Virtual Content Creation
Using Dynamic Omnidirectional Texture Synthesis

Chih-Fan Chen* Evan Suma Rosenberg*

Institute for Creative Technologies, University of Southern California

ABSTRACT

We present a dynamic omnidirectional texture synthesis (DOTS) approach for generating real-time virtual reality content captured using a consumer-grade RGB-D camera. Compared to a single fixed-viewpoint color map, view-dependent texture mapping (VDTM) techniques can reproduce finer detail and replicate dynamic lighting effects that become especially noticeable with head tracking in virtual reality. However, VDTM is very sensitive to errors such as missing data or inaccurate camera pose estimation, both of which are commonplace for objects captured using consumer-grade RGB-D cameras. To overcome these limitations, our proposed optimization can synthesize a high resolution view-dependent texture map for any virtual camera location. Synthetic textures are generated by uniformly sampling a spherical virtual camera set surrounding the virtual object, thereby enabling efficient real-time rendering for all potential viewing directions.

Keywords: virtual reality, view-dependent texture mapping, content creation.

Index Terms: Computing methodologies—Computer Graphics—Graphics systems and interfaces—Virtual reality; Computing methodologies—Computing graphics—Image manipulation—Texturing; Computing methodologies—Computer Graphics—Image manipulation—Image-based rendering

1 INTRODUCTION

Creating photorealistic virtual reality content has gained more importance with the recent proliferation of head-mounted displays (HMDs). However, manually modeling high-fidelity virtual objects is not only difficult but also time consuming. An alternative is to scan objects in the real world and render their digitized counterparts in the virtual world. Reconstructing 3D geometry using consumer-grade RGB-D cameras has been an extensive research topic, and many techniques have been developed with promising results. However, replicating the appearance of reconstructed objects is still an open question. Existing methods (e.g., [4]) compute the color of each vertex by averaging the colors from all captured images. Blending colors in this manner results in lower fidelity textures that appear blurry especially for objects with non-Lambertian surfaces. Furthermore, this approach also yields textures with fixed lighting that is baked onto the model. These limitations become especially noticeable when viewed in head-tracked virtual reality displays, as the surface illumination (e.g. specular reflections) does not change appearance based on the user's physical movements.

To improve color fidelity, techniques such as View-Dependent Texture Mapping (VDTM) have been introduced [1]. In this approach, the texture is dynamically updated in real-time using a subset of images closest to the current virtual camera position. Although these methods typically result in improved visual quality, the dynamic transition between viewpoints is potentially problematic, especially for objects captured using consumer RGB-D cameras. This is due to the fact that the input sequences often cover only a limited range of viewing directions, and some frames may only partially capture the target the object. In this paper, we propose dynamic omnidirection texture synthesis (DOTS) in order to improve the smoothness of viewpoint transitions while maintaining the visual quality provided by VDTM techniques. Given a target virtual camera pose, DOTS is able to synthesize a high-resolution texture map from the input stream of color images. Furthermore, instead of using traditional spatial/temporal selection, DOTS uniformly samples a spherical set of virtual camera poses surrounding the reconstructed object. This results in a well-structured triangulation of synthetic texture maps that provide omnidirectional coverage of the virtual object, thereby leading to improved visual quality and smoother transitions between viewpoints.

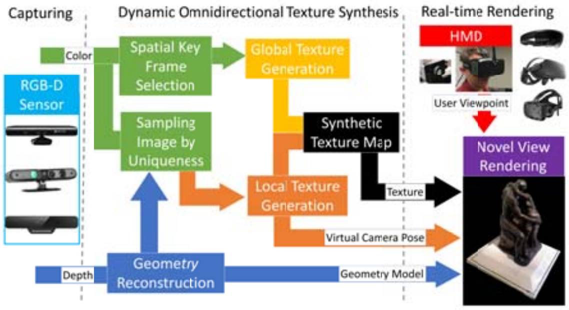


Figure 1: Overview of the DOTS content creation pipeline. Color and depth image streams are captured from a RGB-D camera. The geometry is reconstructed from depth information and is used to uniformly sample a set of virtual camera poses surrounding the object. For each camera pose, a synthetic texture map is blended from the global and local texture images captured near the camera pose. The synthetic texture maps are then used to dynamically render the object in real-time based on the user's current viewpoint in virtual reality.

2 OVERVIEW

Overall Process The system pipeline is shown in Figure 1. Given an RGB-D video sequence, the geometry is first reconstructed from the original depth stream. A set of key frames are selected from the entire color stream, and a global texture is be generated from those key frames. Next, a virtual sphere is defined to cover the entire 3D model and the virtual camera poses are uniformly sampled and triangulated on the sphere's surface. For each virtual camera pose, the corresponding texture is synthesized from several frames and the pre-generated global texture maps. At run-time, the user viewpoints provided by a head-tracked virtual reality display is used for selecting the synthetic maps to render the model in real-time.

Geometric Reconstruction and Global Texture We use Kinect Fusion [3] to construct the 3D model from the depth sequences. Using all color images I of the input video for generating

---

*e-mail: {cfchen, suma}@ict.usc.edu

2018 IEEE Conference on Virtual Reality and 3D User Interfaces
18-22 March, Reutlingen, Germany
978-1-5386-3365-6/18/$31.00 ©2018 IEEE

the global texture is inefficient. In our DOTS framework, we chose a set of key frames that maximize the variation of viewing angles of the 3D model. The color mapping optimization [4] is used to find the optimized camera poses of each key frame.

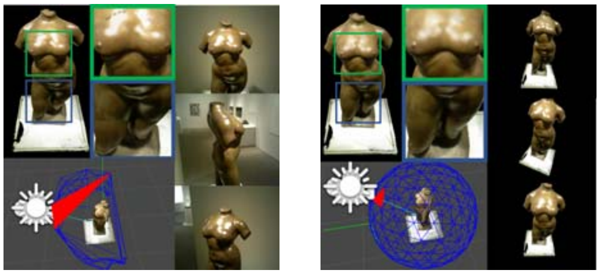
Synthetic Texture Map Generation and Rendering Our objective is to replicate high-fidelity models with smooth transitions between viewpoints. Instead of selecting images directly from the original video, we uniformly sample virtual cameras surrounding the reconstructed geometry in 3D and set the size of the sphere large enough to cover the entire object in each synthesized virtual view. To generate the synthetic view s_i of each virtual camera, all frames I are weighted and sorted based on their uniqueness with respect to the virtual camera pose t_{s_i}. Using only the local texture is insufficient for the entire model because the closest images might have only a partial view of the object (e.g., the three selected frames shown in 2(a) right). Thus, we introduce a weighted global texture to our objective function, but kept their camera poses unchanged since it is already optimized in the previous section. The optimized texture maps s_j not only maximize the color agreement of locally selected images L but also seamlessly blends the global texture G (e.g., the three synthetic images in 2(b) right). It is worth noting that the resolution of synthetic texture can be set to any arbitrary positive number. We set the resolution as 2048×2048, which is preferred by most game engines without compression.

Real-time Image-based Rendering At run-time, the HMD pose is provided by the Oculus Rift CV1 and two external Oculus Sensors. The traditional VDTM method computes the euclidean distance between the users head position and all camera poses and then selects the closest images for rendering. In contrast, DOTS systematically samples all spherical virtual camera poses surrounding the virtual object. Thus, the vector from the model center to the HMD position would only intersects with one triangle mesh. Barycentric coordinates are used to compute the weight and blend the color from the three synthesized textures of the intersected triangle to generate the novel view.

3 EXPERIMENTAL RESULTS

We tested our system with different models from a public dataset [2], which the RGB-D sequences are captured using a Primesense RGBD camera. In Figure 2, we present a visual comparison of DOTS and VDTM [1]. For both methods, three images from the highlighted triangles are used to render the model. Because the selected virtual cameras for DOTS have similar viewing directions, the viewpoint generated by blending the three corresponding synthetic textures exhibits fewer artifacts compared to VDTM, which uses the three closest images from original capture sequence. Moreover, the synthetic texture maps provide omnidirectional coverage of the virtual object, while the selected key frames in VDTM sometimes only partially capture the target object. Because of this missing texture information, VDTM can only synthesize a partial unobserved view, thereby resulting in sharp texture discontinuities (i.e. seams), while DOTS can render the model without such artifacts.

As shown in Figure 3, the fixed texture method results in a lower fidelity (blurry) appearance. Although the model rendered using VDTM can achieve a more photorealistic visual apperance, the region of high-fidelity viewpoints is limited. Texture defects and unnatural changes between viewpoints are not pleasant during virtual reality experiences where user freedom is encouraged and the view direction cannot be predicted in advance. In contrast to VDTM, DOTS generates omnidirectional synthetic texture maps that produce visually reasonable results even under such conditions.



(a) VDTM (b) DOTS

Figure 2: (a) The key frames selected by VDTM are not uniformly distributed around the 3D model because they are dependent upon the camera trajectory during object capture. Thus, this leads to an irregular triangulation (red) and undesirable visual artifacts. (b) In contrast, the synthetic maps generated by DOTS cover all potential viewing directions and the triangulation is uniform, resulting in seamless view-dependent textures.

Figure 3: The left column shows three example images from original video. The middle column are the geometry model and the results of the fixed texture [4]. The right column are the results of VDTM [1] and the results of DOTS).

ACKNOWLEDGMENTS

This work is sponsored by the U.S. Army Research Laboratory (ARL) under contract number W911NF-14-D-0005. Statements and opinions expressed and content included do not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

REFERENCES

[1] C. Chen, M. Bolas, and E. S. Rosenberg. View-dependent virtual reality content from RGB-D images. In *IEEE International Conference on Image Processing*, pp. 2931–2935, 2017.

[2] S. Choi, Q.-Y. Zhou, S. Miller, and V. Koltun. A large dataset of object scans. *arXiv:1602.02481*, 2016.

[3] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*, pp. 559–568. ACM, New York, NY, USA, 2011. doi: 10.1145/2047196.2047270

[4] Q.-Y. Zhou and V. Koltun. Color map optimization for 3d reconstruction with consumer depth cameras. *ACM Trans. Graph.*, 33(4):155:1–155:10, July 2014. doi: 10.1145/2601097.2601134