A Multi-Scenario Dataset for Long-Term Indoor Localization and Pedestrian Behavior Analysis in Dynamic Environments

Faezeh Sadat Mortazavi[1*], Junyi Wei[2], Tim Schimansky[1], Hangbin Wu[2], Claus Brenner[1], Monika Sester[1]

[1] Institute of Cartography and GeoInformation, Leibniz Universität Hannover, Germany,
(Faezeh.Mortazavi, Tim.Schimansky, Claus.Brenner, Monika.Sester)@ikg.uni-hannover.de
[2] College of Surveying and Geo-informatics, Tongji University, Shanghai, China (weijunyi1997, hb)@tongji.edu.cn

Keywords: Indoor Localization, Long-Term Localization, Dynamic Environments, LiDAR Dataset, Pedestrian Behavior Analysis.

Abstract

Human activity and structural modifications continuously alter shared indoor spaces, leading to challenging conditions for reliable localization and motion understanding. To investigate and analyze the impact of such dynamics on long-term indoor localization, we present a multi-scenario dataset designed under controlled levels of occlusion and environmental change. The data were collected in a university entrance hall configured to simulate a conference environment, with movable poster walls and natural pedestrian activity around them. A movable LiDAR platform was used to collect data within the environment, while four synchronized overhead AI cameras captured multi-view pedestrian motion. The image data from the cameras are synchronized with the LiDAR point clouds, enabling joint analysis of pedestrian behavior in both 2D and 3D domains. Three scenarios, named extreme occluded, semi occluded, and free space, represent increasing levels of structural modification and visibility loss. High-precision ground truth was established using total station tracking. The dataset enables systematic research on localization performance under evolving indoor conditions and supports the analysis of pedestrian behavior and human–robot interaction in shared spaces.

1. Introduction

Dynamic indoor environments pose substantial challenges for autonomous systems and spatial computing applications. Unlike controlled or static settings, indoor spaces are subject to continuous change caused by moving pedestrians, temporary occlusions, and long-term structural modifications such as furniture rearrangements or the introduction of partitions. These dynamics directly affect perception, map consistency, and motion planning, thereby limiting the robustness of algorithms that rely on stable scene structure. Among these factors, pedestrian activity represents the most frequent and unpredictable source of variability. Human motion leads to short-term occlusions, alters line-of-sight to structural elements, and produces highly non-repetitive patterns that complicate scene interpretation. For autonomous systems operating in shared spaces such as service robots in public facilities, mobile platforms in factories, or AR applications in crowded halls, understanding pedestrian behavior is essential not only for safe interaction but also for reliable localization and navigation. Accurate localization remains a central requirement for these tasks. In dynamic environments, algorithms must maintain precise pose estimation despite changing visibility conditions and inconsistencies between live sensor observations and pre-existing maps.

To study these challenges, we introduce a dataset that integrates long-term indoor localization with pedestrian behavior analysis. The data were collected in a university entrance hall that was intentionally configured to simulate a conference setting. Poster walls and movable furniture were arranged to create varying levels of occlusion, while natural pedestrian activities were recorded to capture diverse behaviors. A mobile sensor platform, mounted on a push cart and equipped with LiDAR and IMU, provided ego-centric data for localization benchmarking, while four AI cameras installed in the corners of the hall recorded synchronized overhead views of pedestrian motion. Since pedestrians can be detected both in the images and in the LiDAR point cloud, and these modalities are spatially aligned, the dataset enables research in both 2D and 3D domains. This combination supports the evaluation of localization algorithms under controlled environmental variability and the analysis of natural human activity in shared spaces. By combining ego-centric sensor data with synchronized overhead camera views, the dataset bridges the gap between localization and human activity analysis. It enables benchmarking of algorithms under realistic indoor dynamics and opens new opportunities to study interactions between autonomous systems and pedestrians. We expect this resource to support research on both robust localization and pedestrian behavior, contributing to safer and more reliable autonomy in dynamic indoor spaces.

2. Related Work

The evaluation of localization and mapping algorithms relies heavily on the availability of high-quality datasets with accurate ground truth. Benchmarking frameworks based on real-world sensor data are essential for systematically testing algorithmic robustness and identifying limitations under realistic conditions. While numerous datasets have been developed for outdoor environments, often targeting autonomous driving scenarios, comparatively fewer efforts have focused on indoor localization, despite its critical role in applications such as logistics, assistive robotics, and facility monitoring. Well-known outdoor datasets like KITTI (Geiger et al., 2012) and EuRoC (Burri et al., 2016) have established strong precedents by providing synchronized multi-sensor data and precise trajectory ground truth. However, comparable benchmarks for indoor environments are still evolving. Several indoor datasets have been introduced to support the development and evaluation of SLAM and localization algorithms. These datasets differ in terms of sensor modalities, ground truth acquisition methods, and environmental complexity. Early contributions such as RAWSEEDS (Ceriani et al., 2009), the MIT Stata Center dataset (Fallon et al., 2013),

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-1-W6-2025-155-2025 | © Author(s) 2025. CC BY 4.0 License.

and the COLD database (Pronobis and Caputo, 2009) focused primarily on static or structured indoor spaces, often using external tracking systems to generate accurate pose references.

In recent years, more attention has been given to capturing variability in indoor environments, but the focus has often remained on extending spatial coverage or sensor diversity rather than addressing localization under environmental change. For example, the dataset from Kaveti et al. (2023) includes multi-floor buildings and complex layouts, while others such as the TUM RGB-D dataset (Sturm et al., 2012) provide RGB-D sequences for visual SLAM evaluation under realistic sensor noise and limited motion blur. FusionPortable (Jiao et al., 2022) extends this direction by supporting multi-sensor localization across indoor–outdoor campus scenes with a focus on cross-platform use. The Hilti-Oxford dataset (Zhang et al., 2022) offers high-precision ground truth using LiDAR and IMU data, mainly focusing on mapping accuracy in indoor settings. Event-based sensors have also been included in benchmarks like VECtor (Gao et al., 2022) to test SLAM systems under visually challenging conditions.

Furthermore, while several datasets provide long sequences and heterogeneous data types, they often lack structured variation across recordings. Scene changes are often incidental, making it difficult to isolate the impact of dynamic elements such as occlusion, temporary structures, or moving agents. A small number of works, such as OpenLORIS (Shi et al., 2020), offer repeated recordings in real-world spaces, but do not provide control over the degree of environmental dynamics. The dataset proposed by Trekel et al. (2025) offers a valuable benchmark for long-term indoor localization, covering both short-term dynamics and structural changes across multiple rooms using 2D LiDAR. Building on these efforts, our dataset extends this line of research toward explicitly controlled and multimodal settings. It introduces adjustable levels of environmental change, including severe occlusions and long-term modifications, while combining LiDAR and camera data. By systematically varying the scene configuration, our benchmark enables reproducible evaluation of algorithmic robustness under increasingly dynamic indoor conditions.

In addition to dataset development, algorithmic advances such as scene flow estimation have been proposed to capture short-term dynamics in 3D data. Scene flow represents the point-wise motion between consecutive LiDAR scans and has been explored in recent works (Najibi et al., 2022; Wang et al., 2022) for trajectory inference using unsupervised learning. It has also been integrated into SLAM pipelines to improve pose estimation and model rigid moving objects (Singh et al., 2021; Wang et al., 2020). These approaches motivate the inclusion of dynamic scenarios in our dataset, where short-term changes and object motion are explicitly captured for algorithm evaluation.

3. Data Acquisition and Scenario Design

3.1 Sensor Setup and Configuration

To address the challenges of long-term changes and high dynamics in indoor localization, and to provide multimodal data for pedestrian behavior analysis, we employed both fixed multi-view cameras and a mobile sensing platform. The fixed cameras captured pedestrian activity and long-term structural modifications, while an edge-based camera system, consisting of a Raspberry Pi 5 paired with an IMX500 AI camera running a lightweight YOLOv8 detector, reduced redundant storage by recording frames only when pedestrians were detected.

The sensor platform consists of a multi-sensor setup, primarily featuring two 3D LiDARs and a 360° camera (see Figure 1). The LiDAR configuration comprises an Ouster OS1 as the main sensor, mounted horizontally and spanning a 45° vertical field of view, and a Hesai Pandar 64, slanted forward at approximately 40°, capturing the environment that includes the ground in front, the walls to the sides, and the ceiling at the back. The system also features a Ricoh Theta X 360° camera, which captures 8K video at a lower $5\ Hz$ framerate to achieve high image quality while minimizing data transfer rates. The camera operates in dual fisheye mode, where the internal stitching of the two video feeds is omitted in favor of a stable camera geometry. Both LiDARs are recorded on a mini PC running ROS 1, and stored in `rosbag` format, facilitating a common format regardless of the vendor for further processing, such as SLAM.

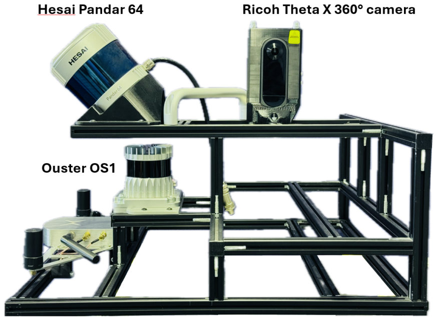

Figure 1. The platform includes a Hesai Pandar64 LiDAR, an Ouster OS1 LiDAR, and a Ricoh Theta X 360° camera mounted on a rigid frame for stable geometry and calibration.

For high-precision mapping and ground-truth validation, we used a Z+F 5016C terrestrial laser scanner and a Leica total station. Scene-wide reference point clouds were generated through multi-station scanning and marker-based registration. The final registered point cloud achieved an average alignment accuracy of 1.0 mm (σ = 0.7 mm), with a maximum target deviation of 3.2 mm. To obtain accurate ground truth, we tracked a 360° reflective prism using the total station while collecting LiDAR data. This setup ensured continuous line-of-sight tracking throughout the sequences.

3.2 Scenario Design

Long-term localization in indoor environments remains a significant challenge, particularly due to changes that disrupt the correspondence between incoming sensor data and pre-existing maps. These changes include both transient dynamics (e.g., pedestrian motion) and semi-permanent modifications such as newly introduced objects or partitions. Such alterations can obscure map features, occlude structural landmarks, and lead to ambiguous scan alignments, resulting in degraded or failed localization. To investigate these challenges in a controlled and reproducible manner, we designed a dynamic indoor environment within a university entrance hall measuring approximately 40×10 meters (see Figure 3).

All data were recorded in the same physical space using an identical sensor configuration. The scenarios differ only in the intentional manipulation of occlusions and scene layout, allowing us to simulate varying levels of environmental change

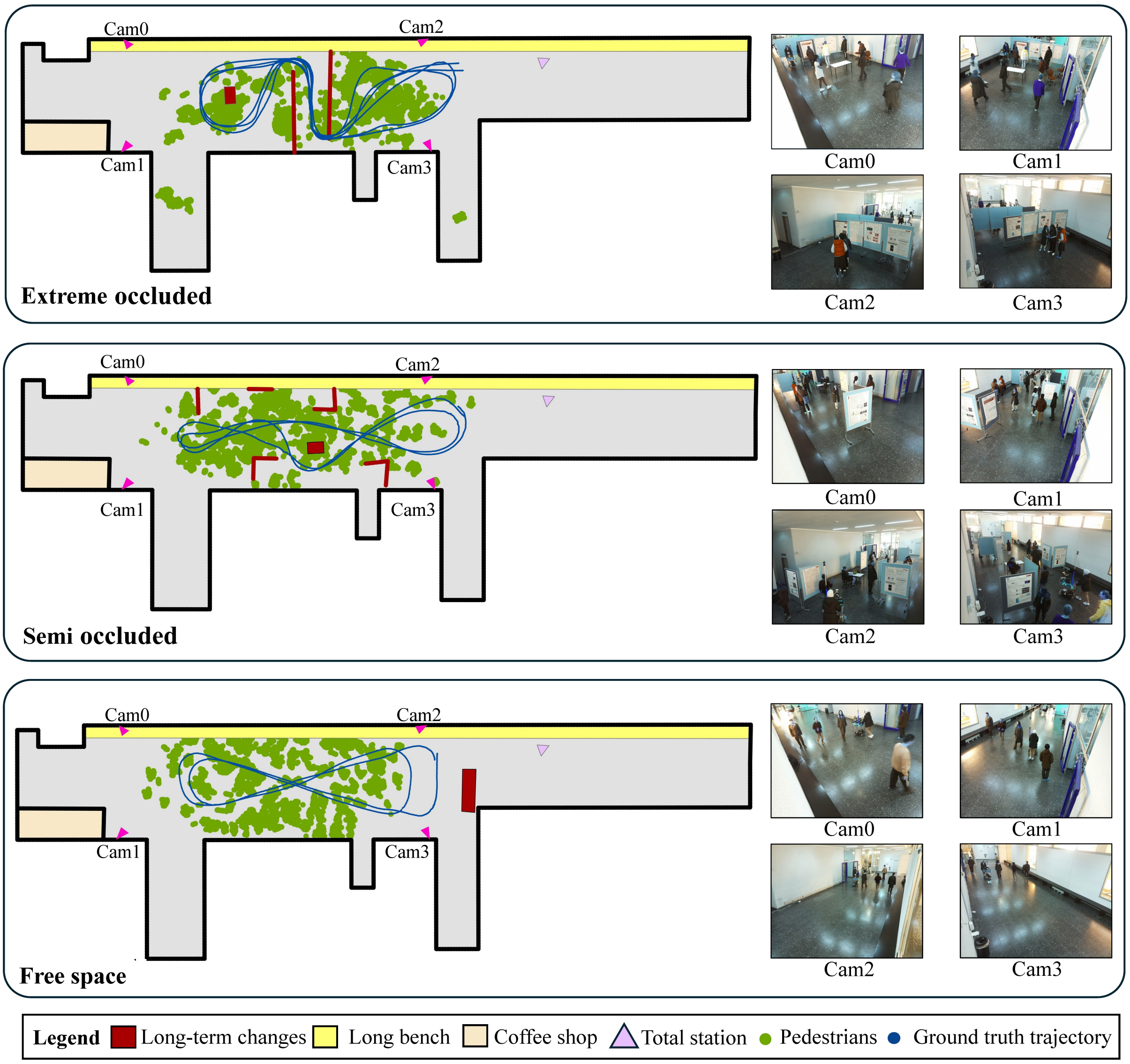

Figure 2. Illustration of the three recording scenarios: extreme occluded, semi occluded, and free space. The left side shows the floor plan with sensor placement, pedestrian distribution, and ground-truth trajectory, while the right side provides sample views from the four fixed cameras.

while maintaining experimental consistency. The three resulting scenarios, extreme occluded, semi occluded and free space, are illustrated in Figure 2, which shows both the floor plan with sensor placement and the sample camera views. Each scenario is evaluated against a static reference map generated via high-resolution terrestrial laser scanning (TLS). This reference represents an uncluttered environment without furniture, poster walls, or pedestrians. Consequently, any additional elements introduced in the dynamic scenarios constitute long-term deviations from the reference, directly influencing the difficulty of localization.

3.2.1 Extreme Occluded Scenario: This scenario presents the most severe modifications. Eight poster walls are used to construct an artificial corridor that obstructs line-of-sight to most original structures. As a result, the LiDAR predominantly perceives newly added elements, while returns from the reference environment are minimal or entirely absent. Pedestrian movement further introduces short-term occlusions. This configuration represents drastic long-term changes, such as renovations or re-partitioned spaces, where scan-to-map alignment becomes highly ambiguous. Successful localization in this setting demands robustness to severe occlusions and the ability to generalize to significantly altered environments.

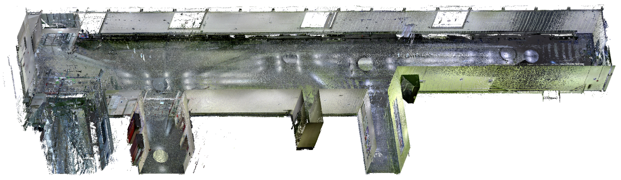

Figure 3. Reference map from terrestrial laser scanning

3.2.2 Semi Occluded Scenario: In this scenario, we introduce L-shaped poster walls and movable objects such as chairs and tables. These additions partially occlude key structural features visible in the reference map, requiring the system to localize using an incomplete and altered subset of the original map.

Sequence	S0	S1	S2	S3	S4	S5	S6	S7
Number of scans	3148	1628	826	1189	1568	1315	1223	1184
Duration [min]	5.25	2.71	1.38	1.98	2.61	2.19	2.04	1.97

Table 1. Number of LiDAR scans and recording duration for each sequence.

This setting reflects common real-world situations; such as rearranged offices or temporary setups, and tests the algorithm's stability under moderate environmental changes.

3.2.3 Free Space Scenario: This baseline scenario mirrors the reference map, with no static obstacles introduced. The only dynamic elements are 10–15 pedestrians exhibiting natural behaviors such as walking, interacting, or pausing. These transient occlusions affect LiDAR visibility but leave the underlying structural layout unchanged. This setting serves as a reference case to assess algorithmic robustness under dynamic, yet structurally consistent, conditions.

To isolate the impact of scene dynamics on localization, all scenarios were recorded under identical lighting conditions, using the same sensor setup. Only the spatial arrangement of objects and pedestrian activity was varied. Each scenario was recorded multiple times with distinct trajectories: two sequences for the extreme occluded (S0, S1), three sequences for the semi occluded case (S2, S3, S4), and three sequences for the free space case (S5, S6, S7). This design enables consistent, reproducible evaluation across varying levels of environmental complexity. Compared to existing indoor datasets, the modifications introduced in our setup are spatially localized and systematically designed to create distinct levels of occlusion, with the goal of analyzing long-term localization performance under increasing environmental change. Table 1 summarizes the number of LiDAR scans and the corresponding recording durations for all sequences.

4. Data Processing and Ground Truth

4.1 Time Synchronization

Time synchronization is a critical requirement for aligning and processing data from multiple sensors. To achieve this, we employed a combination of hardware and software based synchronization methods to establish a unified time reference and ensure consistent timestamps.

First, at the hardware level, we addressed the issue of synchronizing multiple sensors by establishing a common time base. To synchronize the four cameras, the four computers to which they were attached were connected locally via ethernet cables and a network switch. Clock synchronization then was done by running the Network Time Protocol (NTP) service. One of the computers was selected as NTP master and was provided UTC time via a directly connected GPS module. As a result, while the four cameras were free-running, all of them were timestamped in UTC time. Similarly, the mobile platform and the total station each were free-running, but used their own GPS modules to time-stamp measurements using UTC time.

At the software level, we ensured that all data in the dataset could be accurately matched across sensors. To account for the different frame rates of the sensors, the camera group, which had the highest frame rate, was selected as the reference for data-level synchronization. The start time of data acquisition from the master camera was defined as the start time for the entire system. Each of our sensors provides timestamps for every frame with an accuracy better than one millisecond. By matching the nearest frames of each sensor to the master timestamp, corresponding frames were determined.

4.2 Ground Truth Generation

To establish accurate reference trajectories for benchmarking, a total station continuously tracked a prism target rigidly mounted on a pole attached to the sensor platform during data collection. The resulting timestamped 3D trajectories enable systematic evaluation of SLAM algorithms under varying indoor dynamics. The ground-truth generation pipeline consists of the following steps.

4.2.1 Prism Trajectory Transformation into the Reference Map: The total station provided a timestamped 3D trajectory in its local coordinate system. The timestamps were synchronized with the LiDAR sensor to ensure correspondence between prism positions and LiDAR scans. To express this trajectory in the coordinate system of the reference map generated using TLS, we first measured approximately 200 control points distributed across structural elements such as walls, floors, and ceilings. These points were used to determine the spatial transformation between the total station coordinate frame and the reference map. The control points were registered to the reference map using Iterative Closest Point (ICP) alignment, resulting in a root mean square error (RMSE) of 2.3 mm, confirming high geometric consistency. The transformation matrix obtained from this registration defines the rigid-body transformation between the total station and the reference map coordinate systems. Applying this transformation to the prism trajectory expresses it in the reference map frame, forming the basis for subsequent steps in the pipeline, including sensor–prism offset estimation and benchmarking of SLAM-based trajectories.

4.2.2 Sensor-to-Prism Offset Estimation: To convert the globally registered prism trajectory into the corresponding LiDAR sensor trajectory, it is necessary to determine the fixed spatial transformation between the prism and the sensor origin. This transformation, referred to as the sensor-to-prism offset, remained constant throughout all measurement sessions, as the physical configuration of the platform did not change. To estimate the sensor-to-prism offset, we utilized the sequences recorded in the Free Space scenario (S5–S7), which offered minimal occlusions and reliable alignment conditions. In these sequences, global localization was performed using an odometry method that estimates sensor motion based on a constant-velocity model applied between consecutive LiDAR scans. The predicted poses were refined through Point-to-Plane ICP alignment with the reference map to maintain global consistency. Given the synchronized timestamps of the LiDAR and total station data, corresponding pose pairs were identified between the estimated sensor trajectory and the prism trajectory. The fixed spatial offset between the prism and the LiDAR sensor was determined by optimizing the alignment between their corresponding trajectories. The optimization minimized the mean Euclidean distance between matched pose pairs, resulting in the translation parameters ($\Delta x, \Delta y, \Delta z$) with an uncertainty of approximately ± 2 mm (95% confidence) for all components. This offset was then applied to all prism trajectories to obtain globally aligned sensor trajectories for all recorded sequences.

4.3 SLAM Benchmarking and Trajectory Refinement

For each of the eight recorded sequences, four different SLAM algorithms were executed: DDLO, FAST-LIO, LIO-SAM, and KISS-SLAM. These methods represent a diverse range of algorithmic designs. DDLO is designed to handle highly dynamic environments using structured point cloud processing and dynamic object removal (Lichtenfeld et al., 2024). FAST-LIO is a tightly coupled LiDAR–IMU method optimized for real-time performance in structured environments (Xu et al., 2021). LIO-SAM incorporates both LiDAR–IMU fusion and loop closure for improved global consistency (Shan et al., 2020). Finally, KISS-SLAM is a pure point cloud–based method that operates without IMU input, focusing on simplicity and lightweight implementation (Guadagnino et al., 2025). This variety enables a balanced evaluation across different levels of scene dynamics and highlights the comparative strengths and limitations of each approach under complex indoor conditions. The output trajectories from these methods were transformed into the global reference frame by aligning each SLAM-generated map with the reference map. With the known spatial offset between the prism and the LiDAR sensor, along and using synchronized timestamps, we established point-wise correspondences between SLAM-estimated poses and ground truth positions recorded by the total station.

To quantitatively assess SLAM performance, we computed two standard trajectory error metrics: Absolute Trajectory Error (ATE), Relative Pose Error (RPE), and the root mean square of estimated accelerations (Acc_RMS) to assess trajectory smoothness and motion stability. These evaluations were performed separately for each sequence. The benchmarking results are presented in Table 1 and analyzed further in the Experimental Results section. For each sequence, the final refined trajectory was obtained by replacing the translational component of the best-performing SLAM pose with the corresponding positions from total station, while retaining the SLAM-estimated orientation. Since the total station does not provide rotational measurements, this fusion results in a complete 6-DoF trajectory, integrating high-precision positional data with orientation estimates suitable for downstream tasks such as human activity analysis or map-based navigation.

5. Pedestrian Activity Dataset

To demonstrate the potential of this dataset for pedestrian behavior analysis, in this section we provide a concise description and qualitative overview from three aspects: the experimental design, the data-collection protocol, and potential applications.

5.1 Experimental Design

To investigate pedestrian detection, tracking, re-identification, action recognition, and behavior understanding in specific spatial environments, we conducted real-world motion capture experiments in the three previously defined scenarios. A total of fifteen participants took part in the recordings. To compare differences in pedestrian motion and behavioral patterns under high- and low-frequency interaction states (including non-interactive conditions) and to enrich the diversity of actions and behaviors in the dataset, we conducted eight groups of data collection experiments across the three scenarios. During each session, participants were instructed to simulate the assigned interaction condition and were allowed to freely interact with surrounding people or objects, including personal belongings such as mobile phones.

In the extreme occluded and semi occluded scenarios, poster-wall structures were intentionally arranged to create complex visual conditions for pedestrian observation. In the extreme occluded case, a corridor was formed at the center of the scene (as shown in Figure 2), while in the semi occluded case, five display pillars were placed near the corners of the area. These configurations simulate realistic indoor environments with partial visibility, where pedestrians are occasionally captured by only a subset of cameras and their movements are influenced by occlusions from surrounding objects. At the same time, these objects encourage diverse and natural interactions between participants and the environment. The purpose of these two scenarios is to provide challenging yet controlled conditions for studying robust multi-view pedestrian tracking and behavior interpretation.

Within these two scenarios, participants exhibited a wide range of motion and interaction behaviors. In the extreme occluded setup, they frequently walked through narrow passages, sometimes encountering others approaching from the opposite direction, leading to brief avoidance maneuvers, waiting behavior, or spontaneous short exchanges. In the semi occluded setup, some individuals stood near the poster walls or tables, reading or discussing the displayed materials, while others moved between different areas to join or leave small groups. Occasional stop-and-go movements, short conversations, and body rotations around poster walls created realistic crowd dynamics typical of social or exhibition environments. These activities introduce diverse spatial interactions and motion trajectories that reflect indoor behaviors under varying visibility and crowding conditions.

In the free space scenario, participants were instructed to simulate diverse behavioral patterns, including following, joining, avoiding, and confronting. Unlike the other settings, no artificial occlusions were introduced. Instead, this scenario focuses on the recognition and interpretation of multi-view pedestrian motion, actions, and behavior patterns over time under unobstructed visibility conditions.

5.2 Data Statistics and Synchronization

To demonstrate the usability of this dataset for pedestrian-related research, we provide a descriptive overview of the collected data in terms of interaction frequency and camera coverage across the different scenarios. Each recording sequence includes synchronized image streams from four fixed cameras with a resolution of 640×480 px and a frame rate of 30 fps. The detailed number of frames and recording durations for all cameras and sequences are summarized in Table 2. Both high- and low-frequency interaction sessions were recorded for each scenario type to capture a wide range of pedestrian activities.

To further illustrate the characteristics of the collected data, it should be noted that, due to factors such as variations in device startup and shutdown times, viewing angles, and temporary occlusions, the number of frames captured by each camera differs slightly. Nevertheless, since precise multi-sensor time synchronization was performed during data acquisition, accurate multi-view alignment can be achieved in practice by matching

Camera		S0	S1	S2	S3	S4	S5	S6	S7
Cam 0	Number of frames	2982	2628	1247	–	1704	1740	765	1507
	Duration [s]	251.8	231.4	118.2	–	183.2	118.0	49.1	143.8
Cam 1	Number of frames	5599	3876	1791	2663	3073	2018	2034	2280
	Duration [s]	349.7	242.1	113.0	166.3	191.9	126.1	127.0	142.4
Cam 2	Number of frames	5481	3820	1820	2599	3216	1966	2025	2122
	Duration [s]	346.7	239.2	113.6	162.3	201.0	122.9	126.5	132.5
Cam 3	Number of frames	5354	3646	1607	2580	–	1872	1903	2014
	Duration [s]	334.5	227.7	100.3	161.1	–	116.9	118.8	125.8

Table 2. Overview of camera recordings for each sequence. Each pair of rows shows the number of frames and the corresponding recording duration for each camera. A dash (–) indicates that the camera was not operating during that sequence.

corresponding timestamps across devices. The synchronization accuracy of the four-camera system was validated during the calibration phase by simultaneously recording a microsecond-level timer display under a local area network configuration. The results demonstrated a time synchronization error of less than 1 ms. During actual data acquisition, a maximum temporal deviation of approximately 27 ms was observed, primarily due to network latency and hardware timing differences. Considering that the experiments were conducted at a sampling rate of 30 fps, this deviation corresponds to less than one frame interval.

5.3 Potential Applications

The pedestrian dataset provides accurate multi-view visual data, serving as a valuable foundation for reconstructing pedestrian motion information such as 3D trajectories, velocities, and accelerations, as well as for action recognition and behavior understanding in complex indoor environments. As illustrated in Figure 4, occlusions caused by surrounding objects and partial overlaps among pedestrians occasionally lead to missed detections in certain camera views. Rather than being a limitation, these conditions reflect realistic challenges of crowded indoor spaces and make the dataset particularly suitable for developing and evaluating robust multi-view tracking, re-identification, and behavior analysis algorithms under partial visibility and dynamic interaction scenarios.

6. Experimental Results on Localization

In this section, we present a comparative analysis of SLAM performance across all recorded sequences. Results are grouped by scenario type to examine how varying levels of environmental dynamics influence localization and mapping accuracy. The presence of occlusions, structural changes, and moving people introduces challenges such as scan misalignment, missing features, and inconsistent pose estimation. Performance is assessed using the error metrics introduced previously, with a focus on trajectory accuracy, consistency, and motion stability. The quantitative benchmarking results for all methods and sequences are summarized in Table 3.

6.1 Performance in Extreme Occluded Scenario

The extreme occluded sequences (S0, S1) were designed to simulate highly challenging conditions for long-term indoor localization. These scenarios involved frequent occlusions from pedestrian movement and structural changes introduced through the construction of artificial corridors using poster walls. This tunnel-like configuration significantly reduced LiDAR visibility to static elements in the reference map, leading to degraded scan alignment. Among the evaluated methods, DDLO consistently achieved the highest accuracy, with ATE RMSE values of 2.22 cm (S0) and 2.84 cm (S1), demonstrating robustness under severe occlusions. FAST-LIO exhibited stable performance, though with slightly higher error. In contrast, KISS-SLAM and LIO-SAM were more affected by dynamic clutter, resulting in larger errors and reduced consistency. These findings highlight DDLO's effectiveness in handling dynamic indoor environments. Its performance advantage likely stems from the integration of dynamic point filtering, which mitigates the impact of moving agents and temporary structures during map updates and localization.

6.2 Performance in Semi Occluded Scenario

The semi occluded sequences (S2–S4) introduced moderate scene changes, including displaced poster walls and human activity such as walking, lingering, or interacting with visual elements. While the core structure remained partially visible, partial occlusions and environmental clutter introduced challenges to reliable pose estimation. In this setting, FAST-LIO outperformed other methods across all three sequences, achieving the lowest ATE RMSE (1.6 cm in S3), along with the most stable RPE and acceleration profiles. The method's tightly

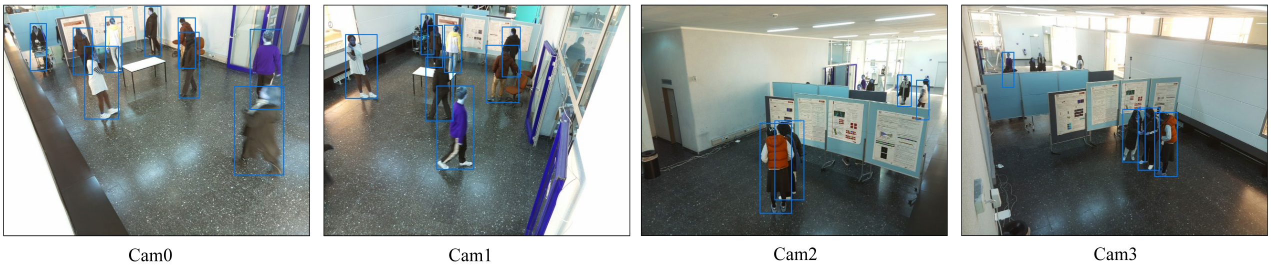

Figure 4. Examples of pedestrian detection from the four fixed cameras under varying viewpoints and occlusion conditions.

Sequence	Method	ATE RMSE [cm]	ATE <5 cm [%]	ATE <10 cm [%]	RPE mean [cm]	Acc RMS [m/s^2]
S0 - extreme occluded	DDLO	**2.222**	**99.181**	**100**	**1.959**	**0.301**
	FAST-LIO	3.202	89.401	100	2.448	0.303
	KISS-SLAM	5.162	63.203	99.144	4.449	0.887
	LIO-SAM	5.229	68.936	93.279	2.278	0.360
S1 - extreme occluded	DDLO	**2.84**	**95.652**	**100**	**2.611**	**0.344**
	FAST-LIO	3.34	91.255	100	2.76	0.349
	KISS-SLAM	5.146	61.447	99.901	4.01	0.909
	LIO-SAM	4.168	78.702	99.847	2.704	0.382
S2 - semi occluded	DDLO	2.198	100	100	2.556	0.325
	FAST-LIO	**2.076**	**100**	**100**	**2.223**	**0.270**
	KISS-SLAM	2.564	97.619	100	2.644	0.822
	LIO-SAM	2.859	95.635	100	2.551	0.331
S3 - semi occluded	DDLO	1.73	100	100	1.6	0.256
	FAST-LIO	**1.596**	**100**	**100**	**1.421**	**0.198**
	KISS-SLAM	3.509	77.485	100	3.354	0.594
	LIO-SAM	2.033	100	100	1.677	0.252
S4 - semi occluded	DDLO	2.076	99.481	100	1.989	0.319
	FAST-LIO	**1.752**	**100**	**100**	**1.720**	**0.224**
	KISS-SLAM	1.777	100	100	2.075	0.665
	LIO-SAM	2.642	98.875	100	1.967	0.299
S5 - free space	DDLO	2.026	98.795	100	2.111	0.306
	FAST-LIO	**1.696**	**100**	**100**	**1.875**	**0.253**
	KISS-SLAM	1.812	100	100	2.135	0.726
	LIO-SAM	2.438	82.018	99.890	2.216	0.305
S6 - free space	DDLO	2.182	99.692	100	2.525	0.347
	FAST-LIO	**1.925**	**100**	**100**	**2.358**	**0.279**
	KISS-SLAM	2.075	100	100	2.572	0.758
	LIO-SAM	2.88	81.274	100	2.765	0.318
S7 - free space	DDLO	2.18	99.65	100	2.399	0.354
	FAST-LIO	**1.95**	**100**	**100**	**2.217**	**0.269**
	KISS-SLAM	2.035	99.767	100	2.515	0.830
	LIO-SAM	2.914	95.804	100	2.578	0.342

Table 3. Benchmarking results for all sequences. Best value in each column per sequence is in bold.

coupled LiDAR–IMU integration enabled robust performance despite moderate disturbances. DDLO also maintained strong performance, with ATE values below 2.1 cm across all sequences, although it showed slightly less stability in RPE and motion consistency. KISS-SLAM and LIO-SAM continued to show weaker performance, likely due to limited dynamic handling and reduced sensor fusion capabilities. These results suggest that in semi occluded environments, robust IMU integration and motion compensation, as employed by FAST-LIO, are more effective than dynamic filtering alone.

6.3 Performance in Free Space Scenario

The Free Space sequences (S5–S7) represent ideal conditions with minimal occlusions and a static environment. Under these circumstances, all SLAM methods achieved acceptable performance. FAST-LIO delivered the highest accuracy, with ATE RMSE values consistently below 1.9 cm and the lowest RPE across sequences. DDLO also produced reliable results, though with slightly higher error values. KISS-SLAM and LIO-SAM performed adequately, with minor inconsistencies appearing in sequences containing occasional human presence.

These results confirm that in structured and stable environments, sensor fusion-based methods such as FAST-LIO provide the most accurate and consistent localization. However, in highly dynamic scenarios, characterized by severe occlusions and significant scene changes, methods that incorporate dynamic point removal, such as DDLO, demonstrate clear advantages. This contrast highlights the increasing importance of dynamic-aware processing as environmental complexity grows, and underscores the need for SLAM systems to adapt to varying degrees of scene dynamics in real-world indoor applications.

7. Conclusion

This study introduced a novel dataset for benchmarking long-term indoor localization under controlled dynamic conditions. Experimental results across varying scenarios demonstrate how different types and degrees of scene dynamics; such as occlusions, structural changes, and pedestrian activity, directly impact the performance of SLAM systems. These findings emphasize the need for adaptive localization strategies capable of handling real-world environmental variability. By capturing multiple sequences in a single environment with systematically varied levels of occlusion and scene modification, we isolate the effect of environmental changes on localization performance. The combination of static TLS-based mapping, high-precision prism tracking, and synchronized multi-sensor data provides a robust foundation for evaluating SLAM approaches in both short-term and long-term scenarios. Beyond localization, the dataset supports broader research directions. The overhead camera views capture natural human behaviors; such as walking, standing, and interacting near posters, providing a valuable resource for pedestrian behavior modeling and human–robot interaction studies. By integrating ego-centric LiDAR data with third-person visual information, this dataset offers a rich foundation for advancing robust, long-term autonomy in dynamic and populated indoor environments.

8. References

Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M.W. and Siegwart, R., 2016. The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research*, 35(10), pp.1157–1163.

Ceriani, S., Fontana, G., Giusti, A., Marzorati, D., Matteucci, M., Migliore, D., Rizzi, D., Sorrenti, D. and Taddei, P., 2009. Rawseeds ground truth collection systems for indoor self-localization and mapping. *Autonomous Robots*, 27(4), pp.353–371.

Fallon, M.F., Johannsson, H., Kaess, M. and Leonard, J.J., 2013. The MIT Stata Center dataset. *The International Journal of Robotics Research*, 32(14), pp.1695–1701.

Gao, L., Liang, Y., Yang, J., Wu, S., Wang, C., Chen, J. and Kneip, L., 2022. VECtor: A Versatile Event-Centric Benchmark for Multi-Sensor SLAM. *IEEE Robotics and Automation Letters*, 7(3), pp.8217–8224.

Geiger, A., Lenz, P. and Urtasun, R., 2012, June. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354–3361. IEEE.

Guadagnino, T., Mersch, B., Gupta, S., Vizzo, I., Grisetti, G. and Stachniss, C., 2025. KISS-SLAM: A Simple, Robust, and Accurate 3D LiDAR SLAM System With Enhanced Generalization Capabilities. *arXiv preprint arXiv:2503.12660*.

Jiao, J., Wei, H., Hu, T., Hu, X., Zhu, Y., He, Z., Wu, J., Yu, J., Xie, X., Huang, H., Geng, R. and Liu, M., 2022. FusionPortable: A Multi-Sensor Campus-Scene Dataset for Evaluation of Localization and Mapping Accuracy on Diverse Platforms. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp.4560–4567.

Kaveti, P., Gupta, A., Giaya, D., Karp, M., Keil, C., Nir, J., Zhang, Z. and Singh, H., 2023. Challenges of Indoor SLAM: A Multi-Modal Multi-Floor Dataset for SLAM Evaluation. In *Proceedings of the IEEE International Conference on Automation Science and Engineering (CASE)*, pp.1234–1240.

Lichtenfeld, J., Daun, K. and von Stryk, O., 2024, October. Efficient Dynamic LiDAR Odometry for Mobile Robots with Structured Point Clouds. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp.10137–10144. IEEE.

Najibi, M., Ji, J., Zhou, Y., Qi, C.R., Yan, X., Ettinger, S. and Anguelov, D., 2022, October. Motion inspired unsupervised perception and prediction in autonomous driving. In European Conference on Computer Vision (pp. 424-443). Cham: Springer Nature Switzerland.

Pronobis, A. and Caputo, B., 2009. COLD: The CoSy Localization Database. *The International Journal of Robotics Research*, 28(5), pp.588–594.

Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C. and Rus, D., 2020. LIO-SAM: Tightly-coupled LiDAR Inertial Odometry via Smoothing and Mapping. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp.5135–5142. IEEE.

Shi, X., Li, D., Zhao, P., Tian, Q., Tian, Y., Long, Q., Zhu, C., Song, J., Qiao, F., Song, L., Guo, Y., Wang, Z., Zhang, Y., Qin, B., Yang, W., Wang, F., Chan, R.H.M. and She, Q., 2020. Are We Ready for Service Robots? The OpenLORIS-Scene Datasets for Lifelong SLAM. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp.3130–3136.

Singh, G., Wu, M., Do, M.V. and Lam, S.K., 2022. Fast semantic-aware motion state detection for visual SLAM in dynamic environment. IEEE Transactions on Intelligent Transportation Systems, 23(12), pp.23014-23030.

Sturm, J., Engelhard, N., Endres, F., Burgard, W. and Cremers, D., 2012. A benchmark for the evaluation of RGB-D SLAM systems. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp.573–580.

Trekel, N., Guadagnino, T., Läbe, T., Wiesmann, L., Aguiar, P., Behley, J. and Stachniss, C., 2025. Benchmark for evaluating long-term localization in indoor environments under substantial static and dynamic scene changes. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).*

Wang, C., Luo, B., Zhang, Y., Zhao, Q., Yin, L., Wang, W., Su, X., Wang, Y. and Li, C., 2020. DymSLAM: 4D dynamic scene reconstruction based on geometrical motion segmentation. IEEE Robotics and Automation Letters, 6(2), pp.550-557.

Wang, Y., Chen, Y. and Zhang, Z.X., 2022. 4d unsupervised object discovery. Advances in Neural Information Processing Systems, 35, pp.35563-35575.

Xu, W., Cai, Y., He, D., Lin, J. and Zhang, F., 2021. FAST-LIO2: Fast Direct LiDAR–inertial Odometry. *arXiv preprint arXiv:2107.06829.*

Zhang, L., Helmberger, M., Fu, L., Wisth, D., Camurri, M., Scaramuzza, D. and Fallon, M., 2022. Hilti-Oxford Dataset: A Millimeter-Accurate Benchmark for Simultaneous Localization and Mapping. *IEEE Robotics and Automation Letters*, 8(1), pp.408–415.