



Article

A Mixed Reality Design System for Interior Renovation: Inpainting with 360-Degree Live Streaming and Generative Adversarial Networks after Removal

Yuehan Zhu, Tomohiro Fukuda * and Nobuyoshi Yabuki

Division of Sustainable Energy and Environmental Engineering, Osaka University, 2-1 Yamadaoka, Suita-shi, Osaka 565-0871, Japan; zhu@it.see.eng.osaka-u.ac.jp (Y.Z.); yabuki@see.eng.osaka-u.ac.jp (N.Y.)

* Correspondence: fukuda.tomohiro.see.eng@osaka-u.ac.jp (T.F.); Tel.: +81-6-6879-7661; Fax: +81-6-6879-7661

Abstract: In contemporary society, “Indoor Generation” is becoming increasingly prevalent, and spending long periods of time indoors affects well-being. Therefore, it is essential to research biophilic indoor environments and their impact on occupants. When it comes to existing building stocks, which hold significant social, economic, and environmental value, renovation should be considered before new construction. Providing swift feedback in the early stages of renovation can help stakeholders achieve consensus. Additionally, understanding proposed plans can greatly enhance the design of indoor environments. This paper presents a real-time system for architectural designers and stakeholders that integrates mixed reality (MR), diminished reality (DR), and generative adversarial networks (GANs). The system enables the generation of interior renovation drawings based on user preferences and designer styles via GANs. The system’s seamless integration of MR, DR, and GANs provides a unique and innovative approach to interior renovation design. MR and DR technologies then transform these 2D drawings into immersive experiences that help stakeholders evaluate and understand renovation proposals. In addition, we assess the quality of GAN-generated images using full-reference image quality assessment (FR-IQA) methods. The evaluation results indicate that most images demonstrate moderate quality. Almost all objects in the GAN-generated images can be identified by their names and purposes without any ambiguity or confusion. This demonstrates the system’s effectiveness in producing viable renovation visualizations. This research emphasizes the system’s role in enhancing feedback efficiency during renovation design, enabling stakeholders to fully evaluate and understand proposed renovations.

Keywords: interior renovation; mixed reality (MR); diminished reality (DR); biophilic design; generative adversarial networks (GANs)



Citation: Zhu, Y.; Fukuda, T.; Yabuki, N. A Mixed Reality Design System for Interior Renovation: Inpainting with 360-Degree Live Streaming and Generative Adversarial Networks after Removal. *Technologies* **2024**, *12*, 9. <https://doi.org/10.3390/technologies12010009>

Academic Editor: Mauro Iacono

Received: 26 October 2023
Revised: 24 December 2023
Accepted: 9 January 2024
Published: 11 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In modern society, there is a growing interest in the health and wellness of indoor spaces. It is said that people spend an average of 87% of their time indoors [1]. Therefore, it is crucial to provide people with a comfortable and healthy indoor environment. While indoor environmental quality (IEQ) and its impact on occupants’ health and comfort have always been considered in new buildings, the majority of buildings in use today were built before these concerns were widely recognized [2]. Existing building stocks retain significant value in terms of social, economic, and environmental aspects. Building stock renovation is in high demand as it gives existing buildings new life instead of constructing new ones. This research developed an MR system that enables the early assessment of renovation plans and improves communication between non-experts and experts, thereby increasing overall efficiency.

In order to reduce the cost and decrease environmental impacts during the renovation process, the use of mixed reality (MR) technology to display interior renovation plans has been proposed [3]. MR refers to a state in which digital information is superimposed to

some extent on the complete real world [4]. Introducing MR technology into renovation design steps enables the simultaneous design of the building plan and environment design that can reduce the construction period with less of a need for coordination, timely feedback, and fewer variation orders. The MR experience enables non-professionals to understand and participate in the design process.

MR encompasses a spectrum from the real to the virtual environment. Augmented reality (AR) enhances the real world with virtual elements, and virtual reality (VR) immerses users entirely in a virtual space. While diminished reality (DR) is a lesser-known term than AR and VR, it is an evolving concept within the field of MR and can be understood as the removal or modification of real-world content [5]. The introduction of DR technology enables users to have an indirect view of the world, where specific objects are rendered invisible, facilitating the presentation of interior renovation plans more effectively, as removing an existing wall is one of the basic steps in building stock renovation.

There are many methods to achieve DR. There are numerous methods for achieving DR. The key aspect is generating the erased part, referred to as the DR background, to enhance the overall satisfaction of the DR result.

In-painting is one of the methods that seek to apply textures and patch details directly from the source image to paint over objects [6]. New algorithms for image in-painting methods are constantly being proposed, such as combining Visual-SLAM (simultaneous localization and mapping) into the algorithm [7], which is a method that segments the background surfaces by using some feature points of the background. However, those methods combining Visual-SLAM are not available for solid-colored or large-area walls. Moreover, the more satisfying the DR results, the more computing resources are needed [8]. Then, the time required for calculations will increase, and real-time DR will be hard to realize. On the other hand, in-painting methods are usually used to erase small and independent objects, rather than for room-based interior renovation designs. This approach can encounter limitations when dealing with intricate indoor environments. Relative to this, building structures such as walls and columns are relatively large objects and are connected to the environment. Therefore, it is hard to use the in-painting methods to erase large objects like an entire wall. In this research, we propose a novel method to generate a DR background that can achieve realistic large-scale DR in real-time.

Another set of methods utilizes pre-captured images or video of a background scene [9]. Subsequently, as new physical elements are introduced into the space, the background images serve as a reference, providing information about the areas obscured by these new objects. Those methods that rely on pre-captured backgrounds cannot display the real-time background and limit the viewing angle provided to users to that of the pre-captured background. Therefore, to achieve realistic large-scale DR, it is necessary to virtually reconstruct a three-dimensional (3D) DR background area.

So far, a system using multiple handheld cameras for DR has been introduced [10]. This system used multiple cameras to capture the same scenes from different angles. AR markers were used to calibrate these cameras, allowing the calculation of occluding content and a reduction in target objects. However, the DR results from this method were not consistently stable, primarily because its pixel-by-pixel comparison algorithm falters when the occluding object lacks color variations on its surface. Another advanced DR technique uses an RGB-D camera to occlude trackable objects [11]. This approach used the RGB-D camera to piece together missing segments and then applied a color correction method to unify the images from the different cameras. Due to the inherent limitations of the RGB-D camera, the reconstructed images often contained gaps that resembled black noise. Although in-painting techniques were implemented to address these inconsistencies, the overall quality of the DR results remained suboptimal, with some residual black noise remaining. The proposed methods pre-capture the existing DR background and focus on removing the physical objects, while this study focuses on representing the renovation plan, which may still be under design. The existing DR background needs to be processed.

With the development of artificial intelligence, generative adversarial networks (GANs) have been used to produce data via deep learning [12], proposing new image generation methods. A stable GANs model was trained that demonstrated models for generating new examples of bedrooms [13]. Also, a study shows that training synthetic images and optimizing the training material can help recover target images better [14], which makes it easier to perform many image editing tasks. Furthermore, identifying different underlying variables can aid in editing scenes in GANs models, including indoor scenes [14,15]. However, there are not many GAN training models specifically for indoor panoramic images. Even when using indoor images synthesized from multiple angles, it is difficult to maintain a consistent style for objects in the same location. Using such synthesized images, it is virtually impossible to accurately reconstruct the indoor environment accurately.

This paper presents a real-time MR system for interior renovation stakeholders, using GANs for DR functions and SLAM-based MR, which provides an intuitive evaluation of the renovation plan and enables improved communication efficiency between non-experts and experts regarding the plan. The system proposes a method that virtually erases a real wall to connect two rooms: the introduction of SLAM-based MR technology to the perform registration, tracking, and display of a stable background at the specified location. The DR background is generated via a GAN method in advance, and the DR results are sent to a head-mounted display (HMD) via a Wi-Fi connection in real time. Users can experience these MR scenes on the HMD and study renovation by making a wall disappear. An earlier version of this paper was presented at the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2019 conference [16]. This paper presents significant subsequent advancements and new validations in the research. An option has been introduced to utilize the GAN method for producing DR backgrounds.

2. Literature Review

Interior renovation is the process of upgrading and modifying existing architectural structures and elements to make them more environmentally friendly or safer, or to meet other new needs. In the early stages of traditional design processes, designers use blueprints to communicate with stakeholders. With sufficient budgets and time, miniature physical models of an interior design, made of materials such as paper or foam, are occasionally created to facilitate more tangible communication with stakeholders. Whether they are hand-drawn blueprints or hand-crafted physical models, they are typically more time-consuming than digital methods, and human errors can introduce inaccuracies in measurement or design. In addition, with hand-drawn designs, any changes can add significant time and increase the budget, and may even require the entire design to be redrawn [17]. This is a common cause of time overruns in architectural projects. In the early stages of the design process, digital methods allow for more convenient sharing with team members or clients. Feedback is accelerated and the cost of changes is reduced [18].

2.1. Interior Renovation with Mixed Reality

Compared to demolition and reconstruction, interior renovation is proving to be a more environmentally friendly, budget-conscious and efficient method of giving existing buildings a new lease on life. A digitized interior model is essential to ensure that clients without specialized knowledge receive renovation results that meet their expectations. Unlike the more technically demanding blueprints, 3D models allow clients to immediately grasp the expected results of the renovation, provide feedback and get involved in the design process.

Compared with traditional methods that use desktop displays to present renovation results in a 2D plane, VR technology that visualizes various data and interior models in 3D space, allowing clients to immerse themselves in the renovation results from a free viewpoint experience, has been proposed to bridge the communication gap between designers and their clients [19]. It also validated the reliability and applicability of using VR technology in the architectural design process [20].

Compared to traditional tools, immersive VR technology offers significant advantages in the interior architectural design process [21]. However, while VR technology has matured significantly in interior design, it also presents some challenges. VR creates an entirely virtual environment, requiring the computer to create a comprehensive 3D model of the building's interior. To provide a more realistic immersive experience, some places that do not require renovation must also be modeled. For example, even if only part of a room is being renovated, the entire room must be modeled to provide a more authentic experience, including objects that serve only as a backdrop. The high up-front costs may cause designers to lose some clients who have not yet signed a contract. In addition, the more realistic the VR experience, the larger and more expensive the equipment required, further increasing the cost of modeling. Examples of such equipment include gigantic, multi-screen, high-definition displays [22], or the Cave Automatic Virtual Environment (CAVE) system, where an entire room transforms into a display [23].

Accordingly, AR technology has been introduced with more compact and flexible [24]. The AR background is the real object itself, requiring modeling only for added items, which significantly reduces the upfront design costs. For simple interior renovation or interior decoration that does not change the structural elements of the house, there may be no need for designers to be involved in the early design stages—even owners without specialized knowledge can manage it. Employing a new automated interior design algorithm, rational and personalized furniture arrangements can be generated [24,25]. AR technology can be used to easily evaluate the expected results of a renovation. For interior renovations that require structural changes to the house, modeling needs to involve the interior structures of the house. Directly importing models from building information modeling (BIM) into the AR system provides a method that is both simple and accurate [26].

However, many interior renovations involve not only the addition of more complex structures and furnishings, but also the removal of existing structures or large, immovable furniture. This is where DR technology becomes essential. Unlike AR technology, which adds information, DR technology removes it [27]. In architectural design, this technology can assist designers and stakeholders in better understanding the impact of design changes that might be imposed on the space. DR technology is widely used in both indoor and outdoor architectural renovation design processes, as well as environmental design processes [28]. DR technology is employed to eliminate buildings and landscapes slated for demolition and can even address pedestrians and vehicles [29], or be used to erase indoor furniture and decorations [30]. Its usability has been validated. This study used multiple cameras, including a depth camera, RGB camera, and 360-degree camera, to achieve real-time DR for large indoor areas. This provides a new approach to consensus building for complex interior renovations. However, due to device specification limitations, real-time streaming suffers from significant latency, resulting in a noticeable delay. The efficiency and accuracy of occlusion generation also needs improvement.

2.2. Generate Virtual Environment Using GANs

With the advancement of artificial intelligence, deep learning has introduced a variety of new and efficient approaches to image and model generation, among which GANs stand out. Briefly, 3D model creation is a complex and costly process that involves 3D capturing, 3D reconstruction, and rendering using computer graphics technology. By using GANs, a large number of house models can be automatically generated based on some simple requirements and prompts provided by the user, such as keywords, sketches, or color blocks [31]. Additionally, utilizing varied training datasets can result in different image effects that exude a unique style. For renovating older houses that lack BIM data, the house point cloud data obtained through the use of a laser scanner or depth camera, or photo data captured by cameras, can be inputted into the GAN method, which can be used to swiftly conduct automatic 3D model reconstruction [32]. These methods were proven to be highly beneficial for the early process of interior renovation design using MR technology, improving the feedback speed throughout the process while reducing the cost of model

creation. In addition, the method of using a 360-degree panoramic camera to capture the target room and the use of panoramic photos or videos as the MR environment was also proposed [33]. In an indoor environment where the structure is not overly complex, the shooting range of a 360-degree camera can cover most of the area observable by the human eye. Compared with the creation of an indoor environment using camera scans or multi-view photos, panoramic photos can complete the collection of environmental data in an instant and have a dominant advantage in terms of generation speed. Although panoramic photos have some shortcomings in terms of image quality, several solutions using GAN technology have been proposed to address this issue [34,35].

3. Proposed MR System with GANs Method

3.1. Overview of the Proposed Methodology

In accordance with Figure 1, this system consists of four well-defined steps, namely real-time data collection, background reconstruction (pre-process or runtime-process), renovation plan digitization, and system integration. Among them, the reconstructed DR background can be switched manually according to the needs of users. An HMD is required that is equipped with a depth camera and an RGB camera, is capable of receiving Wi-Fi data, and has either its own operating system or the ability to exchange data with a computer in real time. After researching HMDs that meet these criteria, at the time of this study, only the HoloLens met these requirements, leading to its selection. However, if an HMD that meets these requirements is developed in the near future, it could also be considered for use. The system includes an HMD (HoloLens Gen 1), a 360-degree camera, a Wi-Fi router, and a computer with a GPU (graphics processing unit) for processing GANs. HoloLens has four “environmental understanding” cameras, an RGB camera, and a gyro sensor. This set of equipment can provide calibration information for registration and occlusion generation, as shown in Figure 2. The streaming data used to reconstruct the background are collected by a 360-degree camera placed in the center of the room to be renovated. These two devices communicate through a Wi-Fi router. Wi-Fi signals can pass through walls, making it possible to collect information in the adjacent room. The renovation plan is to build a BIM model that can provide professional information to designers, making 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 MR result, but also allows the user to have a more realistic MR experience. Except for some pre-processing, the entire MR system operates on the HMD, allowing users to view the MR results through its see-through display. From the user’s perspective, the internal structures to be removed are reduced, the space behind the physical wall is displayed in front of the user, and since the user can walk back and forth, the occlusion of the background space and the physical objects around the reduced wall are correctly displayed.

3.2. Real-Time Data Collection

Real-time reconstruction of the background scene has become a key issue in recent research. Numerous DR methods based on in-painting methods have successfully achieved real-time object exclusion. One such approach has realized a comprehensive image completion pipeline, independent of pose tracking or 3D modeling [36]. However, the results of these DR methods tend to be static, mainly targeting simple objects without complex patterns. This limitation arises because the background information depends largely on the characteristics of the immediate environment. The complexity of the target object directly affects the predictability of the DR results: the more complex the object, the more unexpected the DR result. For this reason, in recent years, there has been a surge in the adoption of methods that use 3D reconstruction for background scene generation. For example, the Photo AR + DR project has developed a reconstruction technique using photogrammetry software that creates a 3D background model by assimilating photographs of the surrounding environment [37]. The generated background is quite realistic, but processing can take up to ten hours or even several days. In a related study, another research

team presented a method for visualizing hidden regions using an RGB-D camera [38]. This method is capable of reconstructing dynamic background scenes. However, the data captured by the RGB-D camera are in the form of point cloud data. Due to their large size, these point cloud files can hinder real-time reconstruction, resulting in occasional delays or missing scene parts. To improve the consistency of DR results, researchers often find it necessary to thin the density of the point cloud data. This paper presents a method for collecting background scene details using a 360-degree camera. Compared to standard RGB and RGB-D cameras, the 360-degree camera provides a wider field of view, allowing it to capture the entire background scene in a single sweep. Importantly, the file size of the resulting panorama is much smaller than the large point cloud data, eliminating the need for mandatory wired connections for data transfer. In this study, the data size of the 360-degree camera is compact enough to facilitate wireless transmission. First, a plugin is developed using the API provided by the camera manufacturer. Then, a Wi-Fi connection is established between the HMD and the 360-degree camera. This connection is based on the `HttpWebRequest` class, with data typically retrieved and sent via GET and POST methods. Using this wireless connection, a panoramic video can be streamed to the HMD in real time.

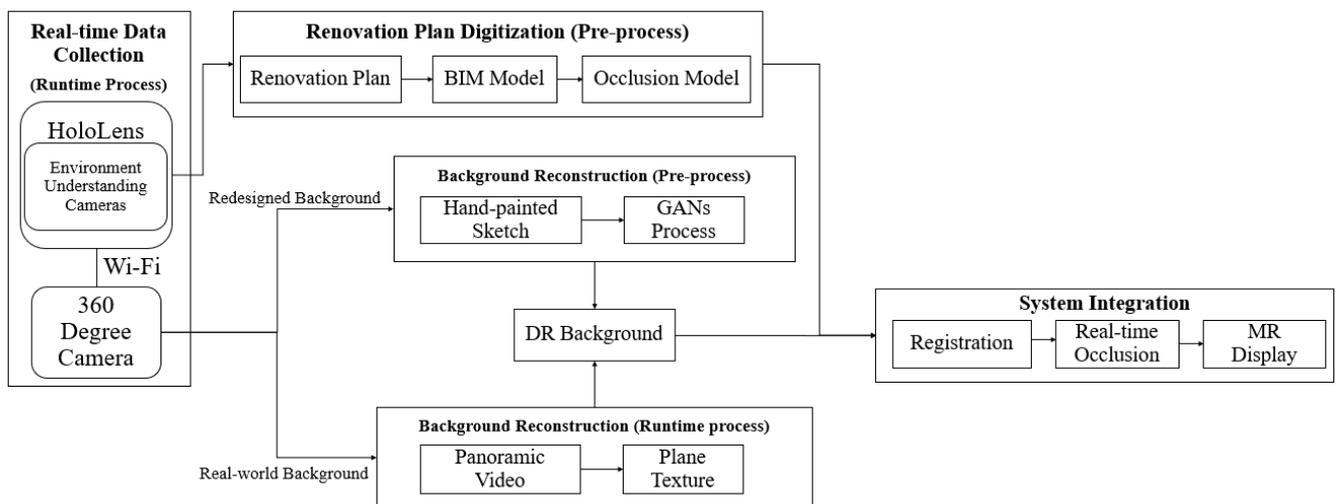


Figure 1. System overview.



Figure 2. Wearing HoloLens Gen 1 (left); HoloLens Gen 1 appearance (right).

3.3. Background Reconstruction

3.3.1. Panorama Conversion

The proposed system receives panoramic video data byte by byte from the 360-degree camera and isolates the image of each frame from the data stream. However, directly using this original panoramic image to reconstruct the background would result in significant image distortion. To overcome this, it is essential to transform the panoramic image into a format compatible with the mask model to minimize distortion. The mask model acts

as a 3D virtual representation of the target to be reduced and determines the segment of the virtual space for DR that is visible in front of the occlusion model. A 1:1 hexahedral space mask model is used to map the background image. The panoramic images are then segmented into six patches and applied to the hexahedron. The algorithm is influenced by a part of a VR study that explores the transformation of a cubic environment map (with 90-degree perspective projections on the face of a cube) into a cylindrical panoramic image [39]. There are two primary solutions for slicing the panoramic image: one uses OpenCV image processing and the other is based on OpenGL. The OpenGL approach adjusts the direction of the LookAt camera to obtain six textures, and then reads and stores the resulting data. For the real-time processing of panoramic images from the data stream required in this study, OpenGL methods could strain GPU computing resources, a known limitation of HMDs. Therefore, OpenCV methods are preferred.

3.3.2. Texture in the Dynamic Mask Model

In this study, the background scene is not represented by a single image or model. Instead, a virtual room is created to replace the reduced wall. The mask model takes the form of a hollow cube with one side missing. The system continuously captures converted images that are dynamically mapped to the mask model, except for the diminished wall that is to be removed in the renovation design. The real-time DR results are then displayed on the mask model. Alternatively, one can switch to the rendered images generated by the pre-trained GANs and map them onto the mask model.

3.4. GAN Generation

Enabling homeowners without design expertise to accurately articulate their design ideas has proven to be a challenge time and time again. These individuals lack the ability to visually represent design concepts or build models the way professional designers do. They also rarely have the luxury of time to devote to these endeavors. Furthermore, relying solely on verbal communication often fails to effectively communicate their vision. In response, a strategy was developed to transform hand-drawn sketches into renderings. Based on an image-to-image translation technique first described by Phillip Isola et al. [40], a model for transforming sketches into objects has been trained. Given a training set of pairs of related images, labeled 'A' and 'B', the model can learn to transform an image of type 'A' (sketches) into an image of type 'B' (renderings).

3.5. Occlusion

Occlusion occurs when one object in 3D space hides another object from view. For a more realistic DR experience, physical objects should interact with virtual objects. Incorrect occlusion can make the DR result appear less realistic, as shown in Figure 3. The system proposed in this study scans and generates real-world object occlusions in real time. It uses four "environment understanding" cameras from HoloLens, which can sense the depth of the scene similarly to an RGB-D camera. In addition, the system removes the real-time occlusion of the wall it is intended to reduce. To ensure that the wall intended for DR is not obscured by the occlusions detected by the "environment understanding" cameras, the scope of real-time occlusion creation is limited. By limiting the scan area of the environmental understanding camera, the occlusion of the target wall is not created.

3.6. System Integration

Except for some pre-processing steps of GANs, which include model training and rendering image generation—performed on a computer equipped with the Ubuntu system—all other modules are integrated into the game engine. This engine supports Universal Windows Platform (UWP) development. In particular, programs developed using UWP are compatible with all Windows 10 devices, including the HMD HoloLens. Because of its support for UWP development and the inclusion of the development kit required for this experiment, Unity (2018.2) will be utilized. The panorama conversion functionality is

rooted in OpenCV (2.2.1), and its library will need to be imported into the Unity asset as a supplemental package.



Figure 3. DR result without occlusion.

4. Implementation

In order to validate the advantages of the proposed method, an evaluation was conducted in rooms 410 and 411 of the M3 building at the Suita Campus of Osaka University. Room 410 covers an area of 42.4 m², while room 411 covers an area of 150 m². The wall designated for the DR target was located between these two rooms, with dimensions of 6.5 m in length and 2.6 m in height. The observation points were to be concentrated mainly on the right half of the room (as shown in Figure 4). The area within the red dotted line is the location of the DR background area in the real world. The red shadow area is the wall that plans to be diminished in this evaluation test. The 360-degree camera, THETA V, was positioned centrally in room 410. The overall arrangement is shown in Figure 4.

4.1. System Data Flow

The system data flow for the DR function is shown in Figure 5. Two different sources provide the reconstruction data: GAN-generated renderings and images captured by the THETA V camera positioned behind the target wall.

In the pre-processing phase, the panoramic image is generated by a computer equipped with a GPU and then uploaded to the HoloLens. Numerous interior renovation renderings and their paired sketches were collected. Using edge extraction processing technology in Photoshop and a set of custom scripts, sketches can be efficiently derived from these renderings. By training the GANs network to transition from sketches to renderings, homeowner sketches can be converted into renderings using this approach.

The captured panoramic video is streamed to HoloLens over a Wi-Fi connection. The transfer protocol is constructed based on the THETA developers' API 2.1 (THETA DEVELOPERS BETA) [41]. Within this method, the `HttpRequest` class is used to set the POST request. Next, the motion-JPEG HTTP stream from the camera is parsed to generate each frame from the streaming data. OpenCV-based techniques are then used to transform these panoramic images into a hexahedral map, as shown in Figure 6.

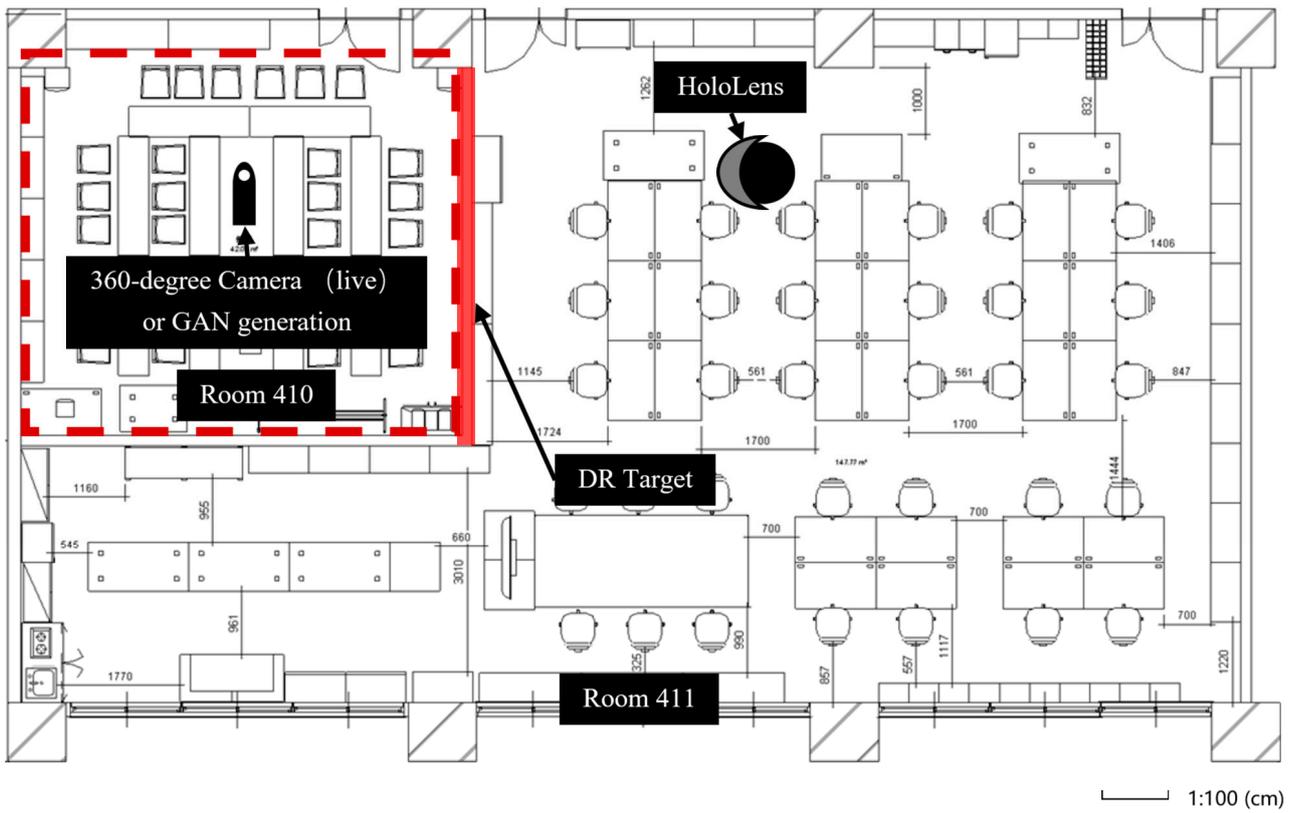


Figure 4. Floor plan of the evaluation test.

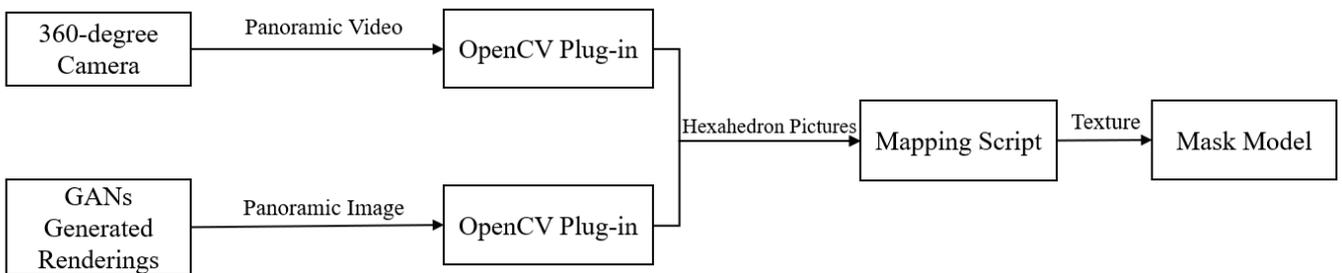


Figure 5. System data flow.

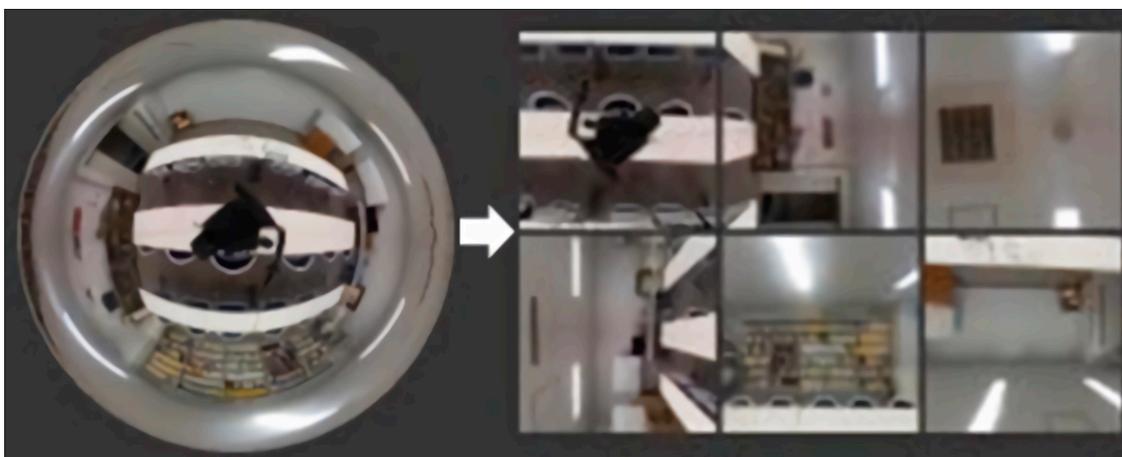


Figure 6. Panorama image conversion.

Each frame is then mapped to a mask model as a texture and awaits processing. Digitizing the restoration plan falls under the MR segment, which provides both the restoration plan model and the occlusion. Once both parts are complete, the combined results are projected onto a translucent screen in HoloLens. This allows individuals to “see through” the DR target wall. The real-world view of an individual experiencing the DR results can be seen in Figure 7. In Figure 7, the portion enclosed by the red frame represents the target wall. A portion of the target wall closer to the entrance is not selected for removal, primarily because of its structural importance.

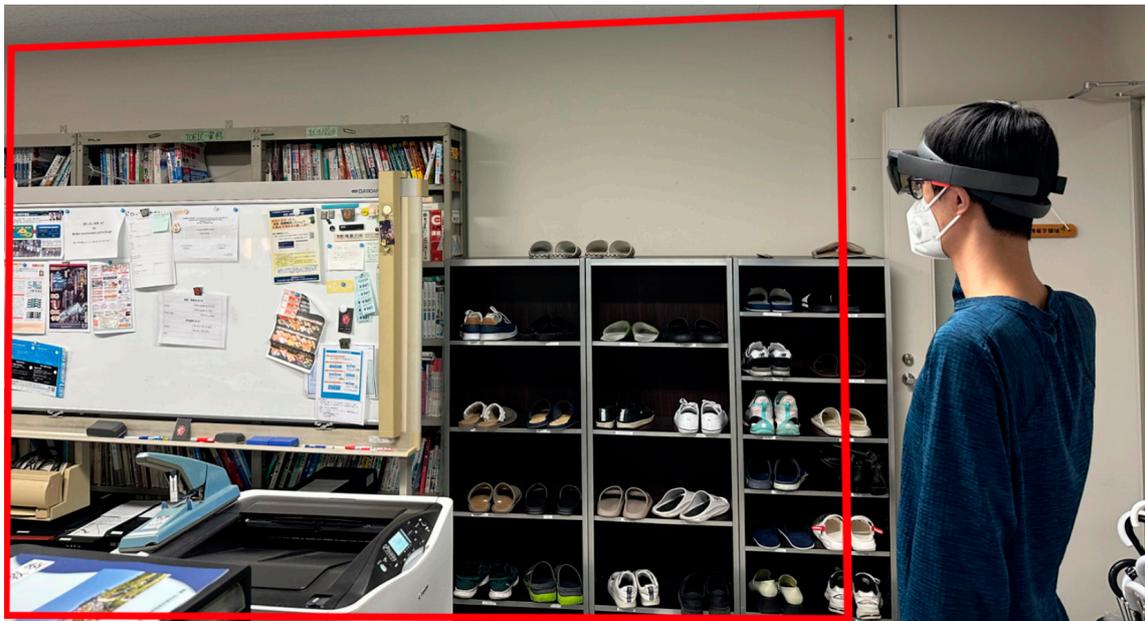


Figure 7. Reality view of using HMD (red frame: DR target wall).

4.2. Renovation Target Room

The BIM model, shown in Figure 8, can provide stakeholders with professional information and simplify the modification of the renovation plan. At the same time, to enable the MR system to operate on some older HMDs that lack real-time scanning and environmental understanding cameras, the BIM model was used to generate mesh and collision created using Autodesk Revit version 2016 software. When the user wears the HMD for observation, the MR system needs to determine the user’s position relative to the room. The BIM model facilitates this by assisting with registration, and when coupled with the collision data, occlusion can be calculated.

4.3. Operating Environment

To verify the feasibility of the proposal, a prototype application was developed based on the methodology proposed in Section 3.1. The BIM modeling, the development of the MR system and the creation of the GAN training materials were performed on a desktop PC (A) (specifications in Table 1). The GAN operations, which required CUDA and PyTorch and the installation of the dominant library, ran on the Ubuntu system of PC (B) (specifications in Table 2). The GPU configuration requirements for PC (B) are relatively high. A GPU with lower specifications may adversely affect the computational speed of GANs, resulting in suboptimal performance in computational and processing tasks. The system environment setup for PC (B) is shown in Table 3.

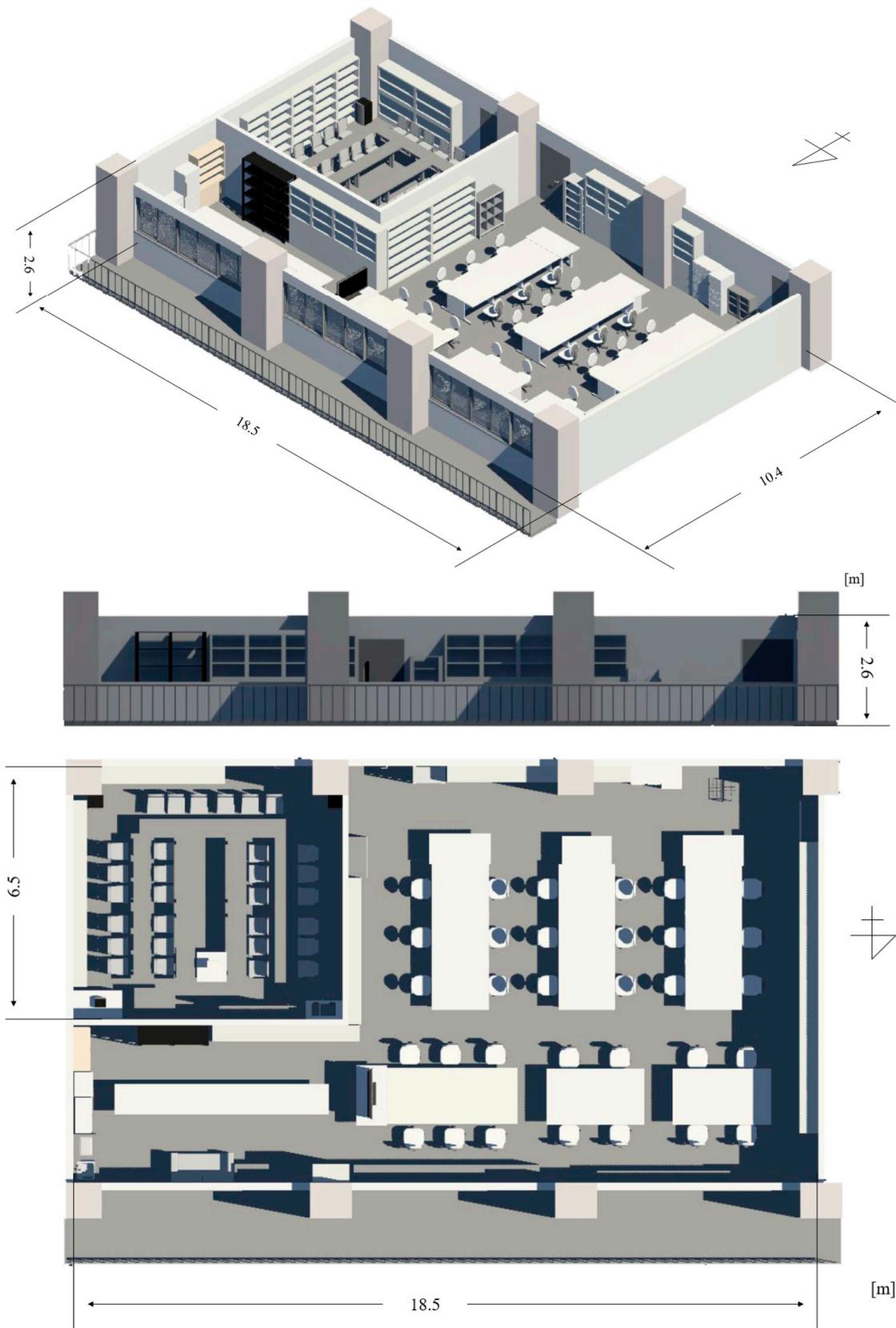


Figure 8. BIM model of room 410 and room 411.

Table 1. Desktop PC (A) specifications.

ITEM	PERFORMANCE
OS	Windows 10 Enterprise 64-bit
CPU	Intel Core i5 7500 @ 3.40 GHz
RAM	16.0 GB Dual-Channel DDR4 @ 2400 MHz
MOTHERBOARD	H270-PLUS
GPU	NVIDIA GeForce GTX 1060 6 G

Table 2. Desktop PC (B) specifications.

ITEM	PERFORMANCE
OS	Ubuntu 16.04
CPU	Intel Core i7 7700K @ 4.2 GHz
RAM	16.0 GB Dual-Channel DDR4 @ 2400 MHz
MOTHERBOARD	Z270-K
GPU	NVIDIA GeForce GTX 2070s

Table 3. System environment of PC (B).

PACKAGE	VERSION
CUDA TOOLKIT VERSION	CUDA 10.0.130_410.48
LINUX X86_64 DRIVER VERSION	NVIDIA driver 410.78
CUDNN	7.4.2.24
ANACONDA3	2021.05
PYTORCH	1.2.0
PYTORCH TORCHVISION	0.4.0

4.4. Simulation of Renovation Proposal

To validate the usability and effectiveness of the MR system proposed in this study, a simulation of the early stage of a renovation project was performed using the MR system for rooms 410 and 411. As shown in Figure 4, the area within the red dotted box is room 410, and the wall between rooms 410 and 411 was removed using DR technology, connecting the two rooms. At the same time, a hand-drawn sketch was created by the participants as the renovation target for room 410. To facilitate the subsequent quality verification tests, the current configuration of room 410 (as shown in Figure 9) serves as the renovation target, prompting participants to create hand-drawn interpretations (as shown in Figure 10). These sketches are then entered into PC (B). After processing through GANs, the generated renderings are loaded into the HMD, allowing participants to observe the simulated renovation results. Additionally, the current state of room 410 is captured by a 360-degree camera and wirelessly transmitted to the HMD in real time, allowing participants to manually switch between the two outcomes. For a more visual explanation of the experimental setup, the system scenarios are shown in Figure 11.

4.5. Verification of GANs Generation Quality

In the proposed simulation, 3044 pairs of panorama images have been inputted as the training dataset. In total, 2236 indoor panorama images were gathered from the HDRI-Haven indoor dataset [42] and Shutterstock dataset [43], as well as 52 indoor panorama images captured by a 360-degree camera. Out of these, 52 images captured by the camera were subjected to data augmentation including flipping, scaling, and cropping. Flipping an image horizontally or vertically increases the variability of the data set by highlighting the different orientations of similar furniture placements.



Figure 9. The current furnishings of room 410.

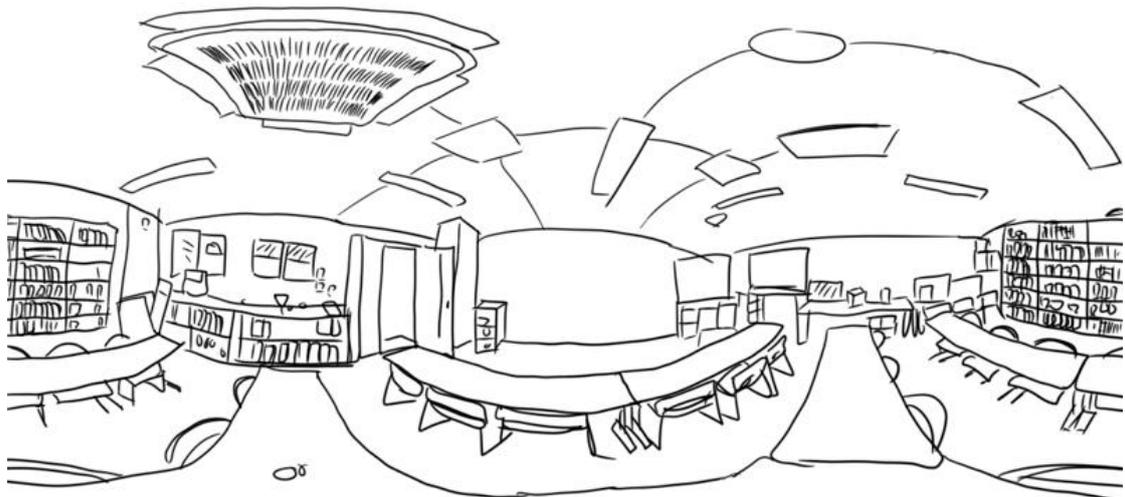


Figure 10. Hand drawn current furnishings of room 410.

Scaling involves resizing images to simulate changes in object size and distance, such as different sizes of vases, paintings, or furniture such as lamps. This is critical for scenarios where objects may appear at different distances or sizes in the real world. It prevents the model from being overly sensitive to certain object sizes seen during training. By cropping images, the model learns to focus on different parts of objects or scenes. This helps improve the model's object recognition regardless of an object's position within the frame, ensuring robustness to changes in object placement. In essence, these enhancements aim to simulate real-world variations that the model may encounter during deployment. By exposing the model to different perspectives, scales, and object arrangements, it becomes more adaptive and capable of making accurate predictions in different scenarios.

After the data augmentation, the number of available images increased to 416. The paired sketches were also generated using a custom Photoshop script. Some of these are shown in Figure 12. To improve the quality of the renderings generated from the hand sketches, we selected 56 photos with reasonable layout and non-repetitive styles from 2236 indoor panorama images and sketched them using the hand drawing method. After data enhancement, 448 paired training data sets were obtained. Some of them are shown in Figure 13.

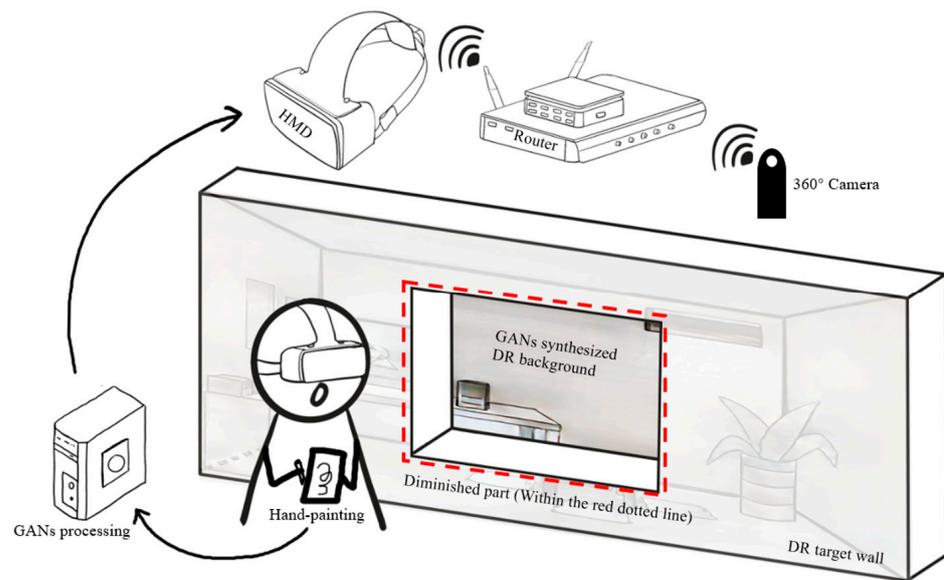


Figure 11. System scenarios.

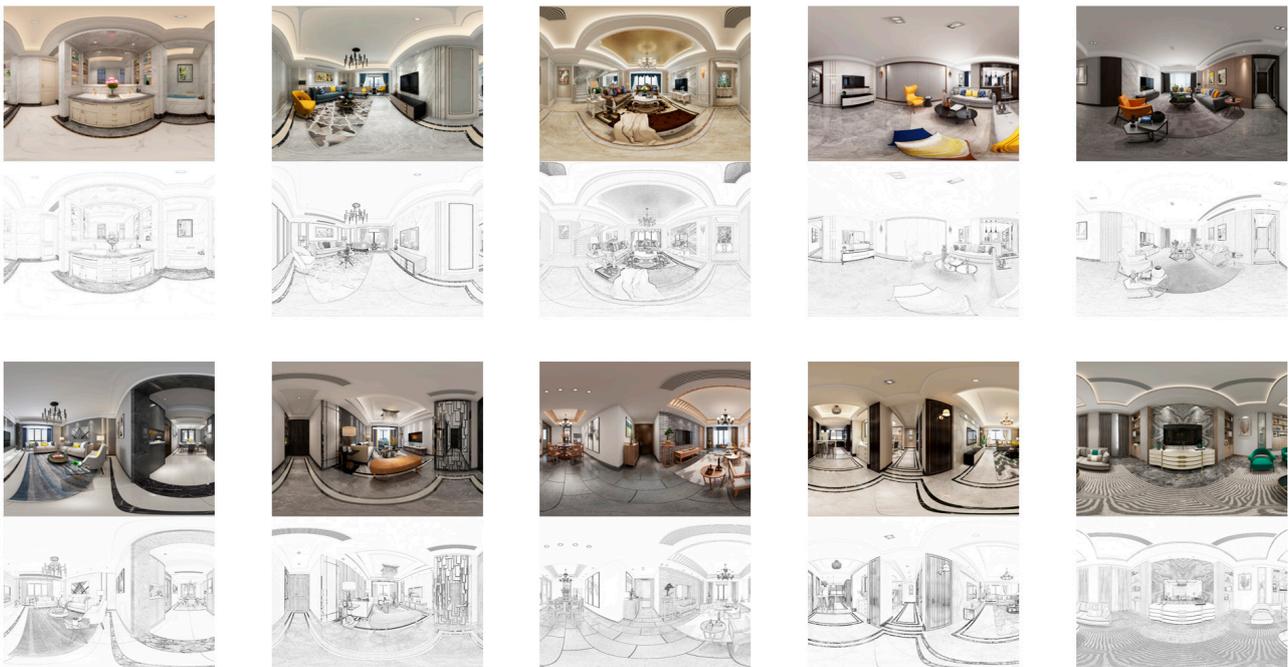


Figure 12. Part of the pairs' training data.

In order to quantitatively assess the quality of the panoramic images generated by GANs, this study intends to compare the results generated by GANs with the correct results. The correct results correspond to the existing layout of room 410, as discussed in Section 4.4 of the thesis. In addition, nine interior panoramic images that were not part of the training dataset were selected for verification. The evaluation will utilize the structural similarity index measure (SSIM) [44] and the peak signal-to-noise ratio (PSNR) [45]. Both are related and complementary full-reference image quality assessment (FR-IQA) methods and are widely recognized as fundamental and universal image evaluation indices [46]. These metrics provide quantitative insights into the quality of the generated panoramic images. The SSIM evaluates the structural similarity between the test and reference images, with values ranging from -1 (no structural similarity) to 1 (perfect similarity). PSNR, on

the other hand, measures the relationship between signal and noise distortion, quantifying reconstruction or compression quality in decibels. Higher PSNR values indicate less distortion and higher image quality, while lower values indicate potential degradation. These metrics serve as robust measures to objectively evaluate the fidelity and similarity of the generated images compared to the correct results.

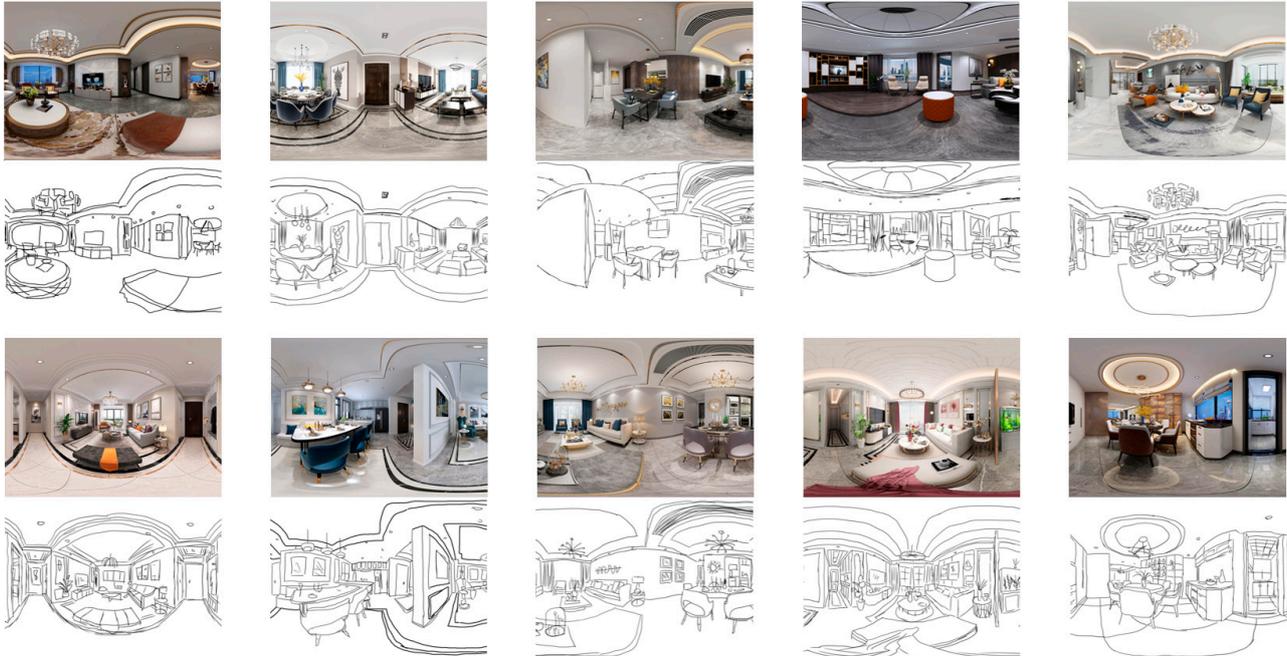


Figure 13. Part of the augmented images with paired hand-drawn images.

5. Results

5.1. Numerical Results

PSNR measures the relationship between the maximum potential value of a signal and the noise distortion that affects its representation quality. A higher PSNR value indicates better reconstruction (or compression) quality due to less distortion or noise. Conversely, a lower PSNR value indicates a decrease in quality. PSNR is expressed in decibels (dB). Based on general guidelines, we can provide the following rough guidelines [47,48]:

- Under 15 dB: Unacceptable.
- 15–25 dB: The quality might be considered poor, with possible noticeable distortions or artifacts.
- 25–30 dB: Medium quality. Acceptable for some applications but might not be for high-quality needs.
- 30–35 dB: Good quality; acceptable for most applications.
- 35–40 dB: Very good quality.
- 40 dB and above: Excellent quality; almost indistinguishable differences from the original image.

The SSIM acts as an index to anticipate the perceived quality of an image compared to that of an original (reference) image. With a value range from -1 to 1 , an SSIM of 1 indicates that the test image matches the reference image perfectly. As the SSIM value approaches 1 , it emphasizes the increasing structural congruence between the two images being compared. The following include the common interpretations of SSIM values [46,47]:

- $SSIM = 1$: The test image is identical to the reference. Perfect structural similarity.
- $0.8 < SSIM < 1$: High similarity between the two images.
- $0.5 < SSIM \leq 0.8$: Moderate similarity. There might be some noticeable distortions, but the overall structure remains somewhat consistent with the reference image.

- $0 < \text{SSIM} \leq 0.5$: Low similarity. Significant structural differences or distortions are present.
- $\text{SSIM} = 0$: No structural information is shared between the two images.

Numerical results to verify the quality of the simulated renovation plans are shown in Table 4. The PSNR and SSIM values show that most of the images generated by GANs are of moderate quality and have a reasonable similarity to the correct result images. The image with serial number 1 was a hand-drawn sketch of room 410, while the rest were processed by scripts and manually corrected hand-drawn sketches. All results hover around 20 dB, which is within an acceptable range. As an intuitive reference, Jing et al. mentioned that for pairs of stego images with embedded secret patterns, the SSIM values ranged from 30 to 52, while the PSNR values ranged from 0.9 to 0.99 [47]. Visually, these image pairs appear identical. This indicates that the GAN methods have been effectively trained and can produce images that are both qualitatively good and structurally consistent with the expected results. While the quality of restoration renderings produced by GANs may not match that of commercial renderings that require modeling and rendering processes, the low cost of a single hand-drawn sketch and its significantly faster generation speed compared to that of traditional renderings offer tremendous advantages in the early stages of the restoration process.

Table 4. PSNR and SSIM scores for 10 image pairs.

IMAGE NUMBER	1	2	3	4	5	6	7	8	9	10
PSNR (dB)	15.12	22.42	21.52	20.95	21.62	23.41	24.29	20.12	18.57	21.49
SSIM	0.7332	0.9265	0.8546	0.8956	0.8103	0.9187	0.9036	0.8452	0.8419	0.8516

The results of GANs generated from training datasets with different numbers of images are shown in Figure 14. At about 50 images, it was difficult to distinguish the types of furniture; only the rough layout of the room was visible. At about 200 images, some of the larger pieces of furniture became distinguishable, but the finer details appeared blurred. After more than 3000 images, the edges of various objects went from blurry to relatively clear. Most of the furniture and decorations could be identified, although some densely arranged items remained difficult to distinguish. The more images used for training, the better the resulting image quality. In addition, as the level of detail in the hand-drawn sketches increases, the quality of the resulting image also improves.

5.2. Visualization Results

The original state of the target wall is shown in Figure 7, and Figure 15 presents the experimental results using the proposed MR system, showing 2D segmentation maps. Within these maps, the gray areas represent real-world objects, while the blue areas represent the background scene. The five planes of the mask model, as well as the walls, ceiling, and floor of room 410 are seamlessly aligned in their real-world positions. As a result, users can accurately perceive the spatial sense of virtual room 410 as they move about. When viewed through the HMD, the DR results give the user the impression that the target wall has disappeared. The two rooms are visually connected, and as the user moves around, the occluded areas and the DR background (room 410) dynamically adjust to the user's perspective. As an intuitive explanation of occlusion, the result without occlusion at the similar position is shown in Figure 3. Users experience almost the same sensation as "taking down a wall" during the renovation process. In addition, activities in room 410 behind the target wall are instantly reflected in virtual room 411. The alignment function remains stable, ensuring that the virtual object is displayed without drift or disappearance.



Figure 14. Results generated by GANs with different training datasets: 50 images (**top**), 200 images (**middle**), and 3044 images (**bottom**).

However, the occlusion is not always accurately represented. An analysis of the segmentation map shows that parts of the occlusion edge are misrepresented, and certain virtual elements that should not be present in the DR results appear sporadically. This inaccuracy is due to the fact that the occlusion generated via real-time scanning is imprecise. The primary culprit is the lack of scanning accuracy. In addition, given the performance limitations of the HoloLens, there are inherent systematic errors in the results. Beyond the uneven edge, the whiteboard that should be in front of the virtual room is obscured by the virtual components. Even though the range for real-time occlusion generation is set between the whiteboard and the target wall, the proximity between the two is so minimal that it occasionally leads to errors.

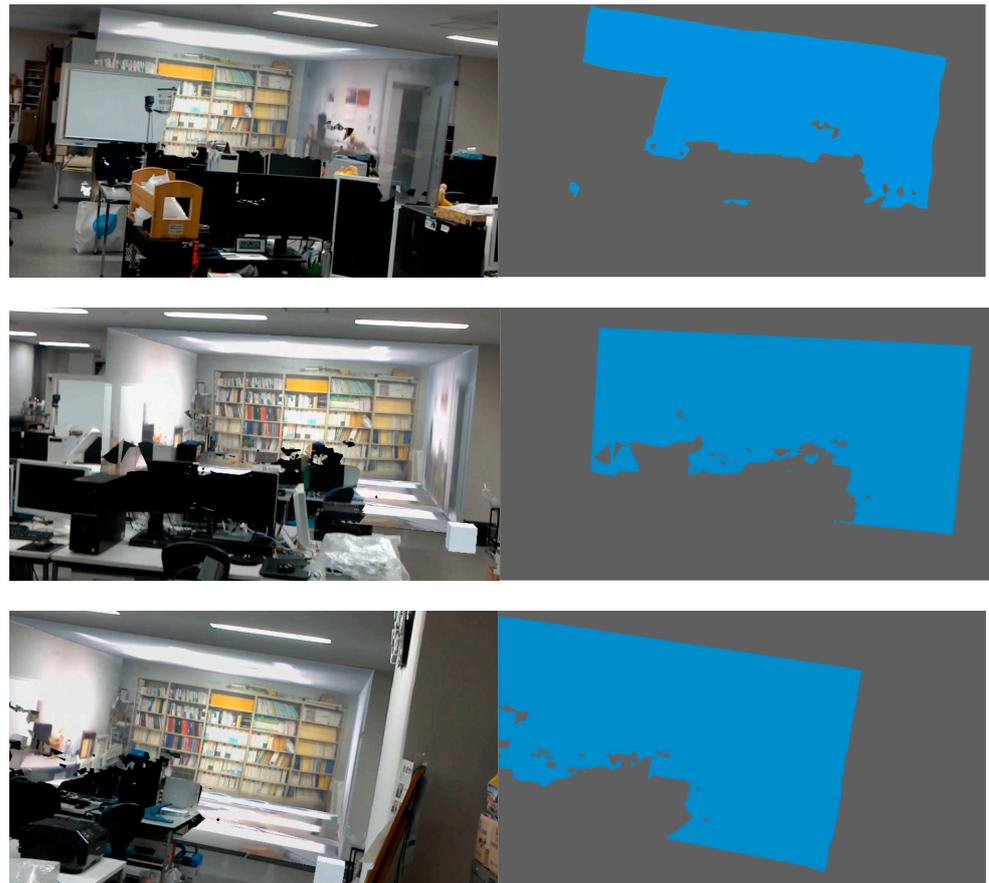


Figure 15. DR result and its segmentation map in position A (above), B (middle), and C (bottom).

In addition, the DR result is dynamic, but the refresh rate remains low. When testing this DR system on a PC platform, the process runs smoothly in real time. Latency challenges are common with streaming media transmissions. One cause of latency is that the sampling rate of the sensors inside the HoloLens host is much higher than the frequency of regular Wi-Fi transmission, resulting in the HoloLens server buffer constantly holding a large amount of image data. As a result, the longer the socket connection time, the more data accumulate. This means that the longer the program runs, the higher the latency. However, the image transmission speed is limited by the bandwidth of HoloLens, resulting in reduced frame rates.

6. Discussion

Using only algorithms to infer the background makes it difficult to generate content for complex backgrounds, especially when there are no repeating patterns. Therefore, this research introduced the approach of using 360-degree camera data to reconstruct the background. By using a 360-degree camera, the entire DR background can be quickly scanned. Wi-Fi data transmission expands the range of deployment scenarios, ensuring that even in rooms with unique layouts, distance constraints are minimized and real-time DR backgrounds can still be provided. This method is effective not only for complex backgrounds, but also for large diminished targets. Especially for large DR targets, there is no need for multiple cameras or multiple shots of the DR background from different angles. Panoramic images can more conveniently reconstruct the DR background. However, if the DR target is too large, the quality of the result may be compromised due to insufficient image resolution.

While GAN-generated renovation renderings as backgrounds may not be able to compete with commercial renderings, the speed and low cost of hand-drawn sketches

offer significant advantages in the early stages of the renovation process. Furthermore, the method of creating renderings from a simple sketch not only improves the efficiency of communication between designers and clients, but also increases the client's enthusiasm for participation. Even younger people can potentially provide valuable design insights. In addition, using panoramic cameras to capture images or collecting royalty-free panoramas for training data sets is time-consuming and often results in inconsistent quality. Due to the lack of extraterrestrial data, methods for collecting training information on artificially constructed lunar terrains within a VR environment have already been presented [49]. If the BIM model of a house is available, introducing it into a VR environment to collect panoramic images for training data is a viable approach. Compared to relying solely on basic data augmentation techniques, this method yields significantly higher quality training data. In addition, the models trained using this approach more closely match the style of the house before the renovation.

At times, the MR system experiences noticeable delays and latencies, especially during extended operation. This problem is common in streaming scenarios, and the most direct solution is to reduce the amount of data being transmitted. However, in this research, compromising video quality is not an option, as a blurry virtual room would negatively impact the overall DR experience. As a temporary solution, the program was manually adjusted to update the DR background. A button has been added for the user to refresh the DR background. Additionally, reducing the sampling rate to reduce the size of the data stream or upgrading to a higher bandwidth HMD may also resolve this issue. After this implementation was completed, Microsoft Corporation released HoloLens Gen 2, which boasts improvements in hardware performance and software algorithms. Specific verification may be conducted in future work.

7. Conclusions

This paper presents an innovative MR system that revolutionizes spatial visualization by enabling real-time wall reduction, seamlessly revealing live background scenes. The novelty of the system lies in its intricate occlusion capabilities that seamlessly merge virtual and real environments, thereby enriching the MR experience with enhanced realism, and the method of integrating GANs can generate MR environments according to user requirements. In particular, this system enables designers to quickly modify renovation plans and gain vivid insights into the expected results.

Currently, the system operates within an LAN, which is functional, but highlights a future direction: the development of a relay server for WAN operation, which will enable remote room reconstruction, a feature not available in current MR systems. In addition, the meticulous manual setting of segment parameters during reconstruction ensures precise alignment between wall edges and mask model surfaces, underscoring the system's attention to detail and accuracy.

Furthermore, the utilization of GANs for image generation marks a notable advancement. As shown in Figure 14, there is a noticeable improvement in image quality as more training images are included. Even though thousands of images were used, the results still show a significant difference when compared to photos taken by cameras or rendered images.

In future work, it is important to consider more efficient methods for generating training sets. Manually created sketches are time consuming and costly, but yield the best results; scripted sketches are fast to produce, but the models trained from them are not entirely adept at processing sketches drawn by non-professionals.

It is important to note that GAN algorithms do not distinguish between panoramic images and standard photos in their learning and computation processes. However, the modification of renovation plans on panoramic images poses specific challenges for those without expertise. The method used in this study involves converting images into cube maps for modifications. However, this technique is primarily effective for small image adjustments. In addition, dealing with elements located at the edges of the image poses

complications due to distortion problems. This is a significant challenge that needs to be addressed in future research efforts.

Author Contributions: Conceptualization, Y.Z. and T.F.; methodology, Y.Z. and T.F.; software, Y.Z.; resources, T.F. and N.Y.; writing—original draft preparation, Y.Z.; writing—review and editing, T.F. and N.Y.; visualization, Y.Z. and T.F.; funding acquisition, T.F. All authors have read and agreed to the published version of the manuscript.

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

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author (accurately indicate status).

Acknowledgments: During the preparation of this work, the authors used ChatGPT to improve the readability and language of the text. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Klepeis, N.E.; Nelson, W.C.; Ott, W.R.; Robinson, J.P.; Tsang, A.M.; Switzer, P.; Behar, J.V.; Hern, S.C.; Engelmann, W.H. The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. *J. Expo. Sci. Environ. Epidemiol.* **2001**, *11*, 231–252. [[CrossRef](#)]
2. Kovacic, I.; Summer, M.; Achammer, C. Strategies of building stock renovation for ageing society. *J. Clean. Prod.* **2015**, *88*, 349–357. [[CrossRef](#)]
3. Zhu, Y.; Fukuda, T.; Yabuki, N. Slam-based MR with animated CFD for building design simulation. In Proceedings of the 23rd International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Tsinghua University, School of Architecture, Beijing, China, 17–19 May 2018; pp. 17–19. [[CrossRef](#)]
4. Milgram, P.; Kishino, F. A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* **1994**, *77*, 1321–1329.
5. Nakajima, Y.; Mori, S.; Saito, H. Semantic object selection and detection for diminished reality based on slam with viewpoint class. In Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Nantes, France, 9–13 October 2017; pp. 338–343. [[CrossRef](#)]
6. Herling, J.; Broll, W. Pixmix: A real-time approach to high-quality diminished reality. In Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Atlanta, GA, USA, 5–8 November 2012; pp. 141–150. [[CrossRef](#)]
7. Kawai, N.; Sato, T.; Yokoya, N. Diminished reality considering background structures. In Proceedings of the 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Adelaide, SA, Australia, 1–4 October 2013; pp. 259–260. [[CrossRef](#)]
8. Sasanuma, H.; Manabe, Y.; Yata, N. Diminishing real objects and adding virtual objects using a RGB-D camera. In Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), Merida, Mexico, 19–23 September 2016; pp. 117–120. [[CrossRef](#)]
9. Cosco, F.I.; Garre, C.; Bruno, F.; Muzzupappa, M.; Otaduy, M.A. Augmented touch without visual obstruction. In Proceedings of the 2009 8th IEEE International Symposium on Mixed and Augmented Reality, Orlando, FL, USA, 19–22 October 2009; pp. 99–102. [[CrossRef](#)]
10. Enomoto, A.; Saito, H. Diminished reality using multiple handheld cameras. In Proceedings of the 2007 First ACM/IEEE International Conference on Distributed Smart Cameras, Vienna, Austria, 25–28 September 2007; Volume 7, pp. 130–135. [[CrossRef](#)]
11. Meerits, S.; Saito, H. Real-time diminished reality for dynamic scenes. In Proceedings of the 2015 IEEE International Symposium on Mixed and Augmented Reality Workshops, Fukuoka, Japan, 29 September–3 October 2015; pp. 53–59. [[CrossRef](#)]
12. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 784–794. [[CrossRef](#)]
13. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434. [[CrossRef](#)]
14. Zhu, J.; Shen, Y.; Zhao, D.; Zhou, B. In-domain gan inversion for real image editing. In *European Conference on Computer Vision*; Springer International Publishing: Cham, Switzerland, 2020; pp. 592–608. [[CrossRef](#)]
15. Yang, C.; Shen, Y.; Zhou, B. Semantic hierarchy emerges in deep generative representations for scene synthesis. *Int. J. Comput. Vis.* **2021**, *129*, 1451–1466. [[CrossRef](#)]

16. Zhu, Y.; Fukuda, T.; Yabuki, N. Synthesizing 360-degree live streaming for an erased background to study renovation using mixed reality. In Proceedings of the 24th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Victoria University of Wellington, Faculty of Architecture & Design, Wellington, New Zealand, 15–18 April 2019; pp. 71–80. [[CrossRef](#)]
17. Chan, D.W.; Kumaraswamy, M.M. A comparative study of causes of time overruns in Hong Kong construction projects. *Int. J. Proj. Manag.* **1997**, *15*, 55–63. [[CrossRef](#)]
18. Watson, A. Digital buildings—Challenges and opportunities. *Adv. Eng. Inform.* **2011**, *25*, 573–581. [[CrossRef](#)]
19. Ozdemir, M.A. Virtual reality (VR) and augmented reality (AR) technologies for accessibility and marketing in the tourism industry. In *ICT Tools and Applications for Accessible Tourism*; IGI Global: Hershey, PA, USA, 2021; pp. 277–301. [[CrossRef](#)]
20. Woksepp, S.; Olofsson, T. Credibility and applicability of virtual reality models in design and construction. *Adv. Eng. Inform.* **2008**, *22*, 520–528. [[CrossRef](#)]
21. Lee, J.; Miri, M.; Newberry, M. Immersive Virtual Reality, Tool for Accessible Design: Perceived Usability in an Interior Design Studio Setting. *J. Inter. Des.* **2023**, *48*, 242–258. [[CrossRef](#)]
22. Chandler, T.; Cordeil, M.; Czauderna, T.; Dwyer, T.; Glowacki, J.; Goncu, C.; Klapperstueck, M.; Klein, K.; Marriott, K.; Schreiber, E.; et al. Immersive analytics. In Proceedings of the 2015 Big Data Visual Analytics (BDVA), Hobart, TAS, Australia, 22–25 September 2015; pp. 1–8. [[CrossRef](#)]
23. Ohno, N.; Kageyama, A. Scientific visualization of geophysical simulation data by the CAVE VR system with volume rendering. *Phys. Earth Planet. Inter.* **2007**, *163*, 305–311. [[CrossRef](#)]
24. Zhang, W.; Han, B.; Hui, P. Jaguar: Low latency mobile augmented reality with flexible tracking. In Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Republic of Korea, 22–26 October 2018; pp. 355–363.
25. Kan, P.; Kurtic, A.; Radwan, M.; Rodriguez, J.M.L. Automatic interior Design in Augmented Reality Based on hierarchical tree of procedural rules. *Electronics* **2021**, *10*, 245. [[CrossRef](#)]
26. Wang, X.; Love, P.E.; Kim, M.J.; Park, C.S.; Sing, C.P.; Hou, L. A conceptual framework for integrating building information modeling with augmented reality. *Autom. Constr.* **2013**, *34*, 37–44. [[CrossRef](#)]
27. Mori, S.; Ikeda, S.; Saito, H. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Trans. Comput. Vis. Appl.* **2017**, *9*, 1–14. [[CrossRef](#)]
28. Eskandari, R.; Motamedi, A. Diminished reality in architectural and environmental design: Literature review of techniques, applications, and challenges. In *ISARC: Proceedings of the International Symposium on Automation and Robotics in Construction, Dubai, United Arab Emirates, 2–4 November 2021*; Dubai, UAE, IAARC Publications 2021; Volume 38, pp. 995–1001. [[CrossRef](#)]
29. Kido, D.; Fukuda, T.; Yabuki, N. Diminished reality system with real-time object detection using deep learning for onsite landscape simulation during redevelopment. *Environ. Model. Softw.* **2020**, *131*, 104759. [[CrossRef](#)]
30. Siltanen, S. Diminished reality for augmented reality interior design. *Vis. Comput.* **2017**, *33*, 193–208. [[CrossRef](#)]
31. Liu, M.Y.; Huang, X.; Yu, J.; Wang, T.C.; Mallya, A. Generative adversarial networks for image and video synthesis: Algorithms and applications. *Proc. IEEE* **2021**, *109*, 839–862. [[CrossRef](#)]
32. Yun, K.; Lu, T.; Chow, E. Occluded object reconstruction for first responders with augmented reality glasses using conditional generative adversarial networks. In *Pattern Recognition and Tracking XXIX*; SPIE: Philadelphia, PA, USA, 2018; Volume 10649, pp. 225–231. [[CrossRef](#)]
33. Teo, T.; Lawrence, L.; Lee, G.A.; Billingham, M.; Adcock, M. Mixed reality remote collaboration combining 360 video and 3d reconstruction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Scotland, UK, 4–9 May 2019; pp. 1–14. [[CrossRef](#)]
34. Li, Q.; Takeuchi, O.; Shishido, H.; Kameda, Y.; Kim, H.; Kitahara, I. Generative image quality improvement in omnidirectional free-viewpoint images and assessments. *IEEE Trans. Image Electron. Vis. Comput.* **2022**, *10*, 107–119. [[CrossRef](#)]
35. Lin, C.H.; Chang, C.C.; Chen, Y.S.; Juan, D.C.; Wei, W.; Chen, H.T. Coco-GAN: Generation by parts via conditional coordinating. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 4512–4521. [[CrossRef](#)]
36. Herling, J.; Broll, W. Advanced self-contained object removal for realizing real-time diminished reality in unconstrained environments. In Proceedings of the 2010 IEEE International Symposium on Mixed and Augmented Reality, Seoul, Republic of Korea, 13–16 October 2010; pp. 207–212. [[CrossRef](#)]
37. Inoue, K.; Fukuda, T.; Cao, R.; Yabuki, N. Tracking Robustness and Green View Index Estimation of Augmented and Diminished Reality for Environmental Design. In Proceedings of the CAADRIA, Beijing, China, 17–19 May 2018; pp. 339–348. [[CrossRef](#)]
38. Meerits, S.; Saito, H. Visualization of dynamic hidden areas by real-time 3d structure acquisition using RGB-D camera. In Proceedings of the 3D Systems and Applications Conference, Marrakech, Morocco, 17–20 November 2015. [[CrossRef](#)]
39. Bourke, P. Converting to/from Cubemaps. 2020. Available online: <https://paulbourke.net/panorama/cubemaps/> (accessed on 5 October 2023).
40. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134. [[CrossRef](#)]
41. THETA Developers' API. Available online: <https://github.com/ricohapi/theta-api-specs> (accessed on 5 October 2023).
42. HDRIHAVEN. HDRIHaven. Available online: <https://polyhaven.com/hdris/indoor> (accessed on 5 October 2023).

43. Shutterstock. Shutterstock. Available online: <https://www.shutterstock.com/ja/search/indoor-panorama> (accessed on 5 October 2023).
44. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)]
45. Huynh-Thu, Q.; Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electron. Lett.* **2008**, *44*, 800–801. [[CrossRef](#)]
46. Horé, A.; Ziou, D. Is there a relationship between peak-signal-to-noise ratio and structural similarity index measure? *IET Image Process.* **2013**, *7*, 12–24. [[CrossRef](#)]
47. Jing, J.; Deng, X.; Xu, M.; Wang, J.; Guan, Z. HiNet: Deep image hiding by invertible network. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 4733–4742. [[CrossRef](#)]
48. Cavigelli, L.; Hager, P.; Benini, L. CAS-CNN: A deep convolutional neural network for image compression artifact suppression. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 752–759. [[CrossRef](#)]
49. Franchi, V.; Ntagiou, E. Augmentation of a virtual reality environment using generative adversarial networks. In Proceedings of the 2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), Taichung, Taiwan, 15–17 November 2021; pp. 219–223. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.