Abstract. To assess an environmental design, augmented and diminished reality (AR/DR) have a potential to build a consensus more smoothly through the landscape simulation of new design visualization of the items to be assessed, such as the green view index. However, the current system is still considered to be impractical because it does not provide complete user experience. Thus, we aim to improve the robustness of the AR/DR system and to integrate the estimation of the green view index into the AR/DR system on a game engine. Further, we achieve an improved stable tracking by eliminating the outliers of the tracking reference points using the random sample consensus (RANSAC) method and by defining the tracking reference points over an extensive area of the AR/DR display. Additionally, two modules were implemented, among which one module is used to solve the occlusion problem while the other is used to estimate the green view index. The novel integrated AR/DR system with all modules was developed on the game engine. A mock design project was developed in an outdoor environment for simulation purposes, thereby verifying the applicability of the developed system.

Keywords. Environmental Design; Augmented Reality (AR); Diminished Reality (DR); Green View Index; Segmentation.

1. Introduction

Augmented reality (AR) involves displaying virtual objects to provide an enhanced experience of the physical world through virtual objects its application to exterior construction has gained popularity (Klinker et al., 2001). AR can facilitate the process of building a consensus on the basis of landscape design because it is capable to simulate real-scaled new structures in three-dimensional (3D) views. However, a drawback of AR is that it cannot be used to simulate new buildings or structures while the old structure is still present. To address this issue, diminished reality (DR) can be used to visually eliminate an existing object from a scene by
overlaying an appropriate background image on the area that is occupied by the object (Mori et al., 2017). While AR is often interpreted in a limited sense to represent only the visual effects that are overlaid on a screen, it can further augment the physical world with digital information. In a landscape, AR can be used to visualize the 3D virtual models of the new structures and to display the items to be assessed so that the stakeholders can easily build a consensus more smoothly.

Urban vegetation is a key element of landscape and urban design. It has been used to tackle the various architectural and urban problems, such as urban heat islands, biodiversity, and resident comfortability. The amount of urban vegetation can be quantified using various metrics that can be used by the environmental stakeholders to advocate for the addition of more vegetation. One metric for evaluation is the percentage of green space, which is suitable for assessing large areas, but a drawback of this metric is that it does not consider the experience of the people on the ground. Another metric is the green view index, which is defined as the ratio of the green area to total area in the field of view of a person, and it is an effective and efficient metric for assessing the visual effects of the green areas in increasing urban comfort (Yang et al., 2009).

The study focuses on a method for simultaneously simulating the building and vegetation designs and assessing the landscape by estimating the green view index. In a previous study, PhotoAR+DR2016 was developed to integrate the simulation by augmented and diminished reality (AR/DR) and the real-time estimation of the green view index (Fukuda et al., 2017). However, this system is impractical because it does not provide a remarkable user experience, such as the lack of the robustness of tracking, the occlusion expression, and the system usability.

Therefore, the objective of this research is to improve the robustness of the AR/DR system and to integrate the estimation of the green view index into the system on a game engine. Further, we achieved improved stable tracking by eliminating the outliers of the tracking reference points using the random sample consensus (RANSAC) method and by defining the tracking reference points across an extensive range of the image. Furthermore, we observe that the user experience is affected not only by the robustness of the system, but also by the perceived reality and usability of the AR/DR system. Therefore, the burden on the user is reduced by using game engine in this study. The AR/DR system was designed on a game engine, thereby, the amount of redundant code described by the user for parameter adjustment is reduced. Additionally, two modules were implemented. The first module solves the occlusion problem using the 3D model reconstructed by photogrammetry. This module enables the user to easily perceive depth. The second module estimates the green view index. This module was implemented based on a process that is different from that used in the previous system. Finally, a mock design project was conducted to validate the applicability of the developed system.
2. Improving the tracking of the AR / DR system

2.1. OUTLIER ELIMINATION

The virtual objects must be properly aligned with respect to the physical world while the camera is moved. However, stable tracking is difficult in an outdoor environment due to the inconstant illumination and many noises. Therefore, maintaining stable tracking over an extensive period of time can be a critical problem to improve the user experience of outdoor AR/DR system.

In our tracking methodology, camera motion is computed by solving the *perspective* *n*-points (*PnP*) problem. This is a problem of estimating the camera pose from *n* 3D-to-2D point correspondences. In the proposed method, the *n* 3D points are in advance defined as tracking reference points and their corresponding 2D points are traced by estimating the optical flow (Tomasi and Kanade, 1992). In the previous system, the camera pose was estimated so that the sum of the errors of all the tracking reference points would be minimized. Thus, the previous system was observed to be susceptible to the influence of the outliers. Additionally, it was difficult to accurately estimate the accurate camera pose for a long period of time because of the accumulated errors. Therefore, we achieved more stable tracking by eliminating the outliers in this study.

To detect the outliers, the RANSAC method was applied as depicted in Figure 1 (Fraundorfer and Scaramuzza, 2012). First, several (in this case, five) tracking reference points are randomly selected. Second, the camera pose was estimated by solving the PnP problem using the tracking reference points. Third, all of the tracking reference points were projected from 3D coordinates to 2D coordinates. Fourth, the errors (called the reprojection errors) between the projected and the original points were calculated. Fifth, the tracking reference points containing reprojection errors, which are observed to be greater than the threshold (in this case, five pixels), were classified as outliers.

Figure 2 depicts comparison between the previous tracking flow (the red wire) and the tracking flow that is proposed in this study (the blue wire). The yellow lines represented the optical flows of the tracking reference points. In the previous tracking flow of the red wire model, the camera pose estimation was observed to be inaccurate after 15 s of tracking. However, the camera pose estimation observed to be accurate for 45 s in the novel tracking flow of the blue wire model.

![Figure 1. The flow that is used by the outlier detection module.](image-url)
2.2. ARRANGEMENT OF THE TRACKING REFERENCE POINTS

In our proposed method, it is necessary to arbitrarily define various tracking reference points in advance. Therefore, the influence of the arrangement of the tracking reference points on the tracking stability was investigated. The portion of the image was occupied by the bounding box of the tracking reference points was used as an index to arrange the points. The error in camera pose estimation in the virtual world was measured while varying the portion that is occupied by the tracking reference points, which is observed to be in the range from 2% to 100%. This was performed on a 1024 × 576 pixel image. First, the tracking reference points were defined to reach an arbitrary portion that is occupied by the tracking reference point. Next, an error of the optical flow estimation was given artificially; in this situation, this value was 1 pixel or 2 pixels. Finally, the camera pose was estimated and the result was compared with that of the correct camera pose.

The results are depicted in Figure 3. It was confirmed that the larger the portion of that is occupied by the tracking reference points, the smaller is the average error in camera pose estimation. Therefore, it was concluded that, tracking can be more stable by defining the tracking reference points to maximize the portion that is occupied by the points. However, when the tracking reference points are defined across an extensive range of the image, it is difficult to keep estimating their optical flow because the points are likely to go out of the screen by moving the camera. Therefore, the restoration module of the tracking reference points whose optical flow could not be estimated was implemented to perform the tracking for a long
period of time.

Figure 3. The influence of the arrangement on tracking reference points on the tracking stability.

3. Integrating the module into the game engine

3.1. SYSTEM ENVIRONMENT

Considering the application of our system to various practical projects, it is important to minimize the burden on the user. The previous system was developed using Visual Studio (Visual C++, OpenGL, OpenCV), which is an integrated development environment. In order to render the 3D models more graphically and realistically, it is necessary to finely tune the rendering settings, such as the light source definition, the material settings, and so on. Every time the usage environment and time changes, the user should repeatedly adjust several parameters and verify the quality of the rendered images to improve the photometric consistency between the real world and rendered 3D models. However, to adjust the parameters, the user must define the redundant codes. For the application of our system to various practical projects, a large burden is placed on the user. Therefore, we developed this system using Unity (a game engine) and OpenCV for Unity (a computer vision plugin for Unity). In this system, the amount of code that is described by the user is drastically reduced because the rendering settings can be adjusted easily and efficiently using the developed graphical user interface (GUI).

In our proposed system, the 3D models for AR/DR are manufactured by reconstructing the photographs of the surrounding environments using photogrammetry software. If the 3D models are not reconstructed by photogrammetry, our system cannot be used. Therefore, the photographs of the surrounding environments were often retaken. In the developed system, Agisoft PhotoScan was used instead of OpenMVG, which was used in the previous system, because it can reconstruct 3D models more stably and accurately. This reduces the probability that photographs have to be retaken.
3.2. OCCLUSION PROBLEM

Occlusion greatly influences the depth perception of the user (i.e., the relationship between the physical and virtual world). This is one of the elements that must be accurately expressed to assess an environmental design. The occlusion problem is an AR/DR challenge of rendering the real objects in front of the 3D models.

In order to solve the occlusion problem, information about the depth of the surrounding environment is required. Portalés et al. (2010) introduced a low-cost outdoor mobile AR application with solving the occlusion problem using High-accuracy 3D photo-models. In our proposed system, depth information can be acquired from a reconstructed 3D model of the surrounding environment, which is referred to as an occlusion model. The virtual world consists of the 3D virtual model to be superimposed for AR and a part of 3D virtual model of the physical world. In this virtual world, it is possible to determine which part of the 3D virtual model for AR is hiding behind the occlusion model from the viewpoint. The occlusion problem is solved by not rendering the pixels of the 3D models that are hidden by the occlusion model, as depicted in Figure 4. This module enables the user to easily perceive the depth.

![Figure 4. Example of a solution to the occlusion problem.](image)

3.3. ESTIMATION OF THE GREEN VIEW INDEX

In the previous system, the green view index was automatically estimated based on the three filtering steps that include Gaussian, mean-shift, and hue, saturation, and value (HSV) filtering methods (Ding et al., 2016). However, this mean-shift filtering could not be used in the new development environment. Therefore, a novel flow using median filtering instead of Gaussian and mean-shift filtering was implemented as depicted in Figure 5. Further, logical operation and morphology operation were added behind HSV filtering. In the logical operation, the image applied to the median and HSV filtering and the original image applied to HSV filtering are used to obtain the per-pixel bitwise logical conjunction. This operation is defined as bitwise and filtering and it is used to reduce the noise such as the green area reflected on the window. In the morphology operation, the result image of
bitwise and filtering is used to remove small holes in the detected green areas. This operation is defined as closing filtering and it is used to improve the accuracy rate.

To compare the previous flow with the proposed flow, the green area was detected by both the flows. Further, the accuracy rate was calculated using 30 images. The resulting accuracy rate was observed to be 93.8 % for the previous flow and 95.7 % for the proposed flow. The results obtained by the proposed flow were observed to be similar to those that were obtained from the previous flow.

Figure 5. The flow that is used by the green area detection module.

4. Case Study

PhotoAR+DR2017 (photogrammetry-based augmented and diminished reality) was developed and integrated with all the modules that are described in chapters 2 and 3. PhotoAR+DR2017 is an AR/DR system for simultaneously simulating the building and vegetation designs and estimating the green view index. To verify the applicability of PhotoAR+DR2017, a mock design project was conducted in an outdoor environment. A laptop PC, GALLERIA GKF1060GF with Intel Core i7-7700HQ with a 2.80 GHz of CPU, GTX1060 with 6 GB of GPU, 8.0 GB of RAM, an operating system of Microsoft Windows 10 Professional (64 bit), and a web camera, LogiCool HD Pro Webcam C920r, was used. The aim of this project was to dismantle a two-storey building, which is the Welfare Hall on Poplar Avenue located at the Osaka University, Suita Campus, and to construct a new object boating of various plants around the structure. Further, the green view indices of the existing and new structures were compared. The arrangements of the current and the new structures are depicted in Figure 6.

The results by applying PhotoAR+DR2017 are depicted in Figure 7. In this case study, the average green view index was observed to increase from 17% to 30%. In PhotoAR+DR2017, the tracking was observed to be more stable for a longer duration. Additionally, it was easy to adjust the rendering settings, such as the light source definition, the material settings, and so on using this system.

The aim of PhotoAR+DR2017 is to create better environment for all the stakeholders by assessing the environmental design during the initial stages of the project, such as predesign, schematic design, and design development. The proposed system eases the assessment of environmental designs in terms of the building size, planting arrangement, glass transparency, position and direction of the lighting equipment, and so on.
Figure 6. Arrangement of the existing buildings and structures to be constructed.

Figure 7. Design simulation using PhotoAR+DR2017: (Top) AR+DR view and (Bottom) detected green areas of planned vegetation (yellow) and of existing vegetation (white).

5. Improvement PhotoAR+DR2017 by Deep Learning Segmentation

In PhotoAR+DR2017, it is difficult to detect small green areas, such as trunks, branches, and leaves. To accurately detect the such areas, it is necessary to manually adjust several kinds of parameters depending on a variety of factors such as weather, sunlight, the type of trees, and so on. In PhotoAR+DR2017, the parameters can be easily adjusted using a slide bar and the changes are immediately reflected in the view. However, manual parameter adjustment is always practiced in the current system framework. Therefore, the module in PhotoAR+DR2017 for estimating the green view index cannot be easily generalized.

Recently, image segmentation systems based on deep learning have attracted considerable research attention (Long et al., 2015; Ronneberger et al., 2015). These systems can be broadly generalized and can segment images into several classes. Therefore, using the classified pixels, it is expected that not only the green view index, but also other landscape indices, such as sky factor, can be calculated. Furthermore, image segmentation is expected to solve the occlusion problem for moving various objects such as pedestrians and vehicles, which can be recognized only in a real-time scenario. Further, to enhance the connection between digital
information and the physical world, it is necessary to recognize various elements in the environment.

Therefore, a pilot system model was constructed using SegNet, which is a deep learning segmentation technology (Badrinarayanan, 2015), as depicted in Figure 8. SegNet can classify each pixel of an urban street image into one of the twelve classes (building, sky, tree, pedestrian, etc.,) by inputting real-time images or video. In this model, a real-time video is captured on the laptop and sent to a server using a WiFi network. The AR/DR simulation and image segmentation are performed on the server in real-time and the output movie is transmitted to the laptop. Further, the series of flows can be processed in real time. While SegNet has not yet been integrated with PhotoAR+DR2017, but it is possible to segment the result of the landscape simulation image and real-time video.

6. Conclusions and Future Works
Here, we developed an AR/DR system (PhotoAR+DR2017) that can simultaneously simulate the building and vegetation designs and estimate the green view index to assess the landscape. The contributions of this research are as follows:

- To improve the robustness of the AR/DR system, a relatively stable tracking was achieved by eliminating the outliers by the RANSAC method and by defining the tracking reference points over an extensive area of the AR/DR display.
- By integrating the AR/DR system and estimation of the green view index on a game engine, it is easy to adjust the rendering settings such as the light source definition, the material settings, and so on.
- We have constructed a system model to calculate the green view index automatically using SegNet. This system model can be used on a laptop, with only basic specifications in a real-time scenario. While SegNet has not yet been integrated into PhotoAR+DR2017 in real time, it has been confirmed that it can be used to automatically calculate the green view index by inputting the simulated landscape image and a video.
For future studies, we note that further improvement of the tracking stability is necessary to allow the web camera to be moved over a long distance to assess a large and continuous landscape. In PhotoAR+DR2017, the occlusion problem is solved using a 3D model that was reconstructed by photogrammetry. However, the 3D models of the trees and moving objects may be observed to be different from the shapes of these objects, when used in practical applications. Therefore, it is necessary to recognize various objects in real time and solve the occlusion problem. This will lead to an enhanced connection between digital information and the physical world.

Acknowledgements
This research has been partly supported by the research grant of ARMO ARCHITECTS & ENGINEERS CO., LTD., and by JSPS KAKENHI Grant Number JP16K00707.

References