Infrastructure-Free Localization System for Augmented Reality Registration in Indoor Environments: A First Accuracy Assessment

Leonardo Messi
Department of Construction, Civil Engineering and Architecture (DICEA)
Università Politecnica delle Marche (UNIVPM)
Ancona, Italy
l.messi@staff.univpm.it

Francesco Spegni
Department of Construction, Civil Engineering and Architecture (DICEA)
Università Politecnica delle Marche (UNIVPM)
Ancona, Italy
f.spegni@staff.univpm.it

Massimo Vaccarini
Department of Construction, Civil Engineering and Architecture (DICEA)
Università Politecnica delle Marche (UNIVPM)
Ancona, Italy
m.vaccarini@staff.univpm.it

Alessandra Corneli
Department of Construction, Civil Engineering and Architecture (DICEA)
Università Politecnica delle Marche (UNIVPM)
Ancona, Italy
a.corneli@staff.univpm.it

Leoanrdo Binni
Department of Construction, Civil Engineering and Architecture (DICEA)
Università Politecnica delle Marche (UNIVPM)
Ancona, Italy
l.binni@pm.univpm.it

***Abstract*—Navigation systems combined with Augmented Reality (AR) constitute an effective solution for helping user accessing unfamiliar environments, like indoor public facilities. Implementing AR navigation systems requires determining the 6-Degrees-of-Freedom (6-DoF) localization of the user. In indoor environments, lack of Global Navigation Satellite System (GNSS) signals makes localization more challenging compared to outdoor environments. A variety of positioning systems have emerged for indoor localization which are based on several system strategies, location methods, and technologies. High-accuracy, low-cost, easy to use, and no need for any technical expertise are key features to ensure large-scale application of AR navigation systems. This study provides an answer to these key requirements by proposing a markerless infrastructure-free localization system for AR registration in indoor environments based on the comparison between a query image and a 3D mapping assumed as a reference. Since most people in public facilities are equipped with tablets or smartphone devices, they should be provided with further functionalities that will help them in day-to-day life and work. To this purpose, the proposed localization system was implemented as a web-service and tested in a university campus to assess applicability for AR registration purposes. Experiment results showed very promising localization accuracy and computational efficiency satisfying the "fine-precision" accuracy threshold (i.e., 1°/0.100 m) for near real-time AR applications.**

Keywords—Augmented Reality, 6-DoF Localization, Registration, Image Comparison, Navigation System.

I. Introduction

Mobility can be defined as the ability of people to move, live and interact with the space. In this context, indoor mobility, in terms of indoor localization and wayfinding, is a relevant topic due to the challenges it presents, in comparison with outdoor mobility, where Global Navigation Satellite Systems (GNSS) are hardly exploited. Knowing how to move in an indoor environment can be crucial for people with disabilities, and in particular for blind users, but it can provide several advantages also to any person who is moving in an unfamiliar place [1]. The first problem to address for the implementation of Augmented Reality (AR) navigation systems is retrieving the 6-Degrees-of-Freedom (6-DoF) localization of the user or, more precisely, the device pose [2]. In fact, this step is pivotal for a correct AR registration [3].

Several methodologies for indoor AR registration have been developed and tested so far. Past studies have exhaustively tested marker-based approaches using visible [4] or invisible markers [5], [6]. Even though artificial markers are advantageous in terms of robustness in detection, they should be installed all over the facility before on-site activities occur. In addition, visual markers can trigger aesthetic issues because of their distinctive appearance. Image-based localization is classically tackled by estimating the camera pose from correspondences established between sparse local features [7] and a 3D Structure-from-Motion (SfM) [8] map of the scene [9]. This approach, although may lose accuracy whenever there is lack of context or repetitive elements, has shown great potential to develop indoor AR registration apps with applicability for non-expert users [10]. With the increasing ubiquity of smartphones and other mobile devices equipped with cameras and internet access, users are enabled to shot photos and use them to query their positions to image-based localization systems. Being able to locate users indoor without the need for any particular infrastructure is certainly an attractive prospect, given the obvious advantages regarding the costs of installation and maintenance and the scalability of the whole system [1]. In addition, vision-aided positioning boasts the fact that noise in images is not cumulating over time [11]. Existing studies propose different image-to-BIM solutions to register images to building information models (BIM) indoor. Among them, the ones that apply edge [12], perspective [13], and object [14] detection are strictly dependent on the presence of geometric features or recognizable objects within the image scene. In addition, since these solutions use BIM models as a reference, registration is affected by models accuracy. To cover these gaps, this study proposes a markerless infrastructure-free localization system based on the comparison between a query image and a SfM-based 3D mapping assumed as a reference. In doing this, the current study further develops the approach presented in [15] for inside-out applications. In addition, the proposed localization system was implemented as a web-service and tested in a university campus to assess applicability for AR registration purposes. Experiment results showed a very promising localization accuracy and efficiency satisfying the "fine-precision" accuracy threshold (i.e., 1°/0.100 m) for near real-time AR applications.

The remainder of this paper is structured as follows. Section II reports the methodology adopted for the development of the proposed system. In Section III, experiment design, execution, and results are reported. Finally, Section I is devoted to results discussion and conclusions.

II. Methodology

A. System architecture

This study proposes a markerless infrastructure-free localization system based on the comparison between a query image and a SfM-based 3D mapping assumed as a reference. The system architecture of the proposed approach, established on an AR cloud platform, includes three macro-phases: 3D mapping, 6-DoF localization, and BIM uploading (Fig. 1). The 3D mapping consists in generating, just once for the explored environment, a 3D reconstructed model through SfM to be used as a reference for localization. More in details, reference photos must be shot and uploaded into the AR cloud platform in order to extract and match feature points through computer vision Convolutional Neural Networks (CNN). In this study, Hierarchical Feature Network (HF-Net) technology [16], which simultaneously detect feature keypoints and compute local and global descriptors, is applied. Next step is determining camera intrinsics. They can be estimated from EXIF metadata or retrieved from a camera calibration [8]. Hence, a 3D model can be reconstructed from SfM technology and geo-registered by providing absolute coordinates of at least three feature/shot points [8]. As a results, the 3D reconstructed point cloud is transformed into its real scale, position, and orientation, so that absolute coordinates can be retrieved for each one of its points. The AR cloud platform enables uploading and displaying the geo-referenced BIM model. Displaying images pose, that will be computed in the next steps, along with the 3D geometry of the BIM model will make the user experience more intuitive. Tasks described so far must be executed just once for the explored environment.

The 6-DoF localization consists in retrieving the pose of the device given a query photo. Hence, once a query photo has been shot and uploaded to the AR cloud platform, its feature points can be extracted and matched, through HF-Net technology, with the ones of the 3D mapping reference photos identified before. Next step is determining camera intrinsics. As described above, they can be estimated from EXIF metadata or retrieved from a camera calibration [8]. At this point, the pose, describing the 6-DoF position and orientation of the query photo, can be computed with reference to the 3D geo-registered reconstructed model by applying the Perspective-n-Point (PnP) algorithm [17]. As a result, the frustum of the query image can be displayed in the context of the geometry of the BIM model. This study focuses on implementing tasks no. 1-4 and 6-8 and carrying out an accuracy assessment of the 3D mapping and 6-DoF localization paving the way for future studies including also tasks no. 5 and 9 (Fig. 1).

B. System implementation

The system architecture presented in Section II.A and depicted in Fig. 1 was implemented as a set of micro-services deployed using state-of-the-art containerization technologies (viz. Docker). At the core there is a backend web-service exposing a RESTful API allowing to:

- Initializing, deleting, and renaming a project.
- Uploading more photos into the 3D mapping dataset of the project or delete existing photos.
- Starting a new 3D mapping task, identifying feature points of the dataset photos, identifying the pose of each photo, and assigning (x,y,z) coordinates to each identified feature point, indeed generating a 3D point cloud from the dataset photos.
- Uploading and 6-DoF localizing a query photo that needs to be localized w.r.t. the 3D point cloud generated during the 3D mapping task.

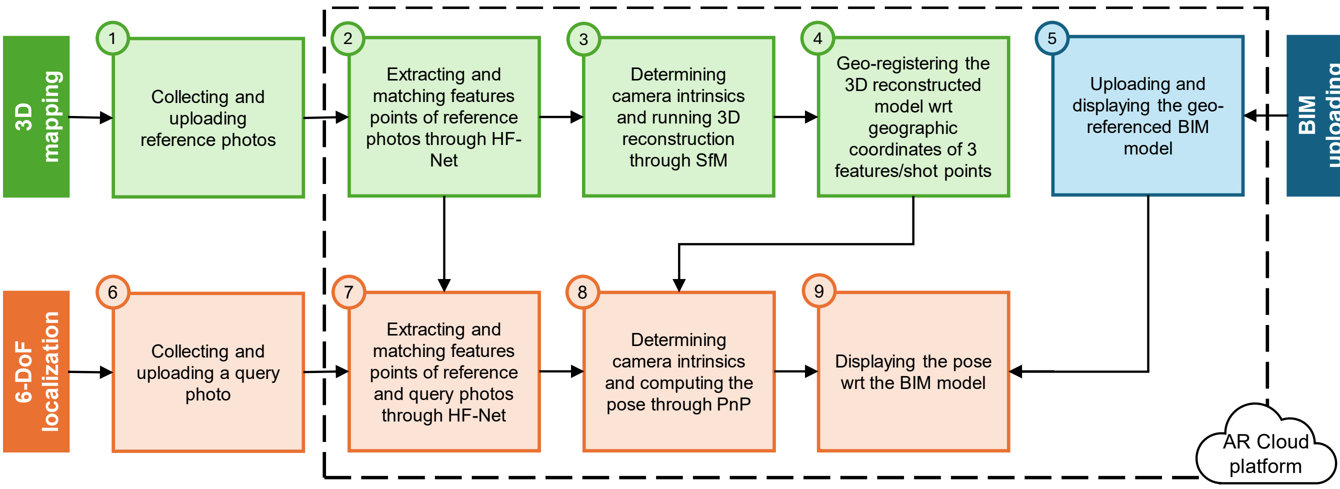

Fig. 1. System architecture of the proposed localization system.

- Exporting the generated 3D point cloud as well as the uploaded dataset and query photos.

The backend web service stores permanent information using volumes provided by a Network-Attached Storage (NAS): while the original pictures are stored as files, the extracted information (e.g., feature points and 3D point clouds) are stored using one database for each project. In addition to the backend web service, a minimal responsive frontend web GUI allows to test the implemented architectures from desktop computers as well as from mobile devices (e.g., smartphones and tablets), that can thus be already employed to run 3D mapping and 6-DoF localization tasks on the field, uploading photos taken using their cameras.

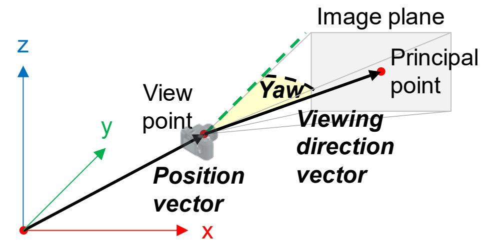

Fig. 2. Overview of computed parameters for accuracy assessment.

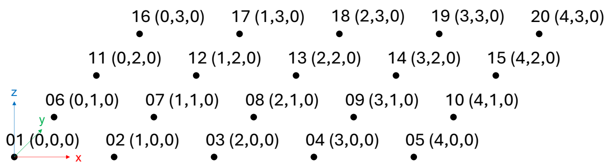

Fig. 3. 5x4 grid describing the positions of the dataset photos.

C. Accuracy assessment

The proposed localization system was tested in order to assess its suitability for application in AR navigation systems. To this purpose, an accuracy assessment of the proposed localization system was carried out by comparing position and orientation (i.e., pose) of photos estimated through the proposed localization system with the ones measured in the real world for the same photos. The localization system provides the viewpoint position of each photo as an estimated position vector, commonly defined "tvec", referred to the coordinate system of the 3D reconstructed model (Fig. 2). The position error of a photo, namely pos_error, is defined by (1) as the Euclidean distance:

$$pos_{error} = est_{pos} - real_{pos} \qquad (1)$$

where est_pos is the position vector estimated for the photo by the proposed localization system, whereas $real\_pos$ is the one measured in the real world during the shooting.

The ways the orientation of each photo is provided is twofold: an estimated quaternion, commonly defined "qvec" [8], and Euler angles described by the estimated viewing direction vector of the camera (Fig. 2). Euler angles are considered in this study since they can be easier measured than quaternions in the real world. Considering the navigation system domain, the rotation around the z axis or yaw rotation is the most relevant since determining the direction of the motion (i.e., left or right). For this reason, the orientation error of a photo is computed with reference to the yaw angle. The latter corresponds to the rotation that the viewing direction vector describes around the z-axis w.r.t. the y-axis (Fig. 2). The yaw error of a photo, namely yaw_error, is defined by (2):

$$yaw\_error = est\_yaw - real\_yaw \qquad (2)$$

where est_yaw is the yaw angle described by the viewing direction vector estimated for the photo by the proposed localization system, whereas $real\_yaw$ is the one measured in the real world during the shooting.

III. Experiments

A. Experiment design and execution

In order to carry out the accuracy assessment of the proposed localization system according to Section II.C, the implemented system (Section II.B) was tested within a public facility. The Department of Construction, Civil Engineering and Architecture at the Università Politecnica delle Marche campus was selected as a case study. Two different smartphones, namely the HuaweiP60Pro and Xiaomi13Pro, were used in the experiment in order to simulate the most generic scenario in which the end user to be localized has a different smartphone from the one adopted during the 3D mapping. The experiment was executed using both smartphone combinations for 3D mapping and 6-DoF localization. In the combination no. 1, HuaweiP60Pro was used for the 3D mapping and the Xiaomi13Pro was used for the 6-DoF localization. In the combination no. 2, smartphones were swapped. Experiment tasks no. 1 and 6 (Fig. 1) were executed according to the following criteria:

- Choosing 20 positions within a 5x4 grid of 1 m unit where placing the camera tripod during the shooting. The positions are numbered from "01" to "20". For the purpose of this study, aiming to assess the 3D mapping and 6-DoF localization accuracy, a local coordinate system was set according to Fig. 2, so that the photo in position no. "01" has coordinates (0,0,0). The rest of the photos are described by the coordinates reported in Fig. 3. According to this, the real position vector (i.e., $real\_pos$ in (1)) can be measured for each photo.

- Choosing 2 directions to target the camera during the shooting. To this purpose, a laser point that is free to rotates with the camera on its tripod was used. The targets are referred to as "A" and "B". According to this, the yaw angle (i.e., $real\_yaw$ in (1)) can be measured for each photo.

- Setting the smartphone on the camera tripod with the camera at 1 m from the ground and aligned with the vertical camera tripod axis. This task was repeated for each one of the two selected smartphones, namely HuaweiP60Pro and Xiaomi13Pro.

- Shooting photos using both smartphones in order to define the 3D mapping datasets (i.e., task no. 1, Fig. 1), including 20 photos. To this purpose, photos were shot for each position from "01" to "20" and alternating targets starting from "A".

- Shooting photos using both smartphones in order to define the 6-DoF localization datasets (i.e., task no. 6, Fig. 1), including 20 photos. To this purpose, photos were shot for each position from "01" to "20" and alternating targets starting from "B".

Experiment tasks no. 2 and 7 (Fig. 1) were carried out by setting respectively 300 and 150 as the maximum feature points numbers and 640 pixels as the maximum resizing dimension. Tasks no. 3 and 8 (Fig. 1) require camera intrinsics to be executed. For the task no. 3, camera intrinsics of both smartphones, applied for the 3D mapping, were determined through a high-resolution calibration (i.e., "HRC_intr", consisting in setting 5000 as the maximum feature points number and 1600 pixels as the maximum resizing dimension). For the task no. 8, instead, camera intrinsics of both smartphones, applied for the 6-DoF localization, were determined according to two different criteria defining the following two scenarios:

1) "EXIF_intr": in this scenario, 6-DoF localization camera intrinsics were estimated from EXIF.

2) "HRC_intr": in this scenario, 6-DoF localization camera intrinsics were determined through a high-resolution calibration (i.e., HRC_intr, consisting in setting 5000 as the maximum feature points number and 1600 pixels as the maximum resizing dimension).

Tasks no. 4 (Fig. 1) was executed by considering the position of 3 shot points. For the purpose of this study, aiming to assess the 3D mapping and 6-DoF localization accuracy, the positions of 3 shot points were determined assuming a local coordinate system as depicted by Fig. 3. Briefly, the execution of tasks no. 1-4 and 6-8 provide, respectively for each photo of the 3D mapping and 6-DoF localization dataset, the estimated position vector (i.e., est_pos in (1)) and yaw angle (i.e., est_yaw in (2)). As reported above, tasks no. 5 and 9 will be addressed in further studies.

B. *Results and discussion*

For both smartphone combinations, position and orientation errors (i.e., pos_error and yaw_error in (1) and (2)) were computed for each one of the 20 photos of the 3D mapping and the 6-DoF localization datasets. Average values and related standard deviations are reported in Table 1. For both smartphone combinations, the 3D mapping accuracy was assessed for the "HRC_intr" camera intrisics evaluation method. In the combination no. 1, an average pos_error of 0.023 m and a yaw_error of 1.021° was determined. The former resulted dispersed for 0.014 m and the latter for 0.180°. In the combination no. 2, an average pos_error of 0.023 m and a yaw_error of 0.402° was determined. The former resulted dispersed for 0.019 m and the latter for 0.298°.

For both smartphone combinations, the 6-DoF localization accuracy was assessed for two camera intrinsics evaluation methods, resulting into the "EXIF_intr" and "HRC_intr" scenarios. In the combination no. 1 and "EXIF_intr" evaluation method, an average pos_error of 0.072 m and a yaw_error of 0.475° was determined. The former resulted dispersed for 0.042 m and the latter for 0.350°. The 6-DoF localization accuracy, for the "HRC_intr" scenario, is described by an average pos_error of 0.043 m and a yaw_error of 0.319°. The former resulted dispersed for 0.016 m and the latter for 0.273°. In the combination no. 2 and "EXIF_intr" evaluation method, an average pos_error of 0.090 m and a yaw_error of 1.165° was determined. The former resulted dispersed for 0.071 m and the latter for 0.348°. The 6-DoF localization accuracy, for the "HRC_intr" scenario, is described by an average pos_error of 0.039 m and a yaw_error of 1.290°. The former resulted dispersed for 0.022 m and the latter for 0.210°.

In summary, in the combination no. 1, the 3D mapping pose estimation performed the best in terms of pos_error and the worst in terms of yaw_error. In the combination no. 2, instead, the 3D mapping pose estimation performed the best in terms of both pos_error and yaw_error. For both smartphone combinations and for the 6-DoF localization, better position accuracy was registered for the "HRC_intr" scenario compared to the "EXIF_intr" one. This confirmed expectations that an accurate initial camera calibration performed only once, based for example on a few photos that the user may be asked to collect, produces better 6-DoF localization results. Existing studies report the "fine precision" accuracy threshold, corresponding to 1°/0.100 m, as the minimum accuracy required for a good AR user experience [18]. As described above, the position accuracy of the proposed localization system fully satisfies that requirement. In particular, the aforementioned accuracy threshold is largely met if camera intrinsics are determined through a high-resolution calibration (i.e., "HRC_intr" scenario). Only the yaw_error values of the 3D mapping for the combination no. 1 and of the 6-DoF localization for the combination no. 2 are on the threshold.

Some final considerations must be done about computing times. Experiments were carried out using a laptop equipped with an Intel® CoreTM i7-8750H CPU 2.20 GHz processor, 16 GB of RAM, and an NVIDIA GeForce GTX 1050. In the combination no. 1, the 3D mapping computing time resulted equals to 21.745 s for the whole dataset and 0.114 s per pair of matched photos. In the combination no. 2, instead, the 3D mapping computing time resulted equals to 20.651 s for the whole dataset and 0.109 s per pair of matched photos. In the combination no. 1, the 6-DoF localization computing time, for the "EXIF_intr" scenario, resulted equals to 1.914 s for the whole dataset and 0.096 s per pair of matched photos. The 6-DoF localization computing time, for the "HRC_intr" scenario, instead, resulted equals to 1.584 s for the whole dataset and 0.079 s per pair of matched photos. In the combination no. 2, the 6-DoF localization computing time, for the "EXIF_intr" scenario, resulted equals to 1.862 s for the whole dataset and 0.093 s per pair of matched photos. The 6-DoF localization computing time, for the "HRC_intr" scenario, instead, resulted equals to 1.679 s for the whole dataset and 0.084 s per pair of matched photos. Briefly, a slower 3D mapping does not represent an issue since it can be done once for all and recalled for localization purposed. Instead, for both smartphone combinations, 6-DoF localization in less than 2 s would ensure near-real time applications within the AR navigation system domain.

Finally, comparable results were obtained for both smartphone combinations in terms of both accuracy and computing times. Hence, on the overall, which device is selected for 3D mapping and 6-DoF localization purposes marginally affect results.

IV. CONCLUSIONS

Combining AR with navigation systems offers an effective solution for assisting users in navigating unfamiliar environments, such as indoor public facilities. Understanding how to navigate indoor spaces is critical for individuals with disabilities, particularly for the visually impaired ones, but it also brings benefits to anyone exploring unfamiliar surroundings. The primary challenge in implementing AR navigation systems is accurately localizing the user in 6-DoF.

TABLE 1. EXPERIMENT RESULTS

		Combination no. 1			Combination no. 2		
		3D mapping HuaweiP60Pro	**6-DoF localization Xiaomi13Pro**		**3D mapping Xiaomi13Pro**	**6-DoF localization HuaweiP60Pro**	
Camera intrinsics evaluation method		***HRC_intr***	***EXIF_intr***	***HRC_intr***	***HRC_intr***	***EXIF_intr***	***HRC_intr***
Maximum feature points no.		***300***	***150***	***150***	***300***	***150***	***150***
Maximum resizing [pixels]		***640***	***640***	***640***	***640***	***640***	***640***
Pos_error [m]	**Average**	0.023	0.072	0.043	0.023	0.090	0.039
	Stand. dev.	0.014	0.042	0.016	0.019	0.071	0.022
Yaw_error [°]	**Average**	1.021	0.475	0.319	0.402	1.165	1.290
	Stand. dev.	0.180	0.350	0.273	0.298	0.348	0.210
Computing time [s]	**Whole dataset**	21.745	1.914	1.584	20.651	1.862	1.679
	Per pair	0.114	0.096	0.079	0.109	0.093	0.084

Unlike outdoor environments, indoor settings lack GNSS signals, presenting unique challenges. Various indoor localization methods have emerged, employing diverse system strategies and technologies. Key attributes for widespread adoption of AR navigation systems include precision, affordability, ease of use, and accessibility for non-technical users.

Addressing these criteria, this study proposes a markerless infrastructure-free localization system for AR registration in indoor environments, leveraging image comparison technology. This system functions by comparing a query image with a SfM 3D mapping serving as a reference. Recognizing the prevalence of tablets and smartphones among individuals in public spaces, additional functionalities should be incorporated to support users in their daily activities. To this end, the proposed localization system was deployed as a web service and tested on a university campus to assess its suitability for AR registration purposes. Considering the 2 smartphone combinations, experimental 6-DoF localization results showed, respectively for the "EXIF_intr"/"HRC_intr" scenarios, average position errors of 0.08/0.04 m and average orientation errors of 0.8/0.8°, within less than 2 s. These results fully meet the accuracy requirements for AR applications, corresponding to 1°/0.100 m and would ensure near real-time applications. Given the very promising localization accuracy and computation efficiency, future studies will further develop the proposed system testing AR registration of BIM models and assessing its accuracy.

ACKNOWLEDGMENT

The research has been partially funded by the European Union - NextGenerationEU under the Italian Ministry of University and Research (MUR) National Innovation Ecosystem grant ECS00000041 – VITALITY - CUP I33C22001330007.

REFERENCES

[1] C. Prandi, G. Delnevo, P. Salomoni, and S. Mirri, "On supporting university communities in indoor wayfinding: An inclusive design approach," *Sensors*, vol. 21, no. 9, May 2021, doi: 10.3390/s21093134.

[2] D. Khan, Z. Cheng, H. Uchiyama, S. Ali, M. Asshad, and K. Kiyokawa, "Recent advances in vision-based indoor navigation: A systematic literature review," *Computers and Graphics (Pergamon)*, vol. 104, pp. 24–45, May 2022, doi: 10.1016/j.cag.2022.03.005.

[3] M. Albahbah, S. Kıvrak, and G. Arslan, "Application areas of augmented reality and virtual reality in construction project management: A scoping review," *Journal of Construction Engineering*, vol. 4, pp. 151–172, 2021, doi: 10.31462/jcemi.2021.03151172.

[4] C. S. Park, D. Y. Lee, O. S. Kwon, and X. Wang, "A framework for proactive construction defect management using BIM, augmented reality and ontology-based data collection template," *Autom Constr*, vol. 33, pp. 61–71, 2013, doi: 10.1016/j.autcon.2012.09.010.

[5] C. Kuo, T. Jeng, and I. Yang, "An invisible head marker tracking system for indoor mobile augmented reality," *Autom Constr*, vol. 33, pp. 104–115, 2013, doi: 10.1016/j.autcon.2012.09.011.

[6] B. Naticchia, M. Vaccarini, A. Corneli, L. Messi, and A. Carbonari, "Leveraging Extended Reality technologies with RFID to enhance on field maintenance of buildings," 2021. Accessed: Aug. 03, 2023. [Online]. Available: http://itc.scix.net/paper/w78-2021-paper-038

[7] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski, "ORB: An efficient alternative to SIFT or SURF," in *2011 International Conference on Computer Vision*, 2011. doi: 10.1109/ICCV.2011.6126544.

[8] J. L. Schönberger and J.-M. Frahm, "Structure-from-Motion Revisited," 2016. [Online]. Available: https://github.com/colmap/colmap.

[9] Y. Li, N. Snavely, D. Huttenlocher, and P. Fua, "Worldwide Pose Estimation using 3D Point Clouds," 2012. doi: https://doi.org/10.1007/978-3-642-33718-5_2.

[10] F. Baek, I. Ha, and H. Kim, "Augmented reality system for facility management using image-based indoor localization," *Autom Constr*, vol. 99, pp. 18–26, Mar. 2019, doi: 10.1016/j.autcon.2018.11.034.

[11] G. Retscher, "Indoor Navigation—User Requirements, State-of-the-Art and Developments for Smartphone Localization," *Geomatics*, vol. 3, no. 1, pp. 1–46, Dec. 2022, doi: 10.3390/geomatics3010001.

[12] C. Kropp, C. Koch, and M. König, "Interior construction state recognition with 4D BIM registered image sequences," *Autom Constr*, vol. 86, pp. 11–32, Feb. 2018, doi: 10.1016/j.autcon.2017.10.027.

[13] K. Asadi, H. Ramshankar, M. Noghabaei, and K. Han, "Real-Time Image Localization and Registration with BIM Using Perspective Alignment for Indoor Monitoring of Construction," *Journal of Computing in Civil Engineering*, vol. 33, no. 5, Sep. 2019, doi: 10.1061/(asce)cp.1943-5487.0000847.

[14] Y. Wei and B. Akinci, "Panorama-to-model registration through integration of image retrieval and semantic reprojection," *Autom Constr*, vol. 140, Aug. 2022, doi: 10.1016/j.autcon.2022.104356.

[15] L. Messi, F. Spegni, M. Vaccarini, A. Corneli, and L. Binni, "Seamless Indoor/Outdoor Marker-Less Augmented Reality Registration Supporting Facility Management Operations," in *Proceedings e report*, Firenze University Press, 2023, pp. 109–120. doi: 10.36253/979-12-215-0289-3.11.

[16] P.-E. Sarlin, C. Cadena, R. Siegwart, and M. Dymczyk, "From Coarse to Fine: Robust Hierarchical Localization at Large Scale," Dec. 2018, [Online]. Available: http://arxiv.org/abs/1812.03506

[17] OpenCV, "Perspective-n-Point (PnP) pose computation." Accessed: Dec. 22, 2023. [Online]. Available: https://docs.opencv.org/4.x/d5/d1f/calib3d_solvePnP.html

[18] T. Sattler *et al.*, "Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions," Jul. 2017. [Online]. Available: http://arxiv.org/abs/1707.09092