Dense Ground Truth for Indoor Localization Competitions: Foot-mounted IMU-Enhanced Evaluation

Antonio R. Jiménez[1,*,†], Joaquín Torres-Sospedra[2,3,†] and Luisa Ruiz-Ruiz[1]

[1]*Centre for Automation and Robotics (CSIC-UPM), Ctra. Campo Real km. 0,2, 28500, Arganda del Rey (Madrid), Spain*
[2]*Department of Computer Science, Universitat de València, Avda. Universitat s/n, 46100 Burjassot, Valencia, Spain*
[3]*VALGRAI, Valencian Graduate School and Research Network of Artificial Intelligence, Campus de Vera, 46022 Valencia, Spain*

Abstract

Indoor localization competitions, such as IPIN Track 3, are crucial for benchmarking smartphone-based localization solutions. However, their current evaluation relies on a limited number of sparse, manually placed, and individually georeferenced ground truth (GT) points. This deployment process is costly and labor-intensive, and the inherent GT sparsity restricts evaluation granularity. It frequently obscures performance nuances and significant position estimation errors that occur between sparse points, consequently limiting comprehensive scoring. This paper proposes a novel method to generate a dense, high-fidelity GT trajectory for competition evaluation without requiring additional manual GT deployment. Our approach fuses a dense, relative trajectory, derived from a foot-mounted Inertial Measurement Unit (IMU) using Zero-Velocity Updates (ZUPT), with the existing sparse, highly accurate surveyed GT points. This fusion is achieved through a robust segment-wise rigid alignment method, precisely translating and rotating individual trajectory segments, followed by a final global trajectory smoothing. The resulting dense GT enables a more granular and continuous evaluation of competitor trajectories at their native IMU output frequency (e.g., 100 Hz). This offers a more comprehensive and fair assessment, providing enhanced diagnostic capabilities. We outline the methodology, discuss key implementation considerations, and propose an evaluation strategy for the generated GT, highlighting its potential to significantly enhance future indoor localization benchmarks.

Keywords

IPIN competition, Track 3, Foot-mounted IMU, Ground-truth generation

1. Introduction

Accurate indoor localization is a critical capability for a wide range of applications, from emergency services and autonomous robotics to augmented reality and asset tracking. Unlike Global Navigation Satellite Systems (GNSS), which are ubiquitous outdoors, indoor positioning faces unique challenges due to signal propagation issues (e.g., multipath, attenuation) and the absence of reliable infrastructure. Competitions like the one organized by the International Conference on Indoor Positioning and Indoor Navigation (IPIN) [1] play a key role by providing standardized benchmarks and encouraging innovation in diverse environments.

Despite their fairness and rigor, current evaluation methodologies—particularly in IPIN Track 3 "Smartphone (offsite-online)"—suffer from a major limitation: the reliance on a sparse set of manually surveyed Ground Truth (GT) points. These highly accurate references are typically spaced 20–50 meters apart due to the cost and effort involved in deploying them across large, multi-floor buildings. Participants submit high-frequency position estimates (e.g., at 2 Hz), which are then evaluated against these sparse GT points.

This sparse evaluation introduces several key challenges:

IPIN-WCAL 2025: Workshop for Computing & Advanced Localization at the Fifteenth International Conference on Indoor Positioning and Indoor Navigation, September 15–18, 2025, Tampere, Finland
*Corresponding author.
†These authors contributed equally.
antonio.jimenez@csic.es (A. R. Jiménez); Joaquin.Torres@uv.es (J. Torres-Sospedra); luisa.ruiz@csic.es (L. Ruiz-Ruiz)
https://lopsi.car.upm-csic.es/ (A. R. Jiménez); http://jtorr.es/ (J. Torres-Sospedra)
0000-0001-9771-1930 (A. R. Jiménez); 0000-0003-4338-4334 (J. Torres-Sospedra); 0000-0003-0316-7781 (L. Ruiz-Ruiz)

- **Masked Error Accumulation:** Localization drift and errors between GT points may go unnoticed or be insufficiently penalized, obscuring actual system performance.
- **Misleading Fidelity Metrics:** Simplified algorithms (e.g., using straight-line interpolation or ignoring lateral movement) can appear accurate at checkpoints, misrepresenting true trajectory quality.
- **Exploitable GT Placement Biases:** The predictable placement of GT points (e.g., near doors or intersections) may lead to overfitting, undermining the fairness of the benchmark.

To overcome these issues, the development of dense, high-fidelity reference trajectories is essential. Foot-mounted Inertial Measurement Units (IMUs), leveraging Inertial Navigation System (INS) principles and Zero-Velocity Updates (ZUPT), can provide dense, relative positioning with high short-term accuracy [2]. However, their performance degrades over time due to drift, requiring fusion with absolute references. While complex fusion techniques such as Kalman filters or graph-based optimization are available, the strong consistency of foot-mounted IMU trajectories allows for more lightweight and robust alternatives.

This paper proposes a novel methodology for generating a **dense, high-fidelity ground truth trajectory** for indoor localization evaluation. It fuses a continuous, drifting trajectory from a foot-mounted IMU with existing sparse GT points using a **segment-wise rigid alignment method**—a combination of localized translation and rotation corrections followed by global trajectory smoothing [3]. The resulting trajectory allows for continuous, fine-grained evaluation at native IMU frequencies (typically over 100 Hz), offering enhanced accuracy and diagnostic power.

The main contributions of this work are:

1. A critical analysis of the limitations of sparse GT-based evaluation in current indoor localization competitions, particularly IPIN Track 3.
2. A robust method to generate dense ground truth by fusing IMU-based relative trajectories with sparse surveyed GT points.
3. A segment-wise rigid alignment strategy that preserves local morphology and corrects drift.
4. An internal validation framework based on held-out GT points to assess the quality of the generated dense GT.

The remainder of this paper is structured as follows: Section 2 describes the enhanced data collection procedure. Section 3 details the proposed fusion method. Section 4 presents experimental results, validation, and implications. Section 5 concludes the paper.

2. Proposed New Data Collection Methodology

The 2025 edition of the IPIN Track 3 competition introduces enhancements to the standard data collection procedure, which gathers smartphone sensor data (WiFi, BLE, etc.) via the GetSensorData App [4, 5]. As a key novelty, the actor is now equipped with a foot-mounted inertial sensor. This setup enables synchronized acquisition of smartphone and IMU measurements under a shared time base. The IMU data allow reconstruction of the actor's trajectory using INS-ZUPT algorithms. Together with sparse ground truth marks (POSI marks) recorded in the mobile logfile when crossing surveyed GT points, this setup provides the required information to fuse both datasets and generate a dense GT reference.

2.1. The New IPIN Track 3 Data Collection Procedure

Between 2015 and 2024, IPIN Track 3 data collection involved deploying physical GT marks on the floor, measuring their distances to nearby structural features (e.g., walls, columns), and extracting their coordinates from georeferenced maps. An actor equipped with one or more smartphones (running GetSensorData [4, 5]) would follow predefined routes, pressing a button to log the exact timestamp when stepping over each GT mark.

Figure 1 shows the upgraded setup used in the 2025 edition.

The main innovations introduced this year are:

Figure 1: Illustration of the data collection setup during the IPIN'25 Competition (Track 3), showing the two smartphones running the GetSensorData App (release June 2025), the GT marks placed on the floor, and the foot-mounted IMU (model IMUE-CSIC).

- **Foot-Mounted IMU:** A custom-made MEMS IMU (model IMUE-CSIC) developed at the Center for Automation and Robotics (CAR) CSIC-UPM [6, 7], attached to the actor's foot. It provides 104 Hz accelerometer and gyroscope readings and transmits data in real-time via BLE to the smartphone.
- **Updated GetSensorData App:** The new Android release (June 2025, https://gitlab.com/getsensordatatools) supports modern phones (e.g., SG24) and logs new data types (e.g., step count, uncalibrated magnetometer, GNSS raw). Crucially, it can now record synchronized streams from the IMUE-CSIC, ensuring all data share a common time reference for fusion.

Figure 2 shows a typical training trajectory mapped onto a satellite view, with the actor passing through 8 GT points.

2.2. Collected Foot-Mounted IMU Data and Estimated Trajectory with Drift

Raw data from the foot-mounted IMU is processed using an Inertial Navigation System (INS) with an Extended Kalman Filter (EKF) [2]. The EKF state vector includes position, velocity, and orientation errors. Zero-Velocity Updates (ZUPT) are applied during foot-flat phases to constrain drift.

This INS-EKF-ZUPT approach yields a dense trajectory $\mathbf{P}_{IMU_j} = (x_j, y_j, z_j)$ at 104 Hz, along with orientation estimates and state covariances. Figure 3 illustrates the drift observed in a training trial (shown previously in Fig. 2).

Even in short trajectories, notable horizontal and vertical drift can be observed. The effect is magnified in longer, multi-floor trials, such as the one in Figure 4, recorded during testing.

These results underscore the need for fusion with absolute GT points to correct drift and enable the IMU data to support rigorous competition evaluation.

3. Proposed Dense Ground Truth Generation

The core of dense Ground Truth (GT) generation lies in effectively fusing the sparse, accurate GT points with the dense, but drifting, IMU trajectory. Various approaches exist for this task, broadly categorised into segment-wise corrections and global optimization techniques. Global optimization methods, such as batch smoothing or graph-based optimization, typically aim to find a globally optimal

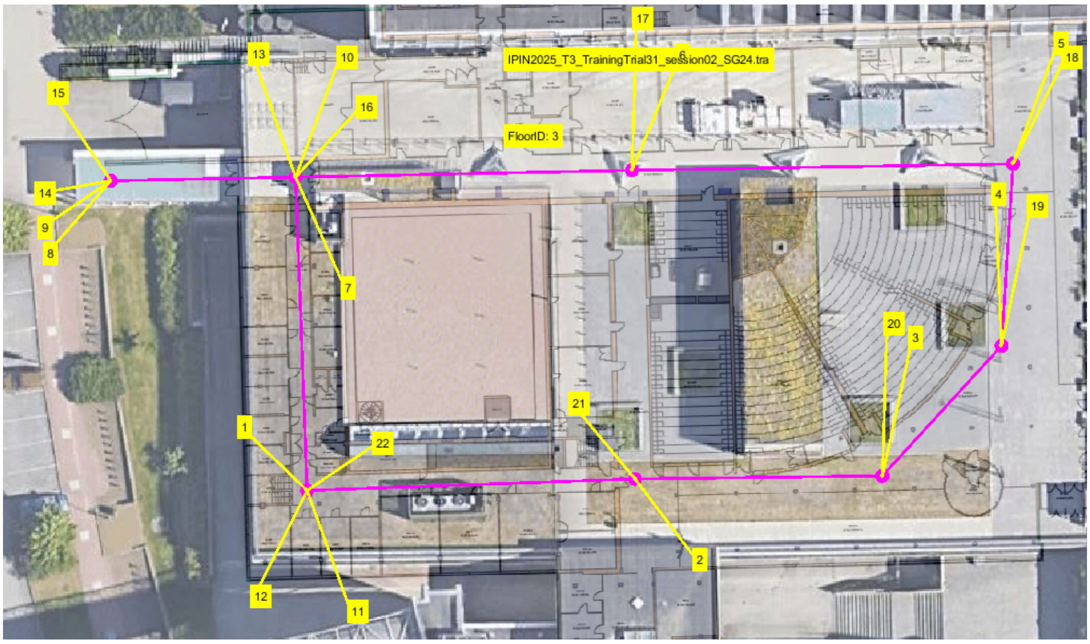

Figure 2: Example of a training trajectory (TrainingTrial31) from IPIN'25 Track 3. The magenta dots show 8 GT points, linked by straight lines. The actor follows a trapezoidal path, starting at GT #1, completing a counter-clockwise loop (up to visit #11), then reversing the route clockwise to return to the start (visit #22).

trajectory by minimizing a cost function that considers all sensor measurements simultaneously. While theoretically powerful and able to distribute errors more smoothly across the entire path, these methods often introduce significant computational complexity and may struggle to preserve sharp, piecewise-linear features of the trajectory without highly sophisticated motion models or specific constraints. Furthermore, some implementations can be computationally prohibitive for very dense IMU data due to the large number of variables involved in the optimization process. Given these considerations and the specific requirement to maintain the precise local geometry of abrupt turns and straight segments, we propose a robust segment-wise correction methodology.

This method provides a robust approach to correct IMU trajectory drift using sparse Ground Truth (GT) points, meticulously preserving the local shape and sharp turns inherent in the original IMU data. The process integrates three main stages: GT point conversion and normalization, segment-wise rigid alignment, and final trajectory smoothing.

3.1. GT Point Conversion and Normalization

Sparse GT points, initially recorded with latitude, longitude, timestamp (t_{GT}), and a FloorID, are transformed into a local Cartesian (X, Y, Z) coordinate system. A planar projection approximation is used for (X, Y) coordinates, where a reference point ($\text{Lat}_{ref}, \text{Lon}_{ref}$), typically the first GT point, serves as the origin:

$$X = R_{earth} \cdot (\text{Lon} - \text{Lon}_{ref}) \cdot \cos(\text{Lat}_{ref})$$

$$Y = R_{earth} \cdot (\text{Lat} - \text{Lat}_{ref})$$

where R_{earth} is Earth's radius and angles are in radians. The Z-coordinate is derived from the FloorID, assuming a constant height per floor, $Z = (\text{FloorID} - \min(\text{FloorID})) \cdot \text{HeightPerFloor}$. Finally, converted GT points are normalized such that their first point aligns with the IMU trajectory's origin.

3.2. Segment-Wise Alignment with Rigid Transformation and Residual Adjustment

This core stage aligns the IMU trajectory, $P_{IMU}(t)$, with the GT points, $P_{GT}(t_{GT})$, while preserving its local geometry. The IMU trajectory is divided into segments defined by consecutive GT points (P_{GT_i} at

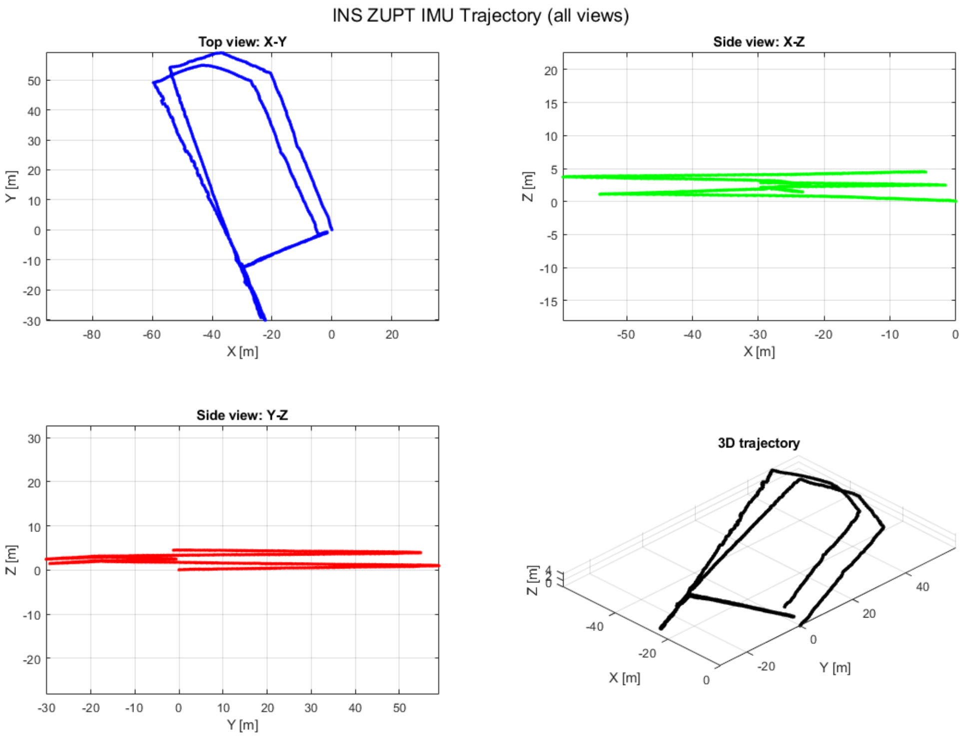

Figure 3: Example of a foot-mounted IMU trajectory (TrainingTrial31) showing drift prior to absolute position correction. Covering a total distance of 463 meters over 9 minutes, the estimated endpoint deviates approximately 6 meters from the true ending position due to accumulated drift. This deviation includes errors of approximately 5 m in Z and -4 m in X. The trajectory represents a closed loop, starting and ending at (0,0,0).

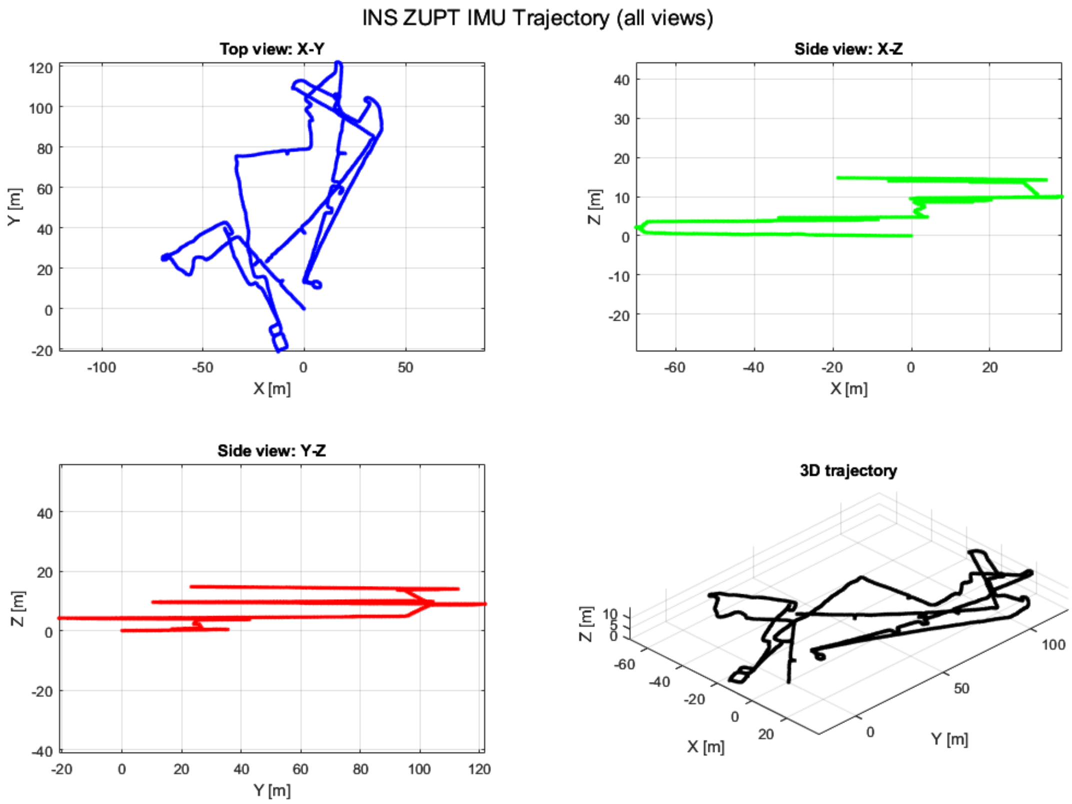

Figure 4: Another foot-mounted IMU trajectory (TestingTrial04) exhibiting drift prior to absolute position correction. This extensive trajectory spans 4 different floors, covering a total distance of 911 meters over 19 minutes. The trajectory represents a closed loop, starting and ending at (0,0,0).

t_{GT_i} and $P_{GT_{i+1}}$ at $t_{GT_{i+1}}$). For each IMU segment $P_{IMU}(t) \in [t_{GT_i}, t_{GT_{i+1}}]$, starting at P_{IMU_i} and ending at $P_{IMU_{i+1}}$:

1. **Initial Translation:** The segment is first translated by $T_1 = P_{GT_i} - P_{IMU_i}$. The translated segment is $P'_{IMU}(t) = P_{IMU}(t) + T_1$.
2. **Rigid Rotation (XY-plane):** A 2D rotation $R(\theta)$ is applied to the XY components of $P'_{IMU}(t)$, pivoting around P'_{IMU_i} (which is now P_{GT_i}). The rotation angle θ aligns the vector $P'_{IMU_i} \to P'_{IMU_{i+1}}$ with $P_{GT_i} \to P_{GT_{i+1}}$ in the XY-plane. The rotated segment is denoted $P''_{IMU}(t)$.
3. **Linear Residual Adjustment:** A final residual translation vector $T_R = P_{GT_{i+1}} - P''_{IMU_{i+1}}$ is computed. This residual is linearly interpolated across the segment and applied:

$$P_{corrected}(t) = P''_{IMU}(t) + T_R \cdot \frac{t - t_{GT_i}}{t_{GT_{i+1}} - t_{GT_i}}$$

For static periods before the first GT and after the last GT, a constant translation offset, derived from the nearest GT point, is applied.

3.3. Final Trajectory Smoothing

To mitigate typical ZUPT position correction at foot stances or any other high-frequency noise while preserving essential trajectory features (e.g., straight lines, sharp turns), a Savitzky-Golay filter [3] is applied to the corrected trajectory. This filter is applied independently to each dimension (X, Y, Z). Optimal results were achieved with a low polynomial order (1) and a window length of 211 points (about 2 seconds for an IMU sampling at 104 Hz).

The outcome of this methodology is a high-fidelity Ground Truth estimate, effectively combining the local robustness of INS-ZUPT with the global accuracy from sparse GT points, yielding a smooth and geometrically faithful path representation, as will see in next section.

4. Quality of Experimental Results and Validation

Our methodology for generating a dense, high-fidelity Ground Truth (GT) will be rigorously validated using data collected from the multi-floor Tampere University building for the IPIN 2025 Track 3 competition. The validation will be conducted through two primary approaches: a) preliminary visual observation of the corrected trajectories and b) a quantitative assessment using a subsample of GT points for fusion and the remaining as an independent test set.

4.1. Visual Observation Validation

This subsection presents visual evidence of the achieved improvements. We will utilize the same IMU trajectories, initially presented with drift in Section 2.2 (Figures 3 and 4), as illustrative examples.

After applying the proposed fusion algorithm described in Section 3, the significant improvement in trajectory accuracy and coherence becomes evident. Figures 5 and 6 visually demonstrate this enhancement. It is clear that the corrected IMU trajectory (represented in blue, now serving as the dense GT) not only precisely aligns with the sparse GT points but also effectively preserves the intricate details and local geometry of the original drifted trajectory (shown in red).

4.2. Quantitative Validation through GT Subsampling

To rigorously validate the accuracy of our generated dense Ground Truth, we propose a cross-validation inspired approach. Specifically, we will intentionally **withhold** a distinct subset of the surveyed GT points from the fusion process. After generating the dense GT trajectory using only the remaining GT points, we will evaluate the accuracy of our fused trajectory at the precise locations of the withheld GT points.

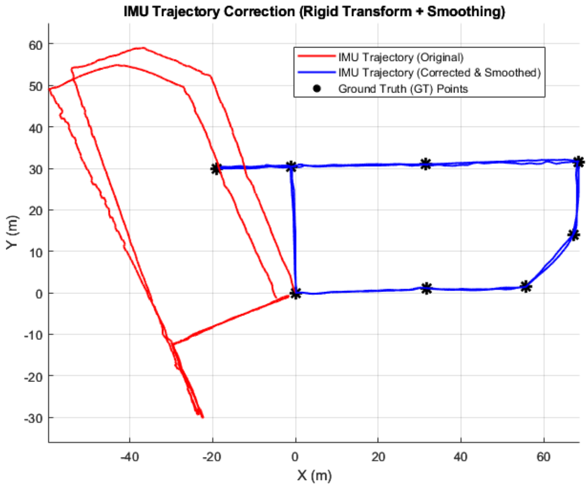

Figure 5: Example of a fused Ground Truth trajectory (blue), showing how the IMU drift (red) is corrected by alignment with sparse GT points (black dots), resulting in a smooth and accurate path that preserves local morphological details.

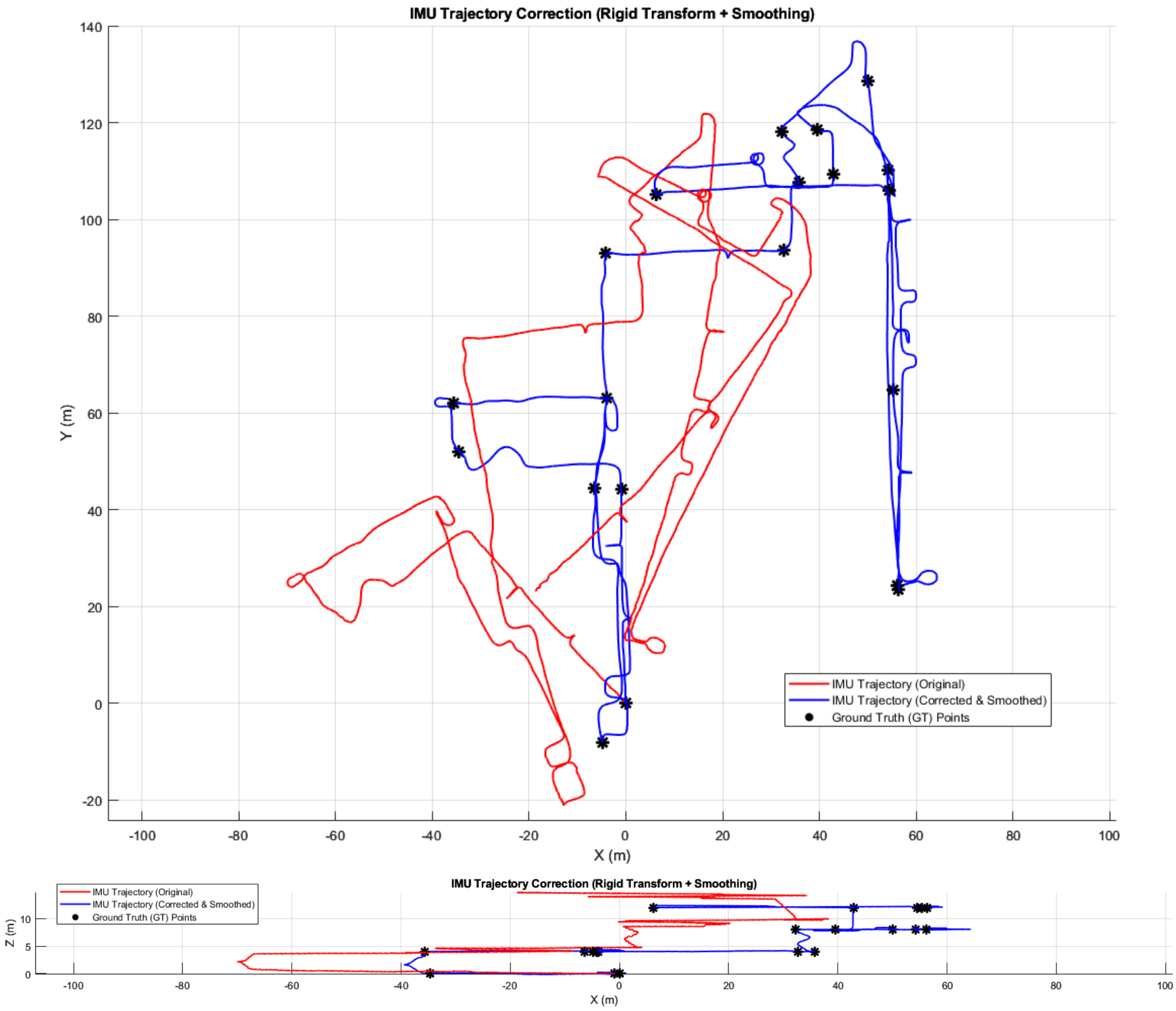

Figure 6: Example of a fused Ground Truth trajectory (blue), showing how the IMU drift (red) is corrected by alignment with sparse GT points (black dots), resulting in a smooth and accurate path that preserves local morphological details. Top plot: X-Y zenithal view, and bottom plot: the Z-X vertical view.

For this validation, 10% of the total GT points were reserved for evaluation in each fold, with the remaining 90% utilized in the fusion process. A total of ten folds were generated for each trajectory, ensuring a comprehensive assessment across different subsets of data.

Table 1 presents the key error metrics (mean Euclidean distance error, median, 3rd quartile, and maximum error) obtained at these withheld GT points, demonstrating the precision of our fused trajectory compared to the actual surveyed points.

Table 1
Error Metrics for Fused GT Validation

Metric	Horizontal Error (m)	Floor ID Error (%)
Mean Error	0.40	0
Median Error	0.37	0
3rd Quartile Error	0.63	0
Max Error	0.74	0

These quantitative error values provide strong evidence of the proposed method's accuracy. Specifically, the maximum observed horizontal error of 0.74 meters, occurring when an intermediate GT point is withheld, confirms that the interpolation between consecutive GTs (even across larger gaps) remains well below 1 meter. Furthermore, the 0.0% Floor ID Error unequivocally demonstrates the method's robustness in maintaining vertical consistency across multiple floors. Given these results, it is reasonable to expect an even lower error when all available GT points are utilized for fusion at their typical density, likely yielding average errors around 0.5 meters. Critically, the error is expected to diminish as the generated dense GT points approach any of the sparse surveyed GT points. Therefore, we confidently conclude that our generated dense GT is indeed suitable as a high-accuracy, continuous reference for rigorous localization system evaluation.

4.3. Impact on Competitor Evaluation: Future Work and Discussion

In the current IPIN 2025 Track 3 competition, our newly proposed dense GT generation methodology will not be used for computing the official scores that determine the competition winners. However, we will compute alternative scores using this dense GT in parallel, specifically to analyze its potential influence on competitor rankings.

Through this future analysis, we aim to systematically compare the evaluation results obtained with the traditional sparse GT against those derived from our proposed dense GT. We expect this comparison to reveal several key differences:

- A different distribution of errors across competitor trajectories, which will expose sustained drifts or localized accuracy issues not evident with sparse GT evaluation.
- Tangible changes in final scoring metrics (e.g., 3rd quartile error, mean error), arguing that the dense GT provides a more representative and robust assessment of real-world continuous localization accuracy.
- Enhanced diagnostic capabilities for competitor algorithms, allowing for clearer identification of their strengths and weaknesses across the entire trajectory, rather than just at sparse checkpoints.

5. Conclusions

This paper has presented a novel and robust methodology for generating a dense, high-fidelity Ground Truth (GT) trajectory, directly addressing the limitations of sparse GT evaluation in indoor localization competitions, particularly within the IPIN Track 3 framework. Our approach effectively fuses a dense, relative trajectory derived from a foot-mounted Inertial Measurement Unit (IMU) using Zero-Velocity Updates (ZUPT) with existing sparse, highly accurate surveyed GT points. The core of our solution lies in a segment-wise rigid alignment method that uniquely preserves the detailed local morphology and sharp turns of the original IMU path, followed by a global trajectory smoothing. This methodology yields a dense GT that significantly enhances the granularity and fairness of localization system evaluation, enabling continuous assessment of competitor trajectories at their native output frequencies. Visual and quantitative experimental results confirm the high accuracy and fidelity of the generated dense GT, proving its suitability as a robust reference for benchmarking and laying the groundwork for more sophisticated and diagnostically rich evaluations in future indoor localization challenges.

Acknowledgments

We thank the local IPIN organizers in Tampere: Lucie, Jari and Roman for their full support during the measurement campaign; Miguel from Université Gustave Eiffel; and the Competition Chairs at CNR: Francesco, Filippo, and Davide, for their assistance with the georeferencing of GT points.

This work was supported by the INDRI Project (PID2021-122642OB-C43 and PID2021-122642OB-C42, funded by AEI/10.13039/501100011033 and FEDER, EU) and the DISTRIMUSE Project (Horizon Europe, KDT-JU-2023-2-RIA, project no. 101139769). J. Torres-Sospedra also acknowledges funding from Generalitat Valenciana (CIDEXG/2023/17, Conselleria d'Educació, Universitats i Ocupació).

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT 4o and Gemini 2.5 Flash in order to: Grammar and spelling check. Figures are generated using Matlab which is the programming language for script implementation. Intellectual ideas are original from the authors. The authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] F. Potorti, A. Crivello, S. Lee, B. Vladimirov, S. Park, Y. Chen, L. Wang, R. Chen, F. Zhao, Y. Zhuge, H. Luo, A. Perez-Navarro, A. R. Jiménez, Offsite evaluation of localization systems: Criteria, systems, and results from ipin 2021 and 2022 competitions, IEEE Journal of Indoor and Seamless Positioning and Navigation 2 (2024) 92–129. doi:10.1109/JISPIN.2024.3355840.

[2] A. R. Jiménez, F. Seco, J. Prieto, J. Guevara, Indoor pedestrian navigation using an ins/ekf framework for yaw drift reduction and a foot-mounted imu, in: 2010 7th Workshop on Positioning, Navigation and Communication, 2010, pp. 135–143. doi:10.1109/WPNC.2010.5649300.

[3] P. A. Gorry, General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method, Analytical Chemistry 62 (1990) 570–573. URL: https://doi.org/10.1021/ac00205a007. doi:10.1021/ac00205a007.

[4] A. R. Jiménez, F. Seco, J. Torres-Sospedra, Tools for smartphone multi-sensor data registration and gt mapping for positioning applications, in: 2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2019, pp. 1–8. doi:10.1109/IPIN.2019.8911784.

[5] J. D. Gutierrez, A. R. Jiménez, F. Seco, F. J. Álvarez, T. Aguilera, J. Torres-Sospedra, F. Melchor, Getsensordata: An extensible android-based application for multi-sensor data registration, SoftwareX 19 (2022) 101186. URL: https://www.sciencedirect.com/science/article/pii/S2352711022001121. doi:https://doi.org/10.1016/j.softx.2022.101186.

[6] L. Ruiz-Ruiz, F. Seco, A. Jiménez, S. Garcia, J. J. García, Evaluation of gait parameter estimation accuracy: a comparison between commercial imu and optical capture motion system, in: 2022 IEEE International Symposium on Medical Measurements and Applications (MeMeA), 2022, pp. 1–2. doi:10.1109/MeMeA54994.2022.9856475.

[7] L. Ruiz-Ruiz, J. J. García-Domínguez, A. R. Jiménez, A novel foot-forward segmentation algorithm for improving imu-based gait analysis, IEEE Transactions on Instrumentation and Measurement 73 (2024) 1–13. doi:10.1109/TIM.2024.3449951.