

## SMARTPHONE LEVEL INDOOR/OUTDOOR UBIQUITOUS PEDESTRIAN POSITIONING 3DMA GNSS/VINS INTEGRATION USING FGO

Hui-Yi Ho<sup>1</sup>, Hoi-Fung Ng<sup>1</sup>, Yan-Tung Leung<sup>1</sup>, Weisong Wen<sup>1</sup>, Li-Ta Hsu<sup>1</sup>, Yiran Luo<sup>2</sup>

<sup>1</sup> Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University – (queenie-hy.ho, hf-ivan.ng, nikkie-yt.leung)@connect.polyu.hk, (welson.wen, lt.hsu)@polyu.edu.hk

<sup>2</sup> Department of Geomatics Engineering, University of Calgary – yiran.luo@ucalgary.ca

**KEY WORDS:** Smartphone, IO, Ubiquitous, Pedestrian Positioning, Sensor Integration, VINS, 3DMA GNSS, FGO

### ABSTRACT:

This paper discusses ubiquitous smartphone pedestrian positioning challenges in urban canyons and GNSS-denied areas such as indoor spaces. Existing sensor-based techniques, including GNSS, INS, and VIO, have limitations that affect positioning accuracy and reliability. A machine learning-based approach is suggested to employ Support Vector Machine (SVM) to classify indoor/outdoor (IO) detection using GNSS measurement data. The proposed system integrates local estimates on VIO and 3D mapping aided (3DMA) GNSS measurements using Factor Graph Optimization (FGO) with an IO detection switch to estimate precise pose and eliminate global drift. The effectiveness of the system is evaluated through real-world experiments that produce notable outcomes.

### 1. INTRODUCTION

**Smartphone positioning is challenging in urban canyons:** Smart mobility has recently been trendy with various sensors, cutting-edge intelligence, and next-generation networks. Among these intelligent techniques, a vast variety of mobile positioning has been proposed over time to support the Location-Based Service (LBS). Smartphones are fast becoming a key instrument in mobile positioning. It has several sensors, for instance, Wi-Fi, an inertial sensor, a magnetometer, a monocular camera, etc. (Sheta et al., 2018). The positioning system can be configured into different combinations of sensors to provide reliable localization information and achieve all-rounded positioning. The challenge of mobile Global Navigation Satellite System (GNSS), positioning is sensor perception in urban areas. Rajak (Rajak et al., 2021a) and Wen (Wen et al., 2021a)(Wen et al., 2019) have validated the unsatisfactory performance of GNSS positioning in the urban environment with conventional positioning methods. Poor positioning accuracy highly affects user experiences, especially for smartphone users. Extra information can be provided to aid the positioning in urban canyons. 3D building models provide the perception of the actual environment as a software-based aided positioning approach for the low-cost, namely 3D mapping aided (3DMA) GNSS (Groves, 2016). Existing studies by (Zhong & Groves, 2022) and (Ng et al., 2021) show a superior positioning performance of 3DMA GNSS in urban canyons. Doppler measurements are often used to integrate with the position solution to provide smoother positioning results. (Ng et al., 2022) integrates 3DMA GNSS and velocity estimated by Doppler frequency as a loosely coupled using Factor Graph Optimization (FGO). This study is also going to adopt the loosely coupled 3DMA GNSS and Doppler velocity to provide more accurate positioning in a global frame.

**GNSS suffering from outages in indoor scenarios:** Augmented with different sensors as a fused solution can improve the performance of the GNSS in urban cities. When the operating environment affects the GNSS, such as the solution outages due to insufficiently visible satellites to resolve the position. The GNSS positioning degrades in the indoor environment because of weak attenuated and scattered signals received by numerous objects in the indoor environment. For example, (Rajak et al., 2021a) found that the signal strength of GPS decreased by 10–12 decibels leading to a rapid reduction in positioning accuracy. In

addition, the scale degeneracy in rotation-only or constant velocity motions happens since the lack of direct distance measurements (Zhu et al., 2019). Hence, IO (Indoor/Outdoor) detection is the key to ubiquitous positioning. To accomplish the ubiquitous indoor localization framework, many researchers conducted IO detection using GPS measurement (Rajak et al., 2021a) (Pei et al., 2009) (Wang et al., 2020). The framework should not require any extra infrastructure but use the existing built-in sensors in the smartphone. The recent developments in IO detection have the potential to assist positioning.

**Visual/inertial odometry can help but is subject to drift over time:** To assist the indoor positioning, visual-inertial odometry is fully employed in the GNSS-denied area for bridging the GNSS gaps (Huai et al., 2015). Although Visual-Inertial Odometry (VIO) is hard to implement within the speed and latency constraints, all stages are refined with a nonlinear optimization (He et al., 2018). In accordance with the different sensor assessments, the effectiveness varies with respect to indoor or outdoor environments. Consequently, the attenuated and scattered signal received will not be adopted to access the indoor position. GNSS, Inertial Measurement Unit (IMU), and the monocular camera are utilized to take advantage of their forte. A loosely coupled GNSS/VINS (Visual Inertial Navigation System) integration is generated in the paper.

**Adaptive 3DMA GNSS/VINS integration is promising:** The multi-sensor integration approach applied for the 3DMA GNSS/VINS is FGO in a loosely coupled way. It is nonlinear optimization represented in probabilistic graphical models. After factorization, it transformed into a factor graph to simulate the relationship between the poses and estimate their value. It satisfies various changes in the dynamic environment (Chen et al., 2016). Moreover, FGO has been demonstrated to use most of the feature constrain to get the optimal trajectory, resulting in higher accuracy and efficiency to achieve a robust estimation (Rajak et al., 2021b).

**The proposed method in this paper:** In this research, smartphone-level ubiquitous mobile pedestrian positioning is proposed. The first objective of the research is to use a machine learning-based method Support Vector Machine (SVM) to classify and detect IO transition by the GNSS measurement. As indoor environments cannot receive satellite signals, the GNSS positioning performance will be degraded going into the indoor

environment from the outdoor environment when achieving seamless positioning. The IO detection can assist in the system switching from GNSS/VINS to VINS. The second objective is the GNSS/VINS alignment. The frame between the global position and local poses must be aligned in the same coordinate system. Through the continuous dynamic motion of the smartphone, the heading of the device can be derived through the accelerations from the accelerometer with the position and velocity measurements from the GNSS receiver. The third objective is the integration of the loosely coupled GNSS/VINS using FGO. FGO constrained all factors regarding GNSS/VINS performance, and IO detection results to smooth indoor and outdoor transitions. It is a cost-deducted system acquiring a precise loosely coupled positioning in a dynamic and complex high-density environment with a monocular camera, Micro Electro Mechanical Systems (MEMS) IMU, and GNSS for performance evaluation. This can enhance the whole accuracy into a sub-meter with cost reduction and accessibility.

This paper is organized as follows: In Section 2, relevant literature will be discussed. In Section 3, a system overview, SVM classification, 3DMA GNSS/VINS, and FGO will be presented. In Section 4, the experiment result of the 3DMA GNSS/VINS will be shown. Finally, the paper will be concluded in Section 5.

## 2. RELATED WORK

### 2.1 Introduction to GNSS/VINS integration for pedestrian positioning

Pedestrian positioning in urban and indoor environments has been a topic of research interest, with GNSS and VINS being two common positioning technologies. The complementary nature of these technologies makes them suitable for integration to provide precise and continuous positioning. However, integrating these technologies faces challenges due to measurement errors, such as multipath, signal obstructions, and drift.

**Improving GNSS performance in challenging areas:** GNSS is widely applied to provide continuous positioning in the global frame in absolute coordinates. However, the performance is usually unsatisfactory due to the blockage or reflection of the signals over the buildings, resulting in the non-line-of-sight (NLOS) reception and multipath effect (Groves, 2013). Improving the GNSS positioning can benefit the whole positioning system, and researchers are trying to mitigate the NLOS error to improve the GNSS performance alone. Recent research has explored integrating VINS for pedestrian positioning using loosely and tightly coupled integration approaches, combining measurements from GNSS, INS, and visual sensors for accurate pose estimation in GPS-denied environments. Evaluation of this approach in real urban scenarios has confirmed its ability to improve positioning accuracy compared to the standalone GNSS or INS systems (Falco et al., 2017). Despite the potential benefits, GNSS/VINS integration still faces several challenges and limitations, including dealing with signal loss, reducing computational complexity, and improving robustness in various environmental conditions. One significant challenge of GNSS/VINS integration is the occurrence of multipath errors.

### 2.2 Smartphone-level pedestrian positioning: challenges and opportunities

Smartphone-level pedestrian positioning is essential for LBS. However, providing accurate positioning in real-time poses challenges due to smartphones' limited processing power, memory, and battery life. State-of-the-art positioning techniques, such as GNSS and VIO, have been used in smartphone-level pedestrian positioning. Recently, various techniques have been proposed to improve pedestrian positioning accuracy using smartphones. One such technique is Pedestrian Dead Reckoning (PDR), using inertial sensors such as accelerometers and gyroscopes. An efficient PDR algorithm was developed using a low-cost MEMS IMU for smartphones and showed that it is low-cost, simple, and easy to use compared to other methods (Jimenez et al., 2009). Even though the most significant advantage of PDR is an infrastructure-independent (Pratama et al., 2012), it also has some limitations, such as low accuracy and drift.

Integrating VIO with other sensors has been proposed to address these challenges and achieve more precise localization. One such approach is to integrate VIO with Light Detection and Ranging (LiDAR) using Simultaneous Localization and Mapping (SLAM) algorithms (Debeunne & Vivet, 2020). SLAM algorithms use visual and LiDAR data to generate a more accurate and detailed map of the surroundings, improving accuracy, robustness, and efficiency. LiDAR-based pedestrian tracking is also a promising technique due to its high accuracy and robustness in various environments. However, this approach is limited by high computational complexity and cost.

One popular approach to improve GNSS positioning in urban canyons is using a 3D building model to identify and correct the NLOS reception error, namely 3DMA GNSS (Groves, 2016). 3DMA GNSS usually demonstrate as a particle-based approach. Measurements are modeled as the prediction at each distributed position hypothesis candidate. The candidate with the highest similarity between modeled and actual received measurements is assumed to be the receiver location. 3DMA GNSS can commonly divide into shadow-matching and ranging-based 3DMA GNSS. Shadow matching (Wang et al., 2013)(Wang et al., 2015) matches the satellite visibility across distributed locations. Meanwhile, ranging-based 3DMA GNSS provides reflection delays for the NLOS-predicted pseudo-range modelling. The delay estimation can be done in a geometrical approach, such as ray-tracing GNSS (Hsu et al., 2016)(Miura et al., 2015) and Skymask 3DMA(Ng et al., 2020). Both approaches validate the signal transmission path and calculate the reflection delay with the predicted reflecting point. The likelihood-based ranging (Zhong & Groves, 2022) statistically uses a skew-normal distribution to model the NLOS delay measurements. Then it remaps the errors to the LOS one with the normal distribution. Extending the single epoch positioning approach to temporal connected can increase the robustness of the positioning performance. (Zhong & Groves, 2022) adopts a grid filter to distribute positioning candidates evenly to improve the smoothness of the solution. An alternative way is using the FGO to connect the temporal domain as a batch optimization to increase the overall robustness and smoothness. (Ng et al., 2022) integrates 3DMA GNSS with velocity estimated by Doppler measurements as a loosely coupled solution, and states are optimized via FGO. The results show that it can provide a more robust trajectory for pedestrian applications.

**VINS positioning and navigation:** VINS are integrated navigation systems that combine visual and inertial sensors to estimate a platform's position and orientation. Despite the promising results of VINS in various applications, they face several challenges, including time drift and accumulated error, particularly in complex environments. Two residual errors

associated with visual and inertial measurements include the discrepancy between the prediction by the IMU and the position estimated by visual odometry. Different algorithms have been evaluated to address these issues, such as keyframes-based methods, inertial measurements preintegration, nonlinear optimization methods, and machine learning-based methods (Leutenegger et al., 2015)(Zhang & Scaramuzza, 2018) (Forster et al., 2017) (Han et al., 2019). For instance, preintegrating IMU measurements and feature observations can obtain high accuracy and efficiency by preintegrating inertial measurements between selected keyframes into single relative motion constraints and using keyframes to reduce computational complexity and improve accuracy (Forster et al., 2017). While the nonlinear optimization method can attain a highly accurate state estimation (Zhang & Scaramuzza, 2018), real-time optimization quickly becomes infeasible as the trajectory grows over time as the inertial measurements come at a high rate (Forster et al., 2017). Deepvio has explored the potential applications of VIO in various fields and identified future research directions, such as deep learning techniques to improve performance (Han et al., 2019). Nonetheless, time drifting is the primary uncertainty when using VINS.

### 2.3 Review of existing GNSS/VINS integration approaches for pedestrian positioning

The Extended Kalman Filter (EKF) is an effective method for sensor fusion, but it relies on a linear approximation of system dynamics, leading to reduced accuracy in nonlinear systems. Furthermore, the EKF's computational complexity scales quadratically with the number of 3D landmarks, limiting its scalability (He et al., 2018). In dense urban areas, EKF fails to achieve optimal performance due to the accumulation of Gaussian errors (Wen et al., 2021b). Nonlinear optimization methods have been proposed to address this issue. FGO is another approach for pedestrian positioning that models relationships between observed measurements and unknown states of the system, leading to high accuracy and efficiency. FGO handles noisy or incomplete data better than EKF and is more equipped to handle changing dynamics. EKF, on the other hand, struggles with nonlinearities in system dynamics. However, FGO's computational requirements may be higher, and accurate modeling of system dynamics is necessary for optimal performance. Ultimately, the choice of approach depends on the specific application and environmental conditions. To mitigate challenges associated with integrating VINS and GNSS, including errors in VINS measurements (e.g., drift) and errors in GNSS measurements (e.g., multipath and signal obstructions), the integration of 3DMA GNSS and VINS using FGO can provide accuracy, robustness, and reliability enhancement in complex and challenging environments.

### 2.4 Development of a framework for Integrating 3DMA GNSS and VINS using FGO for Pedestrian Positioning

In summary, this research is to develop a framework for smartphone-level pedestrian positioning technologies to improve accuracy, robustness, and efficiency in urban and indoor environments. Several challenges such as signal obstructions, drift, and multipath limit the effectiveness of the existing techniques. To address these challenges, the proposed framework suggests integrating 3DMA GNSS and VINS using FGO with accuracy, robustness, and reliability enhancement in complex and challenging environments. The framework emphasizes the importance of IO detection, using a machine learning-based method SVM to distinguish IO, and increasing the robustness of the positioning performance.

## 3. RESEARCH METHODOLOGY

### 3.1 Overview

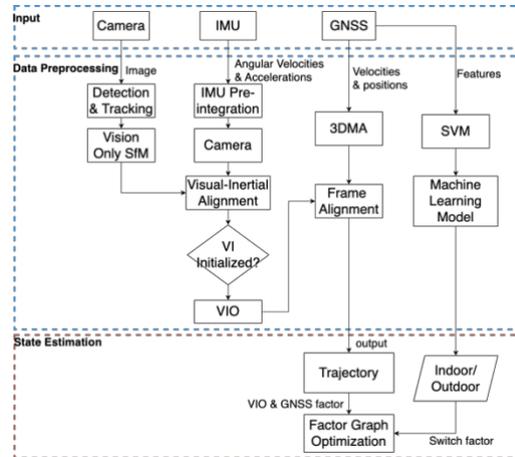


Figure. 1 Flowchart of the proposed system GNSSVINS-IO FGO

The positioning framework proposed in this paper is shown in Figure. 1. This paper demonstrates a machine learning-based method using the SVM to classify IO. Features from GNSS measurements, such as the received and used satellite numbers and elevation angle, are extracted for IO classification. Improving the GNSS positioning accuracy is the key to satisfactory performance under global coordinates. This study integrates the state-of-the-art 3DMA GNSS algorithms on shadow matching and likelihood-based ranging 3DMA GNSS. The 3DMA GNSS solution is then optimized with Doppler measurements using FGO in a loosely coupled fashion to increase the robustness. The full implementation can be found in 3DMA GNSS distributes the hypothesis positioning candidates around the initial position. Measurements are then simulated as each candidate and compared with the received measurements. And this study integrates shadow matching and likelihood-based ranging 3DMA GNSS, as  $\mathbf{x}_{t,3DMA}$ , for FGO integration.

Meanwhile, receiver velocity,  $\mathbf{v}_t$ , and clock drift,  $c\delta t_t$ , is estimated by the Doppler measurements of every satellite  $i$  at epoch  $t$ ,  $d_t = [d_t^1, \dots, d_t^i]$ , via the least-squares (LS) method (Wen & Hsu, 2021). The 3DMA GNSS also integrates with the Doppler measurement estimated velocity using FGO to increase the position robustness. The FGO structure consists of three factors. The first error factor constrains the 3DMA GNSS solution,  $\mathbf{x}_{t,3DMA}$ , and optimized state,  $\mathbf{x}_t$ ,

$$\|e_{t,3DMA}\|_{\sigma_{3DMA}^2}^2 = \|\mathbf{x}_t - \mathbf{x}_{t,3DMA}\|_{\sigma_{3DMA}^2}^2 \quad (1)$$

where  $\sigma_{3DMA}^2$  is a diagonal variance matrix of the 3DMA GNSS. While another two factors connect two consecutive epochs based on the motion propagation model,  $\|e_{t,v}\|_{\sigma_{v,t}^2}^2$ , and constant velocity motion model (Li et al., 2018),  $\|e_{t,v}\|_{\sigma_v^2}^2$ , given by,

$$\|e_{t,v}\|_{\sigma_{v,t}^2}^2 = \left\| \mathbf{v}_t - \frac{1}{\Delta t} (\mathbf{x}_{t+1} - \mathbf{x}_t) \right\|_{\sigma_{v,t}^2}^2 \quad (2)$$

$$\|e_{t,\bar{v}}\|_{\sigma_v^2}^2 = \left\| \frac{1}{2}(\mathbf{v}_t + \mathbf{v}_{t+1}) - \frac{1}{\Delta t}(\mathbf{x}_{t+1} - \mathbf{x}_t) \right\|_{\sigma_v^2}^2 \quad (3)$$

where  $\Delta t$  is the time difference between epoch  $t$  and  $t + 1$ .  $\sigma_v^2$  is a diagonal covariance matrix associated with the velocity  $\mathbf{v}_t$  at  $x$ -,  $y$ -, and  $z$ -axis, respectively.  $\sigma_v^2$  is the averaged diagonal covariance matrix at time  $t$  and  $t + 1$ . Therefore, FGO minimizes the total error of three cost functions of the loosely coupled 3DMA GNSS as,

$$\chi^* = \operatorname{argmin}_{\chi} \sum_k \|e_{t,3DMA}\|_{\sigma_{3DMA}^2}^2 + \|e_{t,v}\|_{\sigma_{v,t}^2}^2 + \|e_{t,\bar{v}}\|_{\sigma_v^2}^2 \quad (4)$$

where  $\chi = [\mathbf{x}_1, \dots, \mathbf{x}_t]$  is the state set of the receiver. And  $\chi^*$  denotes the optimal states set.

To assist with indoor positioning, visual-inertial odometry is fully employed in the GNSS-denied area for bridging the GNSS gaps (Leutenegger et al., 2015). As the input, the images, the angular velocities, and accelerations are pre-processed. A pose graph is defined by skipping frames, pre-integrate IMU between keyframes, and then representing most of the IMU measurements into a single pose constraint so that the IMU can be manageable. The IMU is pre-integrated by adding posterior IMU bias correction and alignment with visual measurement to generate VIO to provide times, positions, orientations, and velocities. It then loosely fuses with GNSS and performs global optimization. All stages are refined with a nonlinear optimization (Shen et al., 2015). Besides the visual and inertial factors, the effectiveness varies with respect to indoor or outdoor environments in accordance with the different sensor assessments. Consequently, the switch factor will also be applied when the attenuated and scattered signal is not received in the indoor position. A loosely coupled GNSS/VINS integration is then generated to achieve pedestrian positioning when GNSS, IMU, and the monocular camera are utilized to take advantage of their forte.

This research focuses on providing reliable sensor fusion twofold: 1) selecting the reliable sensor for integration; 2) providing complete robustness during integration. As a result, reliable smartphone-level ubiquitous mobile pedestrian positioning is proposed. There are three contributions to this study:

- Develop a machine learning-based IO method based on GNSS measurement as the feature to select the reliable sensor during integration to maximize the positioning performance.
- Coordinate frame alignment between INS in a local frame and GNSS in a global frame.
- Loosely integrating solutions on GNSS and VINS as a batch using FGO to provide complete robustness.

### 3.2 Indoor Outdoor Detection Approach

**IO detection:** Indoor refers to a physically confined area; Outdoor refers to a non-completely confined area. (Bai et al., 2022) (Yan et al., 2019)

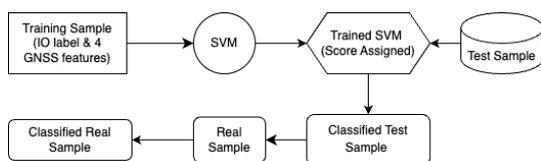


Figure. 2 Flowchart of the SVM

To classify IO, SVM is adopted by GNSS features, including the received and used satellite numbers, the received average carrier-

to-noise ratio (C/N0), and the elevation angle are extracted for IO classification. It is a supervised learning approach that transforms into a binary result. The data was collected from a university campus, indoor/outdoor, fully implementable on the phone. Figure. 2 illustrates the algorithm framework of the proposed IO detection. The algorithm relies on provided GNSS information about pre-existing relationships among a subset of the elements, such as the received and used satellite numbers, the received average C/N0, and the elevation angle that need to be categorized as Indoor or Outdoor. The binary classification result is based on the initial assumption of the relationships between the elements as the expression pattern data, which is either belonging to Indoor or not belonging to Indoor.

The whole framework consists of two primary phases. The training phase is the first phase to utilize the presumptive classification (IO) and the expression data (GNSS measurements) as inputs to generate a set of weights. These weights are then utilized in the subsequent phase. The presumptive classification of IO was done manually for ground truth information. The second phase is the classification phase to utilize the weights from the training phase and the expression data, which are used to assign a score to each element. Each element is classified into the relevant category or excluded depending on the score.

### 3.3 GNSS/VINS estimator

#### 3.3.1 VINS

After detection and tracking of the camera image feature and matching these features to the next image, a visual trajectory is constructed from SfM. It is required to align with IMU to recover the scale, velocity, gravitational acceleration, and IMU deviation (Zhu et al., 2019). However, every IMU rotation in the world coordinate system is required when the speed is updating; those IMU speeds and rotations in the world coordinate system are required when the translation is updating. Knowing that the IMU is in hundreds of Hz, it is unreasonable to update all the states every time. Therefore, the pose graph is defined by skipping frames, pre-integrate IMU between keyframes, and then representing most of the IMU measurements into a single pose constraint.

**IMU pre-integrations:** utilizing the continuous-time quaternion-based approach as previously developed in (Qin et al., 2018), with the inclusion of IMU biases as described in (Qin et al., 2019)(Qin & Shen, 2018). To avoid redundancy, this section provides a concise overview of our method in this section.

IMU measurements, which are measured in the body frame, combine the force for countering gravity and the platform dynamics, and are affected by acceleration bias  $\mathbf{b}_a$ , gyroscope bias  $\mathbf{b}_w$ , and additive noise. It is postulated that the additional noise present in the measurements obtained from the acceleration and gyroscope instruments follows a Gaussian white noise distribution. Moreover, the biases associated with the measurements obtained from these instruments are represented as random walks, whose derivatives are modeled as Gaussian white noise. The raw gyroscope and accelerometer measurements, denoted as  $\hat{\boldsymbol{\omega}}$  and  $\hat{\mathbf{a}}$ , respectively, can be expressed as functions of

$$\begin{aligned} \hat{\mathbf{a}}_t &= \mathbf{a}_t + \mathbf{b}_{a_t} + \mathbf{R}_w^t \mathbf{g}^w + \mathbf{n}_a \\ \hat{\boldsymbol{\omega}}_t &= \boldsymbol{\omega}_t + \mathbf{b}_{w_t} + \mathbf{n}_w. \end{aligned} \quad (5)$$

This study assumes the acceleration and gyroscope measurements error follows a Gaussian white noise distribution. We use a random walk process to model the acceleration and

gyroscope biases, whose derivatives are also Gaussian white noise. To propagate the covariance of the orientation angles  $\alpha$ ,  $\beta$ , and  $\gamma$ , we use equations

$$\mathbf{a}_{b_{k+1}}^{b_k} = \iint_{t \in [t_k, t_{k+1}]} \mathbf{R}_t^{b_k} (\hat{\mathbf{a}}_t - \mathbf{b}_{a_t}) dt^2 \quad (6)$$

$$\boldsymbol{\beta}_{b_{k+1}}^{b_k} = \int_{t \in [t_k, t_{k+1}]} \mathbf{R}_t^{b_k} (\hat{\mathbf{a}}_t - \mathbf{b}_{a_t}) dt. \quad (7)$$

$$\boldsymbol{\gamma}_{b_{k+1}}^{b_k} = \int_{t \in [t_k, t_{k+1}]} \frac{1}{2} \boldsymbol{\Omega} (\hat{\boldsymbol{\omega}}_t - \mathbf{b}_{\omega_t}) \boldsymbol{\gamma}_t^{b_k} dt \quad (8)$$

The Equations 6, 7, and 8 are computed using only the IMU measurements within the time span  $[t_k, t_{k+1}]$  between two consecutive frames  $b_k, b_{k+1}$ , where  $b_k$  is the reference frame given the bias.

### 3.3.2 GNSS VINS Alignment

**Yaw offset:** Three-dimensional position coordinates and quaternion are outputted by the VINS with the acceleration of gravity of IMU, the local frame can be aligned with the ENU coordinate system using two degrees of freedom. To recover the 4-DoF local and global frames transformation, we aim to determine the yaw offset between the ENU frame and the local world frame.

The alignment between VINS and GNSS uses the VIO's velocity,  $\mathbf{v}_{VIO}$ , and relative position of consecutive epochs from GNSS, and minimizes the cost function,

$$\operatorname{argmin}_{\psi} \left\| \frac{1}{\Delta t} (\mathbf{x}_{t+1} - \mathbf{x}_t) - \mathbf{R}_{ECEF}^{ENU} \mathbf{R}_{ENU}^{local}(\psi) \mathbf{v}_{VIO} \right\|^2 \quad (9)$$

Where  $\mathbf{R}_{ECEF}^{ENU}$  and  $\mathbf{R}_{ENU}^{local}(\psi)$  are the rotational matrix from ENU to ECEF, and local frame to ENU, respectively. After this step, the transformation between the ENU and local world frames is completely calibrated. Therefore, the final positions estimated are in a global frame. Although the yaw angle will change when every time the system restarts, the yaw offset estimation between the ENU and local world frames is needed.

### 3.4 Factor Graph Optimization

The multi-sensor integration approach applied for the GNSS/VINS is FGO in a loosely coupled way. It is non-linear optimization represented in probabilistic graphical models and transformed into a factor graph to simulate the relationship between the poses and estimate their value. It can improve the accuracy of the system and mitigate drift and error accumulation issues. After the global and local frame alignment, the result will be further optimized. The VINS estimates are represented as factors in the graph, while the GNSS measurements are incorporated as constraints to optimize the estimates.

**Formulation of the FGO problem for pedestrian positioning:**

$$\min_x \frac{1}{2} \sum_i p_i \left( \|f_i(x_i, \dots, x_{i_k})\|^2 \right) \quad \text{s.t.} \quad l_j \leq x_j \leq u_j \quad (10)$$

Equation. 10 shown above, is Non-linear Least Square where  $p_i$  is the loss function,  $f_i$  is the cost function. First, define the parameter block  $(x_i, \dots, x_{i_k})$  and residual block  $p_i \left( \|f_i(x_i, \dots, x_{i_k})\|^2 \right)$  to compute the residual value. Then, set up all parameter blocks and residual blocks to build the problem. The solver used here is Ceres solver. The structure is shown in Figure. 3.

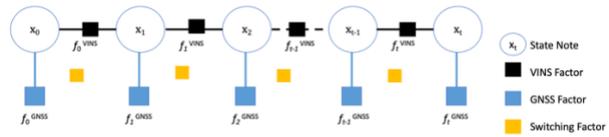


Figure. 3 System state illustration.

$$X = [x_0, x_1, x_2, \dots, x_t] \quad (11)$$

$$X_{0,t} = (p_{0,t}, v_{0,t}, q_{0,t})^T \quad (12)$$

where  $p$  is position,  $v$  is velocity, and  $q$  is orientation

$$\|f_t^{VINS}\|_{\Sigma_t}^2 = \left\| \begin{bmatrix} \mathbf{q}_{t-1}^{l-1} \mathbf{R}(\mathbf{p}_t^l - \mathbf{p}_{t-1}^l) \\ \mathbf{q}_{t-1}^{w-1} \mathbf{R}(\mathbf{p}_t^w - \mathbf{p}_{t-1}^w) \end{bmatrix} \right\|_{\Sigma_t}^2 \quad (13)$$

$$\|f_t^{GNSS}\| = \|\mathbf{p}_t^{GNSS} - \mathbf{p}_t\|_{\Sigma_t}^2 \quad (14)$$

$$\|f_t^{SW}\| = \left\| s_t f_{t,GNSS} + ((1-s_t) f_{t,VINS}) \right\|_{\Sigma_t}^2 \quad (15)$$

The system states will be estimated in a set  $X$  in Equation to initialize the system. 11,  $x_t$  is what we currently want to estimate. Each epoch we calculate is to estimate the position, velocity, and orientation. There are three factors, including the VINS factor in Equation. 13, the GNSS factor in Equation. 14, and switching factor in Equation. 15. And we factories all measurement.

$$X^* = \operatorname{argmin}_X \sum_t \|f_t^{VINS}\|_{\Sigma_t}^2 + \|f_t^{GNSS}\|_{\Sigma_t}^2 + \|f_t^{SW}\|_{\Sigma_t}^2 \quad (16)$$

Finally, the whole problem is organized with a series of factors formulated by the estimation from VINS, GNSS, and Switching I/O. The optimum system state would be stated as Equation. 16 after the estimation problem has been transformed and every cost corresponding to their specific measurement is minimized. To become optimal, the optimum system gives all measurements to be as the state at maximizes a posterior (MAP) (Cao et al., 2022). And it optimizes all states at the same time.

## 4. EXPERIMENT



Figure. 4 Trajectory A



Figure. 5 Trajectory B

In the experiment, there are two scenarios conducted in two places at Hong Kong Polytechnic University to assess the performance. The two scenarios, first scenario is starting from indoor, having a transition to outdoor and going back to indoor. Vice versa, the second scenario is starting outdoor, having a transition indoor and going back indoor. The trajectories A and B are shown in Figures 4, 5 respectively. This section compares

1. VINS-MONO: (Qin et al., 2018) is a monocular visual-inertial state estimator to obtain VIO
2. 3DMA GNSSVINS-IO FGO: Proposed method using the monocular camera, 3DMA GNSS, IMU and FGO
3. 3DMA GNSSVINS FGO: Proposed method without using the IO switching factor in FGO
4. 3DMA GNSS

#### 4.1 Experiment Setup



Figure. 6 Xiaomi Mi8

We used the Xiaomi Mi8 smartphone in Figure. 6 to obtain the IMU, GNSS, and Image data in this experiment. There is a Triple-axis MEMS-IMU (TDK-InvenSense ICM-20690) at 100Hz, GNSS receivers of Broadcom BCM47755 chip, and a 1280 × 640 resolution monocular camera with 1.4 μm pixel size at 30 FPS equipped in the Mi8 smartphone. GNSS receivers have a Broadcom BCM47755 chip that receives GPS (L1+L5), Galileo (E1+E5a), QZSS (L1+L5), GLONASS (L1), Beidou (B1) at 1 Hz. In our system, the intrinsic and extrinsic parameters of the camera and IMU of Xiaomi are calibrated, and the window size is set to 10, as same as VINS-Mono, but the loop closure function is disabled. The library chosen is Ceres for the optimization part. And the experiments conducted are executed on a desktop personal computer equipped with an Intel i9-9900K at 3.6 GHz and 31.2-GB memory.

#### 4.2 Experiment Results

Experiment	Algorithm	RMSE(m)	SVM
TrajectoryA Senario1 (A1)	3DMA GNSS	12.87	84%
	VINS FGO		
	3DMA GNSS VINS-IO FGO	8.34	
	VINS MONO	8.80	
	3DMA GNSS	11.94	
TrajectoryA Senario2 (A2)	3DMA GNSS	44.60	77%
	VINS FGO		
	3DMA GNSS VINS-IO FGO	18.63	
	VINS MONO	37.47	
	3DMA GNSS	13.00	
TrajectoryB Senario1 (B1)	3DMA GNSS	6.80	94%
	VINS FGO		
	3DMA GNSS VINS-IO FGO	5.73	
	VINS MONO	10.29	
	3DMA GNSS	7.29	
TrajectoryB Senario2 (B2)	3DMA GNSS	18.35	97%
	VINS FGO		
	3DMA GNSS VINS-IO FGO	3.91	
	VINS MONO	13.86	
	3DMA GNSS	8.20	

Table. 1 The RMSE of absolute positioning error along with the time and SVM accuracy comparison between 3DMA GNSSVINS-IO FGO, 3DMA GNSSVINS FGO, VINS-MONO, and 3DMA GNSS in all experiments

In general, GNSSVINS FGO provides the worst positioning performance with 4m or more. This is because the VINS-Mono accumulates the positioning drift to the operation time. After integrating the GNSS in the position solution domain, the positioning performance is improved to within 4 m average. Together with the IO switching factor, the RMSE of GNSSVINS-IO-FGO outperforms others across all experiments. Also, It can

be seen that trajectory A gets a relatively low accuracy result in IO detection than trajectory B. It is because trajectory A has a more complex environment than trajectory B.

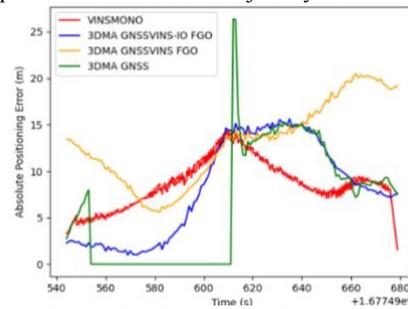


Figure. 7 The A1 absolute positioning error comparison.

After that, this section also selects some of the experiment results as a case study. Figure. 7 shows the absolute positioning error of Experiment A1. It can be observed that the VINS-GNSS-IO suppressed the positioning error in general. However, a wrong estimated covariance of 3DMA GNSS may degrade the optimization results, such as in epochs around 620 to 650 (+1.677e9), due to the covariance cannot bound to the actual positioning error, FGO wrongly trusting the wrong 3DMA GNSS and degrades the performance. The VINS-GNSS FGO even affected the performance seriously by the large GNSS error with small covariance occurring around 615 (+1.677e9), and the whole optimized result deformed. Besides, IO detection can successfully detect the GNSS's worst epoch and mitigate the effect of large positioning errors.

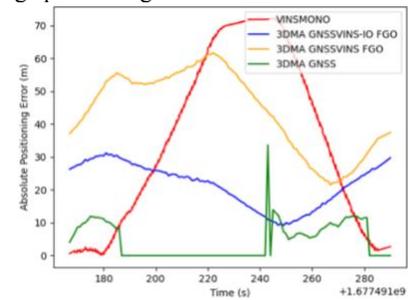


Figure. 8 The A2 absolute positioning error comparison.

Figure. 8 shows the absolute positioning error of Experiment A2. It can be observed that the VINS-GNSS-IO recovered the positioning while no GNSS signals. However, the VINS error did not be suppressed. Since it is added to FGO, the proposed method gets a worse result than 3DMA GNSS.

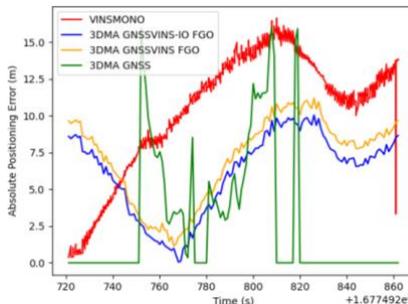


Figure. 9 The B1 absolute positioning error comparison

The third study is Experiment B1, the position error is shown in Figure 9. It can be observed that the VINS/GNSS integration can mitigate the drift caused by the VINS itself and provide a better positioning performance. After integrating the IO detection, the positioning performance is generally improved more than without IO detection.

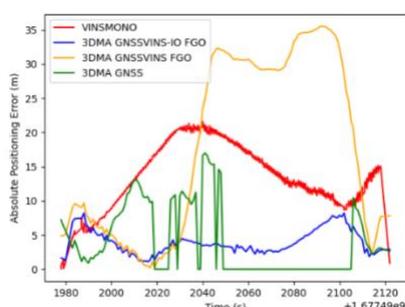


Figure. 10 The B2 absolute positioning error comparison

The last study is the B2 positioning error, the error plot is shown in Figure. 10. There is a large positioning error between times 2,050 to 2,100. However, the positioning error can be successfully suppressed after integrating with GNSS. The result with IO detection can provide a generally better positioning performance, especially at epochs around 2,030. The peak positioning error is successfully suppressed, resulting in the overall positioning error of VINS-GNSS-IO being much lower than the result without IO.

#### Experimental validation:

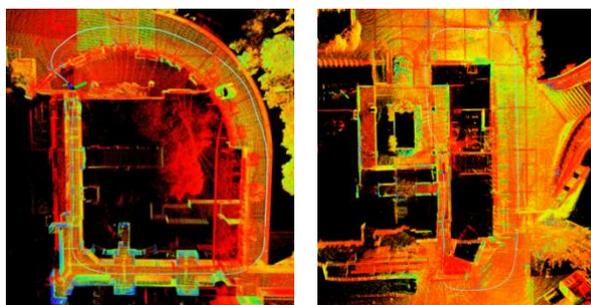


Figure. 11 LiDAR point cloud: Trajectory A      Figure. 12 LiDAR point cloud: Trajectory B

LiDAR can provide precise positioning with centimetre-level accuracy, so we use it for ground truth referencing. This study used HDL 32E Velodyne LiDAR with 360 HFOV, +10 ~ -30 VFOV, 80m range recorded at 10Hz, and the Tamagawa-seiki TAG264 IMU model with a frequency of 50 Hz. Figures 11 and 12 show the LiDAR point cloud generated by the LIO-SAM for two trajectories A and B, respectively.

## 5. CONCLUSION

This study introduces a framework based on factor graph optimization that integrates local estimates derived from prior research on VIO and 3DMA GNSS measurements with an IO detection switch. The proposed system enables precise estimation of pose locally and eliminates global drift. To evaluate the performance of the system, The effectiveness of this system is demonstrated through real-world experiments, which produced notable outcomes. However, there are still some limitations. The training model is not applicable to all smartphone devices as there is only one smartphone used in the experiment. Since the dataset is only covered on the campus, there is still an unexplored area for dealing with different environments. It is believed that there will have improvements in resolving environmental factors in these dynamic and complex urban canyons.

## 6. REFERENCES

- 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>
- 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>
- Chen, W., Zeng, Q., Liu, J., Chen, L., & Wang, H. (2016). *Research on the Multi-sensor Information Fusion Method Based on Factor Graph*. 502–506. <http://www.ion.org/publications/abstract.cfm?jp=p&articleID=13503>
- Debeunne, C., & Vivet, D. (2020). A Review of Visual–LiDAR Fusion based Simultaneous Localization and Mapping. *Sensors*, 20(7), Article 7. <https://doi.org/10.3390/s20072068>
- Falco, G., Pini, M., & Marucco, G. (2017). Loose and Tight GNSS/INS Integrations: Comparison of Performance Assessed in Real Urban Scenarios. *Sensors*, 17(2), Article 2. <https://doi.org/10.3390/s17020255>
- 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>
- Groves, P. D. (2013). GNSS Solutions: Multipath vs. NLOS signals. How Does Non-Line-of-Sight Reception Differ From Multipath Interference. *Inside GNSS (Magazine)*, 8, 40–42.
- Groves, P. D. (2016). It's Time for 3D Mapping–Aided GNSS. *Inside GNSS Magazine*, 50–56.
- Han, L., Lin, Y., Du, G., & Lian, S. (2019). *DeepVIO: Self-supervised Deep Learning of Monocular Visual Inertial Odometry using 3D Geometric Constraints* (arXiv:1906.11435). arXiv. <https://doi.org/10.48550/arXiv.1906.11435>
- 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), Article 4. <https://doi.org/10.3390/s18041159>
- 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>
- Huai, J., Toth, C. K., & Grejner-Brzezinska, D. A. (2015). *Stereo-inertial Odometry Using Nonlinear Optimization*. 2087–2097. <http://www.ion.org/publications/abstract.cfm?jp=p&articleID=12975>
- Jimenez, A. R., Seco, F., Prieto, C., & Guevara, J. (2009). A comparison of Pedestrian Dead-Reckoning algorithms using a low-cost MEMS IMU. *2009 IEEE International Symposium on Intelligent Signal Processing*, 37–42. <https://doi.org/10.1109/WISP.2009.5286542>
- Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R., & Furgale, P. (2015). Keyframe-based visual–inertial odometry using nonlinear optimization. *The International Journal of Robotics Research*, 34(3), 314–334. <https://doi.org/10.1177/0278364914554813>
- 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>
- Miura, S., Hsu, L.-T., Chen, F., & Kamijo, S. (2015). GPS Error Correction With Pseudorange Evaluation Using Three-Dimensional Maps. *IEEE Transactions on Intelligent Transportation Systems*, 16(6), 3104–3115.  
<https://doi.org/10.1109/TITS.2015.2432122>
- 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), Article 17. <https://doi.org/10.3390/s22176533>
- 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>
- Ng, H.-F., Zhang, G., Luo, Y., & Hsu, L.-T. (2021). Urban positioning: 3D mapping-aided GNSS using dual-frequency pseudorange measurements from smartphones. *NAVIGATION: Journal of the Institute of Navigation*, 68(4), 727–749.  
<https://doi.org/10.1002/navi.448>
- Pei, L., Chen, R., Chen, Y., Leppäkoski, H., & Perttula, A. (2009). Indoor/Outdoor Seamless Positioning Technologies Integrated on Smart Phone. *2009 First International Conference on Advances in Satellite and Space Communications*, 141–145.  
<https://doi.org/10.1109/SPACOMM.2009.12>
- Pratama, A. R., Widyawan, & Hidayat, R. (2012). Smartphone-based Pedestrian Dead Reckoning as an indoor positioning system. *2012 International Conference on System Engineering and Technology (ICSET)*, 1–6.  
<https://doi.org/10.1109/ICSEngT.2012.6339316>
- 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>
- Qin, T., Pan, J., Cao, S., & Shen, S. (2019). *A General Optimization-based Framework for Local Odometry Estimation with Multiple Sensors* (arXiv:1901.03638). arXiv. <https://doi.org/10.48550/arXiv.1901.03638>
- Qin, T., & Shen, S. (2018). *Online Temporal Calibration for Monocular Visual-Inertial Systems* (arXiv:1808.00692). arXiv.  
<https://doi.org/10.48550/arXiv.1808.00692>
- Rajak, S., Panja, A. K., Chowdhury, C., & Neogy, S. (2021a). 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](https://doi.org/10.1007/978-981-15-9927-9_67)
- Rajak, S., Panja, A. K., Chowdhury, C., & Neogy, S. (2021b). 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](https://doi.org/10.1007/978-981-15-9927-9_67)
- Shen, S., Michael, N., & Kumar, V. (2015). Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft MAVs. *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 5303–5310.  
<https://doi.org/10.1109/ICRA.2015.7139939>
- Sheta, A., Mohsen, A., Sheta, B., & Hassan, M. (2018). Improved Localization for Android Smartphones Based on Integration of Raw GNSS Measurements and IMU Sensors. *2018 International Conference on Computer and Applications (ICCA)*, 297–302.  
<https://doi.org/10.1109/COMAPP.2018.8460352>
- Wang, L., Groves, P. D., & Ziebart, M. K. (2013). GNSS Shadow Matching: Improving Urban Positioning Accuracy Using a 3D City Model with Optimized Visibility Scoring Scheme. *NAVIGATION*, 60(3), 195–207. <https://doi.org/10.1002/navi.38>
- 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>
- Wang, L., Sommer, L., Riedel, T., Beigl, M., Zhou, Y., & Huang, Y. (2020). NeuralIO: Indoor Outdoor Detection via Multimodal Sensor Data Fusion on Smartphones. In H. Santos, G. V. Pereira, M. Budde, S. F. Lopes, & P. Nikolic (Eds.), *Science and Technologies for Smart Cities* (pp. 127–138). Springer International Publishing.  
[https://doi.org/10.1007/978-3-030-51005-3\\_13](https://doi.org/10.1007/978-3-030-51005-3_13)
- Wen, W., & Hsu, L.-T. (2021). Towards Robust GNSS Positioning and Real-time Kinematic Using Factor Graph Optimization. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 5884–5890.  
<https://doi.org/10.1109/ICRA48506.2021.9562037>
- Wen, W., Kan, Y. C., & Hsu, L.-T. (2019). *Performance Comparison of GNSS/INS Integrations Based on EKF and Factor Graph Optimization*. 3019–3032.  
<https://doi.org/10.33012/2019.17129>
- Wen, W., Pfeifer, T., Bai, X., & Hsu, L.-T. (2021a). Factor graph optimization for GNSS/INS integration: A comparison with the extended Kalman filter. *NAVIGATION: Journal of the Institute of Navigation*, 68(2), 315–331. <https://doi.org/10.1002/navi.421>
- Wen, W., Pfeifer, T., Bai, X., & Hsu, L.-T. (2021b). Factor graph optimization for GNSS/INS integration: A comparison with the extended Kalman filter. *NAVIGATION: Journal of the Institute of Navigation*, 68(2), 315–331. <https://doi.org/10.1002/navi.421>
- Yan, J., Diakité, A. A., & Zlatanova, S. (2019). A generic space definition framework to support seamless indoor/outdoor navigation systems. *Transactions in GIS*, 23(6), 1273–1295.  
<https://doi.org/10.1111/tgis.12574>
- Zhang, Z., & Scaramuzza, D. (2018). A Tutorial on Quantitative Trajectory Evaluation for Visual(-Inertial) Odometry. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 7244–7251.  
<https://doi.org/10.1109/IROS.2018.8593941>
- Zhong, Q., & Groves, P. D. (2022). Multi-Epoch 3D-Mapping-Aided Positioning using Bayesian Filtering Techniques. *NAVIGATION: Journal of the Institute of Navigation*, 69(2). <https://doi.org/10.33012/navi.515>
- Zhu, Y., Luo, H., Wang, Q., Zhao, F., Ning, B., Ke, Q., & Zhang, C. (2019). A Fast Indoor/Outdoor Transition Detection Algorithm Based on Machine Learning. *Sensors*, 19(4), Article 4.  
<https://doi.org/10.3390/s19040786>