

# Automatic Generation of Dynamically Relightable Virtual Objects with Consumer-Grade Depth Cameras

Chih-Fan Chen\*

University of Southern California

Evan Suma Rosenberg†

University of Minnesota

## ABSTRACT

This research demo showcases the results of novel approach for estimating the illumination and reflectance properties of virtual objects captured using consumer-grade RGB-D cameras. This method is implemented within a fully automatic content creation pipeline that generates photorealistic objects in real-time virtual reality scenes with dynamic lighting. The geometry of the target object is first reconstructed from depth images captured using a handheld camera. To get nearly drift-free texture maps of the virtual object, a set of selected images from the original color stream is used for camera pose optimization. Our approach further separates these images into diffuse (view-independent) and specular (view-dependent) components using low-rank decomposition. The lighting conditions during capture and reflectance properties of the virtual object are subsequently estimated from the specular maps. By combining these parameters with the diffuse texture, reconstructed objects are then rendered in a real-time virtual reality demo that plausibly replicates the real world illumination and showcases dynamic lighting with varying direction, intensity, and color.

**Index Terms:** Computing methodologies—Computer graphics—Graphics systems and interfaces—Virtual reality; Computing methodologies—Computer graphics—Rendering—Reflectance modeling

## 1 INTRODUCTION

With the recent proliferation of consumer head-mounted displays, the research topics in virtual and mixed reality have been developing vigorously. In many applications, customized virtual content are one of the essential elements for building virtual worlds. Using ready-made virtual objects can potentially compromise the desire of developers or decrease the interests of users. However, manual creation of high-fidelity virtual content requires expert knowledge of 3D modeling, and is an often expensive and time consuming endeavor. A promising alternative is scanning physical objects and replicating their photorealistic appearance in the virtual environment. With the development of consumer-grade RGB-D sensors such as the Microsoft Kinect or Intel Realsense, grab-and-go 3D scanning is becoming increasingly accessible for general users. However, replicating the photorealistic appearance of reconstructed virtual objects from an RGB-D sequence is still an open question. Common methods typically average the colors from all captured images to compute the color of each vertex, which often results in a lower fidelity appearance and texture maps with baked lighting. However, for non-Lambertian objects, this approach is not ideal for real-time virtual reality applications. Ideally, surface illumination (e.g, specular reflections) should dynamically change based on the user’s viewpoint, and the absence of these dynamic visual cues can be especially noticeable in head-tracked virtual reality.

\*e-mail: chihfanc@usc.edu

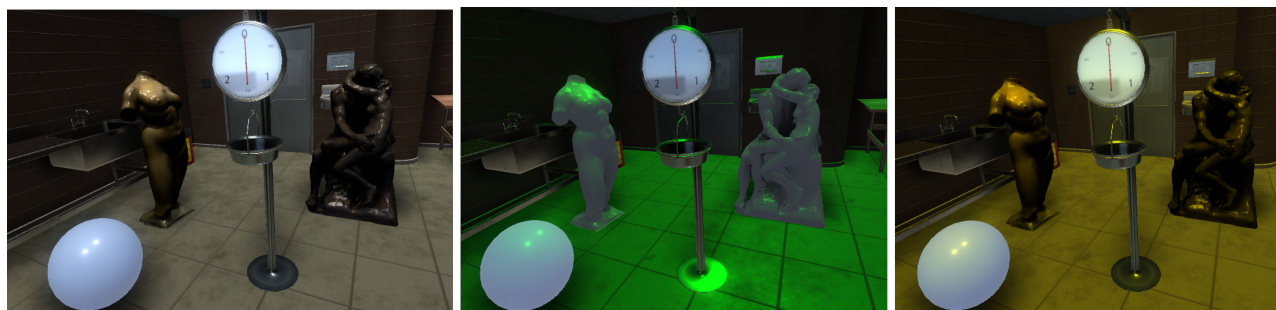
†e-mail: suma@umn.edu



Figure 1: Comparison of the rendered virtual object (top row) and original capture frames (bottom row). The highlights on the sphere visualize the direction and color of the virtual lights that were computed to replicate the illumination in the real world scene.

To overcome these limitations, View-Dependent Texture Mapping (VDTM) techniques have been introduced with promising results [1, 3, 4]. In this approach, the object’s appearance is dynamically rendered using a subset of captured images closest to the current virtual camera position. These methods can result in photo-realistic quality because the surface illumination of the virtual model will change based on the user’s viewpoint. However, the synthesized specular reflections are still problematic since the reflectance of an object is strongly related to its geometry, surface material properties, and environmental lighting conditions. The synthetic views created from interpolating between color images may exhibit inconsistencies and visual artifacts, especially for capture data with sparse camera trajectories. Moreover, the reconstructed objects are incompatible with virtual environments that include dynamic lighting.

In this research demo, we showcase the results of a novel capture-to-rendering content creation pipeline that estimates diffuse and specular reflectance of the object and lighting conditions during capture. Given an RGB-D sequence, the geometry of the object is first reconstructed and several color frames are selected from the original stream. We utilize a novel low-rank decomposition method that simultaneously optimizes the camera poses of each frame and separates each color frame into a diffuse map and a specular map. Using the specular maps from many different viewing directions, the surface material properties and the lighting conditions can be derived. At run-time, the optimized diffuse textures and specular reflectance parameters can then be used to synthesize the captured object’s appearance from an arbitrary viewpoint. Our method can generate plausible results even for unseen views that were not present in the capture dataset. Moreover, the reconstructed virtual content can be readily integrated with virtual scenes and dynamically relit with different lights of varying direction, color, and intensity. The fully automatic pipeline does not require expert knowledge or manual human effort, and the reconstructed models are suitable for integration with industry-standard virtual environments.



(a) Rendering with different light direction and color



(b) Rendering with different light intensities

Figure 2: Demonstration of two reconstructed virtual objects (The Kiss and Torso of Elevation) in a virtual scene with dynamic illumination. The proposed reflectance estimation method can provide plausible results with varying virtual light direction, color, and intensity. Furthermore, the specular highlights on the virtual object will smoothly change in real-time as the user moves between different viewpoints. A virtual sphere is also displayed to more easily visualize the color and the direction of virtual lights.

## 2 REAL-TIME DEMO

Users will experience the live demo of dynamically relightable scanned virtual objects using an Acer Windows Mixed Reality Headset with inside-out-tracking. Because the reconstruction and optimization stages of the proposed content creation pipeline take several hours to compute on a laptop, our live demo showcases the real-time rendering results of objects captured offline. The virtual objects shown in our demo were reconstructed from a public dataset that contains thousands of RGB-D sequences captured by non-experts using a Primesense camera [2]. We selected three objects that exhibit strong view-dependent specular reflections: Torso of Elevation, the Kiss by Rodin, and an antique leather chair (corresponding to IDs 3887, 4252, and 5989 in the database, respectively). The virtual environment was developed in Unity 2017.2.0f3. Our method can render the models in 10-15 milliseconds (i.e., 70-90 fps) on a MacBook Pro with an Intel i7-4850HQ CPU, Nvidia GeForce GT750M GPU, and 16 GB of RAM.

The virtual reality experience is designed for a 3mx3m physical space. Although the virtual objects can be observed while seated, the view-dependent illumination effects are more compelling to observe while physically walking between multiple viewpoints. The demonstration has several visualization modes that can be displayed to the user. First, users can observe the original capture video as a texture billboard in the virtual reality scene. For comparison, the reconstructed virtual object will be displayed with synthetic lights that were computed to replicate the real world illumination (see Figure 1). In the second mode, users will be able to freely explore a virtual environment with several reconstructed objects placed in an existing scene. Furthermore, they will be able to directly manipulate the intensity, direction, and color of the virtual lights using the handheld controller (see Figure 2). In both of these modes, users will also be able to toggle between the proposed reflectance

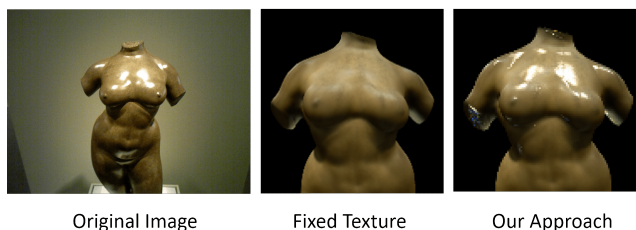


Figure 3: Comparison of our method with fixed textures. Note that the fixed texture results in lower fidelity (blurriness) due to averaging the observed images.

rendering approach and a fixed texture map generated by averaging color images in the capture sequence (see Figure 3).

## REFERENCES

- [1] C.-F. Chen and E. Suma Rosenberg. Dynamic omnidirectional texture synthesis for photorealistic virtual content creation. In *Adjunct Proceedings of the IEEE International Symposium for Mixed and Augmented Reality*, 2018.
- [2] S. Choi, Q.-Y. Zhou, S. Miller, and V. Koltun. A large dataset of object scans. *arXiv:1602.02481*, 2016.
- [3] P. Debevec, Y. Yu, and G. Boshokov. Efficient view-dependent image-based rendering with projective texture-mapping. Technical report, University of California at Berkeley, Berkeley, CA, USA, 1998.
- [4] Y. Nakashima, Y. Uno, N. Kawai, T. Sato, and N. Yokoya. Ar image generation using view-dependent geometry modification and texture mapping. *Virtual Reality*, 19(2):83–94, 2015.