A Landscape Simulation Method with One-by-one Dynamic Occlusion Using Instance Segmentation in Mixed Reality

Image Generation Method Focusing on the Grounding for Occlusion

○Mizuki Nakabayashi＊1, Tomohiro Fukuda＊2 and Nobuyoshi Yabuki＊3

*1 Graduate Student, Div. of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University

*2 Assoc. Prof., Div. of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Ph.D.

*3 Prof., Div. of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, Osaka University, Ph.D.

Keywords: Landscape simulation; Instance segmentation; Dynamic occlusion handling; Mixed Reality; Deep learning

1. Introduction

It is essential to simulate landscape and to build a consensus in an environmental planning and design process. It is important for stakeholders to visualize the expected landscape and to discuss the project contents clearly. Mixed Reality (MR) is a technique to merge real and virtual environments. The three-dimensional (3D) model of a new structure can be superimposed on the real image of the construction site. It is possible to examine the change of the landscape directly on the real scale, and realistic landscape simulation can be expected [1].

One of the challenges of MR is occlusion. It is hard to render the relationship between real and virtual environments correctly. When physical objects that are in the foreground are rendered behind virtual objects, the AR output may confuse user due to the incorrect depth perception. Thus, it is crucial to handle occlusion appropriately in MR. A city simulation method using MR could handle occlusion using depth information from an occlusion model [2]. In this system, the occlusion problem is solved by the pre-defined occlusion model. However, if a physical object such as vegetation changes its shape over time, it can be difficult to appropriately handle occlusion by changing the occlusion model's shape. In addition, a method exists for occlusion processing by recognizing the relationship between the real and virtual objects to be studied using depth information [3], but it is difficult to perform a large-scale landscape simulation because the scope of application is limited in outdoor landscape simulation due to the limited range of depth information that can be obtained. Kido *et al.* [4] proposed an MR system by using semantic segmentation which is one of the image recognition methods using deep learning and produced a mask image for landscape dynamic occluded simulation. However, this MR system has a problem that the same kind of objects that are grouped into one or overlapped are recognized as only one object by semantic segmentation. Thus the system cannot simulate correctly at the place where there are many buildings and trees. In our previous research, Nakabayashi *et al.* [5] proposed an MR system using instance segmentation which is a technique of object detection plus semantic segmentation and enables to distinguish the boundaries between objects with the same label. The system enables to simulate in the parking lot that there are many cars. However, due to the processing speed of the system, it is only an image-based simulation.

This study aims to develop a landscape simulation method that contains one-by-one real-time dynamic occlusion and to consider the method of generating a correct mask image for the landscape simulation method.

2. Proposed Method

2.1. PROPOSED SYSTEM

Our proposed system puts into practice real-time landscape simulation with occlusion using instance segmentation. Figure 1 shows the conceptual diagram of our proposed system and Figure 2 shows the flowchart of our proposed method.

2.2 INSTANCE SEGMENTATION TECHNIQUE

We used Mask R-CNN [6] which is one of the instance segmentation techniques in our previous research [5]. It has a slow processing speed of 8.3 fps [7]. It was difficult to simulate

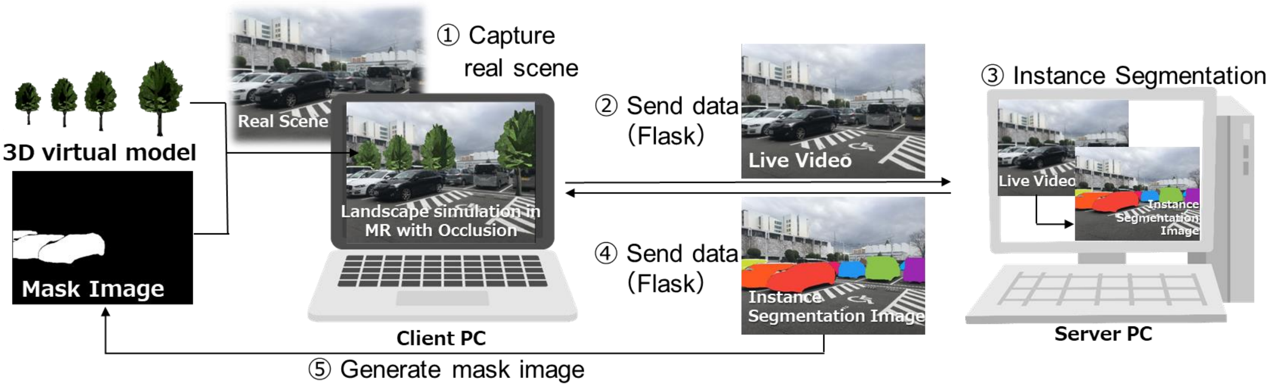


Figure 1. Conceptual Diagram of Our Proposed System

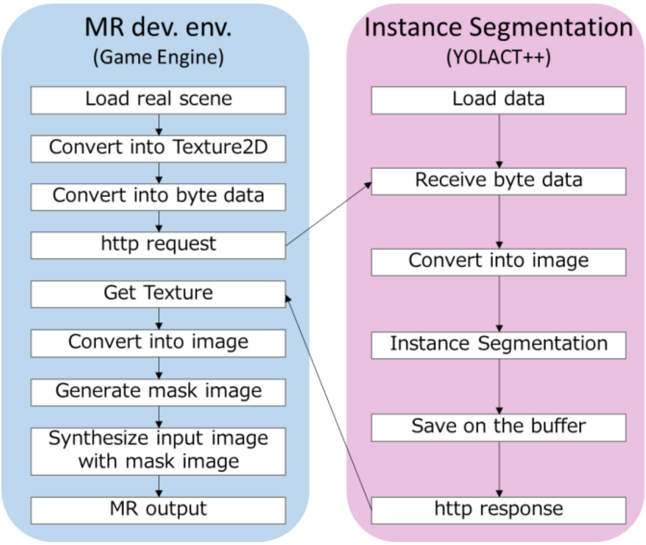


Figure 2. Flowchart of Our Proposed Method

the landscape in real time because of the slow speed and high load. In this system, we used YOLACT++ [7] which is one of the instance segmentation techniques. YOLACT++ has the same high detection accuracy as Mask R-CNN, the processing speed is as fast as 33.5 fps [7] and the load is low. Thus, it enables real-time landscape simulation with MR.

2.3. CONNECTION BETWEEN MR DEV. ENV. AND INSTANCE SEGMENTATION

Real-time instance segmentation processing is very heavy processing and it requires a high-end desktop computer. However, landscape simulations are often conducted outdoors, and the proposed system requires to allow operation on a laptop computer or tablet that is easy to carry around. We divide the system into client PC (laptop computer or tablet) used for MR landscape simulation and server PC (desktop PC) used for instance segmentation. The proposed system is connected instance segmentation with the game engine which is the MR development environment by communication. Real-time video taken by the client PC is sent server PC frame by frame, it is processed instance segmentation by server PC, and instance segmentation image is sent to the server PC frame by frame. We used a Python web application framework known as Flask for communication. With this procedure, we connected instance segmentation and the game engine.

2.4 GENERATE MASK IMAGE

This is because instance segmentation identifies objects of the same class as other objects, there is no fixed color for each class, and the mask color of the instance segmentation image is randomly chosen and displayed. As it is, we cannot generate the correct mask image for the occlusion handling because it is impossible to judge the anteroposterior relationship with the 3D virtual model superimposed by MR in instance segmentation image. Thus, we improved the instance segmentation images so that the before and after the relationship can be judged.

Under the following conditions, the object in front always appears lower in the camera image than the object at the back. In other words, the object with a larger y-coordinate in the camera image is in front of the object in the camera image.

- The ground is almost zero crosswise direction degrees to the camera.
- The object under occlusion is in contact with the ground (not in the air or on top of any other object).
- Camera roll rotation is 0 degrees.

We present a method for generating mask images for landscape MR simulation under these conditions (Figure 3).

1. The number of real objects in front of the 3D virtual model in a landscape MR simulation situation is determined and is the threshold value (manual).
2. Get information of the bounding box (rectangle) of all real objects segmented (automatic).

3. Line up all real objects in order of their y-coordinates at the bottom of the bounding box (automatic).
4. For those arranged in 3, segment the real objects up to the threshold in blue ((R, G, B) = (0, 0, 255)) and apply nothing to the real objects after the threshold (automatic).

Using this method, we succeeded in separating real objects that exist in front of the 3D virtual model from objects that exist behind the 3D virtual model. For the segmentation image in which only the object in front was drawn in blue, the mask image was generated by drawing the blue ((R, G, B) = (0, 0, 255)) pixels in the image in white ((R, G, B) = (255, 255, 255)) and the other pixels in black ((R, G, B) = (0, 0, 0)).

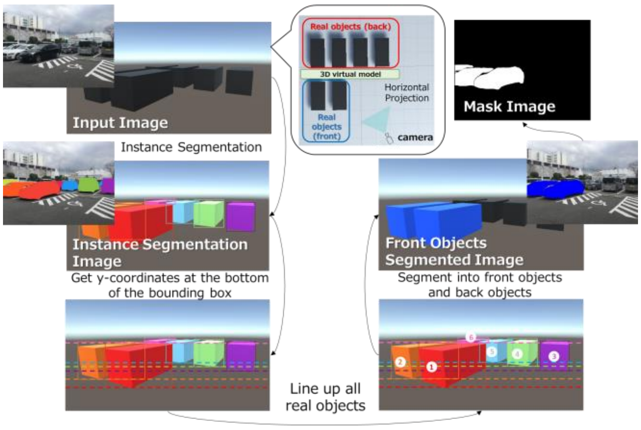

Input Image
Real objects (back)
3D virtual model
Horizontal Projection
Real objects (front)
camera
Mask Image
Instance Segmentation
Instance Segmentation Image
Get y-coordinates at the bottom of the bounding box
Front Objects Segmented Image
Segment into front objects and back objects
Line up all real objects

Figure 3. Method of Generating Mask Image

2.5 MR LANDSCAPE SIMULATION WITH OCCLUSION

An occlusion image is synthesized by the mask image and 3D virtual model. MR landscape simulation is performed by overlapping the occlusion image and the input image and taking them with an MR camera (Figure 4).

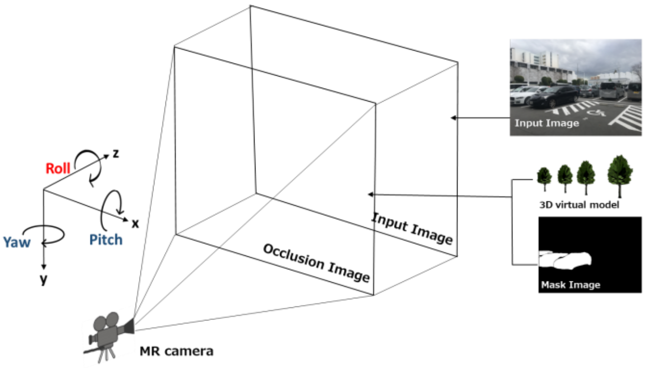
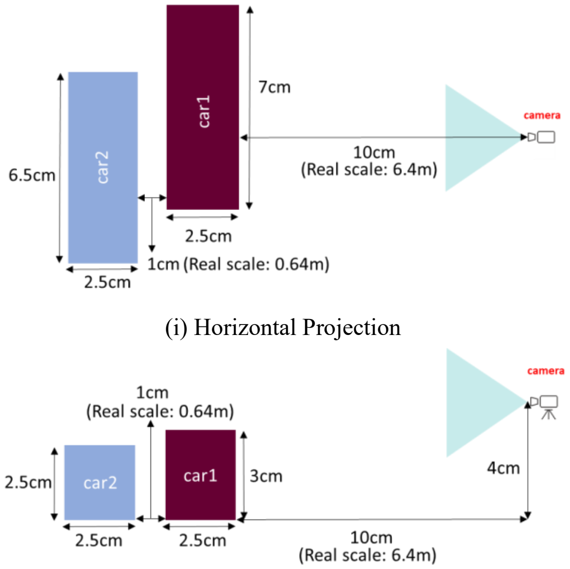


Figure 4. Occlusion Handling

3. Experiment and Results

3.1 EXPERIMENT

In this system, it is important to judge the overlapping objects as different objects. For this purpose, we used a 1/64th scale toy cars to verify the situation in Figure 5 by changing the area of overlap between the two toy cars. We detected an object in the foreground (car1) and an object in the background (car2) and investigated the relationship between the area where the two car toys overlap and IoU (Intersection over Union). The tests were conducted in situations where the two car toys were facing each other and facing the same direction. The results are shown in Figure 6. It was found that the larger the area of overlap, the lower IoU.



(i) Horizontal Projection

(ii) Elevation

Figure 5. Experimental Situation

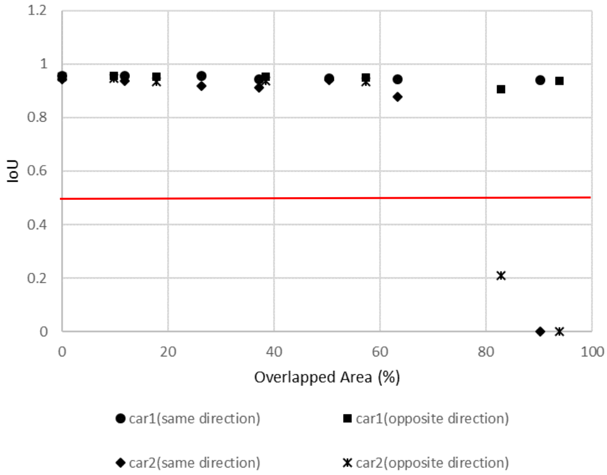


Figure 6. the Relationship Between the Overlapped Area and IoU

3.2 RESULTS

The current system is required to operate in LAN environment. Therefore, we performed the landscape simulation in MR with occlusion of a parking lot with trees in a parking lot where many same class objects (car) are superimposed on each other, using a model that looks like a parking lot. The results of generating mask image are shown in Figure 7 and the results of landscape simulation in MR with occlusion are shown in Figure 8. The processing speed of the whole system was 6-7 fps.

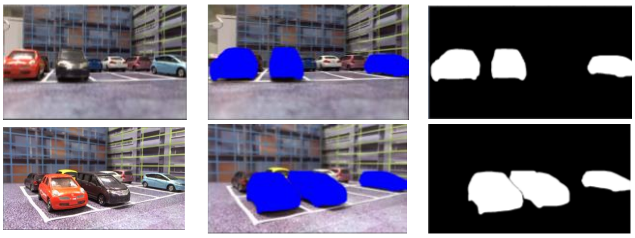

(a) Input Image (b) Front Objects Segmented Image (c) Mask Image

Figure 7. Generate Mask Image

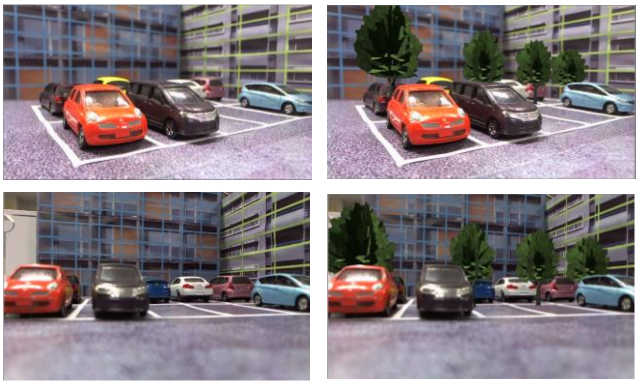

(a) Input Image (b) Output Image

Figure 8. Landscape Simulation in MR with Occlusion

4. Discussion

In the result of the experiment, the detection rate of car1 is satisfactory as IoU = 0.9 or more in all cases, exceeding the reference value of IoU = 0.5 [8]. However, car2 was found to have an IoU below the threshold in cases where there was more than 80% overlap. These findings indicate that if we can see the whole object, this system can recognize the object behind it as a separate object, no matter how much it overlaps. Therefore, when only car1 is in the front of the 3D virtual model, we can simulate correctly, but when both car1 and car2 are in the front of the 3D virtual model and the overlapping area is 80% or more, we cannot simulate correctly. In the simulation, it is not necessary to pay attention to the visibility of real objects that are behind the 3D virtual model, but it should be noted that real objects in front of the 3D virtual model are visible at least 20% of the area.

In addition, although the processing speed of YOLACT++ is 33.5 fps, the processing speed of the whole system is 6-7 fps, which is largely due to the Flask used for server PC to client PC communication, and it is necessary to minimize the data sent and received to increase the processing speed.

5. Conclusion

The conclusions of the present study are shown below.

- The anteroposterior relationship of the detected objects was recognized and a mask image was generated.
- By changing the instance segmentation method and improving the mask image generation method, we were able to distinguish between the same class objects in real time and simulate corrreclty.

Future works include simulations in more diverse situations, such as urban areas, and the implementation of Internet communications.

Acknowledgment

This research has been partly supported by JSPS KAKENHI Grant Number JP19K12681.

References

1) Haynes, P., Lange, S.H. and Lange, E.: Mobile Augmented Reality for Flood Visualisation. Environmental Modelling & Software, 109: 380-389. doi: 10.1016/j.envsoft.2018.05.012

2) Portalés, C., Lerma, J., L., and Navarro, S.:Augmented reality and photogrammetry: A synergy to visualize physical and virtual city environments, ISPRS Journal of Photogrammetry and Remote Sensing, 65 (1), 134-142.2010.

3) Zhu, J., Pan, Z., Sun, C., and Chen, W.: 2010, Handling occlusions in video-based augmented reality using depth information, COMPUTER ANIMATION AND VIRTUAL WORLDS, 21, 509-521.

4) Kido, D., Fukuda, T. and Yabuki, N.: Development of a Semantic Segmentation System for Dynamic Occlusion Handling in Mixed Reality for Landscape Simulation, Proceedings of the 37th eCAADe and 23rd SIGraDi Conference - 1, 641-648, 2019

5) Nakabayashi, M., Fukuda, T. and Yabuki, N.: An On-site Landscape Simulation Method that Enables One-by-one Dynamic Occlusion in Mixed Reality with Instance Segmentation, AIJ KANTO, Summaries of Technical Papers of Annual Meeting, Information Systems Technology, 225-226, 2020.

6) He, K., Gkioxari, G., Dollar, P. and Girshick, R.: Mask R-CNN, In Proc. International Conference on Computer Vision, 2017.

7) Bolya D, Zhou C, Xioa F, Lee Y: "YOLACT++ Better Real-time Instance Segmentation", arXiv preprint arXiv:1912.06218, 2019.

8) Jabbar, A., Farrawell, L., Fountain, J. and Chalup, S. K.: Training Deep Neural Networks for Detecting Drinking Glasses Using Synthetic Images, Neural Information Processing, Part 2, 354–363, 2017.