ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XI-4-2026
XXV ISPRS Congress 2026 “From Imagery to Understanding”, Commission IV, 4–11 July 2026, Toronto, Canada

Situated Augmented Reality for Urban Planning: A Privacy-Aware On-Device Localization Pipeline

Muhammad Alfakhori*, Sven Schneider, Volker Coors

Centre for Geodesy and Geoinformatics, Stuttgart Technical University of Applied Sciences (HFT Stuttgart), Stuttgart, Germany
(muhammad.alfakhori, sven.schneider, volker.coors)@hft-stuttgart.de

Keywords: Augmented Reality, Urban Planning, On-Device Localization, Point Cloud Registration, Mobile LiDAR, Public Participation.

Abstract

Accurate spatial alignment is a key requirement for situated Augmented Reality (AR) in urban planning, where citizens and planners can visualize proposed designs in real outdoor environments. However, existing AR localization approaches often rely on smartphone GNSS, vendor-specific cloud anchors, or cloud-based visual positioning, which introduce accuracy limitations, privacy concerns, or dependencies that restrict their use in participatory planning workflows. This paper presents a privacy-aware on-device localization pipeline for outdoor urban planning scenarios. The approach aligns LiDAR scans captured on smartphones with prescanned reference point cloud tiles to enable stable and accurate placement of urban planning models. Approximate GNSS is used only to retrieve a relevant reference tile, while all preprocessing and registration steps are performed locally on the device. The pipeline combines voxel downsampling, local geometric descriptors, and global registration to estimate alignment without relying on GNSS for pose estimation or on cloud-based visual localization services. A mobile demonstrator was developed to support situated AR in urban planning scenarios, allowing users to explore design proposals directly in context. Initial validation under controlled conditions showed that the system can recover translations and rotations with errors on the order of a few centimeters, while processing times remained suitable for mobile use. The approach was also deployed in an urban planning case study and enabled stable outdoor visualization of planning elements on-site.

1. Introduction

Citizen participation plays an important role in planning processes for the design of public spaces. To this end, there are established procedures in which urban planning drafts are presented to the public at an early stage and discussed in participation forums. This enables people with the opportunity to express opinions on plan alternatives and propose modifications. Although all citizens groups are called upon to participate in these decision-making processes, participatory urban planning is only partially successful in this way (Bäumer et al., 2025).

Meetings at which plans are presented and discussed must take place at specific times that may be inconvenient for some participants, such as families. People who do not regularly deal with plans in the form of two-dimensional, technically dimensioned drawings sometimes find it difficult to reconstruct the three-dimensional objects from them. Another problem is that fewer and fewer young people feel addressed by these methods of participation. These participation forums are also time-consuming to organize, which is why they have so far mainly been used for central development projects, particularly in larger cities. Temporary road layouts due to roadworks, for example, have so far generally been planned without public participation, although sometimes the residents of an entire district are affected by the change over a longer period of time (Othengrafen et al., 2022).

Modern 3D visualization techniques can be used to create realistic images and videos of virtual objects and scenes that support the imagination. However, the perspective representation used makes it difficult to estimate sizes and distances. There are approaches for integrating Virtual Reality (VR) and Augmented Reality (AR) technologies into participatory urban planning processes (Bäumer et al., 2025, Dwarkadas et al., 2023) and thus creating a high degree of realism in the visualization of urban planning designs. The addition of multi-user and Mixed Reality (MR) technologies can further improve the realistic representation and also achieve greater independence from time and place in citizen participation processes.

However, a complex digital system can only establish itself sustainably if it offers a low-threshold service based on widespread media (especially smartphones) and intuitive (Alghisi and Biagi, 2023), multimodal interaction tools for a large number of citizens of all ages and genders as well as diverse population groups. Another prerequisite for broad acceptance is the robustness and real-time capability of the MR urban planning system, which should be continuously developed in a participatory and user-centered manner, taking into account ethical, legal and social implications.

Situated AR has recently emerged as a powerful tool for urban planning and design, enabling stakeholders to explore proposals directly in the physical environment where changes will occur (Bäumer et al., 2025). Modern AR research demonstrates that spatially situated visualization can significantly improve citizens' understanding of design alternatives and support more informed decision making in planning workflows (Wang and Lin, 2023). However, deploying AR outdoors remains challenging because precise and stable localization is required, especially in dense urban environments where GNSS signals and cloud-based visual positioning systems suffer from major accuracy and reliability limitations (Alghisi and Biagi, 2023). Additionally, recent studies highlight the privacy implications of cloud-based localization, where user scan data must be uploaded to external servers, limiting use in planning processes that require trust and compliance with data

---

* Corresponding author

https://doi.org/10.5194/isprs-annals-XI-4-2026-205-2026 | © Author(s) 2026. CC BY 4.0 License.

governance frameworks (Moon et al., 2024). These challenges create a critical barrier preventing the wider adoption of AR in real-world urban planning and public participation contexts.

The aim of this work is therefore to enable Situated AR for outdoor urban planning by providing a precise, privacy-aware localization pipeline in which preprocessing and registration are executed locally on mobile devices. Instead of relying on GNSS for precise pose estimation, cloud anchors, or cloud-based visual positioning, the proposed approach aligns LiDAR scans captured on smartphones with pre-scanned reference point cloud tiles. Approximate GNSS is used only to retrieve a relevant reference tile, while the registration itself is performed on-device. This enables accurate six degrees of freedom (6-DOF) localization even in dense urban areas, allowing planning models to be visualized in correct spatial context without transmitting raw user scan data to external services. The resulting localization capability forms the technical foundation for robust and trustworthy Situated AR applications in public participation and urban design workflows.

2. State of the Art

In the following, the state of the art will be outlined for three subject areas that are of particular importance for this work:

2.1 Participatory urban planning and digitalization

Public participation plays an important role in the planning and design of public spaces. As the population in urban areas increases, so does the importance of accessible participatory processes that enable different citizen groups to contribute their perspectives. However, traditional participation formats such as information evenings or planning exhibitions often take place at fixed times and locations, which disproportionately exclude families, working people, and other groups with limited availability (Othengrafen et al., 2022). In addition, many participants struggle to interpret technically complex two-dimensional plans, which makes it difficult to understand spatial impacts or envision future urban environments (Crooks and See, 2022).

Digital participation formats have therefore gained increasing relevance in recent years, supported by advances in mobile technologies and the growing availability of urban data. These tools offer more flexible forms of engagement and can help reduce participation barriers by enabling asynchronous access and interactive exploration of planning content (Othengrafen et al., 2022). Modern visualization techniques, including 3D models and street-level imagery, further support citizens in understanding the spatial context of planning proposals (Crooks and See, 2022).

In parallel, immersive technologies such as VR and AR have been integrated into participatory urban planning processes, providing more realistic representations of proposed designs and enabling users to experience possible interventions at eye level (Bäumer et al., 2025). Recent work shows that AR-based participation formats can enhance citizens’ spatial understanding, increase engagement, and support more informed decision making through interactive in-situ visualization of planning alternatives (Wang and Lin, 2023). Multi-user and MR approaches also offer the potential to support collaborative exploration and provide greater flexibility regarding time and location.

However, despite these advantages, digital participation systems must be accessible, easy to use, and robust in real-world conditions to achieve broad acceptance among diverse user groups. They also require sufficient technical performance and reliable localization to ensure accurate spatial alignment when immersive representations are used outdoors. These requirements highlight the need for Situated AR solutions that allow citizens to explore urban planning proposals directly within their physical context.

Digital tools, especially AR and VR, offer a very good opportunity here to make urban planning processes more transparent and participatory (Sabitha et al., 2024). In addition to discussing planning variants, citizens also have the opportunity to model their own designs and contribute their own ideas. This potential has also been explored in our previous work, where we used AR to support public participation through occlusion screening based on 3D city models as a reference database for mobile applications (Alfakhori et al., 2023).

2.2 Mobile Augmented Reality and Tracking

Mobile, location-based AR enriches the physical environment with digital information by combining camera images, inertial measurements, and location data. Modern smartphones support AR applications through integrated sensors such as GNSS receivers, magnetometers, accelerometers and gyroscopes. Sensor-based tracking is widely used outdoors, but its accuracy is highly dependent on environmental conditions and device-specific limitations. GNSS positioning on smartphones often suffers from multipath effects, signal blockage and limited satellite visibility in dense urban areas, resulting in poor spatial alignment for AR use cases (Zangenehnejad and Gao, 2021, Alghisi and Biagi, 2023). Although the great mobility of this solution makes it suitable for outdoor AR, these accuracy limitations restrict its applicability in urban planning scenarios that require stable and precise visualization.

Simultaneous Localization and Mapping (SLAM) algorithms, including PTAM (Klein and Murray, 2009) and KinectFusion (Newcombe et al., 2011), enable more accurate motion tracking by continuously estimating the device pose relative to a locally built map. They are integrated in major AR frameworks (e.g., ARKit, ARCore) and can produce dense or sparse point clouds of the environment. While SLAM is effective for relative tracking, it lacks global spatial reference and is sensitive to environmental changes such as moving objects, vegetation, or lighting conditions, which reduces its reliability outdoors.

Commercial platforms offer “cloud anchors” or similar services that share AR reference maps across devices. These solutions rely on centralized 3D maps or image descriptors stored on servers provided by platform vendors. However, cloud-based visual localization introduces privacy concerns, as environment scans or imagery must be uploaded to external servers (Moon et al., 2024). This is problematic in public-sector planning workflows, where data sovereignty, long-term stability and GDPR-compliance are essential. Moreover, their effectiveness depends on the availability and spatial coverage of proprietary map data, which may be incomplete or outdated in many urban environments.

2.3 Geometric processing of point clouds

The automated analysis of 3D point clouds relies on a range of geometric, feature-based, topology-aware, and learning-based methods depending on the target application (Xiong et al., 2017, Boulch et al., 2017). In urban and infrastructure contexts, point clouds are used for tasks such as segmentation, object reconstruction, scene interpretation, and registration. A major challenge in the processing, analysis, and visualization of 3D point clouds is the handling of extremely high data volumes, which result from repeated acquisitions and increasingly dense sampling. Efficient processing and visualization therefore often require out-of-core methods based on spatial indexing structures and multi-level caching strategies (Discher et al., 2019). This enables large point clouds to be provided efficiently as a service for analysis and visualization, regardless of their actual size (Discher et al., 2019, Martinez-Rubi, 2016).

For registration tasks, local geometric descriptors such as Fast Point Feature Histograms (FPFH) are widely used because they offer a compact description of neighborhood structure (Rusu et al., 2009); however, their performance degrades in environments with limited geometric diversity or repeating planar structures, which is often the case in outdoor street environments. This increases the importance of preprocessing steps such as downsampling, normal estimation and outlier removal to ensure robust matching.

Point cloud registration methods can be broadly divided into local refinement approaches and global registration approaches. Local methods such as Iterative Closest Point (ICP) and Point-to-Plane ICP achieve high accuracy when a good initial alignment is available but are sensitive to large pose differences (Besl and McKay, 1992, Chen and Medioni, 1991). Global methods, including Fast Global Registration (FGR), estimate correspondences more robustly under larger initial misalignments and are therefore well suited to coarse localization scenarios (Zhou et al., 2016). In parallel, more recent learning-based and vision-based localization approaches have been proposed, but these often require larger training resources, image feature databases, or cloud-supported infrastructures. For the privacy-aware mobile AR scenario addressed in this paper, an FPFH- and FGR-based pipeline offers a practical compromise between computational efficiency, robustness, and local execution on the device.

3. Methodology

3.1 System Overview

The objective of the proposed method is to provide precise localization and orientation estimation for outdoor AR applications by matching locally acquired point clouds with pre-scanned reference point clouds. This system uses 3D point cloud registration as its primary localization mechanism, in contrast to traditional techniques that depend on GNSS, optical markers, or image-based SLAM. The reference point clouds are pre-acquired via high-precision terrestrial laser scanning (TLS). In real-time operation, a point cloud is captured using a smartphone’s LiDAR sensor (e.g., iPhone with ToF or LiDAR capabilities) and employed to determine the user’s current location through registration.

The novelty of this approach lies not in introducing a new registration algorithm but in integrating a privacy-aware localization pipeline for outdoor AR in which approximate GNSS is used only for reference tile retrieval, while preprocessing and registration are performed locally on the device. This design directly addresses limitations of current localization methods used in mobile AR, particularly their susceptibility to GNSS degradation for precise pose estimation and privacy issues related to cloud-based visual localization. Figure 1 illustrates a schematic depiction of the system components and their interrelations.

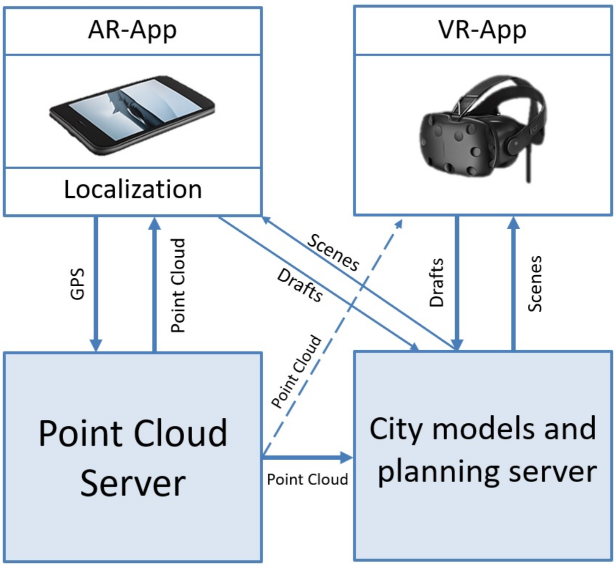


Figure 1. Overview of the on-device localization workflow combining LiDAR capture, preprocessing, registration, and AR visualization.

3.2 Data Flow and Communication

The system uses a hybrid data flow model, combining GNSS-based server access with fully local processing. Based on the user’s GNSS location, a reference point cloud tile is requested via an API and downloaded from a central server in PCD format. These tiles are optimized for mobile use through tiling and lightweight storage.

The user captures the environment using a smartphone’s LiDAR sensor. After one or more scans, the user manually initiates the registration process. All steps—including preprocessing and Fast Global Registration (FGR) (Zhou et al., 2016)—are performed locally to preserve privacy. This avoids uploading environmental scans to external servers, which is a known privacy challenge in cloud-based AR localization workflows (Moon et al., 2024). To reduce data volume and download time, the reference point clouds were downsampled prior to transmission. Tests showed that the original high-density TLS scans exceeded the resolution needed for robust feature matching.

The resulting transformation matrix is transferred to the Unity AR environment via shared memory, enabling low-latency alignment of virtual content with the real environment. This on-device workflow eliminates the need to upload raw scans to external servers and supports fast, privacy-preserving computation suitable for outdoor AR scenarios.

3.3 Platform Integration and Tools

The AR interface, developed in Unity, functions as the front-end for content placement and user engagement. Point cloud registration was implemented in Python using the Open3D library and was integrated into the iOS app via embedded Python support in Swift. This allowed for direct control of the registration logic from the native app environment without the need for reimplementation in C++.

On iOS, Swift manages LiDAR-based point cloud acquisition and facilitates the registration process. Communication between Swift and Unity is facilitated through shared memory, allowing for real-time transmission of the transformation matrix. The solution is developed for smartphones equipped with LiDAR or comparable depth sensors and can additionally use onboard GNSS and inertial measurement unit (IMU) data for coarse positioning and consistency checks, while the precise pose estimation is derived from point cloud registration.

3.4 Point Cloud-Based Localization

The general workflow is illustrated in Figure 2, which shows the loading of the source and target point clouds into memory. These clouds come from local scans captured on the smartphone using LiDAR or are downloaded from a server in small tiles based on GNSS coordinates. Before registration begins, a format check ensures that both point clouds contain compatible data fields (e.g., x, y, z and, if available, normals and color). If normal vectors are missing, they are computed during preprocessing. Color data, however, cannot be generated later and is only relevant for specific methods such as ColorICP (Korn et al., 2014), which was evaluated but not used.

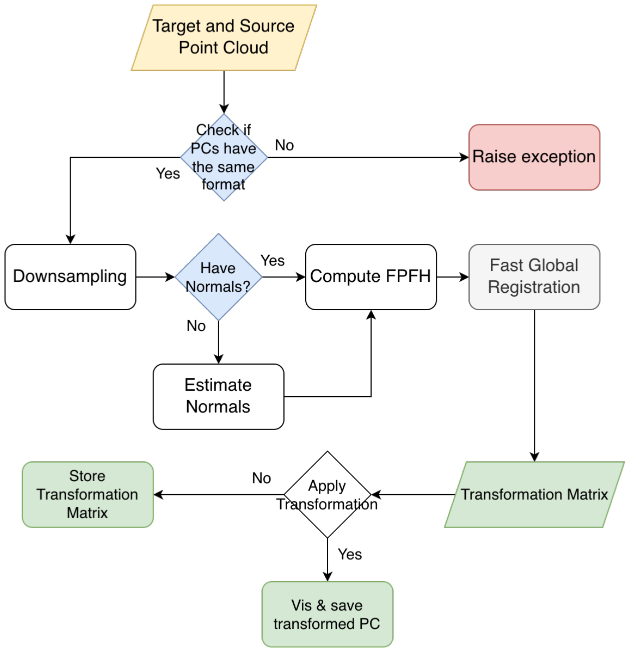


Figure 2. Processing pipeline from point cloud loading to pose estimation using voxel downsampling, FPFH, and FGR.

During preprocessing, the point clouds are voxelized to reduce data size and computational load. Voxel size determines the spatial resolution of the downsampled point cloud, directly affecting both the level of geometric detail and the computational performance of the registration pipeline. Normals are calculated if missing, and Fast Point Feature Histograms (FPFH) are generated to describe local geometric properties (Rusu et al., 2009). While FPFH descriptors are efficient to compute, they are sensitive to repetitive planar structures, which are common in outdoor street environments. Careful selection of voxel size and normal estimation parameters is therefore essential to ensure robust matching.

Registration is performed using FGR, which matches the point clouds based on their FPFH descriptors. FGR is a global registration method that directly estimates the rigid transformation between two point clouds by minimizing a robust correspondence-based objective function. Unlike local methods that require a good initial alignment, FGR can handle large pose differences and partial overlaps efficiently. It is designed to be fast and scalable, making it ideal for real-time applications on mobile devices. FGR is particularly well-suited for partially overlapping 3D scans and offers robust results without requiring an initial pose estimate. It matches the accuracy of local refinement methods such as Iterative Closest Point (ICP) (Besl and McKay, 1992) and Point-to-Plane ICP (Chen and Medioni, 1991), while being significantly faster.

The registration process outputs a 4×4 transformation matrix in homogeneous coordinates, representing both rotation and translation between the captured and reference scans. This matrix is passed to the Unity-based AR application via shared memory, enabling accurate anchoring of virtual planning content within the user’s physical environment.

The method is implemented using the Open3D library (Zhou et al., 2018) and integrated into a Swift-based mobile application that handles LiDAR scanning, preprocessing, registration, and communication with the Unity AR interface as described in Figure 3.

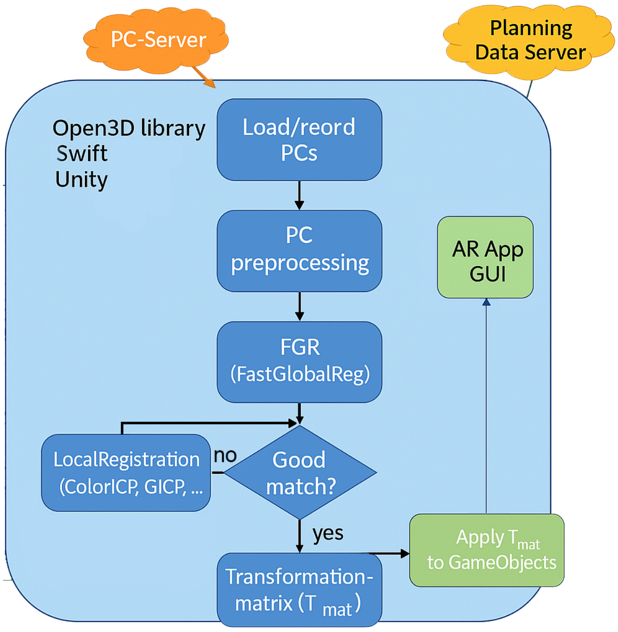


Figure 3. Integration of Swift (LiDAR and registration) and Unity (AR rendering) via shared memory for low-latency visualization.

4. Results and Evaluation

4.1 System Integration and Demonstrator Architecture

The developed AR localization system was implemented as a mobile localization pipeline and integrated into a demonstrator for field testing. After retrieval of a reference point cloud tile based on approximate GNSS location, all preprocessing and registration steps were executed locally on the device.

The use of a shared-memory interface between Swift (LiDAR acquisition and registration) and Unity (AR rendering) provided a low-latency communication channel that allowed virtual planning elements to be updated with minimal delay. This design ensured stable alignment of virtual content even during user motion—an essential requirement for Situated AR in urban planning.

All registration steps—including voxel downsampling, normal estimation, FPFH computation, and FGR—ran locally on the device, ensuring that no raw environment data or point clouds were transmitted to external servers. This fulfills privacy and data sovereignty requirements relevant to public-sector planning workflows.

4.2 Point Cloud Matching Accuracy

To provide an initial technical validation of the point cloud-based registration method, two controlled test cases were conducted. In the first case, a subset of a larger point cloud was extracted and translated by a known vector of $[x = 3, y = 1, z = 4]$ meters. The goal was to assess whether the registration algorithm could recover this transformation using FGR.

The resulting transformation matrix obtained through FGR was:

```
[ 9.9999 0.0047  0.0012 3.0003]
[-0.0047 9.9998 -0.0043 0.9522]
[-0.0012 0.0043  9.9999 3.9888]
[ 0.0000 0.0000  0.0000 1.0000]
```

The Euler angles for X,Y and Z axes in degrees are [-0.27115942, 0.06725352, 0.24689715] respectively. And the translation vector: [3.00033444, 0.95221449, 3.98883637] units in meters.

This result closely approximated the expected shift, with a translation vector of approximately [3.000, 0.952, 3.989] m. The difference from the ground truth was within a few centimeters, confirming the high accuracy of the method even in the absence of local refinement.

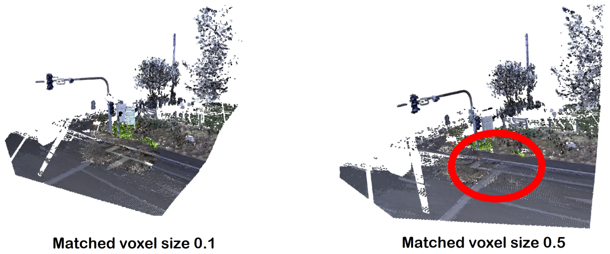
In a second test case, the same subset was modified by applying both a translation and a 10° rotation around the X and Z axes. The global registration still achieved a strong alignment:

```
[ 0.9833  0.1801 0.0241 2.9601]
[-0.1816  0.9703 0.1600 0.9875]
[ 0.0055 -0.1617 0.9868 3.9558]
[ 0.0000  0.0000 0.0000 1.0000]
```

The Euler angles for X,Y and Z axes in degrees are [-10.38017288, 1.3782177, -9.2093472] respectively. The translation vector is [-2.96012747, -0.98748311, -3.95579524] units in meters.

The recovered translation vector [-2.960, -0.987, -3.956] m again showed strong consistency with the ground truth of [-3, -1, -4] m. The successful registration in this case demonstrated the robustness of FGR even when the initial misalignment includes rotation.

Additional experiments were conducted to evaluate the effect of voxel size on registration accuracy. As shown in Figure 4, the choice of voxel resolution during downsampling has a direct impact on the registration outcome. Coarser voxel sizes lead to faster processing but may compromise accuracy, while finer voxel sizes preserve more detail at the cost of increased computation. The selected resolution was therefore balanced to ensure robust feature extraction without significantly affecting real-time performance.



Matched voxel size 0.1 Matched voxel size 0.5

Figure 4. Effect of voxel resolution on registration performance and geometric detail.

4.3 Runtime Performance and Stability

Runtime performance was evaluated on two datasets representative of dense and semi-structured urban scenes. The analysis measured the processing time for each major step of the pipeline: point cloud loading, voxel downsampling, normal estimation, FPFH computation, and FGR registration.

Table 1 summarizes the properties of the two test datasets used to benchmark runtime performance. These included variations in file format, point count, and availability of normal vectors. Table 2 shows the timing breakdown of key processing steps across these datasets.

Table 1. Overview of point cloud datasets used for runtime evaluation, including format, size, and availability of normals.

Dataset	File Name	Format	Points / Normals
Dataset 1-1	Cound_bin_0	PCD	198,835 / Yes
Dataset 1-2	Cound_bin_1	PCD	137,833 / Yes
Dataset 2-1	1004N	PCD	269,856 / Yes
Dataset 2-2	part_reduced	XYZ	494,748 / No

The runtime analysis showed that the one-time library initialization required approximately 0.55–0.59 seconds, after which point cloud loading took between 0.06 and 0.30 seconds, depending on the file format and point density. Preprocessing steps were comparatively lightweight: voxel downsampling required 0.01–0.04 seconds, while normal estimation and FPFH descriptor computation ranged from 0.0015 to 0.40 seconds, with higher values observed for denser datasets. The subsequent FGR registration step completed in 0.57–0.71 seconds, representing the main computational component of the localization pipeline.

These results show that the entire registration process can be completed in approximately 1.5 seconds on a modern smartphone, making it suitable for user-triggered localization events in outdoor AR. Although this approach does not yet support continuous localization at video framerates, the latency is sufficiently low for Situated AR scenarios where the user aligns the virtual model once before exploring the scene.

4.4 Application in Urban Planning Scenarios

The final demonstrator was deployed in an outdoor planning scenario to evaluate usability under real environmental condi-

Table 2. Runtime of key processing steps across two point cloud datasets.

Phase	Set1 - Time (s)	Set2 - Time (s)
Library initialization	0.5927	0.5538
Point Clouds loading	0.0616	0.2982
Downsampling	0.0056	0.0133
Normals estimation	0.0003	0.0373
FPFH computing	0.0015	0.4047
FGR	0.5756	0.7115
Total	**1.2325**	**2.3221**

tions. Users scanned a streetscape using a smartphone LiDAR sensor, triggered the on-device registration, and viewed proposed design elements such as street furniture, façade extensions, and temporary installations directly in the physical environment.

In the field demonstration, the system produced visually stable overlay results under variable lighting, pedestrian activity, and partial occlusion. The situated visualization allowed users to move freely around virtual planning content while maintaining accurate alignment. Participants reported that experiencing planning alternatives in context improved their understanding of spatial impacts and scale, consistent with previous findings on AR-supported urban participation (Bäumer et al., 2025).

This indicates that the proposed localization approach is promising not only in controlled tests but also in real-world environments relevant for public engagement.

Figure 5 shows the final demonstrator in action, where the live AR interface overlays planning models onto the real cityscape. This setup offers a compelling method for participatory urban design and public consultation.

Figure 5. Situated AR demonstrator showing aligned planning models in an outdoor environment.

5. Discussion

Situated AR offers significant potential for improving public participation in urban planning by enabling citizens to experience proposed interventions directly within their physical context. However, its practical deployment outdoors depends on accurate and trustworthy localization, especially in dense urban environments where GNSS and cloud-based visual positioning systems face well-documented limitations (Zangenehnejad and Gao, 2021, Alghisi and Biagi, 2023). The presented approach addresses this gap by providing a privacy-aware localization workflow based on point cloud registration, in which approximate GNSS is used only for reference tile retrieval while preprocessing and registration are performed locally on the device. This reduces dependencies on cloud-based localization services and mitigates privacy concerns associated with external processing of environmental scans (Moon et al., 2024).

Instead of introducing a new registration algorithm, the contribution of this work lies in the integration of a robust, privacy-preserving pipeline tailored to Situated AR for outdoor planning scenarios. The evaluation demonstrates that this pipeline can achieve centimeter-level accuracy under both translational and rotational offsets while maintaining runtimes compatible with user-triggered alignment events on mobile devices. These results indicate that point cloud registration, when designed with appropriate preprocessing and feature extraction strategies, is sufficiently efficient for real-world AR localization tasks and supports the spatial precision required for in-situ visualization of planning models.

From the perspective of urban participation, the ability to anchor planning content reliably in outdoor environments is essential for fostering trust, reducing abstraction barriers, and improving citizens’ understanding of spatial impacts (Wang and Lin, 2023, Bäumer et al., 2025). The demonstrator’s field deployment showed that users could explore virtual planning alternatives in context without noticeable drift, supporting earlier findings that AR-based participation formats can complement traditional planning communication.

Despite these benefits, several limitations remain. The approach depends on the availability and quality of reference point clouds, which are typically obtained through terrestrial laser scanning and require initial acquisition effort. Environmental changes such as vegetation growth, temporary street furniture, or construction activities can reduce registration robustness if they substantially alter the local geometry. Furthermore, while the current implementation supports responsive user-triggered localization, continuous real-time tracking based purely on registration remains computationally demanding on mobile hardware. Future work may therefore focus on combining on-device registration with lightweight visual-inertial updates, integrating semantic filtering to improve robustness under environmental change, and evaluating user experience more systematically in participatory planning contexts.

Overall, the results demonstrate that on-device point cloud registration offers a viable and privacy-conscious foundation for Situated AR applications in urban planning, enabling citizens and planners to interact with design proposals directly in the real world without relying on GNSS accuracy or cloud infrastructure.

6. Conclusion

This paper presented a privacy-aware on-device localization pipeline that enables Situated AR for outdoor urban planning scenarios by aligning smartphone LiDAR scans with pre-scanned reference point clouds. The contribution of this work lies in the integration of a robust, privacy-preserving registration pipeline that does not rely on GNSS or cloud-based visual positioning systems for precise pose estimation—two common sources of inaccuracy and privacy concern in mobile AR applications.

The evaluation demonstrated that the approach achieves centimeter-level accuracy under both translational and rotational offsets while maintaining runtimes suitable for user-triggered alignment on mobile devices. Field deployment in an urban planning scenario further showed that stable, context-accurate visualization of planning content is feasible when point cloud preprocessing and registration are performed locally on the device after retrieval of a relevant reference tile, supporting more intuitive and trustworthy communication of spatial proposals.

While the method depends on the availability and geometric stability of reference point clouds, and does not yet support continuous real-time localization, it provides a practical foundation for integrating Situated AR into public participation processes. Future work will focus on combining global registration with lightweight visual-inertial tracking for continuous updates, improving robustness under environmental changes, and expanding user studies to assess the impact of Situated AR on citizen engagement and decision-making.

Overall, the results indicate that point cloud-based on-device localization is a viable and scalable solution for enabling privacy-conscious, accurate Situated AR in real outdoor urban planning contexts.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT (GPT-4-turbo, version as of April 2025) provided by OpenAI to assist with rephrasing and grammar correction. After using this tool, the author(s) thoroughly reviewed and edited the content to ensure accuracy and take full responsibility for the final version of the publication.

ACKNOWLEDGEMENTS

This work was funded by the German Federal Ministry of Research, Technology and Space (BMFTR) under the grant numbers 16SV8744 (INSPIRER) and 13FH9I05IA (HIRE). The work on INSPIRER was carried out as part of a joint research project. We would like to express our sincere thanks to all project partners for their valuable collaboration and input throughout the project. The responsibility for the content of this publication lies with the authors.

The authors would also like to thank the Department of Surveying at HFT Stuttgart and Mr. Jörg Hepperle for their support with laser scanning. The authors would also like to thank the anonymous reviewers who helped to improve the quality of the paper.

References

Alfakhori, M., Sardi Barzallo, J. S., Coors, V., 2023. Occlusion Handling for Mobile AR Applications in Indoor and Outdoor Scenarios. *Sensors*, 23(9). https://www.mdpi.com/1424-8220/23/9/4245.

Alghisi, M., Biagi, L., 2023. Positioning with GNSS and 5G: Analysis of Geometric Accuracy in Urban Scenarios. *Sensors*, 23(4). https://www.mdpi.com/1424-8220/23/4/2181.

Bäumer, T., Huber, S., Simon-Philipp, C., Coors, V., Alfakhori, M., 2025. Virtual Reality und Augmented Reality als digitale Partizipationsinstrumente. *Transforming Cities*, 10(Sonderausgabe). http://dx.doi.org/10.24053/TC-2025-0032.

Besl, P. J., McKay, H. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 239–256.

Boulch, A., Saux, B. L., Audebert, N., 2017. Unstructured Point Cloud Semantic Labeling Using Deep Segmentation Networks. I. Pratikakis, F. Dupont, M. Ovsjanikov (eds), *Eurographics Workshop on 3D Object Retrieval*, The Eurographics Association.

Chen, Y., Medioni, G., 1991. Object modeling by registration of multiple range images. *Proceedings. 1991 IEEE International Conference on Robotics and Automation*, 3, 2724–2729.

Crooks, A., See, L., 2022. Leveraging Street Level Imagery for Urban Planning. *Environment and Planning B: Urban Analytics and City Science*, 49(3), 773-776.

Discher, S., Richter, R., Trapp, M., Döllner, J., 2019. Service-oriented processing and analysis of massive point clouds in geoinformation management. *Advances in 3D Geoinformation*, Springer, Cham, 43–61.

Dwarkadas, A. L., Krishna Challa, R., Talasila, V., K G, S., 2023. Augmented/virtual reality: Technological advancement with use cases. *2023 Global Conference on Information Technologies and Communications (GCITC)*, 1–7.

Klein, G., Murray, D., 2009. Parallel tracking and mapping on a camera phone. *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, IEEE, 83–86.

Korn, M., Holzkothen, M., Pauli, J., 2014. Color supported generalized-icp. *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, 3, 592–599.

Martinez-Rubi, O., 2016. Using modular 3D digital earth applications based on point clouds for the study of complex sites. *International Journal of Digital Earth*, 9(12), 1135–1152.

Moon, H., Lee, C., Hong, J. H., 2024. Efficient privacy-preserving visual localization using 3d ray clouds. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 9773–9783.

Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A., 2011. Kinectfusion: Real-time dense surface mapping and tracking. *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, 127–136.

Othengrafen, F., Reinecke, E., Sievers, L., 2022. Digitale Beteiligungsformate in der Stadtentwicklung. Aktuelle Anwendungen und Einsatzmöglichkeiten der E- und M-Partizipation. *RaumPlanung*, 217(3/4), 62–69.

Rusu, R. B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (fpfh) for 3d registration. *IEEE International Conference on Robotics and Automation*, 3212–3217.

Sabitha, R., S, G., S, Y., M, G. D., Babuji, R., Murugan, S., 2024. Augmented reality for public engagement in sustainable city planning: Cloud and machine learning integration. *2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE)*, 1–6.

Wang, Y., Lin, Y.-S., 2023. Public participation in urban design with augmented reality technology based on indicator evaluation. *Frontiers in Virtual Reality*, 4.

Xiong, X., Xia, J., Wu, B., 2017. Topologically Aware Building Rooftop Reconstruction From Airborne Laser Scanning Point Clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 55(10), 5820–5832.

Zangenehnejad, F., Gao, Y., 2021. GNSS smartphones positioning: advances, challenges, opportunities, and future perspectives. *Satellite Navigation*, 2(1), 24. https://doi.org/10.1186/s43020-021-00054-y.

Zhou, Q.-Y., Park, J., Koltun, V., 2016. Fast global registration. B. Leibe, J. Matas, N. Sebe, M. Welling (eds), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, 766–782.

Zhou, Q.-Y., Park, J., Koltun, V., 2018. Open3D: A Modern Library for 3D Data Processing. *arXiv preprint arXiv:1801.09847*.