

# Interactive Exploration of Augmented Aerial Scenes with Free-Viewpoint Image Generation from Pre-Rendered Images

Fumio Okura\*  
Nara Institute of Science  
and Technology (NAIST)  
JSPS Research Fellow

Masayuki Kanbara\*  
Nara Institute of Science  
and Technology (NAIST)

Naokazu Yokoya\*  
Nara Institute of Science  
and Technology (NAIST)

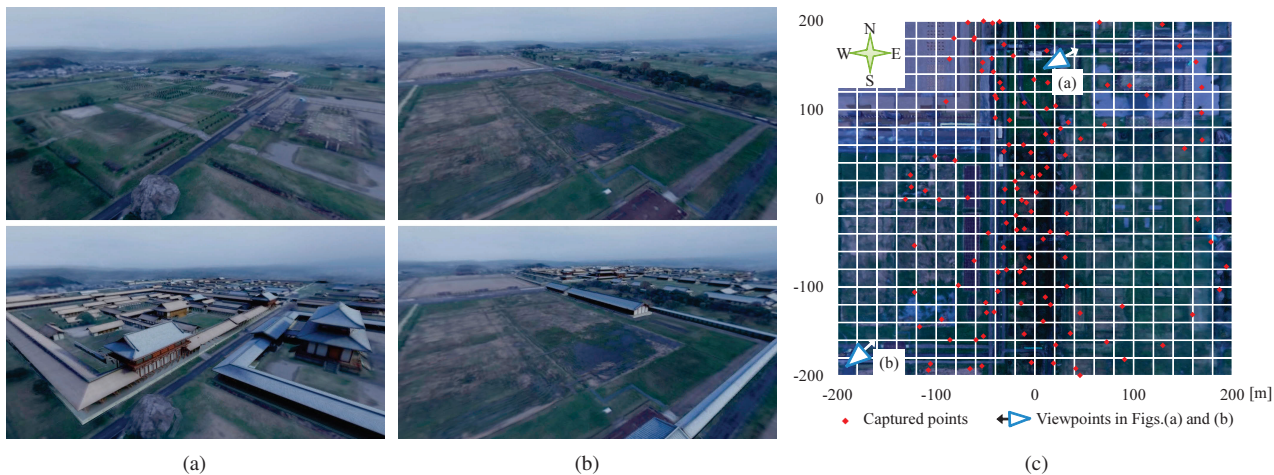


Figure 1: (a), (b): Free-viewpoint images at shown in (c), with/without photorealistic superimposition of virtual objects generated based on proposed framework. (c): Captured points of omnidirectional images and grid viewpoints as well as designated free-viewpoints for (a) and (b).

## 1 INTRODUCTION

Exploration of the real world by flying in its virtual simulation is one of the most important applications of augmented virtuality (AV) due to the popularization of Virtual Globe applications [1] (e.g., Google Earth) as well as large-scale three-dimensional (3D) reconstruction approaches. To improve the appearance of the automatically constructed 3D models, free-viewpoint (sometimes referred to as novel- or arbitrary-viewpoint) image generation using a hybrid of image- and model-based rendering [2] provides promising solutions. In some applications that allow interactive virtual exploration of the real world, virtual objects can be superimposed for visualization of disaster areas or non-existent buildings [5]. In such applications, the virtual objects have to be rendered frame-by-frame. Real-time rendering approaches have been studied for photometric registration of augmented reality (AR) [3] and have become one of the most important research fields in MR. The goal of these research has always been to achieve the same quality as offline rendering using high-quality global illumination (GI), but it has not been achieved thus far.

This study proposes a framework to photorealistically synthesize virtual objects and virtualized real-world. We combine the offline rendering of virtual objects and the free-viewpoint image generation to take advantage of the higher quality of offline rendering without the computational cost of online computer graphics (CG) rendering; i.e., it incurs only the cost of the online computation for the free-viewpoint image generation. In addition, the generation of structured viewpoints (e.g., at every grid point) reduces the com-

putational costs required to online process. The advantages of our framework are:

- High-quality superimposition of virtual objects by combining pre-rendering and free-viewpoint image generation.
- Reducing computational cost in the online free-viewpoint image generation using structured viewpoint generated in offline.

This paper also describes a practical application that provides free-viewpoint on a two-dimensional (2D) plane based on our framework, which superimposes lost buildings onto real scenes of that are virtualized using aerial omnidirectional images captured preliminarily. In this application, environmental maps for image-based lighting (IBL) are completed by removing missing areas in omnidirectional images (e.g., where scenery is occluded by a vehicle mounting the camera), and virtual objects are rendered with GI. The application generates a viewpoint at every grid point and reduces the computational cost of the online process using the structured view and the simplified 3D models.

## 2 PHOTOREALISTIC SYNTHESIS OF FREE-VIEWPOINT IMAGES AND VIRTUAL OBJECTS

The proposed application generates scenes in which virtual objects are photorealistically superimposed, as viewed from a position and a direction that are freely configured on a designated two-dimensional plane, such as in Figure 1. Our framework is divided into two parts: the offline process and the online process. In the offline process, the augmented images are generated at the structured viewpoints (at every grid point in our application) from the real-world images captured at hundreds of position and 3D models of the virtual objects. The online process generates perspective images from freely configured viewpoints in real-time, using pre-generated augmented views at the structured viewpoints and simplified 3D models of both physical and virtual objects.

\*e-mail: {fumio-o,kanbara,yokoya}@is.naist.jp



Figure 2: Aerial omnidirectional image used as input.



Figure 3: Augmented omnidirectional scene at a grid point.

## 2.1 Offline process: Augmented structured viewpoint generation

Omnidirectional images are captured from a number of positions in the sky using an omnidirectional camera and an aerial vehicle [4], and used as input of the offline process. Figure 2 shows an example of an input image.

The flow of the offline process is as follows:

1. 3D reconstruction
  2. Environmental map completion using sky model
  3. Pre-generation of real scene images for structured viewpoints
  4. Pre-rendering of virtual objects for structured viewpoints
- 1) 6 degrees-of-freedom (DoF) camera pose (3 DoF positions and 3 DoF orientations) and dense 3D shapes of the real world are first reconstructed using structure-from-motion and multi-view-stereo. 2) In aerial imaging, parts of the omnidirectional images could be missing as shown in upper half in Figure 2, due to occlusion of the camera or the vehicle, for example. Intensity of these missing areas is estimated using the camera pose and a statistical model of the sky luminance. 3) The structured views (on every grid points in our application) are estimated from input images, using a free-viewpoint image generation which is the hybrid approach of model- and image-based rendering, as known as view-dependent texture mapping (VDTM) [2]. 4) The virtual objects are finally superimposed onto the structured views, as shown in Figure 3. We employ an offline rendering with IBL using the completed environmental maps, similar to the approach in [5].

## 2.2 Online process: Light-weight free-viewpoint image generation

We enabled the free-viewpoint image generation on a 2D plane using the pre-generated viewpoints at the grid points. VDTM-based free-viewpoint image generation was improved to reduce online computation. In the online process, the views at four grid points neighboring to the virtual viewpoint to be generated are projected on 3D surfaces and blended with the bilinear weights, which were calculated from the positions of the grid points and the virtual viewpoint, as illustrated in Figure 4.

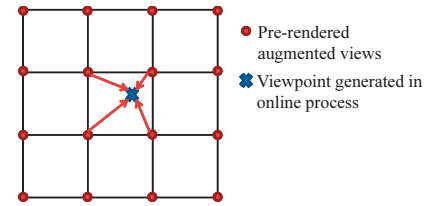


Figure 4: Online free-viewpoint image generation based on a bilinear weighting function which requires  $O(1)$  calculation.

Our online process achieved  $O(1)$  (i.e., four times) calculation for the weight to blend textures; unless the worst case in regular VDTM approach requires  $O(n * m)$  with  $n$  cameras and  $m$  polygons. In addition, pre-generation of dense, structured viewpoints enables to simplify the 3D models of both real and virtual objects. In our application, they were combined and simplified down to 1% of the original numbers of polygons. Although the simplification of 3D models does not reduce the cost of calculating the weight, the rendering cost for the 3D models is reduced in proportion to the number of polygons. Note that, this process can be easily implemented on the GPU, such as GLSL.

## 3 EXPERIMENTS

To confirm that the application based on the proposed framework generates the appropriate scenes in a practical environment, augmented free-viewpoint scenes were generated on a  $400\text{m} \times 400\text{m}$  plane at an altitude of approximately 50 m from the ground. The augmented free-viewpoint scenes generated by the online process have been shown in Figure 1. The setting for the application was based on Heijo-kyo, which is an ancient capital in Japan, and its old palaces no longer exist in the palace site. The grid points, which were used for the positions of pre-generated viewpoints, were designated at every  $20\text{m} \times 20\text{m}$ , as shown in Figure 1(c). The input omnidirectional images were captured at various altitudes, and the altitude of the grid plane was designated as the average altitude of the captured points.

The offline process performed at approximately 1 fps using the GPU implementation on an NVIDIA GeForce 690 (2 GB texture memory) with 174 omnidirectional images and 3,290,880 reconstructed real-world 3D polygons. In the online process was implemented on the same GPU used in the offline process, using 3D models of both real and virtual worlds were combined and simplified to a model of 104,054 polygons. It performed at approximately 400 fps. The online process generated reasonable results with high performance.

**Acknowledgements** This research was supported by JSPS Grants-in-Aid for Scientific Research Nos. 24700208, 23240024, Grant-in-Aid for JSPS Fellows No. 25-7448, and the “Ambient Intelligence” project granted by MEXT. The CG models of the Heijo palace are courtesy of Toppan Printing Co., Ltd.

## REFERENCES

- [1] D. Butler. Virtual globes: The web-wide world. *Nature*, 439(7078):776–778, 2006.
- [2] P. Debevec, C. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. In *Proc. ACM SIGGRAPH’96*, pages 11–20, 1996.
- [3] P. Lensing and W. Broll. Instant indirect illumination for dynamic mixed reality scenes. In *Proc. IEEE ISMAR’12*, pages 109–118, 2012.
- [4] F. Okura, M. Kanbara, and N. Yokoya. Augmented telepresence using autopilot airship and omni-directional camera. In *Proc. IEEE ISMAR’10*, pages 259–260, 2010.
- [5] F. Okura, M. Kanbara, and N. Yokoya. Fly-through Heijo palace site: Augmented telepresence using aerial omnidirectional videos. In *Proc. ACM SIGGRAPH’11 Posters*, page 78, 2011.