# Self-Contained Pedestrian Navigation Fusing ML-selected GNSS Carrier Phase and Inertial Signals in Challenging Environments

Ziyou Li, Ni Zhu, Valérie Renaudin

# Self-Contained Pedestrian Navigation Fusing ML-selected GNSS Carrier Phase and Inertial Signals in Challenging Environments

Ziyou Li, Ni Zhu *Member, IEEE* and Valérie Renaudin, *Member, IEEE*

***Abstract*— The performance of the Global Navigation Satellite System (GNSS)-based navigation is usually degraded in challenging environments such as deep urban and light indoors. In such environments, the satellite visibility is reduced, and the complex propagation conditions perturb the GNSS signals with attenuation, refraction, and frequent reflection. This paper presents a novel Artificial Intelligence (AI)-based approach, to tackle the complex GNSS positioning problems in deep urban, even light indoors. The new approach, called LIGHT, i.e., Light Indoor GNSS macHine-learning-based Time difference carrier phase, can select healthy GNSS carrier phase data for positioning thanks to Machine Learning (ML). The selected carrier phase data are fed into a Time Difference Carrier Phase (TDCP)-based Extended Kalman Filter (EKF) to estimate the user's velocity. Four trajectories including shopping mall, railway station, shipyard, as well as urban canyon scenarios over a 3.2 km total walking distance with a handheld device are tested. It is shown that at least half of the epochs are selected as usable for light indoor GNSS TDCP standalone positioning, and the accuracy of the velocity estimates can improve up to 88% in terms of the 75$^{th}$ percentile of the absolute horizontal velocity error compared with the state-of-the-art non-ML approach. Furthermore, a newly designed hybridization filter LIGHT-PDR that fuses the LIGHT algorithm and Pedestrian Dead Reckoning (PDR) solution is applied to perform seamless indoor/outdoor positioning in a more robust manner.**

***Index Terms*— GNSS carrier phase, Urban/indoor pedestrian navigation, Complex environment, PDR**

## I. Introduction

With the explosive trend of wearable smart devices, the demands for ubiquitous infrastructure-free pedestrian navigation are constantly growing. The hybridization between PDR and GNSS is a classic way for seamless indoor/outdoor positioning. On the one hand, GNSS can provide satisfactory navigation solutions outdoors, especially in open sky areas. On the other hand, PDR is a classic way to estimate pedestrian trajectories using inertial and magnetic measurements. Therefore, these two approaches are often hybridized together to mitigate PDR error propagation using GNSS signals.

However, urban and indoor environments where most of the human activities take place present huge challenges for GNSS and PDR positioning. On the one hand, in such environments, the satellite visibility is degraded, and the remaining tracked GNSS signals suffer severely from reflection and refraction. Also, the indoor GNSS signal is attenuated through its propagation channel, and the attenuation results in the need for a longer coherent integration time that increases the magnitude of the noise [1]. Therefore, common sense is that GNSS signals are not usable for indoor positioning. On the other hand, the processing of the magnetometer measurements is complicated if metallic materials are in the vicinity of the pedestrian, perturbing its use for PDR heading estimation. In the GNSS/PDR hybridization approach, GNSS is usually used for correcting the PDR solution, thus the GNSS positioning quality is essential to guarantee a satisfactory hybridized solution.

This work aims to exploit the degraded GNSS signal in challenging environments as much as possible, and also hybridize it with the inertial signal, to meet the expectation of ubiquitous infrastructure-free pedestrian navigation. In this paper, we extend the operational limits of GNSS to light indoor environments by using GNSS carrier phase measurement. Light indoor environment here refers to the indoor area where GNSS signals can be tracked by the receiver, although the number of tracked satellites and the signal power could be lower than outdoors. A new approach called LIGHT (**L**ight **I**ndoor **G**NSS Mac**H**ine-learning-based **T**ime Difference Carrier Phase) is proposed to exploit GNSS carrier phase signals in challenging environments. Prior research has shown that GNSS carrier phase measurement outperforms other GNSS measurements: it is less affected by multipath effects compared to code measurement [2], and provides better navigation performance compared to Doppler [3]. However, most GNSS signals in deep urban and light indoors are severely disturbed by Non-Line-Of-Sight (NLOS) or multipath effects. The ability to select healthy signals is important to achieve a better navigation solution. In the proposed LIGHT algorithm, the GNSS carrier phase signal is selected by machine learning (ML). The selected signals are fed into the Time Difference Carrier Phase (TDCP) algorithm, which removes several error terms by differencing the carrier phase within two consecutive epochs [4]. TDCP is a time gradient-based approach that provides a more accurate estimate of pedestrian velocity.

In real-world applications, both "light indoor"' and "deep indoor"' environments can occur together. The GNSS-only solution may not be continuous because of the lack of

Ziyou Li, Ni Zhu, and Valérie Renaudin are with the GEOLOC laboratory, AME Department of Gustave Eiffel University, Bouguenais, 44344 France (email: {ziyou.li, ni.zhu, valerie.renaudin}@univ-eiffel.fr). Corresponding author: Ziyou Li

measurements in deep indoor environments and some transition areas between outdoor and indoor environments. Therefore, the PDR solution needs to be fused to obtain a continuous solution. In this paper, an Extended Kalman Filter (EKF)-based navigation filter called LIGHT-PDR is designed to combine the LIGHT and PDR algorithms to further improve the estimated pedestrian trajectory.

This paper is an extended version of the paper [5]. It describes further the LIGHT approach in terms of the ML model training process. Also, 2 more test scenarios are added.

The remaining part of this paper is organized as follows: Section II presents the state-of-the-art of indoor GNSS use and the ML-based GNSS. Then, in section III, we analyze how the signal attenuation impacts the receiver tracking process, and further define the "light indoors" according to our needs for indoor positioning. The detailed LIGHT algorithm and the structure of the LIGHT-PDR hybridization filter are presented in Section IV. In Section V, the proposed LIGHT / LIGHT-PDR approaches are evaluated in several light indoor environments, such as shopping mall and railway station, and in deep urban environments for the handheld device use case. The conclusion and perspective are given in Section VI.

## II. State-of-the-art

The analysis of the current state-of-the-art focuses on two aspects: 1) indoor GNSS use and 2) artificial intelligence (AI)-based GNSS signal quality discrimination.

Indoor GNSS positioning is still an open research topic. In the state-of-the-art, there are mainly two general approaches: using high-sensitivity receivers (HS) and implementing GNSS repeaters or simulators. In [6], an HS-GNSS software receiver for indoor navigation is proposed. Code measurements are fed into a Kalman filter to compute the navigation solution. On average, the HS receiver can track 8 GPS L1 C/A satellites and achieve a 2D positioning accuracy (Root Mean Square, RMS) of 2.2 m for the dynamic test in the residential building. [7] proposes a hybridization approach between a direct vector-based HS-GNSS software receiver using GPS L1 Doppler measurements and PDR. The proposed HSGPS/PDR approach outperforms the conventional HSGPS/PDR approach with a 24% and 60% improvement in north and east RMS positioning errors, respectively.

Regarding the GNSS repeater, [8] and [9] use the repeated carrier phase measurement to achieve a $<$0.5 m horizontal positioning error for parking lot and laboratory environments respectively. [10] downconverts the 1575.42 MHz GPS L1 signal to 433 MHz for transmission, then at the receiver level, the signal is upconverted back to the L1 frequency. In this way, the system is immune to NLOS conditions. Horizontal positioning accuracy below one meter has been achieved.

The current state-of-the-art in indoor GNSS relies heavily on additional hardware, while AI provides the ability to select the healthy GNSS signal by enhancing the software. In challenging GNSS signal propagation environments, such as cities, AI is usually used for signal classification. To accomplish this task, features can be extracted at two different levels: (1) at the raw measurement level and (2) at the receiver correlator output level, since the correlation shape of NLOS/multipath signals is distorted compared to healthy signals. For example, [11] uses a neural network (NN) to detect the NLOS signal based on three features of the correlation outputs, i.e., signal strength as a function of elevation, number of local correlation maxima, and delay of maximum correlation. The result shows that the NN method can achieve an average NLOS classification accuracy of 97.7% in urban street canyons. [12] extracts the features from the measurements, namely signal-to-noise ratio, normalized pseudorange residuals, elevation, and pesudorange rate consistency. The tested SVM model achieves a 91.5% LOS/NLOS classification rate in urban street canyons. In addition, the detection of indoor/outdoor transition stage is critical for seamless navigation. In [13], a fast transition detection method for smartphone GNSS receivers is proposed by merging different ML models. 4 models are used in the base layer: XGBoost, LightGBM, AdaBoost and Random Forest; and logistic regression is used in the second layer to make the final decision. 36 features, including C/N0, number of satellites, Dilution of Precision (DOP), azimuth and their statistical values are used for training. For indoor/outdoor classification, an accuracy of 97% for 17,199 testing epochs that contains both indoor or outdoor walking scenarios is achieved. And among 11,218 epochs walking in the transition area, the accuracy of indoor/outdoor transition detection is 92.8%. What is more, to better interpret the measurement quality and improve the navigation solution, [14] proposes an approach to weight GNSS measurements using the Long-Short Term Memory Neural Network (LSTM NN). The predicted weights are used in the Weighted Least Square (WLS) algorithm to estimate the user's position. The features are composed of satellite elevation, carrier phase lock time, carrier-to-noise density ratio $C/N_0$, $C/N_0$ variance, $C/N_0$ mean value, the window size of calculating the mean and the variance, and pseudorange residual. Compared to the state-of-the-art Fault Detection Exclusion (FDE) approach, the proposed method provides a 1.2 m improvement in 68% horizontal position error cumulative density function (CDF) and a 0.66 m improvement in 68% vertical position error CDF in urban environments.

Globally, the majority of existing indoor GNSS research is infrastructure-dependent, and AI-based GNSS research still mainly focuses on outdoor environments. The main contribution of this research work is to extend the possibility of using GNSS carrier phase measurements in light indoors with the help of AI techniques, following an infrastructure-free strategy.

## III. Definition of Light Indoors

There is no clear definition of the light indoor environment in the literature, therefore it is necessary to analyze its nature, especially how the signal is attenuated in different indoor environments. In the introduction, we have briefly defined the "light indoor" environment from the GNSS receiver signal tracking point of view, i.e., the light indoor environment refers to the indoor spaces where the receiver can successfully track a few satellites, contrary to "deep indoors" where no satellite can be successfully tracked at all.

Since the GNSS carrier phase measurements are used in this paper, the carrier phase tracking performance from the receiver

TABLE I
ATTENUATION THROUGH DIFFERENT BUILDING MATERIALS FOR GPS L1 SIGNAL [17]

| Materials | Attenuation [dB] |
|---|---|
| Drywall | 2.40 |
| Plywood | 0.93 |
| Glass | 2.43 |
| Tinted glass | 24.44 |
| Lumber | 2.40 |
| Brick | 5.19 |
| Concrete | 9.57 |
| Reinforced concrete | 16.70 |

Phase Tracking Loop (PLL) is being analyzed. According to [15], the thermal noise is usually considered the only error source for carrier phase tracking, and the standard derivation of the PLL thermal noise jitter $\sigma_{PLLt}$ (with the unit of degrees) can be written as:

$$\sigma_{PLLt} = \frac{\lambda}{2\pi}\sqrt{\frac{B_n}{C/N_0}\left(1+\frac{1}{2T*C/N_0}\right)} \quad (1)$$

where $\lambda$ is the carrier wavelength, $B_n$ is the PLL loop bandwidth, $T$ is the receiver integration time, and $C/N_0$ is the carrier to noise density ratio. Suppose $B_n$ and $T$ keep constant for the same receiver, and the received signals have the same frequency, the wavelength is also constant. Therefore, the PLL thermal noise jitter only depends on $C/N_0$, which is the strength of the received signal. If the signal is strongly attenuated, its $C/N_0$ is low, then it results in a high PLL thermal noise jitter, thus the phase tracking may fail. In practice, it has been shown that the carrier phase tracking would fail if the $C/N_0$ is below 25 dB-Hz for a second-order PLL, where the noise jitter is more than 15 degrees [16].

Table I compares the attenuation through different building materials for GPS L1 signal [17]. It has shown that some of the building materials such as drywall, plywood, glass, and lumber can make small attenuation on the received signal, while some others such as tinted glass and reinforced concrete have a strong impact on attenuating the signal.

Therefore, according to our needs of indoor GNSS positioning, we can define the term "light indoor environment" as follows:

*The light indoor environment is an enclosed indoor space surrounded by the building materials that cause small levels of GNSS signal attenuation, such that the GNSS signals can be successfully tracked with an acceptable level of tracking jitter.*

In modern architectures, materials with low GNSS signal attenuation such as glass and other composite materials are widely used instead of simply using the concrete. Also, wooden architecture is still very common in residential constructions. This allows the use of GNSS signals in such environments to become possible. However, the signal strength does not totally reflect the signal quality. Even though in light indoors the signal can be successfully tracked, it suffers strongly from the multipath. Being able to select the tracked signal with good quality is essential to make the GNSS positioning system fully functional.

## IV. METHODOLOGY

Fig. 1 shows the flow chart of the proposed approach. Each part will be detailed in the following subsections. For GNSS processing, the LIGHT algorithm is proposed: a machine learning model is trained to select qualified carrier phase measurements for TDCP. Then, the LIGHT-PDR hybridization filter is proposed to combine the LIGHT and PDR solutions as well as barometer measurements to provide the pedestrian trajectory.

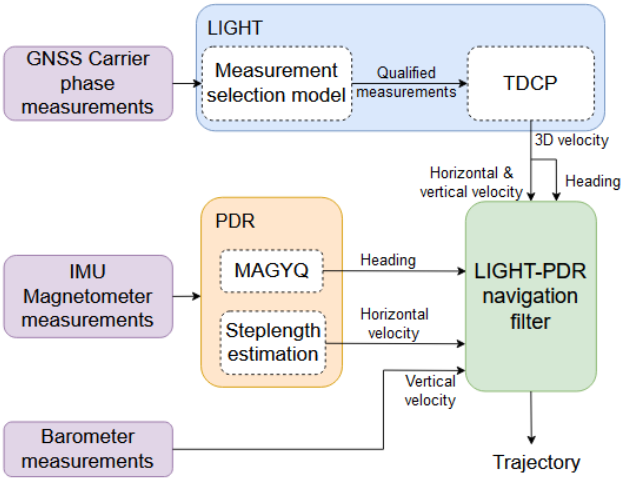


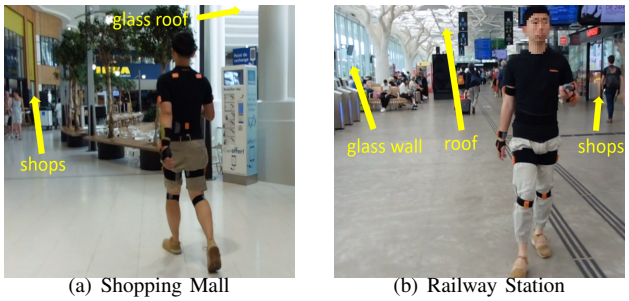

Fig. 1. General flow chart of the proposed approach

### A. LIGHT: *Light Indoor GNSS MacHine-learning-based Time Difference Carrier Phase Positioning Approach*

In challenging environments like urban and light indoors, the obstacles are common in the GNSS signal propagation channel, thus NLOS measurements can lead to inaccurate navigation solutions. For pedestrian navigation, the walking speed of the subject is relatively slow compared to other transport modes, so the GNSS measurement errors do not change significantly within the sampling intervals. Therefore, it is possible to remove the measurement errors by using the time-difference approach. Our previous research work [4] has already shown satisfactory navigation performances in urban environments by using GNSS carrier phase measurements with TDCP algorithm together with the FDE. However, the indoor environment is much more complex than outdoors. On the one hand, the satellite visibility is degraded, and the magnitude of the tracked signal power is attenuated [1]. On the other hand, indoors, where lots of reflections and refractions take place, LOS signals can hardly be received, and fast-changing NLOS signals may dominate the tracked signals. Despite the fact that the excess delay of GNSS signal reflection indoors (around 1 $\mu$s) is much shorter than outdoors (up to 100 $\mu$s) [18], indoor signal receptions involve more frequent reflections, even multiple reflections for a single reception. This increases significantly the complexity for indoor GNSS signal classification. What is more, every indoor space is different from each other, so it is not possible to mathematically model the indoor GNSS signal propagation channel accurately. All those limitations complicate the indoor

TABLE II
COLLECTED DATASET CHARACTERISTICS FOR THE ML MODEL

| Environment | Light indoor/urban canyon epochs / duration [s] | Total epochs / total duration [s] |
|---|---|---|
| Shopping Mall | 6851 / 1370.2s | 8171 / 1634.2s |
| Railway Station | 7609 / 1521.8s | 10430 / 2086s |
| Office | 1441 / 288.2s | 2293 / 458.6s |
| Urban | 2527 / 505.4s | 3723 / 744.6s |
| ***Total*** | ***18428 / 3685.6s*** | ***24617 / 4923.4s*** |




(a) Shopping Mall (b) Railway Station

(c) Office (d) Urban Canyon

Fig. 2. Different environments for data collection

navigation with standalone GNSS positioning, thus quite few research works are addressing this problem. Therefore, machine learning (ML) is expected to tackle the complex GNSS indoor positioning problem.

The proposed LIGHT algorithm builds an ML model capable of identifying usable GNSS carrier phase measurements for TDCP in challenging surroundings. The procedure of constructing the LIGHT model involves four key steps: (1) Dataset preparation, to collect the data; (2) Data labeling, to label the collected GNSS carrier phase measurements as "usable" and "not usable"; (3) Feature selection, to select the most appropriate features for training; and (4) Model selection, to test different models and select one with the best performance. The details of each step are presented in the following subsections.

*1) Dataset Preparation:* To build a diverse dataset for machine learning, GNSS carrier phase data were collected at 5 Hz in different challenging environments: shopping mall with glass roof (Fig. 2 (a)), railway station (Fig. 2 (b)), office building (Fig. 2 (c)), and urban canyons (Fig. 2 (d)) where the signals' reflection and refraction are more frequent compared with other outdoor spaces. Table II details a summary of the dataset recorded to build the ML model.

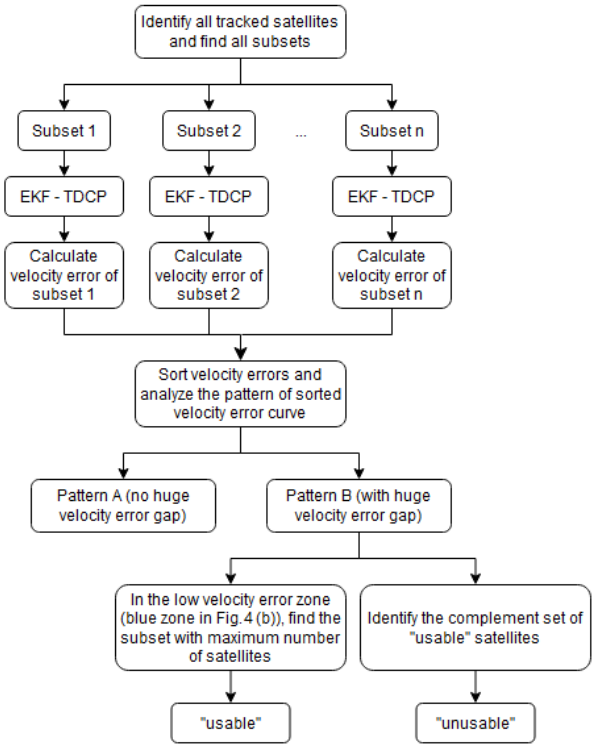
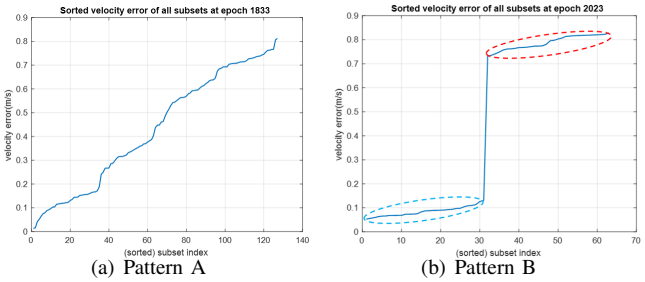



Fig. 3. Flowchart of labeling method at each epoch




(a) Pattern A (b) Pattern B

Fig. 4. Two typical patterns on sorted velocity errors of all subsets *(x axis: subset index after sorting the velocity error; y axis: sorted velocity error)*

*2) Data Labeling:* In the state-of-the-art ML-based GNSS vehicle positioning in urban environments, such as [19] [20], the ray-tracing methods are used for labeling LOS/NLOS signals. However, in light indoors, it is hard to track LOS signals. Even though sometimes there is a direct link between the receiver and the glass roof or the glass wall, the signal is also NLOS since the refraction takes place while it penetrates through the glass. Moreover, since TDCP is a time-differenced algorithm, if the NLOS signal does not change very fast within 2 consecutive epochs, the multipath errors will be eliminated, thus the NLOS could become useful. Therefore, the ray-tracing methods that identify LOS/NLOS signals are not suitable in our case.

The objective of developing the LIGHT algorithm is to select GNSS carrier phase measurements that provide satisfactory velocity estimates for TDCP in challenging environments such as deep urban and light indoors. Therefore, assessing the velocity errors estimated by different satellite combinations could be a clue for our data labeling. Since it is a classification

problem, two categories of satellites must be identified: 1) satellites with usable measurements that contribute to low velocity errors, and 2) satellites with unusable measurements that result in high velocity errors. Also, there must be a clear boundary seen on the velocity errors between the two categories, to make sure that the data labeling procedure does not make any ambiguity.

In this paper, we propose an exhaustive satellite subset velocity error searching approach to label the data as shown in Fig. 3. Since the number of tracked satellites is reduced in light indoor environments, it is feasible to find all possible combinations of tracked satellites at each epoch. Each combination is called a "subset". Suppose that at epoch $k$, $N$ satellites are tracked by the receiver, therefore if we remove the zero set, the total number of subsets is $2^N-1$. Some subsets could have only one satellite, but with the Kalman filter, it is possible to estimate the TDCP velocity with only one measurement.

Then an EKF-based TDCP without any fault exclusion or outlier removal technique is applied to each subset to estimate the velocity. Knowing the reference velocity, the velocity error $\varepsilon_{v_i}$ $(i = 1...n,\ n \in [1, 2^N - 1]))$ of each subset can be calculated. If we sort the velocity errors $\varepsilon_{v_i}$ in ascending order, 2 types of patterns on the sorted velocity error curves can be observed. Fig. 4 illustrates an example of these 2 patterns. In pattern A (Fig. 4 (a)), the sorted velocity error curve increases smoothly and there is no clear boundary between low and high velocity errors; while in Pattern B, this boundary exists, which is the gap in the sorted velocity error curve. On the left side of the Pattern B curve (highlighted by the blue circle in Fig. 4 (b)), all the subsets contribute to small velocity errors, which means that the subsets in the blue zone only contain the usable satellites. On the right side of the curve (highlighted by the red circle in Fig. 4 (b)), the velocity errors are much larger, which means that all subsets in the red zone must contain at least one "unusable" satellite whose carrier phase measurement is inaccurate.

Our training set was built based on the epochs whose sorted velocity error curves correspond to Pattern B in which the minimum velocity error is usually lower than 0.2 m/s. Within the blue zone, the subset with the maximum number of satellites is identified as "usable" set. The complementary set of the "usable" set contains the "unusable" satellites.

In fact, the case of Pattern B is not common in light indoor environments. Table III details the number of epochs that Pattern B can be observed for each environment. The complete "usable" epochs dataset corresponds to 9392 observations (1 observation = 1 feature set of one satellite). 70% of the observations (6574) are used for training the model, and 30% of the observations (2818) are used for testing the model classification.

*3) Feature Selection:* After summarizing the state-of-the-art on ML-based GNSS signal classification and analyzing the needs of this research, 15 candidate features are firstly proposed. These candidate features can be divided into 4 categories: 4 signal-nature-based features, 3 raw-measurement-based features, 7 time-relevant features, and 1 TDCP Kalman filter-based feature. The explanation of each feature and the feature selection analysis are as follows.

TABLE III
SUMMARY OF THE USABLE DATASET FOR ML MODEL

| Environment | Usable epochs | Light indoor/urban canyon epochs | Total epochs |
|---|---|---|---|
| Shopping Mall | 399 | 6851 | 8171 |
| Railway Station | 265 | 7609 | 10430 |
| Office | 42 | 1441 | 2293 |
| Urban | 183 | 2527 | 3723 |
| ***Total*** | ***889*** | ***18428*** | ***24617*** |

**(A) Signal-nature-based features:**

(F1) Carrier-to-noise density ratio (C/N0),
(F2) Satellite elevation (elev),
(F3) Satellite azimuth (azimuth),
(F4) Number of tracked satellites (n_sat).

For C/N0, as shown in Fig. 6 (a), the distributions of the "usable" and "unusable" classes are more separable than others. In Fig. 5 (a), C/N0 has a 55% linear correlation with the label. However, the distributions of the other signal-nature-based features are overlapped (Fig. 6(b)(c)(d)). In the case of light indoor GNSS positioning, the signal quality depends on the geometry of the building. For example, in the shopping mall with the glass roof, the good measurements come from higher elevations, while in the railway station with the large glass walls, the good measurements come from lower elevations. Similar to the azimuth. Therefore, C/N0 is the only selected feature in this category.

**(B) Raw-measurement-based features:**

In challenging environments especially light indoors, the fast-changing multipath is the main GNSS measurement error source. Although the TDCP algorithm is effective against multipath by taking the time difference, the multipath change time could be even shorter (minimum 1 ms [18]) than the TDCP time interval (200 ms for 5 Hz sampling frequency). Therefore, some multipath errors could remain after doing the time difference. Therefore, several candidate features that are based on the raw measurements must be found such that the ML model can identify those problematic measurements.

(F5) Phase-range rate consistency (tdcp_dopp):

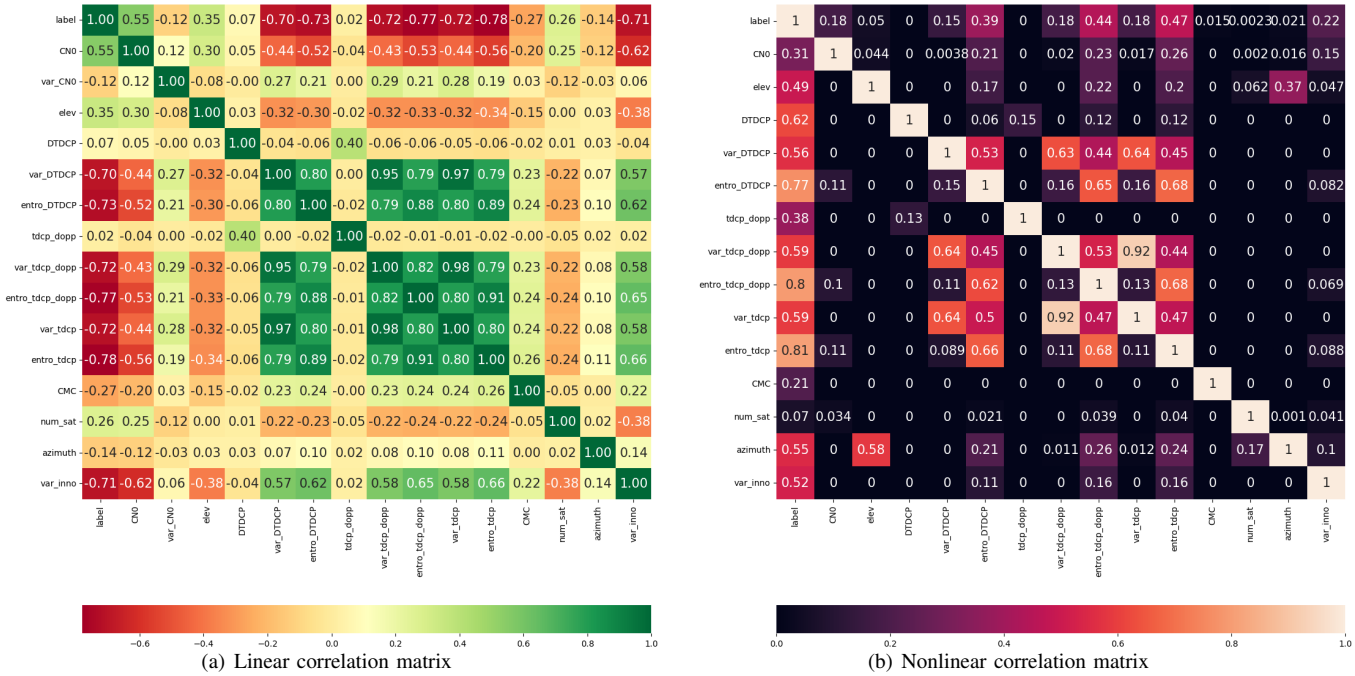
It is the difference between the phase-range rate $d\rho_{\phi}^{i,k}$ and the Doppler-range rate $d\rho_{D}^{i,k}$. If there is no measurement error, the Doppler shift of the signal is the time derivative of the carrier phase [21], therefore the measurement error can be accessed through the phase-range rate consistency. Phase-range rate is the change rate of carrier phase (converted into range) within 2 consecutive epochs, it is also the TDCP measurement used in EKF. The phase-range rate (or TDCP measurement) is expressed as follows:

$$d\rho_{\phi}^{i,k} = \lambda\phi^{i,k} - \lambda\phi^{i,k-1} \tag{2}$$

where $\phi^{i,k}$ is the carrier phase measurement of satellite $i$ at epoch $k$, and $\lambda$ is the wavelength of the carrier phase signal. Doppler-range rate is derived from Doppler:

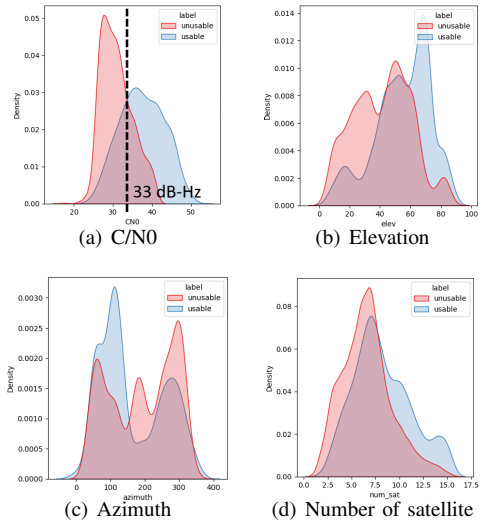
$$d\rho_{D}^{i,k} = -\lambda f_{D}^{i,k}\Delta t \tag{3}$$

where $f_{D}^{i,k}$ is Doppler frequency of satellite $i$ at epoch $k$.




(a) Linear correlation matrix (b) Nonlinear correlation matrix

Fig. 5. Correlation matrices of candidate features




(a) C/N0 (b) Elevation (c) Azimuth (d) Number of satellite

Fig. 6. Distributions of category A features *(blue: usable, red: unusable)*

(F6) Double time differenced carrier phase (DTDCP):

The phase measurements are twice differenced. It measures the difference of TDCP measurements between 2 epochs, as expressed in (8):

$$\begin{aligned} DTDCP &= d\rho_{\phi}^{i,k} - d\rho_{\phi}^{i,k-1} \\ &= (\lambda\phi^{i,k} - \lambda\phi^{i,k-1}) - (\lambda\phi^{i,k-1} - \lambda\phi^{i,k-2}) \end{aligned} \tag{4}$$

Fig. 7 shows an example of both TDCP and DTDCP measurements from the same satellite for an urban/light indoor mixed environment dataset. The x-axis of Fig. 7 is the epoch index. 5 consecutive epochs imply one second. We can see that if the signal is not perturbed, small fluctuations that depict the receiver motion can be observed in both TDCP and DTDCP measurements, as highlighted by the zoomed signals in Fig. 7. When the signal is perturbed, i.e., the carrier phase signal suffers from multipath or cycle slip, spikes can be observed from the TDCP measurement. However, the TDCP measurement is not stationary as its mean value keeps increasing over time. By doing the DTDCP, the measurement becomes stationary, which allows us to better assess the magnitude of TDCP perturbations.

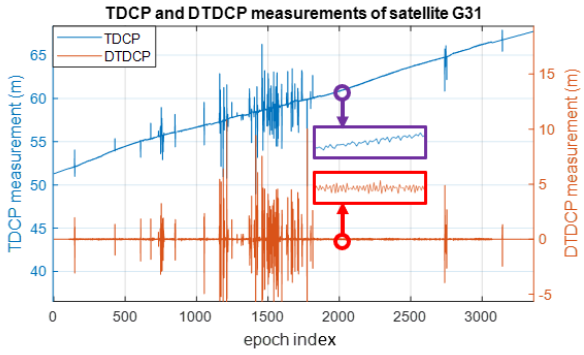
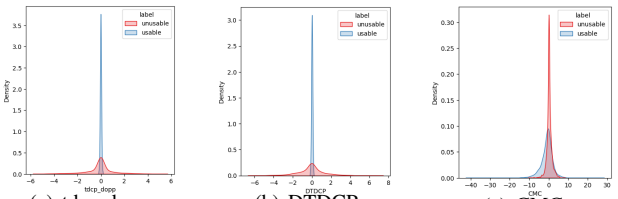



Fig. 7. TDCP and DTDCP measurements of satellite G31 for an urban/light indoor mixed dataset

(F7) Code Minus Carrier (CMC):

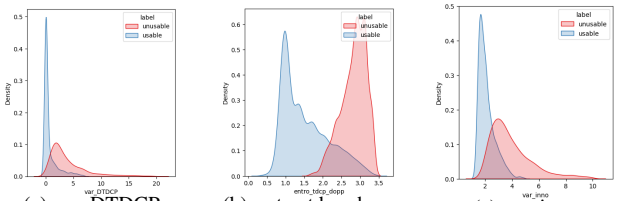
The CMC is the difference between the code measurement $R^{i,k}$ and carrier phase measurement $\lambda\phi^{i,k}$. They can be respectively modeled as:

$$R^{i,k} = d^{i,k} + c(\delta t_s^i - \delta t_u^k) + d_{tropo}^{i,k} + d_{iono}^{i,k} + \epsilon_{Rk}^i \tag{5}$$




(a) tdcp_dopp (b) DTDCP (c) CMC

Fig. 8. Distributions of category B features *(blue: usable, red: unusable)*




(a) var_DTDCP (b) entro_tdcp_dopp (c) var_inno

Fig. 9. Distributions of selected categories C and D features *(blue: usable, red: unusable)*

$$\lambda\phi^{i,k} = d^{i,k} + c(\delta t_s^i - \delta t_u^k) + \lambda N + d_{tropo}^{i,k} - d_{iono}^{i,k} + \epsilon_{\phi_k}^i \quad (6)$$

where $R^{i,k}$ is the code measurement in meters, $\phi^{i,k}$ is the carrier phase in cycles, $\lambda$ is the signal wavelength, $d^{i,k}$ is the true geometrical distance between satellite and receiver, $c$ is the speed of light, $\delta t_s^i$ is the satellite clock offset, $\delta t_u^k$ is the receiver clock offset, $d_{tropo}^{i,k}$ is the tropospheric error, $d_{iono}^{i,k}$ is the ionospheric error, $\epsilon_{Rk}^i$ $\epsilon_{\phi_k}^i$ represent other code and phase errors respectively, such as multipath and noise.

By taking the difference between the code measurement (5) and the phase measurement (6), we get the CMC measurement as follows:

$$CMC^{i,k} = R^{i,k} - \lambda\phi^{i,k} = 2d_{iono}^{i,k} + \lambda N + \epsilon_{Rk}^i + \epsilon_{\phi_k}^i \quad (7)$$

It has been widely used on multipath detection [22] as well as cycle slip detection [23].

These 3 candidate features based on raw measurements are all related to the carrier phase. However, because the code measurement is more vulnerable to noise and multipath effects compared to the carrier phase [2], the CMC cannot correctly reflect the quality of the carrier phase measurement. This can also be observed in Fig. 8 (c), where the CMC distributions of the two classes are highly overlapped. For the others, in Fig. 8 (a)(b), the usable distributions are centered in the middle, while the unusable distributions have much larger values. Therefore Phase-range rate consistency (tdcp_dopp) and Double time differenced carrier phase (DTDCP) features are finally chosen.

**(C) Time-relevant features:**

(F8) Variance of carrier-to-noise density ratio (var_CN0)

(F9) Variance of phase-range rate consistency (var_tdcp_dopp)

(F10) Entropy of phase-range rate consistency (entro_tdcp_dopp)

(F11) Variance of double time differenced carrier phase (var_DTDCP)

(F12) Entropy of double time differenced carrier phase (entro_DTDCP)

(F13) Variance of TDCP measurement (var_TDCP)

(F14) Entropy of TDCP measurement (entro_TDCP).

Time-relevant features are able to measure the magnitude of fluctuation of the signal within a time window. Here, all time-relevant features are calculated within a sliding window with a length of 10 samples (for 5 Hz data). However, the distributions of the var_CN0 feature for the two target classes mostly overlap, therefore it is excluded. For the other features, most of them are highly correlated with each other, as we can see in Fig. 5, either linear or non-linear. Finally we select one variance (var_DTDCP) and one entropy (entro_tdcp_dopp), with the most separable distributions, as shown in Fig. 9 (a)(b).

**(D) TDCP Kalman filter-based feature:**

(F15) EKF innovation variance (var_inno)

In the TDCP filter [4], the innovation is the difference between the measured and the predicted phase-range rate derived from the user-satellite geometry. The outlier in the innovation may also indicate bad measurements. At each epoch $k$, the innovation variance can be directly extracted from the Kalman filter:

$$\boldsymbol{C}_k = diag(\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \mathbf{R}_k) \quad (8)$$

where, $\mathbf{H}_k$ is the EKF observation matrix, $\mathbf{P}_k$ is the state covariance matrix, and $\mathbf{R}_k$ is the measurement covariance matrix.

This feature has a relatively separable distribution (see Fig. 9 (c)), and a strong correlation with the label (see Fig. 5). It is therefore selected as one of the final features.

At the end, 6 features are selected to train the model, namely 5 non-filter-based features: Carrier-to-noise density ratio (C/N0), Phase-range rate consistency (tdcp_dopp) together with its entropy (entro_tdcp_dopp), Double time differenced carrier phase (DTDCP) together with its variance, and one filter-based feature: EKF innovation variance (var_inno).

*4) Model Selection:* Five candidate models are tested, namely Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting (GB). The metrics such as Accuracy, Precision, F1 score, and Recall score on the test set as well as Accuracy on the training set are evaluated to make the final decision. The confusion matrices on the test set of each candidate model are shown in Fig. 10. The synthesis of the candidate models' performances is given in Table IV. The highest score for each metric is highlighted in the table. The GB model gives the highest scores in terms of accuracy and F1 score on the test set and the highest accuracy on the training set.

Although the RF model has a suboptimal performance with similar metric values compared to the GB model, its training took 3.82 s while the GB took 2.76 s. What is more, GB's prediction time is much shorter than that of RF (e.g., for the shopping mall data described in the following section, GB takes 18.50 s, while RF takes 67.89 s). For this reason, the GB model is finally selected.

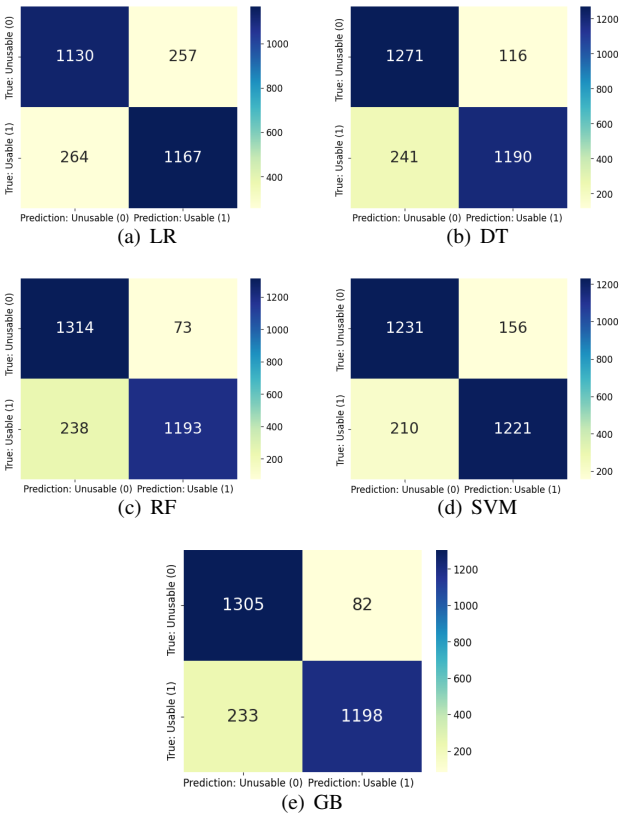



Fig. 10. Confusion matrices of each candidate model

TABLE IV
PERFORMANCES OF THE CANDIDATE MODELS

| Model | Testing Set | | | | Training Set |
|---|---|---|---|---|---|
| | Accuracy | Precision | F1 score | Recall | Accuracy |
| LR | 0.8151 | 0.8195 | 0.8175 | 0.8155 | 0.9234 |
| DT | 0.8733 | 0.9112 | 0.8696 | 0.8316 | 0.9627 |
| RF | 0.8896 | **0.9423** | 0.8847 | 0.8337 | 0.9745 |
| SVM | 0.8701 | 0.8867 | 0.8697 | **0.8532** | 0.9587 |
| GB | **0.8957** | 0.9417 | **0.8918** | 0.8470 | **1** |

*5) Time Difference Carrier Phase (TDCP) Positioning Algorithm:* Once the ML model has been built, the qualified carrier phase measurements will be used for TDCP positioning algorithm. The advantage of TDCP is that, by taking the time difference, the satellite clock offset $\delta t_s^i$, ambiguity factor $N$, atmospheric errors $d_{tropo}^{i,k}$ and $d_{iono}^{i,k}$, noise and multipath error $\epsilon_{\phi_k}^i$ in (6) are supposed to be eliminated, since they do not change fast within the sampling interval [4].

Therefore, by taking the time difference of (6), we can get:

$$\lambda \Delta \phi_{k,k-1}^i = \Delta d_{k,k-1}^i + c(\Delta \delta t_u^k) \tag{9}$$

By making the geometry analysis, for each satellite, the range difference $\Delta d_{k,k-1}^i$ in (9) can be calculated as:

$$\begin{aligned} \Delta d_{k,k-1}^i &= (\mathbf{e}_k^{s_i,u})^T \cdot (\mathbf{x}_k^{s_i} - \mathbf{x}_k^u) - (\mathbf{e}_{k-1}^{s_i,u})^T \cdot (\mathbf{x}_{k-1}^{s_i} - \mathbf{x}_{k-1}^u) \\ &= \Delta S - \Delta G - (\mathbf{e}_k^{s_i,u})^T \cdot \Delta \mathbf{d}_{k,k-1} \end{aligned} \tag{10}$$

where $\mathbf{e}_k^{s_i,u}$ and $\mathbf{e}_{k-1}^{s_i,u}$ are the unit vectors point from the user's receiver towards satellite $i$, $\mathbf{x}_k^{s_i}$ and $\mathbf{x}_{k-1}^{s_i}$ are the positions of satellite $i$ in the ECEF (Earth Centered Earth Fixed) frame, $\mathbf{x}_k^u$ and $\mathbf{x}_{k-1}^u$ are the positions of user's receiver in the ECEF frame, at epochs $k$ and $k-1$ respectively. $\Delta \mathbf{d}_{k,k-1}$ is user's displacement vector between epoch $k-1$ and epoch $k$. We denote $\Delta S = (\mathbf{e}_k^{s_i,u})^T \cdot \mathbf{x}_k^{s_i} - (\mathbf{e}_{k-1}^{s_i,u})^T \cdot \mathbf{x}_{k-1}^{s_i}$ as *Satellite Doppler correction term*, $\Delta G = (\mathbf{e}_k^{s_i,u})^T \cdot \mathbf{x}_k^u - (\mathbf{e}_{k-1}^{s_i,u})^T \cdot \mathbf{x}_{k-1}^u$ as *Geometry correction term* [24].

By taking the combination of (9) and (10), and divided by the time interval between two consecutive phase measurements, for each observed satellite, we can get a function of velocity in ECEF frame:

$$\frac{\lambda \Delta \phi_{k,k-1}^i}{\Delta t} = \frac{\Delta S - \Delta G + c(\Delta \delta t_u^k)}{\Delta t} - (\mathbf{e}_k^{s_i,u})^T \cdot \Delta \mathbf{v}_{k,k-1}^{ECEF} \tag{11}$$

In (11), the velocity $\Delta \mathbf{v}_{k,k-1}^{ECEF}$ is the average velocity of GNSS receiver within two consecutive epochs. The terms of user clock offset and velocity are unknown, they can be solved by using the Extended Kalman Filter (EKF). The details of the used TDCP navigation filter can be found in [4]. The state vector consists of the 3-dimensional position in the North-East-Down (NED) frame, 3-dimensional velocity in the NED frame, and the receiver clock bias for GPS and Galileo respectively. The constant velocity state transition model is applied in the prediction stage.

Fig. 11 shows the overall flow chart of the proposed LIGHT-based approach. At each epoch, once all tracked satellites have been identified, 5 non-filter-based features will be calculated also the EKF will make its prediction. After the prediction stage, the filter-based feature i.e., the EKF innovation variance will be calculated and extracted. All these features will be fed to the pre-trained ML model. The ML model will make the decision on which satellites can be used for EKF update to further calculate the navigation solution.

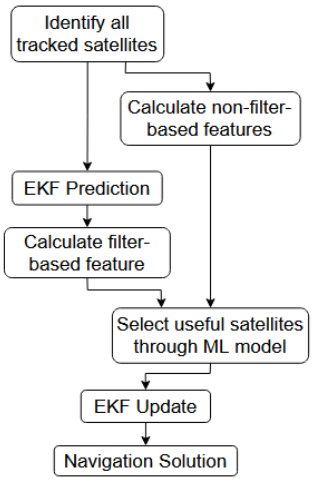



Fig. 11. Flow chart of the LIGHT-based approach at each epoch

## B. LIGHT-PDR: Hybridization Navigation Filter

The proposed LIGHT algorithm extends the use of GNSS signals to light indoor environments for positioning. However,

at certain epochs all measurements may be rejected by the ML model, or at some epochs there may even be no GNSS measurement available at all. Therefore, other sensors must be used to provide a continuous solution. For this purpose, the PDR heading and step length computed from inertial and magnetic measurements as well as the height change rate calculated from barometer measurements, are fused together with the LIGHT solution. The PDR heading is computed using the "MAGYQ" algorithm, which stands for "Magnetic, Acceleration fields and GYroscope Quaternion" [25]. In the MAGYQ filter, the gyroscope error is modeled in the quaternion space to reduce linearization and ambiguity problems. In addition, the quasi-static magnetic and acceleration field error models are used to correct the attitude errors. Together with the "SmartStep" step instance detection algorithm [26] and the handheld device step length model [27], we can iteratively compute the trajectory of the pedestrian. The next sections detail the structure of the proposed LIGHT-PDR hybrid filter.

*1) State Vector:* The state vector comprises the 3D position in the North-East-Down (NED) frame $\boldsymbol{P}_{ned}$, horizontal velocity $v_{2d}$, vertical (Down) velocity $v_d$, and heading error $\delta\theta$, since the MAGYQ filter estimates the pointing direction of the device and not the walking direction of the pedestrian.

$$\mathbf{x}^T = [\boldsymbol{P}_{ned}^T \quad v_{2d} \quad v_d \quad \delta\theta] \tag{12}$$

*2) Prediction:* The dead reckoning approach is adopted in the prediction stage of the hybrid filter. The running rate of the hybrid filter is 5 Hz, which is identical to the GNSS data rate and higher than the general step frequency (around 2 Hz). Therefore, we must calculate the equivalent "step length" during the filer prediction time interval ($\Delta t = 0.2s$) using the horizontal velocity $v_{2d}$ multiplied by the time interval $\Delta t$. Also, the MAGYQ heading $\theta_{MAGYQ}$ is corrected by heading error $\delta\theta$.

$$\hat{P}_{n_k} = P_{n_{k-1}} + v_{2d_{k-1}} * \Delta t * sin(\theta_{MAGYQ_{k-1}} + \delta\theta_{k-1}) \tag{13}$$

$$\hat{P}_{e_k} = P_{e_{k-1}} + v_{2d_{k-1}} * \Delta t * cos(\theta_{MAGYQ_{k-1}} + \delta\theta_{k-1}) \tag{14}$$

$$\hat{P}_{d_k} = P_{d_{k-1}} + v_{d_{k-1}} * \Delta t \tag{15}$$

Constant velocity model is applied for velocity components:

$$\hat{v}_{2d_k} = v_{2d_{k-1}} \tag{16}$$

$$\hat{v}_{d_k} = v_{d_{k-1}} \tag{17}$$

The heading error is assumed as constant during the interval of the prediction stage:

$$\hat{\delta\theta}_k = \delta\theta_{k-1} \tag{18}$$

*3) Updates:* In the proposed LIGHT-PDR hybrid filter, 3 types of updates are involved: the *TDCP* update corrects horizontal velocity, vertical velocity, and heading; the *Step* update corrects horizontal velocity; and the *Barometer* update corrects vertical velocity.

**(1) *TDCP* update** With the help of the proposed LIGHT algorithm, TDCP can provide the 3-dimensional velocity light indoors in the North-East-Down (NED) frame. At each epoch, if a TDCP solution is available, the correction (innovation) can be calculated as follows:

$$\boldsymbol{inno}_{TDCP} = \begin{bmatrix} v_{2d_{TDCP_k}} - \hat{v}_{2d_k} \\ v_{d_{TDCP_k}} - \hat{v}_{d_k} \\ \widetilde{\delta\theta}_k - \hat{\delta\theta}_k \end{bmatrix} \tag{19}$$

where,

$$v_{2d_{TDCP_k}} = \sqrt{v_{n_{TDCP_k}}^2 + v_{e_{TDCP_k}}^2} \tag{20}$$

$$\widetilde{\delta\theta}_k = arctan\left(\frac{v_{e_{TDCP_k}}}{v_{n_{TDCP_k}}}\right) - \theta_{MAGYQ_k} \tag{21}$$

**(2) *Step* update** The step frequency is around 2 Hz, which is lower than the 5 Hz filter prediction frequency. Since we assume that the horizontal velocity remains constant during the step interval, the horizontal velocity derived from the step length at the current step instant can be used for the incoming epochs until the next step instant. The innovation on horizontal velocity provided by the *Step* update can be written as:

$$inno_{step} = \frac{S_k}{\Delta t_{step}} - \hat{v}_{2d_k} \tag{22}$$

where $S_k$ is the PDR step length at epoch $k$ calculated according to the model in [25], and $\Delta t_{step}$ is the time interval between two steps.

**(3) *Barometer* update** The ambient air pressure and temperature measurements can be used to correct the vertical velocity:

$$inno_{baro} = \frac{\Delta h_{baro}}{\Delta t} - \hat{v}_{d_k} \tag{23}$$

where $\Delta h_{baro}$ is the height difference between 2 consecutive epochs. It is a function of pressure and temperature [28]:

$$\Delta h_{baro} = \frac{T_{k-1}}{T_{grad}}\left(1 - \left(\frac{P_k}{P_{k-1}}\right)^{T_{grad}\cdot\frac{G_c}{g}}\right) \tag{24}$$

Where $T_{k-1}$ is the measured temperature at the previous epoch. $T_{grad}$ is the temperature change rate over altitude, which is a constant with the value of $0.0065K/m$. $P_k$ and $P_{k-1}$ are the measured air pressures at current and previous epochs respectively. $G_c$ is the gas constant ($287.052J/K.kg$). $g$ is the gravitational acceleration. Here we compute the relative altitude with the pressure measurements at current and previous epochs, to reduce the effects of measurement noise and bias in the barometer measurements.

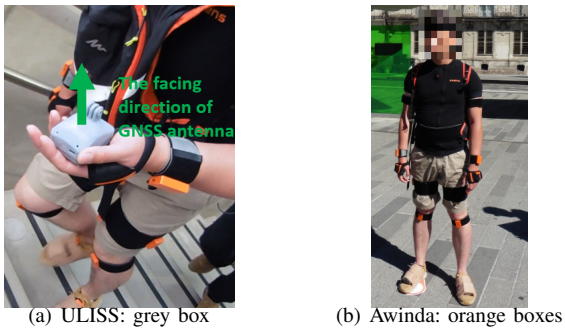
## V. Results and Discussions

### A. Experimentation setup

*1) Equipment:* A wearable device "ULISS" (Ubiquitous Localization with Inertial Sensors and Satellites) was used for data collection (Fig. 12 (a)). It contains a GNSS receiver (UBlox F9P, GPS+Galileo, dual-frequency, sampling rate 5 Hz), inertial sensors and magnetometer (Xsens Mti-7, sampling rate 200 Hz), and a barometer (Bosch BMP280, sampling rate 200 Hz). The ULISS device was held in a "texting" position, which means the device was held in hand as if texting a message on the phone. The facing direction of the GNSS antenna inside of

the ULISS device is perpendicular to the device's main surface, i.e., in the case of texting, the GNSS antenna is always facing toward the zenith.

The reference solution is provided by the Xsens MTw Awinda system, which is a set of seventeen high-precision full-body motion-capture IMU sensors (orange boxes in Fig. 12), with a 60 Hz sampling frequency. The advantage of using such a system is that the reference information of every single body part can be retrieved. The hand velocity and trajectory are extracted from the "Xsens MVN Analyze" software, and then intensively post-processed with map-matching to obtain highly accurate references.




(a) ULISS: grey box (b) Awinda: orange boxes

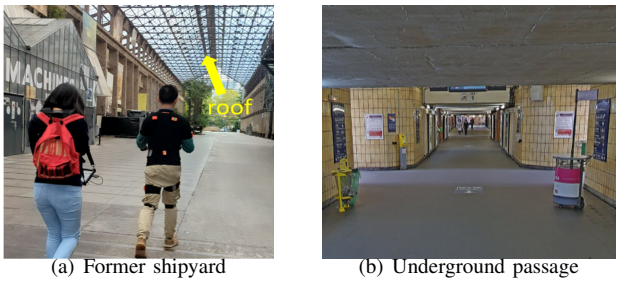
Fig. 12. Experimental devices

*2) Tested trajectories:* 4 trajectories are tested in this paper by covering diverse challenging environments: two in a shopping mall (light indoor), one in a railway station including halls, mezzanine and underground corridor (light and deep indoor), as well as one in deep urban environments through a former shipyard with a triangle-shaped roof (urban and light indoor).

More details of the tested environments can be found in Fig. 2 and Fig. 13. For the shopping mall, see Fig. 2 (a), the subject was walking on the path below the edge of the glass roof, surrounded by trees, pillars, and also some other obstacles like walking customers, escalators, statues, etc. In the mezzanine of the railway station, obstacles like artificial trees, glass-structured shops, panels, and also walking travelers can be seen in Fig. 2 (b). Fig. 2 (d) shows a classic deep urban environment with narrow streets surrounded by tall buildings and trees. Inside the former shipyard building, there were not many obstacles as shown in Fig. 13 (a) but the roof is split into small pieces by the metal structures. The underground passage of the railway station is shown in Fig. 13 (b).

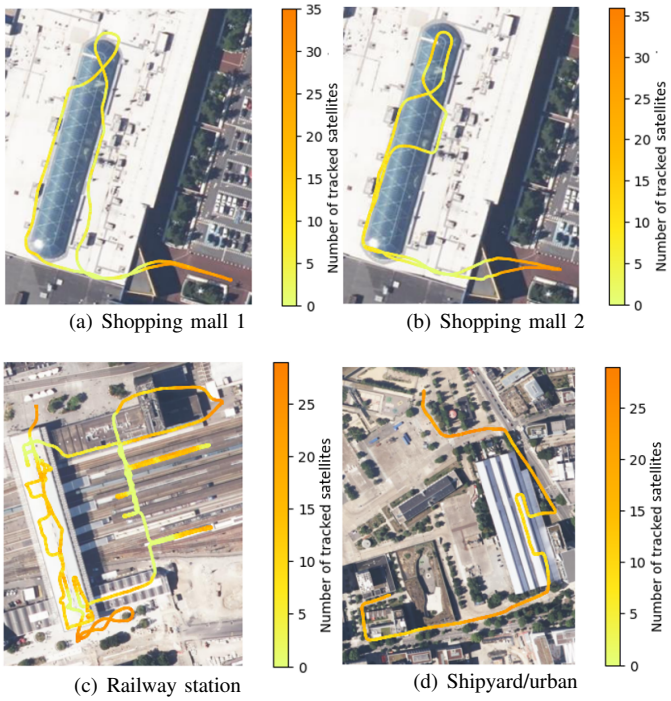
Fig. 14 depicts the ground truth of the 4 tests, as well as the number of tracked satellites along the trajectory. Thanks to the background satellite image, we can easily identify the glass roof of the shopping mall (in dark blue of Fig. 14 (a)(b)), the mezzanine of the railway station (enclosed waiting room that crosses the rail tracks, big white rectangle on the left of Fig. 14 (c)), and the roof of the shipyard (big lilac rectangle on the right of Fig. 14 (d)) respectively. We can also identify that the number of tracked satellites in all light indoor scenarios is far less than outdoors (light yellow trajectory instead of dark orange). The number of tracked satellites in the shopping mall is even lower than that in the other two light indoor scenarios.

Moreover, the min, max, and mean $C/N_0$ evolution of all




(a) Former shipyard (b) Underground passage

Fig. 13. Details of the former shipyard and underground environments




(a) Shopping mall 1 (b) Shopping mall 2

(c) Railway station (d) Shipyard/urban

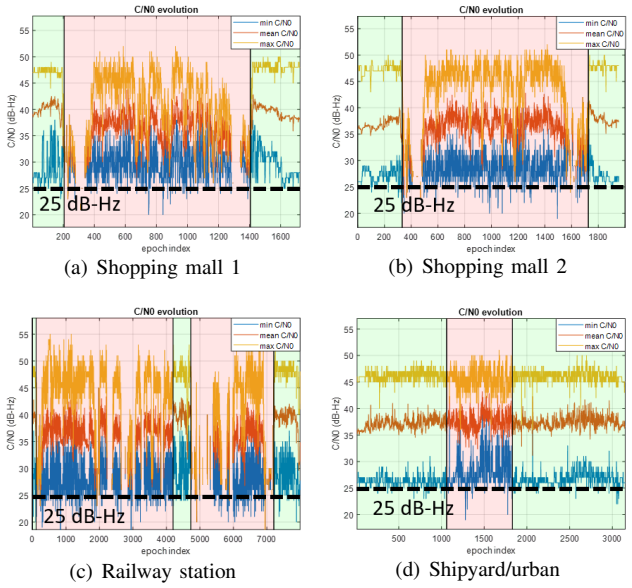
Fig. 14. Visualization of the ground truths and the number of tracked satellites along the tested trajectories *(Background satellite image source: Géoportail France)*

tracked satellites over time for the 4 datasets are studied in Fig. 15. The indoor periods (both light indoor and deep indoor) are highlighted by pink backgrounds, contrary to the outdoor periods highlighted by green backgrounds. Several pieces of information can be read from Fig. 15: 1) the discontinuities in the curves imply that no satellite was tracked, i.e. the deep indoor scenario. For the shopping mall datasets, we also observe the $C/N_0$ discontinuities just after entering the building and just before exiting the building, i.e., the indoor/outdoor transition area. 2) For most of the time, the minimum $C/N_0$ of tracked satellites is larger than or equal to 25 dB-Hz, even though a few spikes that $C/N_0$ is less than 25 dB-Hz can be observed. This phenomenon is consistent with the common second-order PLL tracking performance [16]. (3) The $C/N_0$ curves are more fluctuated indoors than outdoors. (4) The attenuation in indoor spaces compared with outdoors is visible through the magnitude of $C/N_0$ curves.

The synthesis of 4 tested trajectories regarding the environment natures, total walking distance, total duration and light

TABLE V
SYNTHESIS OF TESTED TRAJECTORIES

| Dataset | Environments | Vertical levels | Total walking distance | Total duration | Light indoor duration | Mean $C/N_0$ of tracked satellites outdoors | Mean $C/N_0$ of tracked satellites light indoors | Average number of tracked satellites light indoors |
|---|---|---|---|---|---|---|---|---|
| Shopping mall 1 | Semi-open-sky, light indoor | Single floor | 419 m | 5'44" | 4'6" | 39.7 dB-Hz | 35.4 dB-Hz | 4.82 |
| Shopping mall 2 | Semi-open-sky, light indoor | Single floor | 415 m | 5'39" | 4'45" | 37.2 dB-Hz | 35.9 dB-Hz | 6.47 |
| Railway station | Semi-open-sky, light indoor, deep indoor | Multi-floor, with stairs and ramps | 1545 m | 26' | 18'13" | 39.3 dB-Hz | 35.6 dB-Hz | 9.69 |
| Urban/ ship-yard | Semi-open-sky, deep urban, light indoor | Single floor | 810 m | 9'53" | 2'37" | 37.3 dB-Hz | 37.6 dB-Hz | 9.91 |




(a) Shopping mall 1 (b) Shopping mall 2

(c) Railway station (d) Shipyard/urban

Fig. 15. Min, max and mean C/N0 evolution of the tracked satellites for four tested datasets *(Blue: min C/N0, yellow: max C/N0, red: mean C/N0. Green background: outdoor period, pink background: indoor period)*

indoor duration, mean outdoors $C/N_0$, mean light indoors $C/N_0$, and the average number of tracked satellites light indoors is given in Table V. According to the last three columns of Table V, we can conclude that the "shopping mall 1" dataset is the most challenging one, since it has the fewest average tracked satellites light indoors, and also the most attenuation from outdoors to light indoors. For the "urban/shipyard" dataset, the mean outdoor and light indoor $C/N_0$ do not change much, since there are not many obstacles in the shipyard building in addition to its roof.

The first author is the human subject who carried out the experiments in this paper. The railway station handheld dataset used in this paper was collected simultaneously with one of the foot-mounted datasets for Indoor Positioning and Indoor Navigation (IPIN) 2022 competition track 4 [29].

### *B. Performance evaluation of the LIGHT algorithm*

First of all, as we can see in Table VI, even in the dataset for the shopping mall 1, where less than 5 satellites can be

TABLE VI
USABLE LIGHT INDOOR EPOCHS FOR THE LIGHT ALGORITHM

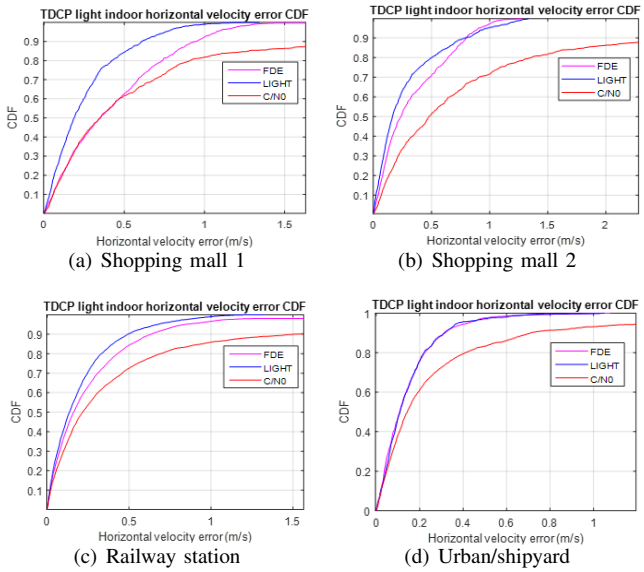
| Dataset | Usable epochs | Total light indoor epochs | Usable rate |
|---|---|---|---|
| Shopping Mall 1 | 690 | 1230 | 56.10% |
| Shopping Mall 2 | 996 | 1428 | 69.75% |
| Railway Station | 4061 | 5467 | 74.28% |
| Shipyard | 792 | 796 | 99.50% |

tracked on average, at least half of the epochs can be used for GNSS TDCP positioning. Associating with Table V, it can be seen that the GNSS usable rate depends on the average number of tracked satellites: the more satellites tracked, the more epochs can be used. It also depends on the mean $C/N_0$ value of all tracked satellites. For example, the railway station and the shipyard datasets have a similar average number of satellites, but more usable epochs in the shipyard where the mean $C/N_0$ is higher.

Since the LIGHT approach aims at selecting healthy carrier phase measurements, benchmark comparisons are made with two non-ML-based satellite selection approaches. The first baseline method is to select satellites through a hard $C/N_0$ threshold. The threshold here is fixed at 33dB-Hz, which is determined through the $C/N_0$ distribution in Fig. 6. Another one uses the state-of-the-art hypothesis test-based Fault Detection Exclusion (FDE) [30] to remove the satellites with faulty measurements. It uses the EKF innovation normalized by its variance as the test statistic, to iteratively remove the faulty measurements [30]. For simplicity, the three methods for comparison will be designated as "LIGHT", "$C/N_0$", and "FDE" respectively in the following.

Fig. 16 shows the Cumulative Density Function (CDF) of the velocity errors, where our proposed LIGHT algorithm outperforms both $C/N_0$ and FDE approaches. It achieves significant improvements in terms of velocity accuracy especially in light indoor environments. Table VII reports the 75% velocity error for each method. By applying the LIGHT approach, the velocity accuracy can improve by up to 88% compared with the state-of-the-art FDE, and up to 251% compared with the $C/N_0$ thresholding baseline method.

It proves that the $C/N_0$ thresholding approach provides the




(a) Shopping mall 1 (b) Shopping mall 2

(c) Railway station (d) Urban/shipyard

Fig. 16. TDCP light indoor horizontal velocity error CDF for LIGHT, C/N0 threshold, and FDE approaches *(blue: LIGHT, magenta: FDE, red: C/N0)*

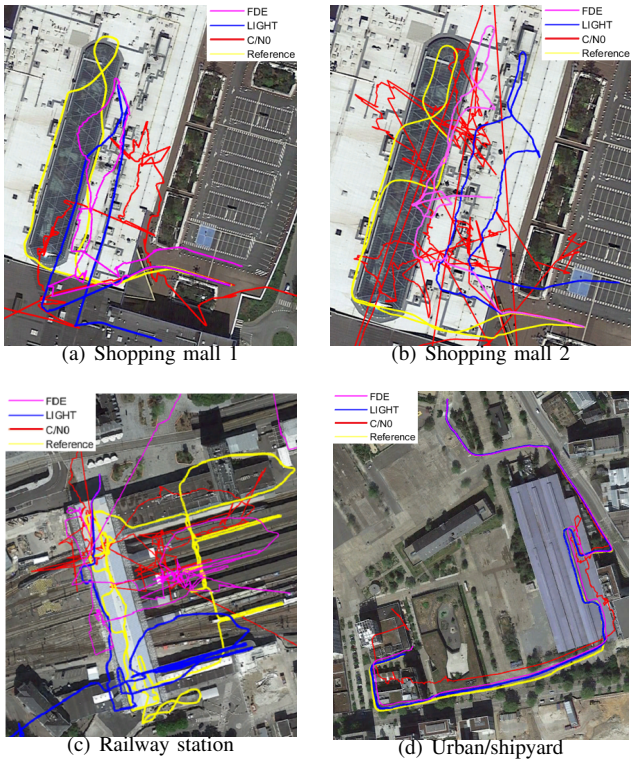
TABLE VII
75% TDCP LIGHT INDOOR HORIZONTAL VELOCITY ERROR CDF FOR LIGHT AND BENCHMARK APPROACHES

| Dataset | Benchmark errors | | LIGHT error | Accuracy improvement |
|---|---|---|---|---|
| **Shopping mall 1** | $C/N_0$ | 0.78 m/s | 0.35 m/s | 122.86% |
| | FDE | 0.66 m/s | | 88.57% |
| **Shopping mall 2** | $C/N_0$ | 1.09 m/s | 0.31 m/s | 251.61% |
| | FDE | 0.46 m/s | | 48.39% |
| **Railway station** | $C/N_0$ | 0.55 m/s | 0.28 m/s | 96.43% |
| | FDE | 0.37 m/s | | 32.14% |
| **Urban/shipyard** | $C/N_0$ | 0.33 m/s | 0.20 m/s | 65.0% |
| | FDE | 0.20 m/s | | 0% |

worst results. Even by selecting the satellites with relatively higher $C/N_0$ values, it is still very risky to come across bad measurements, therefore the signal strength cannot reflect the quality of the measurements in challenging environments. This also justifies the necessity of finding multiple features, as done in the LIGHT approach.

The FDE approach, which is measurement redundancy-based, can partly identify the bad measurements in light indoors. Its performance highly depends on the number of tracked satellites because the more satellites are tracked, the higher the redundancy. The Urban/shipyard dataset which has the highest average number of tracked satellites and the lowest indoor/outdoor attenuation makes the FDE approach achieve similar velocity accuracy as the LIGHT approach, as in Fig. 16 (d). However, the Shopping mall 1 dataset in which only in average 5 satellites are tracked light indoors, the performance of FDE is far behind that of LIGHT, see Fig. 16 (a).

The estimated trajectories using GNSS standalone for all tested satellite selection approaches are shown in Fig. 17. Since the $C/N_0$ approach provides the worst velocity estimates, the trajectories which were the integration of velocity also have rapid drifts, as the trajectories for shopping mall datasets (red trajectories in Fig. 17 (a)(b)) and for railway station dataset




(a) Shopping mall 1 (b) Shopping mall 2

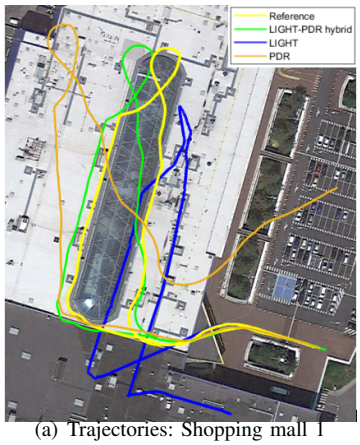
(c) Railway station (d) Urban/shipyard

Fig. 17. TDCP-only trajectories for LIGHT, $C/N_0$ threshold, and FDE approaches for tested datasets *(blue: LIGHT, magenta: FDE, red: $C/N_0$, yellow: reference) (Background satellite image source: Google Earth)*

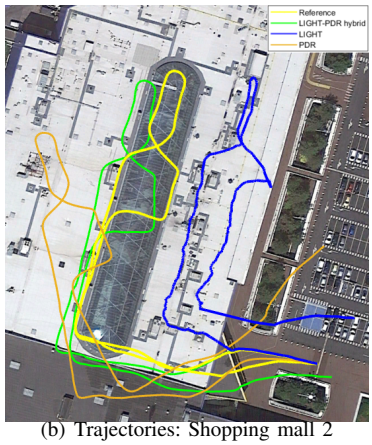
(red trajectory in Fig. 17 (c)) do not make too much sense. In the most favorable urban/shipyard dataset, the red trajectory in Fig. 17 (d) is also strongly perturbed. In general, the proposed LIGHT algorithm can effectively correct the heading estimation when the carrier phase measurements are available. This can be clearly seen in the shopping mall datasets (Fig. 17 (a)(b)). Also, the LIGHT trajectories are less biased compared with the FDE trajectories in the more favorable railway station and urban/shipyard datasets, as shown in Fig. 17 (c)(d).

However, during the deep indoor period and the transition phase between indoor and outdoor, either no TDCP measurement is selected or no satellite is tracked, the LIGHT solution is not continuous. It can be identified in the blue trajectories of Fig. 17: the trajectories of deep indoor and transition parts are missing: the bottom part of the shopping mall trajectories are shorter (see Fig. 17 (a)(b)), and the trajectory on the railway station platforms overlap together (see Fig. 17 (c)). This justifies the need for the hybridization with other sensors.

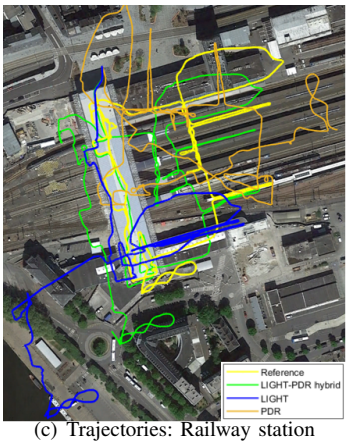
In the end, the computational performance of the LIGHT algorithm on four tested trajectories is summarized in Table VIII. The evaluation is performed on a Dell Latitude 5520 computer equipped with an Intel i7-1165G7 (2.80 GHz) processor and a RAM of 32 GB. Matlab R2021b is used for running the code, with the interaction of the pre-trained ML model from Python at each epoch. The mean processing time per epoch of all tested datasets is lower than the sampling interval (0.2 s for 5 Hz data frequency), which is feasible for real-time implementation.



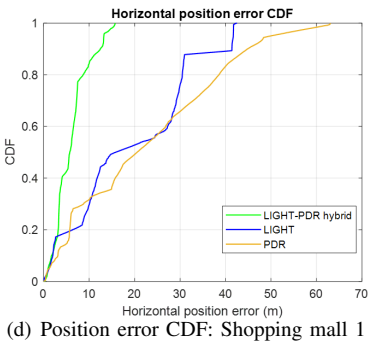

(a) Trajectories: Shopping mall 1



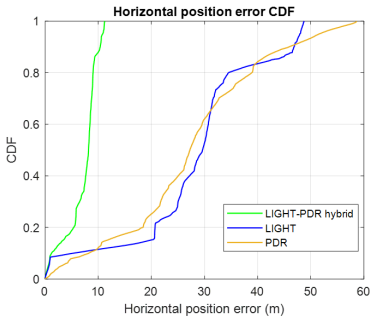

(b) Trajectories: Shopping mall 2



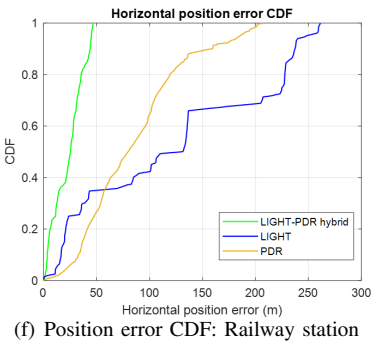

(c) Trajectories: Railway station




(d) Position error CDF: Shopping mall 1




(e) Position error CDF: Shopping mall 2




(f) Position error CDF: Railway station

Fig. 18. Trajectories and position error CDF for shopping mall 1&2 and railway station datasets. *(green: LIGHT-PDR hybrid, blue: LIGHT-only, orange: PDR-only, yellow: reference) (Background satellite image source: Google Earth)*

TABLE VIII
COMPUTATIONAL PERFORMANCE OF LIGHT ALGORITHM ON THE TESTED TRAJECTORIES

| Dataset | Total epochs | Total run time | Mean run time per epoch | Max CPU consumption | Max RAM consumption |
|---|---|---|---|---|---|
| Shopping Mall 1 | 1720 | 9.04 s | 0.0053 s | 43.5% | 0.5% (∼160 MB) |
| Shopping Mall 2 | 1698 | 8.69 s | 0.0051 s | 44.1% | 0.4% (∼128 MB) |
| Railway Station | 7803 | 41.37 s | 0.0053 s | 39.6% | 0.9% (∼288 MB) |
| Urban/ Shipyard | 2967 | 19.25 s | 0.0065 s | 38.7% | 1.0% (∼320 MB) |


Fig. 19. Setup of the ULISS device and Xiaomi Mi 8 smartphone

### C. Performance evaluation of the LIGHT-PDR filter

As in Fig. 18 (a)(b)(c), the LIGHT solution (in blue) drifts when no usable GNSS carrier phase measurement is selected, and the inertial-only heading drift on the PDR-only trajectory (in orange) is visible. However, with the LIGHT-PDR hybridization filter, seamless indoor-outdoor positioning can be performed more robustly: the hybrid trajectory (in green) is continuous, and the heading is also successfully corrected by LIGHT when the TDCP update is available.

Therefore, the final position accuracy is significantly improved, as shown in the position errors CDF in Fig. 18 (d)(e)(f), and Table IX. For example, the LIGHT-PDR hybridization can achieve a <9 m position accuracy for the two shopping mall

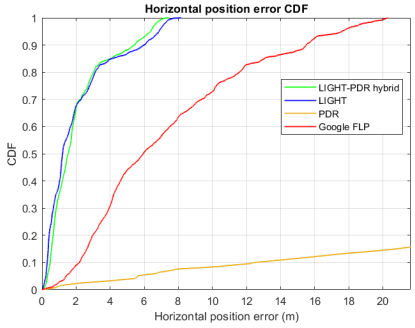



Fig. 20. Position error CDF for Urban/shipyard dataset *(green: LIGHT-PDR hybrid, blue: LIGHT-only, orange: PDR-only, yellow: reference, red: Google FLP) (Background satellite image source: Google Earth)*

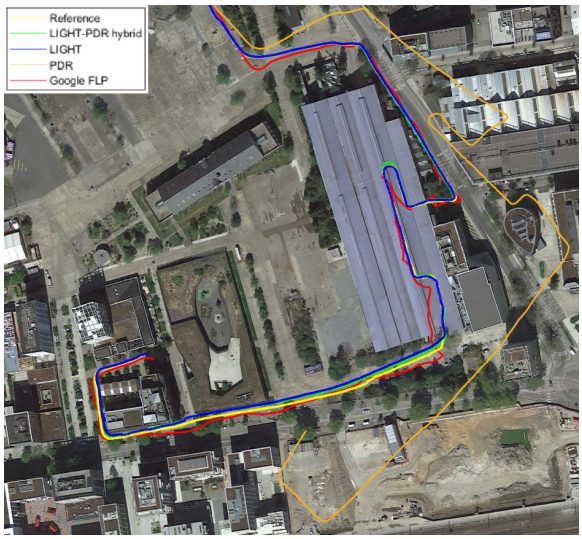


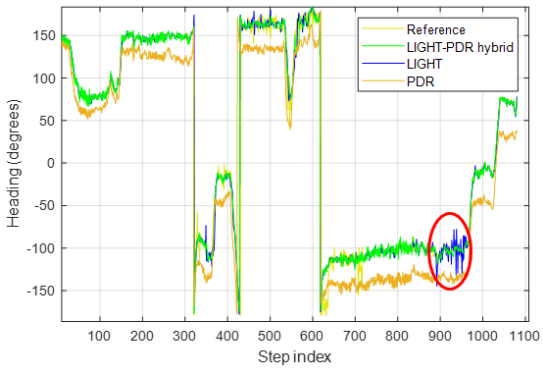

Fig. 21. Trajectories for Urban/shipyard dataset *(green: LIGHT-PDR hybrid, blue: LIGHT-only, orange: PDR-only, yellow: reference, red: Google FLP) (Background satellite image source: Google Earth)*

datasets, while the accuracy of the LIGHT-only and PDR-only solutions is at least 30 m. The proposed LIGHT-PDR hybrid filter takes advantage of both individual filters in a complementary way to provide a seamless positioning solution in all challenging environments.

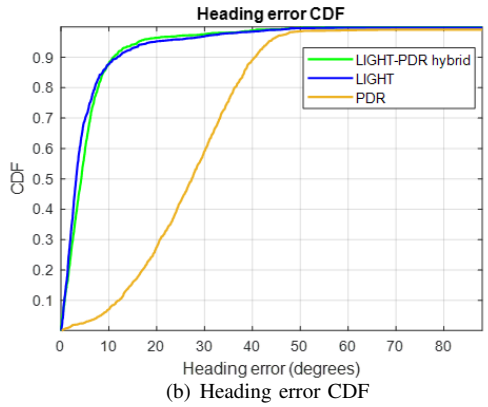
As for the urban/shipyard dataset, the hybridization performance is similar to that of the LIGHT-only solution. However, we can still see some minor differences in the CDF as shown in Fig. 20: the hybridization solution outperforms the LIGHT-only thanks to its capability of eliminating huge error in high percentile CDF.

In addition, another comparison with the trajectory provided by Google Fused Location Provider (FLP) of a Xiaomi Mi 8 smartphone was performed for this dataset. The Google FLP fuses four-constellation GNSS, IMU, magnetometer measurements in the smartphone and potentially Wifi information to provide the navigation solution. We are going to compare the FLP solution with our two-constellation GNSS-only LIGHT solution, as well as the LIGHT-PDR hybrid solution. Both ULISS and the smartphone are in the "texting" position, as shown in Fig. 19. When computing the position error, the distance between ULISS and the smartphone was removed. Our solution (either LIGHT-only and LIGHT-PDR hybrid) outperforms the FLP solution, as shown in the trajectory (Fig. 21) and the CDF plot (Fig. 20). 2.67 m and 2.77 m 75% positioning accuracy for LIGHT-only and LIGHT-PDR solutions respectively are achieved, while the FLP solution has a positioning accuracy of 10.23 m. We can also observe that in Fig. 21, the FLP trajectory drifts strongly compared to LIGHT-only or LIGHT-PDR hybrid solution, especially during the light indoor period inside the shipyard.

Regarding the heading correction performance brought from the LIGHT solution, let us take an example of the urban/shipyard dataset. Fig. 22 (a) compares the heading (in degrees) obtained by each method (LIGHT, PDR, LIGHT-PDR hybrid) with the reference for each step instant over time. The PDR heading starts drifting from the beginning, while the TDCP (LIGHT) heading can effectively correct it ensuring the overall hybridization heading remains correct. Highlighted by the red circle in Fig. 22 (a), the LIGHT-PDR hybridization filter can also denoise the LIGHT heading in some extreme cases where the LIGHT heading has too many oscillations. The heading error CDF is plotted in Fig. 22 (b). For LIGHT and LIGHT-PDR hybrid methods, the 75% heading error CDF is 6.04° and 6.85° respectively, while PDR has a 75% heading error of 34.70°.




(a) Heading comparison




(b) Heading error CDF

Fig. 22. Heading comparison and heading error CDF for urban/shipyard dataset *(green: LIGHT-PDR hybrid, blue: LIGHT-only, orange: PDR-only, yellow: reference)*

There are still some limitations of the hybridization solution when there is no available carrier phase selected by LIGHT. In Fig. 18 (a)(b), the bottom part of the hybrid trajectory is slightly longer than the reference, because no available TDCP updates could be applied at indoor/outdoor transition. The hybrid filter only relied on the PDR solution. The errors from the PDR step length estimation introduced a scale factor in the estimation of the trajectory. Also, in Fig. 18 (c), the hybrid heading cannot be sufficiently corrected during the deep indoors (underground) stage where there is no GNSS at all.

## VI. Conclusion and Perspectives

GNSS carrier phase measurement is a powerful tool for positioning. With the help of artificial intelligence, it becomes possible to use GNSS carrier phase even in challenging environments. In this paper, a ML model, i.e., LIGHT, is trained to select healthy GNSS carrier phase measurements

TABLE IX
75% HORIZONTAL POSITION ERROR CDF OF LIGHT-PDR HYBRID FILTER, LIGHT-ONLY, PDR-ONLY AND GOOGLE FLP

| Dataset | PDR | LIGHT | LIGHT-PDR hybrid | Google FLP |
|---|---|---|---|---|
| Shopping Mall 1 | 36.03 m | 29.87 m | 7.43 m | - |
| Shopping Mall 2 | 35.70 m | 32.96 m | 8.94 m | - |
| Railway Station | 111.60 m | 225.27 m | 33.64 m | - |
| Urban/ shipyard | 117.43 m | 2.67 m | 2.77 m | 10.23 m |

among a majority of unhealthy measurements to improve the navigation solution. The texting carrying mode is investigated in different challenging environments, i.e., light indoor and deep urban. Even with quite few tracked satellites, LIGHT can provide significant improvements in terms of TDCP velocity estimation accuracy compared to the traditional non-ML-based algorithms. For example, in the most challenging shopping environment, the 75% velocity error CDF drops from 0.66 m/s and 0.78 m/s for the state-of-the-art FDE approach and the baseline $C/N_0$ thresholding approach respectively, to 0.35 m/s with the LIGHT algorithm, i.e., up to 88% of improvement compared with FDE method was achieved, and up to 255% of improvement compared with C/N0 method was achieved.

What is more, the LIGHT-PDR hybridization filter is proposed to further improve the pedestrian trajectory by integrating Magneto-Inertial Measurement Unit (MIMU) and barometer. For the shopping mall datasets, a $<$9m position accuracy can be achieved through LIGHT-PDR hybridization, while the accuracy before hybridization is at least 30m. For the urban/shipyard dataset, both LIGHT and LIGHT-PDR solutions can provide a $<$3m position accuracy, while the accuracy of the state-of-the-art Google FLP solution is larger than 10m.

The main limitation of the current LIGHT algorithm is in the transition phase between indoor and outdoor, when either no useful carrier phase measurements are selected, or no satellites are tracked. This leads to the non-continuity of the LIGHT solution. By using the PDR information, the LIGHT-PDR hybridization filter fills in the missing portion of the LIGHT-only trajectory, and the PDR heading drift is corrected by the velocity estimated using LIGHT. Future work will improve the LIGHT algorithm during the transition phase between indoor and outdoor to achieve a more continuous solution.

## REFERENCE


[1] G. Lachapelle, "GNSS Indoor Location Technologies," *Journal of Global Positioning Systems*, vol. 3, no. 1-2, pp. 2–11, 2004.

[2] G. Lachapelle and P. Gratton, "GNSS Precise Point Positioning with Android Smartphones and Comparison with High Performance Receivers," *2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, pp. 1–9, 2019.

[3] A. Angrisano, M. Vultaggio, S. Gaglione, and N. Crocetto, "Pedestrian localization with PDR supplemented by GNSS," *2019 European Navigation Conference (ENC)*, pp. 1–6, 2019.

[4] Z. Li, N. Zhu and V. Renaudin, "Velocity Protection Level for Wearable Devices on TDCP-based Pedestrian Navigation," *2022 International Conference on Localization and GNSS (ICL-GNSS)*, 2022.

[5] Z. Li, N. Zhu and V. Renaudin, "LIGHT-PDR: Light Indoor GNSS Carrier Phase Positioning with Machine Learning and Inertial Signal Fusion for Pedestrian Navigation," *2023 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2023.

[6] T. Lin, M. Ma, A. Broumandan and G. Lachapelle, "Demonstration of a high sensitivity GNSS software receiver for indoor positioning," *Advances in Space Research*, vol. 51, no. 6, pp. 1035–1045, 2013.

[7] Z. He, V. Renaudin, M. G. Petovello and G. Lachapelle, "Use of High Sensitivity GNSS Receiver Doppler Measurements for Indoor Pedestrian Dead Reckoning," *Sensors*, vol. 13, no. 4, pp. 4303–4326, 2013.

[8] X. Li, "GNSS Repeater Based Differential Indoor Positioning With Multi-Epoch Measurements," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 803–813, 2023.

[9] I. Selmi, N. Samama and A. Vervisch-Picois, "A new approach for decimeter accurate GNSS indoor positioning using Carrier Phase Measurements," *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–6, 2013.

[10] A. Uzun, F. A. Ghani, A. M. Ahmadi Najafabadi, H. Yenigun and I. Tekin, "Indoor Positioning System Based on Global Positioning System Signals with Down- and Up-Converters in 433 MHz ISM Band," *Sensors*, vol. 21, no. 13, 2021.

[11] T. Suzuki and Y. Amano, "NLOS Multipath Classification of GNSS Signal Correlation Output Using Machine Learning," *Sensors*, vol. 21, no. 7, 2021.

[12] H. Xu, A. Angrisano, S. Gaglione and L.-T. Hsu, "Machine learning based LOS/NLOS classifier and robust estimator for GNSS shadow matching," *Satellite Navigation*, vol. 1, no. 15, 2020.

[13] Y. Zhu, H. Luo, F. Zhao, et al, "A Fast Indoor/Outdoor Transition Detection Algorithm Based on Machine Learning," *Sensors*, vol. 19, no. 786, 2019.

[14] I. Sbeity, C. Villien, B. Denis, E. V. Belmega, "RNN-Based GNSS Positioning using Satellite Measurement Features and Pseudorange Residuals," *2023 International Conference on Localization and GNSS (ICL-GNSS)*, 2023.

[15] E. D. Kaplan and C. J. Hegarty, *Understanding GPS: Principles and Applications, Second Edition*. ARTECH HOUSE, INC., 2006.

[16] M. Irsigler and B. Eissfeller, "PLL tracking performance in the presence of oscillator phase noise," *GPS solutions*, vol. 5, no. 4, pp. 45–57, 2002.

[17] G. Hein, M. Paonni, V. Kropp and A. Teuber, "GNSS Indoors: Fighting the Fading (Part 2)," *Inside GNSS*, vol. May/June, pp. 47–53, 2008.

[18] G. Hein, M. Paonni, V. Kropp and A. Teuber, "GNSS Indoors: Fighting the Fading (Part 1)," *Inside GNSS*, vol. March/April, pp. 43–52, 2008.

[19] L.-T. Hsu, "GNSS multipath detection using a machine learning approach," *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6, 2017.

[20] Y. Sun and L. Fu, "Stacking Ensemble Learning for Non-Line-of-Sight Detection of Global Navigation Satellite System," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, April 2023.

[21] European Space Agency, "GNSS Basic Observables (online)." `https://gssc.esa.int/navipedia/index.php/GNSS_Basic_Observables`, 2011.

[22] M. Irsigler, "Characterization of multipath phase rates in different multipath environments," *GPS Solutions*, vol. 14, no. 4, pp. 305–317, 2010.

[23] European Space Agency, "Examples of single frequency Cycle-Slip Detectors (online)." `https://gssc.esa.int/navipedia/index.php/Examples_of_single_frequency_Cycle-Slip_Detectors`, 2011.

[24] Z. Zhu and E. Vinande and M. U. De Haag, "Multi-constellation time-differenced carrier phase solution with protection from multiple failures," *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pp. 336–348, 2018.

[25] V. Renaudin and C. Combettes, " Magnetic, Acceleration Fields and Gyroscope Quaternion (MAGYQ) Based Attitude Estimation with Smartphone Sensors for Indoor Pedestrian Navigation," *Sensors*, vol. 14, no. 12, pp. 22864–22890, 2014.

[26] N. A. Abiad, Y. Kone, V. Renaudin and T. Robert, "Smartstep: A Robust STEP Detection Method Based on SMARTphone Inertial Signals Driven by Gait Learning," *IEEE Sensors Journal*, vol. 22, no. 12, pp. 12288–12297, 2022.

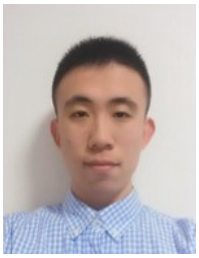
[27] V. Renaudin, M. Susi and G. Lachapelle, "Step Length Estimation Using Handheld Inertial Sensors," *Sensors*, vol. 12, no. 7, pp. 8507–8525, 2012.

[28] K.-W. Chiang, H.-W. Chang, Y.-H. Li, et al, "Assessment for INS/GNSS/Odometer/Barometer Integration in Loosely-Coupled and Tightly-Coupled Scheme in a GNSS-Degraded Environment," *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3057–3069, 2020.

[29] F. Potortì et al, "Offsite evaluation of localization systems: criteria,

systems and results from IPIN 2021–22 competitions," *IEEE Journal of Indoor and Seamless Positioning and Navigation*, vol. 2, pp. 92–129, 2024.

[30] N. Zhu, J. Marais, D. Bétaille, and M. Berbineau, "Evaluation and Comparison of GNSS Navigation Algorithms Including FDE for Urban Transport Applications," in *Proceedings of the 2017 International Technical Meeting of The Institute of Navigation*, pp. 51–69, 2017.



**Ziyou Li** received the M.Sc. degree in aeronautic navigation and telecommunication from the French Civil Aviation University (Ecole Nationale d'Aviation Civile, ENAC), Toulouse, France, in 2021. He is currently working toward the Ph.D. degree in GNSS pedestrian navigation and integrity monitoring in challenging environments with GEOLOC Laboratory, University Gustave Eiffel, Bouguenais, France.

**Ni Zhu** (Member, IEEE) received the Engineering degree in aeronautic telecommunications from the French Civil Aviation University (Ecole Nationale d'Aviation Civile, ENAC), Toulouse, France, in 2015, and the Ph.D. degree in science of information and communication from the University of Lille, Lille, France, in 2018. She is currently a Research Fellow with the Laboratory GEOLOC, University Gustave Eiffel, Bouguenais, France. Her research interests include specialization in GNSS channel propagation modeling in urban environments, positioning integrity monitoring for terrestrial safety-critical applications, and multisensory fusion techniques for indoor/outdoor pedestrian positioning assisted by artificial intelligence. Since 2020, she has been the Co-Chair of the foot-mounted IMU-based positioning track of indoor positioning and indoor navigation competition.

**Valérie Renaudin** (Member, IEEE) received the M.Sc. degree in geomatics engineering from the École Supérieure d'ingénieurs Géomètres et Topographes, Le Mans, France and the Ph.D. degree in computer, communication, and information sciences from the École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, in 1999 and 2009, respectively. She is currently a Professor with Gustave Eiffel University, Bouguenais, France. She was the Technical Director with SWISSAT, Schwyz, Switzerland, where she developed real-time positioning solutions based on a permanent global network of satellite navigation systems (GNSS), and a Senior Research Associate with the University of Calgary, Calgary, AB, Canada. She currently heads the Geopositioning Laboratory, Gustave Eiffel University, where she has built a team specializing in the positioning and navigation of travelers. Her research interests include indoor/outdoor navigation methods and systems using GNSS, as well as inertial and magnetic data, especially for pedestrians to improve sustainable personal mobility. Dr. Ranaudin was the recipient of several awards, including a Marie Curie European Grant for smartWALK project. She founded NAV4YOU, Bouguenais, France, in 2021, developing location-based services for the safety of firefighters in intervention, defence, and underground activities. She is the Editor-in-Chief of the new open-access IEEE JOURNAL OF INDOOR AND SEAMLESS POSITIONING AND NAVIGATION that she launched in 2022. She is also a member of the steering committee of the international conference "Indoor Positioning and Indoor Navigation."