PDR mechanism, the resulting system achieves higher accuracy and robustness in indoor tracking and positioning. This can substantially improve the user experience, particularly in complex indoor environments where traditional dead reckoning methods can result in significant inaccuracies due to cumulative errors and signal interference.

To correspondingly estimate walking speed and detect handheld modes, the accelerometer and gyroscope's smoothed data are employed as input features in the MLP. The detected handheld modes derived from these features are then utilized to train a Bi-LSTM framework that can predict walking velocity information under different handheld modes. For increasing the speed estimation precision, an input vector of 3 s in overall length is chosen from a dataset sampled at a rate of 50 Hz. These enhanced input features can then be used to generate more precise estimates of walking speed while also detecting handheld modes with greater accuracy.

Finally, the MLP network $MLP()$ is applied as the first layer of the developed deep-learning model, the raw sensor data are extracted as the input vector of MLP, and the four different handheld modes are adopted as the output vector of the MLP network [39]:

$$\beta_i = MLP(\alpha_i) \quad (1)$$

where β_i indicates the output vector of the MLP, which contains four different handheld modes of smartphones, and α_i is the input vector of the sensor data.

Next is the Bi-LSTM layer. The input vector of the Bi-LSTM model is the pre-processed sensor data and the detected handheld mode of the smartphones. The relationships among the different Bi-LSTM parameters are presented as follows [40]:

$$\begin{cases} f_t = \sigma \cdot (W_f \cdot [h_{t-1}, X_t] + b_f) \\ i_t = \sigma \cdot (W_i \cdot [h_{t-1}, X_t] + b_i) \\ \tilde{C}_t = \tanh \cdot (W_C \cdot [h_{t-1}, X_t] + b_C) \\ o_t = \sigma \cdot (W_o \cdot [h_{t-1}, X_t] + b_o) \\ h_t = o_t \cdot \tanh(C_t) \end{cases} \quad (2)$$

In the following equation i_t , f_t , and o_t denote the input, forget, and output units of the Bi-LSTM network, respectively. X_t represents the input vector of the Bi-LSTM layer under each time period, and h_t denotes the hidden state vector, which is considered the output of the Bi-LSTM layer and is further applied as the input of the full connection layer. The sigmoid function is denoted by σ , while C_t is the candidate vector that is merged with the output vector to form the memorized state under the recorded time.

In our deep learning-based speed estimator framework, we utilize walking velocity as the expected output vector of the model training procedure. It is important to note that the initial predicted walking speed lacks real-world geospatial reference and can only be considered as the forward velocity. This forward speed is defined as follows:

$$\omega^b = [\omega_{forward}^b \quad 0 \quad 0]^T \quad (3)$$

where $\omega_{forward}^b$ is presented according to the step-length calculation result.

To obtain the pedestrian's forward speed, it is necessary to transform the estimated walking speed according to the results obtained from handheld mode detection. The calculation for obtaining the forward velocity under the n-frame follows the formula outlined in [7]. This step ensures that the final output accurately reflects the pedestrian's true forward speed and takes into account any identified variations in handheld modes during the data collection:

$$\omega^n = \chi_e^n \chi_{e_1}^e \chi_b^{e_1} \omega^b \quad (4)$$

In the above equation, ω^n represents the NHC-based speed that has been converted. The attitude matrix calculated between the b-frame and the ENU frame is denoted by $\chi_b^{e_1}$. The translation matrix related to handheld modes, denoted by $\chi_{e_1}^e$, converts the axis related to heading into the reading mode-based heading-related axis in accordance with the results of the handheld mode recognition. Finally, the translation matrix from the ENU coordinate system to the NED coordinate system is indicated by χ_e^n .

The estimated location of the pedestrian can be determined by combining both the heading and walking velocity values in a time period.

$$r_i = r_{i-1} + \int_{t=i-1}^{t=i} \omega_i^n \quad (5)$$

In the above equation, ω_i^n represents the walking velocity value provided in a recognized step period and needs to be converted into the n-frame. The real-world position r_i is then calculated according to the formal location information.

2.2. Data and Model Dual-Driven Trajectory Estimator

INS mechanization is proposed for inertial sensor-based localization. The information of the acceleration and angular rate acquired from MEMS sensors are integrated by INS

mechanization for the estimation of the 3D position, velocity, and attitude of the moving object with a high update rate, which is shown below [5]:

$$\begin{bmatrix} \dot{p}^n \\ \dot{v}^n \\ \dot{C}_b^n \end{bmatrix} = \begin{bmatrix} \omega^{-1}v^n \\ C_b^n f^b - (2\omega_{ie}^n + \omega_{en}^n)v^n + g^n \\ C_b^n(\omega_{ib}^b - \omega_{in}^b) \end{bmatrix} \tag{6}$$

In the following equation, $p^n = [p_N \ p_E \ p_D]^T$ denotes the real-time 3D location of the pedestrian. The 3D velocity is represented by $v^n = [v_N \ v_E \ v_D]^T$, while C_b^n indicates the rotation matrix of the b-frame and n-frame. g^n represents the collected local gravity vector, and ω_{ie}^n denotes the rotation angular rate between the ECEF frame and i-frame. Additionally, ω_{en}^n indicates the rotation angular rate between the navigation coordinate system and the ECEF coordinate system. A 3×3 matrix related to the latitude and the ellipsoidal altitude of the selected object is denoted by ω^{-1} .

Due to the low precision of MEMS sensors, the Earth-related angular rate error vectors, ω_{ie}^n and ω_{en}^n , can be ignored. Hence, the simplified error model of INS can be described as follows [5]:

$$\begin{cases} \delta \dot{p}^n = -\omega_{en}^n \times \delta p^n + \delta v^n \\ \delta \dot{v}^n = -(2\omega_{ie}^n + \omega_{en}^n)\delta v^n + f^n \times \psi + C_b^n(\varepsilon_a + w_{ba}) \\ \dot{\psi} = -(\omega_{ie}^n + \omega_{en}^n) \times \psi - C_b^n(\varepsilon_g + w_{bg}) \\ \dot{\varepsilon}_g = \varepsilon_g / \tau_{bg} + w_{bg} \\ \dot{\varepsilon}_a = -\varepsilon_a / \tau_{ba} + w_{ba} \end{cases} \tag{7}$$

In the following equation, δp^n , δv^n , and ψ denote the measured errors of the 3D position, velocity, and attitude information, respectively. ε_g and ε_a represent the biases of the gyroscope and accelerometer, respectively. f^n indicates the converted acceleration data in the n-frame, while τ_{bg} and τ_{ba} represent the parameters related to sensor noise. Additionally, w_{bg} and w_{ba} are the measured noises of ε_g and ε_a , respectively.

Using the INS error model presented above, the state vector can be described in the AUKF as follows:

$$\delta x = \left[(\delta p^n)_{1 \times 3} \quad (\delta v^n)_{1 \times 3} \quad \psi_{1 \times 3} \quad (\varepsilon_g)_{1 \times 3} \quad (\varepsilon_a)_{1 \times 3} \right]^T \tag{8}$$

The discrete-time EE-UKF system equation and observation equation are:

$$\begin{cases} \delta x_t = F_{t-1,t} \delta x_{t-1} + v_t \\ \delta z_t = G_t \delta x_t + \zeta_t \end{cases} \tag{9}$$

where δx_t and δz_t represent the state vector and observed vector at the moment t ; G_t indicates the observation matrix at the moment t . v_t and ζ_t indicate the state noise and measurement noise at the timestamp t ; $F_{t-1,t}$ indicates the 15×15 state transition matrix, which is shown below:

$$F_{t-1,t} = \begin{bmatrix} \mathbf{I}_{3 \times 3} & \mathbf{I}_{3 \times 3} \times \Delta t & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{I}_{3 \times 3} & (f_k^n \times) \cdot \Delta t & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{I}_{3 \times 3} & -C_{b,k}^n \cdot \Delta t & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{I}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{0}_{3 \times 3} & \mathbf{I}_{3 \times 3} \end{bmatrix} \tag{10}$$

In the above equation, Δt represents the update interval of the INS mechanization, while f_k^n represents the collected accelerometer vector under the n-frame.

To address the influence of the indoor artificial magnetic fields added to the accuracy of the heading calculation, a pseudo-observation is extracted to constrain the heading divergence error. This observation is obtained by calculating the heading deviation from the straightforward motion mode during the quasi-static magnetic field (QSMF) periods [5].

Incorporating this observation helps to reduce any interference caused by indoor magnetic fields and improves the overall precision of the heading estimation.

$$\tilde{\theta}_k - \hat{\theta}_0 = \delta\theta_k + n_\theta \quad (11)$$

In the above equation, $\hat{\theta}_0$ and $\tilde{\theta}_k$ denote the heading observation extracted from the detected QSMF period data for the first epoch and other epochs, respectively. n_θ represents the measurement noise.

The observation model for the calculated walking speed under the n-frame in the deep learning model is modeled as follows:

$$\delta z_\omega^n = \omega_{INS}^n - \omega_{DL}^n \quad (12)$$

In the following equation, ω_{INS}^n represents the INS mechanization-provided walking velocity, while ω_{DL}^n denotes the results of deep learning-based velocity prediction. The observation model for position increment is expressed as follows:

$$\delta z_r^n = r_{INS}^n - r_{DL}^n \quad (13)$$

In the above equation, r_{INS}^n denotes the position cumulated by INS mechanization, while r_{DL}^n denotes the results of the location update based on the deep learning model.

3. Floor Detection, Network Matching, and Intelligent Fusion

To enhance the efficiency and precision of indoor network information extraction and intelligent integration, a deep-learning approach was developed. This framework aims to achieve autonomous floor recognition, indoor network matching, and calibration, as well as the EE-UKF-based integration of indoor network and MEMS sensors.

3.1. Bi-LSTM-Enhanced Floor Recognition

To utilize useful observations obtained from daily-life trajectory data and local indoor environments, a time-continuous approach for detecting floors is proposed in this work. This method employs a Bi-LSTM network that considers a period of trajectory data to enhance the precision of floor recognition. The input vector for the Bi-LSTM network is constructed by combining features from Wi-Fi, barometer, and magnetic sources:

- (1) To capture the wireless features of the selected floor, the modeled RSSI values obtained from some representative Wi-Fi access points (APs) are utilized as a part of the input vector in the training procedure of the Bi-LSTM model. These collected RSSI values are deemed to be the most representative and are critical for ensuring accurate predictions:

$$\left\{ \phi_{RSSI}^1 \phi_{RSSI}^2 \cdots \phi_{RSSI}^k \right\} \geq Th_{RSSI} \quad (14)$$

In the following description, ϕ_{RSSI}^k represents the collected RSSI vector of a selected Wi-Fi access point, while Th_{RSSI} denotes the set standard for RSSI filtering.

- (2) The mean RSSI value of the selected representative RSSI values: In order to capture the overall description index of the RSSI collection, the mean RSSI value is also computed and included as an input value for the developed Bi-LSTM model. This additional input helps to provide a more comprehensive understanding of the RSSI vector, enabling the model to make more accurate predictions based on both individual signal strengths and the overall average signal strength:

$$\phi_{RSSI}^{Ave} = \sum_{i=1}^k \phi_{RSSI}^i \quad (15)$$

In the above description, ϕ_{RSSI}^{Ave} denotes the calculated mean RSSI value.

- (3) Deviation between the collected representative RSSI values: The calculated deviation in the scanned RSSI vector can significantly capture the dynamic changes in the surrounding buildings and is therefore considered a crucial feature. By utilizing this feature, the proposed model can better adapt to variations in the environment and provide more accurate estimations as a result. Incorporating real-time differences between the scanned RSSI vectors enhances the model's ability to detect subtle changes in the signal strength over time, allowing it to generate more reliable predictions:

$$\phi_{RSSI}^{Diff} = \left| \phi_{RSSI}^k - \phi_{RSSI}^{k-1} \right| \quad (16)$$

where ϕ_{RSSI}^{Diff} represents the RSSI deviation vector.

- (4) The norm vector of the extracted local magnetic data is calculated as follows:

$$M_{Norm} = \sqrt{m_x^2 + m_y^2 + m_z^2} \quad (17)$$

where m_x , m_y , and m_z represent the magnetometer outputted data.

- (5) The barometric pressure increment values based on the initial data are calculated as follows:

$$\Delta\varphi_{Baro}^i = \varphi_{Baro}^i - \varphi_{Baro}^0 \quad (18)$$

where $\Delta\varphi_{Baro}^i$ indicates the pressure increment compared to the initial data, and φ_{Baro}^0 and φ_{Baro}^i are the real-world acquired pressure values.

- (6) The deviation between the adjacent collected pressure data is calculated as follows:

$$\varphi_{Baro}^{Diff} = \left| \varphi_{Baro}^k - \varphi_{Baro}^{k-1} \right| \quad (19)$$

where φ_{Baro}^k and φ_{Baro}^{k-1} represent the pressure outputs obtained at two consecutive timestamps, and their variation can indicate changes in the elevation during a pedestrian's movement. This information is crucial for indoor positioning systems that estimate the user's vertical displacement and determine their current floor level, with applications in wayfinding, asset tracking, and environmental monitoring.

3.2. Indoor Network Matching and Trajectory Calibration

Pedestrian network information extracted from an indoor map can provide effective location references in the procedure of pedestrian walking. The principle of indoor network matching is to find the optimal reference point in the network to provide accurate location observation to the data and model dual-driven trajectory estimator, which can effectively reduce the cumulative error of the integrated sensor-based positioning approach. Compared to the traditional map-matching approach, network matching can significantly reduce the calculation complexity and provide effective location references. This research focuses on the representation of indoor pedestrian networks from multiple floors in the form of matrices. The network is extracted by marking intersection points as matrix elements, with each element containing information on the heading and the length between two intersections. This approach allows for an accurate depiction of the entire indoor network, enabling efficient navigation and localization. Moreover, this method enables the creation of a corresponding network matrix, which provides a clear visual representation of the indoor network's topology and facilitates further analysis. Figure 3 presents a detailed illustration of the extracted indoor network and its corresponding matrix, highlighting the usefulness of this representation method.

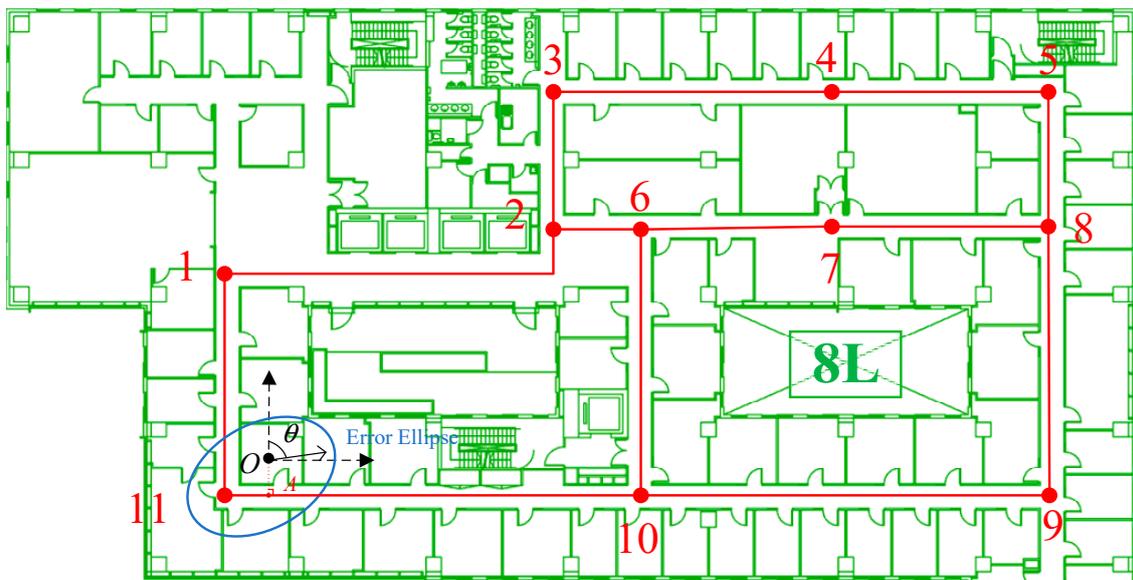


Figure 3. Presentation of pedestrian network and error ellipse.

Figure 3 demonstrates the division of the indoor pedestrian network into a combination of straight lines and turning points. Each straight line is characterized by two features: heading and gait length, which serve as the feature-matching parameters for comparing with the results provided by the deep-learning-based trajectory estimator. By utilizing the described approach, we can accurately estimate the user’s position in the indoor environment and determine their trajectory. In the network matrix generated from this representation, adjacent turning points are marked as 1, while non-adjacent turning points are identified as 0. This allows for an efficient and intuitive representation of the indoor network’s topology, which can facilitate path-planning and navigation tasks. Overall, this method provides a comprehensive and reliable solution for indoor localization and positioning, with potential applications in various fields such as logistics, security, and healthcare.

$$\psi = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \tag{20}$$

The above equation describes the relationship between each pair of turning points, which is divided as connected-1 and non-connected-0. The network matrix generated from the indoor pedestrian network extraction process has dimensions equivalent to the amount of turning points in the network. For each pair of adjacent turning points, the related characteristics are recorded, including the heading and gait length during a normal walking period. This information can be utilized for various applications, such as path planning, navigation, and localization. By representing the indoor environment as a matrix, we can efficiently capture and process data, facilitating real-time decision making and improving the user’s experience. The grid search algorithm is proposed for network matching and compares the extracted reference network information in Equation (20) and the real-time collected and divided trajectory. Overall, this approach offers a comprehensive

and reliable solution for indoor positioning and navigation that can be applied in diverse settings from healthcare to logistics:

$$\psi(i, j) = [1, \theta_{ij}, \zeta_{ij}] \quad (21)$$

In which θ_{ij} and ζ_{ij} denote the heading and distance between each pair of adjacent intersection points, calculated based on the user's normal walking trajectory between the two turning points. This information is crucial for accurately representing the indoor pedestrian network and facilitating indoor positioning and navigation. By utilizing this approach, we can efficiently capture and process data, enabling real-time decision making and improving the user's experience. Furthermore, this representation provides a comprehensive solution for various applications, such as asset tracking, security, and emergency response. Overall, this method offers a reliable and versatile approach for indoor localization and positioning.

In this study, a grid search algorithm was developed to perform trajectory matching and calibration of the estimated trajectory and the extracted network information from a constructed network matrix. By utilizing this approach, the model can effectively identify optimal matches between the two datasets, allowing for more accurate localization and tracking of pedestrians within indoor environments. By leveraging the insights gained from the constructed network matrix, the proposed approach can help to improve the overall accuracy of trajectory matching and calibration, making it a valuable tool for indoor navigation and location-based services. This method utilizes the information contained in the network matrix to accurately estimate the user's position and trajectory in real-time. By implementing this technique, we can enhance the precision and reliability of IPSs, improving the user's experience and facilitating various applications, such as asset tracking and emergency response. Overall, this approach offers a comprehensive and practical solution for indoor localization and navigation tasks:

- (1) To detect turning points during pedestrian movement, a hybrid deep-learning framework is utilized to detect changeable handheld modes and determine the walking direction. This approach is particularly useful in complex environments where pedestrians may adopt various postures or holding positions. After identifying the forward direction, the turning points are calculated by peak recognition, similar to a step detection method proposed previously [1]. By incorporating these techniques, we can accurately estimate the user's position and trajectory in indoor environments, facilitating effective navigation and positioning tasks. Moreover, this approach can be applied in diverse settings, such as healthcare, logistics, and security, enhancing the overall performance of indoor positioning systems;
- (2) To reduce the incidence of false matching, we exclusively considered the results provided by the trajectory estimator that contained more than three turning points, which can be further applied for network matching purposes. This algorithm was proven to significantly increase the precision of trajectory matching;
- (3) To effectively match trajectories with the existing indoor network, the proposed approach employs both the correlation coefficient value and the dynamic time warping (DTW) index. These two techniques are used in tandem to provide a more comprehensive understanding of the similarities and differences between different trajectories, enabling the model to identify optimal matches with greater accuracy. By incorporating both the correlation coefficient index and the DTW index, the proposed approach can enhance the overall performance of trajectory matching, paving the way for more precise location-based services within indoor environments. These indices enable us to identify similar trajectories by analyzing the detected turning points in each trajectory [14]. By employing this method, we can enhance the overall performance of indoor positioning systems, making them suitable for various applications such as logistics, security, and healthcare.

$$\begin{aligned} DTW(\beta_{\tau-1}, \beta_{\tau}) \\ = Dist(p_j, s_k) + \min[D(s_{j-1}, p_k), D(s_j, p_{k-1}), D(s_{j-1}, p_{k-1})] \end{aligned} \quad (22)$$

where $DTW(\delta_{refer}, \delta_k)$ represents the cumulated distance covered between two turning points vectors, while $Dist(q_n, c_m)$ represents the Euclidean distance calculated by each pair of points within these distributions.

$$\begin{aligned} \rho_{cor}(x, y) &= \rho_{cor}(x_{\tau-1}, x_{\tau}) + \rho_{cor}(y_{\tau-1}, y_{\tau}) \\ &= \frac{\sum_{i=1}^M (x_{\tau-1}^i - \bar{x}_{\tau-1})(x_{\tau}^i - \bar{x}_{\tau})}{\sqrt{\sum_{i=1}^M (x_{\tau-1}^i - \bar{x}_{\tau-1})^2} \sqrt{\sum_{i=1}^{2m+1} (x_{\tau}^i - \bar{x}_{\tau})^2}} \\ &\quad + \frac{\sum_{i=1}^M (y_{\tau-1}^i - \bar{y}_{\tau-1})(y_{\tau}^i - \bar{y}_{\tau})}{\sqrt{\sum_{i=1}^M (y_{\tau-1}^i - \bar{y}_{\tau-1})^2} \sqrt{\sum_{i=1}^{2m+1} (y_{\tau}^i - \bar{y}_{\tau})^2}} \end{aligned} \quad (23)$$

where $\rho_{cor}(x_{\tau-1}, x_{\tau})$ and $\rho_{cor}(y_{\tau-1}, y_{\tau})$ indicate the results of the correlation coefficient on the x - and y -axes, respectively.

- (4) Following the map-matching phase, the matched turning points on the extracted pedestrian network are utilized as absolute references for the trajectory calibration phase of straight lines. This approach enables us to accurately estimate the user's position and trajectory, improving the overall performance of indoor positioning systems. By incorporating this method, we can facilitate various applications, such as asset tracking, navigation, and emergency response. Overall, this technique achieves a robust and practical approach for indoor localization tasks:

$$\hat{x}_{k-1|k} = \hat{x}_{k-1} + P_{k-1} \phi_k^T (P_{k-1}^-)^{-1} (\hat{x}_k - \hat{x}_k^-) \quad (24)$$

$$P_{k-1|k} = P_{k-1} - (P_{k-1} \phi_k^T (P_{k-1}^-)^{-1}) (P_k - P_k^-) \cdot (P_{k-1} \phi_k^T (P_{k-1}^-)^{-1})^T \quad (25)$$

where x_k and z_k indicate the state and measurement vectors presented in Equation (8).

3.3. Error Ellipse-Enhanced UKF for Intelligent Fusion

In this section, the error ellipse-enhanced unscented Kalman filter (EE-UKF) is used to integrate all of the results provided by the integrated sensors, pedestrian indoor network matching, and floor detection for achieving meter-level indoor localization precision. The error ellipse is generated based on the previous integration result of the indoor network observation and data and model dual-driven trajectory estimator.

First, the INS mechanization applied in the data and model dual-driven trajectory estimator is applied as the state equation of the final EE-UKF, described in Equation (6). Then, the deep-learning-based walking speed prediction results, multi-level human motion constraints, and reference location and floor information provided by the network matching and floor detection are adopted as the observations in order to decrease the cumulative error of the INS mechanization and maintain the positioning accuracy under complex human motion and handheld modes and in disturbed indoor environments.

Furthermore, the enhancement of the localization performance involves implementing constraints on the indoor grid and mapping. Initially, the algorithm identifies the closest pair of adjacent turning points of observation using the current location data acquired through the integration of the sensory inputs and indoor network information. The segment of the indoor network is then represented in the model as follows:

$$\lambda_1 x + \lambda_2 y + \lambda_3 = 0 \quad (26)$$

Hence, the current coordinates (x_1, y_1) can be utilized to compute the closest point of observation in the indoor network. Consequently, the resulting representation of the indoor network is defined as follows:

$$\begin{cases} \delta z_r^n = r_{\text{network}}^n - r_{\text{MEMS}}^n \\ \delta z_\omega^n = \omega_{\text{network}}^n - \omega_{\text{MEMS}}^n \end{cases} \quad (27)$$

where r_{network}^n and $\omega_{\text{network}}^n$ represent the indoor network-provided location and walking velocity values, r_{MEMS}^n and ω_{MEMS}^n denote the trajectory estimator-provided location and velocity prediction results.

Lastly, with respect to the characterization of confidence ellipses in engineering [41], the focal point of the error ellipse is designated as the origin of the major semi-axis of the ellipse, which can be expressed as follows:

$$\zeta_1 = s_e \cdot \sqrt{0.5(\sigma_N^2 + \sigma_E^2) + \sqrt{0.25(\sigma_E^2 - \sigma_N^2)^2 + \sigma_{NE}^2}} \quad (28)$$

The minor semi-axis of the error ellipse is calculated as:

$$\zeta_2 = s_e \cdot \sqrt{0.5(\sigma_N^2 + \sigma_E^2) - \sqrt{0.25(\sigma_E^2 - \sigma_N^2)^2 + \sigma_{NE}^2}} \quad (29)$$

The azimuth of the major semi-axis of the error ellipse is described as:

$$\kappa = 0.5 \tan_4^{-1}(2\sigma_{NE} / (\sigma_E^2 - \sigma_N^2)) \quad (30)$$

Following the error ellipse-assisted network constraint check, eligible observations will be automatically incorporated into the UKF (unscented Kalman Filter) update procedure. This step helps to improve the overall accuracy of the filter by allowing it to integrate new observations into its predictions with greater precision. By incorporating eligible observations after the error ellipse-assisted network constraint check, the proposed approach can help to minimize errors and improve the quality of location estimates within indoor environments.

4. Experimental Results of ML-ISNM

This section outlines a series of comprehensive experiments devised to assess the efficacy of ML-ISNM. To this end, one publicly available dataset and a multi-floor indoor environment representative of real-world scenarios were chosen as experimental sites. Google Pixel 3 and Google Pixel 4 were applied as the experimental terminals, and Samsung Galaxy A7 was adopted for data collection in a public dataset. The proposed techniques, including trajectory estimator Bi-LSTM-based floor recognition, network matching, calibration, and EE-UKF, were tested and compared with state-of-art approaches or frameworks.

In configuring the model setting, the Adam optimizer was utilized due to its efficiency in handling large amounts of training data. Specifically, the learning rate was set at 0.002 for optimal performance. Additionally, the deep-learning-based speed estimator module employed an input vector dimension of 11, consistent with that of the integrated sensor data. The output hidden state from the Bi-LSTM layer was set at a dimension of 30, while the dimension of the input vector remained at 11. These settings were carefully selected to ensure optimal performance in the subsequent evaluations.

For the model training of the proposed data and model dual-driven trajectory estimator, a daily-life dataset containing a time period of more than 56 h of trajectory data provided by a number of 30 users under 4 different handheld modes (reading, phoning, swaying, and pocket) was collected with the reference of a Lidar-based positioning system (LPS) as the benchmark trajectory. For the accuracy evaluation, the trajectories estimated by the data and model dual-driven estimator were compared with the LPS for positioning

error calculation, and then compared with the state-of-art algorithms based on the same test route and handheld modes. In our experiments, a public dataset was first applied to evaluate the long-term performance of the proposed trajectory estimator, and the existing state-of-the-art algorithms were also applied for the robustness and significance comparison of ML-ISNM. In addition, a real-world scene was further adopted for evaluation and comparison purposes.

4.1. Performance Evaluation of Trajectory Estimator

To assess the efficacy of the proposed data and model dual-driven trajectory estimator, a public dataset from the IPIN-2018 competition [42] was utilized to estimate the accuracy with changeable handheld modes. For this purpose, different indoor environments containing multiple floors were selected, as illustrated in Figure 4. The evaluation process involved starting at point A and conducting tests over an extended duration of approximately 20 min to evaluate the long-term performance of the proposed estimator. The chosen experimental setup allowed for a thorough assessment of the algorithm's ability to accurately estimate the trajectories across varying conditions. The use of a publicly available dataset ensured that the results obtained were reproducible and could be compared with those of other studies conducted using the same dataset. Furthermore, the selection of a multi-floor environment enabled the evaluation of the proposed technique's effectiveness in accurately estimating trajectories in complex indoor settings.

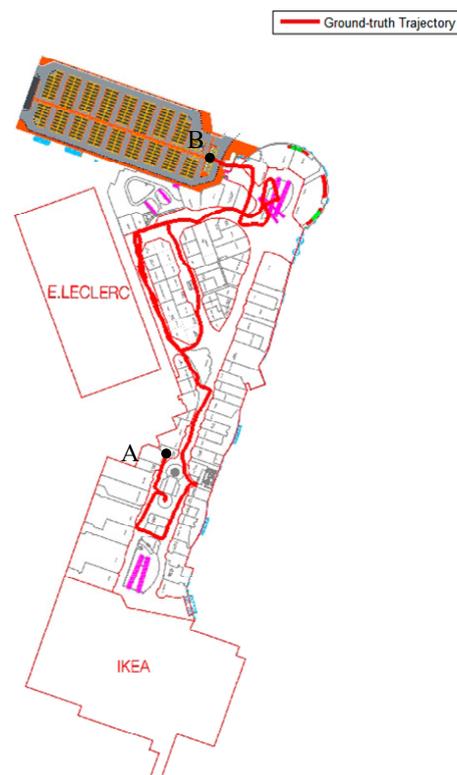


Figure 4. Ground-truth walking route of public dataset.

In order to assess the trajectory estimator's long-term performance, a comparison was conducted with an enhanced INS-PDR framework [5], utilizing identical smartphone data and a shared walking route. High-precision control points were utilized to evaluate the positioning errors. The respective estimated trajectories produced by the two algorithms were then compared, with the results presented in Figure 5.

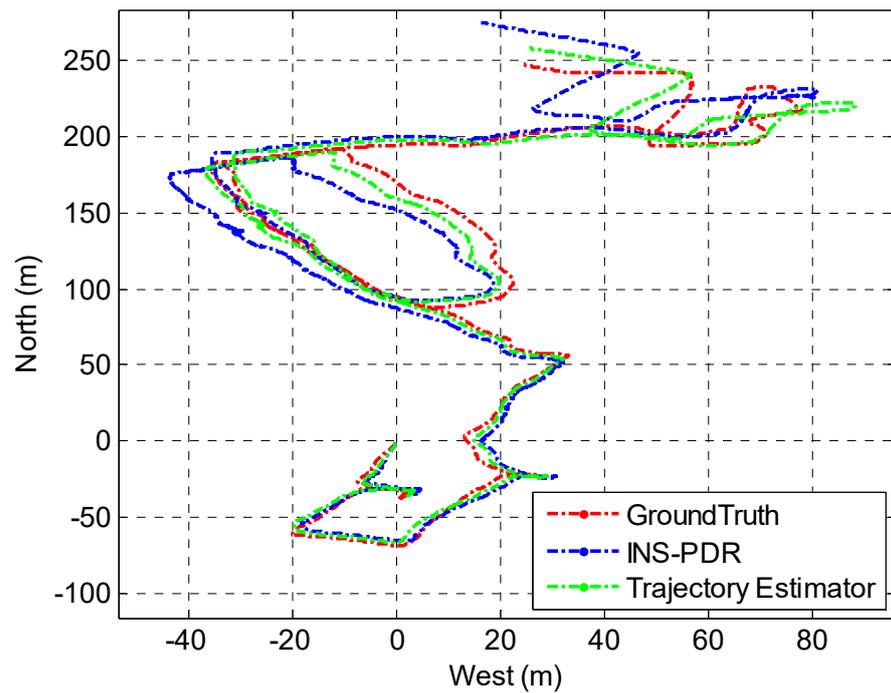


Figure 5. Estimated trajectory of inertial odometry.

Based on the data presented in Figure 5, it is apparent that the suggested inertial odometry achieved noticeably superior and accurate long-term location determination in comparison to the lone PDR mechanism. In order to obtain a more comprehensive evaluation of the positioning precision of both algorithms, a group of ten individuals retraced the same walking path multiple times, and their calculated positioning errors are depicted in Figure 6.

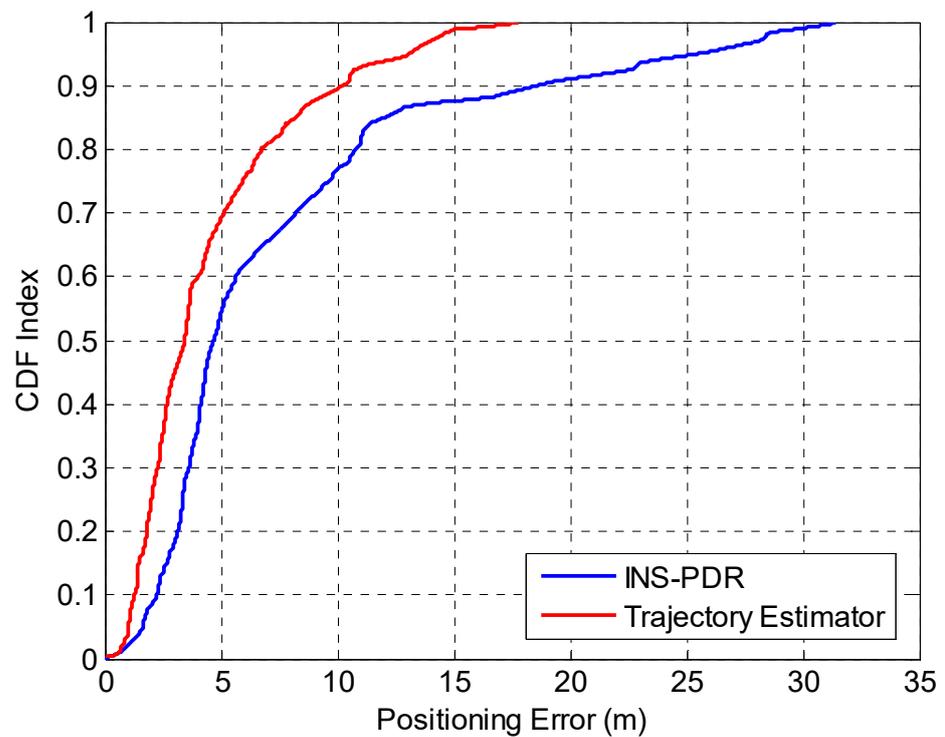


Figure 6. Error comparison of trajectory estimator and INS-PDR.

The results described in Figure 6 illustrate that the developed trajectory estimator maintained a long-term error of no more than 5.95 m in 75% of cases, which is superior to the INS-PDR approach, which had an error of up to 9.73 m in 75% of cases. This improvement can be attributed to the hybrid observations and constraints utilized by the proposed estimator.

To further assess the performance of the data and model dual-driven trajectory estimator under varying handheld modes, we compared it with the state-of-the-art 3D navigation framework (3D-NF) [14]. The mean positioning inaccuracies were computed for four distinct portable modes utilizing the identical pedestrian pathway specified in the reference material [14], and the resultant mean positioning inaccuracies were contrasted.

According to Figure 7, the proposed path estimator attained notably superior positioning precision across all four distinct portable modes, with mean positioning inaccuracies of 1.02 m (reading), 1.15 m (calling), 1.78 m (swaying), and 1.21 m (pocket) along the prolonged test pathway. This stands in contrast to the 3D-NF algorithm, which registered mean positioning inaccuracies of 1.25 m (reading mode), 1.32 m (calling mode), 2.08 m (swaying mode), and 1.35 m (pocket mode).

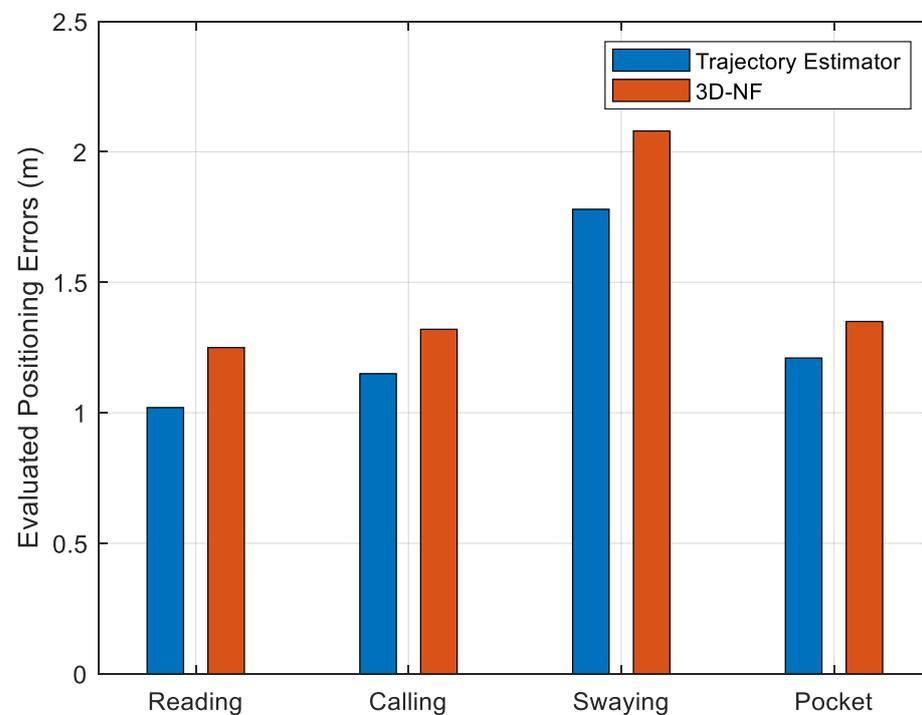


Figure 7. Accuracy indices of four handheld modes.

4.2. Performance Evaluation of Floor Recognition

This manuscript addresses the issue of lacking an initial floor observation in raw estimated trajectories. To overcome this limitation, a Bi-LSTM network was implemented to improve floor recognition precision on multiple floors in a contained building by incorporating data from local wireless signals and environment-related characteristics before the map-matching phase. To ensure comprehensive coverage of the indoor scenes, a 2.5 h trajectory dataset from five different floors was collected for training purposes, while a 0.5 h trajectory dataset was utilized for accuracy evaluation. The developed Bi-LSTM-based floor recognition model was compared with the classical 1D-CNN model [16] and the classical MLP model [1], with the accuracy comparison results presented in Figure 8.

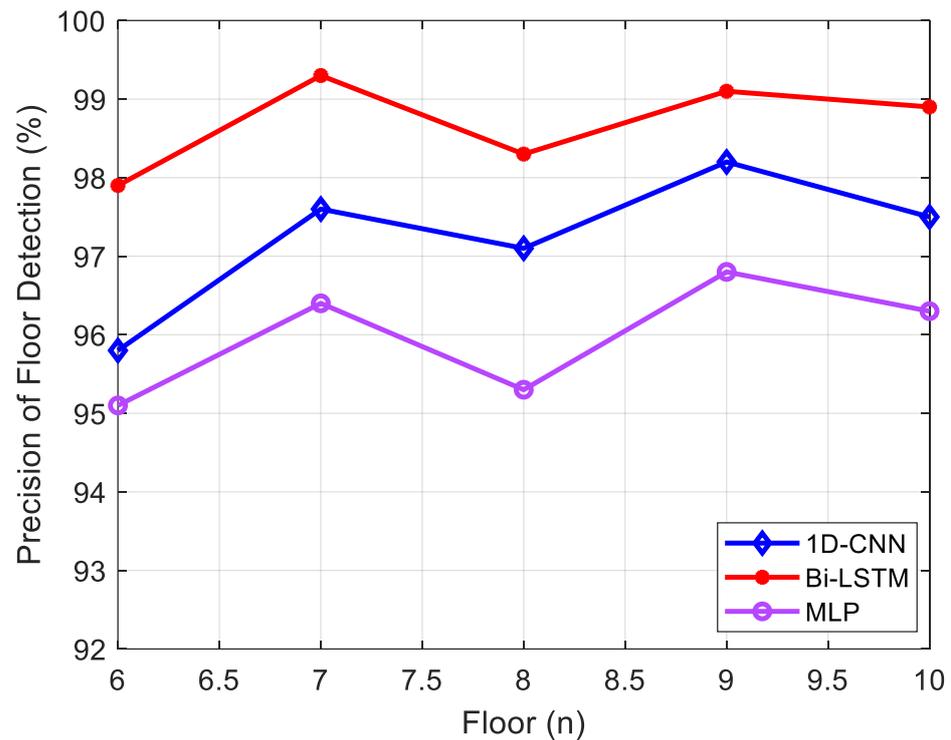


Figure 8. Accuracy comparison of different models.

As shown in Figure 8, the Bi-LSTM-based floor recognition model proposed in this study demonstrated superior accuracy compared to the 1D-CNN and MLP algorithms. The test dataset achieved an average accuracy of over 98.7% using the proposed approach, whereas the 1D-CNN-based approach and MLP-based approach resulted in average accuracies of 97.2% and 95.98%, respectively.

4.3. Precision Evaluation of Error Ellipse-Assisted UKF

In order to assess the efficacy of map-matching and the fusion of the trajectory estimator and pedestrian indoor network, a multi-floor 3D indoor environment comprising four distinct floors with corresponding pedestrian indoor networks (as depicted in Figure 9) was selected for evaluation. In this scenario, the indoor networks collected from different floors were modeled as a combination of nodes and segments and applied for trajectory matching, as illustrated in Figure 10:

As illustrated in Figure 10, the extracted indoor network comprehensively covers all pedestrian walking route and turning point information within multi-floor contained indoor building environments. Therefore, this study proposes an indoor network matching algorithm to enable absolute turning point references for optimizing and calibrating raw trajectory estimator results, while the floor recognition algorithm is employed to achieve dynamic floor indices of the trajectory estimator.

To evaluate the efficacy of the final ML-ISNM structure, a 3D indoor environment containing four adjacent floors was selected. The evaluator proceeded with uninterrupted walking from the 6th floor to the 9th floor, following the explicit pedestrian path delineated in Figure 9. In this setting, an EE-UKF-driven amalgamation algorithm was suggested to astutely merge diverse location origins, involving MEMS sensors integrated into smartphones and indoor pedestrian network data. To refine the path estimator, the error ellipse was utilized to delimit the exploration range of the indoor network and to correlate relevant indoor network data. A comparison between the single trajectory estimator (TE), EE-UKF-based TE, and the indoor network integration model was conducted by examining the estimated trajectories.

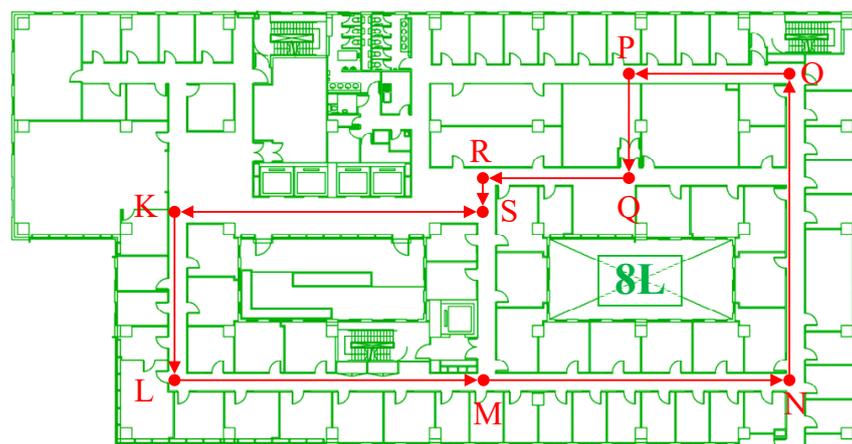
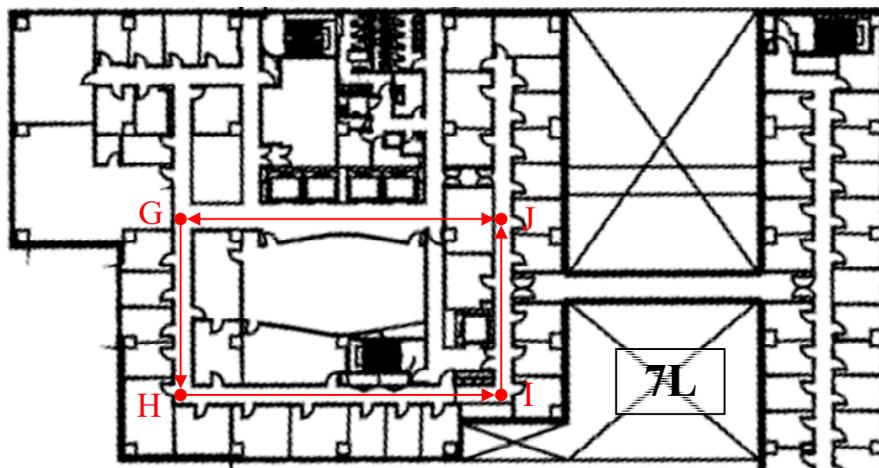
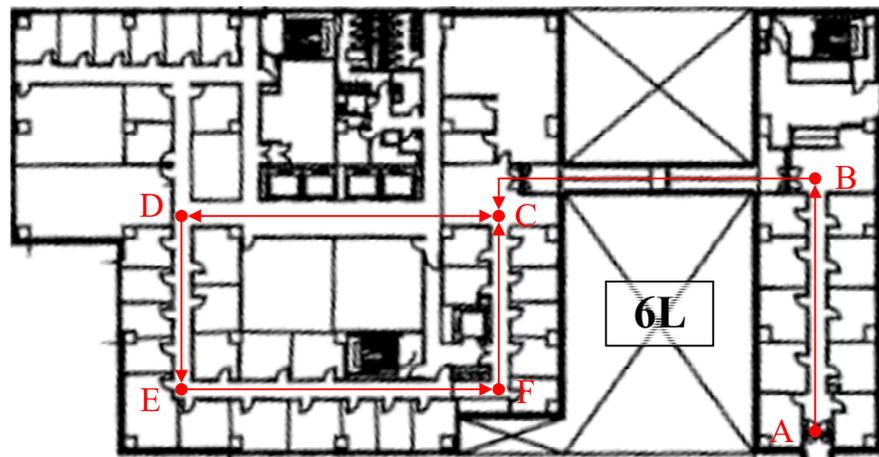


Figure 9. Cont.



Figure 9. Multi-floor contained real-world indoor environment. (a) 6th Floor. (b) 7th Floor. (c) 8th Floor. (d) 9th Floor.

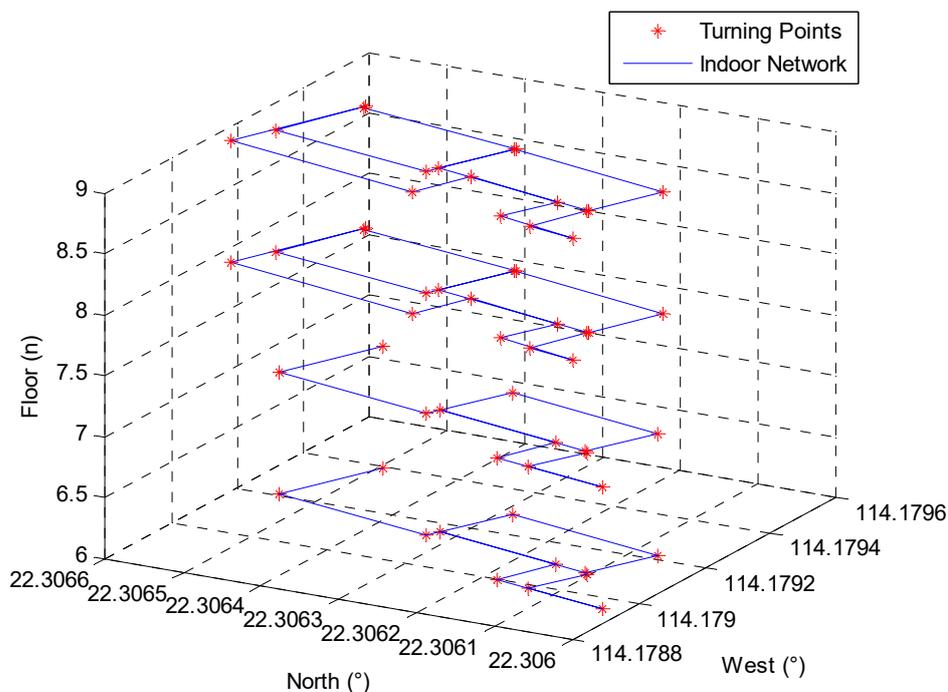
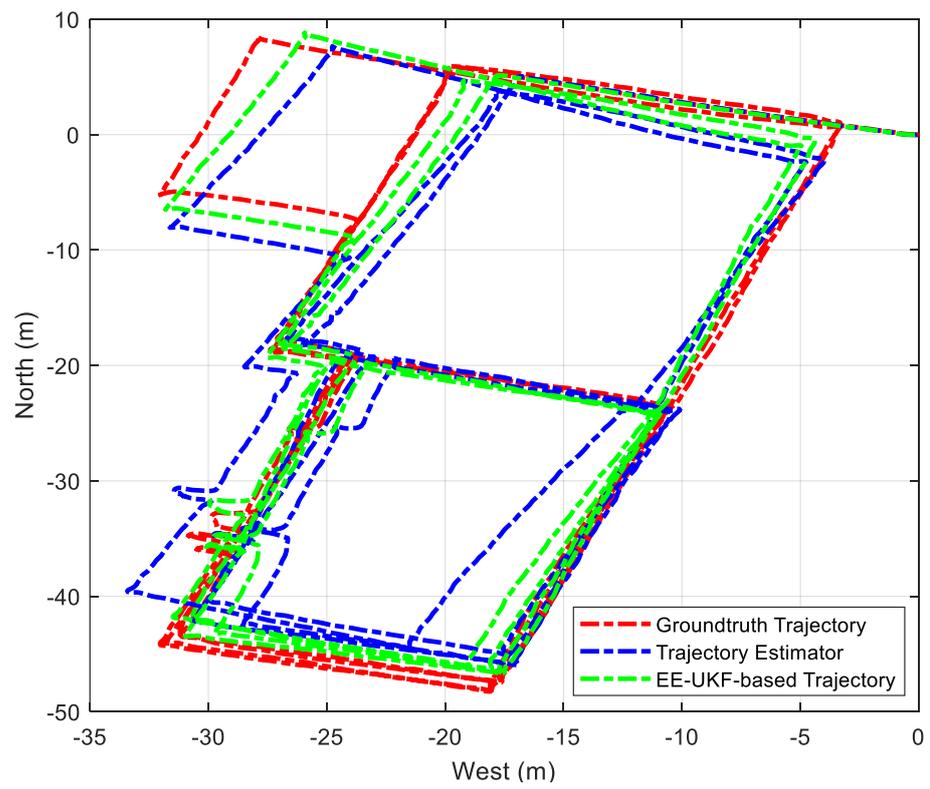


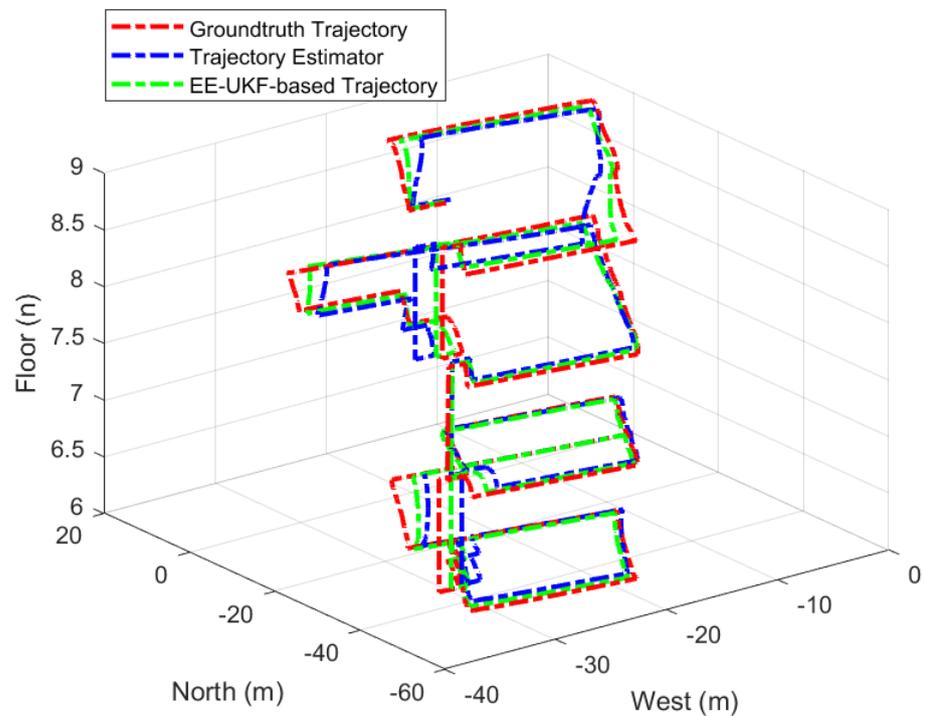
Figure 10. Extracted pedestrian indoor network.

As shown in Figure 11, despite the multi-level based constraints and observations, the trajectory estimator remained susceptible to cumulative errors. However, the integration model of the pedestrian indoor network and trajectory estimator significantly enhanced single TE performance, while the integration of pedestrian network data further reduced the TE localization error and brought the estimated trajectory closer to the ground-truth reference. A localization error comparison between the two different combinations of location sources is presented below.

As depicted in Figure 12, the amalgamation of indoor networks achieved indoor positioning accuracy at the meter level, with a positioning inaccuracy of less than 1.35 m in 75% of cases, outperforming the single trajectory estimator, which acquired a precision of 2.73 m in 75% of cases.



(a)



(b)

Figure 11. (a). Two-dimensional trajectory comparison between TE and EE-UKF. (b). Three-dimensional trajectory comparison between TE and EE-UKF.

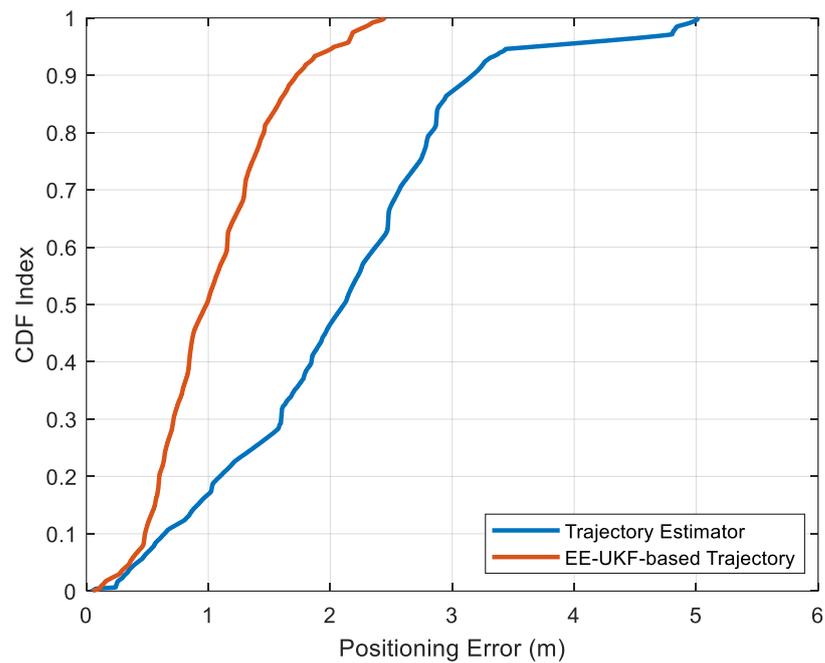


Figure 12. Positioning error comparison between TE and EE-UKF.

Lastly, to ensure impartiality, the proposed multi-source fusion algorithm based on EE-UKF was compared with two cutting-edge systems: the HTrack system [23] and the map-assisted particle filter (MA-PF) [43], utilizing the same pedestrian path. A comparison of the 3D positioning inaccuracies among these three distinct algorithms is presented in Table 2.

Table 2. Positioning error comparison.

Models	Maximum Error	75th Percentile	Average Error
EE-UKF	2.39 m	1.35 m	1.13 m
MA-PF	2.87 m	1.62 m	1.31 m
HTrack	3.42 m	2.07 m	1.68 m

In accordance with Table 2, the proposed EE-UKF technique attained superior integration performance of multi-sourced data in contrast to the MA-PF and HTrack methodologies, thanks to the error ellipse-oriented UKF amalgamation strategy. The realized average positioning inaccuracy was lower than 1.13 m, which was remarkably lower compared to the MA-PF and HTrack schemes (which registered average inaccuracies of 1.31 m and 1.68 m, respectively).

5. Conclusions

To improve the positioning ability in complex and changeable urban buildings, this paper presents an autonomous multi-floor localization framework using smartphone-integrated sensors and pedestrian network matching (ML-ISNM). A robust data and model dual-driven pedestrian trajectory estimator is proposed for accurate integrated sensor-based positioning under different handheld modes and disturbed environments. The Bi-LSTM network was further developed for floor recognition using extracted environmental features and pedestrian motion features, and further combined with the indoor network matching algorithm for acquiring accurate location and floor observations. In the multi-source fusion phase, an error ellipse-enhanced unscented Kalman filter is presented to realize the multi-source fusion of the trajectory estimator, human motion constraints, and indoor network information. The empirical outcomes validate that the proposed ML-ISNM achieves self-directed and accurate 3D indoor localization performance in intricate and extensive indoor

settings, with an average estimated positioning inaccuracy below 1.13 m, and without any reliance on wireless amenities.

The advantages of the proposed ML-ISNM contain the following aspects. Firstly, no additional local facilities are required for positioning purposes. The proposed ML-ISNM framework uses only smartphone-integrated sensors and extracted indoor network information for accurate multi-floor localization. Secondly, the proposed ML-ISNM has proven accurate and stable positioning performance under complex human motion and handheld modes and in disturbed indoor environments based on the EE-UKF fusion model. Thirdly, the proposed ML-ISNM can acquire precise initial location and floor information according to network matching and floor detection results, which is more autonomous than the existing approaches and applicable to providing universal location-based services.

The proposed ML-ISNM also has some limitations. The human trajectory is usually disordered and complex, which decreases the performance of pedestrian network matching. Thus, a more robust network matching algorithm is expected to enhance the matching accuracy for disordered trajectories. In addition, more complex human motion and handheld modes need to be considered to improve the robustness and precision of the data and model dual-driven method, which is more suitable for real-world applications.

The future work of this research includes the enhancement of the current ML-ISNM framework, improving the adaptability to more complex human motion and handheld modes, and developing more robust network matching and floor detection algorithms, which can be applied for the fast calibration of initial location/floor information and for providing more accurate observation information to the trajectory estimator. In addition, other advanced deep-learning frameworks [44] can also be attempted for human trajectory prediction and estimation using multi-sensor data, for instance, a new automatic pruning method—sparse connectivity learning (SCL)—proposed by Tang et al. [45], a channel-pruning method via class-aware trace ratio optimization (CATRO) proposed by Hu et al. [46], and the weight-quantized SqueezeNet model developed by Huang [47].

Author Contributions: This paper is a collaborative work by all the authors. C.S. and W.T. proposed the idea and implemented the system. Y.Y. and Y.Z. performed the experiments, analyzed the data, and wrote the manuscript. R.C., Q.L. and L.C. aided in proposing the idea, gave suggestions, and revised the rough draft. All authors have read and agreed to the published version of the manuscript.

Funding: The work was funded in part by the National Natural Science Foundation of China under Grant 41901392, the Natural Science Foundation of Hubei Province under Grant 2019CFB098, the Fundamental Research Funds for the Central Universities under Grant 2019kfyXJJS142 (HUST), the Hong Kong Polytechnic University (1-ZVN6, 4-BCF7), the State Bureau of Surveying and Mapping, P.R. China (1-ZVE8), and the Hong Kong Research Grants Council (T22-505/19-N).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no conflict of interest to disclose.

References

1. Yu, Y.; Chen, R.; Chen, L.; Li, W.; Wu, Y.; Zhou, H.T. H-WPS: Hybrid Wireless Positioning System Using an Enhanced Wi-Fi FTM/RSSI/MEMS Sensors Integration Approach. *IEEE Internet Things J.* **2022**, *9*, 11827–11842. [[CrossRef](#)]
2. Aranda, F.J.; Parralejo, F.; Álvarez, F.J.; Paredes, J.A. Performance analysis of fingerprinting indoor positioning methods with, B.L.E. *Expert Syst. Appl.* **2022**, *202*, 117095. [[CrossRef](#)]
3. Ridolfi, M.; Kaya, A.; Berkvens, R.; Weyn, M.; Joseph, W.; Poorter, E.D. Self-calibration and collaborative localization for UWB positioning systems: A survey and future research directions. *ACM Comput. Surv. CSUR* **2021**, *54*, 1–27. [[CrossRef](#)]
4. Liu, T.; Niu, X.; Kuang, J.; Cao, S.; Zhang, L.; Chen, X. Doppler shift mitigation in acoustic positioning based on pedestrian dead reckoning for smartphone. *IEEE Trans. Instrum. Meas.* **2020**, *70*, 1–11. [[CrossRef](#)]
5. Kuang, J.; Niu, X.; Chen, X. Robust pedestrian dead reckoning based on MEMS-IMU for smartphones. *Sensors* **2018**, *18*, 1391. [[CrossRef](#)]
6. An, Z.; Lin, Q.; Zhao, X.; Yang, L.; Zheng, D.; Wu, G.; Chang, S. One tag, two codes: Identifying optical barcodes with NFC. In Proceedings of the 27th Annual International Conference on Mobile Computing and Networking, New Orleans, LA, USA, 25–29 October 2021; pp. 108–120.

7. Chen, L.; Zhou, X.; Chen, F.; Yang, L.L.; Chen, R. Carrier phase ranging for indoor positioning with 5G NR signals. *IEEE Internet Things J.* **2021**, *9*, 10908–10919. [[CrossRef](#)]
8. Pan, Q.; An, Z.; Yang, X.; Zhao, X.; Yang, L. RF-DNA: Large-scale physical-layer identifications of RFIDs via dual natural attributes. In Proceedings of the 28th Annual International Conference on Mobile Computing and Networking, Sydney, Australia, 17–21 October 2022; pp. 419–431.
9. Shi, J.; Wong, T.T.; He, Y.; Li, L.; Zhang, R.; Yung, C.S.; Wang, L.V. High-resolution, high-contrast mid-infrared imaging of fresh biological samples with ultraviolet-localized photoacoustic microscopy. *Nat. Photonics* **2019**, *13*, 609–615. [[CrossRef](#)]
10. Zhuang, Y.; Hua, L.; Qi, L.; Yang, J.; Cao, P.; Cao, Y.; Haas, H. A survey of positioning systems using visible LED lights. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 1963–1988. [[CrossRef](#)]
11. Couture, O.; Hingot, V.; Heiles, B.; Muleki-Seya, P.; Tanter, M. Ultrasound localization microscopy and super-resolution: A state of the art. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2018**, *65*, 1304–1320. [[CrossRef](#)]
12. Ouyang, G.; Abed-Meraim, K. A Survey of Magnetic-Field-Based Indoor Localization. *Electronics* **2022**, *11*, 864. [[CrossRef](#)]
13. Oliveira, A.S.S.; dos Reis, M.C.; da Mota, F.A.X.; Martinez, M.E.M.; Alexandria, A.R. New trends on computer vision applied to mobile robot localization. *Internet Things Cyber-Phys. Syst.* **2022**, *2*, 63–69. [[CrossRef](#)]
14. Yu, Y.; Chen, R.; Shi, W.; Chen, L. Precise 3D Indoor Localization and Trajectory Optimization Based on Sparse Wi-Fi FTM Anchors and Built-In Sensors. *IEEE Trans. Veh. Technol.* **2022**, *71*, 4042–4056. [[CrossRef](#)]
15. Yu, Y.; Chen, R.; Chen, L.; Li, W.; Wu, Y.; Zhou, H. A robust seamless localization framework based on Wi-Fi FTM/GNSS and built-in sensors. *IEEE Commun. Lett.* **2021**, *25*, 2226–2230. [[CrossRef](#)]
16. Shi, W.; Yu, Y.; Liu, Z.; Chen, R.; Chen, L. A deep-learning approach for modelling pedestrian movement uncertainty in large-scale indoor areas. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *114*, 103065. [[CrossRef](#)]
17. Zou, H.; Jin, M.; Jiang, H.; Xie, L.; Spanos, C.J. WinIPS: WiFi-based non-intrusive indoor positioning system with online radio map construction and adaptation. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 8118–8130. [[CrossRef](#)]
18. Yan, H.; Shan, Q.; Furukawa, Y. RIDI: Robust IMU double integration. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 621–636.
19. Herath, S.; Yan, H.; Furukawa, Y. Ronin: Robust neural inertial navigation in the wild: Benchmark, evaluations, new methods. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 3146–3152.
20. Guo, G.; Chen, R.; Ye, F.; Chen, L.; Pan, Y.; Liu, M.; Cao, Z. A pose awareness solution for estimating pedestrian walking speed. *Remote Sens.* **2018**, *11*, 55–73. [[CrossRef](#)]
21. Zhang, P.; Chen, X.; Ma, X.; Wu, Y.; Jiang, H.; Fang, D.; Ma, Y. SmartMTra: Robust indoor trajectory tracing using smartphones. *IEEE Sens. J.* **2017**, *17*, 3613–3624. [[CrossRef](#)]
22. Poulouse, A.; Senouci, B.; Han, D.S. Performance analysis of sensor fusion techniques for heading estimation using smartphone sensors. *IEEE Sens. J.* **2019**, *19*, 12369–12380. [[CrossRef](#)]
23. Wu, Y.; Chen, P.; Gu, F.; Zheng, X.; Shang, J. HTrack: An efficient heading-aided map matching for indoor localization and tracking. *IEEE Sens. J.* **2019**, *19*, 3100–3110. [[CrossRef](#)]
24. Xia, H.; Zuo, J.; Liu, S.; Qiao, Y. Indoor localization on smartphones using built-in sensors and map constraints. *IEEE Trans. Instrum. Meas.* **2018**, *68*, 1189–1198. [[CrossRef](#)]
25. Li, Y.; He, Z.; Gao, Z.; Zhuang, Y.; Shi, C.; El-Sheimy, N. Toward robust crowdsourcing-based localization: A fingerprinting accuracy indicator enhanced wireless/magnetic/inertial integration approach. *IEEE Internet Things J.* **2018**, *6*, 3585–3600. [[CrossRef](#)]
26. Zhao, F.; Luo, H.; Zhao, X.; Pang, Z.; Park, H. HYFI: Hybrid floor identification based on wireless fingerprinting and barometric pressure. *IEEE Trans. Ind. Inform.* **2015**, *13*, 330–341. [[CrossRef](#)]
27. Shao, W.; Luo, H.; Zhao, F.; Tian, H.; Huang, J.; Crivello, A. Floor identification in large-scale environments with Wi-Fi autonomous block models. *IEEE Trans. Ind. Inform.* **2021**, *18*, 847–858. [[CrossRef](#)]
28. Shen, G.; Chen, Z.; Zhang, P.; Moscibroda, T.; Zhang, Y. Walkie-Markie: Indoor pathway mapping made easy. In Proceedings of the NSDI'13: 10th USENIX Symposium on Networked Systems Design and Implementation, Lombard, IL, USA, 2–5 April 2013; pp. 85–98.
29. Luo, C.; Hong, H.; Chan, M.C. PiLoc: A self-calibrating participatory indoor localization system. In Proceedings of the 13th International Symposium on Information Processing in Sensor Networks, Berlin, Germany, 15–17 April 2014; pp. 143–154.
30. Luo, C.; Hong, H.; Chan, M.C.; Li, J.; Zhang, X.; Ming, Z. MpiLoc: Self-calibrating multi-floor indoor localization exploiting participatory sensing. *IEEE Trans. Mobile Comput.* **2018**, *17*, 141–154. [[CrossRef](#)]
31. Li, T.; Han, D.; Chen, Y.; Zhang, R.; Zhang, Y.; Hedgpeth, T. IndoorWaze: A crowdsourcing-based context-aware indoor navigation system. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 5461–5472. [[CrossRef](#)]
32. Yu, N.; Zhan, X.; Zhao, S.; Wu, Y.; Feng, R. A Precise Dead Reckoning Algorithm Based on Bluetooth and Multiple Sensors. *IEEE Internet Things J.* **2018**, *5*, 336–351. [[CrossRef](#)]
33. Feng, D.; Wang, C.; He, C.; Zhuang, Y.; Xia, X.G. Kalman-filter-based integration of IMU and UWB for high-accuracy indoor positioning and navigation. *IEEE Internet Things J.* **2020**, *7*, 3133–3146. [[CrossRef](#)]
34. Yu, Y.; Chen, R.; Chen, L.; Guo, G.; Ye, F.; Liu, Z. A robust dead reckoning algorithm based on Wi-Fi FTM and multiple sensors. *Remote Sens.* **2019**, *11*, 504. [[CrossRef](#)]

35. Xie, H.; Gu, T.; Tao, X.; Ye, H.; Lu, J. A reliability-augmented particle filter for magnetic fingerprinting based indoor localization on smartphone. *IEEE Trans. Mob. Comput.* **2016**, *15*, 1877–1892. [[CrossRef](#)]
36. Huang, Q.; Zhang, Y.; Ge, Z.; Lu, C. Refining Wi-Fi based indoor localization with Li-Fi assisted model calibration in smart buildings. *arXiv* **2016**, arXiv:1602.07399.
37. Chen, P.; Kuang, Y.; Chen, X. A UWB/improved PDR integration algorithm applied to dynamic indoor positioning for pedestrians. *Sensors* **2017**, *17*, 2065. [[CrossRef](#)]
38. Chen, J.; Ou, G.; Peng, A.; Zheng, L.; Shi, J. An INS/WiFi indoor localization system based on the Weighted Least Squares. *Sensors* **2018**, *18*, 1458. [[CrossRef](#)]
39. Nosrati, L.; Fazel, M.S.; Ghavami, M. Improving Indoor Localization Using Mobile UWB Sensor and Deep Neural Networks. *IEEE Access* **2022**, *10*, 20420–20431. [[CrossRef](#)]
40. Liu, Z.; Shi, W.; Yu, Y.; Chen, P.; Chen, B.Y. A LSTM-based approach for modelling the movement uncertainty of indoor trajectories with mobile sensing data. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *108*, 102758. [[CrossRef](#)]
41. Wan, Q.; Yu, Y.; Chen, R.; Chen, L. Map-Assisted 3D Indoor Localization Using Crowd-Sensing-Based Trajectory Data and Error Ellipse-Enhanced Fusion. *Remote Sens.* **2022**, *14*, 4636. [[CrossRef](#)]
42. Renaudin, V.; Ortiz, M.; Perul, J.; Torres-Sospedra, J.; Jiménez, A.R.; Perez-Navarro, A.; Park, Y. Evaluating indoor positioning systems in a shopping mall: The lessons learned from the IPIN 2018 competition. *IEEE Access* **2019**, *7*, 148594–148628. [[CrossRef](#)]
43. Du, X.; Liao, X.; Gao, Z.; Fan, Y. An enhanced particle filter algorithm with map information for indoor positioning system. In Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6.
44. Gholami, A.; Kim, S.; Dong, Z.; Yao, Z.; Mahoney, M.W.; Keutzer, K. A survey of quantization methods for efficient neural network inference. *arXiv* **2021**, arXiv:2103.13630.
45. Tang, Z.; Luo, L.; Xie, B.; Zhu, Y.; Zhao, R.; Bi, L.; Lu, C. Automatic sparse connectivity learning for neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *in press*. [[CrossRef](#)]
46. Hu, W.; Che, Z.; Liu, N.; Li, M.; Tang, J.; Zhang, C.; Wang, J. Channel Pruning via Class-Aware Trace Ratio Optimization. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *in press*. [[CrossRef](#)]
47. Huang, Q. Weight-quantized squeezeNet for resource-constrained robot vacuums for indoor obstacle classification. *AI* **2022**, *3*, 180–193. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.