Markerless Augmented Reality for Facility Management: Automated Spatial Registration based on Style Transfer Generative Network

Junjie Chen [a], Shuai Li [b], Weisheng Lu [a], Donghai Liu [c], Da Hu [b], and Maohong Tang [a]

[a] Department of Real Estate and Construction, The University of Hong Kong, Hong Kong, China
[b] Department of Civil and Environmental Engineering, The University of Tennessee, Knoxville, TN, USA
[c] State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300350, China
E-mail: chenjj10@hku.hk, sli48@utk.edu, wilsonlu@hku.hk, liudh@tju.edu.cn, dhu5@vols.utk.edu, litang.luke@gmail.com

This is the authors' version of the paper:

Chen, J., Li, S., Lu, W., Liu, D., Hu, D., & Tang, M. (2021). Markerless Augmented Reality for Facility Management: Automated Spatial Registration based on Style Transfer Generative Network. In Proceedings of the 38th International Symposium on Automation and Robotics in Construction (ISARC2021), 1-5 Nov 2021, Dubai, UAE. DOI: 10.22260/ISARC2021/0064

The final version of this paper is available at: https://doi.org/10.22260/ISARC2021/0064

Abstract

On-demand and real-time building information is of great value to support facility management. Such information can be easily retrieved from an up-to-date building information model (BIM), and then intuitively presented to facility managers or inspectors by augmented reality (AR). However, effective spatial registration into BIM so as to align the virtual and real content still remains an unresolved challenge. Leveraging recent development in the field of generative networks, this paper proposes a markerless registration approach that can automatically align BIM with the real view captured by a mobile device without any manual operation. A mobile AR application is develop based on the proposed registration approach. Our field experiments demonstrate the effectiveness of the proposed approach for automated BIM registration. The successful registration thus allows users to access the rich building information, especially invisible utility such as the mechanical, electrical and plumbing (MEP) system, in the real-life context for better facility management practice.

Keywords: Augmented reality; Facility management; Markerless spatial registration; Building information model (BIM); Generative adversarial network (GAN).

1. Introduction

The importance of facility management (FM) can never be overstated for built environments such as buildings and civil infrastructure [1, 2]. Without periodical inspection and maintenance, undesired defects or safety issues can continuously deteriorate, eventually leading to disastrous consequences in terms of economics, life, and social impacts [3]. During FM, inspectors or facility managers rely on relevant building information (e.g., equipment specifications, building materials, and spatial layout) to support decision making and condition assessment. This is especially true for utilities such as the mechanical, electrical and plumbing (MEP) system as they are normally invisible to humans and thus requires information on their layout and geometry to assist maintenance. In traditional FM practice, such information is manually retrieved from an archive of paper-based documents or materials such as drawings. The practice is laborious, time-consuming and cumbersome, calling for a more effective solution.

The rapid development of building information model (BIM) and its wide adoption provide a unified source where rich building information can be easily retrieved. Driven by the development, many mobile applications based on tablets and smart phones have been developed to bring the abundant information to the field for assisting decision-making [4, 5]. Although these applications have significantly streamlined the information retrieval process, they still suffer from the several limitations: a) require manual search of relevant information from the mobile device; b) insufficient intuitiveness due to a lack of connection between the virtual content and the real-life context. The limitations can be addressed by the application of augmented reality (AR), which aims at connecting the cyber space with the physical world by superimposing virtual content onto the videos or photos of the real world [6-10].

A core component of AR is spatial registration that can align the virtual and real world according to their relative positions [1]. Most of existing AR solutions used 2D fiducial markers, or 3D objects with certain patterns as a target for registration. Such solutions require significant efforts in deploying and maintaining the markers in the building of interest [11]. Another stream of AR registration approaches seeks to utilize the global navigation satellite system (GNSS), Bluetooth, ultra-wideband (UWB), or other radio signals to estimate the position and orientation of an AR device in the coordinate frame of the virtual model. The estimated posture is then used to align the model with the real-world view. Such systems rely on external signal emission/receival infrastructure, which may not always be available (e.g., GNSS is not applicable in buildings), and require additional time and efforts for installation and maintenance as well.

BIM, as a readily available asset, shares similar visual appearance to its physical counterpart in the real world. Leveraging the similarity, there is potential to directly register a real-life photo into BIM for subsequent AR facility management applications, avoiding the reliance on markers or external infrastructure [12]. This study presents a markerless image-to-BIM registration approach that does not require manual initialization or pre-calibration. It does so by innovatively leveraging a style transfer model based on generative adversarial networks (GAN) to address the cross-domain gap between the virtual and real contents. A prototype mobile application has been developed to demonstrate the effectiveness of the registration approach to enable AR-based FM. The field experiments show that actional information can be easily retrieved from BIM with our AR application to support instant decision-making.

2. Literature review

2.1. Spatial registration for AR application

To align the virtual contents in BIM with the real-world view, spatial registration is an indispensable module for AR systems. According to the differences of registration principles, existing approaches can be divided into two major categories, i.e., marker-based and markerless registration [1]. The markers used in the former approaches are usually 2D images composed of distinctive visual features, which are referred to as fiducial markers [13-15]. In some cases, 3D objects with distinctive features can be pre-mapped to serve as registration markers as well [16]. Although marker-based approaches can generally achieve precise and robust registration, they require manual efforts to install and maintain the markers [17, 18]. In addition, the approaches are subject to the drifting issue as users go farther away from the markers [1]. For many markerless solutions, an AR device's location in the virtual world is determined by calculating its relative position to external radio signal emission/receival infrastructure, e.g., GNSS [19], Wi-Fi [8], and UWB [20]. Orientation of the AR device is estimated with data collected by its onboard sensors such as gyroscope, electrical compass and accelerometer. Combining the calculated location and orientation, registration can then be realized. These solutions rely on an external radio signal system, which is usually expensive to deploy and maintain. In some cases, such systems can be even inaccessible (e.g., GNSS in indoor environments).

Research efforts have been made to develop registration techniques without the need of deploying markers or infrastructure. A line of major endeavors attempts to utilize computer vision and photogrammetry to recover the camera pose based on data retrieved from pre-mapped visual assets of the building of interests [18, 21]. However, significant efforts are still required to pre-map the environments. BIM shares similar visual appearance to its real-life counterpart, which can serve as a potential database of both the spatial and visual information of the environments to be registered, thus eliminating the requirements of pre-mapping. However, due to distinctive differences between the visual representations of BIM (plain texture) and real photos (abundant texture and details), there is a perception gap across the two domains. In [11, 22], manual initialization is required to designate a rough location, and then precise alignment was achieved by iterative closest point or perspective alignment. To overcome the cross-domain gap, Ha et al.

[12] and Baek et al. [6] proposed an image retrieval algorithm for indoor localization based on deep learning features. However, camera localization based on image retrieval is subject to the issue of viewpoint change, which can undermine the performance of registration [23]. The limitations of existing research call for a vision-based markerless approach that can automatically and precisely register into BIM across domains.

2.2. AR-assisted facility management

Facility management is important to ensure the operation of buildings and infrastructure. As a knowledge-intensive activity, relevant building information is required to support FM decision-making. Traditionally, such information is retrieved from paper-based design documents such as drawings. The cumbersome practice has been alleviated with the proliferation of BIM, from which rich building information can be easily accessed via mobile devices such as tablets and smart phones [4, 5]. The practice can be further improved by aligning the virtual objects and their relevant information in BIM with the real scenario captured by cameras through the use of AR [6-10].

In the area of civil and infrastructure engineering, many research attempts have been made to incorporate AR in FM for better efficiency and productivity. Kamat and El-Tawil [24] investigated the feasibility and potential of AR in evaluating structural damages. Chen et al. [8] devised a location aware AR framework for facility maintenance, which integrated Wi-Fi fingerprinting and room identification to achieve location awareness and can stream the site situation and inspection video back to remote control center. Chen et al. [9] integrated AR and BIM to facilitate the maintenance of fire safety equipment using mobile devices. Baek et al. [6] developed an AR system to assist FM based on the indoor localization algorithm proposed by [12]. The system used a client/server structure, and can visualize hidden pipelines in the field of view of Microsoft HoloLens.

Despite the significant progress made by previous studies, the BIM registration of existing systems either relies on external signal infrastructure for localization [8, 9, 24] or requires certain extent of manual operation for location initialization [11, 12]. In this study, we will explore how a markerless vision-based camera pose estimation approach proposed by [25] can be integrated in AR application to support FM.

3. Cross-domain spatial registration to style-transfer BIM

A cross-domain gap between textureless virtual models and real-life scenarios hinders the direct registration into BIM. As shown by Figure 1, a prior research by [25] indicates that style transfer based on generative adversarial networks (GAN) such as CycleGAN can bridge the cross-domain gap to enable meter-level indoor localization. Based on the findings, this study proposes an automated image-to-BIM registration method without the need of manual initialization.

(1) Style transfer based on generative adversarial networks

CycleGAN, a powerful unpair image-to-image translation GAN model [26], is used to train a deep learning model to realize style transfer between the BIM domain and the Real domain. To train the model, images from both domains need to be collected. In this study, images from 7 locations (around 2800 images for each domain) were taken in an indoor space of 112.8 m2. After training, one of the generators is then used to convert the BIM-rendered images in the same indoor environments to photorealistic ones with vivid texture. The generator has a near real-time performance, which can convert a given image in less than 1 s. Hence, processing the entire dataset (~2800 images) consumes no more than one hour. For more details on how the GAN model was trained, readers are suggested to refer to our previous work [25].

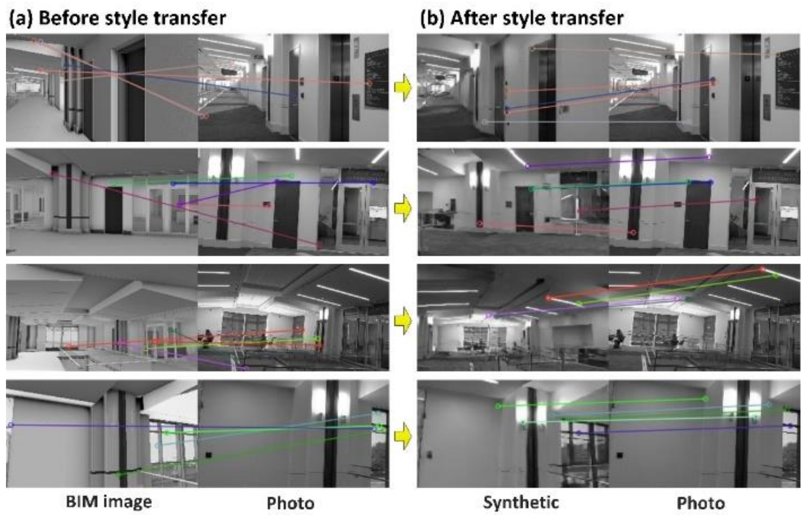

Figure 1. The effects of style transfer in addressing cross-domain gap

(2) Construction of a visual-spatial database

In the second step, a database of both visual and spatial information of the indoor environment is constructed based on the BIM images after style transfer. The visual information includes global and local visual features extracted from the BIM images. Chen et al. [25] demonstrates the superiority of edge histogram descriptor (EHD) over other global features in retrieving the style-transfer images; thus it will be used in our study. The local features used by us is scale invariant feature transform (SIFT). Spatial information refers to the real-world coordinates of SIFT keypoints detected from the BIM images. Such spatial coordinates can be extracted from BIM, as geometry layout of the environment is known priori in BIM. We developed a tool based on the application programming interface (API) provided by Autodesk Forge Viewer to automatically retrieve real-world coordinates corresponding to the keypoints from BIM in batch. Note that all the works mentioned in this step are implemented in an offline manner, which means they do not occur in the BIM registration process. Rather, its purpose is to prepare a database for registration in the subsequent step.

(3) BIM registration based on similarity comparison and photogrammetry

A newly captured photo (referred to as a query photo) in an indoor environment can be registered into the corresponding BIM by the following steps. First, the most similar style-transfer image to the query photo is retrieved from the database based on the cosine similarity between their EHD. Then, SIFT keypoint correspondences will be detected between the query photo and the retrieved BIM images after style transfer. As the 3D world coordinates of the keypoints have been extracted in last step, several pairs of 2D-3D correspondences with known pixel and world coordinates can be obtained. Based on these correspondences, the query photo's camera position and orientation will then be calculated by solving a classical perspective-n-points (PnP) problem in photogrammetry theory. Finally, the estimated camera posture can be used by the AR device to align BIM and the real world.

4. Development of the markerless AR application

4.1. Overall development framework

Figure 2 shows the overall development framework of the markerless AR mobile application (APP) for facility management. Unity is used as the core development platform, under which the Android Studio needs to be installed to provide environment support for building and running the APP. We use Visual Studio to write and edit C# script for the realization of required functionalities such as taking photos and spatial registration. For model preparation, third-party software, SimLab Composer, is used as a middleware to ensure the BIM created by Revit can be safely imported to Unity without losing texture. To use image processing and CV related algorithms (e.g., SIFT detection and matching) in Unity environment, an external package called "OpenCV for Unity" also needs to import. After properly setting up the scenario and coding, the project can be built to release an APP that can run on any Android devices.

4.2. Model preparation and importing

Revit files cannot be directly imported into Unity. Instead, they need to be first converted to formats compatible for Unity, e.g., .fbx and .obj. It seems viable to first export a Revit model to a FBX format, and then re-import it to Unity. However, doing so will lead to undesired loss of materials and textures, as shown by Figure 3 (a). To solve the problem, we adopt the following solution. First, the model is exported as a .dwfx file from Revit, which is then processed in SimLab Composer to generate a pack of files with both FBX and relevant materials extracted. Finally, the FBX model can be imported into Unity with proper texture and materials attached, as shown in Figure 3 (b).

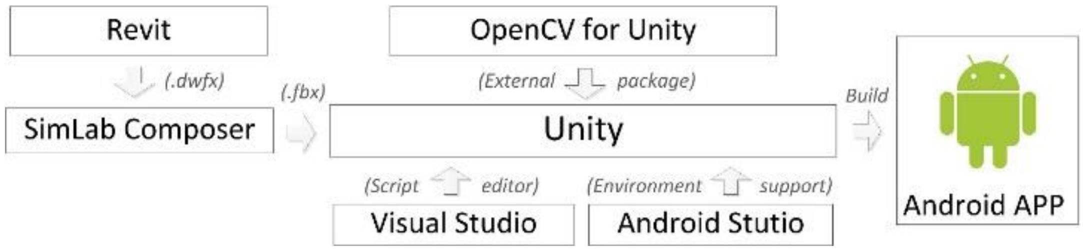

Figure 2. Overall technical framework for the application development

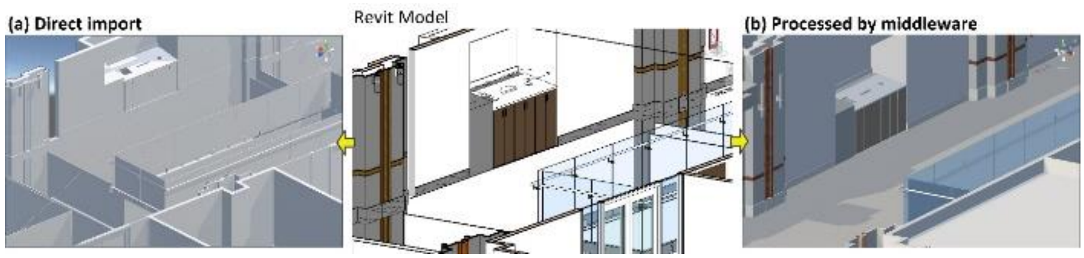

Figure 3. Results comparison between different model preparation approaches: (a) direct import of FBX file exported from Revit; (b) processed by third-party middleware before importing

During the above process, it is important to ensure the consistency of unit and up axis among different model processing platforms. It should be noted that Unity uses a left-handed coordinate system with the y axis as the up vector. To place the model in the same position as it is in Revit (so as to ensure a consistent coordinate system), we follow the below procedure: (a) re-assign pivot point of the model to a point with known coordinates; (b) specify the "transform.position()" attribute of the model as the coordinates of the known point; (c) adjust the rotation of the model as appropriate (e.g., 180° around the y axis in our case).

4.3. Realization of image-to-BIM registration

Since we intend to develop a standalone application without using servers, the visual-spatial database extracted from the style-transfer BIM needs to be properly incorporated as part of the software. In Unity, the Resources class allow developers to load and access assets of various formats, e.g., images, and text document. The EHD, SIFT features, image ID, and 2D and 3D coordinates of the key points are stored in separate text files. Note that sometimes a file might be too large to be loaded by the software loader in a time. In such cases, we divided data in the file and saved it into multiple different files, and added an indexing file on the top to ensure the easy retrieval of these files.

Next step is to retrieve correspondences based on visual features of the input query photo. A C# implementation of EHD [27] is used to calculate EHD features of the query photo. The calculation and matching of SIFT descriptors are realized by a third-party package called "OpenCV for Unity". Based on the 2D-3D coordinate correspondences retrieved by feature matching, the camera posture can be estimated with the Calib3d.solvePnPRansac() function from the OpenCV for Unity package. Finally, we can align the virtual model with the real view in the query image by passing the estimated camera posture to the transform.position() and transform.eulerAngles() function of the camera object in Unity.

4.4. Superimposing virtual objects onto real view

To superimpose the virtual BIM model onto real camera view of a mobile device, two camera objects in Unity need to be designated cooperatively. One camera is to capture the current view of the BIM model, and the other is used to display the streaming image of the real-life scenario recorded by the mobile phone camera in real time. The two cameras are designated to be present on the same display (i.e., phone screen), and the camera background canvas is placed behind the model by assigning a larger value to its depth attribute, thus ensuring the virtual model are overlaid onto the real view.

4.5. Device posture tracking

After initial registration, the device posture will be continually tracked to update the virtual model view to ensure alignment. The estimated camera posture in section 4.3 is used as the initial camera state, to which the relative camera orientation is then accumulated based on the gyroscope and accelerometer information obtained from the mobile device. Note that at current stage, translational movement of the device is not considered, but will be incorporated in the future by integrating other robust tracking techniques such as simultaneous localization and mapping (SLAM) or visual odometry.

5. Field experiments

To validate the effectiveness of the proposed approach, a field test was implemented in the student union (SU) complex of the University of Tennessee, Knoxville (UTK). The experiment site is at the northwest corner of the third floor of the SU, which covers an area of 12.0 m × 9.4 m.

5.1. Evaluation of the markerless registration approach

The data collection scheme and training of the style transfer model followed the settings and parameters mentioned in [25]. After training, the style transfer model was used to generate synthetic photorealistic images based on the given BIM-rendered images, and the visual-spatial database was then constructed.

To evaluate the efficacy of the proposed registration approach, 18 test points were set up, scattering evenly inside the experiment area, as shown in Figure 4 (a). Figure 4 (b)~(s) compare the real-life camera view taken by the mobile device and the resulting BIM view after registration at each test point. As can be seen from the figure, a great majority of the images have been successfully registered, leading to alignments with the virtual BIM model. Table 1 quantitatively analyzes the registration performance, showing a successful registration rate of 77.8%. It took 4.3 s in average to register an image. The results demonstrated the efficacy of the proposed markerless registration approach.

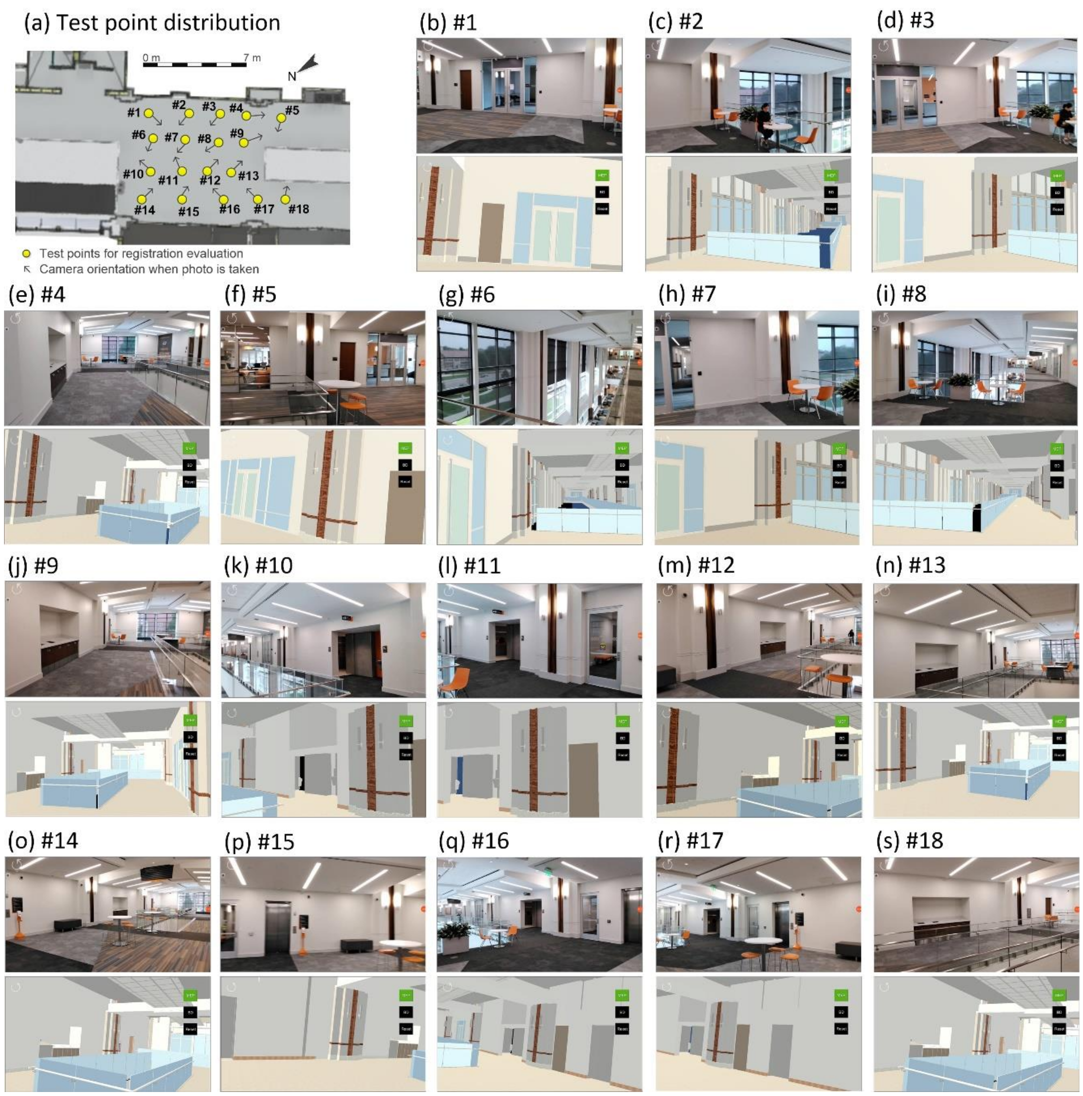

Figure 4. Setup and results of registration evaluation experiments: (a) distribution of test points in the experiment site; (b)~(s) image pairs showing level of alignment between virtual and real content after registration at each test point

Table 1. Summary of the registration performance

Type	Quantity	Proportion
Success	14	77.8%
Failure	4	22.2%
Total	18	100%

5.2. Demonstration of the mobile AR application

Figure 5 demonstrates the general procedure of using our mobile AR application for facility management. The application can be installed on tablets or any other mobile devices that support the Android operating system. After opening the application, users can click the “Photo” button at the bottom of the initial interface to enter the photo shooting page. Then, a photo of the surrounding environment should be taken by clicking the “SHOOT” button. Based on the taken photo, our proposed markerless image-to-BIM registration algorithm is executed to retrieve BIM model corresponding to the current camera view. Finally, after hiding the building structures, the effect of virtual-real alignment can be realized. By the use of our APP, the invisible utility such as MEP can be intuitively visualized, thus assisting decision-making for facility inspection.

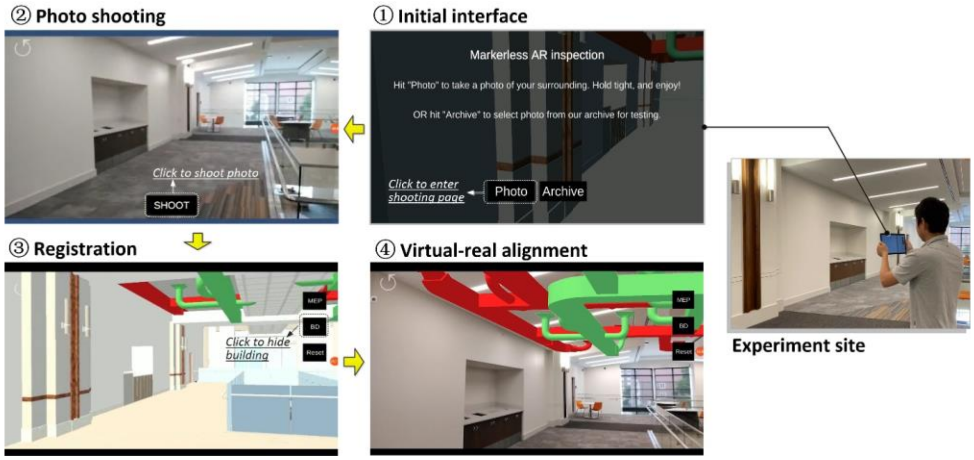

Figure 5. Workflow and corresponding interface of the developed mobile APP for facility management

6. Conclusions

Augmented reality is a useful technique to assist facility management of buildings and civil infrastructure. However, the promise and potential of AR has not been fully fulfilled due to the challenging issue of spatial registration. To tackle the challenge, this study proposes an image-to-BIM registration approach that requires neither fiducial markers nor external signal emission/receival infrastructure. The registration approach leverages CycleGAN style transfer to bridge the cross-domain gap between virtual and real content, and can estimate six degree-of-freedom camera pose by using classical photogrammetry theory. Based on the registration approach, a mobile AR application is developed to support facility management. The development used “Unity + Android studio” as the basic technical framework. Relevant development issues such as model preparation, realization of registration, virtual-real fusion, and device posture tracking were introduced subsequently. Experiments have been carried out to validate the effectiveness of the proposed approach and the AR application.

The proposed registration approach is challenged by several issues, e.g., illumination, deviation between BIM and real scenes, and architectural self-similarities. For the former two challenges, our previous research [25] has demonstrated the approach’s robustness against varying lighting conditions and certain changes of the physical environments (e.g., adding furniture that does not exist in BIM). For the last challenge, we acknowledge it as a downside commonly experienced by many computer vision applications. To overcome, it is suggested to integrate the approach with other sensors (e.g., barometers and gyroscope) to reduce the ambiguities among indoor spaces with similar designs or layout. Future research should also address the issue of motion tracking, and test out the approach’s generalizability to different types of indoor spaces, e.g., conference rooms and corridors.

References

[1] Cheng J. C. P., Chen K. Y. and Chen W. W. State-of-the-Art Review on Mixed Reality Applications in the AECO Industry. Journal of Construction Engineering and Management, 146(2), 12, 2020.

[2] Palmarini R., Erkoyuncu J. A., Roy R. and Torabmostaedi H. A systematic review of augmented reality applications in maintenance. *Robotics and Computer-Integrated Manufacturing*, *49*, 215-228, 2018.

[3] Chen J. and Liu D. Bottom-up image detection of water channel slope damages based on superpixel segmentation and support vector machine. *Advanced Engineering Informatics*, *47*, 101205, 2021.

[4] Park J., Cai H. B. and Perissin D. Bringing Information to the Field: Automated Photo Registration and 4D BIM. *Journal of Computing in Civil Engineering*, *32*(2), 2018.

[5] Lin K.-Y., Tsai M.-H., Gatti U. C., Je-Chian Lin J., Lee C.-H. and Kang S.-C. A user-centered information and communication technology (ICT) tool to improve safety inspections. *Automation in Construction*, *48*, 53-63, 2014.

[6] Baek F., Ha I. and Kim H. Augmented reality system for facility management using image-based indoor localization. *Automation in Construction*, *99*, 18-26, 2019.

[7] Carozza L., Valero E., Bosché F., Banfill G., Mall R. and Nguyen M. UrbanPlanAR: BIM Mobile Visualisation in Urban Environments with Occlusion-Aware Augmented Reality. In *Proceedings of the Joint Conference on Computing in Construction (JC3)*, pages 229-236, Heraklion, Greece, 2017.

[8] Chen K. Y., Chen W. W., Li C. T. and Cheng J. C. P. A BIM-Based Location Aware AR Collaborative Framework for Facility Maintenance Management. *Journal of Information Technology in Construction*, *24*, 360-380, 2019.

[9] Chen Y. J., Lai Y. S. and Lin Y. H. BIM-based augmented reality inspection and maintenance of fire safety equipment. *Automation in Construction*, *110*, 18, 2020.

[10] Kwon O.-S., Park C.-S. and Lim C.-R. A defect management system for reinforced concrete work utilizing BIM, image-matching and augmented reality. *Automation in Construction*, *46*, 74-81, 2014.

[11] Kopsida M. and Brilakis I. (2016). Markerless BIM Registration for Mobile Augmented Reality Based Inspection. In *Proceedings of the International Conference on Smart Infrastructure and Construction*, pages. 1-6. 2016.

[12] Ha I., Kim H., Park S. and Kim H. Image retrieval using BIM and features from pretrained VGG network for indoor localization. *Building and Environment*, *140*, 23-31, 2018.

[13] Feng C. and Kamat V. Augmented Reality Markers as Spatial Indices for Indoor Mobile AECFM Applications. In *Proceedings of International Conference on Construction Applications of Virtual Reality (CONVR)*, Taipei, Taiwan, 2012.

[14] Yabuki N., Miyashita K. and Fukuda T. An invisible height evaluation system for building height regulation to preserve good landscapes using augmented reality. *Automation in Construction*, *20*(3), 228-235, 2011.

[15] Lee S. and Akin Ö. Augmented reality-based computational fieldwork support for equipment operations and maintenance. *Automation in Construction*, *20*(4), 338-352, 2011.

[16] Hu D., Hou F., Blakely J. and Li S. Augmented Reality Based Visualization for Concrete Bridge Deck Deterioration Characterized by Ground Penetrating Radar. In *Construction Research Congress 2020: Computer Applications*, pages 1156-1164, Arizona, US, 2020.

[17] Davila Delgado J. M., Oyedele L., Demian P. and Beach T. A research agenda for augmented and virtual reality in architecture, engineering and construction. *Advanced Engineering Informatics*, *45*, 101122, 2020.

[18] Carozza L., Tingdahl D., Bosché F. and van Gool L. Markerless Vision-Based Augmented Reality for Urban Planning. *Computer-Aided Civil and Infrastructure Engineering*, *29*(1), 2-17, 2014.

[19] Behzadan A. H., Timm B. W. and Kamat V. R. General-purpose modular hardware and software framework for mobile outdoor augmented reality applications in engineering. *Advanced Engineering Informatics*, *22*(1), 90-105, 2008.

[20] Teizer J., Venugopal M. and Walia A. Ultrawideband for Automated Real-Time Three-Dimensional Location Sensing for Workforce, Equipment, and Material Positioning and Tracking. *Transportation Research Record*, *2081*(1), 56-64, 2008.

[21] Bae H., Golparvar-Fard M. and White J. High-precision vision-based mobile augmented reality system for context-aware architectural, engineering, construction and facility management (AEC/FM) applications. *Visualization in Engineering*, *1*(1), 3, 2013.

[22] Asadi K., Ramshankar H., Noghabaei M. and Han K. Real-Time Image Localization and Registration with BIM Using Perspective Alignment for Indoor Monitoring of Construction. *Journal of Computing in Civil Engineering*, *33*(5), 04019031, 2019.

[23] Piasco N., Sidibé D., Demonceaux C. and Gouet-Brunet V. A survey on Visual-Based Localization: On the benefit of heterogeneous data. *Pattern Recognition*, *74*, 90-109, 2018.

[24] Kamat V. R. and El-Tawil S. Evaluation of augmented reality for rapid assessment of earthquake-induced building damage. *Journal of Computing in Civil Engineering*, *21*(5), 303-310, 2007.

[25] Chen J., Li S., Liu D. and Lu W. Indoor camera pose estimation via style-transfer 3D models. *Computer-Aided Civil and Infrastructure Engineering*, 2021.

[26] Zhu J.-Y., Park T., Isola P. and Efros A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. pages 2223-2232, 2017.

[27] Lux M. and Chatzichristofis S. A. Lire: lucene image retrieval: an extensible java CBIR library. pages 1085-1088, 2008.