SYNTHESIZING 360-DEGREE LIVE STREAMING FOR AN ERASED BACKGROUND TO STUDY RENOVATION USING MIXED REALITY

YUEHAN ZHU[1], TOMOHIRO FUKUDA[2] and
NOBUYOSHI YABUKI[3]
[1,2,3]*Osaka University, Suita, Osaka, Japan*
[1]*zhu@it.see.eng.osaka-u.ac.jp* [2,3]*{fukuda|
yabuki}@see.eng.osaka-u.ac.jp*

Abstract. In a modern society, people spend more time indoors. Indoor Environmental Quality (IEQ) and its effect on occupants' health and comfort has become an important area of study. Many existing building stocks still have huge social, economic, and environmental value. There is a high demand for stock renovation, which gives existing buildings new lives, rather than building new ones. In the early stage of the renovation design, it is essential to achieve a timely feedback process as bring together stakeholders. Introducing Mixed Reality (MR) with Diminished Reality (DR) provides users with an indirect view of the world where some objects have been made invisible which makes it easier to display indoor renovation plans. This paper describes the development of an MR system for architectural designers that integrates DR results into the MR system. Aiming to provide a stable, realistic and real-time DR results for enhancing feedback efficiency during renovation design which can help stakeholders better understand or evaluate the renovation plan.

Keywords. Building stock renovation; mixed reality (MR); diminished reality (DR); real-time background update.

1. Introduction

In an advanced society, the world is gradually interested in health and wellness inside buildings. It is said that people spend an average of 87% of their time indoors (Klepeis et al., 2001). Therefore, it is important to provide people with a comfortable and healthy indoor environment. Indoor Environmental Quality (IEQ) and its effect on occupants' health and comfort has always been considered in new buildings. However, the proportion of building stock is much larger than new buildings (Kovacic et al., 2015). Many existing building stocks still have huge social, economic, and environmental value. There is a high demand for stock renovation, which gives existing buildings new lives, rather than building new ones.

In order to reduce the cost and decrease environmental impacts during the renovation process, using Mixed Reality (MR) technology to display indoor renovation plans has been proposed (Zhu et al., 2018). MR is the one where real

Intelligent & Informed, Proceedings of the 24th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2019, Volume 2, 71-80.

world and virtual world objects coexist (Milgram and Kishino, 1994). Introducing MR technology into the renovation design steps enables the simultaneous design of the building plan and environment design which can reduce the construction period with less coordination, timely feedback and fewer variation orders. The MR experience makes non-professionals can also understand and participate in the design process. MR including augmented reality (AR), virtual reality (VR) and diminished reality (DR). In contrast to AR which superimpose virtual information on the real scene, DR is the process of removing, eliminating, or diminishing the number of objects from the real world (Nakajima et al., 2017). The introduction of DR technology provides users an indirect view of the world where some objects have been made invisible which makes it easier to display indoor renovation plans, as removing an existing wall is one of the basic steps in the building stock renovation.

There are many methods to achieve DR. In-painting is one of the methods that attempt to paint over objects using texture and patch information from the source image itself (Jan and Broll, 2014). New algorithms of the image in-painting methods are constantly being proposed. Such as combining Visual-SLAM into the algorithm, which is a method to segment the background surfaces uses some feature points of the background (Kawai et al., 2013). However, those method combining Visual-SLAM are not available on the solid-colored wall. And the more satisfied the DR results, the more computing resources are needed (Sasanuma et al., 2017). Then, the time required for calculations will increase, and real-time DR will be hard to realize. On the other hand, in-painting methods usually used to erase small and independent objects. Relative to this, building structure such as wall and column is a relatively large object and connects to the environment. Therefore, it is hard to use the in-painting methods to erase an entire wall. In this research, we propose a method that can achieve realistic large-scale DR in real-time.

Another set of methods utilizes pre-captured images or video of a background scene. Then, when new physical items are incorporated into space, the background images can be used as a reference for obtaining background information obstructed by the new objects. Those methods utilize pre-captured background cannot display the real-time background and observation angle provided to users is limited to the angle of the pre-captured background. Therefore, to achieve a realistic large-scale DR, reconstructing the background area virtually in real-time is necessary.

Previously, a system for DR with multiple handheld camera system has been proposed (Akihito et al., 2007). They used multiple cameras to capture the same scene from different directions, and used the AR marker to calibrate the multiple cameras, so that they can computing the occluded content, diminished the target object. However, its DR results are not stabilized because of its algorithm based on pixel-by-pixel comparison does not work if the occluding object has no color variation across its surface. Another advanced DR method utilizes an RGB-D camera to hide arbitrary trackable objects (Meerits and Saito, 2015a). They used RGB-D camera to reconstruct the missing parts then introduce color correction technique to merge image content originating from different cameras. In this method, because of the performance of RGB-D camera itself, reconstruction results would have many missing parts which seem like black noise. Although they used in-painting way to repair those black noise, the quality of its DR results still has its limitations, and some black noise still exist.

This paper describes the development and application of a real-time MR system for architectural designers that includes DR function. The system proposes a method that erases a real wall virtually to connect two rooms. Introducing SLAM based MR technology to do the registration and display a stable background at the specified location. The background reconstruction is operated by a 360-degree camera, and the reconstruction results are sent to Head Mounted Display (HMD) through a Wi-Fi connection in real-time. Users can experience the MR system on the HMD and study renovation with a wall disappearing. This MR experience can help them better understand or evaluate the renovation plan flexibility.

2. The Proposed MR System

In accordance with Figure 1, this system consists of four well-defined steps, namely, real-time data collection, background reconstruction, renovation plan digitization, and system integration. The system includes one HMD (HoloLens) and one 360-degree camera. The HoloLens has four "environment understanding" cameras, one RGB camera and one Gyro sensor. This set of equipment can provide calibration information for registration and occlusion generation, as seen in Figure 2. The streaming data used to reconstruct the background is collected, via a 360-degree camera, which is placed in the central location of different rooms from the HMD. These two devices communicate via Wi-Fi connection. Wi-Fi signals can pass through walls, which makes it possible to collect background data behind the target wall. The renovation plan is to build a BIM model that can provide professional information by designers that will make it faster and easier for designers to modify the renovation plan. Synchronizing the BIM model with the 3D coordinates and occlusion information not only ensures the accuracy of the DR result but also allows the user to have a more realistic MR experience. The entire system operates on HMD, and so, users can watch the DR results from its see-through display. From the user's perspective view, the target wall will be diminished, the room behind the physical wall will be displayed in front of the user, and as the user can walk back and forth, the occlusion of the background room and the physical objects around the diminished wall will be correctly displayed.

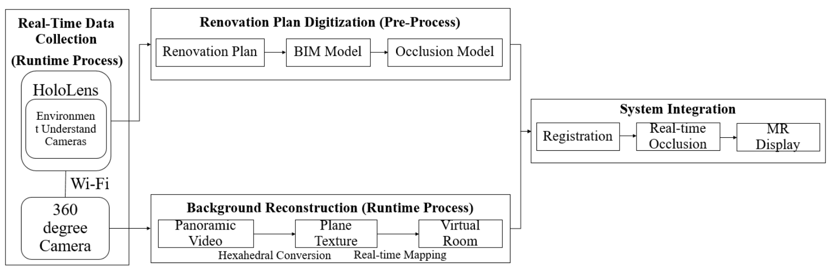

Figure 1. System workflow.

Figure 2. HoloLens side view (left); front view (right).

2.1. REAL-TIME DATA COLLECTION

The structure of the background scene is reconstructed in real-time. Many existing DR approaches that are based on in-painting methods have achieved real-time object elimination. One of the approaches realized an entire image completion pipeline without relying on pose tracking or 3D modelling (Herling and Broll, 2010). However, the DR results of this methods are static and the DR target is limited to simple objects without a complex pattern. This is because its background information is based on the characteristics of the surrounding environment. The more complex the target object is, the more unexpected the DR result comes out. Therefore, methods for generating background scenes using 3D reconstruction have been frequently used in recent years. PhotoAR+DR project proposed a reconstruction method using photogrammetry software which manufactures background 3D model by reconstructing the photographs of the surrounding environments (Inoue et al., 2018). The generated background is pretty realistic but it takes a long time to process for ten hours or even days. Another research team presented a general method of visualizing hidden areas using an RGB-D camera (Meerits and Saito, 2015b). This method can also reconstruct dynamic background scenes. However, the construction data captured from RGB-D camera is point cloud data. The point cloud data files are extremely large; thus, the real-time reconstruction can be delayed or part of the scene may be missing. Researchers have to reduce the density of point cloud data in order to correct the stability of DR results. Overall, in this paper, we will propose a method to collect background scene information using a 360-degree camera. In comparison to the general RGB camera and RGB-D camera, the 360-degree camera has a wider field of view, which can scan the entire background scene all at once. The file size of the resulting panorama is significantly smaller than the point cloud data, where a wired connection is mandatory in order to transfer data. In our research, 360-degree camera data size is small enough to use wireless transmission. First, we will need to create a plugin using the API provided by the camera manufacturer, and then, we need to build a Wi-Fi connection between the HMD and 360-degree camera. The connection method is based on the HttpWebRequest class, and the data is usually acquired and submitted through GET and POST. With this wireless network transmission connection, panoramic video can be transferred to the HMD in real-time.

2.2. BACKGROUND RECONSTRUCTION

2.2.1. Panorama Conversion

Our proposed system gets the original panoramic video data byte-by-byte from the 360-degree camera, and separates the image of each frame from the data stream. However, this original panoramic image cannot be directly used to rebuild the background. Because this will have a serious image distortion. Converting the panorama image into another form that fits the mask model is necessary to reduce the distortion. Mask model is a 3D virtual model of the diminished target, and it can determine which part of the virtual room for DR is exposed in front of the occlusion model. A 1:1 hexahedron room mask model to map the background image. Then, the panoramic photos are converted into six patches and attached to the hexahedron. The algorithm refers to a part of a VR study that discusses the transformation of a cubic environment map (90-degree perspective projections onto the face of a cube) into a cylindrical panoramic image (Paul Bourke, 2016). Basically, there are two solutions that can split the panoramic image. The first of which is based on OpenCV image processing and the other is based on OpenGL. The latter changes the direction of the LookAt camera, obtain six textures, and then, read and save the return data. In this study, we need to process the panoramic image from the data stream in real-time. The OpenGL methods would cost GPU computing resources, which is a weak point of HMD. Hence, the OpenCV methods are chosen.

2.2.2. The Dynamic Texture on the Mask Model

In this study, the background scene should not be a static image or a model, a dynamic virtual room is created to replace the diminished wall. The mask model is a hollow cube with one empty side. The system continually receives the converted images that are dynamically mapped to the mask model (except the diminished wall that is planned to be removed in the renovation design). After that, the real-time dynamic DR results are shown on the mask model.

2.3. OCCLUSION

Occlusion happens when one object in a 3D space is blocking another object from view. In order to convey a more realistic MR experience, the physical objects should interact with virtual objects. The occlusion of incorrect objects makes the DR result look somewhat unreal, which can be seen in Figure 3. In this study, our proposed system not only scans and generates the real-world objects occlusion in real-time, using four “environment understanding” cameras from HoloLens, which can perceive the depth of the scene like RGB-D camera; the system also deletes the real-time occlusion of the diminished wall. In order to prevent the occlusion of the DR target wall, which is scanned by the “environment understanding” cameras from covering the background scene, the range of real-time occlusion generation should be limited. The occlusion of the target wall is not generated by limiting the scan range of the “environment understanding camera”.

Figure 3. DR result without occlusion.

2.4. SYSTEM INTEGRATION

Every module is integrated into a game engine, Unity, which supports Universal Windows Platform (UWP) development. Moreover, programs developed using UWP can operate on all Windows 10 devices including the HMD-HoloLens. The panorama conversion function is based on OpenCV (2.2.1) and its library needs to be imported into the Unity Asset as optional attachment packages.

3. System Evaluation

To validate the advantages of the proposed method, an evaluation test was carried out in room 411 and room 410, M3 building, Suita Campus, Osaka University. The DR target wall was between the two rooms and the length and height of the target wall are 6.5 m and 2.6 m respectively. The observational positions were set at A, B, and C in the room 411 (Figure 4). The 360-degree camera, THETA V, was placed in the center location of the room 410. The overall layout is shown in Figure 4. The system data flow of the DR part was shown in Figure 5. The captured panoramic video was streamed from THETA V to HoloLens via Wi-Fi connection. We used the method refer to THETA developers' API 2.1 (THETA developers) to build the transfer protocol. In this method, we first used HttpWebRequest class and set the POST-request; then, parsed motion-jpeg HTTP stream from the camera and generated each frame from streaming data. After that, we used OpenCV-based methods to convert the panoramic images into a Hexahedral map. This can be seen in Figure 6. At last, we mapped each frame onto a mask model as a texture and waited for it to finish processing. After all the steps were complete, the integrated results were projected on a translucent screen (HoloLens). This is so that people could "perspective" through the DR target wall. The real-world perspective of the user experiencing the MR system was shown in Figure 7. In the Figure 7, the part that is wrapped in the red frame was the target wall. A part of the target wall that is close to the entrance is not taken as a target of elimination. This is simply because it was a bearing part.

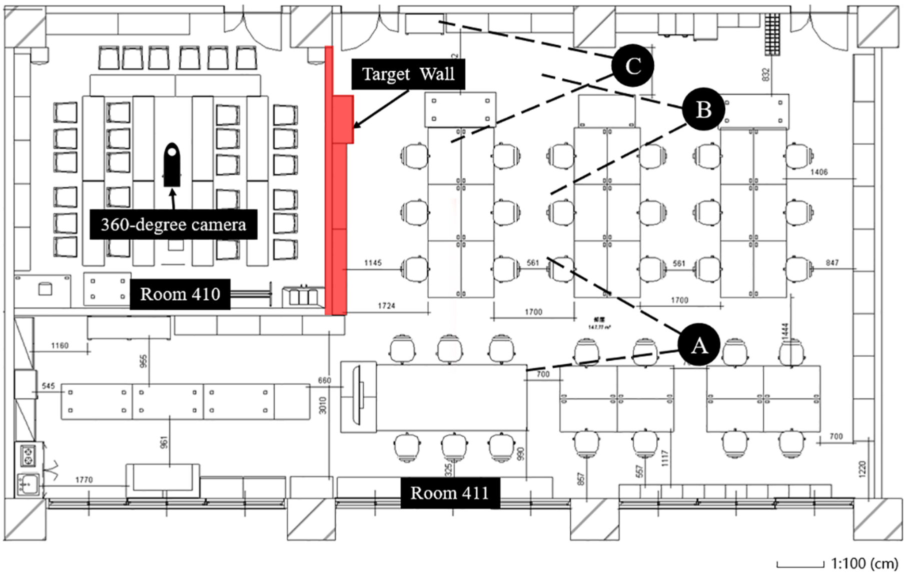

Figure 4. Room floor plan.

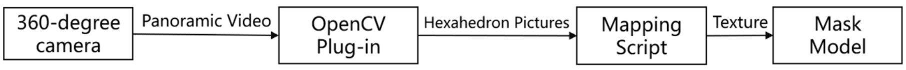

Figure 5. System data flow.

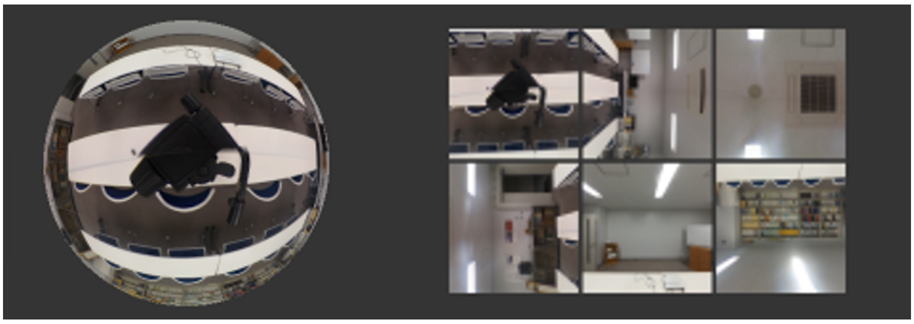

Figure 6. Panorama Image Conversion.

Figure 7. Reality view (red frame: DR target wall and shelves).

3.1. RESULTS

The original condition of the target wall is shown in Figure 7. The experimental results using the proposed MR system are shown in Figure 8, which contains 2D segmentation maps. In the segmentation maps, the grey parts are real-world objects while the blue parts belong to the background scene. The five planes of the mask model and the walls, the ceiling and the floor of room 410 are coordinated to coincident in the real-world position. Therefore, when the user walks around, they can experience the correct sense of space within the virtual room 410. The DR results which are observed through HMD made users feel that the target wall has disappeared. In room 410, the movements behind the target wall are also displayed in the virtual room 411 immediately. The registration function is stabilized while the virtual object shows no drift and disappearance.

However, the occlusion is not fully displayed correctly. From the segmentation map, we can conclude out that a part of the occlusion edge was shown incorrectly, and several virtual parts that should not appear in the DR results are scattered in the results. The reason for this result is that the occlusion result generated by real-time scanning is not accurate. The lack of scanning accuracy is the main cause of this problem. In addition, due to the performance limitation of the HoloLens performance, the results itself do contain systematic errors. Besides the irregular edge, the whiteboard, which should be in front of the virtual room, is covered by the virtual part. Although the range of real-time occlusion generation is set between the whiteboard and the target wall, the distance between the whiteboard and the wall is too close, and because of this, occasional errors occur.

Furthermore, the DR results are dynamic, and the refresh rate is still low. When testing this MR system on a PC platform, the entire process is fast enough to run in real-time. However, the image transmission speed is limited by HoloLens' bandwidth, which, in turn, lowers the frame rates.

3.2. DISCUSSION

If we only rely on algorithms to guess the background, generating correct content for complex backgrounds is quite difficult, and probably impossible since there are no repeating patterns. As a result, we proposed the method of using 360-degree camera data to reconstruct the background. This method is not only suitable for complex backgrounds but also works on large, diminished targets.

Latency issues are common problems in streaming media transmission. In this study, the video quality cannot be declined because the virtual room will become blurred, which, in turn, affects the entire MR experience. Therefore, we chose to reduce the refresh rate. The program is changed manually in order to refresh the background. We set a button on the user interface, the background will be refreshed each time when the button is pressed.

Figure 8. MR result and its segmentation map in position A (above), B (middle), C (bottom); left: (MR Result) right (segmentation map).

4. Conclusions and Future Work

In this paper, we demonstrated a novel MR system that diminishes a wall and shows that details the movement of the background scene in real-time. The correct occlusion mixes the virtual world and the physical world which makes the MR experience appear more realistic. For designers, they can change the renovation plan in a short period of time and observe the effect in the MR system. For occupants, the MR experience helps them to fully understand the designer's intentions and have a clear concept of the renovation results.

Currently, the system is operated in LAN. As for future work, if we create a relay server, the system would be operated on WAN and it can rebuild a remote room in front of the user. During image conversion processing, in order to ensure the edge of each wall is accurately assigned to each surface of the mask model, segment parameters are set manually. However, with the rise of artificial intelligence in recent years, we believe the best approach to assign the correct parts is by using machine learning to train and figure out correct room parts, and to segment the panorama.

Acknowledgement

This research has been partly supported by JSPS KAKENHI Grant Number JP16K00707.

References

"Converting to/from cubemaps" : 2018. Available from Paul Bourke<http://paulbourke.net/miscellaneous/cubemaps/> (accessed 10th December 2018).

"THETA developers' API" : 2018. Available from <https://developers.theta360.com/en/docs/> (accessed 10th December 2018).

Akihito, E. and Hideo, S.: 2007, Diminished reality using multiple handheld cameras, *Proceedings of ICDSC-07*, Vienna, 251-258.

Herling, J. and Broll, W.: 2012, PixMix: A real-time approach to high-quality diminished reality, *Proceedings of IEEE ISMAR 2012*, 141-150.

Inoue, K., Fukuda, T., Cao, R., and Yabuki, N.: 2018, Tracking Robustness and Green View Index Estimation of Augmented and Diminished Reality for Environmental Design, *Proceedings of CAADRIA 2018*, Beijing, 339-348.

Jan, H. and Broll, W.: 2010, Advanced Self-Contained Object Removal for Realizing Real-Time Diminished Reality in Unconstrained Environments, *Proceedings of IEEE ISMAR 2010*, 207-212.

Neil E. Klepeis, K.l.e.p.e.i.s., William C., N.e.l.s.o.n., Wayne R., O.t.t., John P., R.o.b.i.n.s.o.n., Andy M., T.s.a.n.g. and Paul, S.w.i.t.z.e.r.: 2001, The national human activity pattern survey, *J. Expo. Anal. Environ. Epidemiol.*, **11**, 231-252.

Kovacic, I., Summer, M. and Achammer, C.: 2015, Strategies of building stock renovation for ageing society, *Journal of Cleaner Production*, **88**, 349-357.

Meerits, S. and Saito, H.: 2015a, Real-time diminished reality for dynamic scenes, *Proceedings of IEEE ISMAR-W 2015*, Fukuoka, 53-59.

Meerits, S. and Saito, H.: 2015b, Visualization of Dynamic Hidden Areas by Real- Time 3D Structure Acquistion Using RGB-D Camera, *Proceedings of 3D Systems and Applications Conference*, 2-5.

Milgram, P. and Kishino, A. F.: 1994, "Taxonomy of Mixed Reality Visual Displays, *IEICE Transactions on Information and Systems*, **E77-D**, 1321-1329.

Nakajima, Y., Mori, S. and Saito, H.: 2017, Semantic Object Selection and Detection for Diminished Reality based on SLAM with Viewpoint Class, *Proceedings of IEEE ISMAR-A 2017*, Natnes, 338-343.

Norihiko, K., Sato, T. and Yokoya, N.: 2013, Diminished Reality Considering Background Structures, *Proceedings of ISMAR 2013*, 259-260.

Sasanuma, H., Yoshitsugu, M. and Noriko, Y.: 2017, Diminishing Real Objects and Adding Virtual Objects Using a RGB-D Camera, *Proceedings of IEEE ISMAR-A 2016*, 117-120.

Zhu, Y., Fukuda, T. and Yabuki, N.: 2018, SLAM-Based MR with Animated CFD for Building Design Simulation, *Proceedings of CAADRIA 2018*, Beijing, 409-418.