Framework for Low-Cost Indoor Smartphone Camera-based Localization Using Spherical Panorama and BIM

Max Jwo Lem Lee, Ju Lin, Guohao Zhang, *Member, IEEE*, and Li-Ta Hsu, *Senior Member, IEEE*

***Abstract*—This paper presents a cost-effective framework for indoor localization that integrates spherical panoramas, smartphone cameras, and Building Information Modeling (BIM). We propose two methods for panorama pose estimation: Panorama Pose Estimation via Attribute Matching (PPE-AM) and Panorama Pose Estimation via Genetic Algorithm (PPE-GA). PPE-AM provides an initial pose estimate by matching attributes extracted from the panorama with those rendered from the BIM. PPE-GA refines this estimate using a genetic algorithm that aligns semantic and depth maps. Our approach eliminates the need for pre-mapping and on-site data collection, offering a practical alternative to expensive solutions such as spherical panorama localization based on point cloud or 3D line maps. Experimental results in a typical indoor environment demonstrate the system's accuracy, with PPE-AM achieving a mean 3D error of 0.74 meters, and PPE-GA further improving it to 0.47 meters. Smartphone Pose Estimation (SPE), leveraging the localized panoramas, achieved a mean 3D error of 0.92 meters. The proposed framework offers a robust and scalable solution for accurate indoor localization, with potential applications in augmented reality, facility management, and digital twin technologies.**

***Index Terms*—Indoor Positioning, Spherical Panorama, BIM Model, Camera-based Localization, Genetic Algorithm**

I. Introduction

THE construction and facilities management industry is undergoing a significant transformation driven by the rapid integration of automation technologies, leading to enhanced operational efficiency and a reduction in human errors [1]. A cornerstone of this advancement is the development of accurate digital twins—virtual representations that mirror physical assets and environments—enabling seamless integration between the physical and digital realms, particularly under dynamic operational conditions [2]. However, achieving precise and reliable camera-based localization remains a critical bottleneck in realizing the full potential of digital twins within these industries.

Various methodologies have been explored to determine the position of a smartphone image, each with inherent strengths and weaknesses. Traditional approaches such as image registration, Global Navigation Satellite Systems (GNSS), and Light Detection and Ranging (LiDAR) have been extensively investigated [3]–[6]. Image registration techniques, while conceptually straightforward, are prone to cumulative errors, limiting their accuracy and robustness. GNSS, despite its widespread availability, suffers from signal attenuation and multipath effects, particularly in dense urban and indoor environments, rendering it unreliable for precise indoor localization [7]. LiDAR technology offers high accuracy and detailed spatial information, but its adoption is often hindered by the high sensor cost and operational complexities associated with data acquisition and processing, especially for large-scale or continuous monitoring applications [8].

In response to the limitations of traditional methods, camera-based visual localization techniques, including Visual Odometry (VO) and Simultaneous Localization and Mapping (SLAM), have emerged as promising alternatives for indoor positioning [9], [10]. VO sequentially estimates camera motion by analyzing frame-to-frame changes in visual information, while SLAM concurrently builds a map of the environment and localizes the camera within it. Despite their potential, VO and SLAM methods are susceptible to drift accumulation over extended trajectories, demanding substantial computational resources, and often falter in environments characterized by repetitive visual patterns or suboptimal lighting conditions [11], [12]. More recently, panoramic localization techniques, exemplified by algorithms like image matching panorama localization [13], Colored Point Cloud-based Localization (CPO) [14] and Feature-based Geometric Panorama Localization (FGPL) [15], have leveraged spherical panoramas to enhance camera pose estimation. These methods have demonstrated improvements in accuracy and runtime efficiency; however, they typically rely on pre-existing 3D point cloud data, often acquired through costly and time-consuming LiDAR scanning or photogrammetry, introducing significant data acquisition overhead and operational complexities [20]. Thus, while panoramic approaches offer advancements, their dependence on extensive pre-captured 3D data limits their scalability and cost-effectiveness for widespread deployment.

Building Information Modeling (BIM) has increasingly become a focal point in indoor positioning research, primarily due to its potential to mitigate the costs and complexities associated with localization in intricate indoor spaces [16]. Unlike conventional localization methods that necessitate extensive mapping and manual data collection, BIM provides a comprehensive digital representation of a building's physical and functional attributes. This rich, pre-existing digital framework

Manuscript received March 6, 2025; revised Month Date, 2025. *(Corresponding author: Guohao Zhang.)*

Max Jwo Lem Lee, Ju Lin, Guohao Zhang, and Li-Ta Hsu are with the Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University, Hong Kong (e-mails: maxjl.lee@connect.polyu.hk; ju.lin@connect.polyu.hk; gh.zhang@polyu.edu.hk; lt.hsu@polyu.edu.hk).

Digital Object Identifier

offers the opportunity to circumvent repetitive mapping efforts and substantially reduce both time and financial investments in localization infrastructure [17]. Furthermore, recognizing its significant benefits for project lifecycle management and operational efficiency, BIM adoption has been mandated for building projects in many countries worldwide, making it a readily available and increasingly standardized resource. Lee et al.'s Building Information Positioning System (BIPS) exemplifies this direction, integrating smartphone-based visual localization with BIM models to enhance indoor positioning accuracy. Their approach utilizes deep learning to extract and compare semantic information from smartphone images with BIM data, aiming to improve the precision and reliability of indoor navigation [1]. However, the practical deployment of such BIM-integrated methods in real-world scenarios encounters significant challenges. Discrepancies between the as-built environment and the BIM model, coupled with potential feature scarcity in certain BIM representations, can severely undermine the effectiveness of these systems [18]. These limitations highlight the need for robust and adaptable BIM-integrated localization strategies that can perform reliably under varied and less controlled real-world conditions. While recent BIM-integrated systems like BIPS [17], demonstrate smartphone-BIM fusion for localization, our framework introduces three critical innovations:

- *Mapping-free Panoramic BIM Attribute Matching (PPE-AM):* PPE-AM estimates initial pose by matching semantic attributes (object class, spherical centroid, region size) from captured panoramas with BIM-rendered attributes. Unlike CPO [14] and FGPL [15], which require pre-mapped point clouds or 3D lines, PPE-AM needs no pre-mapping or on-site data collection, yet achieves comparable accuracy by leveraging structured BIM semantics.
- *BIM-based Panorama-Mediated Smartphone Positioning:* Our two-stage approach first localizes spherical panoramas using the BIM, exploiting their richer environmental information for higher accuracy than direct smartphone-to-BIM matching. These localized panoramas then act as a high-fidelity reference for smartphone Perspective-n-Point (PnP) positioning, mitigating feature scarcity issues encountered in direct BIM matching (e.g., BIPS [17]).
- *Hybrid Semantic and Depth Optimization via Genetic Algorithm (PPE-GA):* PPE-GA refines the PPE-AM pose by minimizing a weighted combination of semantic and depth map discrepancies between captured and BIM-rendered panoramas. This hybrid approach, using a genetic algorithm (GA), leverages complementary data strengths and avoids local minima common in gradient-based methods. Weights were determined empirically via CDF analysis of correlations with pose differences, ensuring a balanced, data-driven optimization.

The remainder of this paper is organized as follows: Section II presents the proposed solution, Section III details the methodology, Section IV discusses the experimental results, and Section V concludes the paper.

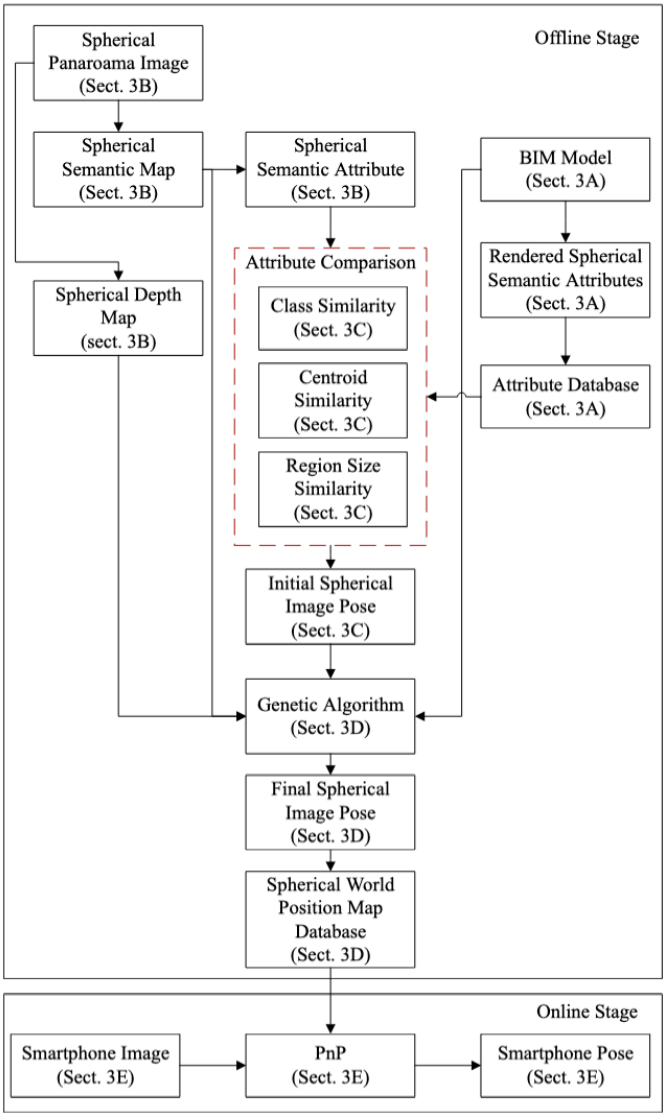


Fig. 1. Overview of the proposed solution.

II. Proposed Solution

The proposed solution for enhancing indoor smartphone camera-based localization involves two main stages as shown in Figure 1: offline and online.

A. Offline Stage

During the offline stage, the process begins with obtaining the BIM from The Hong Kong Polytechnic University. This BIM model is crucial as it provides a detailed 3D representation of the building, incorporating extensive information about each element, categorized by class. These classes are extracted to create a comprehensive list that aids in classifying real-world objects in subsequent stages.

To facilitate indoor localization, spherical semantic attributes are synthetically rendered from the BIM model at a granularity of 2×2 m. These geo-tagged attributes are stored in a database, establishing a foundational layer for later stages involving spherical panorama matching and localization.

The next step involves capturing spherical panorama images with a camera, which provide a 360-degree view of the indoor environment. These images are captured from various locations within the building. Our system is designed to automatically estimate the initial pose of these images using attributes derived from the semantics of the captured panoramas.

The initial pose of each spherical image is then refined iteratively using the genetic algorithm. This method minimizes discrepancies between the captured and BIM-rendered panoramas by aligning their semantic and depth maps. Following this refinement, a spherical world position map is generated from the BIM model using the refined pose, enabling each pixel from the captured panorama to correspond with a world position.

B. Online Stage

In the online stage, new incoming images from smartphones are matched directly against the captured panoramas. Since these panoramas now contain world position information, feature matching and PnP techniques are employed to achieve pose estimation of the smartphone images.

This two-stage process ensures a robust framework for indoor localization, leveraging detailed information from BIM models and advanced image processing techniques.

C. Definitions of Key Concepts

- **Semantic Map**: Maps each azimuth and elevation point within a panorama with semantic labels (e.g., 'wall', 'door'), offering a class representation in spherical coordinates.
- **Depth Map**: Maps the metric distance of scene objects to each azimuth and elevation point in spherical coordinates.
- **Semantic Attribute**: Records detailed attributes of objects—including class, centroid and size within spherical coordinates. This structured data streamlines the matching process and enhances computational efficiency.
- **World Position Map**: Maps each azimuth and elevation point in panorama images to 3D coordinates in spherical coordinates, essential for accurate smartphone localization and navigation.

III. Methodology

A. Attribute Extraction from BIM Model

The methodology begins by acquiring the BIM model of the Hong Kong Polytechnic University, as depicted in Figure 2. This model provides a detailed 3D digital representation of physical and terrain structures.

This comprehensive BIM model is employed in various applications, from land assessment to detailed engineering presentations. In this study, the model is specifically used for pose estimation in static and geometrically defined indoor environments using spherical panoramas.

Within the BIM model, each object is embedded with metadata that facilitates the extraction of specific classes, ensuring consistency and precision with minimal preprocessing. To accurately generate and extract attributes from the BIM model, we utilize Blender, an open-source 3D creation suite [21]. Blender's powerful scripting and rendering capabilities allow us to automate the creation of a structured 2D grid system that overlays the BIM model, characterized by a uniform separation of 2x2 meters at a fixed elevation (z) of 1.5 meters above ground level and fixed heading (θ) of 0 degrees, as shown in Figure 3. At each grid intersection, defined by coordinates (x, y, z, θ), attributes are systematically generated using Blender and stored in an organized database $\mathcal{A}$. The attribute generation process is defined as follows:

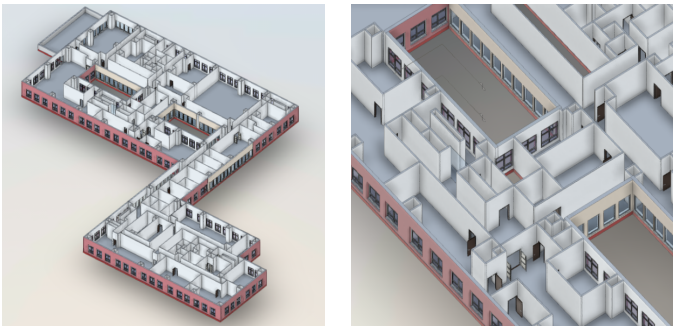


Fig. 2. BIM Model of The Hong Kong Polytechnic University Campus Z Core floor: (Left) An overview and (Right) a detailed close-up view.

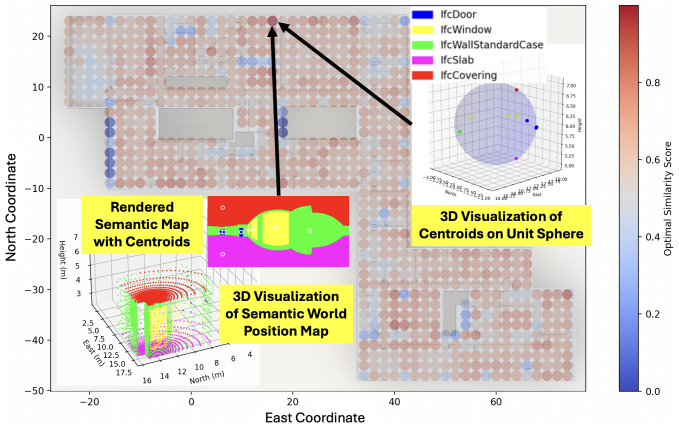


Fig. 3. 2D grid system on the BIM model with blue dots indicating sampling points at a 1.5-meter elevation.

$$\mathcal{A}(x, y, 1.5, 0) = \Big\{(\text{Class}, \text{Centroid}_{\phi}, \text{PixelSize})_{obj} \mid obj \in \text{Objects at grid point}(x, y, 1.5, 0)\Big\} \quad (1)$$

For each object obj within the line of sight at the grid point $(x, y, 1.5, 0)$, attributes are defined as follows to facilitate modeling and interaction:

- **Class**: Determined based on object's metadata, aiding in accurate categorization within the classes of the BIM model.
- **Centroid**: Calculated as the average spherical position of all pixels constituting the object in azimuth (ϕ) and elevation (ψ). An example is shown in the 3D visualization of centroids on unit sphere in Figure 3.

- **Region Size**: Counts the total number of pixels making up the object, crucial for assessing the object's scale and visual prominence.

These attributes are meticulously cataloged in $\mathcal{A}(x, y, 1.5, 0)$ for each line-of-sight object, providing the necessary detail for rough pose estimation through matching with captured spherical panoramas. Figure 3 also presents the rendered semantic map of a specific location. An example of the attributes for various objects in the rendered semantic map are tabulated in Table I:

TABLE I
ATTRIBUTES OF VARIOUS OBJECTS EXTRACTED FROM THE BIM-RENDERED SEMANTIC MAP

Class	Centroid Azimuth (ϕ)	Centroid Elevation (ψ)	Region Size
Door	85	95	51
Door	36	93	45
Door	41	93	58
WallStandardCase	260	93	17481
Covering	38	31	21451
Slab	38	149	19077
Window	173	93	5645
Window	118	92	849
Window	104	91	143

Table I enumerates the attributes of various objects within a specific location, as generated from the BIM model. The table does not follow predefined classes; instead, the classes emerge from the BIM model's specifications. Each row in the table corresponds to distinct objects identified within the line of sight, detailing their centroids in azimuth (ϕ) and elevation (ψ) and the region size they occupy. This comprehensive tabulation plays a crucial role in the efficient matching of real-world captured panoramas' attributes with their corresponding digital attributes in the BIM model. By cataloging and utilizing these attributes directly rather than storing the full semantic map, the system significantly reduces data storage requirements and enhances computational efficiency. This approach not only streamlines the matching process but also ensures scalability and faster response times. Next, we describe how semantic maps and attributes are extracted from spherical panorama images.

B. Semantic Map and Attribute Extraction from Spherical Panorama Image

Spherical panorama images capture a 360-degree view of the environment as shown in Figure 4, crucial for camera-based localization and environmental understanding. These images are captured using the Insta360 X3, which provide detailed visuals of the surroundings [22]. The images are then processed to extract semantic maps, attributes and depth maps that align with those generated from the BIM model. The segmentation is enhanced by integrating a Segment Anything Model 2 (SAM 2) that leverages contextual information and detailed object classes from the BIM model to improve segmentation accuracy [23]. Additionally, depth estimation is refined using the Depth Anything V2, which advances the state-of-the-art in depth prediction by enabling more depth map generation across diverse scenes and objects [24].

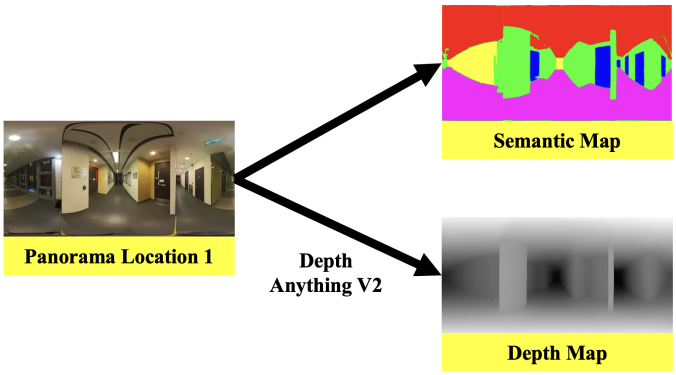


Fig. 4. Example of predicted Semantic and Depth map extraction.

$$\mathcal{S}_p = \text{Segmentation}(\text{Image}_p, \text{Classes}_{\text{BIM}}) \tag{2}$$

In this equation, $\mathcal{S}_p$ represents the semantic map obtained from the spherical panorama, Image_p is the input image from the camera, and $\text{Classes}_{\text{BIM}}$ is the object classes from the BIM model. The function Segmentation refers to the process of labeling each pixel in the image with a class from the predefined set in the BIM model.

Following the segmentation, attributes are extracted for each identified object within the panorama. These attributes include the object's class, centroid, region size, similar to those extracted from the BIM model. The centroid calculation for spherical images considers the spherical coordinates, converting them into azimuth (ϕ) and elevation (ψ) for consistency with the BIM attribute system.

$$\mathcal{A}_p(x_{\text{unk}}, y_{\text{unk}}, z_{\text{unk}}, \theta_{\text{unk}}) = \Big\{(\text{Class}, \text{Centroid}_\phi, \text{Centroid}_\psi, \text{PixelSize})_{\text{obj}} \mid \text{obj} \in \text{Objects in } \mathcal{S}_p\Big\} \tag{3}$$

Here, $\mathcal{A}_p(x_{\text{unk}}, y_{\text{unk}}, z_{\text{unk}}, \theta_{\text{unk}})$ denotes the set of attributes extracted from the spherical panorama at the unknown coordinates $(x_{\text{unk}}, y_{\text{unk}}, z_{\text{unk}})$ and unknown heading (θ_{unk}), emphasizing that these are the parameters to be determined.

C. Attribute Comparison

The attribute comparison between the spherical panoramas ($\mathcal{A}_p$) and the BIM model database ($\mathcal{A}$) is crucial for the initial estimation of the camera pose. This step entails aligning the centroid of each object in the panorama with the closest centroid of the same class from the BIM model. The matching criterion is the smallest angular distance in spherical coordinates, with ϕ representing azimuth and ψ representing elevation. This approach ensures that the comparison leverages the most geographically and categorically relevant objects, thereby increasing the accuracy of the pose estimation.

Subsequently, the process optimizes the camera heading θ for each considered camera position (x, y, z) to enhance the similarity score. This can be represented as a single optimization problem:

$$\max_{x,y,z,\theta} \text{Similarity}(\mathcal{A}_p, \mathcal{A}(x, y, z, \theta)) \tag{4}$$

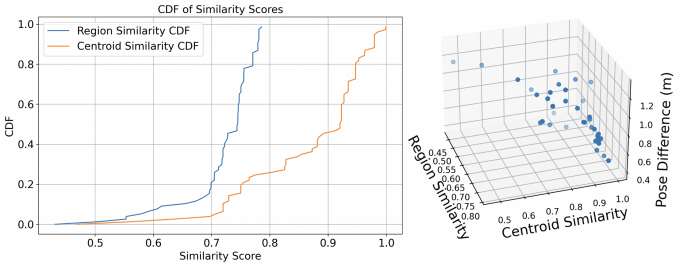


Fig. 5. Combined plot showing the CDF of Region and Centroid Similarity Scores, and the 3D Scatter plot of Region Similarity vs. Centroid Similarity vs. Pose Difference.

The optimal camera pose is then selected based on the highest similarity score obtained from this optimization across all the positions. Figure 3 depicts this selection process through a heatmap, showcasing how each potential position correlates with a specific heading that yields the maximum similarity score.

where the similarity function is defined by:

$$\text{Similarity}(\mathcal{A}_p, \mathcal{A}(x,y,z,\theta)) = w_{\text{centroid}} \cdot S_{\text{centroid}} + w_{\text{size}} \cdot S_{\text{size}} \tag{5}$$

The components of the similarity function include:

- **Centroid Similarity** (S_{centroid}):

$$S_{\text{centroid}} = 1 - \frac{1}{\pi} \arccos(\sin\psi_{\mathcal{A}_p} \sin\psi_{\mathcal{A}} + \cos\psi_{\mathcal{A}_p} \cos\psi_{\mathcal{A}} \cos(\phi_{\mathcal{A}_p} - \phi_{\mathcal{A}})) \tag{6}$$

- **Region Size Similarity** (S_{size}):

$$S_{\text{size}} = 1 - \frac{|\text{Pixel Size}_{\mathcal{A}_p} - \text{Pixel Size}_{\mathcal{A}}|}{\max(\text{Pixel Size}_{\mathcal{A}_p}, \text{Pixel Size}_{\mathcal{A}})} \tag{7}$$

Through CDF analysis, we have determined the correlation between centroid similarity and pose difference to be -0.7546, indicating a strong inverse relationship. In contrast, the correlation between region size similarity and pose difference is -0.3456, which suggests a moderate inverse relationship. These correlations are illustrated in Figure 5.

To define the weights for centroid and region size similarities, which are crucial for improving the alignment of spherical panoramas with the BIM model, we calculate the normalized weights based on the absolute values of these correlations. The heuristically fixed normalized weights are computed as follows:

$$w_{\text{centroid}} = \frac{|\text{-}0.7546|}{|\text{-}0.7546| + |\text{-}0.3456|} \approx 0.69,$$
$$w_{\text{size}} = \frac{|\text{-}0.3456|}{|\text{-}0.7546| + |\text{-}0.3456|} \approx 0.31.$$

These weights form a combined metric for enhancing the alignment of spherical panoramas with the BIM model. As the evaluation of this metric is conducted offline, it allows for thorough optimization without the constraints of real-time processing.

D. Pose Refinement Using Genetic Algorithm

After initial pose estimation from attribute comparison, pose refinement is performed using a genetic algorithm. This method leverages discrepancies between observed and rendered semantic and depth maps extracted from a spherical panorama and the BIM model, respectively. The genetic algorithm was selected for pose refinement after evaluating alternative optimization methods, such as gradient-based approaches and non-linear least squares, which struggled with the non-linear and multi-modal nature of the problem. These methods often converged to local minima or required precise initial guesses, which were difficult to guarantee in complex indoor environments. In contrast, the genetic algorithm's population-based exploration and robustness to noise enabled it to handle the hybrid objective function (combining semantic and depth discrepancies) effectively, avoiding premature convergence and achieving higher accuracy. This adaptability, along with its ease of implementation, made it the most suitable choice for refining pose estimates in our framework.

Semantic and depth maps are rendered from the currently estimated pose, denoted as $S_{\text{rendered}}(\phi,\psi)$ and $D_{\text{rendered}}(\phi,\psi)$, covering the full panorama FOV (360° by 180°). The actual maps from the spherical panorama, $S_{\text{p}}(\phi,\psi)$ and $D_{\text{p}}(\phi,\psi)$, are compared to the rendered maps.

To quantify the discrepancies between the rendered and actual semantic and depth maps, we define ΔS and ΔD as follows:

$$\Delta S = \sum_{\phi=0}^{360} \sum_{\psi=-90}^{90} |S_{\text{rendered}}(\phi,\psi) - S_{\text{p}}(\phi,\psi)| \tag{8}$$

$$\Delta D = \sum_{\phi=0}^{360} \sum_{\psi=-90}^{90} |D_{\text{rendered}}(\phi,\psi) - D_{\text{p}}(\phi,\psi)| \tag{9}$$

The objective function for the genetic algorithm's optimization process is a weighted sum of these discrepancies:

$$\Delta T(x,y,z,\theta) = w_{semantic} \cdot \Delta S + w_{depth} \cdot \Delta D \tag{10}$$

where $w_{semantic}$ and w_{depth} are the weighting factors for the semantic and depth discrepancies, respectively. The genetic algorithm seeks to minimize this objective function by finding the optimal pose parameters—position (x,y,z) and heading (θ). The genetic algorithm is configured for efficient pose refinement, with a population size of 10 and a maximum of 30 generations, balancing search diversity with computational speed. Mutation and crossover rates are set at 0.1 and 0.5, respectively. The optimization process may halt early when $\Delta T(x,y,z,\theta)$ reaches 0.90 as benchmarked in Table II.

The refined pose from the genetic algorithm enhances the alignment with the actual capture conditions of the panorama. A world position map, denoted as W_{rendered}, is generated from the BIM model using this refined pose. This map is crucial for subsequent tasks such as accurate smartphone positioning and augmented reality applications.

To define weights for the pose estimation, we analyze semantic ($w_{semantic}$) and depth (w_{depth}) similarity scores with

TABLE II
COMPARISON OF CAPTURED AND RENDERED PANORAMAS ACROSS DIFFERENT ENVIRONMENTS

Loc	ΔT	Environment Description
1	0.92	Rich in semantic content
2	0.85	Rich in semantic content
3	0.88	Rich in semantic content
4	0.90	Rich in semantic content
5	0.87	Moderately semantic
6	0.95	Moderately semantic
7	0.89	Moderately semantic
8	0.86	Moderately semantic
9	0.91	Environmentally complex
10	0.94	Sparse semantic elements
Average	**0.90**	—

the captured panoramas, using a rendered dataset of 60 panoramas within 3 meter pose difference. The scores' distributions are visualized using cumulative distribution functions (CDFs). The CDFs, illustrated in Figure 6, reveal the rarity of high similarity scores, while the 3D scatter plot provides a visual correlation between the two types of similarities and pose differences.

Statistical measures, including mean, median, and standard deviation, describe the central tendency and dispersion of the scores. The semantic similarity scores have a mean of 0.5869, a median of 0.5826, and a standard deviation of 0.1045. For the depth similarity scores, the mean is 0.7440, the median is 0.7193, and the standard deviation is 0.1114.

The correlation coefficients ($\rho_{semantic}$ and ρ_{depth}) inform the predictive strength of the similarity scores with respect to pose discrepancies. For semantic similarity, the correlation with pose differences is $\rho_{semantic} = -0.3812$, while for depth similarity, it is $\rho_{depth} = -0.3955$.

To derive the weighting coefficients for semantic and depth similarities, we normalize the absolute values of these correlation coefficients, using the absolute values of the correlation coefficients, the heuristically fixed normalized weights are computed as follows:

$$w_{semantic} = \frac{|-0.3812|}{|-0.3812| + |-0.3955|} \approx 0.49,$$
$$w_{depth} = \frac{|-0.3955|}{|-0.3812| + |-0.3955|} \approx 0.51.$$

These weights are applied to the similarity scores to calculate a composite metric for pose estimation refinement. This fusion of similarity scores, by prioritizing the more informative measure through weighting, aims to improve the accuracy of the pose estimation process.

The final step of the pose estimation process involves generating a spherical world position map as shown in Figure 7, which is derived from the optimized pose parameters and the BIM model's data. This map encodes the 3D coordinates of each pixel in the captured spherical panorama image, relative to the reference frame of the BIM model. In essence, it provides a geospatial context for every point within the image, establishing a direct linkage between the captured visual data and the virtual model of the environment. From this, 45-degree FOV images and their corresponding 3D positions

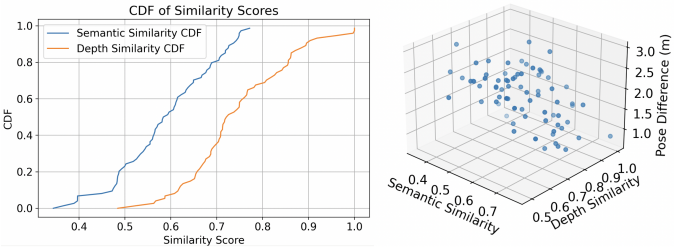


Fig. 6. Combined plot displaying the Cumulative Distribution Function (CDF) of Semantic and Depth Similarity Scores, and the associated 3D Scatter plot of Semantic Similarity vs. Depth Similarity vs. Pose Difference.

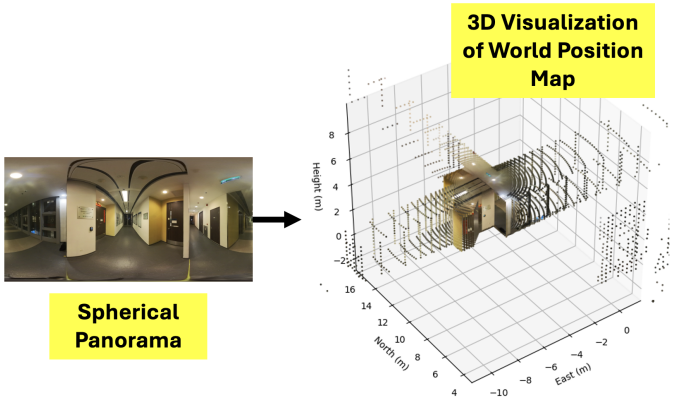


Fig. 7. 3D Visualization of World Position Map.

are extracted and pre-computed to store NetVLAD global descriptors and OmniGlue local features, enabling efficient real-time smartphone positioning [25].

E. Smartphone Positioning

For smartphone image localization, the system first performs a coarse search to identify the most similar database images based on pre-computed global descriptors. OmniGlue features are then extracted from the query image and matched with pre-computed features from the selected database images [26]. The smartphone pose is estimated using the PnP algorithm, which minimizes the re-projection error between 2D feature positions in the query image and their associated 3D coordinates from the 45-degree FOV images. This localization process is computationally efficient, requiring approximately 0.3 seconds (Table VII). While direct low-light accuracy metrics for OmniGlue are not provided, its design principles and overall performance suggest it is likely to be effective in low-light scenarios. While strategies such as adaptive thresholding or fine-tuned models could improve low-light performance, these enhancements fall outside the scope of this work.

The mathematical objective for minimizing the reprojection error is expressed as:

$$\min_{\mathbf{R},\mathbf{t}} \sum_i \|\mathbf{p}_i - \text{proj}\left(\mathbf{K}[\mathbf{R}|\mathbf{t}]\mathbf{P}_i\right)\|^2 \tag{11}$$

In this equation, $\mathbf{p}_i$ signifies the 2D feature points identified in the smartphone image, $\mathbf{P}_i$ represents the 3D coordinates from the 45-degree FOV images in the database, and $\mathbf{K}$ is

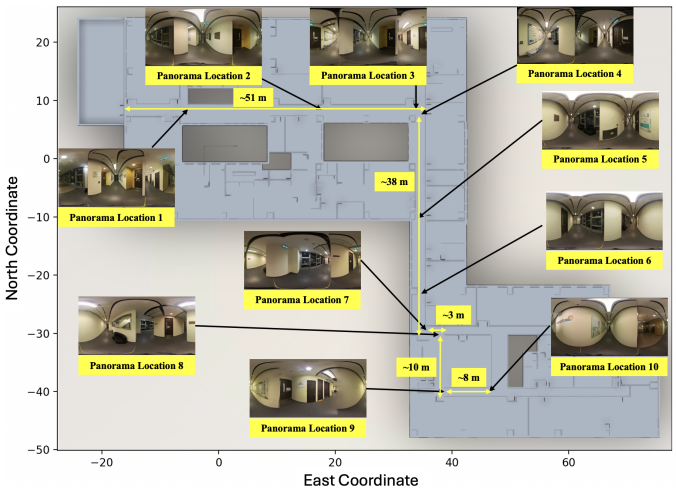


Fig. 8. Data Collection Pathway.

the camera calibration matrix that encompasses the intrinsic parameters of the smartphone's camera.

IV. Experimentation

TABLE III
Descriptions of Measurement Sensors

Method	Description and Hardware Used
Ground Truth (GT)	Data accuracy verification collected at landmark locations within 0.1 meter precision, alongside the indoor floor plan from The Hong Kong Polytechnic University.
GNSS [27]	The U-Blox F9P GNSS receiver is used for performance comparison, despite its limitations in indoor environments.
UWB [28]	The NoopLoop LinkTrack Ultra Wideband-based positioning system, though offering high-accuracy indoor localization through time-of-flight measurements of radio waves, is an expensive solution and serves as a benchmark for evaluating the performance of our more cost-effective proposed methods.
CPO [14]	Colored Point Cloud-based Localization for panoramic localization, which requires colored point clouds for pose estimation.
FGPL [15]	Feature-based Geometric Panorama Localization method for panoramic localization, which requires 3D line maps for pose estimation.
Stage 1: Panorama Pose Estimation by Attribute Matching (PPE-AM)	Utilizes spherical panorama matching based on the BIM model for initial pose estimation. This stage leverages attribute extraction and comparison from both the BIM model and captured panoramas to determine the most probable initial camera positions and headings.
Stage 2: Panorama Pose Estimation by Genetic Algorithm (PPE-GA)	Enhances the initial pose estimation through differential rendering. This method refines the pose estimation by aligning semantic and depth maps from the BIM model and the captured panoramas, thus improving the accuracy of the panorama localization.
Smartphone Pose Estimation (SPE)	Applies the localized spherical panoramas to precisely estimate smartphone poses. This involves matching features from smartphone images to features in the panorama, followed by employing PnP to accurately estimate the smartphone's pose.

A. Experimental Setup

The experimental setup was designed to assess the efficacy of the proposed indoor localization system, which utilizes spherical panorama images and smartphone images. The venue was a corridor within The Hong Kong Polytechnic University, selected due to its typical indoor navigation characteristics such as varied wall textures, multiple doorways, and consistent lighting conditions. For the experiment, a total of 10 spherical panorama images were captured at designated points along the corridor as shown in 8. Additionally, 10 smartphone images were acquired at the same location of the panorama capture points to evaluate the system's capability in accurately localizing smartphone images.

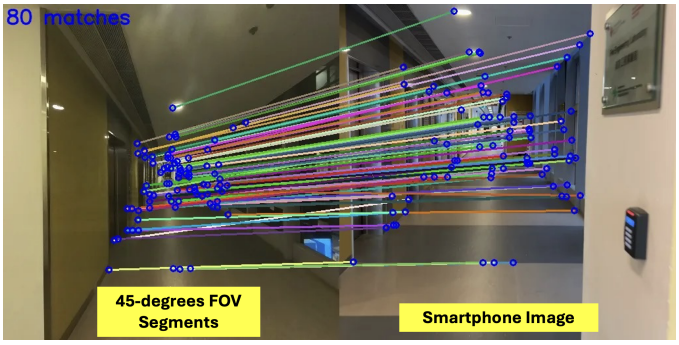


Fig. 9. Feature Correspondence Visualization.

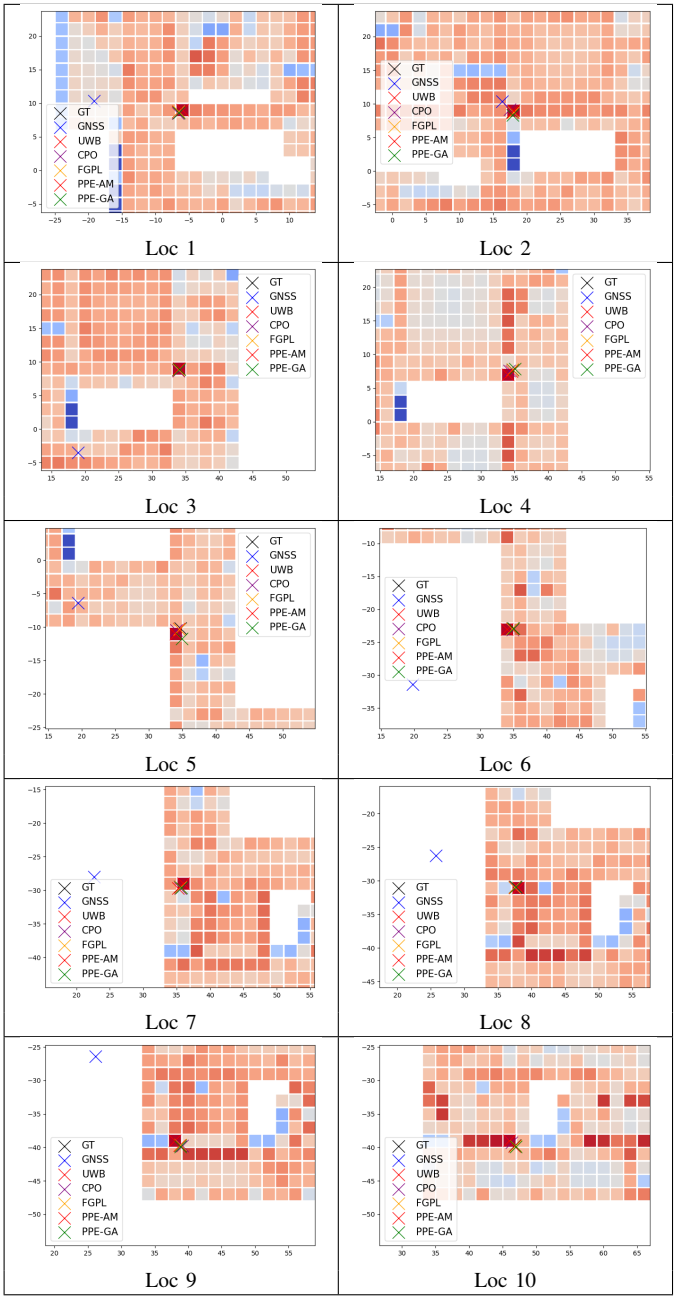
The evaluation process was divided into two main phases: first, determining the accuracy of pose estimation for spherical panorama images, and second, localizing smartphone images with respect to the panoramas. This structured experimental approach enabled an in-depth analysis of the system's performance in a controlled indoor environment. It is important to acknowledge that this experiment was conducted in a single corridor, which represents a limitation in terms of generalizability to more complex indoor environments. Future work will address this by evaluating the system in diverse settings, including larger spaces and multi-floor buildings.

Seven distinct measurement methodologies were employed to gather diverse data, which are summarized in Table III.

TABLE IV
Errors for Spherical Panorama Pose Estimation with GNSS, UWB, CPO, FGPL, PPE-AM (Stage 1), PPE-GA (Stage 2) and SPE

Loc	GNSS Error (m)	UWB Error (m)	CPO Error (m)	FGPL Error (m)	PPE-AM Error (m)	PPE-GA Error (m)	SPE Error (m)
1	12.72	0.17	0.35	0.10	0.67	0.02	0.29
2	2.17	0.05	0.45	0.12	0.22	0.53	0.75
3	19.46	0.20	0.28	0.09	0.22	0.29	0.51
4	21.50	0.19	0.50	0.15	0.99	0.75	1.54
5	15.84	0.07	0.65	0.25	1.06	1.46	2.94
6	17.47	0.08	0.30	0.11	0.91	0.30	0.48
7	13.04	0.20	0.40	0.10	0.71	0.31	0.44
8	12.66	0.18	0.32	0.08	0.51	0.26	0.39
9	18.35	0.17	0.55	0.16	1.06	0.55	1.33
10	27.81	0.07	0.48	0.13	1.06	0.28	0.56
Mean	16.10	0.14	0.43	0.13	0.74	0.47	0.92
Std Dev	6.41	0.06	0.11	0.05	0.32	0.38	0.78
Min	2.17	0.05	0.28	0.08	0.22	0.02	0.29
Max	27.81	0.20	0.65	0.25	1.06	1.46	2.94

TABLE V
PPE-AM SIMILARITY SCORE HEATMAPS WITH COMPARATIVE METHOD LOCATIONS



B. Panorama Pose Estimation Results

The GNSS-based localization methods, while ubiquitous in outdoor settings, have long been known to face significant challenges in indoor environments. The results in Table IV confirm these difficulties, with GNSS showing a mean error of 16.10 meters. In contrast, UWB technology, known for its high accuracy and robustness in indoor localization, performed significantly better with a mean error of 0.14 meters. The UWB system's ability to measure the time of flight of radio waves with high precision enables more accurate positioning even in complex indoor scenarios. However, despite its high accuracy, the cost of UWB technology becomes a major limiting factor, especially as the size of the deployment area grows. The installation, calibration, and maintenance of an extensive network of UWB beacons can be prohibitively expensive, making it a less attractive solution for large-scale indoor environments and cost-sensitive applications.

The performance of CPO and FGPL further underscores advancements in spherical panorama-based indoor localization techniques. CPO and FGPL leverages pre-existing colored point cloud data and 3D line maps to align captured panoramas. This approach achieves high precision of 0.43m and 0.13m by minimizing discrepancies between observed and rendered scenes, resulting in mean errors comparable to those achieved by UWB systems. However, the reliance on dense point clouds introduces significant data acquisition overhead, often requiring costly and time-consuming LiDAR scanning or photogrammetry processes. This dependency limits the scalability of CPO and FGPL for widespread deployment in large-scale or cost-sensitive applications.

The results of the experiment demonstrated the performance of PPE-AM and PPE-GA methods as shown in Table IV. The PPE-AM method achieved sub-meter accuracy, with a mean error of approximately 0.74 meters and a standard deviation of 0.32 meters. Visualization of the estimated poses, depicted in Table V, highlights the method's reliability and spatial consistency. PPE-AM expedites the localization process by comparing attributes extracted from both the BIM model and captured panoramas.

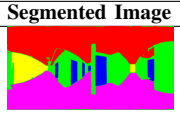
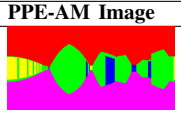
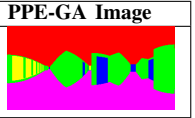
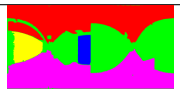
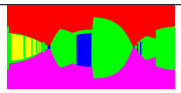
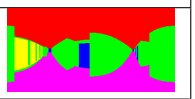
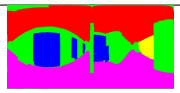
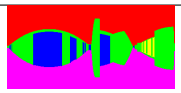
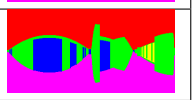
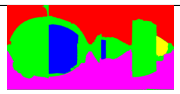
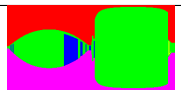
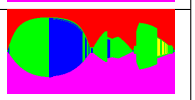
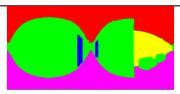
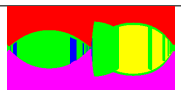
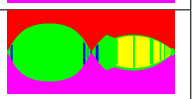
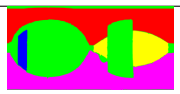
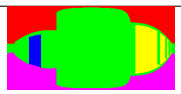
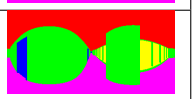
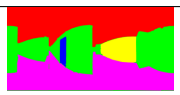
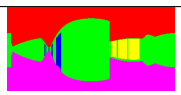
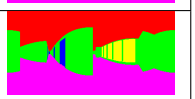
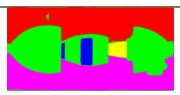
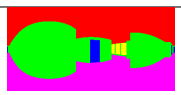
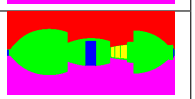
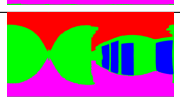
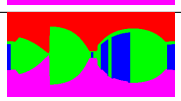
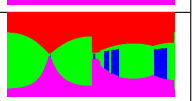
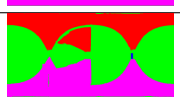
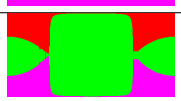
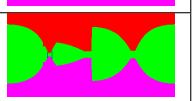
The PPE-GA method further refined the initial pose estimations from PPE-AM, with a mean error of 0.47 meters and a standard deviation of 0.38 meters. It employs a genetic algorithm to optimize the alignment of semantic and depth data between the BIM model and captured panoramas, as evidenced by the corrected images in Table VI.

These refined poses enable the generation of a detailed 3D map, where each pixel from the spherical panorama images is accurately associated with a corresponding location in the BIM model's coordinate system. This map serves as a foundational dataset for applications such as virtual reality, indoor navigation, and spatial analysis. Figure 10 provides a visual representation of the 3D map generated from the ten spherical panorama images using the refined poses from the PPE-GA method. This detailed point cloud highlights the precision and potential of the system in transforming 2D visual information into a comprehensive 3D spatial context.

In addition, the smartphone localization experiments were conducted to validate the accuracy of the proposed SPE method. The SPE method demonstrated a remarkable level of precision, achieving a mean positioning error of 0.92 meters with a standard deviation of 0.78 meters across all tested smartphone images, as summarized in Table IV. Localization errors for smartphone images at each capture point indicated that the SPE method maintained consistent accuracy. To further understand the impact of 3D position discrepancies, we conducted a sensitivity analysis, as discussed in Sec IV-C.

[0]For GNSS Error, the reported values are based on 2D horizontal positioning, as the Z-axis (vertical) accuracy is known to be significantly lower and less reliable indoors. All other reported errors (UWB, CPO, FGPL, PPE-AM, PPE-GA, SPE) represent 3D positioning accuracy.

TABLE VI
PANORAMA POSE ESTIMATION BY PPE-AM AND PPE-GA RENDERED IMAGES

Loc	Segmented Image	PPE-AM Image	PPE-GA Image
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			

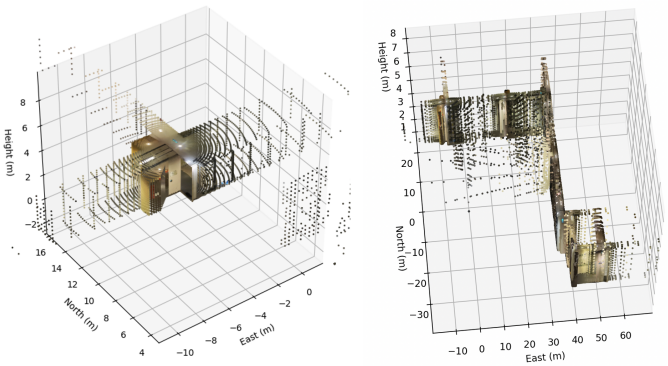


Fig. 10. 3D map generated from ten spherical panorama images using refined poses from the PPE-GA method: (Left) A single panorama image is converted into a 3D point cloud; (Right) An aggregate point cloud combining ten panorama images, demonstrating the comprehensive coverage and integration capabilities

To evaluate the computational efficiency of our proposed methods, we analyzed the trade-offs between accuracy, processing time, and resource requirements, with computations performed using a Nvidia GeForce RTX 3090 GPU. Table VII summarizes the key aspects of PPE-AM, PPE-GA, CPO, FGPL, and SPE. While CPO and FGPL achieve higher accuracy (0.43m and 0.13m, respectively), they require dense point clouds or 3D line features, resulting in significant preprocessing. While CPO and FGPL are less scalable for large environments, PPE-AM, PPE-GA, and SPE offer more cost-effective and scalable solutions. PPE-AM and PPE-GA are particularly suitable in settings with pre-existing BIM data, making them good choices for resource-constrained applications. Although the processing times for PPE-AM and PPE-GA are longer (and currently unoptimized), this is acceptable as these are offline computations; the resulting optimized panorama poses are leveraged for efficient, real-time SPE positioning.

TABLE VII
COMPUTATIONAL TRADE-OFFS SUMMARY

Aspect	CPO	FGPL	PPE-AM	PPE-GA	SPE
Data Representation	Dense colored point clouds	3D line features	Pre-computed BIM attributes	BIM Model	Pre-computed Panoramas with 3D Features
Pre-processing	High (3D mapping required)	High (3D mapping required)	Minimal (automated BIM extraction)	Minimal (same as PPE-AM)	Low (Panorama capture)
Pose Estimation	Dense point cloud alignment	Feature-based alignment	Attribute matching	Genetic algorithm optimization	3D Feature matching
Scalability	Limited by dense data	Moderately scalable	Highly scalable	Highly scalable	Highly Scalable
Accuracy (Mean Error)	0.43m	0.13m	0.74m	0.47m	0.92m
Cost	High (mapping)	High (mapping)	Low (BIM-based)	Low (BIM-based)	Low (PPE-GA-based)
Processing Time	2.2-2.4 seconds	1.1-1.3 seconds	1.2-1.4 seconds	6.3-10.9 seconds	0.3 second

Overall, the experimental results demonstrate that the proposed PPE-AM, PPE-GA and SPE methods offer a balance between accuracy, scalability, and cost-effectiveness.

C. Impact of 3D Position Discrepancies between Captured and Rendered Panoramas on PnP Accuracy

The accuracy of PnP pose estimation is fundamentally contingent upon the fidelity of the 3D coordinates of reference points used in the computation. Ideally, these 3D points are derived from a spherical world position map that precisely mirrors the physical environment captured in the panorama. However, discrepancies between the 3D positions of points in the captured panorama and the corresponding points in the rendered panorama—synthesized from the BIM model—can introduce significant errors into the PnP solution.

To quantify the impact of these discrepancies, we conducted a sensitivity analysis. We introduced controlled, synthetic

errors into the 3D coordinates of the reference points used in the PnP algorithm. These errors were introduced in the form of 3D Gaussian noise with varying standard deviations (σ) along each axis (x, y, z). The standard deviations were set to represent different levels of positional error: 0.05m, 0.10m, 0.25m, 0.50m, and 1.00m. For each error level, we performed 1000 PnP computations using different sets of reference points randomly selected from the world position map. The results of this analysis, summarized in Table VIII, demonstrate a correlation between the magnitude of 3D position discrepancies and the resulting error in the estimated camera pose. As the standard deviation of the introduced noise increased, the mean and standard deviation of the translational and rotational errors in the PnP solution also increased.

TABLE VIII
SENSITIVITY ANALYSIS OF PNP POSE ESTIMATION TO 3D POSITION ERRORS

3D Error	Translational Error (m)		Rotational Error (deg)	
Std Dev (σ)	Mean	Std Dev	Mean	Std Dev
0.05	0.08	0.03	0.5	0.2
0.10	0.15	0.06	1.1	0.4
0.25	0.38	0.15	2.8	1.0
0.50	0.75	0.30	5.5	2.1
1.00	1.51	0.62	11.2	4.3

For instance, with a standard deviation of 0.10m in the 3D position error, the mean translational error was found to be 0.15m, while the mean rotational error was 1.1 degrees. When the standard deviation of the 3D error was increased to 0.50m, the mean translational error grew to 0.75m and the mean rotational error to 5.5 degrees. These results highlight the sensitivity of the PnP algorithm to inaccuracies in the 3D reference point data. The observed errors can be attributed to:

1) **Misalignments between the actual scene and the BIM model:** These can arise from inaccuracies in the BIM model itself, changes in the environment post-BIM creation (e.g., moved furniture), or the presence of unmodeled elements.
2) **Errors in the panorama capture and rendering process:** Imperfect camera calibration, distortions in the spherical projection, or inaccuracies in the pose refinement process can contribute to discrepancies.
3) **Limitations of the feature matching process:** While OmniGlue is a robust feature matching algorithm, it may still produce some incorrect matches, especially in areas with repetitive patterns or poor lighting.

These findings underscore the critical importance of minimizing 3D position discrepancies to achieve accurate and reliable PnP-based SPE. The accuracy of the spherical world position map, the fidelity of the BIM model to the real-world environment, the precision of the panorama capture, and rendering process directly impact the system performance.

V. CONCLUSION

This paper introduces a cost-effective framework for indoor localization using spherical panoramas, smartphone imagery, and BIM. Our proposed methods, PPE-AM and PPE-GA, demonstrate superior cost-effectiveness compared to traditional techniques like UWB, CPO and FGPL in indoor environments. Experiments at The Hong Kong Polytechnic University yielded mean localization errors of 0.74 meters for PPE-AM, 0.47 meters for the refined PPE-GA, and 0.92 meters for the smartphone-based SPE.

The genetic algorithm in PPE-GA proved crucial for aligning BIM and panorama data, while the generated 3D spherical world position map offers a foundational dataset for applications such as VR/AR, indoor navigation, and spatial analysis. This research contributes a cost-effective and scalable alternative to existing methods, reducing reliance on extensive pre-mapping or expensive infrastructure. The system's ability to combine BIM data with advanced image processing provides a robust solution for navigation in complex indoor spaces.

VI. FUTURE DIRECTIONS

Building upon the accuracy and cost-effectiveness of PPE-GA and SPE, our future research will focus on the following key areas to enhance the robustness and applicability:

1) **Testing in Diverse Environments:** We will evaluate the system in diverse settings, including multi-floor buildings, and outdoor-indoor transition areas, to validate its generalizability under varying environmental conditions. We will compare the localization accuracy (mean error, standard deviation, and error distribution) across these environments to quantify the system's performance.
2) **Real-time BIM Adaptation:** We will develop algorithms for dynamic BIM updates using real-time visual input, enabling the system to adapt to changes like furniture repositioning or room reconfigurations, ensuring accuracy in evolving indoor spaces.
3) **Novel Applications:** We will explore advanced applications, such as real-time asset tracking, augmented reality-assisted maintenance, and semantic-rich indoor navigation, to demonstrate the broader impact of our framework in digital twin technologies.

REFERENCES

[1] Y. Liu, H. Jiang, G. Nader, Z. Wu, T. Tanasnitikul, and P. Lasang, "Local topology constrained point cloud registration in Building Information Modelling," *IEEE Sensors Journal*, vol. 24, no. 3, pp. 4036-4046, 2024, doi: 10.1109/JSEN.2023.3341218.

[2] H. Stuckey, A. Al-Radaideh, L. Sun, and W. Tang, "A spatial localization and attitude estimation system for unmanned aerial vehicles using a single dynamic vision sensor," *IEEE Sensors Journal*, vol. 22, no. 15, pp. 15497–15507, 2022, doi: 10.1109/JSEN.2022.3187423.

[3] T. L. Baldi, F. Farina, A. Garulli, A. Giannitrapani, and D. Prattichizzo, "Upper body pose estimation using wearable inertial sensors and multiplicative Kalman filter," *IEEE Sensors Journal*, vol. 20, no. 1, pp. 492–500, 2019, doi: 10.1109/JSEN.2019.2940612.

[4] F. Gu, K. Khoshelham, J. Shang, F. Yu, and Z. Wei, "Robust and accurate smartphone-based step counting for indoor localization," *IEEE Sensors Journal*, vol. 17, no. 11, pp. 3453–3460, 2017, doi: 10.1109/JSEN.2017.2685999.

[5] J. Li, B. Zhang, S. Li, S. Zlatanova, Z. Yang, M. Bai, and D. Wang, "MLA-MFL: A smartphone indoor localization method for fusing multi-source sensors under multiple scene conditions," *IEEE Sensors Journal*, vol. 24, no. 16, pp. 26320-26333, 2024, doi: 10.1109/JSEN.2024.3420727.

[6] P. Zheng, Z. Li, S. Zheng, H. Zhang, X. Liang, Y. Li, and X.-d. Zou, "Range-aided LiDAR cooperative localization and mapping," *IEEE Sensors Journal*, vol. 24, no. 9, pp. 14608-14622, 2024, doi: 10.1109/JSEN.2024.3374346.

[7] H.-F. Ng, G. Zhang, Y. Luo, and L.-T. Hsu, "Urban positioning: 3D mapping-aided GNSS using dual-frequency pseudorange measurements from smartphones," *Navigation*, vol. 68, no. 4, pp. 727–749, 2021, doi: 10.1002/navi.448.

[8] W. Wen, G. Zhang, and L.-T. Hsu, "Correcting NLOS by 3D LiDAR and building height to improve GNSS single point positioning," *Navigation*, vol. 66, no. 4, pp. 705–718, 2019, doi: 10.1002/navi.335.

[9] M. Aladem and S. A. Rawashdeh, "A combined vision-based multiple object tracking and visual odometry system," *IEEE Sensors Journal*, vol. 19, no. 23, pp. 11714–11720, 2019, doi: 10.1109/JSEN.2019.2937304.

[10] Q. Zeng, C. Gao, Z. Chen, Y. Jin, and Y. Kan, "Robust mono visual–inertial odometry using sparse optical flow with edge detection," *IEEE Sensors Journal*, vol. 22, no. 6, pp. 5260–5269, 2021, doi: 10.1109/JSEN.2021.3070603.

[11] V. Kachurka, B. Rault, F. I. I. Muñoz, D. Roussel, F. Bonardi, J.-Y. Didier, H. Hadj-Abdelkader, S. Bouchafa, P. Alliez, and M. Robin, "WeCo-SLAM: Wearable cooperative SLAM system for real-time indoor localization under challenging conditions," *IEEE Sensors Journal*, vol. 22, no. 6, pp. 5122–5132, 2021, doi: 10.1109/JSEN.2021.3101121.

[12] H. Saleem, R. Malekian, and H. Munir, "Neural network-based recent research developments in SLAM for autonomous ground vehicles: A review," *IEEE Sensors Journal*, vol. 23, no. 13, pp. 13829–13858, 2023, doi: 10.1109/JSEN.2023.3273913.

[13] J.-Y. Huang, S.-H. Lee, and C.-H. Tsai, "A fast image matching technique for the panoramic-based localization," in *Proc. IEEE/ACIS 15th Int. Conf. Computer and Information Science (ICIS)*, Okayama, Japan, 2016, pp. 1–6, doi: 10.1109/ICIS.2016.7550792.

[14] J. Kim, H. Jang, C. Choi, and Y. M. Kim, "CPO: Change robust panorama to point cloud localization," in *Computer Vision – ECCV 2022*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds., Lecture Notes in Computer Science, vol. 13669. Cham, Switzerland: Springer, 2022, pp. 189–206, doi: 10.1007/978-3-031-20077-9_11.

[15] J. Kim, J. Jeong, and Y. M. Kim, "Fully geometric panoramic localization," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2024, pp. 20827–20837, doi: 10.1109/CVPR52733.2024.01968.

[16] R. Tomasi, F. Sottile, C. Pastrone, M. M. Mozumdar, A. Osello, and L. Lavagno, "Leveraging BIM interoperability for UWB-based WSN planning," *IEEE Sensors Journal*, vol. 15, no. 10, pp. 5988–5996, 2015, doi: 10.1109/JSEN.2015.2451736.

[17] M. J. L. Lee, H. Y. Ho, L.-T. Hsu, and S. L. M. Au, "BIPS: Building information positioning system," in *Proc. Int. Conf. Indoor Positioning and Indoor Navigation (IPIN)*, Lloret de Mar, Spain, Nov.–Dec. 2021, pp. 1–7, doi: 10.1109/IPIN51156.2021.9662575.

[18] Z. Zhao, H. Yu, C. Lyu, W. Yang, and S. Scherer, "Attention-enhanced cross-modal localization between spherical images and point clouds," *IEEE Sensors Journal*, vol. 23, no. 19, pp. 23836-23845, 2023, doi: 10.1109/JSEN.2023.3306377.

[19] Z. Yu, C. Xiao, and G. Zhou, "Multi-objectivization-based localization of underwater sensors using magnetometers," *IEEE Sensors Journal*, vol. 14, no. 4, pp. 1099–1106, 2013, doi: 10.1109/JSEN.2013.2287915.

[20] P. Sridhar, S. K. Thangavel, L. Parameswaran, and V. R. M. Oruganti, "Fire sensor and surveillance camera-based GTCNN for fire detection system," *IEEE Sensors Journal*, vol. 23, no. 7, pp. 7626–7633, 2023, doi: 10.1109/JSEN.2023.3244833.

[21] Blender Foundation, "Blender – a 3D modelling and rendering package," Blender Software, 2021. [Online]. Available: https://www.blender.org. [Accessed: Dec. 9, 2024].

[22] Insta360, "Insta360 X3," *Insta360*. [Online]. Available: https://www.insta360.com/product/insta360-x3. [Accessed: Dec. 10, 2024].

[23] N. Ravi *et al.*, "SAM 2: Segment anything in images and videos," *arXiv preprint* arXiv:2408.00714, 2024. [Online]. Available: https://arxiv.org/abs/2408.00714. [Accessed: Dec. 9, 2024].

[24] L. Yang *et al.*, "Depth Anything V2," *arXiv preprint* arXiv:2406.09414, 2024. [Online]. Available: https://arxiv.org/abs/2406.09414. [Accessed: Dec. 9, 2024].

[25] P.-E. Sarlin, C. Cadena, R. Siegwart, and M. Dymczyk, "From coarse to fine: Robust hierarchical localization at large scale," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 12708–12717, doi: 10.1109/CVPR.2019.01300.

[26] H. Jiang *et al.*, "OmniGlue: Generalizable feature matching with foundation model guidance," *arXiv preprint* arXiv:2405.12979, 2024. [Online]. Available: https://arxiv.org/abs/2405.12979. [Accessed: Dec. 9, 2024].

[27] u-blox, "ZED-F9P module," [Online]. Available: https://www.u-blox.com/en/product/zed-f9p-module. [Accessed: May 21, 2024].

[28] Nooploop, "UWB high-precision positioning: LinkTrack P-A series," [Online]. Available: https://www.nooploop.com/en/linktrack/. [Accessed: May 21, 2024].

Max Lee received a Bachelor of Engineering (Honours) in Aviation Engineering from The Hong Kong Polytechnic University in 2018. He is currently a PhD student in the Department of Aeronautical and Aviation Engineering at The Hong Kong Polytechnic University. He is a recipient of the Hong Kong PhD Fellowship. Max has been named to Forbes' 30 Under 30 Asia list and has received other awards for his work in the field, including the AAE Outstanding Alumni Award. His research interests include seamless positioning, smart cities, and drones. Max is committed to advancing technology and its applications in the low altitude economy, continuing to explore new frontiers in drone technology and positioning systems.

Ju Lin is a final-year undergraduate pursuing a Bachelor of Engineering (Honours) in Aviation Engineering, with a minor in Electronics and Information Engineering at the Hong Kong Polytechnic University. He is exploring his interests in novel computer science algorithms, seamless positioning, GNSS positioning, space exploration, and the low-altitude economy.

Guohao Zhang received his bachelor's degree in mechanical engineering and automation from University of Science and Technology Beijing, China, in 2015. He received his master's degree in Mechanical Engineering and his Ph.D. degree in Aeronautical and Aviation Engineering from The Hong Kong Polytechnic University, Hong Kong, in 2017 and 2022, respectively. He is currently a Research Assistant Professor with the Department of Aeronautical and Aviation Engineering, The Hong Kong Polytechnic University. He won the Best Presentation Award in 2018 and the ION GNSS+ Student Paper Award in 2019 from the Institute of Navigation (ION). His research interests include GNSS urban positioning, collaborative positioning, machine learning-aided GNSS, signal propagation modelling and remote sensing.

Li-Ta Hsu is an associate professor at the Department of Aeronautical and Aviation Engineering of Hong Kong Polytechnic University. He is Limin Endowed Young Scholar in Aerospace Navigation. He received the B.S. and Ph.D. degrees in Aeronautics and Astronautics from National Cheng Kung University, Taiwan, in 2007 and 2013, respectively. He was a Visiting Researcher with the Faculty of Engineering, University College London and Tokyo University of Marine Science and Technology, in 2012 and 2013, respectively. In 2013, he won a Student Paper Award and two Best Presentation Awards from the Institute of Navigation (ION). He was selected as a Japan Society for the Promotion of Sciences Postdoctoral Fellow with the Institute of Industrial Science, The University of Tokyo and worked from 2014 to 2016. He is an Associate Fellow in the Royal Institute of Navigation. Dr. Hsu currently is a member of ION and IEEE and serves as a member of the editorial board and reviewer in professional journals related to GNSS.