Received: 1 April 2024 | Revised: 13 September 2024 | Accepted: 26 October 2024 | Published online: 31 October 2024

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2024, Vol. 00(00) 1–17
DOI: 10.47852/bonviewJDSIS42022961

BON VIEW PUBLISHING

Enhancing Smartphone-based Pedestrian Positioning: Using Factor Graph Optimization with Indoor/Outdoor Detection for 3DMA GNSS/Visual-inertial State Estimation

Hiu-Yi Ho[1], Hoi-Fung Ng[1,*], Weisong Wen[1], Yanlei Gu[2] and Li-Ta Hsu[1]

[1]*Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, China*

[2]*Graduate School of Advanced Science and Engineering, Hiroshima University, Japan*

Abstract: This paper explores the pervasive challenges of pedestrian positioning using smartphones in densely populated urban environments where global navigation satellite system (GNSS) signals are inaccessible, for example, in indoor areas. Existing sensor-based positioning methods, such as inertial navigation systems, GNSS, and visual-inertial odometry (VIO), suffer from inherent restrictions that compromise the accuracy and reliability of the positioning performance. An approach based on machine learning is proposed to address these limitations, employing the support vector machine (SVM) algorithm to accurately distinguish indoor/outdoor (IO) based on the measurement of GNSS. The proposed approach in this study seamlessly incorporates 3D mapping aided (3DMA) GNSS measurements and localized estimations derived by VIO via factor graph optimization (FGO), complemented by an IO detection switch, to achieve accurate pose estimation and effectively eliminate global drift. The system's effectiveness and robustness are rigorously assessed through comprehensive extensive real-life experiments, with an average reduction of 4 meters, leading to noteworthy and statistically significant findings.

Keywords: FGO, pedestrian positioning, smartphone, sensor integration, IO, VINS, 3DMA GNSS

1. Introduction

Smart mobility faces challenges in accurate smartphone positioning within urban canyons. The development makes use of various sensors, advanced intelligence, and next-generation networks to improve performance. A diverse range of mobile positioning techniques has been studied to facilitate location-based services, with smartphones emerging as key instruments. Equipped with various sensors such as Wi-Fi, inertial sensors, magnetometers, and monocular cameras [1–3], smartphones offer reliable localization information and comprehensive positioning through different sensor combinations. However, urban areas pose significant obstacles to the effectiveness of mobile global navigation satellite system (GNSS) positioning. Studies by Rajak et al. [4] demonstrated the limitations of conventional GNSS localization methods in cityscapes, highlighting the adverse impact of low positioning accuracy on user experiences, particularly for smartphone users. To overcome these challenges, additional data enhance urban positioning. One such approach is the utilization of 3D building models as a software-based aid for low-cost positioning, known as 3D mapping aided (3DMA) GNSS. Noteworthy research [5] showcased the superior 3DMA GNSS urban positioning results. Doppler measurements are frequently integrated with the position solution to enhance the precision and reliability of the positioning. The fusion of velocity and 3DMA GNSS estimated from Doppler frequency is proposed [6]. Building upon these foundations, the present study adopts a loosely coupled approach to integrating Doppler velocity and 3DMA GNSS, striving to offer precise positioning on a global scale.

In scenarios where GNSS signals are compromised, such as indoor environments, GNSS positioning is significantly affected, leading to outages and degraded accuracy [7]. Indoor positioning suffers from weakened and scattered signals caused by various objects within the environment. Studies have revealed a decline in GNSS signal strength of approximately 10–12 decibels, rapidly reducing positioning accuracy [4]. Detection of indoor/outdoor (IO) environments becomes crucial for achieving ubiquitous positioning [8]. Several researchers have worked on IO detection using GNSS measurements to create a framework for indoor localization that relies on only smartphone sensors, requiring no additional infrastructure.

*Corresponding author: Hoi-Fung Ng, Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, China. Email: hf-ivan.ng@connect.polyu.hk

To enhance indoor positioning capabilities, visual-inertial odometry (VIO) has been extensively employed in GNSS-denied areas to bridge the gaps in the absence of GNSS signals [9]. VIO poses implementation challenges due to the speed and latency constraints; however, nonlinear optimization techniques address these challenges [10, 11]. In the case of scattered and weakened signals encountered, standalone GNSS is not considered for indoor positioning. Instead, the system uses a combination of inertial navigation system (INS) and VO, leveraging inertial measurement unit (IMU) and monocular camera sensors. This paper explores the complementarity between GNSS and visual-inertial navigation systems (VINS) for pedestrian positioning.

Promising results of the GNSS/INS fusion using factor graph optimization (FGO) are shown [12]. FGO is an advanced method for nonlinear optimization that uses probabilistic graphical models and integrates GNSS/INS data. By factorizing the problem and transforming it into a factor graph, the process simulates the connection between poses and estimates the values. This approach effectively accommodates alterations in the ever-changing environment [13, 14]. Additionally, FGO optimally utilizes feature constraints, leading to higher accuracy and efficiency in achieving robust estimations [15].

This paper is an extension of the proceedings paper [16]. The paper is structured as follows: Section 2 reviews relevant literature; Section 3 provides a system overview, covering support vector machine (SVM) classification, 3DMA GNSS/VINS, and FGO; Section 4 presents the experimental results of 3DMA GNSS/VINS-IO FGO, and finally, Section 5 concludes the paper.

2. Literature Review

2.1. Integration of GNSS and VINS for pedestrian positioning

Pedestrian localization in urban and indoor settings has gained attention, with GNSS and VINS as key technologies. Integrating these methods enhances accuracy and continuous positioning. Recent studies focus on improving GNSS by incorporating INS and combining inertial, visual sensors, and GNSS for precise positioning in GNSS-denied environments. [17, 18]. Evaluation in actual urban environments has convincingly proven its effectiveness for positioning accuracy compared to standalone INS or GNSS systems [19]. Despite its potential, the VINS and GNSS fusion encounters various obstacles and restrictions, such as addressing signal loss, streamlining computational complexity, and enhancing robustness across diverse environmental conditions.

2.2. Challenges and opportunities of GNSS/VINS integration

Accurate real-time positioning in smartphone pedestrian applications is challenging due to noisy measurements and limited processing power. While GNSS is widely used for its accessibility and affordability, it struggles with accuracy and availability in urban and indoor environments due to signal interference and blockage [20]. Additionally, multipath interference frequently manifests, especially with low-cost sensors. Multipath interference has the potential to contaminate GNSS measurements, imposing a substantial impact on smartphone positioning accuracy [21]. Advanced techniques like VIO, 3DMA GNSS, and pedestrian dead reckoning (PDR), which use smartphone inertial sensors, are employed to address these challenges for enhanced pedestrian positioning. One notable approach is PDR, which exploits the use of inertial sensors, e.g., accelerometers and gyroscopes. Bluetooth and received signal strength indicator are also leveraged [2, 22, 23]. By leveraging a low-cost micro-electro mechanical system (MEMS) IMU in smartphones, an efficient PDR algorithm has been developed, exhibiting advantages such as affordability, simplicity, and user-friendliness when compared to alternative methods [24–26]. Despite the noteworthy PDR advantage of being independent of infrastructure requirements [27], it is important to acknowledge certain limitations associated with this technique, including diminished accuracy and drift. Cost-effective MEMS IMUs encounter a range of inaccuracies, with biases on time-varying [28].

To overcome the challenges and achieve heightened precision in localization, researchers have proposed integrating VIO with supplementary sensors [29–31]. One notable approach involves integrating light detection and ranging (LiDAR) and VIO through the utilization of simultaneous localization and mapping (SLAM) [29]. SLAM technology leverages LiDAR and visual data to create maps of the surrounding environment that are more accurate and detailed, thereby enhancing accuracy, robustness, and overall efficiency. LiDAR for pedestrian tracking exhibits considerable promise due to its exceptional accuracy and resilience across diverse environmental conditions. Nevertheless, it is imperative to acknowledge that this integration approach encounters challenges related to high computational complexity and associated costs.

In the context of urban canyons, the utilization of 3D building models has emerged as a prominent approach to improve GNSS positioning. Specifically, a technique known as 3DMA GNSS [32] has gained considerable popularity. 3DMA GNSS typically adopts a particle-based methodology, wherein measurements are modeled as predictions at various hypothetical positions [33]. The candidate that exhibits the greatest similarity between the modeled and received measurements is deemed to represent the location of the receiver.

Within 3DMA GNSS, two common variants can be identified: shadow matching [34] and ranging-based approaches. Shadow matching involves the matching of satellite visibility patterns across distributed locations. On the other hand, ranging-based 3DMA GNSS employs NLOS-predicted pseudo-range modeling by estimating reflection delays. Utilizing geometric methodologies like Skymask 3DMA [35] and ray-tracing GNSS [36, 37] enables the determination of reflection delays through the validation of signal transmission paths and the computation of reflection delays contingent on predicted reflection points. An alternative technique called likelihood-based ranging [5] adopts a skew-normal distribution for the statistical representation of NLOS delay measurements. It transforms these discrepancies into LOS measurements via a normal distribution.

Moreover, expanding the single-epoch positioning method to incorporate temporal linkages can substantially fortify positioning robustness. Zhong and Groves [5] propose the utilization of a grid filter to evenly distribute positioning candidates, thereby improving solution smoothness. Alternatively, the integration of 3DMA GNSS and velocity estimates obtained through Doppler measurements, coupled with the adoption of FGO for batch optimization, has been explored by Ng et al. [6] as a loosely coupled solution. The results demonstrate that this approach yields a more resilient trajectory suitable for pedestrian applications.

VINS has become an integrated navigation system that leverages the synergy between inertial and visual sensors to calculate a platform's orientation and position. However, due to technological and technical constraints, MEMS gyroscope bias

instability commonly exceeds 15°/h [38]. Consequently, there is a worse result of MEMS in the process of heading initialization. Compared with INS, VINS harnesses the synergies between inertial and visual data, resulting in a proven enhancement of stability and precision in local pose estimation [39]. It is possible to integrate an IMU for the development of VINS and retrieve scale information [28]. While VINS shows promise across various applications, it faces challenges like accumulated errors and time drift in complex environments. Two residual errors stem from IMU predictions and visual odometry estimates. Algorithms such as keyframe-based methods, inertial measurement pre-integration, and machine-learning approaches have been explored to address these. Pre-integrating IMU measurements into relative motion constraints within keyframes improves precision and computational efficiency, enhancing accuracy while reducing complexity [40].

On the other hand, the nonlinear optimization approach can attain highly accurate state estimation [41]. Nonetheless, real-time optimization becomes increasingly challenging as the trajectory expands over time due to the high rate of incoming inertial measurements [40]. DeepVIO, an innovative approach, explores the potential VIO applications in various fields and highlights future research directions, including the utilization of deep learning techniques to enhance performance [42]. Nevertheless, when employing VINS, time drifting continues to be a notable cause of uncertainty.

In summary, Figure 1 highlights the integration of various technologies to enhance location accuracy. Radio-based methods such as Wi-Fi and Bluetooth leverage signal strength indicators for positioning. Sensor-based techniques including IMUs, VIO, and LiDAR integrated with SLAM aid navigation in complex environments. GNSS-based systems, including 3DMA GNSS and Doppler velocity integration, are optimized using FGO. The integration of GNSS and VINS addresses challenges such as signal loss and computational complexity, particularly in urban and indoor settings. Advanced optimization techniques, including nonlinear algorithms and machine learning, refine accuracy. Comparing the extended Kalman filter (EKF) and FGO underscores advantages in nonlinear systems.

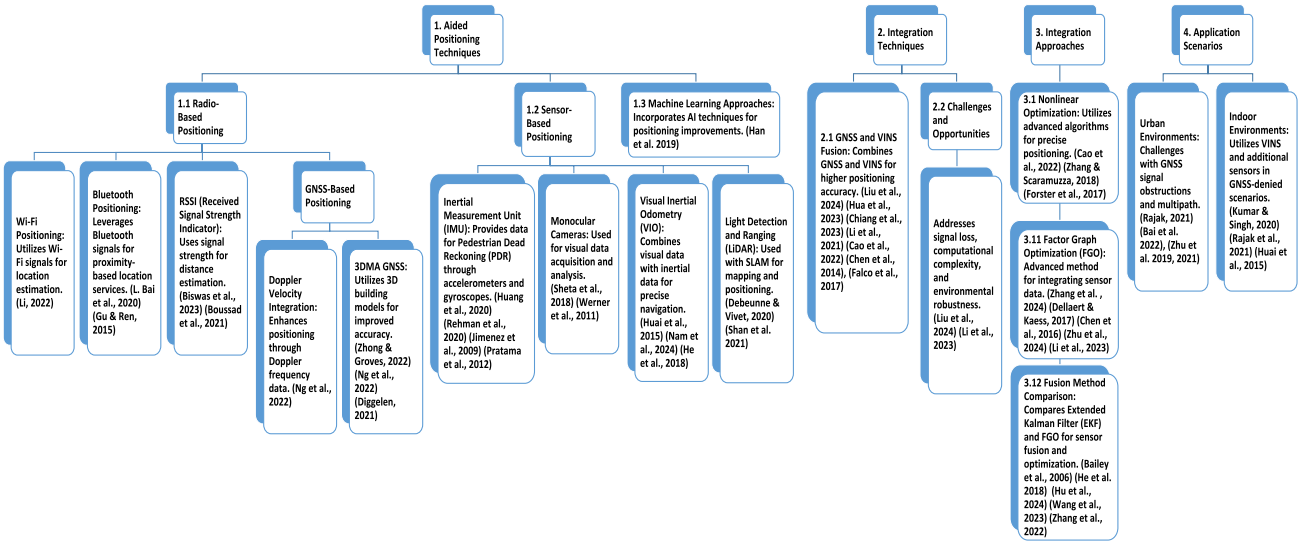
2.3. Review of GNSS/VINS integration in pedestrian positioning

The EKF has been widely utilized for sensor fusion [43–45]; however, its linear approximation of system dynamics compromises accuracy when applied to nonlinear systems. Computational demands of the EKF surge quadratically as the number of 3D landmarks grows, imposing restrictions on its scalability [10]. In densely populated urban regions, the EKF falls short of attaining peak performance because of the accumulation of Gaussian errors [46]. To overcome these limitations, nonlinear optimization methods are proposed as alternatives.

Another promising method for pedestrian positioning is FGO [47], which models the relationships between observed measurements and unknown system states, resulting in high accuracy and efficiency. FGO is superior at managing noisy or incomplete data-changing dynamics than EKF. FGO is better suited for nonlinear problems, using iterative solvers like Gauss-Newton or Levenberg-Marquardt to refine the solution which shows that FGO is a more appropriate method than the EKF. FGO may necessitate greater computational resources, and precise system dynamics modeling is critical. The environmental conditions and specific applications ultimately determine the choice of the positioning system. FGO-based methods show a superior capacity for achieving dependable and uniform state estimation than traditional filter-based techniques due to FGO's adeptness at tackling nonlinear optimization challenges through iterative processes and relinearization [39]. FGO is better equipped to handle the inherent nonlinearities in the pedestrian positioning problem compared to filter-based techniques like EKF. Additionally, FGO adeptly exploits time correlation, which significantly improves accuracy and robustness. This global optimization approach is a key advantage over EKF. The entire graph optimizes simultaneously, leading to a globally consistent solution.

Standalone GNSS and VINS have complementary strengths and limitations. GNSS offers global positioning but struggles with errors in urban and indoor settings, while VINS provides accurate relative positioning but suffers from drift over time. Integrating

Figure 1
Literature mapping



VINS with 3DMA GNSS through FGO enhances accuracy, robustness, and reliability. This combination, along with an IO switch to detect indoor or outdoor environments, allows for more precise and reliable positioning by adjusting algorithms accordingly, offering a robust solution for pedestrian positioning.

2.4. Our contributions

Three problems and key objectives were identified:

Contribution 1: Development of a machine-learning-based method for IO detection using GNSS measurements as features. It facilitates the selection of the most reliable sensor during fusion to maximize smartphone positioning.

Contribution 2: Loosely integrate smartphone-based solutions of VINS and 3DMA GNSS in a batch using FGO and a switching factor to provide complete robustness.

Contribution 3: Validate 3DMA GNSS/VINS performance using FGO and an IO switching factor via actual smartphone experiments in an urban canyon. The code is also open source on https://github.com/queenie-ho/3DMAGNSSVINS-IOFGO.

Our study proposes a comprehensive framework for smartphone-level pedestrian positioning to significantly enhance robustness, efficiency, and accuracy in urban and indoor environments. Existing techniques encounter various challenges, including signal obstructions, drift, and multipath, which hinder their effectiveness. To overcome these challenges, our framework focuses on the integration of VINS and 3DMA GNSS using FGO, aiming to achieve substantial improvements in reliability, accuracy, and robustness, particularly in complicated and demanding environments. It emphasizes IO detection, leveraging machine learning with SVM to accurately differentiate IO conditions.

Our methods directly enhance smartphone navigation accuracy, particularly in urban areas prone to signal obstruction and multipath effects. Reliable indoor positioning is crucial for applications, including wayfinding in large buildings such as shopping malls and airports, asset tracking, and emergency response. VIO and improved indoor navigation systems enhanced positioning accuracy for AR applications, which rely on precise location data to overlay digital information onto the real world. Leveraging smartphone sensors and existing GNSS infrastructure makes the solution cost-effective and widely accessible.

3. Definition and Notation

3.1. Definition

Local frame: A local world frame is an unchanging reference frame employed to represent the orientation and position of objects in the visual-inertial system.

Global frame: A reference frame utilized to represent the orientation and position of objects in an ECEF frame where all axes are fixed with respect to the Earth.

IO (Indoor/Outdoor): The term "indoor" pertains to physically confined areas, while "outdoor" refers to spaces lacking complete confinement [8].

3.2. Notations

A superscript or subscript of the system state and transformation representing a frame, it can be *w* in the local world frame, *b* in the body frame, *c* in the camera frame, and *i* in the IMU frame. A subscript *t* of the frame refers to the frame at a specific time. Table 1 stated the notations.

Table 1.
Notations of frame transformation

Symbol	Explanation
$\mathbf{R}_w^{b_t}$	A rotational transformation from the local world frame to the body frame
$\mathbf{p}_{b_{t_k}}^w$	The position of the body frame with respect to the local world frame at time k (t_k)
$\mathbf{v}_{b_{t_k}}^w$	The velocity of the body frame with respect to the local world frame at t_k
$\mathbf{q}_{b_{t_k}}^w$	The orientation of the body frame with respect to the local world frame at t_k

4. Research Methodology

4.1. Overview

Figure 2
System flowchart of the proposed 3DMA GNSSVINS-IO FGO

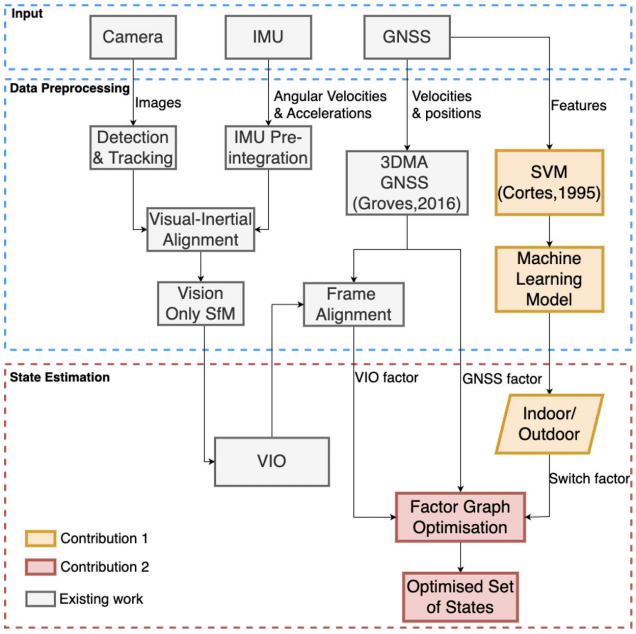


Figure 2 illustrates the positioning framework from this study, featuring an SVM-based machine-learning approach for IO classification. It uses GNSS measurement features, such as satellite numbers received and elevation angle, for categorization. This research combines advanced 3DMA GNSS algorithms involving likelihood-based ranging [5] and shadow matching [34] for improved GNSS performance in a loosely coupled approach [6].

To address GNSS-denied areas and ensure reliable indoor positioning, VIO is used to fill GNSS gaps. The system processes images, pre-integrated angular velocities, and accelerations, avoiding repeated IMU reintegration. A pose graph is defined by selectively pre-integrating IMU measurements between keyframes and skipping frames, resulting in a manageable representation of IMU data. The pre-integrated IMU measurements, combined with visual measurements, generate VIO estimates that provide

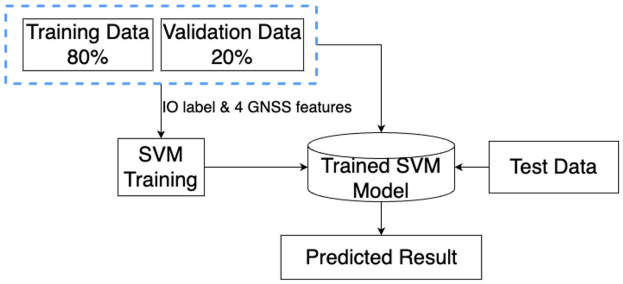
timestamps, position, orientation, and velocity. These estimates are then loosely fused with GNSS measurements and refined through global optimization using nonlinear techniques. It necessitates that the switch factor handles attenuated or scattered signals that may not be received indoors. A loosely coupled integration of GNSS and VINS achieves pedestrian positioning by leveraging the strengths of GNSS, IMU, and monocular camera sensors.

This research aims to deliver reliable sensor fusion: 1) selecting the most reliable sensor for integration and 2) ensuring robustness integration. Consequently, a reliable smartphone-level ubiquitous pedestrian positioning solution can be achieved. The study developed a machine-learning-based IO transition method based on the GNSS feature to select the reliable sensor during incorporation, maximizing the performance of the positioning system. It coordinated frame alignment between the GNSS measurement in a global frame and VINS positioning in a local frame. This alignment ensures consistent and accurate positioning information across different coordinate systems. It loosely integrated VINS and GNSS solutions as a batch using FGO, enhancing the overall robustness of the system for smartphones. FGO enables the fusion of VINS and GNSS measurements, leading to improved positioning accuracy and reliability. Finally, the system's performance was validated through experiments in urban canyons.

4.2. SVM for IO classification

To classify IO conditions, we adopt SVM with GNSS features, including satellite numbers, elevation angle, and average carrier-to-noise ratio (C/N_0). SVM [48] is a supervised learning approach that enables binary classification by identifying a hyperplane to separate different types of data. SVMs are widely utilized in classification, regression, and outlier detection tasks. Using support vectors, SVMs achieve memory efficiency. The flexibility of SVMs allows different kernel functions in the decision function. SVMs for IO classification using GNSS measurements offer a promising solution to the challenges of positioning pedestrians with smartphones in urban areas and where GNSS signals are not available, including indoor areas.

Figure 3
SVM flowchart for IO classification



We collected data from The Hong Kong Polytechnic University encompassing both indoor and outdoor environments, suitable for smartphone use. Figure 3 shows the algorithm structure for IO detection, which relies on GNSS data and predefined relationships among elements like elevation angle, average C/N_0, and the used and received satellite numbers. These elements are categorized as indoor or outdoor based on their expression patterns, which indicate their belonging to the indoor or non-indoor category.

The framework comprises two main stages. In the training stage, the SVM is trained with labeled sample data. IO classification (presumptive data) and the measurements from GNSS are utilized to create a series of weights, then utilized in the next stage. IO conditions are manually labeled as the training data at the training phase.

In the classification stage, the trained SVM assigns scores to new data points based on their similarity to the training data. It employs weights from the previous stage and new GNSS measurements to classify results. Therefore, SVM can classify the new inputs based on their scores. SVM has been widely used in various classification tasks due to its effectiveness in handling high-dimensional data. SVM distinguishes between indoor and outdoor environments by analyzing signal strengths from received satellites or other relevant parameters. This method is both computationally efficient and easy to implement.

Convolutional neural networks (CNNs) have demonstrated success in image-related tasks like object recognition and scene understanding. With labeled indoor and outdoor images, CNN can learn intricate patterns and capture the visual characteristic features that discriminate between these environments [49]. Comparing CNNs with SVMs in this study evaluates the performance of deep learning-based techniques against conventional machine-learning approaches. SVMs have been widely used for various classification tasks, including indoor-outdoor detection; they often rely on handcrafted features manually designed and selected, which can limit their ability to capture the full complexity of visual information. In contrast, CNNs learn relevant visual representations directly from image data, without the need for extensive feature engineering [50]. This study compares CNN's sophisticated feature learning capabilities with conventional techniques, such as SVMs, in indoor-outdoor classification. This comparative evaluation assesses whether visual information improves indoor-outdoor detection accuracy, leveraging the rich visual information present in images, for more accurate and robust detection performance. The classification method is Places365 CNN developed by researchers from the Massachusetts Institute of Technology [49]. The model is trained on a large-scale dataset named Places365, which contains millions of labeled images across a wide range of scene categories, including both indoor and outdoor environments. With the Places365 CNN model, an image is passed through the network to analyze the output probabilities across different scene categories. Examining probabilities tied to indoor and outdoor scene labels, the decision rule classifies the input image as either indoor or outdoor.

4.3. Loosely coupled 3DMA GNSS using FGO

Although VINS is integrated to help outdoor positioning and recovery when GNSS is unavailable, it still relies on GNSS to provide reliable positioning in an absolute coordinate frame. As a result, improving GNSS performance can positively influence the overall performance of the proposed positioning framework. 3DMA GNSS provides a huge potential for improving positioning performance in urban canyons [32]. It is a software-based approach that aids in positioning with 3D building models. Unlike approaches such as the consistency check [51], to exclude the NLOS unhealthy measurement, 3DMA GNSS also uses NLOS measurements as features for positioning.

The cutting-edge 3DMA GNSS algorithms on shadow matching [34] and likelihood-based ranging 3DMA GNSS [5] are combined in this study. To increase the robustness, the 3DMA GNSS solution incorporates Doppler measurements through FGO in a loosely coupled manner. Details are available in Ng et al. [6].

3DMA GNSS spreads out potential positioning candidates around the initial position. It simulates measurements for each candidate and compares them with actual measurements. This research combines shadow matching and likelihood-based ranging for 3DMA GNSS, denoted as $\mathbf{x}_{k,3DMA}$, in the context of FGO.

Shadow matching [34] uses LOS and NLOS satellite visibility as the feature for positioning. Non-received satellites in the ephemeris are assumed to be NLOS satellites, while the received satellites are evaluated in their LOS probability based on the C/N$_0$ value, as $p(LOS|C/N_0)^i$, to mitigate the NLOS reception effect. The LOS probability is then compared with the predicted visibility at each distributed candidate, with the elevation angle of the building boundary at the corresponding azimuth angle, as $p(LOS|BB)_j^i$. The consistency of the satellite at the candidate can be formed,

$$P_j^i = p(LOS|C/N_0)^i p(LOS|BB)_j^i + [1 - p(LOS|C/N_0)^i]\left[1 - p(LOS|BB)_j^i\right] \quad (1)$$

Shadow matching likelihood score for the candidate j is then calculated by the geometrical mean of the visibility consistency of all satellites' $i = 1 \ldots I$, P_j^i,

$$S_{j,SDM} = \left(\prod\nolimits_i P_j^i\right)^{1/I} \quad (2)$$

The candidate with the highest similarity is determined as the receiver's location. The detailed implementation of shadow matching follows in Wang et al. [34].

Likelihood-based ranging 3DMA GNSS [5] determines the user's location by comparing the modeled pseudo-range at each distributed candidate with the measurements. The implementation of likelihood-based ranging 3DMA GNSS follows in pseudo-range, $\widehat{\rho}_j^i$, obtains the smallest difference between the measured, $\hat{\rho}^i$. This forms the pseudo-range innovation of satellite i at candidate j, as $\Delta\rho_j^i$,

$$\Delta\rho_j^i = \hat{\rho}^i - \widehat{\rho}_j^i \quad (3)$$

For the satellite that predicted NLOS at a corresponding candidate, the pseudo-range innovation, $\Delta\rho_j^i$, is inputted to a skewed distribution and then remapped to the normal distribution to eliminate the NLOS delay error. While the predicted LOS satellite at candidate keeps the pseudo-range innovation remains unchanged, such that,

$$\Delta\rho_{j,LBR}^i = \begin{cases} \Delta\rho_j^i & LOS \\ Remapping\left(\Delta\rho_j^i, C/N_0{}^i\right) & NLOS \end{cases} \quad (4)$$

where $C/N_0{}^i$ is the C/N_0 measurements of satellite i. This study selects the highest elevation angle satellite as the reference satellite. $Remapping(\cdot)$ is the function to remap the pseudo-range innovation using a skewed distribution to a normal distribution and become NLOS delay-free pseudo-range innovation, the full implementation can be found in Zhong and Groves [5]. Therefore, an array consists of all pseudo-range innovations of available satellites, from 1 to i, as $\Delta\boldsymbol{\rho}_{j,LBR} = \left[\Delta\rho_{j,LBR}^1 \ldots \Delta\rho_{j,LBR}^i\right]^{\mathrm{T}}$, can be used to calculate the score based on their consistency,

$$S_{j,LBR} = \exp\left(-\left[\frac{1}{\sum \mathbf{Q}}\left(\Delta\boldsymbol{\rho}_{j,LBR}^{\mathrm{T}}\mathbf{Q}^{-1}\Delta\boldsymbol{\rho}_{j,LBR}\right)\right]^{1/I}\right) \quad (5)$$

where $\mathbf{Q}$ is the weighting matrix. Diagonal elements are the uncertainty of each satellite, such that $\sum \mathbf{Q}$ is the summation of the uncertainty of all satellites.

The scores from shadow matching and likelihood-based ranging 3DMA GNSS are then taken from the square root of their product as the integrated score, such as $S_{j,3DMA} = \sqrt{S_{j,SDM} \times S_{j,LBR}}$. As a result, the 3DMA GNSS solution, $\mathbf{p}_{k,3DMA}$, can be obtained by weighted averaging of all distributed candidates, $\mathbf{p}_{j=1\ldots J}$, with their score, $S_{j=1\ldots J,3DMA}$,

$$\mathbf{p}_{k,3DMA} = \frac{\sum_{j=1}^{J} \mathbf{p}_j S_{j,3\mathrm{DMA}}}{\sum_{j=1}^{J} S_{j,3\mathrm{DMA}}} \quad (6)$$

To increase the robustness of positioning, this study integrates 3DMA GNSS with Doppler measurements using FGO. Receiver's velocity and clock drift, $\mathbf{v}_k$, and $c\dot{\delta}t_k$, respectively, are estimated by i-th satellite's Doppler measurements at epoch k, $d_k = \left[d_k^1, \ldots, d_k^i\right]$, via the least-squares method [52]. The FGO framework includes three components. The initial error component limits the Euclidean distance between the 3DMA GNSS solution and the optimized state, denoted as $\mathbf{x}_{k,3DMA}$ and $\mathbf{x}_k$, respectively. The error function can be expressed as,

$$\left\|e_{k,3\mathbf{DMA}}\right\|_{\boldsymbol{\sigma}_{3\mathbf{DMA}}^2}^2 = \left\|\mathbf{p}_k - \mathbf{p}_{k,3DMA}\right\|_{\boldsymbol{\sigma}_{3\mathbf{DMA}}^2}^2 \quad (7)$$

where $\boldsymbol{\sigma}_{3\mathbf{DMA}}^2$ is a diagonal variance matrix of the 3DMA GNSS at x-, y-, and z-axis, respectively.

Two factors are introduced to constraint consecutive epochs, utilizing both the motion propagation model and the constant velocity motion model [53], expressed with error functions, $\left\|e_{k,\mathbf{v}}\right\|_{\boldsymbol{\sigma}_{\mathbf{v},k}^2}^2$ and $\left\|e_{k,\bar{\mathbf{v}}}\right\|_{\boldsymbol{\sigma}_{\bar{\mathbf{v}}}^2}^2$, respectively. They are given by,

$$\left\|e_{k,\mathbf{v}}\right\|_{\boldsymbol{\sigma}_{\mathbf{v},k}^2}^2 = \left\|\mathbf{v}_k - \frac{1}{\Delta t}(\mathbf{p}_{k+1} - \mathbf{p}_k)\right\|_{\boldsymbol{\sigma}_{\mathbf{v},k}^2}^2 \quad (8)$$

$$\left\|e_{k,\bar{\mathbf{v}}}\right\|_{\boldsymbol{\sigma}_{\bar{\mathbf{v}}}^2}^2 = \left\|\frac{1}{2}(\mathbf{v}_k + \mathbf{v}_{k+1}) - \frac{1}{\Delta t}(\mathbf{p}_{k+1} - \mathbf{p}_k)\right\|_{\boldsymbol{\sigma}_{\mathbf{v},k}^2}^2 \quad (9)$$

where Δt represents the time gap between epoch k and $k+1$. $\boldsymbol{\sigma}_{\mathbf{v}}^2$ is a diagonal covariance matrix of velocity $\mathbf{v}_t$,while $\boldsymbol{\sigma}_{\bar{\mathbf{v}}}^2$ is the averaged diagonal covariance matrix at epoch k and $k+1$. As a result, FGO aims to minimize the collective error of three cost functions related to the loosely coupled 3DMA GNSS as,

$$\boldsymbol{\chi}^* = \mathrm{argmin}_{\boldsymbol{\chi}} \sum_k \left\|e_{k,3\mathbf{DMA}}\right\|_{\boldsymbol{\sigma}_{3\mathbf{DMA}}^2}^2 + \left\|e_{k,\mathbf{v}}\right\|_{\boldsymbol{\sigma}_{\mathbf{v},t}^2}^2 + \left\|e_{k,\bar{\mathbf{v}}}\right\|_{\boldsymbol{\sigma}_{\bar{\mathbf{v}}}^2}^2 \quad (10)$$

where $\boldsymbol{\chi} = [\mathbf{p}_1, \ldots, \mathbf{p}_k]$ is the state set of the receiver, while $\boldsymbol{\chi}^*$ represents the optimal set of states. The optimal set of 3DMA GNSS is then integrated with the VINS results.

4.4. GNSS/VINS integration

Establishing a visual trajectory involves detecting and tracking camera image features, matching images. SLAM techniques construct a visual trajectory using structure-from-motion principles [54]. IMU data help estimate scale, velocity, and IMU deviation [55]. Updating every state with every IMU measurement at high frequency is impractical. Therefore, a pose graph consolidates the majority of IMU measurements into a single pose constraint. Pose graph integration with keyframes and pose constraints improves scalability and computational efficiency.

IMU pre-integrations use a continuous-time quaternion-based approach, building upon previous work, and incorporating IMU [53]. Measurements from the IMU include force and platform dynamics with gravity. These measurements include additive noise $\mathbf{n}_a$, acceleration bias $\mathbf{b}_w$, and gyroscope bias $\mathbf{b}_a$. The additional noise in the acceleration and gyroscope measurements follows a Gaussian white noise distribution. The biases follow a random walk and their derivatives are Gaussian white noise. The raw gyroscope and accelerometer measurements are $\widehat{\boldsymbol{\omega}}$ and $\widehat{\mathbf{a}}$, expressed in Equation (11).

$$\begin{aligned}\widehat{\mathbf{a}}_t &= \mathbf{a}_t + \mathbf{b}_{a_t} + \mathbf{R}_w^{b_t}\mathbf{g}^w + \mathbf{n}_a \\ \widehat{\boldsymbol{\omega}}_t &= \boldsymbol{\omega}_t + \mathbf{b}_{w_t} + \mathbf{n}_w\end{aligned} \tag{11}$$

where $\mathbf{a}_t$ represents the expected accelerometer measurements, $\mathbf{g}^w$ represents the gravity in the world frame, $\mathbf{n}_a$ represents the additive noise, ω_t represents the expected accelerometer measurements gyroscope, $\mathbf{R}_{\mathrm{w}}^{b_t}$ represents a rotation matrix of the transformation from the local world frame to the body frame b_k at time t, and both $\mathbf{n}_a$ and $\mathbf{n}_w$ are the additive noise, calculated according to IMU specification.

Gaussian white noise distributions the measurement errors of acceleration and gyroscope. A random walk process to model the biases of acceleration and gyroscope, using Gaussian white noise. To transmit the covariance of the orientation angles α, β, and γ:

$$\boldsymbol{\alpha}_{b_{k+1}}^{b_k} = \iint_{t\in[t_k,t_{k+1}]} \mathbf{R}_t^{b_k}\left(\widehat{\mathbf{a}}_t - \mathbf{b}_{a_t}\right)dt^2 \tag{12}$$

$$\boldsymbol{\beta}_{b_{k+1}}^{b_k} = \int_{t\in[t_k,t_{k+1}]} \mathbf{R}_t^{b_k}\left(\widehat{\mathbf{a}}_t - \mathbf{b}_{a_t}\right)dt \tag{13}$$

$$\gamma_{b_{k+1}}^{b_k} = \int_{t\in[t_k,t_{k+1}]} \frac{1}{2}\boldsymbol{\Omega}\left(\widehat{\omega}_t - \mathbf{b}_{w_t}\right)\gamma_t^{b_k}dt \tag{14}$$

$$\Omega(\omega) = \begin{bmatrix} 0 & -\omega_z & \omega_y & \omega_x \\ \omega_z & 0 & -\omega_x & \omega_y \\ -\omega_y & \omega_x & 0 & \omega_z \\ \omega_x & \omega_y & \omega_z & 0 \end{bmatrix} \tag{15}$$

The equations apply to $[t_k, t_{k+1}]$ of two consecutive frames between b_k and b_{k+1}, where b_k denoted the reference frame with the bias, $\mathbf{R}_t^{b_k}$ represents the rotation matrix of the transformation from the body frame to the reference frame b_k at time t, and ω_x, $\omega_{y,}$ and ω_z represents the angular velocities in the body frame.

4.5. Probabilistic framework formulation

4.5.1. Inertial factor

In this context, b_{t_k} refers to the body frame at time t_k, and the set $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}\}$ denotes the relative position, velocity, and rotational information of b_{t_k} and $b_{t_{k+1}}$.

Lastly, the residual that establishes the relationship between the system states and pre-integrated IMU [53] measurements is expressed as

$$\mathbf{r}_{\mathcal{B}}\left(\tilde{\mathbf{z}}_{b_{t_{k+1}}}^{b_{t_k}}, \mathcal{X}\right) = \begin{bmatrix} \delta\boldsymbol{\alpha}_{b_{t_{k+1}}}^{b_{t_k}} \\ \delta\boldsymbol{\beta}_{b_{t_{k+1}}}^{b_{t_k}} \\ \delta\boldsymbol{\theta}_{b_{t_{k+1}}}^{b_{t_k}} \\ \delta\mathbf{b}_a \\ \delta\mathbf{b}_g \end{bmatrix} \tag{16}$$

$$= \begin{bmatrix} \mathbf{R}_w^{b_{t_k}}\left(\mathbf{p}_{b_{t_{k+1}}}^w - \mathbf{p}_{b_{t_k}}^w + \frac{1}{2}\mathbf{g}^w\Delta t_k^2 - \mathbf{v}_{b_{t_k}}^w\Delta t_k\right) - \widehat{\boldsymbol{\alpha}}_{b_{t_{k+1}}}^{b_{t_k}} \\ \mathbf{R}_w^{b_{t_k}}\left(\mathbf{v}_{b_{t_{k+1}}}^w + \mathbf{g}^w\Delta t_k - \mathbf{v}_{b_{t_k}}^w\right) - \widehat{\boldsymbol{\beta}}_{b_{t_{k+1}}}^{b_{t_k}} \\ 2\left[\mathbf{q}_{b_{t_k}}^{w^{-1}} \otimes \mathbf{q}_{b_{t_{k+1}}}^w \otimes \left(\widehat{\boldsymbol{\gamma}}_{b_{t_{k+1}}}^{b_{t_k}}\right)^{-1}\right]_{xyz} \\ \mathbf{b}_{ab_{t_{k+1}}} - \mathbf{b}_{ab_{t_k}} \\ \mathbf{b}_{wb_{t_{k+1}}} - \mathbf{b}_{wb_{t_k}} \end{bmatrix}$$

where $\mathbf{p}_{b_{t_k}}^w$, $\mathbf{v}_{b_{t_k}}^w$, $\mathbf{q}_{b_{t_k}}^w$ are the position, velocity, and orientation, respectively, of the body frame with respect to the local world frame at t_k; $\mathbf{R}_w^{b_{t_k}}$ transforms from the body frame to local world frame at t_k; $\delta\theta_{b_{t_{k+1}}}^{b_{t_k}}$ denotes the error in the relative rotation within the 3-D Euclidean space. The operator $\left[\mathbf{q}_{b_{t_k}}^{w^{-1}} \otimes \mathbf{q}_{b_{t_{k+1}}}^w \otimes \left(\widehat{\boldsymbol{\gamma}}_{b_{t_{k+1}}}^{b_{t_k}}\right)^{-1}\right]_{xyz}$ extracts the imaginary component of a quaternion.

4.5.2. Visual factor

In our system, the visual measurement use of sparse feature points extracted from image frames. These points are identified as strong corners [56] and are tracked using the iterative Lucas-Kanade method [57]. After applying distortion correction [58], the projection process is modeled as $\tilde{\mathcal{P}} = [u, v]^T$

$$\tilde{\mathcal{P}} = \pi_c\left(\mathbf{R}_b^c\left(\mathbf{R}_w^b\mathbf{x}^w + \mathbf{p}_w^b\right) + \mathbf{p}_b^c\right) + \mathbf{n}_c \tag{17}$$

where $\tilde{\mathcal{P}}$ denotes the feature coordinates in the image plane, $\mathbf{x}^w$ represents the corresponding 3-D landmark position in the local world frame, $\pi_c(.)$ denotes the camera projection function in the camera frame, and $\mathbf{n}_c$represents the measurement noise.

$$\mathbf{r}_C(\tilde{\mathbf{z}}_l, \mathcal{X}) = \tilde{\mathcal{P}}_l^{a_j} - \pi_e\left(\widehat{\mathbf{x}}_l^{\Theta_j}\right) \tag{18}$$

$$\widehat{\mathbf{x}}_l^{c_j} = \mathbf{R}_b^c\left(\mathbf{R}_w^{b_j}\left(\mathbf{R}_{b_i}^w\left(\mathbf{R}_c^b\frac{1}{\rho_l}\pi_c^{-1}\left(\widetilde{P}_l^{c_i}\right) + \mathbf{p}_c^b\right) + \mathbf{p}_{b_i}^w\right) + \mathbf{p}_w^{b_w}\right) + \mathbf{p}_b^c \tag{19}$$

To link the observations of a feature point l with inverse depth $\mathcal{P}_l$ in two frames i and j, we use the residual $\{\mathbf{R}_b^c, \mathbf{t}_c^b\}$, which represents the transformation between the IMU and the camera. This residual is expressed as the difference between the projected feature point in frame j and the estimated feature point in frame i. The estimated feature point is obtained by applying a sequence of rotation and translation transformations to the projected feature point in frame

j. These transformations relate the camera frame to the IMU frame and the local world frame to the IMU frame.

4.5.3. GNSS VINS alignment

Since VINS completes gravitational acceleration alignment, only the yaw angle offset between the coordinate systems of VINS and GNSS remains estimated before integration. This study optimizes the positioning solution in the world frame. The alignment process uses the motion of the GNSS antenna and VINS. Velocity, $\mathbf{v}_k$, estimated by Doppler measurements facilitates the alignment process once Doppler measurements are less noisy than pseudo-range measurements.

The relative position from VINS, $\mathbf{v}_{VIO}$, rotates over the yaw, ψ, axis using the rotational matrices from local frame to global frame in ENU and ENU to ECEF, $\mathbf{R}_{ECEF}^{ENU}\mathbf{R}_{ENU}^{local}(\psi)$, aligning VINS with GNSS. An objective function is formulated to minimize the difference between the relative motion of GNSS velocity and VINS by changing the yaw angle, ψ, expressed as,

$$\psi^* = \underset{\psi}{\operatorname{argmin}} \left\| \mathbf{v}_k - \mathbf{R}_{ECEF}^{ENU}\mathbf{R}_{ENU}^{local}(\psi)\mathbf{v}_{VIO} \right\|^2 \tag{20}$$

As a result, the transformation between the local frame and the global frame is completely calibrated. The estimated yaw angle, ψ, is used in the optimization process for VINS to rotate to the global frame and integrate with GNSS.

4.5.4. FGO

FGO integrates multiple sensors in the GNSS/VINS system, using a loosely coupled approach. It employs nonlinear optimization via factor graphs to capture interdependencies and enhance accuracy, reducing challenges related to drift and error accumulation. Following the alignment of the global and local frames, the obtained results undergo further optimization. The VINS estimates become factors in the graph, with GNSS measurements incorporated as constraints to refine these estimates.

The problem is constructed by setting up the parameter blocks and residual blocks. The Ceres Solver [59] is employed as the optimization solver for this study, and the system structure is illustrated in Figure 4.

$$X = [\mathrm{x}_0, \mathrm{x}_1, \mathrm{x}_2, \ \ldots, \ \mathrm{x}_\mathrm{t}] \tag{21}$$

$$\mathrm{x}_{0,\mathrm{t}} = \left(\mathbf{p}_{0,\mathrm{t}}, \mathbf{v}_{0,\mathrm{t}}, \mathbf{q}_{0,\mathrm{t}}\right)^{\mathrm{T}} \tag{22}$$

Figure 4 showcases the system state representation, where X represents system states in Equation (21). The system states at each epoch t, as x_t, are composed of position ($\mathbf{p}$), velocity ($\mathbf{v}$), and orientation ($\mathbf{q}$) parameters denoted as $\mathrm{p}_{0,\mathrm{t}}, \mathrm{v}_{0,\mathrm{t}}, \mathrm{q}_{0,\mathrm{t}}$ in Equation (22).

$$\left\| f_t^{VINS} \right\|_{\Sigma_t^{\mathrm{VINS}}}^2 = \left\| \left[\frac{\mathbf{q}_{t-1}^{l\,-1}\mathbf{R}(\mathbf{p}_t^l - \mathbf{p}_{t-1}^l)}{\mathbf{q}_{t-1}^{l\,-1} - \mathbf{q}_t^l} \right] - \left[\frac{\mathbf{q}_{t-1}^{w\,-1}\mathbf{R}(\mathbf{p}_t^w - \mathbf{p}_{t-1}^w)}{\mathbf{q}_{t-1}^{w\,-1} - \mathbf{q}_t^w} \right] \right\|_{\Sigma_t^{\mathrm{VINS}}}^2 \tag{23}$$

$$\left\| f_t^{3DMA\ GNSS} \right\|_{\Sigma_t^{3\mathrm{DMA\ GNSS}}}^2 = \left\| \mathbf{p}_t^{3DMA\ GNSS} - p_t \right\|_{\Sigma_t^{3\mathrm{DMA\ GNSS}}}^2 \tag{24}$$

$$\left\| f_t^{SW} \right\|_{\square}^2 = \left\| s_t f_{t,3DMA\ GNSS} + \left((1 - s_t) f_{t,VINS}\right) \right\|_{\square}^2 \tag{25}$$

where Σ_t^* is the covariance matrix of the corresponding methods, $*$, at epoch t.

The term $\left\| f_t^{VINS} \right\|_{\Sigma_t^{\mathrm{VINS}}}^2$ in Equation (23) represents the VINS factor, which computes the difference between the estimated orientation using the local frame $\mathbf{q}_{t-1}^l$ and the relative rotation obtained from the position difference $(\mathbf{p}_t^l - \mathbf{p}_{t-1}^l)$, and the global frame $\mathbf{q}_{t-1}^w$ and $(\mathbf{p}_t^w - \mathbf{p}_{t-1}^w)$.

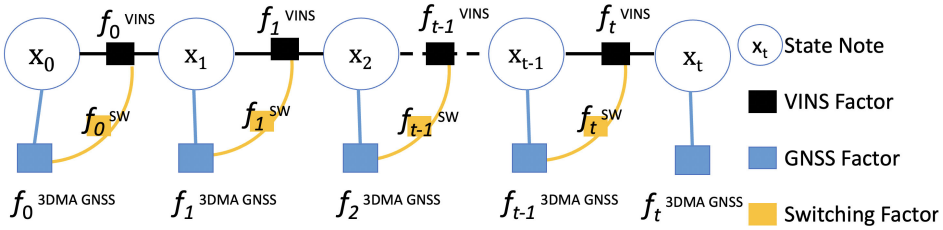
The term $\left\| f_t^{3DMA\ GNSS} \right\|_{\Sigma_t^{3\mathrm{DMA\ GNSS}}}^2$ in Equation (24) represents the GNSS factor, which measures the position difference between the loosely coupled FGO 3DMA GNSS-derived position $\mathbf{p}_t^{GNSS}$ and the estimated position p_t.

The term $\left\| f_t^{SW} \right\|_{\square}^2$ in Equation (25) represents the switching factor, which combines the 3DMA GNSS and VINS factors based on a switching variable s_t. The term $s_t f_{t,GNSS}$ corresponds to the 3DMA GNSS factor contribution, while $(1 - s_t) f_{t,VINS}$ corresponds to the VINS factor contribution.

$$\mathrm{X}^* = \underset{\mathrm{X}}{argmin} \sum_t \left\| f_t^{VINS} \right\|_{\Sigma_t^{VINS}}^2 + \left\| f_t^{3DMA\ GNSS} \right\|_{\Sigma_t^{3DMA\ GNSS}}^2 + \left\| f_t^{SW} \right\|_{\square}^2 \tag{26}$$

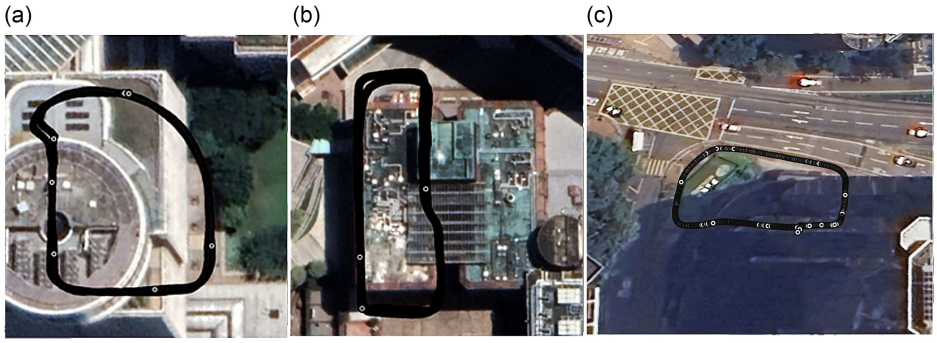
This estimate of the system states X by minimizing the objective function in Equation (26), which comprises the VINS factors, 3DMA GNSS factors, and switching factors. The optimal system state X^* is obtained by solving the estimation problem and minimizing the costs associated with each measurement type. The FGO framework integrates measurements from VINS, 3DMA GNSS, and switching IO to accurately estimate pedestrian position. The optimization process achieves the maximum a posteriori estimation by jointly optimizing all states and minimizing the costs from measurements.

Figure 4
FGO structure on loosely coupling GNSS and VINS with IO switching factor



5. Experiment

Figure 5
(a) Trajectory A, (b) Trajectory B, and (c) Trajectory C



Experiments were conducted in the urban city of Hong Kong. Trajectories A and B were conducted at The Hong Kong Polytechnic University and C in Tsim Sha Tsui. Trajectory A contained both outdoor and indoor environments, with outdoor areas surrounded by trees and indoor areas with sufficient light. Trajectory B includes both open and confined areas outdoors and in a narrow corridor indoors. Trajectory C includes urban outdoor and indoor environments, taking place in the bus terminal, with many people, buses, and insufficient indoor lighting.

All three experiments involved two scenarios to evaluate performance with different positioning. Trajectory A, B, and C were tested under Scenario 1 (starting from indoors, transitioning to outdoors, and returning to indoors) and Scenario 2 (starting from outdoors, transitioning to indoors, and returning to outdoors. We compared the root mean square error (RMSE) of absolute positioning error, standard deviation (std), and SVM accuracy of VINSMONO, 3DMA GNSS, 3DMA GNSSVINS FGO, and 3DMA GNSSVINS-IO FGO in all experiments. The ground truth (GT) of trajectories A, B, and C, accordingly, presented in Google Earth can be seen in Figure 5, respectively. This section presents a comparison of the following methods:

- VINSMONO [53]: A monocular visual-inertial state estimator.
- 3DMA GNSS [6]: Loosely coupled 3DMA GNSS and velocity estimated by weighted least squares.
- 3DMA GNSSVINS FGO: The proposed method without IO switching factor in FGO.
- 3DMA GNSSVINS-IO FGO: The proposed method utilizes VINS, 3DMA GNSS, and FGO in a loosely coupled way, incorporating the IO switching factor.

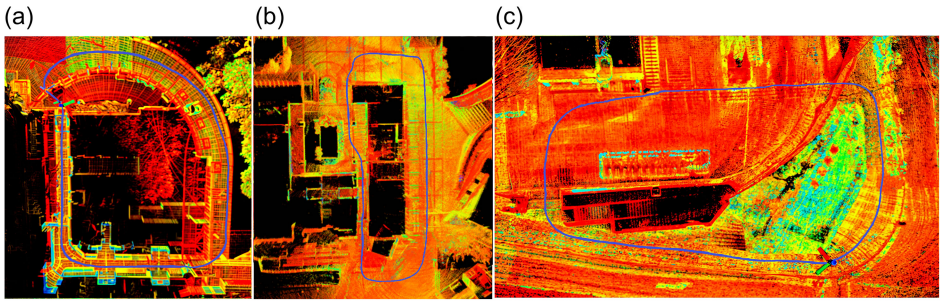
5.1. Experiment setup

In this experiment, we utilized the Xiaomi Mi8 smartphone, to acquire the GNSS measurements as well as inertial and visual data. The smartphone is equipped with a Triple-axis MEMS-IMU (TDK-InvenSense ICM-20690) operating at 100Hz. Additionally, it features GNSS receivers with a Broadcom BCM47755 chip, receiving the signals from GPS (L1+L5), Galileo (E1+E5a), GLONASS (L1), and Beidou (B1) at 1 Hz during the experiments. The smartphone also incorporates a monocular camera with a resolution of 1280 × 640 pixels and a pixel size of 1.4 μm, operating at 30 frames per second. The output rate after integrating with IMU and GNSS is 10 Hz. To ensure accurate data acquisition, we performed calibration of the extrinsic and intrinsic parameters of the IMU and camera of the Xiaomi Mi8. A window size of 10 was employed, consistent with VINS-Mono. However, the loop closure function was not enabled. For optimization purposes, the Ceres library was selected. A desktop personal computer that had an Intel i9-9900K processor operating at 3.6 GHz and 31.2 GB of memory was used.

5.2. Experimental validation

LiDAR technology is employed as the GT reference in our investigation to attain precise centimeter-level positioning.

Figure 6
Ground truth (GT) generated by LIO-SAM: (a) Trajectory A, (b) Trajectory B, and (c) Trajectory C



Specifically, we utilized the HDL 32E Velodyne LiDAR, with a 360-degree horizontal field of view and a vertical field of view spanning from +10° to -30°, with an 80-meter range, recorded at 10Hz. In addition, the Xsens Mti 10 AHRS IMU model, operating at a frequency of 5.0 Hz, was employed. Figure 6 presents the LiDAR GT data generated by the LIO-SAM algorithm for three distinct trajectories, denoted as A, B, and C. By leveraging the information from the LiDAR point cloud map, the IO flag can be accurately labeled, indicating whether the pedestrian is indoors or outdoors as GT for validation.

Table 2.
Classification accuracy (%)

Trajectory (Scenario)	A (1)	A (2)	B (1)	B (2)	C (1)	C (2)
SVM classification accuracy	84	77	94	97	94	99
CNN classification accuracy [45]	91	76	88	63	66	50

Table 2 presents the SVM and CNN classification accuracy for six scenarios: A1, A2, B1, B2, C1, and C2. It includes the accuracy percentage for the SVM and CNN algorithms. SVM achieved better accuracy with 92.8% than CNN with 72.3% on average. SVM is better suited for classifying indoor and outdoor environments.

Trajectory A in SVM IO detection shows lower accuracy compared to trajectories B & C. This discrepancy stems from the complexity of the environment in trajectory A, which poses greater challenges than trajectories B & C.

In contrast, the CNN IO detection results reveal that trajectory A has better accuracy compared to trajectories B & C. This is due to the complexity of the environment in trajectory A, which poses more features for the CNN classification than trajectories B & C.

Overall, SVM performs better than CNN. SVM primarily relies on manually engineered features, which can capture information beyond what is visible in the image. CNNs rely heavily on visual patterns to make classifications. If an outdoor environment is partially covered or occluded, the CNN may not have access to the complete visual information necessary to accurately classify it as outdoor. In such cases, SVM's reliance on manually engineered features is an advantage.

Abrupt changes in light conditions can pose significant challenges for CNNs. If the camera encounters low light or overexposure, the image may be too dark or too bright, reducing CNN accuracy.

Furthermore, GNSS may not function properly for short periods. Accurate IO classification is critical in these scenarios to choose the right sensor to trust, and SVM's robustness makes it more suitable.

By comparing SVM and CNN accuracies, we conclude that SVM is more suitable for IO classification in environments where visual features may be inconsistent or occluded, and where robustness to lighting conditions is crucial. This makes SVM a more reliable choice for accurate indoor-outdoor classification in complex and varying environments.

5.3. Experiment result

Table 3
The RMSE of absolute positioning error and the standard deviation

Experiment	Algorithm	Std (m)	RMSE
Trajectory A Scenario 1 (A1)	VINSMONO	2.47	11.94
	3DMA GNSS	NA	8.80
	3DMA GNSSVINS FGO	4.42	12.87
	3DMA GNSSVINS-IO FGO	5.29	8.34
Trajectory A Scenario 2 (A2)	VINSMONO	26.01	13.00
	3DMA GNSS	NA	37.47
	3DMA GNSSVINS FGO	12.55	44.60
	3DMA GNSSVINS-IO FGO	6.56	18.63
Trajectory B Scenario 1 (B1)	VINSMONO	13.24	7.29
	3DMA GNSS	1.73	10.29
	3DMA GNSSVINS FGO	4.65	6.80
	3DMA GNSSVINS-IO FGO	NA	5.73
Trajectory B Scenario 2 (B2)	VINSMONO	3.93	8.20
	3DMA GNSS	NA	13.86
	3DMA GNSSVINS FGO	3.93	18.35
	3DMA GNSSVINS-IO FGO	2.75	3.91
Trajectory C Scenario 1 (C1)	VINSMONO	2.36	5.31
	3DMA GNSS	NA	5.48
	3DMA GNSSVINS FGO	2.31	3.68
	3DMA GNSSVINS-IO FGO	1.89	2.93
Trajectory C Scenario 2 (C2)	VINSMONO	13.57	13.00
	3DMA GNSS	NA	9.29
	3DMA GNSSVINS FGO	3.28	7.61
	3DMA GNSSVINS-IO FGO	2.49	6.47

Table 3 presents the results of six scenarios for VINSMONO, 3DMA GNSS, 3DMA GNSSVINS FGO, and 3DMA GNSSVINS-IO FGO across trajectories A, B, and C in two different scenarios. The std represents the variability or spread of the positioning errors. The RMSE values indicate the error between the estimated and actual positions, providing an overall measure of accuracy.

The 3DMA GNSSVINS-IO FGO algorithm demonstrates improved performance. The RMSE results indicate 3DMA GNSSVINS-IO FGO improvement of 4.53 meters in A1, 18.84 meters in A2, 4.56 meters in B1, 4.29 meters in B2, 2.93 meters in C1, and 6.47 meters in C2. It achieves lower std, showing more consistent positioning accuracy.

The integration of the IO detection switch in the 3DMA GNSSVINS-IO FGO algorithm significantly contributes to its improved performance. The IO switch helps mitigate challenges related to global drift and enhances the algorithm's ability to accurately estimate local poses, improving positioning accuracy.

In summary, trajectory A demonstrates lower accuracy in IO detection compared to trajectory B and trajectory C, due to the complexity of the environment. Additionally, the 3DMA GNSSVINS-IO FGO algorithm generally performs well with relatively low std, indicating consistent positioning accuracy. This improvement is due to the integration of the IO switching factor. VINSMONO shows lower std in some scenarios. Additionally, the std is not applicable for the 3DMA GNSS algorithm since the GNSS was not available all the time.

Table 4.
3DMA GNSS Estimated Std(m)

Trajectory (Scenario)	A (1)	A (2)	B (1)	B (2)	C (1)	C (2)
3DMA GNSS estimated std	48.19	35.91	28.52	17.89	38.07	34.95

Table 4 highlights the huge 3DMA GNSS Std which 3DMA GNSS might not be throughout the entire period, including indoors and the transition time. IO detection mainly helps during transitions, in that time it is more difficult for GNSS to perform well due to reflections and diffraction.

The IO detection can prevent large errors of GNSS. The IO trusts VINS first indoors, the error and std are small and once the IO detected it as outdoor, the IO trusts GNSS more, and the error may increase during this transition period. When no IO switching factor, it is based on the weighting of the std of GNSS, the GNSS affects the positioning performance even if the error is high. It mistrusted the GNSS position, and the performance got worse while the std of the positioning error remained lower than having IO detection. Therefore, our proposed method effectively addresses this issue, enhancing the overall positioning performance and reliability.

Table 3 shows that the 3DMA GNSSVINS-IO FGO has the highest std and lowest RMSE with values of 5.29 meters and 8.34 meters, respectively, which indicates a greater variability of positioning errors and higher accuracy compared to other algorithms. However, the incorporation of the IO switching factor may introduce additional complexity or noise into the positioning solution, leading to increased variability. The proposed method contributes to reduced variability by incorporating VINS measurements.

Figure 7 presents the absolute positioning error observed in Experiment A1. The integration of VINSGNSS-IO effectively reduces the overall positioning error. It is important to note that erroneous covariance in the 3DMA GNSS module can compromise the optimization outcomes, particularly during epochs spanning 80–100, where the covariance fails to accurately bound the actual positioning error. Consequently, FGO mistakenly relies on inaccurate 3DMA GNSS data, leading to performance degradation. Significant GNSS error with a small covariance around epoch 75 substantially affects the performance of VINSGNSS FGO, resulting in distorted optimized results. The IO detection mechanism effectively identifies the epochs with the highest GNSS errors, mitigating the impact of large positioning errors.

Figure 7
The A1 absolute positioning error comparison with the estimated 3DMA GNSS standard deviation (STD) and IO indicator

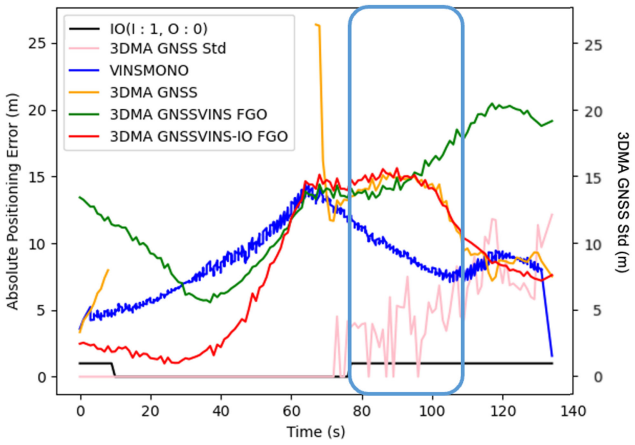


Figure 8
The A1 trajectory comparison

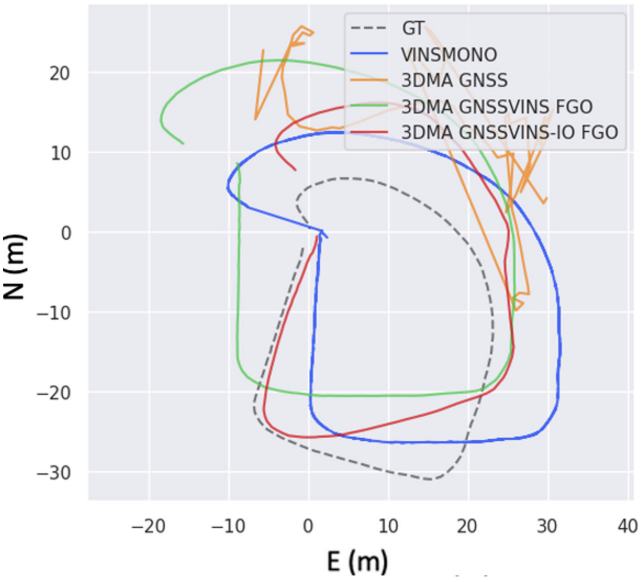


Figure 8 presents the trajectories observed in Experiment A1. We observe that the proposed method's trajectory is closest to the GT.

Table 3 shows that the 3DMA GNSSVINS-IO FGO algorithm exhibits a low std and low RMSE with values of 6.56 meters and 18.63 meters, respectively, suggesting improved accuracy compared to the previous algorithm due to adding the IO factor. 3DMA GNSSVINS-IO FGO demonstrated its capability to recover orientation errors from VINS when IO is indicated as outdoor and the positioning with no GNSS signals in Figure 9. It also shows an increase of 3DMA GNSSVINS FGO error in time 25 to time 55, while there is a loss of GNSS signal and an accumulating error of VINS MONO. The 3DMA GNSS VINS-IO FGO solution maintains better positioning accuracy during transitions from outdoor to indoor environments compared to the standard 3DMA GNSSVINS FGO method. The standard method

Figure 9
The A2 absolute positioning error comparison with the estimated 3DMA GNSS standard deviation (std) and IO indicator

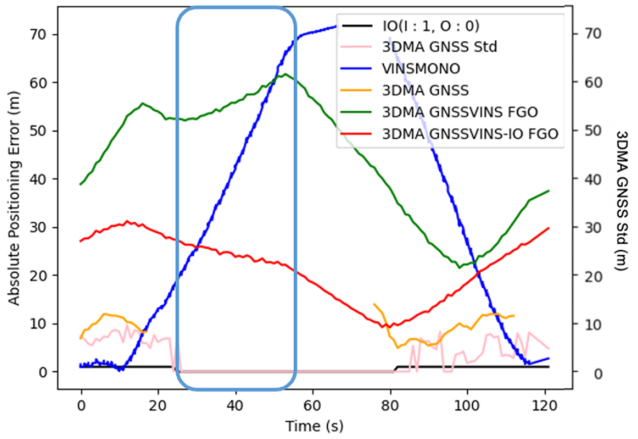
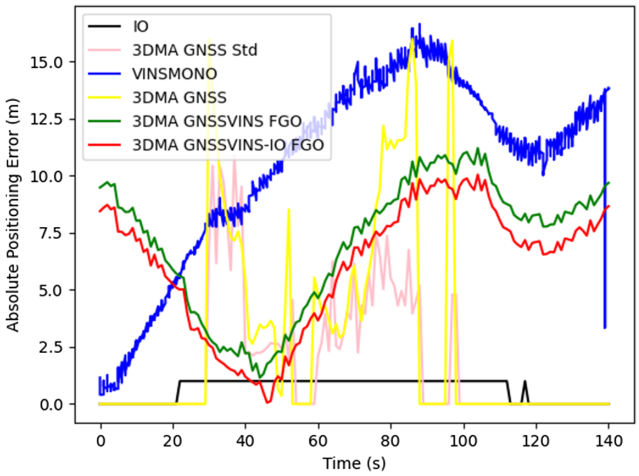


Figure 11
The B1 absolute positioning error comparison



struggles with the loss of GNSS signals and not trusting the VIO sufficiently. 3DMA GNSS uncertainties lead to mistrust in the system and degrade the overall performance when transitioning indoors.

Figure 10 shows that 3DMA GNSS gets the closest epochs when it is available, and our proposed method 3DMA GNSSVINS-IO FGO gets the smooth closest trajectory among these methods to the GT. VINSMONO has an orientation error and affects the 3DMA GNSSVINS FGO method. Our proposed method recovered orientation errors from VINS MONO.

Table 3 shows that the highest std comes from 3DMA GNSSVINS FGO, with a value of 13.24 meters and the lowest RMSE of 6.80 meters, indicating relatively high variability but good accuracy. Meanwhile, the algorithm with the lowest std and RMSE comes from 3DMA GNSSVINS-IO FGO, with a value of 1.73 meters, indicating more consistent performance. The integration of the IO switching factor contributes to this improved performance. The absolute positioning error graph, as shown in Figure 11, shows how the integration of VINS and 3DMA GNSS effectively mitigates the drift from the VINS system, resulting in significantly improved positioning performance. Integrating IO detection further improves positioning performance compared to not using IO detection. Overall, the positioning performance was consistently better when IO detection was employed.

In addition, Figure 12 presents the trajectories of our proposed method, which gets the smoothest trajectory among these methods while getting a close trajectory with the GT.

Table 3 shows that 3DMA GNSSVINS-IO FGO has the lowest std with a value of 2.75 meters, integrating with the IO switching factor contributes to this improved performance. Figure 13 shows a smoother transition into indoor environments for the 3DMA GNSS VINS-IO FGO method **despite large GNSS uncertainty**

Figure 10
The A2 trajectory comparison

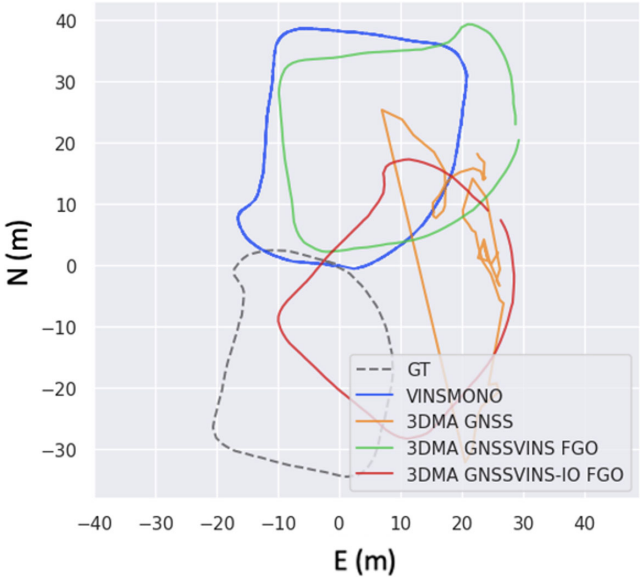


Figure 12
The B1 trajectory comparison

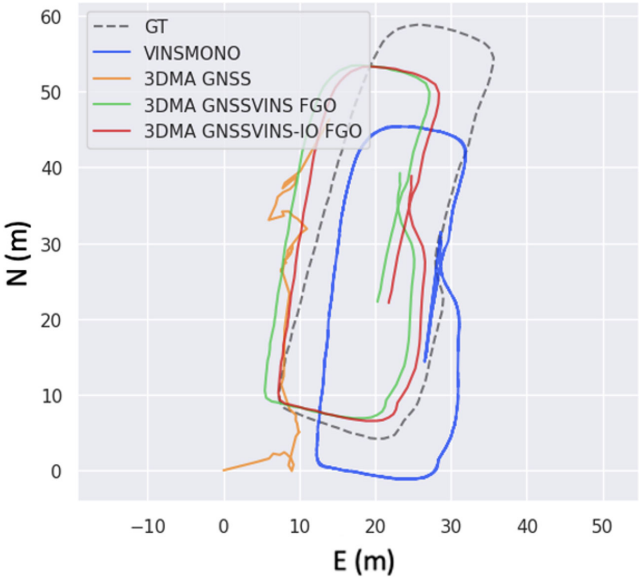
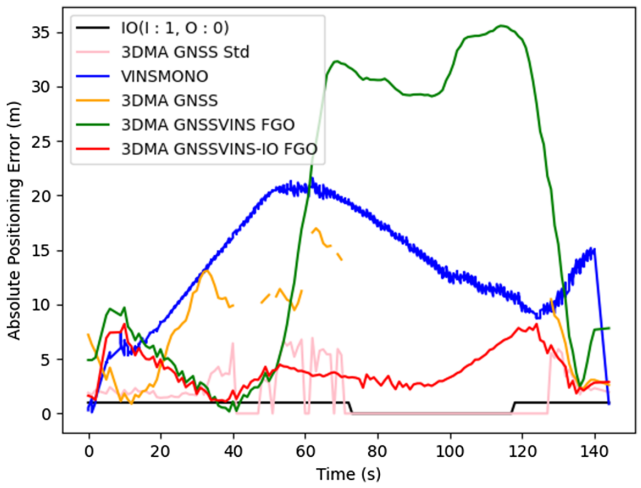


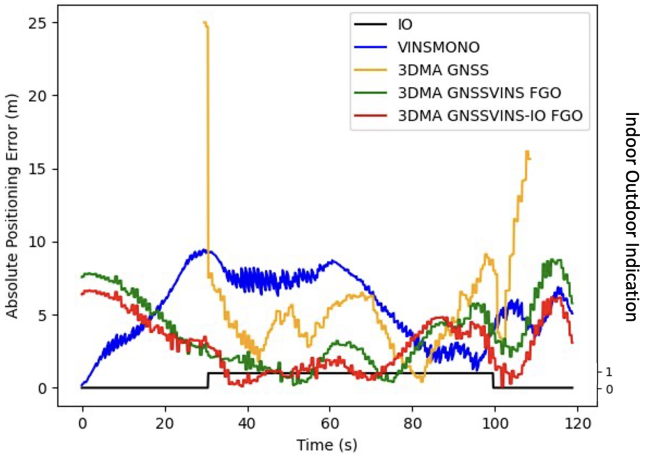
Figure 13
The B2 absolute positioning error comparison with the estimated 3DMA GNSS standard deviation (std) and IO indicator



during the transition from time 40 to 80. Despite the large 3DMA GNSS std, a system without IO remains affected by inaccurate 3DMA GNSS performance. In contrast, the 3DMA GNSS VINS-IO FGO method effectively balances the transition. Particularly, when IO detection is applied, the overall positioning performance is notably improved, especially during epochs 50s to 70s, where the peak positioning error is successfully suppressed. 3DMA GNSSVINS-IO FGO enhances the device's correct heading and significantly reduces the overall positioning error of 3DMA GNSS/VINS-IO FGO compared to not using IO.

Figure 14 shows that our proposed method gets a smooth and closest trajectory among these methods with the GT while 3DMA GNSS gets the closest epochs depending on its available and the

Figure 15
The C1 absolute positioning error comparison



3DMA GNSSVINS FGO method has a wrong device heading which is affected by the VINSMONO error. Our proposed method 3DMA GNSSVINS-IO FGO recovered the heading error and resulted in the closest trajectory.

Table 3 shows that both 3DMA GNSSVINS FGO and VINSMONO have the highest std, with a value of 2.31 and 2.36 meters, respectively. Meanwhile, the algorithm with the lowest std and RMSE is 3DMA GNSSVINS-IO FGO, with a value of 1.89 meters and 2.93 meters, respectively.

There was a large positioning error between times 20s and 70s presented in Figure 15. However, this error was successfully suppressed by integrating GNSS and VINSMONO. The integration of these two systems effectively improved the accuracy and stability of the positioning results during that period. Also, Figure 16 shows that the proposed method achieves the closest trajectory to the GT.

Figure 14
The B2 trajectory comparison

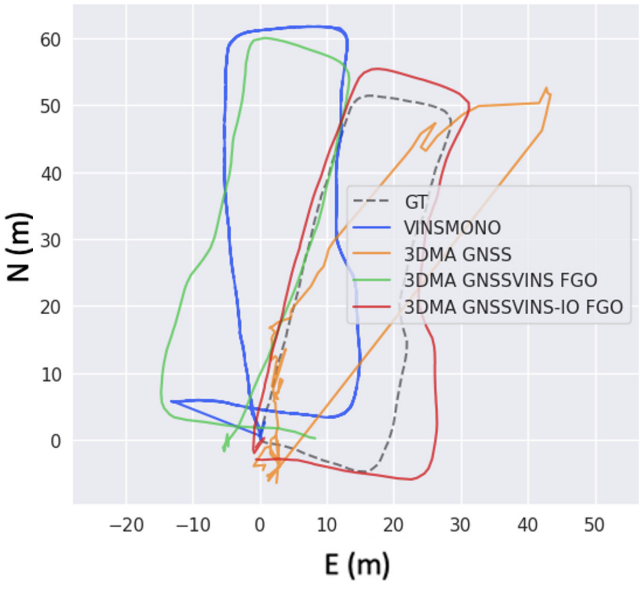


Figure 16
The C1 trajectory comparison

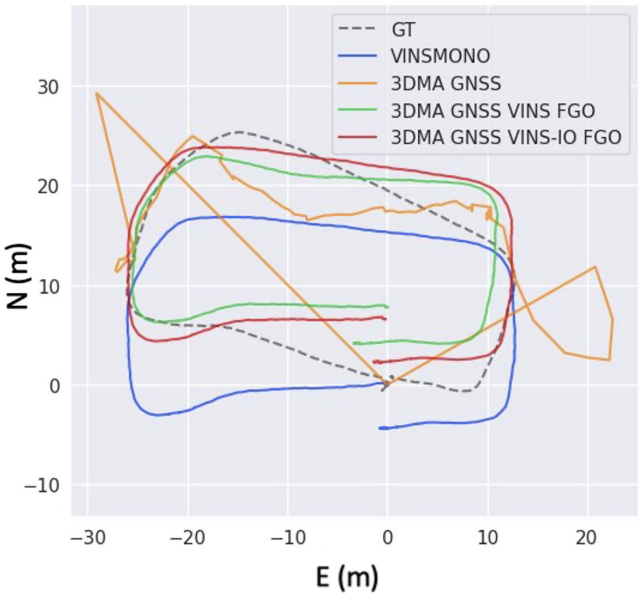


Table 3 presents the lowest std and RMSE from 3DMA GNSSVINS-IO FGO, with a value of 2.49 meters and 6.47 meters, respectively.

Furthermore, the analysis reveals that the 3DMA GNSSVINS-IO FGO algorithm significantly reduced a positioning error during an IO transition from 55s to 100s as shown in Figure 17. This integration successfully mitigated the error stemming from both VINS and GNSS, indicating that the IO switching factor had a profound impact on the smoothness and accuracy of the positioning results. By effectively managing the IO transition, the algorithm demonstrated improved performance and minimized the adverse effects of the error during that period.

Moreover, a substantial positioning error was observed between times 20s and 90s in Experiment C1. However, this error was effectively suppressed by integrating the GNSS and VINS systems. Notedly, the trajectory of the 3DMA GNSSVINS-IO FGO affected by the 3DMA GNSS outliner deviates from the GT, as shown in Figure 18.

Figure 17
The C2 absolute positioning error comparison

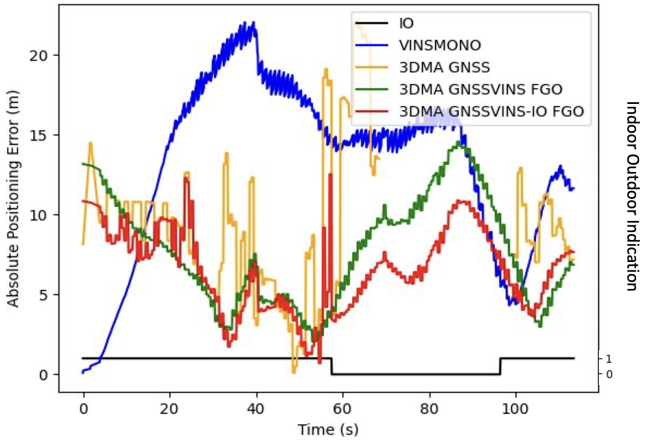
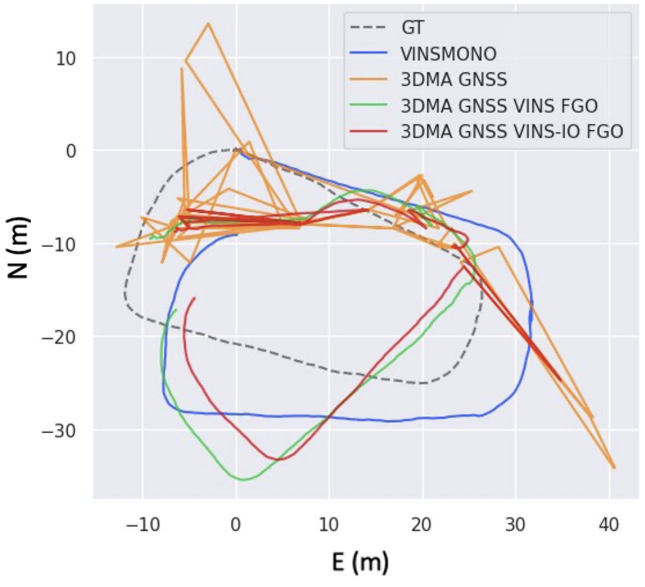


Figure 18
The C2 trajectory comparison



6. Conclusion

In this study, an FGO framework was developed to integrate local pose estimates from previous 3DMA GNSS and VIO research, with an IO detection switch. The proposed system tackles the challenges of global drift and enables accurate local pose estimation. Real-world experiments evaluated the system's performance, yielding promising results. The 3DMA GNSSVINS-IO FGO algorithm, incorporating the IO detection switch, demonstrates superior performance, with lower RMSE values and more consistent positioning accuracy. Combining 3DMA and VINS systems using FGO enhanced the accuracy and stability of the positioning results. By leveraging the complementary strengths of GNSS and VINS, the algorithm successfully mitigated the impact of the error and improved reliability and consistency positioning performance. Adding the IO detection switch significantly mitigated global drift challenges, improving performance with an average error of under 3 meters.

Real-life experiments validate the system's effectiveness, surpassing other methods. However, certain limitations exist. The training model is specific to the smartphone used in the experiment, and its applicability to other devices may be limited. Future work should **focus on enhancing** sensor **self-robustness, especially when** a camera or GNSS **cannot be used, allowing** the IMU propagation to take over further advancements **should target the complexities in** dynamic urban canyons and **better manage** environmental factors.

To address device specificity, future efforts should include cross-device training by using data from various devices, enhancing generalizability across different hardware configurations and manufacturers. This approach will create a larger model that is more robust and applicable to a wider range of devices. The goal is to achieve consistent performance and accuracy, regardless of device, thus extending the proposed system's usability and reliability.

Funding Support

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

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The source code that supports the findings of this study is openly available in 3DMAGNSSVINS-IOFGO at https://github.com/queenie-ho/3DMAGNSSVINS-IOFGO.

Author Contribution Statement

Hiu-Yi Ho and Hoi-Fung Ng: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Weisong Wen:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Yanlei Gu:** Resources, Writing – review & editing, Writing – review & editing. **Li-Ta Hsu:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

References

[1] Biswas, D., Barai, S., & Sau, B. (2023). New RSSI-fingerprinting-based smartphone localization system for indoor environments. *Wireless Networks*, *29*(3), 1281–1297. https://doi.org/10.1007/s11276-022-03188-2

[2] Boussad, Y., Mahfoudi, M. N., Legout, A., Lizzi, L., Ferrero, F., & Dabbous, W. (2021). Evaluating smartphone accuracy for RSSI measurements. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–12. https://doi.org/10.1109/TIM.2020.3048776

[3] Naser, R. S., Lam, M. C., Qamar, F., & Zaidan, B. B. (2023). Smartphone-based indoor localization systems: A systematic literature review. *Electronics*, *12*(8), 1814. https://doi.org/10.3390/electronics12081814

[4] Rajak, S., Panja, A. K., Chowdhury, C., & Neogy, S. (2021). A ubiquitous indoor–outdoor detection and localization framework for smartphone users. In A. E. Hassanien, S. Bhattacharyya, S. Chakrabati, A. Bhattacharya & S. Dutta (Eds.), *Emerging technologies in data mining and information security* (pp. 693–701). Springer. https://doi.org/10.1007/978-981-15-9927-9_67

[5] Zhong, Q., & Groves, P. D. (2022). Multi-epoch 3D-mapping-aided positioning using Bayesian filtering techniques. *NAVIGATION*, *69*(2). https://doi.org/10.33012/navi.515

[6] Ng, H. F., Hsu, L. T., Lee, M. J. L., Feng, J., Naeimi, T., Beheshti, M., & Rizzo, J. R. (2022). Real-time loosely coupled 3DMA GNSS/doppler measurements integration using a graph optimization and its performance assessments in urban canyons of New York. *Sensors*, *22*(17), 6533. https://doi.org/10.3390/s22176533

[7] Kumar, A., & Singh, A. K. (2020). A novel multipath mitigation technique for GNSS signals in urban scenarios. *IEEE Transactions on Vehicular Technology*, *69*(3), 2649–2658. https://doi.org/10.1109/TVT.2019.2962919

[8] Bai, Y. B., Holden, L., Kealy, A., Zaminpardaz, S., & Choy, S. (2022). A hybrid indoor/outdoor detection approach for smartphone-based seamless positioning. *The Journal of Navigation*, *75*(4), 946–965. https://doi.org/10.1017/S0373463322000194

[9] Huai, J., Toth, C. K., & Grejner-Brzezinska, D. A. (2015). Stereo-inertial odometry using nonlinear optimization. In *Proceedings of the 28th International Technical Meeting of the Satellite Division of The Institute of Navigation*, 2087–2097.

[10] He, Y., Zhao, J., Guo, Y., He, W., & Yuan, K. (2018). PL-VIO: Tightly-coupled monocular visual–inertial odometry using point and line features. *Sensors*, *18*(4), 1159. https://doi.org/10.3390/s18041159

[11] Nam, S., Kim, T., Oh, J., & Lee, S. (2024). Cascade-structured visual-inertial navigation system for aerial imaginary. In *Proceedings of the ION 2024 Pacific PNT Meeting*, 143–153. https://doi.org/10.33012/2024.19636

[12] Zhang, H., Chen, C. C., Vallery, H., & Barfoot, T. D. (2024). GNSS/Multi-sensor fusion using continuous-time factor graph optimization for robust localization. *IEEE Transactions on Robotics*, *40*, 4002–4023. https://doi.org/10.1109/TRO.2024.3443699

[13] Li, X., Yu, H., Wang, X., Li, S., Zhou, Y., & Chang, H. (2023). FGO-GIL: Factor graph optimization-based GNSS RTK/INS/LiDAR tightly coupled integration for precise and continuous navigation. *IEEE Sensors Journal*, *23*(13), 14534–14548. https://doi.org/10.1109/JSEN.2023.3278723

[14] Zhu, J., Zhuo, G., Xia, X., Wen, W., Xiong, L., Leng, B., & Liu, W. (2024). FGO-MFI: Factor graph optimization-based multi-sensor fusion and integration for reliable localization. *Measurement Science and Technology*, *35*(8), 086303. https://doi.org/10.1088/1361-6501/ad440a

[15] Jiang, C., Chen, Y., Chen, C., Jia, J., Sun, H., Wang, T., & Hyyppä, J. (2022). Smartphone PDR/GNSS integration via factor graph optimization for pedestrian navigation. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–12. https://doi.org/10.1109/TIM.2022.3186082

[16] Ho, H. Y., Ng, H. F., Leung, Y. T., Wen, W., Hsu, L. T., & Luo, Y. (2023). Smartphone level indoor/outdoor ubiquitous pedestrian positioning 3DMA GNSS/VINS integration using FGO. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *48*, 175–182. https://doi.org/10.5194/isprs-archives-XLVIII-1-W1-2023-175-2023

[17] Cao, S., Lu, X., & Shen, S. (2022). GVINS: Tightly coupled GNSS–visual–inertial fusion for smooth and consistent state estimation. *IEEE Transactions on Robotics*, *38*(4), 2004–2021. https://doi.org/10.1109/TRO.2021.3133730

[18] Chen, C., Chai, W., Zhang, Y., & Roth, H. (2014). A RGB and D vision aided multi-sensor system for indoor mobile robot and pedestrian seamless navigation. In *2014 IEEE/ION Position, Location and Navigation Symposium – PLANS 2014*, 1020–1025. https://doi.org/10.1109/PLANS.2014.6851469

[19] Falco, G., Pini, M., & Marucco, G. (2017). Loose and tight GNSS/INS Tntegrations: Comparison of performance assessed in real urban scenarios. *Sensors*, *17*(2), 255. https://doi.org/10.3390/s17020255

[20] Li, Y., Chen, W., Wang, J., & Nie, X. (2023). Precise indoor and outdoor altitude estimation based on smartphone. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1–11. https://doi.org/10.1109/TIM.2023.3315391

[21] Liu, Q., Gao, C., Shang, R., Gao, W., & Zhang, R. (2023). Multifrequency signal characterization and assessment of GNSS positioning in smartphone. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1–9. https://doi.org/10.1109/TIM.2023.3306519

[22] Bai, L., Ciravegna, F., Bond, R., & Mulvenna, M. (2020). A low cost indoor positioning system using bluetooth low energy. *IEEE Access*, *8*, 136858–136871. https://doi.org/10.1109/ACCESS.2020.3012342

[23] Guo, G., Chen, R., Ye, F., Liu, Z., Xu, S., Huang, L., ..., & Qian, L. (2022). A robust integration platform of Wi-Fi RTT, RSS signal, and MEMS-IMU for locating commercial smartphone indoors. *IEEE Internet of Things Journal*, *9*(17), 16322–16331. https://doi.org/10.1109/JIOT.2022.3150958

[24] Rehman, A., Shahid, H., Afzal, M. A., & Bhatti, H. M. A. (2020). Accurate and direct GNSS/PDR integration using extended Kalman filter for pedestrian smartphone navigation. *Gyroscopy and Navigation*, *11*(2), 124–137. https://doi.org/10.1134/S2075108720020054

[25] Huang, L., Li, H., Yu, B., Gan, X., Wang, B., Li, Y., & Zhu, R. (2020). Combination of smartphone MEMS sensors and environmental prior information for pedestrian indoor positioning. *Sensors*, *20*(8), 2263. https://doi.org/10.3390/s20082263

[26] Khedr, M., & El-Sheimy, N. (2021). S-PDR: SBAUPT-based pedestrian dead reckoning algorithm for free-moving handheld devices. *Geomatics*, *1*(2), 148–176. https://doi.org/10.3390/geomatics1020010

[27] Pratama, A. R., Widyawan, & Hidayat, R. (2012). Smartphone-based pedestrian dead reckoning as an indoor positioning system. In *2012 International Conference on System Engineering and Technology*, 1–6. https://doi.org/10.1109/ICSEngT.2012.6339316

[28] Tang, H., Niu, X., Zhang, T., Wang, L., & Liu, J. (2023). LE-VINS: A robust solid-state-LiDAR-enhanced visual-inertial navigation system for low-speed robots. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1–13. https://doi.org/10.1109/TIM.2023.3260279

[29] Debeunne, C., & Vivet, D. (2020). A review of visual-LiDAR fusion based simultaneous localization and mapping. *Sensors*, *20*(7), 2068. https://doi.org/10.3390/s20072068

[30] Lee, W., Eckenhoff, K., Geneva, P., & Huang, G. (2020). Intermittent GPS-aided VIO: Online initialization and calibration. In *2020 IEEE International Conference on Robotics and Automation*, 5724–5731. https://doi.org/10.1109/ICRA40945.2020.9197029

[31] Chen, D., Wang, N., Xu, R., Xie, W., Bao, H., & Zhang, G. (2021). RNIN-VIO: Robust neural inertial navigation aided visual-inertial odometry in challenging scenes. In *2021 IEEE International Symposium on Mixed and Augmented Reality*, 275–283. https://doi.org/10.1109/ISMAR52148.2021.00043

[32] van Diggelen, F. (2021). *End game for urban GNSS: Google's use of 3D building models*. Retrieved from: https://insidegnss.com/end-game-for-urban-gnss-googles-use-of-3d-building-models/

[33] Hsu, L.-T., Gu, Y., & Kamijo, S. (2016). 3D building model-based pedestrian positioning method using GPS/GLONASS/QZSS and its reliability calculation. *GPS Solutions*, *20*(3), 413–428. https://doi.org/10.1007/s10291-015-0451-7

[34] Wang, L., Groves, P. D., & Ziebart, M. K. (2015). Smartphone shadow matching for better cross-street GNSS positioning in urban environments. *The Journal of Navigation*, *68*(3), 411–433. https://doi.org/10.1017/S0373463314000836

[35] Ng, H. F., Zhang, G., & Hsu, L. T. (2020). A computation effective range-based 3D mapping aided GNSS with NLOS correction method. *The Journal of Navigation*, *73*(6), 1202–1222. https://doi.org/10.1017/S037346332000003X

[36] Lau, L., & Cross, P. (2007). Development and testing of a new ray-tracing approach to GNSS carrier-phase multipath modelling. *Journal of Geodesy*, *81*(11), 713–732. https://doi.org/10.1007/s00190-007-0139-z

[37] Ziedan, N. I. (2017). Urban positioning accuracy enhancement utilizing 3D buildings model and accelerated ray tracing algorithm. In *Proceedings of the 30th International Technical Meeting of The Satellite Division of the Institute of Navigation*, 3253–3268. https://doi.org/10.33012/2017.15366

[38] Zhang, T., Liu, S., Chen, Q., Feng, X., & Niu, X. (2022). Carrier-phase-based initial heading alignment for land vehicular MEMS GNSS/INS navigation system. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–13. https://doi.org/10.1109/TIM.2022.3208646

[39] Wang, X., Li, X., Chang, H., Li, S., Shen, Z., & Zhou, Y. (2023). GIVE: A tightly coupled RTK-inertial–visual state estimator for robust and precise positioning. *IEEE Transactions on Instrumentation and Measurement*, *72*, 1–15. https://doi.org/10.1109/TIM.2023.3282296

[40] Forster, C., Carlone, L., Dellaert, F., & Scaramuzza, D. (2017). On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics*, *33*(1), 1–21. https://doi.org/10.1109/TRO.2016.2597321

[41] Zhang, Z., & Scaramuzza, D. (2018). A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 7244–7251. https://doi.org/10.1109/IROS.2018.8593941

[42] Han, L., Lin, Y., Du, G., & Lian, S. (2019). Deepvio: Self-supervised deep learning of monocular visual inertial odometry using 3D geometric constraints. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 6906–6913. https://doi.org/10.1109/IROS40897.2019.8968467

[43] Bailey, T., Nieto, J., Guivant, J., Stevens, M., & Nebot, E. (2006). Consistency of the EKF-SLAM algorithm. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3562–3568. https://doi.org/10.1109/IROS.2006.281644

[44] Kaczmarek, A., Rohm, W., Klingbeil, L., & Tchórzewski, J. (2022). Experimental 2D extended Kalman filter sensor fusion for low-cost GNSS/IMU/Odometers precise positioning system. *Measurement*, *193*, 110963. https://doi.org/10.1016/j.measurement.2022.110963

[45] Yin, Y., Zhang, J., Guo, M., Ning, X., Wang, Y., & Lu, J. (2023). Sensor fusion of GNSS and IMU data for robust localization via smoothed error state Kalman filter. *Sensors*, *23*(7), 3676. https://doi.org/10.3390/s23073676

[46] Hu, Y., Li, H., & Liu, W. (2024). Robust factor graph optimisation method for shipborne GNSS/INS integrated navigation system. *IET Radar, Sonar & Navigation*, *18*(5), 782–798. https://doi.org/10.1049/rsn2.12521

[47] Zhang, H., Xia, X., Nitsch, M., & Abel, D. (2022). Continuous-time factor graph optimization for trajectory smoothness of GNSS/INS navigation in temporarily GNSS-denied environments. *IEEE Robotics and Automation Letters*, *7*(4), 9115–9122. https://doi.org/10.1109/LRA.2022.3189824

[48] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

[49] Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and*

Machine Intelligence, *40*(6), 1452–1464. https://doi.org/10.1109/TPAMI.2017.2723009

[50] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., . . . , & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 1–74. https://doi.org/10.1186/s40537-021-00444-8

[51] Groves, P., & Jiang, Z. (2013). Height aiding, C/N0 weighting and consistency checking for GNSS NLOS and multipath mitigation in urban areas. *The Journal of Navigation*, *66*(5), 653–669. https://doi.org/10.1017/S0373463313000350

[52] Takasu, T., & Yasuda, A. (2009). Development of the low-cost RTK-GPS receiver with an open source program package RTKLIB. In *International Symposium on GPS/GNSS*, *1*, 1–6.

[53] Li, W., Cui, X., & Lu, M. (2018). A robust graph optimization realization of tightly coupled GNSS/INS integrated navigation system for urban vehicles. *Tsinghua Science and Technology*, *23*(6), 724–732. https://doi.org/10.26599/TST.2018.9010078

[54] Qin, T., Li, P., & Shen, S. (2018). VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, *34*(4), 1004–1020. https://doi.org/10.1109/TRO.2018.2853729

[55] Chen, Q., Lin, H., Guo, R., & Niu, X. (2020). Rapid and accurate initial alignment of the low-cost MEMS IMU chip dedicated for tilted RTK receiver. *GPS Solutions*, *24*(4), 1–13. https://doi.org/10.1007/s10291-020-01032-8

[56] Shi, J., & Tomasi. (1994). Good features to track. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 593–600. https://doi.org/10.1109/CVPR.1994.323794

[57] Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *IJCAI'81: 7th International Joint Conference on Artificial Intelligence*, *2*, 674–679.

[58] Heng, L., Li, B., & Pollefeys, M. (2013). Camodocal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1793–1800. https://doi.org/10.1109/IROS.2013.6696592

[59] Agarwal, S., Mierle, K., & Ceres Solver Team. (2023). Ceres Solver (2.2). Retrieved from: https://github.com/ceres-solver/ceres-solver

How to Cite: Ho, H.-Y., Ng, H.-F., Wen, W., Gu, Y., & Hsu, L.-T. (2024). Enhancing Smartphone-based Pedestrian Positioning: Using Factor Graph Optimization with Indoor/Outdoor Detection for 3DMA GNSS/Visual-inertial State Estimation. *Journal of Data Science and Intelligent Systems*. https://doi.org/10.47852/bonviewJDSIS42022961